ECMWF COPERNICUS REPORT

Copernicus Climate Change Service

Product Quality Assessment Report

Issued by: Met Norway / Jan Griesfeller and DLR / Thomas Popp

Date: 28/05/2021

Ref: C3S_312b_Lot2.2.2.2-v2.1_202102_PQAR_v2.4

Official reference number service contract: 2018/C3S_312b_Lot2_DLR/SC1

Contributors

MET NORWAY
Jan Griesfeller
Michael Schulz

DLR
Thomas Popp

History of modifications

Version	Date	Description of modification	Chapters / Sections
Contract C3S_312a_Lot5 (2016)			
1	11.01.2018	First version	
1.1	12.01.2018	Internal review	
1.2	17.01.02018	Corrections: Section 1.2 IASI: use daytime IASI data to compare with AERONET Section 1.3 POLDER: SSA map adjust colour bar (fig. 2.13)	
1.3	29.03.2018	Add annex (excerpt of Aerosol_cci2 PVIR) Add compliance check against TRD	Annex New section 4
1.4	20.06.2018	Switched to updated AERONET dataset; ATSR: Name change to C3S standard for ADV; added stability analysis IASI: added stability analysis, added global analysis. SLSTR: added analysis of first 6 month total AOD datasets GRASP: version upgrade to 2.01	
2	25.06.2018	Review by PI; minor layout corrections; update of compliance table	
3	26.09.2018	Updated AATSR ADV retrieval to v3.11; updated IASI MAPIR retrieval to v4.1; rewritten chapter 2.2 (IASI assessment); added assessment of	

		GRASP aerosol layer height product; added MAPIR v4.1 to the stability assessment of the IASI retrievals	
Contract C3S_312b_Lot2 (2018)			
1.	21.05.2019	Changed IASI assessment year to year 2018. Changed dust AOD reference data set to AERONET SDA version 3 level 2.0; SLSTR assessment covers now the entire year 2018 and the SLSTR ensemble; Added OLCI_XBAER assessment for the year 2018	
1.1	17.07.2019	Corrections as requested by ASSIMILA (check / align / comment differences in algorithm versions)	
1.2	31.07.2019	Clarify references to “contract” with concrete CDR and ICDR processings and their scheduled date Clarify reference to PQAD	Scope, executive summary, summary and outlook Start of section 1
2.0	25.05.2020	Assessed SLSTR SDV v2.00 retrieval assessed SLSTR ensemble retrieval. Assessed SLSTR year 2019 data (all algorithms) Aeronet reference data sets: updated to more recent version (end of April 2020)	Title page Executive summary Sections 1.5.4, 1.6, 2.2.1, 2.4 and 2.5

		use level 2.0 instead of level 1.5 datasets for the years 2018 and 2019 everywhere. Added IASI year 2019 assessment (all algorithms) assessed OLCI S4O retrieval (2019) Extended assessment of OLCI XBAER retrieval to the year 2019	
2.1	26.05.2020	Review / minor updates by PI, including updated requirements table from latest TRD-GAD in section 3	All as for v2.0 Title page Section 3
2.2	03.06.2020	Addition of SLSTR FM-AOD validation after correction of a bug	Section 2.4.2
2.2a	29.06.2020	Minor corrections as suggested by ASSIMILA	Executive summary Several other minor corrections
2.3		Assessed new versions of ATSR2/AATSR retrievals; assessed new version of GRASP retrieval; assessed new DLR IASI retrieval; updated IASI stability analysis with longer time frame and usage of AERONET version 3 data; use Aeronet version 3 also for ATSR2/AATSR stability analysis	
2.4	31.05.2021	Added ensemble datasets to the analysis; assessed GOMOS version 5.00	

Related documents

Reference ID	Document
D1	Lot2 Product Quality Assurance Document for Aerosol products Version 1.2 (PQAD, C3S_312b_Lot2.2.2.1-v1.2)
D2	Kinne Stefan, Pavel Lytvynov, Jan Griesfeller, Michael Schulz, Kerstin Stebel, Christoph Bruehl, Yong Xue, Thomas Popp, Aerosol_cci2 Product Validation and Intercomparison Report (PVIR), version 3.4, 18.08.2017
D3	Lot2 Target Requirement and Gap Analysis Document for Aerosol products 2019 Version 2.11 (TRD-GAD, C3S_312b_Lot2.1.0-2019(AER))
D4	de Leeuw, G., T. Holzer-Popp, S. Bevan, W. Davies, J. Descloitres, R.G. Grainger, J. Griesfeller, A. Heckel, S. Kinne, L. Klüser, P. Kolmonen, P. Litvinov, D. Martynenko, P.J.R. North, B. Ovigneur, N. Pascal, C. Poulsen, D. Ramon, M. Schulz, R.Siddans, L. Sogacheva, D. Tanré, G.E. Thomas, T.H. Virtanen, W. von Hoyningen Huene, M.Vountas, S. Pinnock, Evaluation of seven European aerosol optical depth retrieval algorithms for climate analysis, Remote Sensing of Environment, 162, 295-315, doi: 10.1016/j.rse.04.023, 2015

Acronyms

Acronym	Definition
AAOD	(Absorption Aerosol Optical Depth) is the vertically normalized atmospheric column integrated aerosol absorption at a certain wavelength (usually at 550 nm, the reference wavelength in global modelling) [note, AAOD = AOD*(1-SSA)]
AeroCom	is an open science initiative founded to inter-compare aerosol modules in global modelling and evaluate overall model performance as well as the treatment of specific aerosol processes against available (and trusted) observations.
AERONET	represents a federated network of globally distributed ground-based CIMEL sun-/sky-photometers, which is maintained (calibration facility, data processing and aerosol and water vapor products access) by NASA (National Aeronautics and Space Administration) and PHOTONS (PHOtométrie pour le Traitement Opérationnel de Normalisation Satellitaire)
AOD	(Aerosol Optical Depth) is the vertically normalized atmospheric column integrated aerosol extinction at a certain wavelength or waveband

	(usually at 550nm, the reference wavelength in modelling). AOD is also often referred to as Aerosol Optical Thickness (AOT).
FM AOD	(Fine-mode Aerosol Optical Depth) is the vertically normalized atmospheric column integrated aerosol extinction at a certain wavelength or waveband (usually at 550nm) of aerosol particles smaller than 0.5um in radius (or smaller 1um in diameter).
CM AOD	(Coarse-mode Aerosol Optical Depth) is the vertically normalized atmospheric column integrated aerosol extinction at a certain wavelength or waveband (usually at 550nm) of aerosol particles larger than 0.5um in radius (or larger 1um in diameter).
ATSR	(Along Track Scanning Radiometer) was a multi-channel imaging radiometer (with dual view capabilities in the visible and near-IR solar spectrum). Two versions are used for aerosol retrieval: ATSR-2 on board of the European Space Agency's ERS-2 satellite (1995-2002) and the advanced ATSR (AATSR) on ESA's ENVISAT satellite (2002-2012).
CALIOP	(Cloud-Aerosol Lidar with Orthogonal Polarization) is a two-wavelength polarization-sensitive backscatter LIDAR that provides high-resolution vertical profiles of aerosols and clouds onboard NASA's CALIPSO satellite.
CF	(Climate and Forecast) naming convention metadata are designed to promote the processing and sharing of files created with the NetCDF API.
CMUG	(Climate Model User Group) is a part of ESA's Climate Change Initiative (CCI) and is composed of members of major climate research institutes in Europe. The group is tasked to oversee the usefulness of new climate data records produced for CCI selected ECVs.
EARLINET	European Aerosol Research Lidar Network. Network of research LIDARs in Europe established in 2000 as a research project with the goal of creating a quantitative, comprehensive, and statistically significant database for the horizontal, vertical, and temporal distribution of aerosols on a continental scale.
ECV	(Essential Climate Variables) are geo-physical quantities of the Earth-Atmosphere-System that are technically and economically feasible for systematic (climate) observations.
ENVISAT	("Environmental Satellite") is a now inoperative ESA polar-orbiting (ca 10am local overpass) satellite, which supplied between 2002 and 2012 atmospheric data, including for aerosol remote sensing relevant AATSR, MERIS and GOMOS sensor data.
ESA	(European Space Agency), is an international space agency based in Paris.
FCDR	(Fundamental Climate Data Records or simply **CDR**) represent long-term records of measurements or retrieved physical quantities from remote sensing. FCDRs require consistency across multiple platforms with respect to (1) calibration, (2) algorithms, (3) spatial and temporal resolution, (4) quantification of errors and biases and (5) data format. FCDRs also need to manifest applied ancillary data.

FMF	(Fine Mode Fraction) is the fraction of the total AOD which is contributed by aerosol particles smaller than 1m in diameter. Due to their smaller size these aerosol particles are referred to as fine-mode aerosol, in contrast to (larger or) coarse mode aerosol particles.
GCOS	(Global Climate Observing System), located at WMO in Geneve, is intended to be a long-term, user-driven operational system capable of providing the comprehensive observations required for (1) monitoring the climate system, (2) detecting and attributing climate change, (3) assessing impacts of, and supporting adaptation to, climate variability and change, (4) application to national economic development and (5) research to improve understanding, modelling and prediction of the climate system
GRASP	(Generalized Retrieval of Aerosol and Surface Properties) is an aerosol retrieval algorithm that processes properties of aerosol- and land-surface-reflectance. It infers nearly 50 aerosol and surface parameters including particle size distribution, the spectral index of refraction, the degree of sphericity and absorption.
GOMOS	(Global Ozone Monitoring by Occultation of Stars) is an instrument on board the European satellite ENVISAT. The main scientific objective of this stellar occultation instrument is to monitor ozone and ozone trends as function of altitude in the upper atmosphere (stratosphere, mesosphere). GOMOS also measures atmospheric parameters related to (stratospheric ozone) chemistry like NO2, NO3, H2O and aerosol as well as ozone dynamics like temperature, air density and turbulence.
IASI	(Infrared Atmospheric Sounding Interferometer) on European MetOp platforms senses the thermal heat emission from the Earth (with a Michelson interferometer) mainly to provide atmospheric temperature and humidity profiles.
ICAP	(International Cooperative for Aerosol Prediction) is an international forum for aerosol forecast centers, remote sensing data providers, and lead systems developers to share best practices and discuss pressing issues facing the operational aerosol community.
MACC	(Monitoring Atmospheric Composition and Climate) were EU-funded projects for the development of a chemical weather forecast service. Now in its operational phase, **Copernicus** predicts global distributions and long-range transports of greenhouse gases (carbon dioxide, methane), of aerosols that result from both natural processes and human activities and of reactive gases (tropospheric ozone, nitrogen dioxide). **Copernicus** also evaluates how these constituents influence climate and estimates their sources and sinks.
MAN	(Marine Aerosol Network) is the ocean branch of the AERONET network, based on handheld solar attenuation measurements with calibrated MICROTOPS-II sun-photometers.
MERIS	(Medium Resolution Imaging Spectrometer) was a solar spectral satellite sensor on ESA's ENIVISAT platform.

MISR	(Multi-angle Imaging Spectro-Radiometer) is a multi-spectral sensor on NASA's EOS Terra platform with (9) multi-directional view capabilities.
MODIS	(Moderate Resolution Imaging Spectro-Radiometer) is a multi-spectral sensor on NASA's EOS Terra and Aqua platforms.
NASA	(National Aeronautics and Space Administration), is an independent agency of the executive branch of the United States federal government responsible for the civilian space program as well as aerocnautics and aerospace resesarch.
OMI	(Ozone Monitoring Instrument) is a UV multi-spectral sensor on NASA's EOS Aura platform.
POLDER	(POLarization and Directionality of the Earth's Reflectances) is a passive optical imaging radiometer and polarimeter for studies on radiative and microphysical properties of clouds and aerosols on the French CNES PARASOL (Polarization and Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar).
SEAWIFS	(Sea-viewing Wide Field-of-view Sensor) was the sensor on the US GeoEye-Satellits Orb-View-2 (SeaStar).
SCIAMACHY	(Scanning Imaging Absorption Spectrometer for Atmospheric ChartographY) was a high spectral resolution passive sensor (in the UV and the visible solar spectral region) with both nadir and limb measurement capabilities on the ESA's ENVISAT platform.
SLTSR	(Sea and Land Surface Temperature Radiometer) on-board SENTINEL-3 is to maintain continuity with the (A)ATSR series of instruments. Additional new features include a wider swath, new channels (including two channels dedicated to fire detection), and higher resolution in some channels.
SSA	(Single Scattering Albedo) quantifies the likelihood of scattering during an attenuation (or 'extinction') event by an atmospheric particle of given size and shape at a certain wavelength (most important at 550 nm, the reference wavelength in global modeling). The remaining fraction, 1-SSA referred to co-single scattering albedo, quantifies the likelihood of absorption during an attenuation (or extinction) event.

General definitions

For evaluations of retrieval skill, aside from side-by-side comparisons and difference plots, the following metrics have been used.

linear correlation coefficient ***K***

$$K = \frac{\sum_{i=1}^{N}(x_{Sat} - < x_{Sat} >)(x_{AER} - < x_{AER} >)}{\sqrt{\sum_{i=1}^{N}(x_{Sat} - < x_{Sat} >)^2 \sum_{i=1}^{N}(x_{AER} - < x_{AER} >)^2}}$$

root mean square error ***RMSE*** *or rms*

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_{Sat} - x_{AER})^2}{N}}$$

standard deviation *STDV* or σ

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}[(x_{Sat} - x_{AER}) - < (x_{Sat} - x_{AER}) >]^2}{N}}$$

bias ***b*** (taken from difference histograms)

$$b = \frac{1}{N}\sum_{i=1}^{N}(x_{SAT} - x_{AER})$$

normalized bias **<b>**

$$< b >= \frac{b}{< x_{AER} >}$$

modified normalized bias ***<mb>***

$$< mb >= \frac{b}{< 0.5 * (x_{SAT} + x_{AER}) >}$$

In the above formulas x_{Sat} corresponds to satellite test data *D* and x_{AER} corresponds to the matching reference data *R*. $< x >$ is the average value and N is the number of data pairs. The bias corrected RMSE subtracts the average bias from each point before calculating the sum of the squared differences.

Table of Contents

Scope of the document

This document provides the product assessment within C3S_312b_Lot2 of all products delivered for this project as defined in [D1], including total column and (parts of) the vertical information products. This report includes the whole set of products processed under this contract from sensors ATSR-2/AATSR, SLSTR, IASI, OLCI, POLDER, and GOMOS, including four 6-monthly extensions (ICDRs) for three sensors (SLSTR, IASI, OLCI), and a full CDR reprocessing for all sensors (including ATSR-2/AATSR, POLDER and GOMOS). Consequently, this report shows evaluation results for 2008 for early sensors (ATSR-2 / AATSR, POLDER, GOMOS), and the analyses of later ICDR datasets for 2018 and 2019 for the recent sensors (SLSTR, IASI, OLCI). In addition, this report contains stability analysis of full records processed.

This document cannot fully cover the validation of vertical resolved products due to missing reference data sets. Foremost, a sufficient amount of independent and high accuracy reference data for validation of vertical layer height or profile information is not available, so that only comparisons to other datasets of the same variable with limited coverage can be made.

Executive summary

This version of the report is the final issue under the current contract and had 3 earlier issues in the predecessor contract. It includes the final assessments of the predecessor contract C3S_312a_Lot5 with its latest CDR data provision 9/2018 and focuses on new ICDRs delivered in 2/2019, 8/2019 and 2/2020 and 8/2020 and CDRs processed in 11/2020 under this contract C3S_312b_Lot2. The datasets comprise of three total column and two vertical information aerosol product series. They include all four GCOS variables (AOD = Aerosol Optical Depth and SSA = Single Scattering Albedo, AEP = aerosol extinction profiles, ALH = aerosol layer height) together with two AEROCOM requested quantities (FMAOD=Fine Mode AOD and Dust AOD) and the Ångström exponent (AE) as another measure of particle size. The products are provided from different sensors with complementary information content; the first three datasets are provided with several independent algorithms:

(1) Nadir view instruments (OLCI): AOD
(2) Dual view instruments (ATSR-2, AATSR; SLSTR): AOD / FMAOD (+ Ångstroem exponent)
(3) Thermal infrared spectrometers (IASI): Dust AOD / dust layer height
(4) Polarimetric multi-angle radiometer (POLDER): AOD / FMAOD / SSA (+ Ångstroem exponent) / aerosol layer height
(5) Star occultation (GOMOS): stratospheric AEP

The datasets for which half-yearly ICDR extensions are provided and assessed cover the first three sensor lines while for the other two sensor lines only a full reprocessing is planned towards the end of this contract.

Investigations involve direct comparisons of the retrieved satellite ECV datasets to trusted (but only local) networks of ground-based reference measurements (AERONET sun photometers, NDACC and EARLINET lidar) and comparisons to retrievals of other satellite sensors (NASA MODIS and CALIPSO).

The validation is done based on the level3 gridded products of the different retrievals. Standard statistical quantities (bias, RMSE, correlation) are assessed.

ATSR ESA's most mature AOD retrievals (derived from changes in solar reflection to space) are those of the AATSR sensor. The latest round robin comparison (de Leeuw, et al., 2015) showed that the retrieval from Swansea University (SU) has the highest skill compared to the retrievals from FMI (ADV) and RAL (ORAC). The error weighted ensemble (EN) derived from the three retrievals performs no better than the SU retrieval and the gain in coverage is not significant.

The retrieval skill of fine mode AOD (FM AOD, linked to aerosol from pollution and wildfires) is slightly poorer than for total AOD. The one of Angström exponent is much poorer than for total AOD.

SLSTR This document covers the years 2018 and 2019 of SLSTR data. The retrieval skill is still a bit worse than the AATSR retrievals from ESA's Aerosol-CCI project, but especially the SU retrieval and the ensemble are close with the newest AATSR retrieval versions. The first assessment of SLSTR Fine Mode AOD also shows good skill.

IASI These AOD retrievals (from changes in IR radiation to space) are only sensitive to larger elevated aerosol sizes, thus mainly to mineral dust aerosol. The evaluation though is done for mid-visible AOD so the skill assessment is not only influenced by the retrieval algorithms but also by assumptions for the spectral conversion. Considering the most recent versions for the year 2018 and 2019, the retrievals from three algorithms by Université Libre de Bruxelles (ULB), Laboratoire de Météorologie Dynamique (LMD) and BIRA (MAPIR Mineral Aerosol Profiling from thermal Infrared Radiances) all show some skill in retrieving dust AOD but all have some problem with certain statistical parameters. The DLR retrieval (IMARS Infrared Mineral Aerosol Retrieval Scheme) displays inconsistencies and lowest correlation and its use cannot be recommended at this stage.

POLDER The GRASP algorithm is an interesting new concept (multi-pixel smoothness constraints, highest number of independent observations) and promises consistency among most retrieval unknowns without relying on (fixed) a priori assumptions. Thus, it is not surprising that the AOD attribution to fine and coarse mode is much better than in most other retrievals. The AOD retrieval skill, which only could be tested over land does very well but fails to provide low AOD values.

OLCI This nadir viewing instrument provides wider swath width and thus better coverage of AOD retrievals than the more sensitive dual view and polarimetric instruments. Accordingly the quality achieved by two algorithms (XBAER Extended Bremen aerosol retrieval by Bremen university and S4O SeaWIFS4OLCI by DLR) is significantly weaker than for SLSTR with XBAER performing better.

GOMOS This star occultation instrument provides complementary stratospheric extinction profiles. Its validation is limited due to the sparsity of suitable reference measurements at typically only few ground-based lidar stations.

1. Product validation methodology

In this section the reference data sets to be assessed of the different satellite products are described. Because the tools used for the assessment are the same as used in the AEROCOM project, it uses the AEROCOM variable naming scheme (which had also been used in the Aerosol_cci and C3S_312a_Lot5 projects). In the following sub chapters the concrete variable naming used for assessment is noted for each product delivered.

The assessment itself is done according to the accompanying Product Quality Assurance Document (PQAD, [D1]) if not noted otherwise. In general three parameters are used for evaluation of all parameters:

- Bias
- Correlation
- RMS

while a fourth quantity is used for Aerosol Optical Depth (AOD):

- GCOS fraction (pixels within GCOS required envelope; defined for AOD only).

1.1 ATSR-2 / AATSR

This analysis uses the level 3 daily gridded products of the ATSR sensors only. Shown monthly, seasonal or yearly means are derived from the daily means. Monthly means provided by data providers are **not** used.

There are three variables delivered from AATSR (AEROCOM variable name in brackets): Total AOD (OD550_AER), FM AOD (OD550LT1_AER) and Ångström exponent (ANG4487_AER).

1.1.1 ADV 4.0

The variables used for validation are:
AOD550_mean as total AOD, *FM_AOD550_mean* as fine mode AOD and *ANG550-670_mean* as Ångström exponent.

1.1.2 ORAC 4.01

The variables used for validation are:
AOD550_mean as total AOD, *FM_AOD550_mean* as fine mode AOD and *ANG550_870_mean* as Ångström exponent. This version 4.01 has only minor differences to the latest version 4.02 specified for the reprocessing in 2020.

1.1.3 SU 4.33

The variables used for validation are:
AOD550_mean as total AOD, *FM_AOD550_mean* as fine mode AOD and *ANG550_870_mean* as Ångström exponent. This version 4.3 has only minor differences to the latest version 4.32 specified for the reprocessing in 2020.

1.1.4 ensemble.v3.0

The variables used for validation are:
AOD550 as total AOD and FM_AOD550 as fine mode AOD. This version adds the fine mode AOD to the ensemble dataset the first time.

1.2 IASI

The evaluation uses the level 3 daily gridded products of the IASI retrievals only. Shown monthly, seasonal or yearly means are derived from the daily means.

There are two variables delivered from IASI (AEROCOM names in brackets): Dust AOD (OD550DUST_AER) and Dust layer height (Z_DUST). Because the assessment is done using the AERONET SDA coarse mode AOD, the comparisons show the AEROCOM variable name for coarse mode AOD (OD550GT1_AER). For clearness the exact variable names are also listed for the IASI data sets.

1.2.1 IASI_DLR.v7.0 (IMARS)

The variables used for validation are:
D_AOD550 as dust AOD.

1.2.2 IASI_LMD_V2.2

The variables used for validation are:
Daod550 as dust AOD and *Mean_dust_layer_altitude* as dust layer altitude.

1.2.3 IASI_MAPIR_v4.1

The variables used for validation are:
D_AOD550_mean as dust AOD and *D_ALT_mean* as dust layer altitude.

1.2.4 IASI_ULB_v8

The variables used for validation are:
D_AOD550_mean as dust AOD.

1.2.5 IASI_ensemble.v1.1

The variable used for validation are:
DAOD550 as dust AOD.

1.3 POLDER / PARASOL

The evaluation uses the level 3 daily gridded products of the PARASOL retrievals only. Shown monthly, seasonal or yearly means are derived from the daily means.

There are four variables in the delivery (aerocom variable names used in brackets): Total AOD (OD550_AER), FM AOD (OD550LT1_AER), Ångström exponent (ANG6787_AER) and SSA (SSA670_AER).

1.3.1 PARASOL_GRASP_V02.10

The variables used for validation are:
AOD550 as total AOD and *FM_AOD565* as fine mode AOD, ANG670_865 as Ångström Exponent , SSA670 as SSA at 670nm and AerosolLayerHeight as aerosol layer height. Because the data providers

consider the 443nm channel of PARASOL too noisy, the usage of any of its measurements is discouraged. The assessment therefore excludes data from this channel.

1.4 GOMOS / AERGOM

The evaluation uses the level 3 five day averages provided by the GOMOS AERGOM algorithm. To calculate correct monthly means, these five day averages are then distributed through their 5 days to form a daily data set. From this daily data set the shown monthly, seasonal or yearly means are derived. The product includes stratospheric extinction profile, extinction from polar stratospheric clouds, stratospheric aerosol optical depth and a size parameter.

1.4.1 GOMOS_AERGOM.v5.00

The variable used for validation is S_AOD550 as stratospheric AOD. This version 5.00 has significant differences to the latest version 4.00 specified for the reprocessing in 2020.

1.5 SLSTR

This analysis uses the level 3 daily gridded products of the SLSTR sensor only. Shown monthly, seasonal or yearly means are derived from the daily means. Monthly means provided by data providers are **not** used.
There are three variables delivered from SLSTR (AEROCOM variable name in brackets): Total AOD (OD550_AER), FM AOD (OD550LT1_AER) and Ångström exponent (ANG4487_AER).

1.5.1 ORAC 1.00

The variables used for validation are:
AOD550_mean as total AOD, *FM_AOD550_mean* as fine mode AOD and *ANG550_870_mean* as Ångström exponent.

1.5.2 SDV 2.10

The variables used for validation are:
AOD550_mean as total AOD, *FM_AOD550_mean* as fine mode AOD and *ANG550-670_mean* as Ångström exponent.

1.5.3 SU 1.12

The variables used for validation are:
AOD550_mean as total AOD, *FM_AOD550_mean* as fine mode AOD and *ANG550_870_mean* as Ångström exponent.

1.5.4 Ensemble 2.1

The variable used for validation is:
AOD550 as total AOD and *FM_AOD550* as fine mode AOD. Other variables have not been calculated.

1.6 OLCI

This analysis uses the level 3 daily gridded product of the OLCI sensor only. Shown monthly, seasonal or yearly means are derived from the daily means. Monthly means provided by data providers are **not** used.

The only variable analyzed from the OLCI data is Total AOD.

1.6.1 OLCI_S4O_v2.0 (SeaWIFS for OLCI)

The variable used for validation is *AOD550* as total AOD.

1.6.2 OLCI_XBAER_V1.0

The variable used for validation is *AOD550_mean* as total AOD. This version 1.0 has significant differences to the latest version 2.3 applied to MERIS.

1.6.3 OLCI_ensemble.v1.1

The variable used for validation *AOD550* as total AOD. This report is the first to cover the OLCI ensemble dataset.

2. Validation results

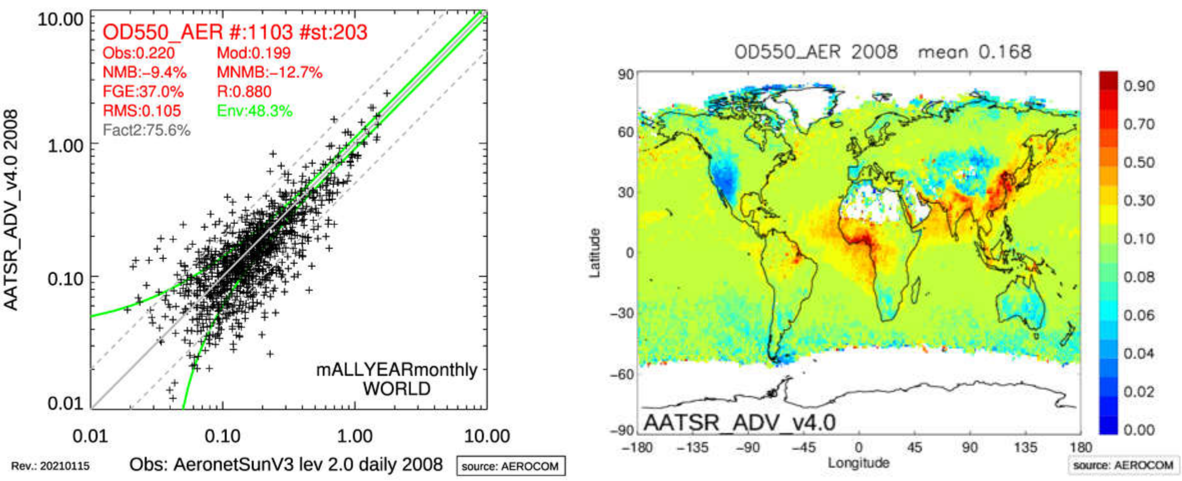
In this section the validation results are provided on per instrument and per ECV basis. Because discussing all years of AATSR retrievals would be out of the scope of this document, this is done for one year only (which shows typical results). The comparison is made for monthly means derived from daily averages. The averages are not weighted with the days per month data was provided by the satellite retrieval. Other years are available on Met Norway's validation web site (https://aerocom.met.no/cgi-bin/surfobs_annualrs.pl ; select as project "C3S-Aerosol").

2.1 AATSR

The AATSR retrievals used in this comparison are ADV_v.4.0 (ADV) from the Finnish meteorological Institute, SU_4.33 from the University of Swansea (SU) and ORAC v4.01 from the Rutherford Appleton Laboratory (ORAC). The year chosen for this validation is the year 2008 which has also been used in the phase 1 of the Aerosol_CCI project and the predecessor project C3S_312a_Lot5.

2.1.1 Total optical depth (AOD)

Figure 2.1 shows the comparison for monthly means of the total AOD for the three AATSR retrievals and the ensemble for the year 2008. On the left-hand side is a log scale scatter plot of the retrievals (y axis) against the Aeronet Sun version 3 level 2.0 product (x axis). On the right-hand side a plot of the yearly average of total AOD (derived from the monthly means) is shown. The green lines indicate the GCOS requirements (and the green %-value the fraction of pixels which fall into the GCOS required envelope), the dashed lines a factor of 2 between the satellite and the observation.



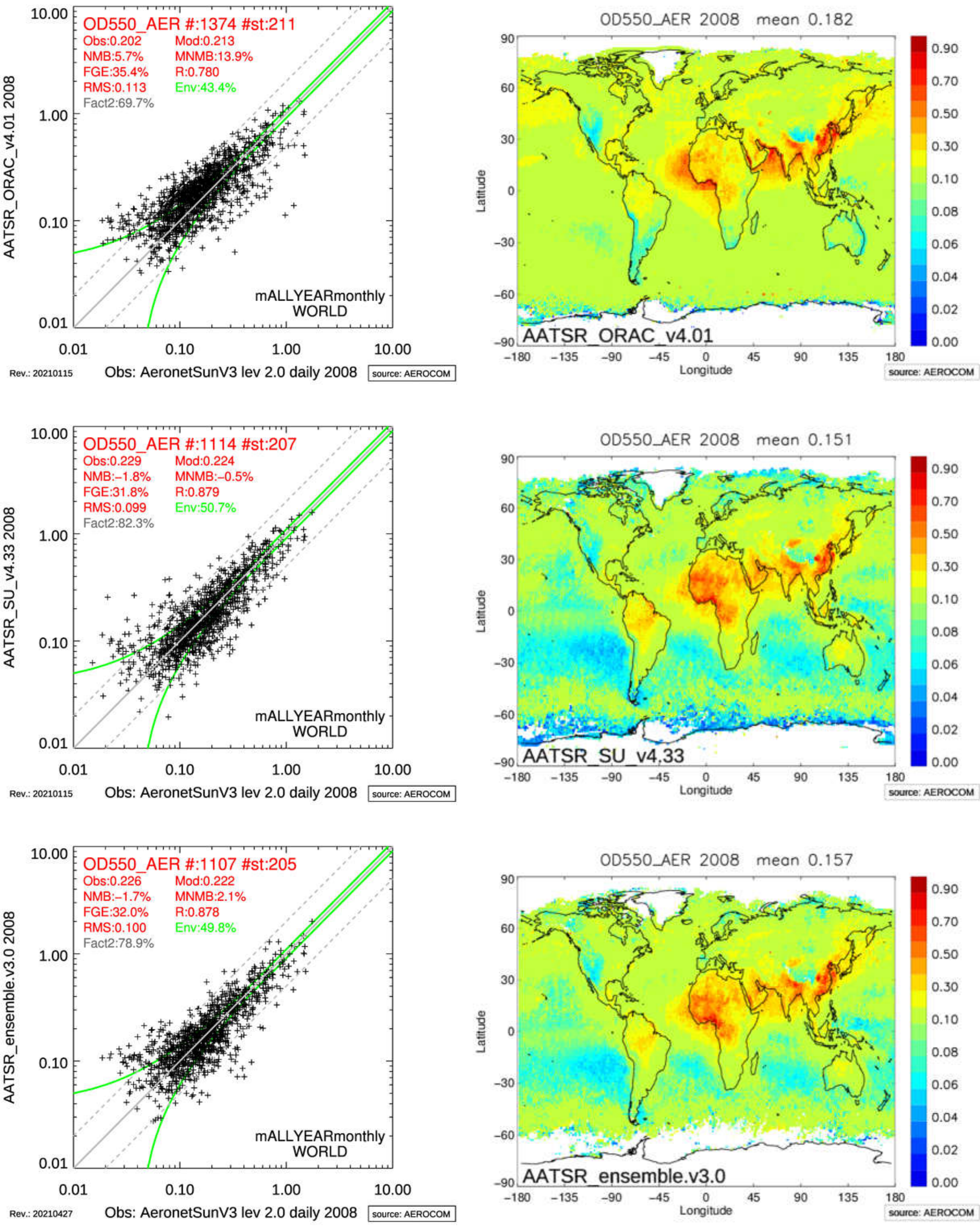


Figure 2.1: total AOD error statistics (left) for different AATSR retrievals and retrieved total annual AOD (right) for the year 2008

As shown in the scatter plots of figure 2.1 (left) and in table 2.1, retrieval averages are within +/- 10% for all the retrievals with the ensemble showing the lowest absolute bias of 1.7% and ADV showing the highest absolute bias of 9.4%.

SU, ADV and the ensemble show higher correlation (correlation of 0.88 / rmse of ~0.1) and lower RMS than ORAC (correlation of 0.78 / rmse of 0.11), but ORAC has roughly 20% more valid months than

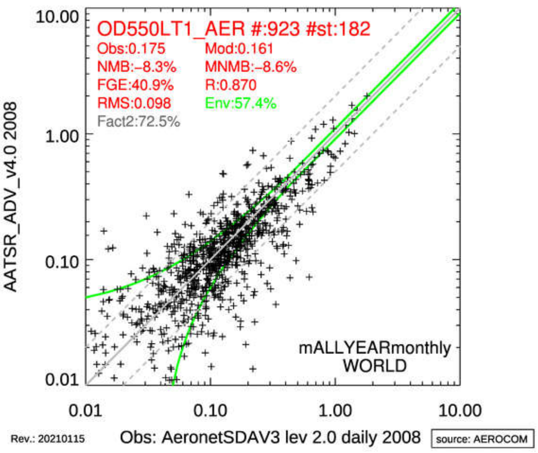
SU, ADV and the ensemble. The ensemble performs as well as SU and ADV in general, but shows no better coverage. The GCOS fraction provides the fraction of pixels which fall into the envelope defined by the GCOS AOD accuracy criterion (max(0.03, 10%)). For normal distribution of errors it should be at 67%, while it is around 50% for ADV, SU and the ensemble and at 43% for ORAC.

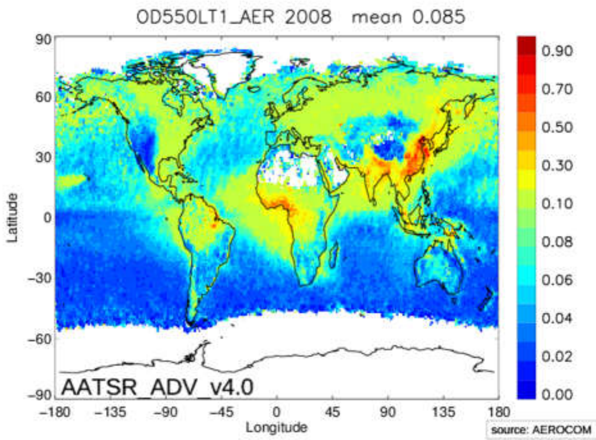
Table 2.1 level 3, year 2008 evaluation statistics for AATSR total AOD retrievals at AERONET sites

(vs. Aeronet)	**Ensemble 3.0**	**SU 4.33**	**ADV 4.0**	**ORAC 4.01**
number of pairs	1107	1114	1103	1374
AERONET AOD avg	0.226	0.229	0.22	0.202
retrieval AOD average	0.222	0.224	0.199	0.213
normalized mean bias	-1.7%	-1.8 %	- 9.4 %	5.7 %
mod. norm. mean bias	2.1%	- 0.5 %	- 12.7 %	13.9 %
correlation coefficient	.878	.879	.88	.78
root-mean square error	.1	.099	.105	.113
GCOS fraction	49.8%	50.7%	48.3%	43.4%

2.1.2 Fine mode optical depth (FM AOD)

The plots in figure 2.2 are shown the same way as for total AOD for the 3 algorithms ADV, ORAC, SU and the ensemble. The ensemble has been included for FM AOD the first time





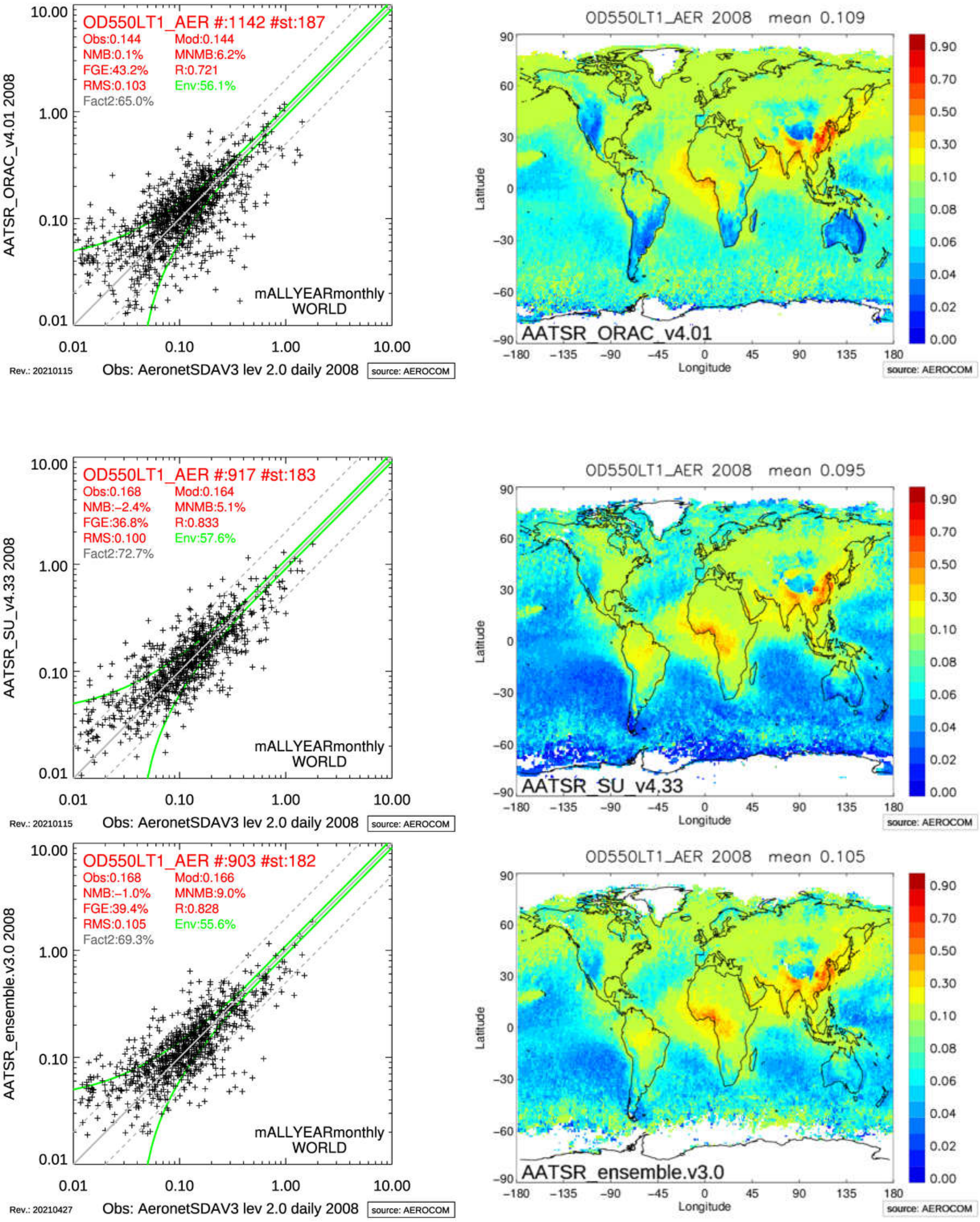


Figure 2.2: fine mode AOD error statistics (left) for different AATSR retrievals and retrieved total annual FM AOD (right) for the year 2008

Retrieval skill is in general comparable to total AOD (slightly lower bias, but lower correlation and worse rmse). Coverage for ORAC is again roughly 20% higher than for the other retrievals, but again ORAC's correlation is also lower than the other retrievals.

Note that the rmse of FM AOD should be lower than the one for total AOD because FM AOD shows in general lower values than total AOD. Because this is not the case, retrieval skill with respect to rmse is actually worse than for total AOD.

Table 2.2 level 3 year 2008 evaluation statistics for AATSR FM AOD retrievals at AERONET sites

(vs. Aeronet)	**Ensemble 3.0**	**SU 4.33**	**ADV 4.0**	**ORAC 4.01**
number of pairs	903	917	923	1142
AERONET FM AOD avg	0.168	0.168	0.175	0.144
retrieval FM AOD average	0.166	0.164	0.161	0.144
normalized mean bias	-1.0%	-2.4%	-8.3%	0.1%
mod. norm. mean bias	9.0%	5.1%	-8.6%	6.2%
correlation coefficient	.828	.833	.87	.721
root-mean square error	.105	.1	.098	.103

2.1.3 Ångström exponent (AE)

The ensemble was not calculated for Ångström exponent and can therefore not be included into the comparison here. The plots in figure 2.3 are shown the same way as for total AOD.

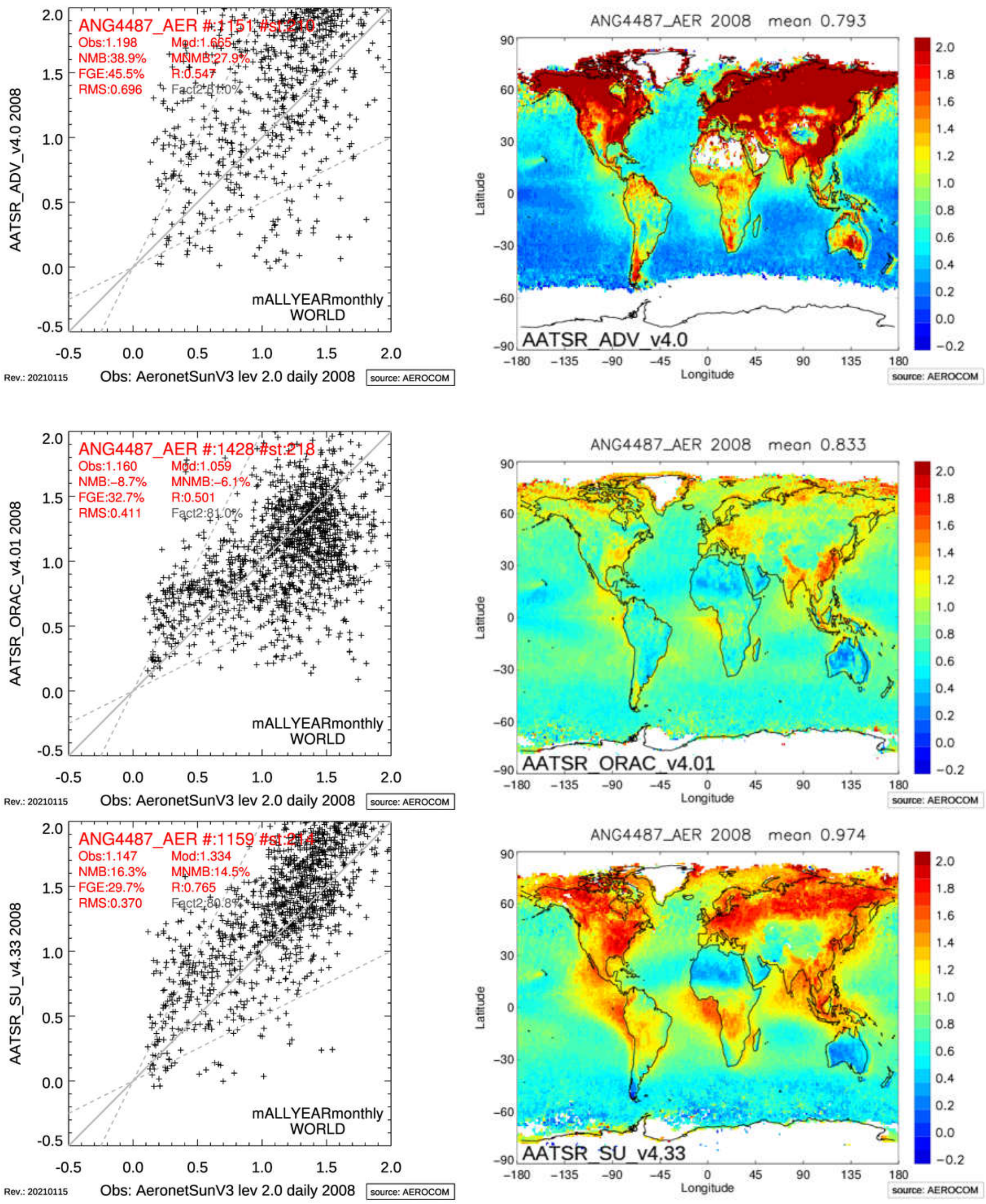


Figure 2.3: Ångström exponent error statistics (left) for different AATSR retrievals and retrieved average annual Ångström exponent (right) for the year 2008

SU and ADV show far too high values of Ångström exponent compared to Aeronet while ORAC only shows a small bias. Looking at correlation SU is best with 0.765 while ADV (0.54) and ORAC (0.5) are much worse. RMS is pretty bad at all retrievals with values ranging from 0.37 for SU to 0.7 for ADV. Looking at the maps does not show much agreement between the three retrievals with ADV showing implausible high values over large regions of the northern hemisphere and central Australia. Out of the three retrievals SU shows most skill at this point, but RMS is still very high.

Table 2.3 level 3, year 2008 evaluation statistics for AATSR Angström exponent retrievals at AERONET sites

(vs. Aeronet)	**SU 4.33**	**ADV 4.0**	**ORAC 4.01**
number of pairs	1159	1151	1428
AERONET Angström exponent average	1.147	1.198	1.160
retrieval Angström exponent average	1.334	1.665	1.059
normalized mean bias	16.3%	38.9%	-8.7%
mod. norm. mean bias	14.5%	27.9%	-6.1%
correlation coefficient	0.765	0.547	0.501
root-mean square error	0.37	0.696	0.411

2.2 IASI

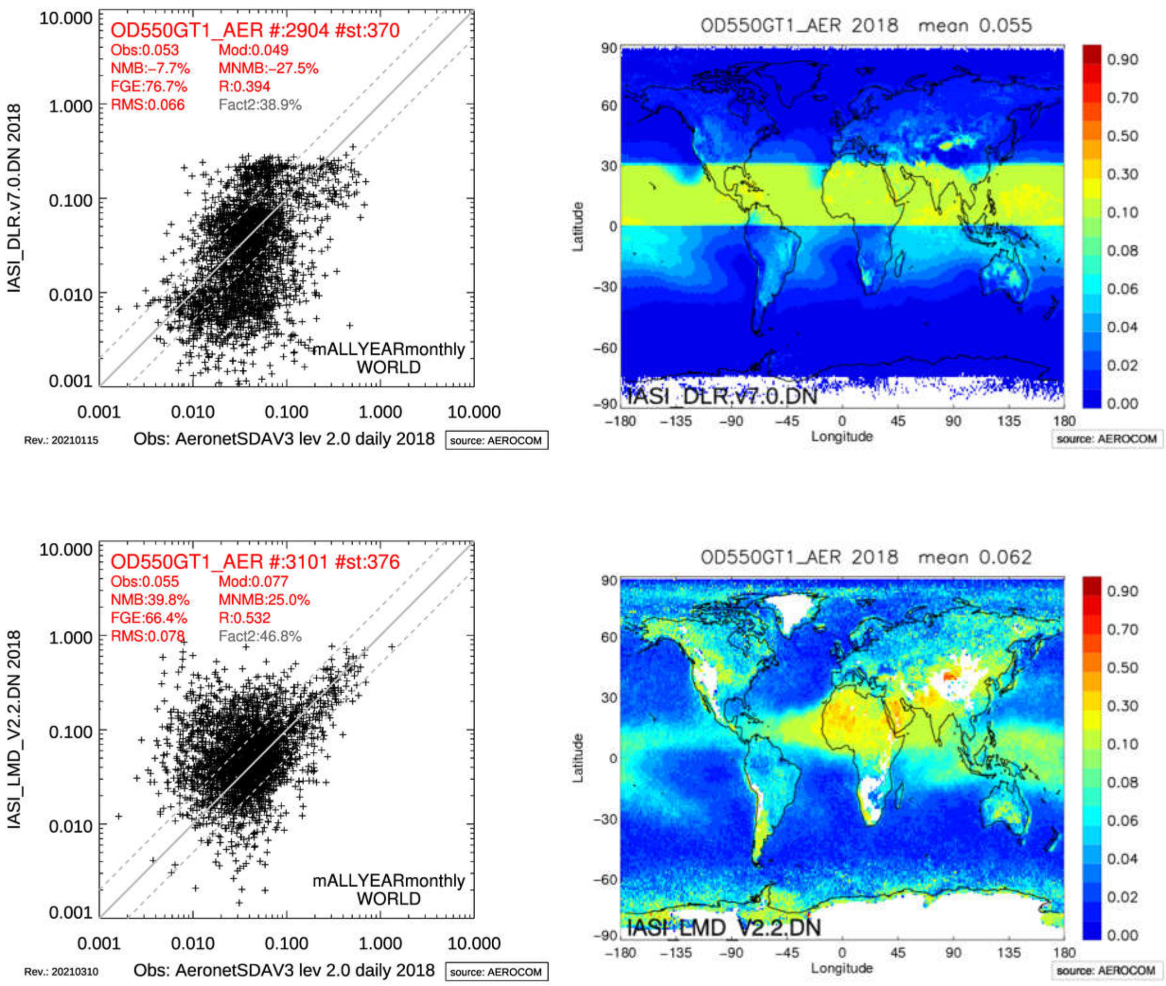
All four IASI retrieval groups submitted three different retrieval results. One of the ascending orbit (named AN at the end of the retrieval name), one for the descending orbit (named DN at the end of the retrieval name) and one for both orbits at once (named All at end of the retrieval name). Because the descending orbit is at daytime when AERONET measures as well, only retrievals using the descending orbit were used in this comparison.
The IASI retrievals used in this comparison are DLR_v7.0.DN (DLR) from the German Aerospace Center, LMD_V2.2.DN from LMD (LMD), MAPIR.v4.1.DN from (BIRA), ULB.v8.DN from ULB (ULB) and an ensemble retrieval calculated from these four retrievals. This report covers the years 2018 and 2019 for dust AOD.
Because there is no suitable aerosol layer height reference dataset for the year 2018 or 2019 freely available yet, the authors decided to keep the dust layer height assessment unchanged at the year 2013.

2.2.1 Dust AOD

During the Aerosol_cci project the data providers and the validation team agreed to validate the dust AOD with a comparison to AERONET SDA coarse mode AOD. The variable noted in the figures is therefore named after the AEROCOM variable name for coarse mode AOD (OD550GT1_AER). Because the earlier used AERONET version 2 SDA product is no longer available, the version 3 level 2.0 product is used as a reference data set in this report. The level 2.0 was used because it applies the most rigid quality control and should therefore avoid potential problems in the reference data set itself. Even though Aeronet version 3 does the quality assessment without human interaction, it still takes time to get from the raw measurement to quality-controlled data. As a result of that, the number of measurements during the years 2018 and 2019 is different and has been changed significantly since the last report. Nevertheless, the validation of both years uses the most recent data available at Met Norway during the time of writing dated from end of January 2021. Although the exact numbers are slightly different from the last report, the conclusions are the same as in the earlier versions of this document.

2.2.1.1 Year 2018

Figure 2.4 shows the comparison for monthly means of the dust AOD for the three IASI retrievals for the year 2018 for the whole globe. On the left-hand side is a log scale scatter plot of the monthly mean of the retrievals (y axis) against the AERONET SDA version 3 level 2.0 product (x axis). On the right-hand side a plot of the yearly average of dust AOD (derived from the monthly means) is shown. The dashed lines in the scatter plot indicate a factor of 2 between the satellite and the observation.



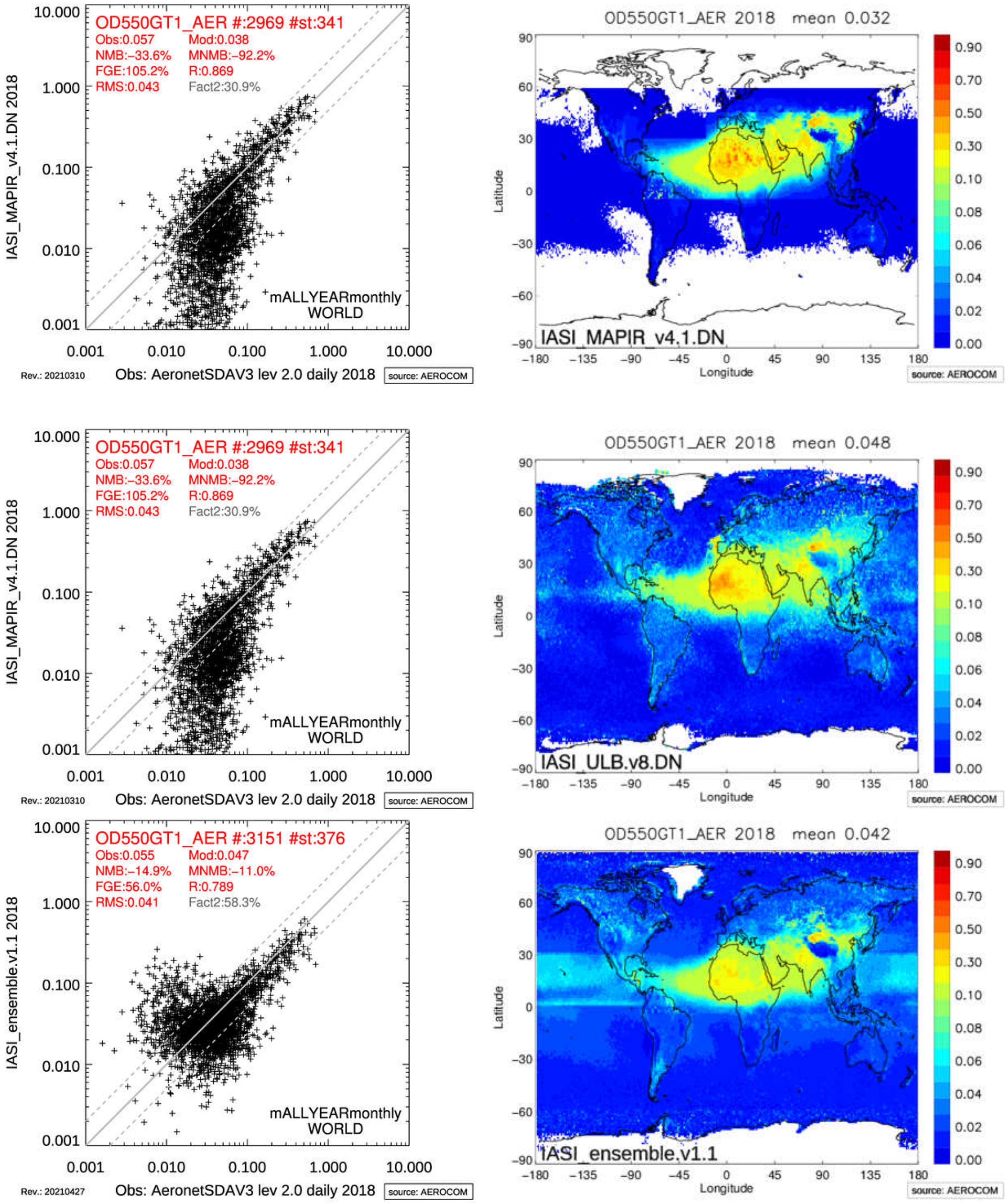


Figure 2.4: dust AOD compared to coarse mode AOD error statistics (left) for different IASI retrievals and retrieved annual average of dust AOD (right) for the year 2018.

Comparing the statistical parameters, the four retrievals can be divided into two groups: one that is performing reasonably well (MAPIR, ULB and the ensemble), and one that has some room for improvement.

The best performing retrievals in this monthly comparison to AERONET SDA are the MAPIR and the ULB retrieval. They show the highest correlations of (MAPIR 0.87, ULB 0.78), especially ULB shows a

very low bias of just -2.4% (MAPIR -33.6%) and the lowest RMS (MAPIR 0.043, ULB 0.059). The downside of ULB is the lowest number of points.
The LMD and the DLR retrieval share a lack of correlation to the measurements (DLR: 0.39, LMD: 0.53), although the latest DLR retrieval has significantly improved from earlier versions. The DLR retrieval shows a low bias of -7.7%, but a high RMS (0.066) while LMD shows a 39.8% bias and the highest RMS of 0.075. Additional statistical parameters are noted in Table 2.5.
The ensemble's retrieval skill is as expected in the middle of the four original retrievals with a bias of -15%, a correlation of 0.79 and an RMS of 0.041. But the coverage (number of datapoints) is better for the ensemble. The authors of this report consider it therefore really beneficial for the IASI dust AOD product.

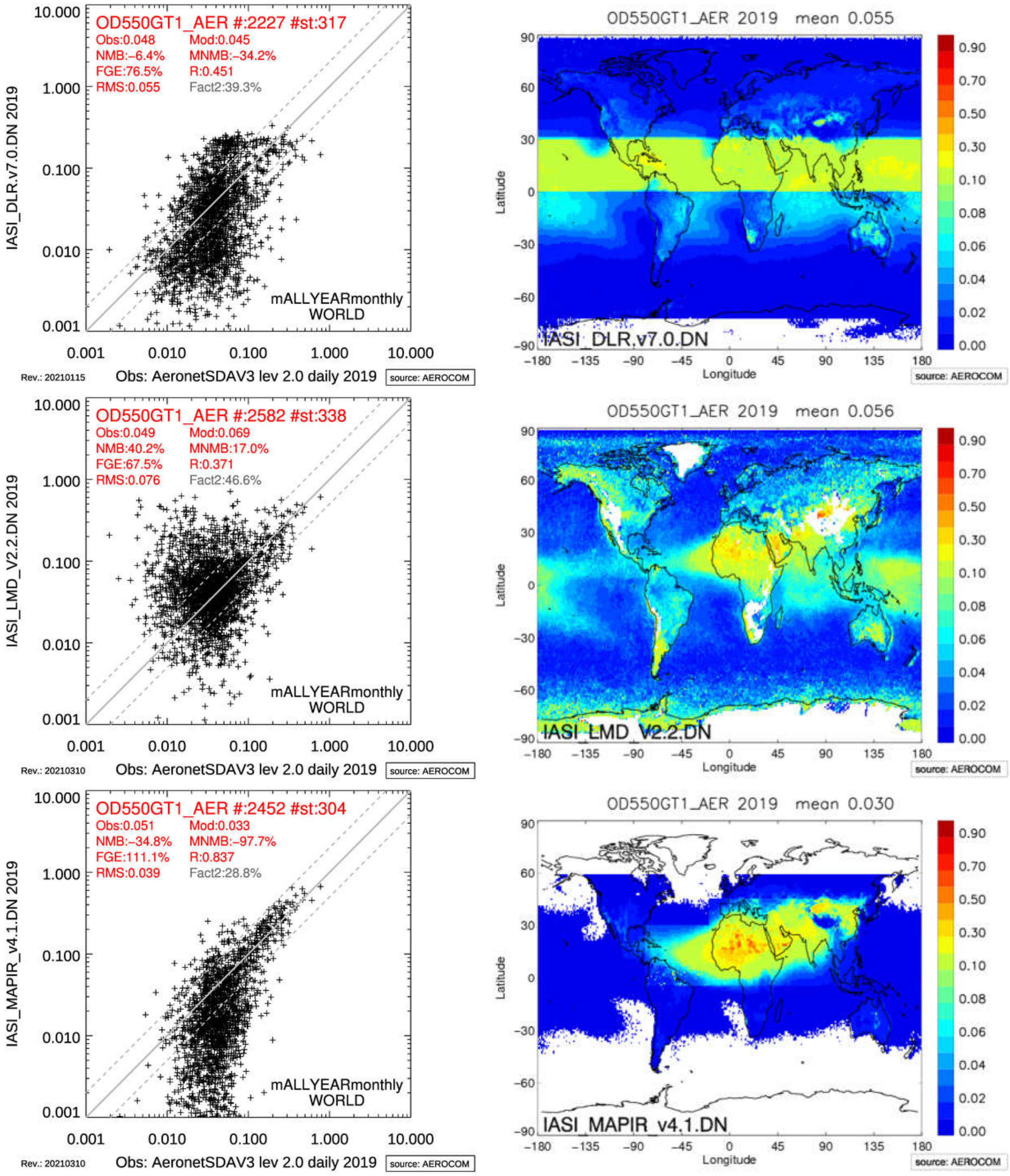
The maps show a stripe of unreasonably high values between the equator and 30 degrees north for the DLR retrieval. This is due to a method split of the dust belt and outside the dust belt to avoid erroneous DAOD outliers outside the dust region. It needs further assessment and correction in a next dataset version.

Table 2.5 level 3, year 2018 WORLD region evaluation statistics for IASI dust AOD retrievals compared to coarse mode AOD at AERONET sites

(vs. Aeronet SDA)	**Ensemble 1.1**	**DLR 7.0**	**LMD 2.2**	**MAPIR 4.1**	**ULB v8**
number of pairs	3151	2904	3101	2969	2065
AERONET coarse mode AOD average	.055	.053	.055	0.57	.067
retrieval dust AOD average	.047	.049	.077	.038	.065
normalized mean bias	-14.9%	-7.7%	39.8%	-33.6%	-2.4%
mod. norm. mean bias	-11%	-27.5%	25%	-92.2%	-17.1%
correlation coefficient	.798	.394	.532	.869	.782
root-mean square error	.041	.066	.078	.043	.059

2.2.1.2 Year 2019

Despite the roughly 20% lower number of measurements compared to those of the year 2018, the statistical parameters for the year 2019 are pretty close to those of the year 2018. The conclusion that MAPIR and ULB are the better retrievals than the two others is valid as well. As during the year 2018 MAPIR has the better coverage, correlation and RMS, while ULB shows the lowest numerical mean bias.


OD550GT1_AER #:2227 #st:317
Obs:0.048 Mod:0.045
NMB:-6.4% MNMB:-34.2%
FGE:76.5% R:0.451
RMS:0.055 Fact2:39.3%
mALLYEARmonthly
WORLD
IASI_DLR.v7.0.DN 2019
Rev.: 20210115
Obs: AeronetSDAV3 lev 2.0 daily 2019
source: AEROCOM
OD550GT1_AER 2019 mean 0.055
IASI_DLR.v7.0.DN
Latitude
Longitude
source: AEROCOM
OD550GT1_AER #:2582 #st:338
Obs:0.049 Mod:0.069
NMB:40.2% MNMB:17.0%
FGE:67.5% R:0.371
RMS:0.076 Fact2:46.6%
mALLYEARmonthly
WORLD
IASI_LMD_V2.2.DN 2019
Rev.: 20210310
Obs: AeronetSDAV3 lev 2.0 daily 2019
source: AEROCOM
OD550GT1_AER 2019 mean 0.056
IASI_LMD_V2.2.DN
Latitude
Longitude
source: AEROCOM
OD550GT1_AER #:2452 #st:304
Obs:0.051 Mod:0.033
NMB:-34.8% MNMB:-97.7%
FGE:111.1% R:0.837
RMS:0.039 Fact2:28.8%
mALLYEARmonthly
WORLD
IASI_MAPIR_v4.1.DN 2019
Rev.: 20210310
Obs: AeronetSDAV3 lev 2.0 daily 2019
source: AEROCOM
OD550GT1_AER 2019 mean 0.030
IASI_MAPIR_v4.1.DN
Latitude
Longitude
source: AEROCOM

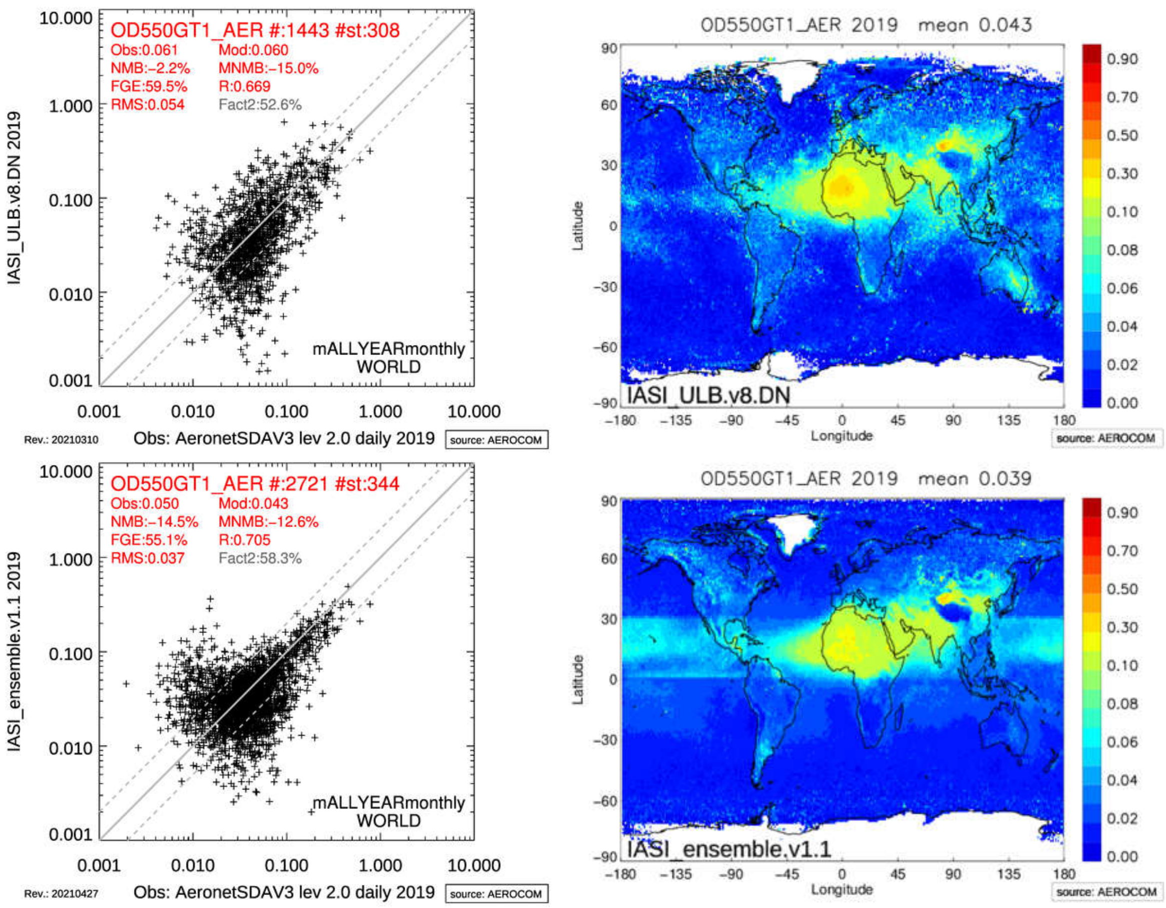


Figure 2.5: dust AOD compared to coarse mode AOD error statistics (left) for different IASI retrievals and retrieved annual average of dust AOD (right) for the year 2019.

Table 2.6 level 3, year 2019 WORLD region evaluation statistics for IASI dust AOD retrievals compared to coarse mode AOD at AERONET sites

(vs. Aeronet SDA)	**Ensemble 1.1**	**DLR 7.0**	**LMD 2.2**	**MAPIR 4.1**	**ULB v8**
number of pairs	2721	2227	2582	2452	1443
AERONET coarse mode AOD average	.05	.048	.049	.051	.061
retrieval dust AOD average	.043	.045	.069	.033	.06
normalized mean bias	-14.5%	-6.4%	40.2%	-34.8%	-2.2%
mod. norm. mean bias	-12.6%	-34.2%	17%	-97.7%	-15%
correlation coefficient	.705	.451	.371	.837	.669
root-mean square error	.037	.055	.076	.039	.054

2.2.2 Aerosol layer height

This ECV has so far only been provided by two of the retrievals: LMD and MAPIR.

At this point there are two reference data sets available for this ECV: A CALIOP dust layer height product and dust layer height derived from EARLINET measurements.

2.2.2.1 Dust layer height from CALIOP

Due to the limited resolution in space and time of the CALIOP instrument (single LIDAR), the product is relatively coarse (2 by 5 degrees, monthly averages only). Furthermore, the information content of a Lidar as an active instrument is much higher than the one of a passive instrument like IASI. CALIOP provides therefore information about more than just one dust layer height since it can under certain circumstances retrieve more than one aerosol layer.

The data providers have limited experience with this CALIOP product, but agreed that CALIOP V3.00 cloud free product *Highest_Aerosol_Layer_Detected_Dust* with the 50% percentile should be the right product to carry out the comparison. Nevertheless, the comparison between the IASI LMD product and CALIOP is qualitative only at this point.

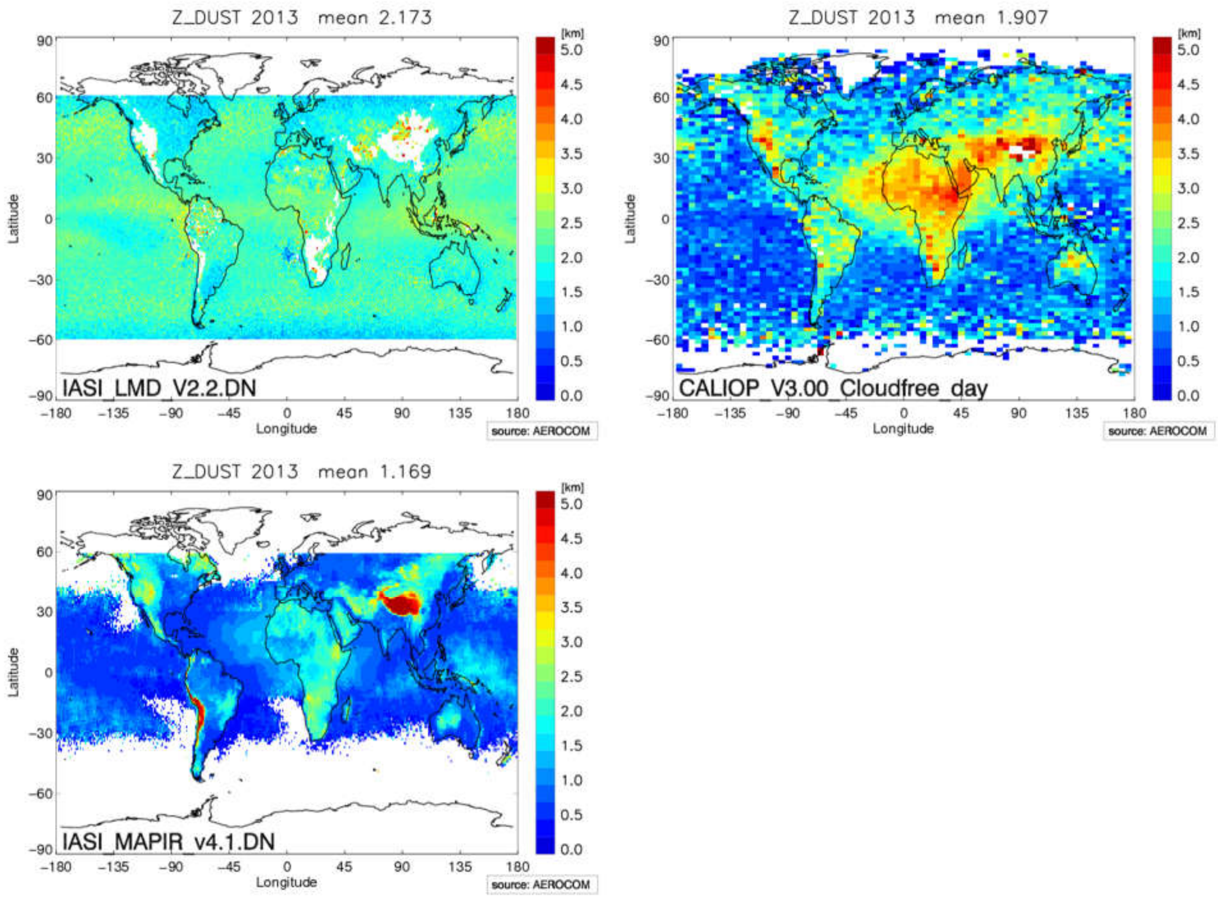


Figure 2.6: annual average of dust layer height of the IASI LMD (left panel) and CALIOP maximum dust layer height.

Figure 2.6 shows yearly averages for the dust layer height provided by LMD (upper left panel), MAPIR (lower left panel) and CALIOP (right panel). The three panels look quite different. While the IASI LMD retrieval is not able to retrieve dust layer height over the Gobi desert, the Rocky Mountains, the Andes and over big parts of southern Africa, MAPIR and CALIOP can. For the northern part of the Andes and the Gobi desert MAPIR and CALIOP agree reasonably well, but MAPIR does show much lower dust layer heights over Africa than CALIOP.

2.2.2.2 Dust layer height from EARLINET

The European Aerosol Research Lidar Network to Establish an Aerosol Climatology (EARLINET) also provides a dust layer height product. Unfortunately, only 12 out of 21 existing stations ever submitted a dust layer height since EARLINET was founded. Out of these, only 11 submitted a dust layer height during the time IASI has been active. All in all there are 220 days of measurements with a valid dust layer height in the data set. For an unknown reason, the latest submission of dust layer height is from the year 2013 while the latest backscatter measurement easily available is from the year 2015. EARLINET has unfortunately not adopted an open data policy so far. Newer data is therefore not available to Met Norway at this time.

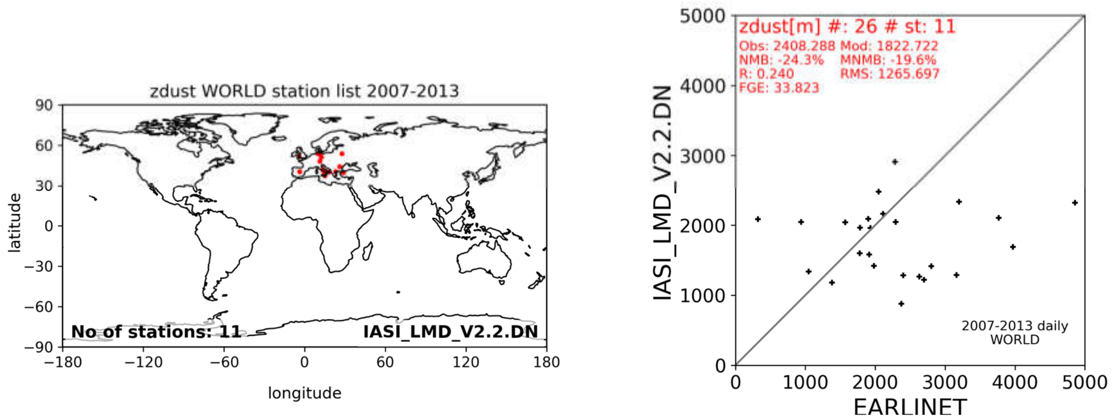


Figure 2.7: IASI LMD dust layer height compared to the EARLINET dust layer height product: locations of stations (left panel) and scatterplot for all days where IASI LMD and EARLINET provided a dust layer height within the same day and grid box.

Figure 2.7 shows a comparison of the IASI LMD dust layer height compared to the EARLINET dust layer height product. Because the L3 production of the MAPIR product has not been finished for the whole time series at the time of this writing, the MAPIR dust layer height product cannot be shown at this time. The panel on the left-hand side shows the location of the eleven EARLINET sites that provided a dust layer height since 2007.

The scatter plot on the right-hand side of Figure 2.7 shows a comparison for all coincidences that could be found. Only 26 coincidences with LMD retrieval within the same grid box and the same day were found. While roughly half of these agree reasonably well within +/- 500m, there are also very big differences that result in low correlation (0.24) and high RMS.

At this point the quality of the reference data set is not clear. The author could not find any earlier usages of this data in the literature besides case studies on certain dust events. The number of measurements is so limited that the resulting statistics is not reliable. For a case study the L3 IASI data that is part of this project is likely not the right product to use due to its averaging over a large area and time frame.

2.3 PARASOL GRASP

The Parasol GRASP retrieval submitted four ECVs for this report: Total AOD, fine mode AOD, Ångström exponent and SSA. The current version 2.10 is the first version that covers the whole globe as the other retrievals in this report. This assessment focusses on the year 2008 data.

To estimate GRASP skill compared to other satellite retrievals, the AATSR SU 4.3 retrieval is used in comparison. Because AATSR does not provide any information about SSA, this cannot be done for that ECV.

2.3.1 Total AOD

Figure 2.8 shows the comparison for monthly means of the total AOD for the GRASP retrieval for the year 2008. On the left-hand side is a log scale scatter plot of the retrievals (y axis) against the AERONET Sun version 3 level 2.0 product (x axis). On the right-hand side a plot of the yearly average of total AOD (derived from the monthly means) is shown. The green lines indicate the GCOS requirements, the dashed lines a factor of 2 between the satellite and the observation.

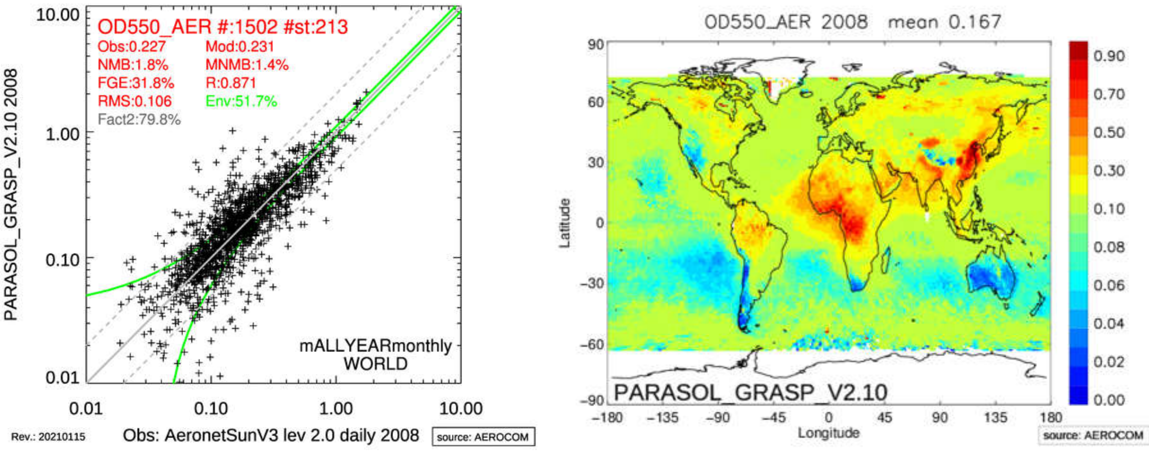


Figure 2.8: Total AOD error statistics (left) for the GRASP retrieval and retrieved annual average AOD (right) for the year 2008

GRASP shows good skill for total AOD: high correlation (0.87), low RMS (0.11) and a very low bias (1.8%).

Figure 2.9 shows the same scatter plot as before for the Northern Africa region for AATSR based SU 4.33 and GRASP. Correlation of the GRASP retrieval is slightly lower (AATSR SU: 0.915 vs. GRASP 0.861), the percentage of data pairs within the GCOS requirements is slightly higher (GRASP: 57.3%, AATSR SU: 50.4%). In addition, GRASP provides a roughly 20% higher number of data points in the North African region than AATSR.

All in all the GRASP retrieval shows a very good performance for total AOD.

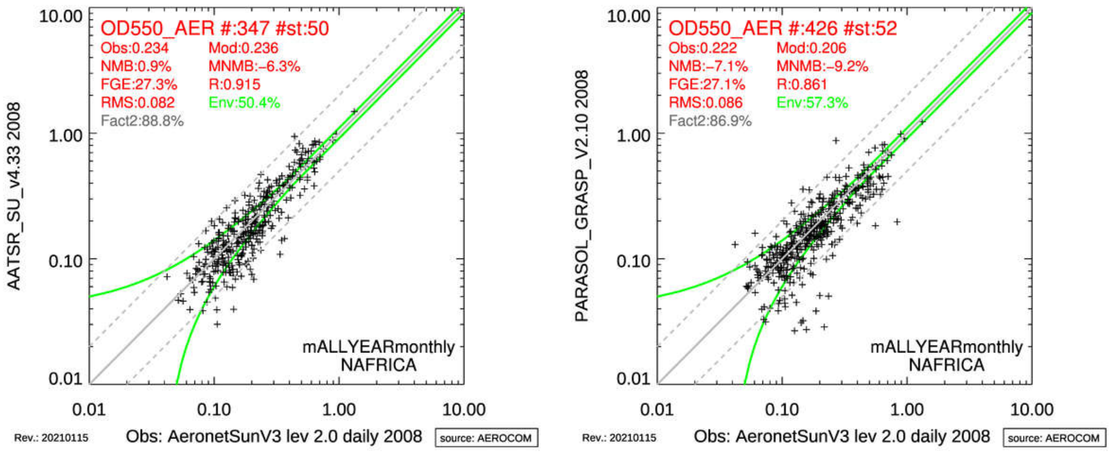


Figure 2.9: scatter plots of total AOD of the AATSR SU 4.3 retrieval (left panel) and of PARASOL GRASP (right panel) for the Northern Africa region at the same time

2.3.2 Fine mode AOD

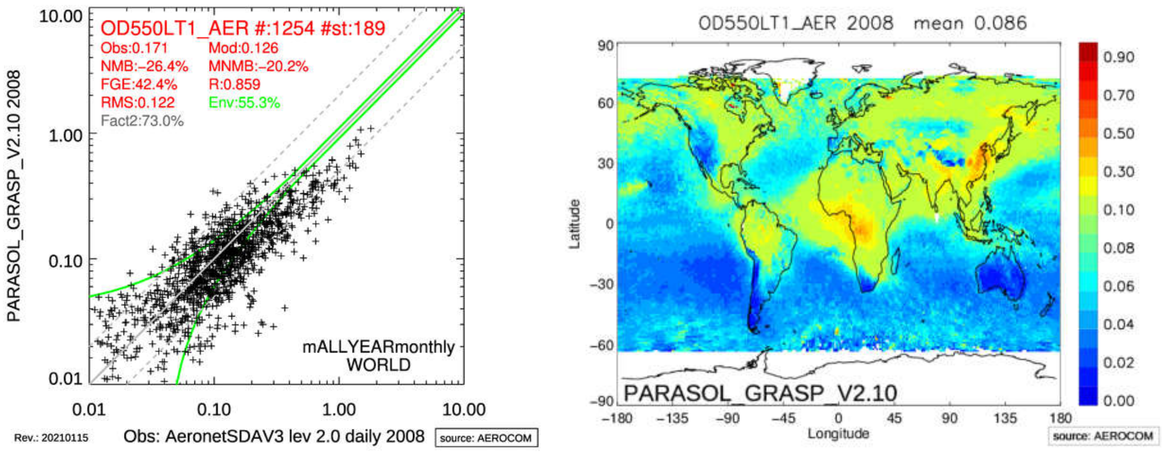
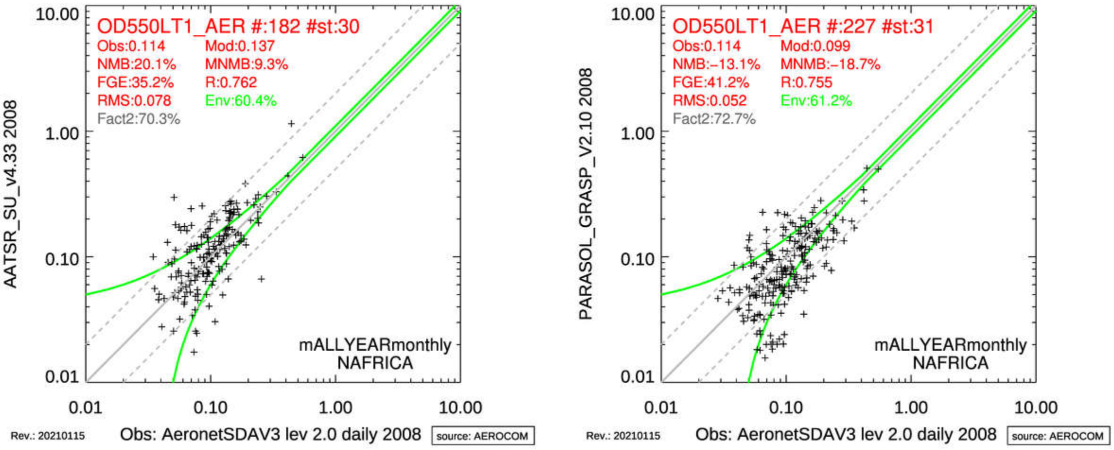


Figure 2.10: Fine mode AOD error statistics (left) for the GRASP retrieval and retrieved annual average FM AOD (right) for the year 2008

Globally GRASP shows good skill compared to AERONET for the FM AOD retrieval. Correlation (0.86) is very high, but RMS and bias are a bit high (0.112 and 22.6%).
Figure 2.11 shows scatter plots of FM AOD of AATSR SU4.3 and GRASP in the Northern Africa region. Comparing the statistical parameters shows that GRASP's skill very much the same for all statistical parameters in that region as the newest AATSR SU retrieval.



670

Figure 2.11: scatter plots of FM AOD of the AATSR SU 4.3 retrieval (left panel) and of PARASOL GRASP (right panel) for the northern Africa region at the same time

2.3.3 Ångström exponent (AE)

GRASP delivered Ångström exponent using the wavelengths of 670nm and 865nm. Because the AEROCOM project usually uses 440nm as the lower wavelength a new reference data set was calculated using the same wavelengths as GRASP from AERONET measurements.

This assessment was done using the delivered ECV as it is (without any data filtering). According to the data provider “retrieved AE values can be trustful for AOD550 > = 0.2 over land and AOD550 >= 0.1 over ocean” only.

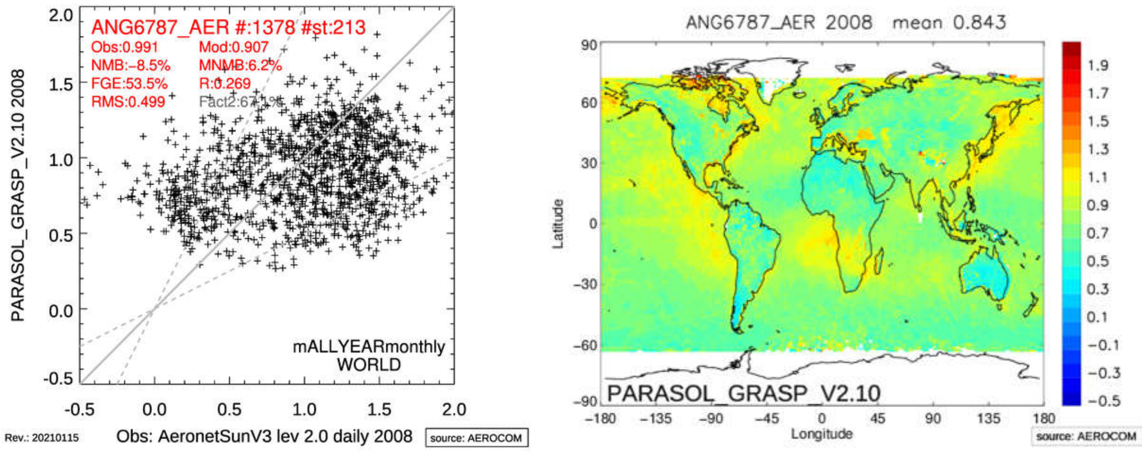


Figure 2.12: Ångström exponent error statistics (left) for GRASP retrieval and retrieved average annual Ångström exponent (right) for the year 2008

The statistical parameters show a poor agreement with AERONET. Correlation is 0.27, bias is -8.5% and the RMS is rather high with 0.5.

Comparing these parameters with the same ones of AATSR SU 4.33 (although AATSR is using a different lower wavelength of 440nm) in the same region (Northern Africa) and the same time in figure 2.13 shows that SU 4.33 has a greater skill. Correlation is higher for SU (0.84 vs. 0.39), RMS is lower for SU (0.29 vs. 0.43) but the bias higher (13.8% for SU vs. 7.9% for GRASP). All in all the GRASP retrieval cannot compete with the SU retrieval for Ångström exponent.

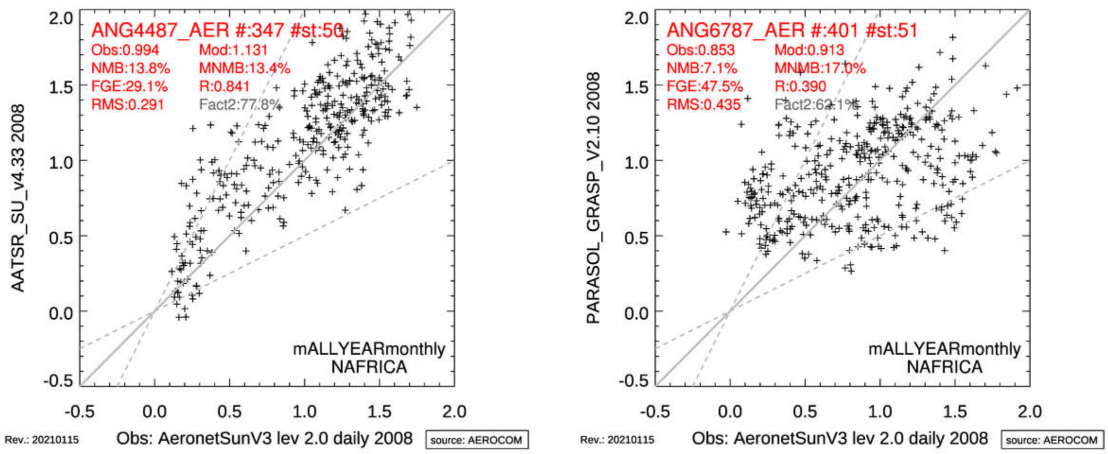


Figure 2.13: scatter plots of Angström exponent of the AATSR SU 4.3 retrieval (left panel) and of PARASOL GRASP (right panel) for the Northern Africa region at the same time

2.3.4 Single scattering Albedo (SSA)

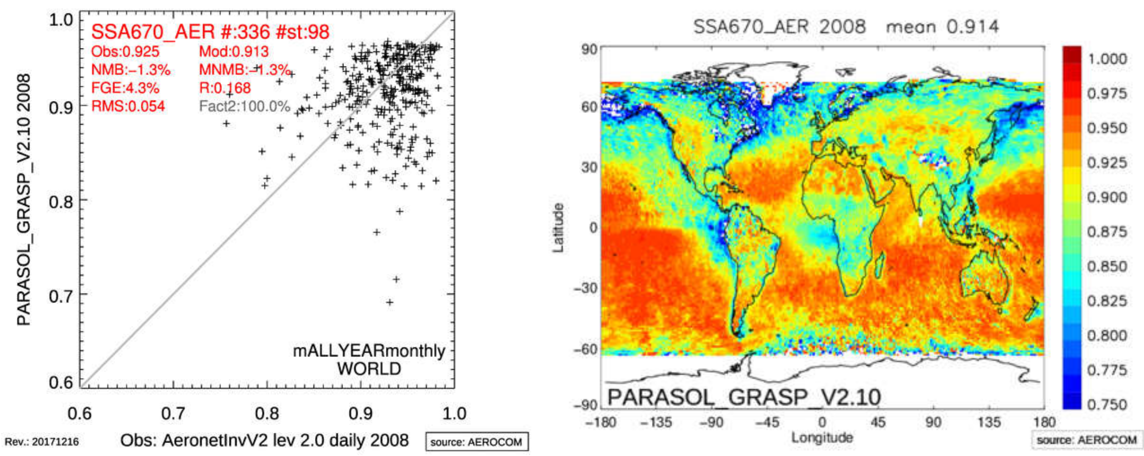


Figure 2.14: SSA @ 670nm error statistics (left) for the GRASP retrieval and retrieved annual average SSA (right) for the year 2008

At this first look the comparison to AERONET version 2 level 2 inversions do not look very promising with a correlation of 0.168. But since even the usage of their SSA data is discouraged by AERONET, the validation team and the data providers need to discuss the validation strategy further. Possible changes are the new version 3 AERONET products or a comparison to model data.

2.3.5 Aerosol layer height

In this sub chapter the aerosol layer height product from GRASP is qualitatively compared to the aerosol layer height fom CALIOP (variable name *Highest_Aerosol_Layer_Detected* at the 50% confidence interval).

Figure 2.15 shows the yearly average of the aerosol layer height for the GRASP retrieval and the CALIOP aerosol layer height product. The patterns over land are very different. Besides central Africa there is hardly any agreement. GRASP does not show high aerosol layer height in northern Africa, central Asia and the Gobi desert as CALIOP does. In addition, the high heights in higher latitudes are not really plausible and are present in the CALIOP product in very few pixels only.

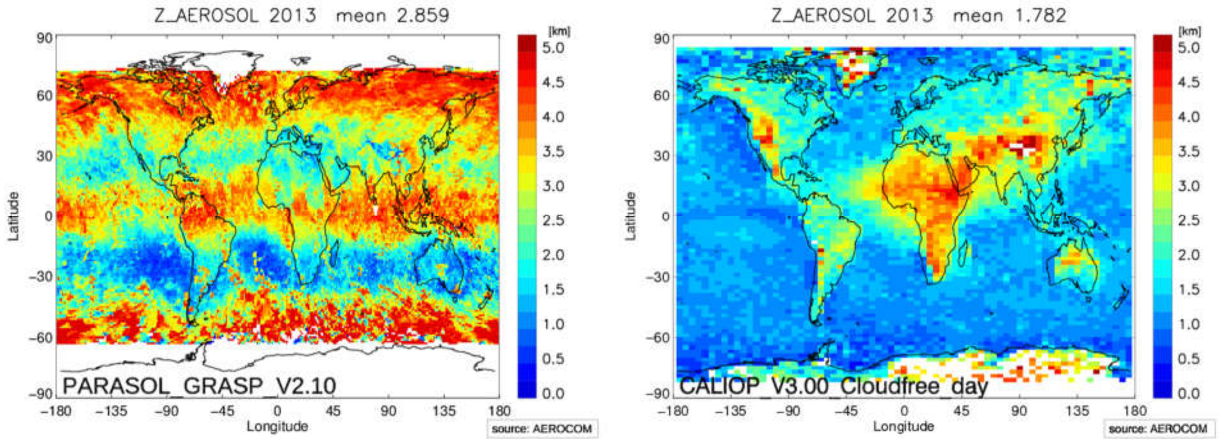


Figure 2.15: Aerosol layer height for the GRASP retrieval (left) and from the CALIOP cloud free product

2.4 SLSTR

The SLSTR retrievals used in this evaluation are ORAC_v1.00 from the Rutherford Appleton Laboratory, SDV_v2.10 from the Finnish meteorological Institute, SU_v1.12 from the University of Swansea and an ensemble that has been calculated out of these three retrievals. This report covers data from the years 2018 and 2019.

2.4.1 Total optical depth

Due to the different setup of the SLSTR instrument with rearward viewing instead of forward viewing of the two ATSR instruments, in theory, retrieval accuracy is expected to be better in the Southern hemisphere (with more suitable geometric conditions for aerosol retrieval) while it is expected to be worse in the Northern hemisphere, where the bulk of AERONET stations are located. Consequently, one expects a somewhat weaker performance of global statistics for SLSTR than for AATSR.

2.4.1.1 Year 2018

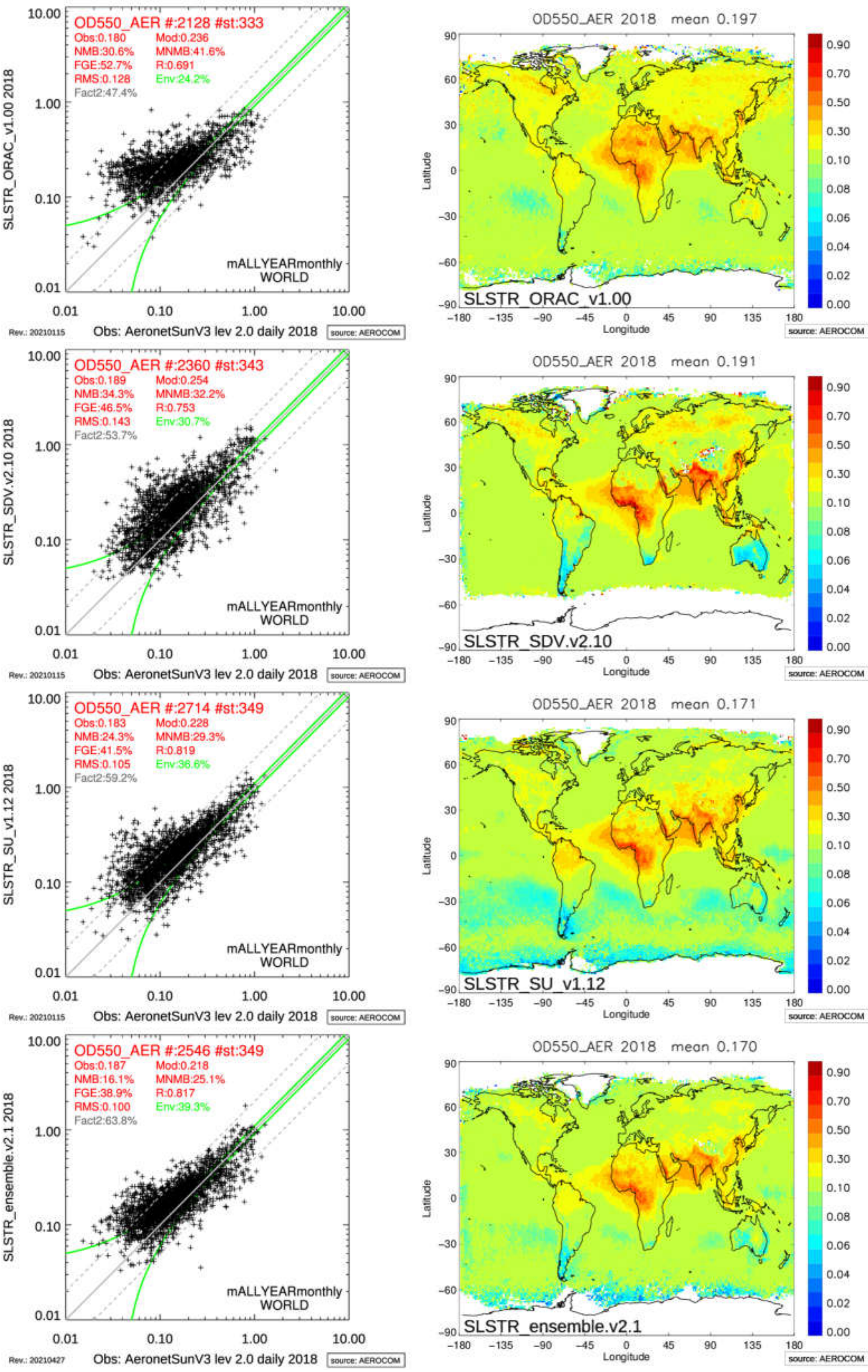


Figure 2.16: total AOD error statistics (left) for different SLSTR retrievals and retrieved total annual AOD (right) for the year 2018

Figure 2.16 shows the comparison for monthly means of the total AOD for the three SLSTR retrievals and the ensemble for the year 2018. On the left-hand side is a log scale scatter plot of the retrievals (y axis) against the AERONET Sun version 3 level 2.0 product (x axis). On the right-hand side a plot of the yearly average of total AOD (derived from the monthly means) is shown. The green lines indicate the GCOS requirements (and the green %-value the fraction of pixels which fall into the GCOS required envelope), the dashed lines a factor of 2 between the satellite and the observation.

The maps for the year 2018 show in general similar patterns but especially in India and along the western coast of Africa the differences are larger with SDV showing the highest values.
Compared to AERONET all three retrievals and the ensemble cover a similar number of AERONET stations with at least one month with the numbers ranging from 2128 for ORAC to 2714 for SU. The normalized mean bias is quite similar for all retrievals with the ensemble showing the lowest bias at 16.1% and SDV showing the highest value of 34.3%. For the correlation coefficient and the root mean square error SU and the ensemble show the best values (SU: R=0.819, RMS=0.105; ensemble: R=0.817, RMS=0.1) while ORAC and SDV are a bit worse (ORAC: R=0.691, RMS=0.128; SDV: R=0.753; RMS=0.143) Table 2.7 shows all statistical parameters for the comparison to AERONET.

Table 2.7: level 3, year 2018 evaluation statistics for SLSTR total AOD retrievals compared AERONET sites

(vs. Aeronet)	**ORAC**	**SDV**	**SU**	**Ensemble 2.1**
number of pairs	2128	2360	2714	2546
AERONET total AOD avg	.18	.189	.183	.187
retrieval total AOD average	.236	.254	.228	.218
normalized mean bias	30.6%	34.3%	24.3%	16.1%
mod. norm. mean bias	41.6%	32.2%	29.3%	25.1%
correlation coefficient	.691	.753	.819	.817
root-mean square error	.128	.143	.105	.1
GCOS fraction	24.2%	30.7%	36.6%	39.3%

All in all, the ensemble shows the best overall performance (lowest NMB, RMS and highest GCOS fraction) with just a 5% worse coverage than single retrieval with the best coverage (SU).

2.4.1.2 Year 2019

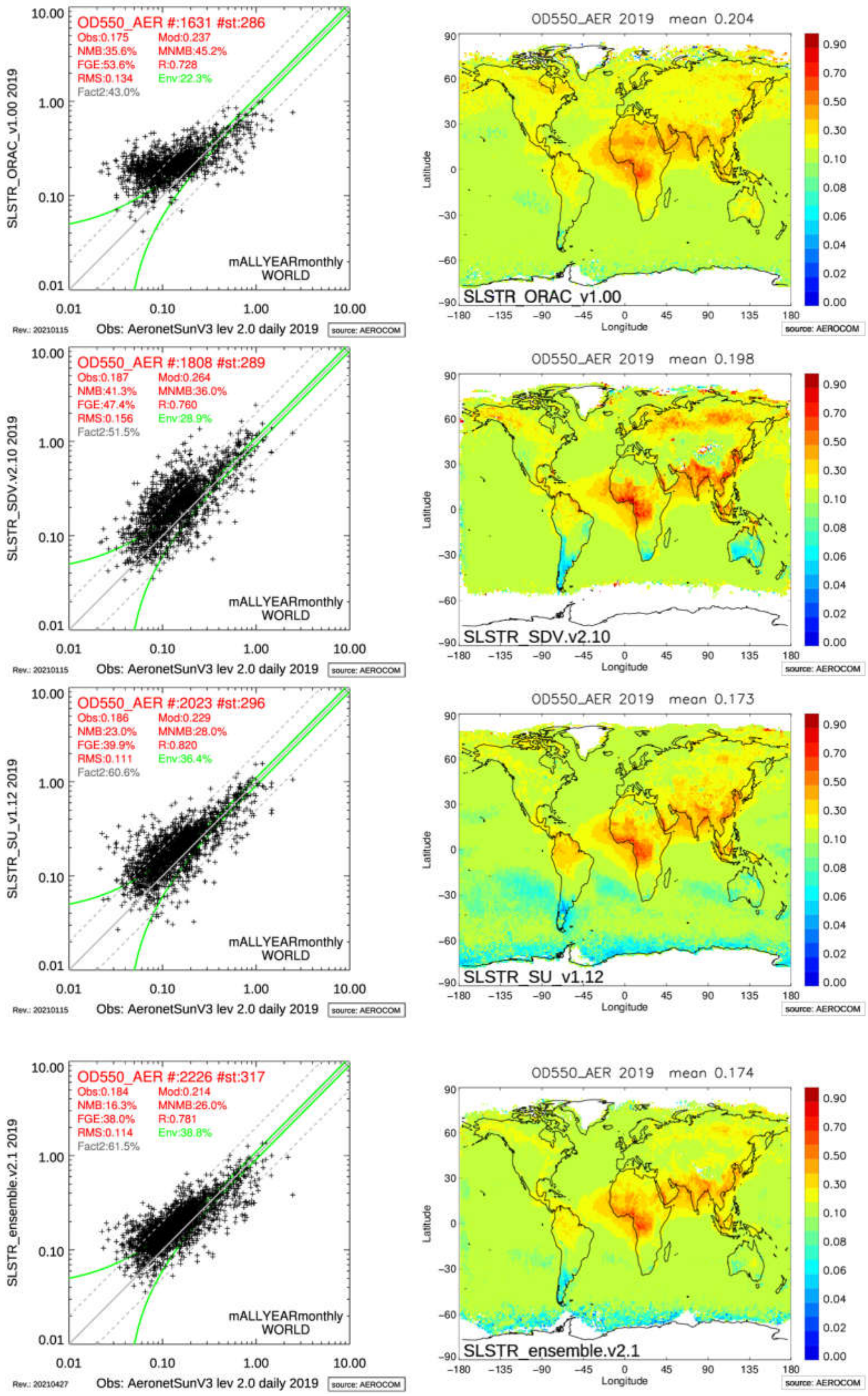


Figure 2.17: same as fig. 2.16 but year 2019

Figure 2.17 and Table 2.8 show the comparison for monthly means of the total AOD for the three SLSTR retrievals and the ensemble for the year 2019. The content is as for the year 2018. The maps show the same patterns as for the year 2018.
Compared to AERONET version 3 level 2.0 data the number of measurement pairs is roughly 20% lower than that for the year 2018. The exact reason for this different number of measurements despite the usage of Aeronet data published in January 2021 is unknown at this point and needs to be further looked upon. It could be due to more clouds during the year 2019 or just some station operators not having submitted their 2019 data yet.
Despite the number of measurements available for comparison, the statistical parameters are close to those of the year 2018 for each retrieval except for the ensemble which is showing a lower correlation coefficient. The SU retrieval is still the one with the best agreement with the AERONET measurements (~20% NMB, correlation of 0.82, RMS of 0.111)

Table 2.8: level 3, year 2019 evaluation statistics for SLSTR total AOD retrievals compared AERONET sites

(vs. Aeronet)	**ORAC**	**SDV**	**SU**	**Ensemble 2.1**
number of pairs	1631	1808	2023	2226
AERONET total AOD avg	.175	.187	.186	.184
retrieval total AOD average	.237	.264	.229	.214
normalized mean bias	35.6%	41.3%	23%	16.3%
mod. norm. mean bias	45.2%	36%	28%	26%
correlation coefficient	.728	.76	.82	.781
root-mean square error	.134	.156	.111	.114

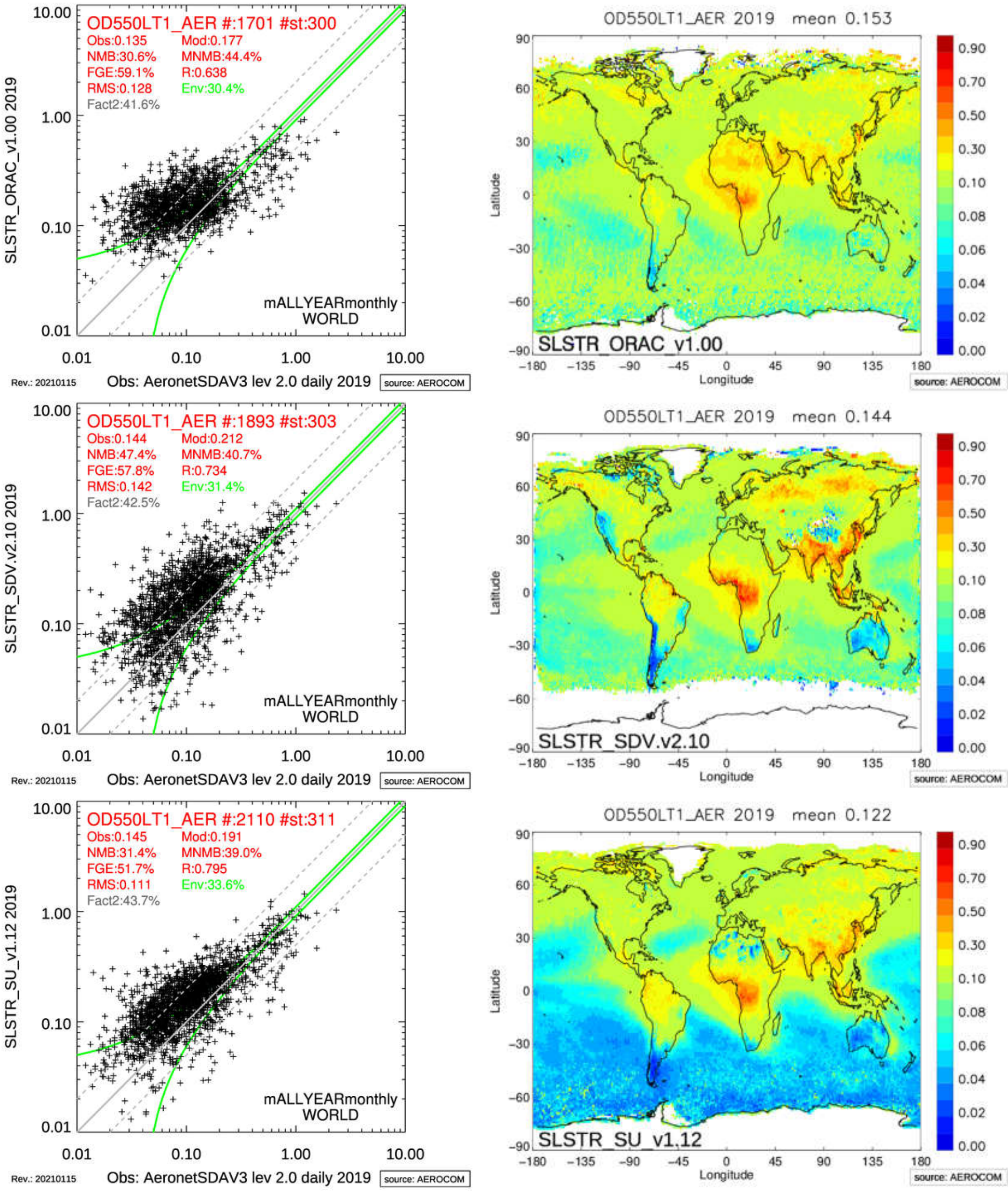
With the most recent retrieval versions, SU and SDV did get much closer to the data quality they had for the AATSR data sets. Especially SU's improvement is quite impressive and is at this point showing the best performance of the three independent retrievals not only regarding the statistical parameters, but also regarding the timely coverage compared to the AERONET stations.

The ORAC retrieval suffers a bit from the fact that it has not updated its aerosol retrieval since the start of the project.

As during the Aerosol-CCI project, the ensemble shows a slightly better performance where the AERONET data set is complete (year 2018).

2.4.2 Fine mode AOD

Figure 2.18 shows the comparison for monthly means of the fine mode AOD for the three SLSTR retrievals ORAC, SDV, SU and the ensemble for the year 2019. The ensemble has not been calculated for fine mode AOD in earlier reports. On the left-hand side is a log scale scatter plot of the retrievals (y axis) against the AERONET SDA version 3 level 2.0 product (x axis). On the right-hand side a plot of the yearly average of fine mode AOD (derived from the monthly means) is shown. The dashed lines indicate a factor of 2 between the satellite and the observation.



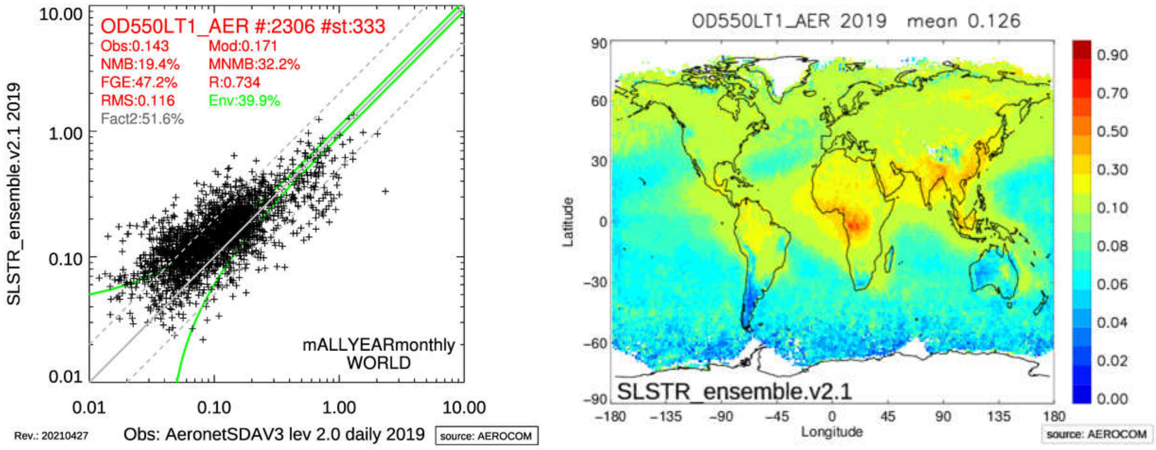


Figure 2.18: fine mode AOD error statistics (left) for different SLSTR retrievals and retrieved fine mode annual AOD (right) for the year 2019.

The map plots show in general a good agreement between the four retrievals in the northern hemisphere with every retrieval having problems in certain areas. ORAC's values in Northern Africa seem a bit too high, SU's values too low. ORAC shows in addition too high values at the Arabian Peninsula. SDV is showing too high values in south east Asia and Siberia and parts of China and has problems in the area of the Gobi dessert.
In the southern hemisphere ORAC and SDV look quite similar, while SU shows much lower values especially over the ocean and Australia.
The ensemble as a combination of all retrievals is in general missing most of the obvious weak points and looks at least visually best.

Compared to the AERONET SDA version 3 level 2.0 data, the retrievals show a similar timely coverage (ORAC 1701 months with data, SDV 1893 and SU 2110). All retrievals are biased high (ORAC: 30.6%, SDV 47.4% , SU 31.4% and the ensemble 19.4% normalized mean bias). Correlation is fair for ORAC (0.64) and good for SDV (0.73), SU (0.80) and the ensemble (0.73). RMS is in general quite high with SU providing the best value of 0.111. All statistical parameters are listed in Table 2.9.

Telling if SU or ORAC and SDV is better in the southern hemisphere is difficult due to the fact that AERONET measurements are sparse in that region. The available five Australian stations with at most 33 months to compare are not enough to make a statistically valid validation (comparison not shown, but available at Met Norway's intercomparison web site). The retrieval with the highest skill there is the ensemble with a NMB of 38.6%, a correlation of 0.77, RMS of 0.03).

Overall the number of months with a valid monthly retrieval mean and an AERONET mean is similar to the one of total AOD in the previous chapter for all retrievals. The bias is also comparable to the one of total AOD while correlation is a bit lower. The statistical parameters do not match those of the latest versions of the AATSR retrievals, but that is expected due to the different measurement geometry. Nevertheless, the fine mode products are ready for scientific use.

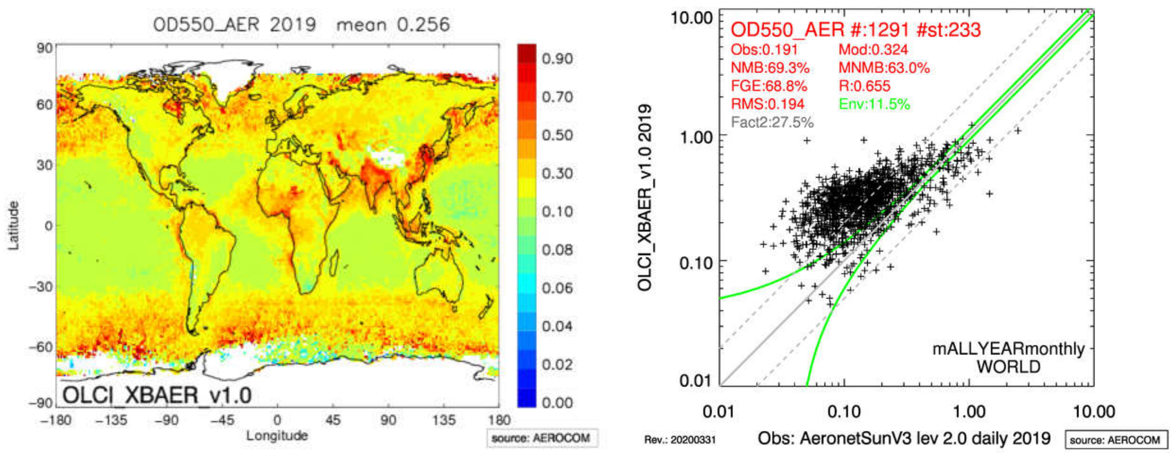
Table 2.9: level 3, year 2019 evaluation statistics for SLSTR fine mode AOD retrievals compared AERONET sites

(vs. Aeronet)	**ensemble**	**ORAC**	**SDV**	**SU**
number of pairs	2306	1701	1893	2110
AERONET fine mode AOD avg	.143	.135	.144	.145
retrieval fine mode AOD average	.171	.177	.212	.191
normalized mean bias	19.4%	30.6%	47.4%	31.4%
mod. norm. mean bias	32.3%	44.4%	40.7%	39%
correlation coefficient	.734	.638	.734	.795
root-mean square error	.116	.128	.142	.111

2.5 OLCI

This report assesses the performance of the OLCI_S4O_v2.0 (SeaWIFS for OLCI; named S4O from this point), from the OLCI_XBAER_v1.0 and from the OLCI ensemble for the year 2019. This is the first look at the S4O version 2 and the OLCI ensemble retrieval.

Figure 2.18 shows a map of the total retrieved annual AOD of the year 2019 for the OLCI_XBAER retrieval (top), the OLCI S4O retrieval , the OLCI ensemble and the SLSTR_SU (as a high quality comparison) retrieval in comparison (bottom) with the annual mean of all retrievals at the left hand side and a scatterplot compared to AERONET version 3 level 2.0 daily means station data on the right hand side.



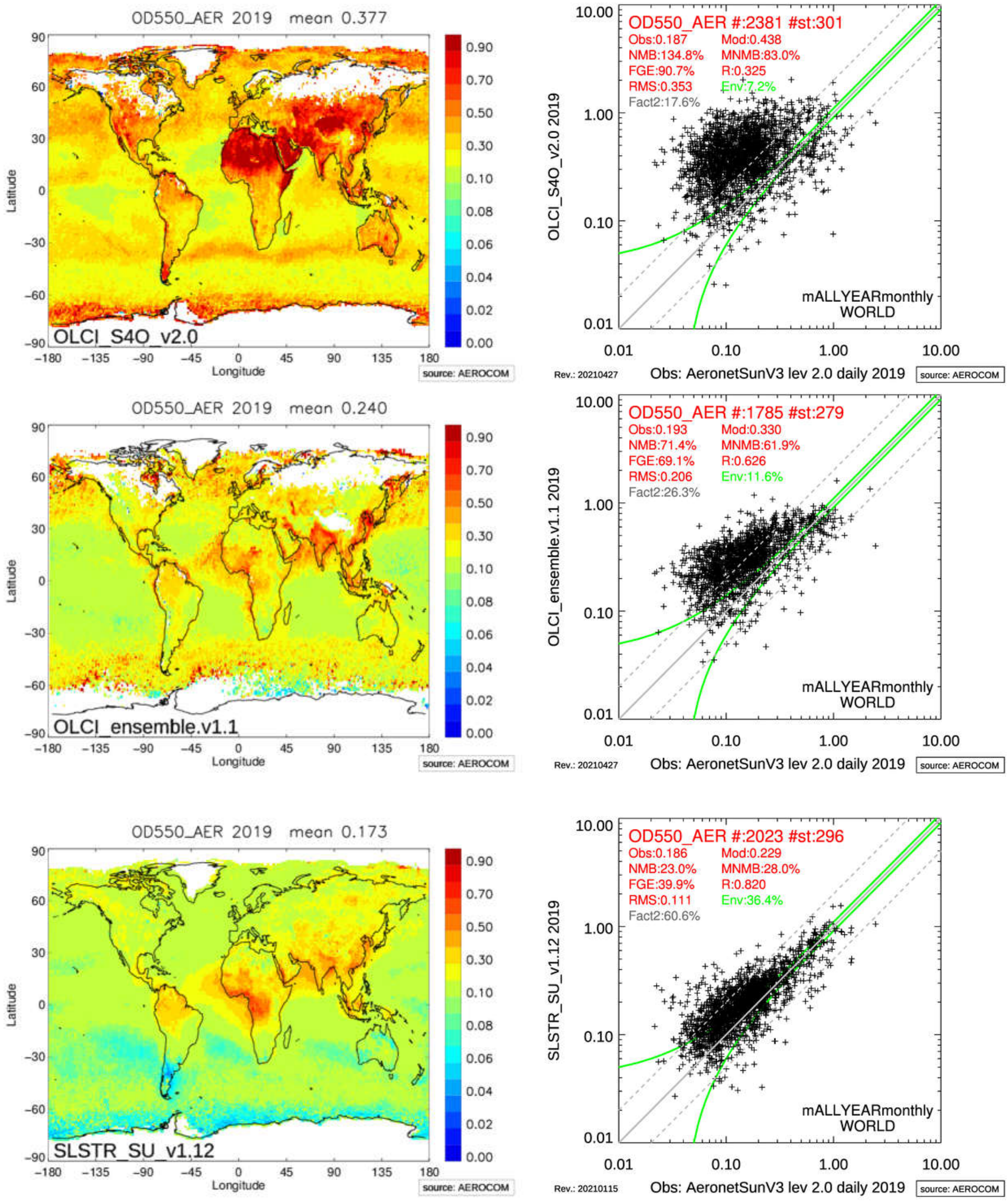


Figure 2.18: OLCI_XBAER, OLCI_S4O and SLSTR SU retrieved total annual AOD (top to bottom left) for the year 2019 and total AOD error statistics (right)

Comparing the maps of the two OLCI retrievals with the one of SLSTR one can see that the OLCI's retrieved values are much higher than those of the known to have good quality SLSTR_SU retrieval. The retrieved values over ocean at latitudes between -30 and -60 and 30 and 70 degrees are far too high for both OLCI retrievals. Over land the difference is not as high as over ocean for the XBAER retrieval, but still visible (e.g. India; Eastern Europe, Australia). The S4O retrieval shows far too high values in the Dust regions (northern Africa, the Arabian Peninsula, parts of Australia) and in eastern Asia. In addition, it is not able to retrieve over land in higher northern latitudes (e.g. wide areas of Siberia, Alaska and a huge part of Canada.

Table 2.10: level 3, year 2019 evaluation statistics for OLCIO total AOD retrievals compared AERONET sites. SLSTR's SU v1.12 retrieval added as comparison.

(vs. Aeronet)	**OLCI ensemble**	**S4O v2**	**XBAER**	**SU v1.12 (SLSTR)**
number of pairs	1785	2381	1986	2023
AERONET total AOD avg	.193	.187	.185	.186
retrieval total AOD average	.33	.438	.322	.229
normalized mean bias	71.4%	135%	74%	23%
mod. norm. mean bias	61.9%	83%	64.7%	28%
correlation coefficient	.626	.325	.674	.82
root-mean square error	.206	.353	.191	.111
GCOS fraction	11.6%	7.2%	9.9%	36.4%

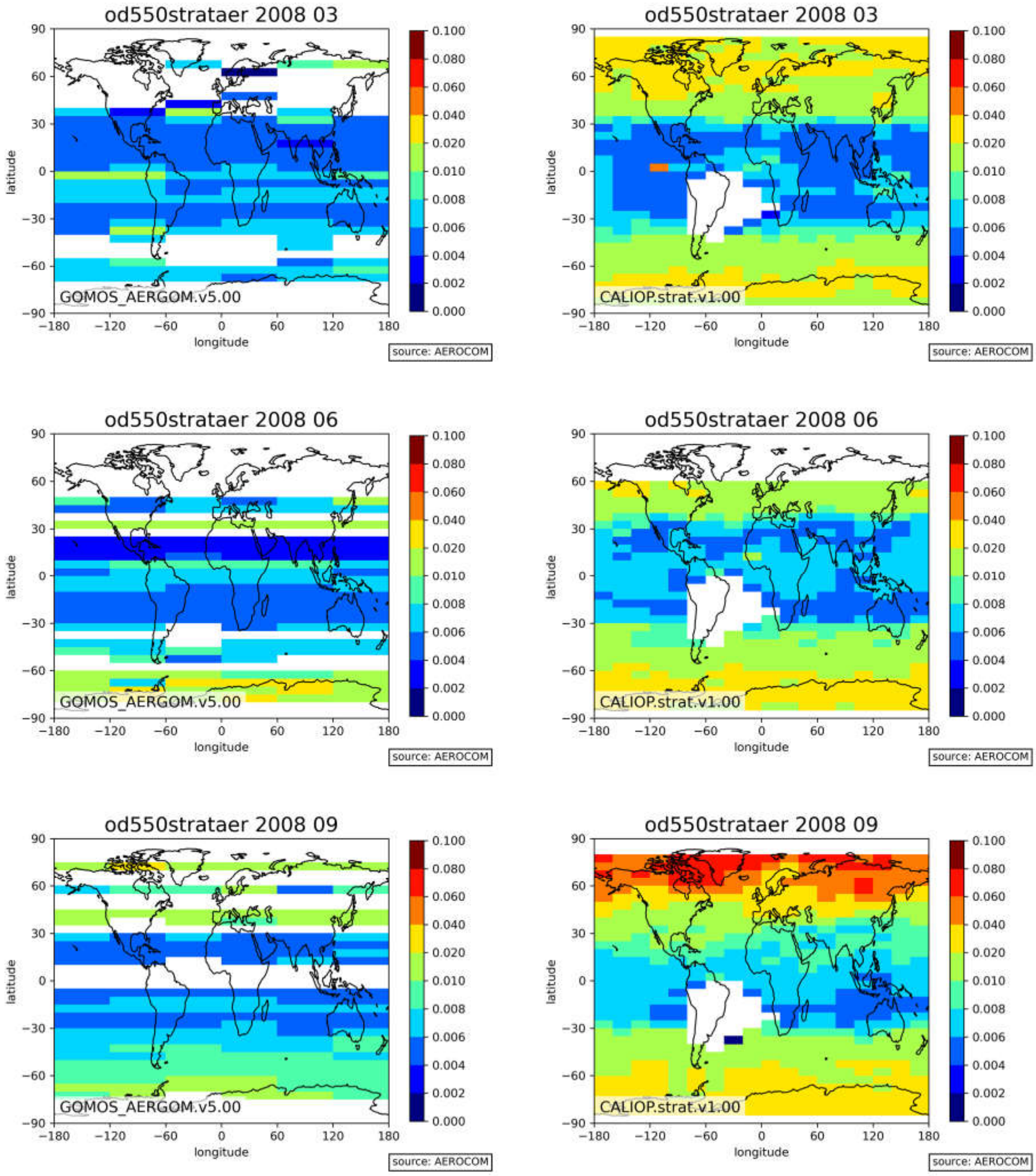
The comparison to AERONET version 3 level 2.0 data shows a very high normalized mean bias for both OLCI retrievals (74% XBAER, 135% S4O, 74% ensemble), a fair correlation of 0.67 for XBAER, and 0.63 for the ensemble but a much lower correlation for S4O, a high root mean square error (0.194 for XBAER, 0.337 for S4O, 0.2 for the ensemble) and a low number of retrievals that lie within the GCOS requirements (10% XBAER, 7.2% S4O, 12% ensemble).
Compared to the initial versions of the SLSTR retrievals, the statistics are comparable for XBAER and therefore a good starting point that shows the potential of the OLCI instrument also for aerosol retrievals. The S4O retrieval is worse than XBAER in nearly every statistical parameter and needs therefore some additional work to be useful. The ensemble brings little overall improvement compared to the sole XBAER retrieval and looks therefore not very helpful in its current version.

2.6 GOMOS

In August 2018 NASA made version 1.00 of the CALIPSO stratospheric aerosol product available. This product also includes a stratospheric aerosol product on a 5 degrees latitude by 20 degrees longitude resolution. Although ESA's Aerosol_cci phase 2 project has proved that a 5 day average product is usable to identify volcanic eruptions, the CALIOP product is provided with monthly resolution only. This somewhat limits the usefulness compared to the 5 day average product from GOMOS, but provides the first data set the GOMOS data can be compared to.

In order to make the five day averages delivered by GOMOS comparable to the monthly product of CALIPSO, the five days have been distributed to five daily averages to form a *daily* GOMOS product. This was then used as input for the analysis tools.

Figure 2.19 shows a map for the months March, June, September and December with the GOMOS data on the left-hand side and the CALIPSO data on the right-hand side.



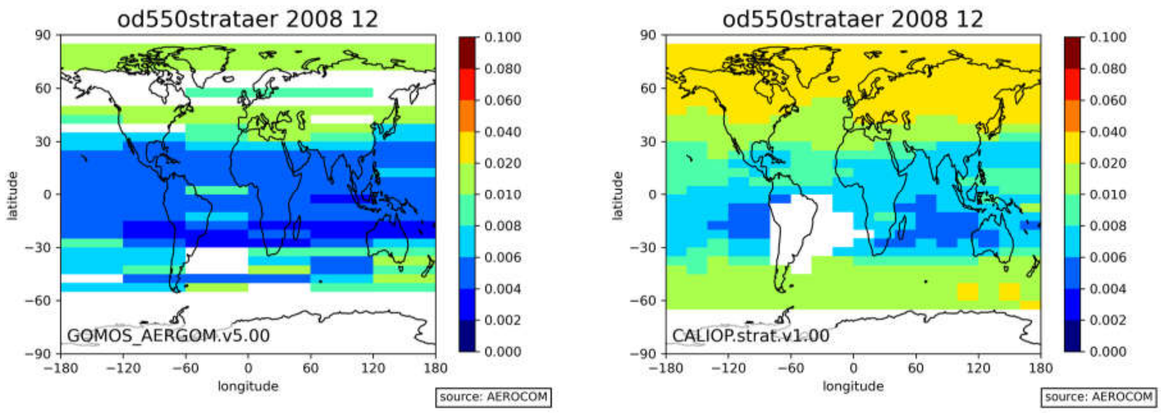


Figure 2.19: 4 months stratospheric AOD from GOMOS (left) and CALIOP (right)

Besides some general commonalities like higher values in higher latitudes and lower values in lower latitudes, the two products have little in common. The values from CALIOP are in general much higher, show more seasonal variation and cover a larger part of the globe.
Unfortunately, the quality of the CALIOP product is unknown at this time. Especially in September the measured values seem too high and are not seen within a reasonable error margin by GOMOS. Knowing that the transfer from the LIDAR measured backscatter to extinction is not trivial and changing over time, the choice of the CALIOP team to use a fixed lidar ratio of 50 for their stratospheric aerosol product is at least questionable. But other reference data sets are not easily accessible at this point.

2.7 Evaluation of stability

To evaluate the stability of the retrieval time series over time a set of reference stations of the AERONET network was identified. These stations needed to have data coverage for at least two years before the year 2001 and two years after the year 2002. In addition, they had to cover major aerosol regimes over the globe. This selection led to 43 stations for the evaluation, but not all stations did provide data all the time during the years 1995 to 2012. Their location is shown in figure 2.20.

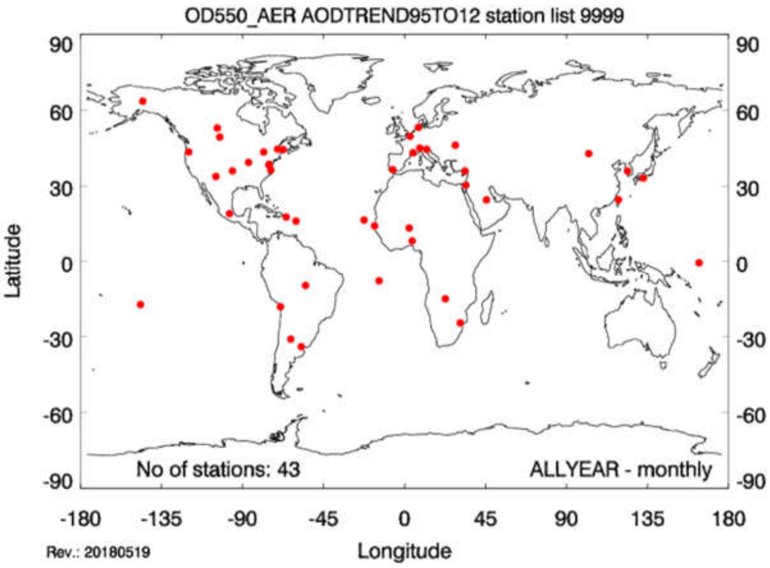


Figure 2.20: Location of the 43 AERONET sites used to evaluate the consistency of the climate record of the ATSR-2 and AATSR retrievals for the total AOD and fine mode AOD

To judge for stability, temporal changes of biases of the three ATSR retrievals were analyzed. To calculate these biases, the AERONET and the satellite data is collocated in space and time.
The resolution of the data products used in this comparison is 1 degree in latitude and longitude and 1 day temporal.

The three retrieval codes do not have the same coverage in space and time and no common point filter was applied to not further reduce data coverage. The daily values of the retrievals and the AERONET data were averaged to a yearly mean and then annual biases were calculated.

In earlier versions of this document Aeronet version 2 data was used. Because that version is no longer available and does not cover the last years for the IASI record Aeronet version 3 data has been used in this report for the stability analysis as well. Although in general there is little difference between Aeronet version 2 and version 3 data, a major difference for the year 1996 was found.

Figure 2.21 shows on the left-hand side the scatterplot of a comparison between the ORAC retrieval of the year 1996 / World region and Aeronet version 2 level 2 data, the right-hand side shows the same plot, but with the Aeronet version 3 level 2 data. The number of data points / station number is 369 / 31 in the version 2 dataset vs only 138 / 16 stations for the version 3 dataset. This basically means that with the version 3 dataset, the number of collocated datapoints between the ORAC retrieval has more than halved and the number of stations providing data has nearly halved by using the version 3 dataset for the year 1996. Although the year 1996 is the most extreme year, it needs to be noted that the Aeronet version 3 data is not the same as the version 2 data. The stability analysis in this report is therefore not comparable with versions in earlier reports.

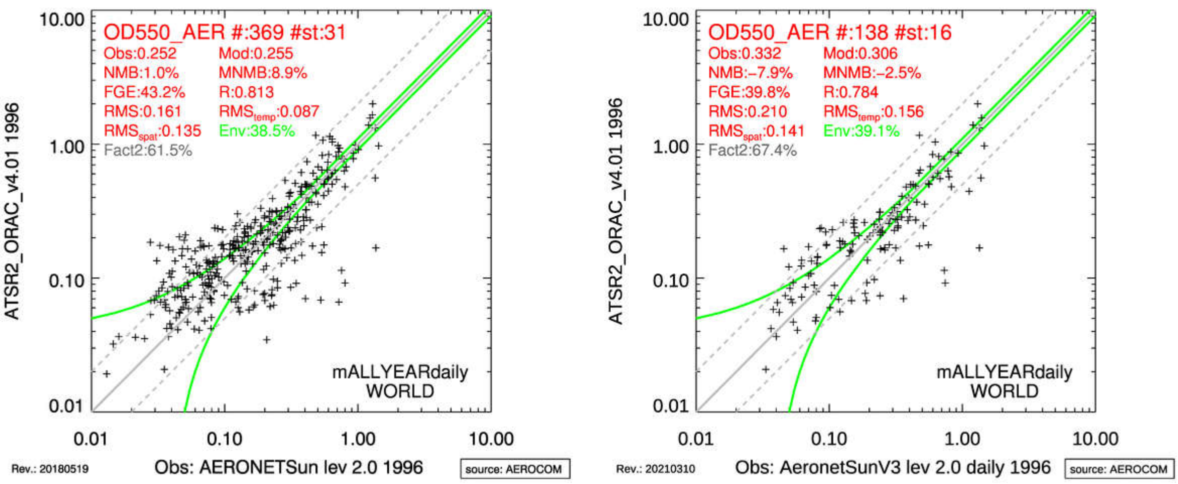


Figure 2.21: yearly comparison of the ORAC retrieval for the year 1996 world region with Aeronet version 2 data on the left-hand side and Aeronet version 3 data on the right-hand side

2.7.1 Stability evaluation of the ATSR record

The ATSR data record consists of data from the ATSR-2 sensor (years 1995 to 2003) and from the AATSR sensor (years 2002 to 2012). To test retrieval stability over time, these data were compared with the above described data record of trusted AERONET stations. Figure 2.22 shows the number of stations per retrieval and year that contributed to this comparison. Before 1999 it is rather limited with less than 20 stations, from the year 2000 until the year 2010 the number of stations is always above 35 per year. Unfortunately, some of the chosen sites did disappear after the year 2010.

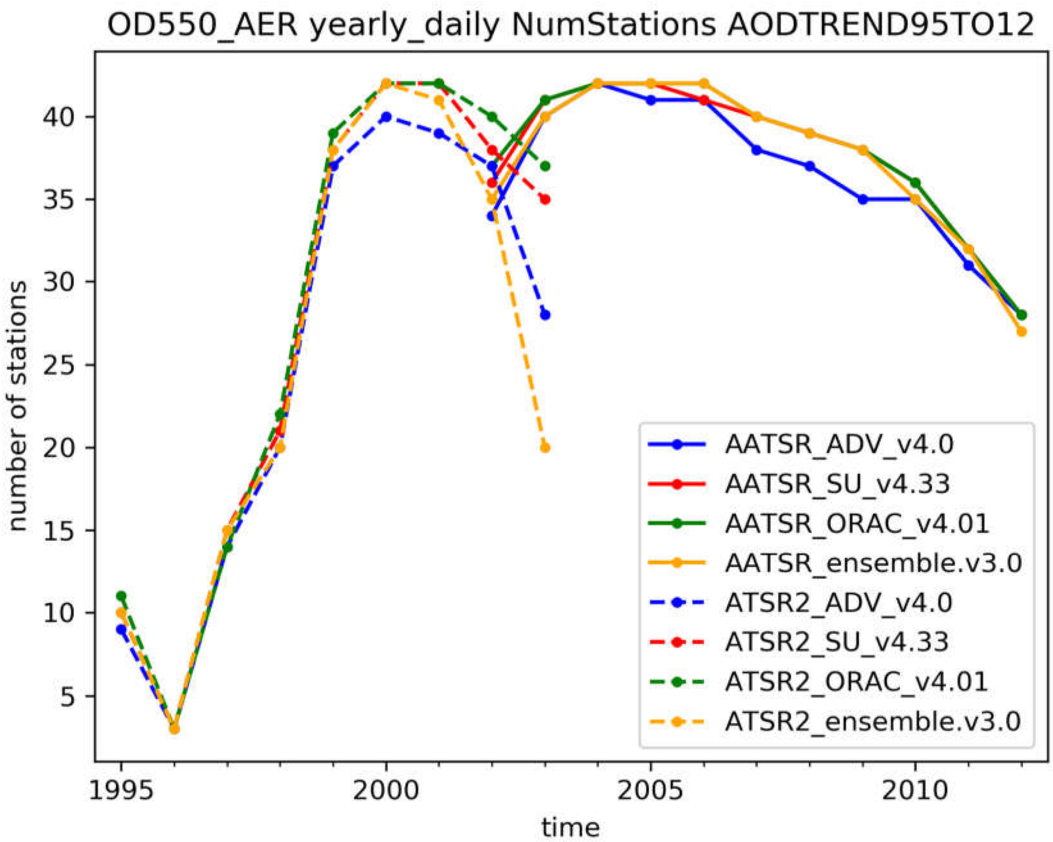


Figure 2.22: number of AERONET sites with at least one match per year. The dashed lines denote the ATSR-2 retrievals, the solid lines the AATSR retrievals. The red lines are the Swansea retrieval, the blue lines the ADV retrieval the green lines are the ORAC retrieval and the orange lines are the retrieval ensemble.

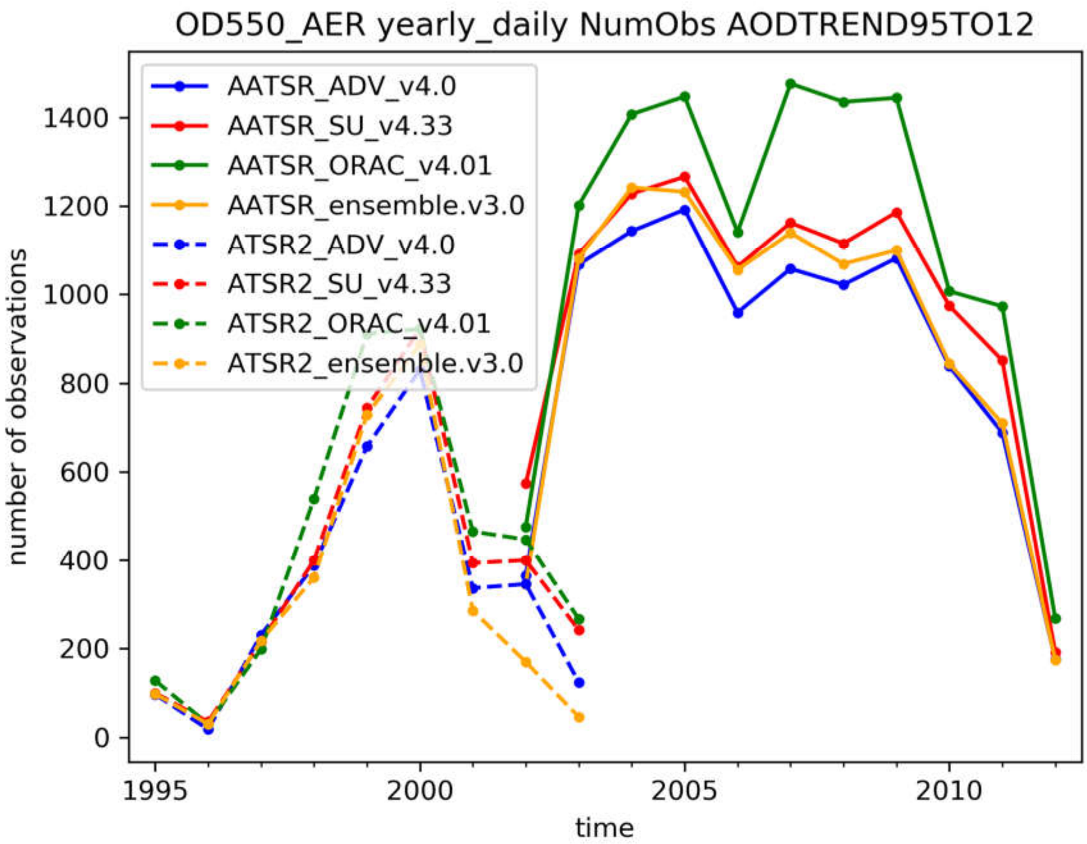


Figure 2.23: total number of spatial and temporal matches per year per retrieval. Line coding as in figure 2.22

Figure 2.23 shows the total number matches per year. It indicates that the AATSR sensor in general has more matches than ATSR-2 due to its wider swath and better coverage. Among the different retrieval codes ORAC nearly always delivers the best coverage. In general, the coverage is best for the years 1998 to 2001 for the ATSR-2 period and between 2003 and 2010 for the AATSR period. It seems a bit strange that the number of coincidences go so much down during the year 2011 where there is still data for the entire year.

Please note that for the years 1995 and 2003 for ATSR-2 and for the years 2002 and 2012 for AATSR there is no data for the entire year. These years are therefore excluded from the analysis.

2.7.1.1 Total AOD stability

The normalized mean bias (MNB; normalized with respect to AERONET) of each year and for each retrieval for total AOD is shown in figure 2.24.

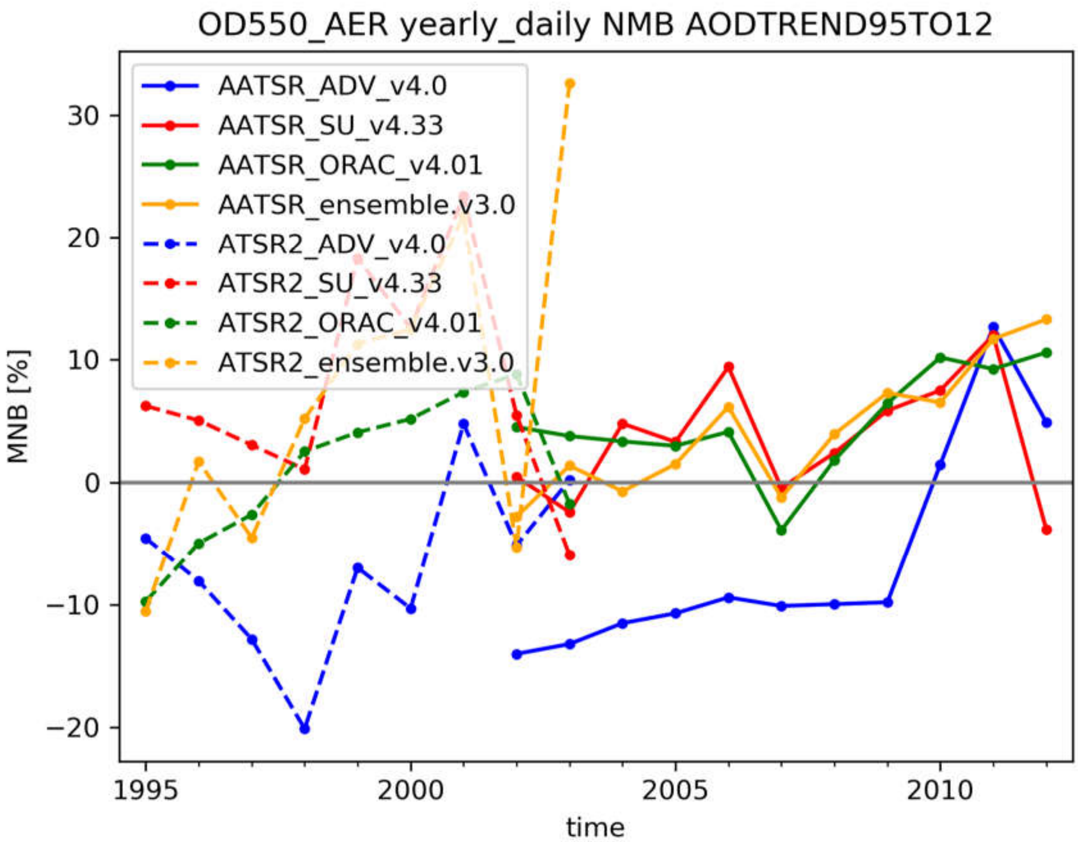


Figure 2.24: time series of the normalized mean bias (MNB) derived from daily values for the ATSR retrievals at up to 43 references sites. Color and line style coding as earlier in this chapter.

The ATSR-2 retrievals show a little higher annual bias variation than the AATSR retrievals. Most stable and therefore most meaningful are the years 1998 to 2010. The overlap period of the years 2002 to 2003 is not really comparable using these yearly statistics since the coverage is not the same because the ATSR-2 and AATSR did not both measure an entire year together.

From this plot one can estimate an upper limit for a trend of the NMB (by stitching the ATSR-2 and AATSR pieces together) assuming that the set of stations over the entire 17 year record is representing the same global mean average (this is not true over the whole period). Visually this quantity is estimated to 30% per 17 years for ADV and around 10% per 17 years for SU, ORAC and the ensemble. These estimates translate (with an average global AOD of 0.15) into stability better than 0.025 / decade for ADV and 0.01 / decade for SU, ORAC and the ensemble.

The monthly NMB plot for the overlap period of ATSR-2 and AATSR at figure 2.25 shows that besides the spike for the ORAC AATSR retrieval in October 2002 and the spike of the ORAC ATSR2 retrieval in November 2002, all retrievals show a similar variation of their results for the ATSR-2 and the AATSR instrument with the AATSR ensemble showing the lowest variation in MNB over the overlap period. The retrieval skill should be considered the same for the ATSR-2 and AATSR instrument.

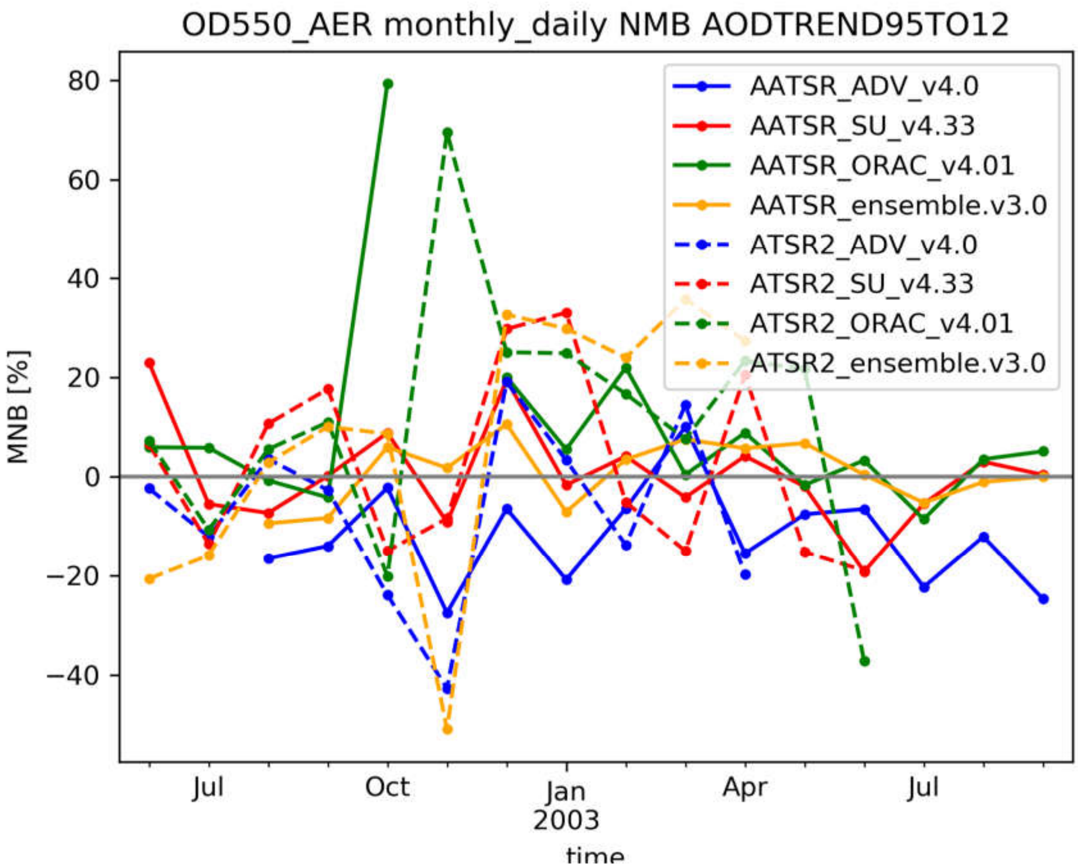


Figure 2.25: monthly evolution of MNB during the overlap time of ATSR-2 and AATSR between June 2002 and June 2003. Dashed lines are the ATSR-2 retrievals, solid lines the AATSR retrievals.

2.7.1.2 Fine mode AOD stability

Figure 2.26 shows the time series of the MNB for fine mode AOD of the ATSR retrievals. The bias variation over the years is higher than for total AOD, but from 1998 when the number of measurements has increased, the tendencies are rather stable until 2010. The much higher bias in 2011 for all retrievals looks unexpected and is therefore excluded from the analysis as is data from the incomplete years.

From this plot one can estimate an upper limit for a trend of the NMB (by stitching the ATSR-2 and AATSR pieces together) assuming that the set of stations over the entire 17 year record is representing the same global mean average (this is not true over the whole period). Visually this quantity is estimated to 30% per 16 years for ADV and around 10% per 16 years for SU, ORAC and the ensemble. These estimates translate (with an average fine mode AOD of 0.09) into stability better than 0.017. / decade for ADV and 0.005 / decade for SU, ORAC and the ensemble.

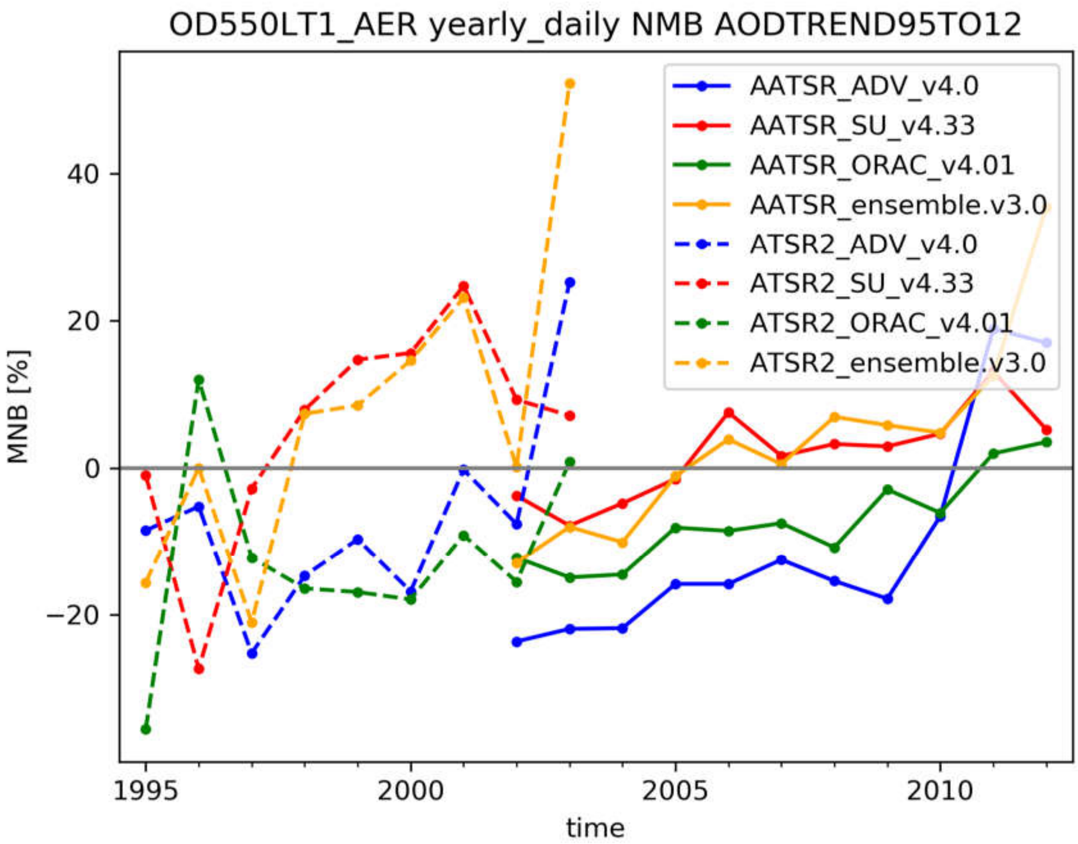


Figure 2.26: time series of yearly normalized mean bias (MNB) for ATSR-2 / AATSR fine mode AOD retrievals at up to 43 AERONET stations.

2.7.2 Stability evaluation of the IASI record

For the stability evaluation of the IASI record the same 43 reference stations were used as for the ATSR record.

Figure 2.27 shows the number of AERONET stations out of the selected 43 long term stations that have conducted at least one daily measurement within the considered year. The figure shows that ULB always provides data at slightly more stations than the other retrievals. The MAPIR retrieval is extremely sensitive to the level 2 data version of IASI (temperature profiles). The data before September 2010 is therefore not usable. To save computation time, BIRA has therefore decided to conduct global retrievals only starting with the year 2011 (and before provide retrievals in the global dust belt only, where good quality IASI L2 temperature profiles are available).

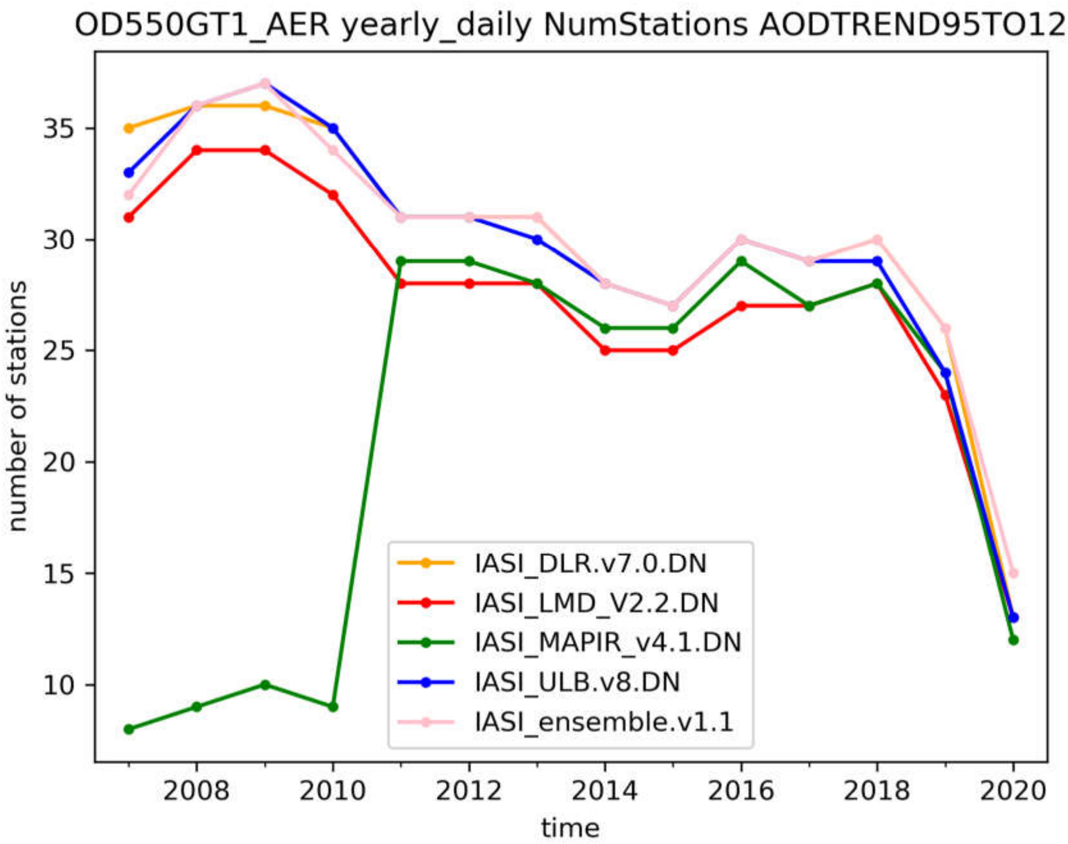


Figure 2.27: number of AERONET sites with at least one match per year. The orange line denotes the DLR retrieval, the red line the LMD retrieval, the blue line the ULB retrieval, the green line the MAPIR retrieval and the pink line the ensemble.

Figure 2.28 shows the number of days per year where the IASI retrievals and the AERONET stations conducted measurements at the same time. The number of measurements per year is stable within a +/- 15% range from the average of the years 2008 to 2018 (only from 2011 onward for MAPIR). For the year 2007 IASI did not measure the whole year, for the year 2020 the retrieval groups did not provide data for the entire year. These years have therefore been excluded from this analysis. For the year 2019 the number of days with measurements per year is roughly 30% lower than for the year 2018 although the data basis used for this analysis has been downloaded in January 2021. This suggests that Aeronet still needs more than a year to get the data of all stations into its database. This year therefore also not considered in the discussion of the stability.

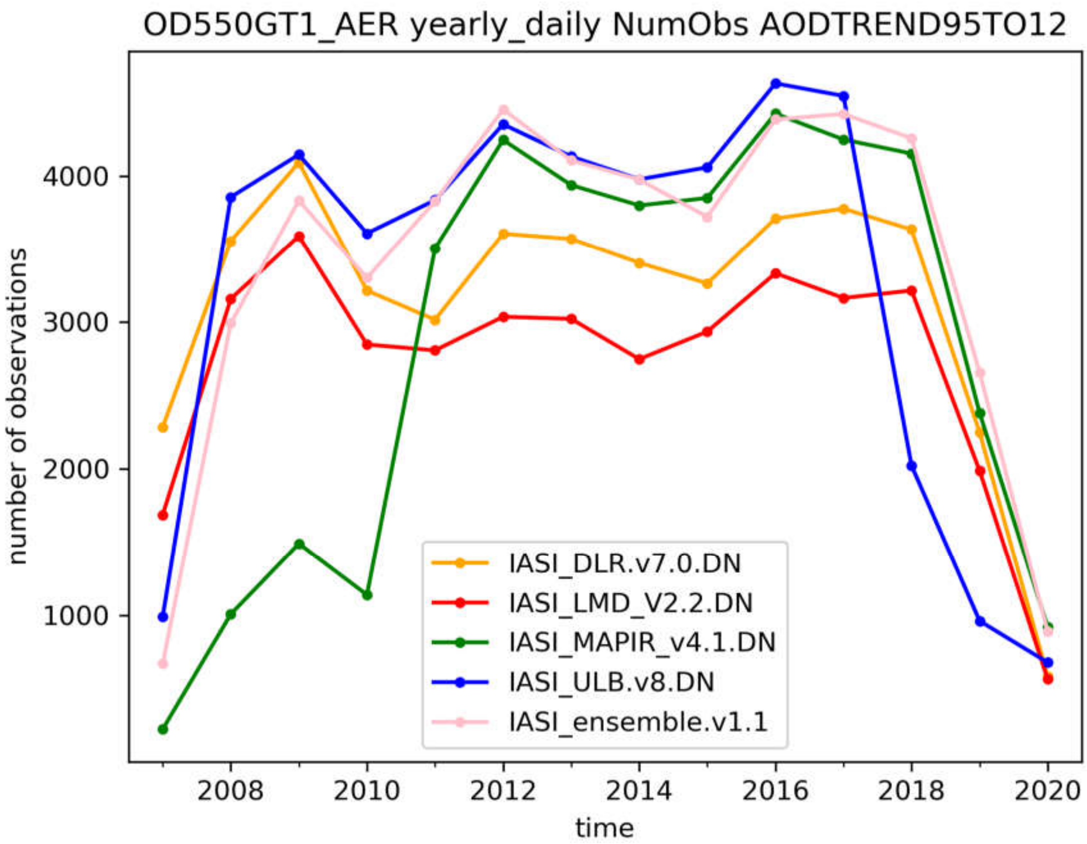


Figure 2.28: Number of days per year where the IASI retrievals and the AERONET stations conducted measurements at the same time. Color coding as before.

Figure 2.29 shows numerical mean bias (NMB) between AERONET SDA version 3 level 2 coarse mode AOD and the IASI dust AOD measurement for each year of the analysis. DLR shows a NMB between -18.1% (year 2018) and -1% (year 2013), LMD shows a NMB between +9.6% (in year 2015) and -44.3% (in the year 2009), MAPIR (considering only years from 2011 onward) shows a MNB between 2.5% (year 2011) and -25% (year 2015) while ULB the NMB varies between -94.8% (in the year 2011) and -1.8% (in the year 2015).
These numbers indicate that considering the years 2008 to 2018 the DLR retrieval is the most stable with a spread of roughly 17% of the NMB for the selected 43 long time stations. MAPIR is (although only from the year 2011 onwards) not far behind with a spread of roughly 22%, while the ULB retrieval has serious stability issues with a very large spread in the NMB of over 90%. The stability of the 2 best IASI dust AOD retrievals is comparable to the total AOD retrievals of the AATSR instrument.

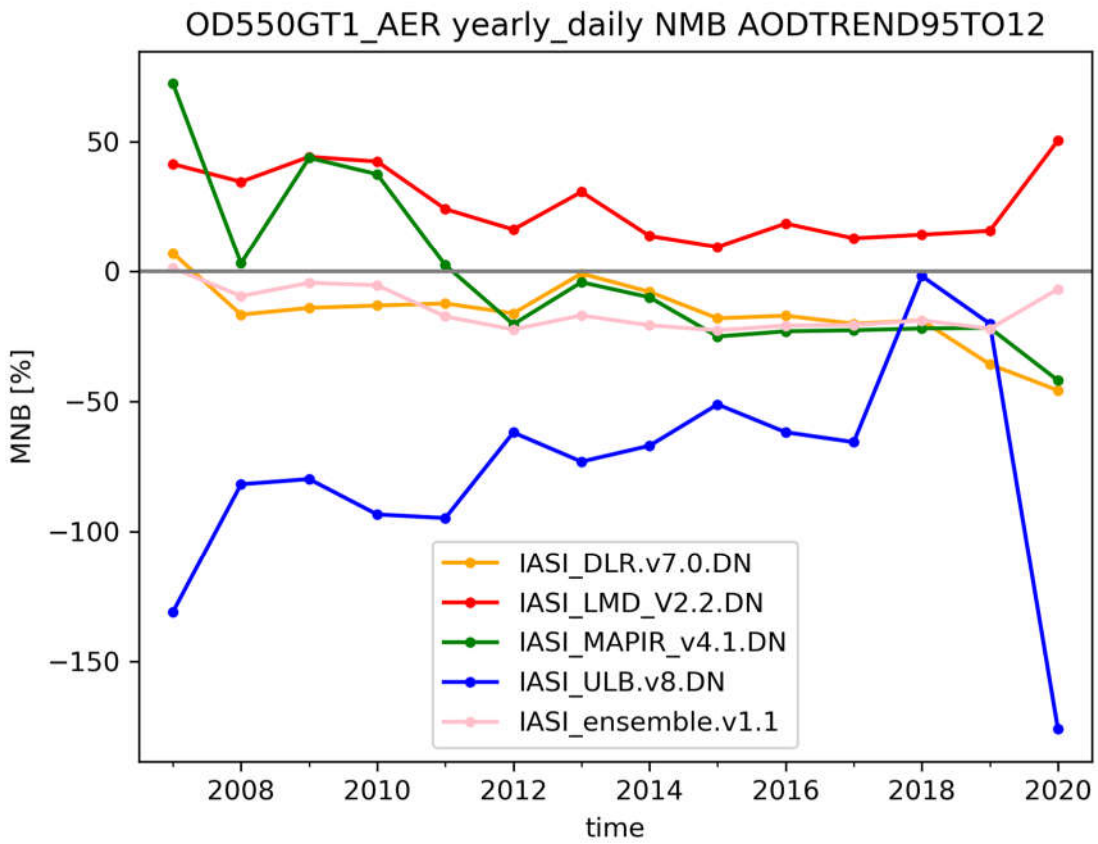


Figure 2.29 Numerical mean bias (NMB) between AERONET SDA coarse mode AOD and the IASI dust AOD measurement. Color coding as before

With a similar analysis as for ATSR, one can estimate the trends of NMB during the decade from 2008 to 2018. The DLR retrieval shows roughly a trend of -10%, the LMD and MAPIR retrieval a trend of roughly -20% and the ULB retrieval of roughly +70%. In AOD quantity (taking 0.06 as global average) this represents a stability of better than 0.006 per decade for the DLR retrieval, 0.012 per decade for the LMD and MAPIR retrieval and 0.04 per decade for the ULB retrieval.

2.7.3 Stability of the SLSTR record

At the time of this writing, the SLSTR retrievals have submitted data from 1. July 2017 to 30 June 2020 leaving only the years 2018 and 2019 as complete years for a stability analysis. Because this time frame is far too short, no stability analysis has been conducted for the SLSTR data record. Nevertheless, similar plots as for the AATSR and the IASI instrument are shown for completeness. Due to the shortness of the dataset, only total AOD is shown in the following analysis.

2.7.3.1 Total AOD

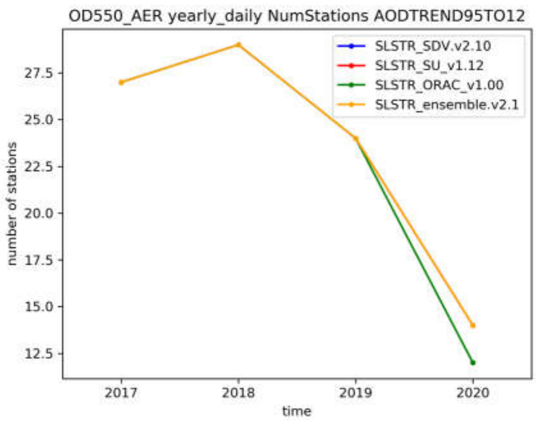


Figure 2.30: number of stations per year and retrieval as indicated in the legend. Note that only the ORAC retrieval has data for a different number of stations than the other retrievals (the orange line has been plotted last).

Figure 2.30 shows the number of stations per year and retrieval. For the years 2018 and 2019 all retrievals provide data at the same number of stations. Unfortunately, the number is significantly lower than for the AATSR retrievals (at max 28 out of 43 for SLSTR while for AATSR there were at max 42 stations present). This suggests that some of the in earlier comparisons selected long term Aeronet stations are not measuring anymore. For future comparisons a new set of stations therefore needs to be selected.

Figure 2.31 shows the evolution of the numerical mean bias of the 4 SLSTR retrievals over time. Compared to the AATRS record, the variation over time is lower than for the AATSR retrievals, but as said before, the time frame is too short to tell that for sure.

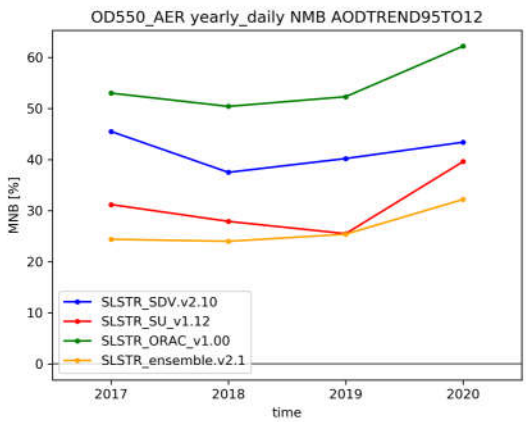


Figure 2.31: Evolution of the numerical mean bias over time for the different SLSTR retrievals

3. Compliance with target service requirements

The following table contains a compliance check of the target requirements in [D3].

Important note (relevant for correct interpretation of quality KPIs): One crucial element of algorithm improvement is to achieve an appropriate balance between the two counter-acting requirements for largest possible coverage (also including pixels with less quality) and for highest possible accuracy (thus reducing coverage), for which the developer team makes different optimal choices. In the quality KPI definition we have aimed at challenging / not yet fulfilled requirements but realisticly within reach over the next decade.

REQ. ID	Requirement content	Compliance - comment
Service definition and scope		
REQ-AER-DEF-1	The C3S products should provide as much as possible the GCOS aerosol variables AOD, SSA, ALH, AEX	Y – by combination of different sensors, algorithms
REQ-AER-DEF-2	The C3S products should provide as much as possible the CMUG aerosol variables AOD (multi-wavelength, several layers), AAOD, depolarization	Y – by combination of different sensors, algorithms
REQ-AER-DEF-3	The C3S aerosol products should as far as possible include following AEROCOM variables: AOD, fine mode AOD, absorbing AOD, stratospheric AOD (and effective radius), dust (or coarse mode) AOD, vertical extinction profiles	Y – by combination of different sensors, algorithms
REQ-AER-DEF-4	The C3S products should contain all information on aerosol type; aerosol types should be matching those used in modelling.	Y – by combination of different sensors, algorithms Matching in parts to the variety of model aerosol types
REQ-AER-DEF-5	The C3S products should provide as far as possible information on aerosol absorption	Y – POLDER product
REQ-AER-DEF-6	The C3S products should provide as far as possible vertical profile information	Y – AEX (GOMOS), ALH (IASI, POLDER)
REQ-AER-DEF-7	For use in aerosol-(water)cloud studies the C3S aerosol products should include fine mode AOD and effective radius	Y- FMAOD
REQ-AER-DEF-8	For use in aerosol-(ice)cloud studies the C3S aerosol products should include coarse mode / dust AOD and dust effective radius	Y- dust AOD
REQ-AER-DEF-9	For use in aerosol-cloud studies the C3S aerosol products should be associated at highest possible temporal resolution with cloud products (joint histograms)	N – out of scope of C3S_312a_Lot5 and C3S_312b_Lot2, which both do not include cloud products
REQ-AER-DEF-10	C3S aerosol and cloud products should be provided together as associated products (joint histograms on regional scale)	N – out of scope of C3S_312a_Lot5 and C3S_312b_Lot2, which both do not include cloud products
Requirements for uncertainty information, accuracy and stability		
REQ-AER-UNC-1	The C3S products should be accompanied with uncertainty estimates for each datum	Y
REQ-AER-UNC-2	Assessments should be made to understand spatial and temporal sampling biases versus reference data	N – is a research need out of scope for C3S
REQ-AER-UNC-3	The C3S products should as far as possible fulfill the GCOS accuracy and stability requirements (table 2.2)	N – highly challenging targets (also determined by trade-off with coverage)
REQ-AER-UNC-4	The C3S products should as far as possible fulfill the CMUG accuracy and stability requirements as function of spatial and temporal grids (table 2.3)	N – highly challenging targets (also determined by trade-off with coverage)
REQ-AER-UNC-5	The C3S aerosol products should fulfill as far as possible the accuracy requirements as function of spatial / temporal resolution (table 2.4 of this TRD)	Partly

REQ-AER-UNC-6	For uncertainties contained in the C3S aerosol products should separate as far as possible random and systematic parts	Y – all systematic corrected as far as known
REQ-AER-UNC-7	Aerosol datasets should as far as possible be bias-free, all known biases should be corrected	Y
Technical requirements: record length		
REQ-AER-TECH-1	The C3S products should provide as long trends as possible with existing satellite instrument records; across subsequent sensors best possible consistency	Y (full mission records) Partly (consistency)
REQ-AER-TECH-2	The C3S stratosphere products should have a record length of at least 10 years and highest possible temporal resolution (~5 days)	Y
REQ-AER-TECH-3	Coverage of C3S aerosol products should be as good as possible to provide sufficient statistics for monthly / seasonal average products (daily or global sampling every few days)	Y
REQ-AER-TECH-4	Record length of C3S aerosol products should be at least 15 years	Y – in some cases 17 years
Technical requirements: sampling, coverage, resolution		
REQ-AER-TECH-5	The service shall have global, multi decadal coverage (minimum 10 years, better several decades, satellite era, up to present), and ensure consistency in the time series.	Y (consistency to be improved)
REQ-AER-TECH-6	The C3S products should provide as large as possible coverage (including high AOD episodes and background levels and covering most continents and oceans)	Y
REQ-AER-TECH-7	The C3S products should comply as much as possible with the MACC (CAMS) data assimilation requirements for temporal sampling, record length, resolution and uncertainty (table 2.5)	Y (except uncertainty)
REQ-AER-TECH-8	The C3S aerosol products should be provided in several resolutions (10km, 100km / 1 degree, 1000 km); vertical resolution should aim to reach 1 km (troposphere) and 0,5 km (stratosphere)	Y – C3S 1 degree gridded data daily / monthly (1000km grid can be derived); 10km underlying data on request
REQ-AER-TECH-9	Level3 gridded products should be provided on 3 horizontal resolutions	N – only 1 degree grid (finer is level2 data, coarser can be derived)
REQ-AER-TECH-10	The C3S products should be provided in lat-lon gridded format with metadata similar to those specified by AEROCOM and CMIP5	Y
Technical requirements: format, metadata		
REQ-AER-TECH-11	The C3S products should be provided in netCDF format with CF compliant metadata	Y
REQ-AER-TECH-12	One (gridded monthly mean for model comparison in AEROCOM) file should only contain one variable and one year of data	N – OGC interoperable interfaces should allow extraction to fulfill user request
REQ-AER-TECH-13	Metadata should be included in the data files of C3S products	Y
REQ-AER-TECH-14	C3S aerosol files should include auxiliary diagnostics (aerosol optical properties assumed or retrieved, surface parameters, cloud fraction) for comparison to models (possibly in separate files)	Y
General requirements: documentation		
REQ-AER-GEN-1	The C3S products should be accompanied with sufficient documentation and a peer reviewed paper as reference	Y in ATBDs, available at the CDS
REQ-AER-GEN-2	The CDS / C3S should provide guidance to users on which products to use for which application and per product on strengths and weaknesses (short concise user guides)	Y – Product User Guide, available at the CDS
REQ-AER-GEN-3	Aerosol types / their optical properties need to be defined precisely and in product related documentation.	Y – in ATBDs, available at the CDS

REQ-AER-GEN-4	The service shall document products by including suitable documentation and ancillary data together with the CDR	Y – in the CDS
REQ-AER-GEN-5	The service shall ensure traceability of provenance of data and workflows, through data collection management tools, linked data principles, and configuration system control.	Y – through metadata in files
	General requirements: data access	
REQ-AER-GEN-6	The C3S datasets should have easy data access (few major upgrades with specific added value)	Y – see work plan with reprocessing every ~2 years
REQ-AER-GEN-7	The C3S should provide stable data access mechanisms and formats to avoid frequent I/O changes on user side	Y – we have adopted the CCI data standards and continue using them in C3S
REQ-AER-GEN-8	Space agencies should support resilience of operational applications by providing operational instrument series	Out of scope for this service
	General requirements: operational	
REQ-AER-GEN-9	C3S should provide capabilities for fast reprocessing	Y
REQ-AER-GEN-10	The service shall set-up a flexible, agile, scalable, and iterative production system capitalizing on existing e-infrastructures and enabling new software and new data streams.	Y – decentralized system
REQ-AER-GEN-11	The service shall ensure a dedicated user support function to deal with user queries interactively including a helpdesk functionality for specialized user queries.	Y – via the CDS and specific level2 user support in C3S_312b_Lot2
REQ-AER-GEN-12	The service shall monitor routinely the performance of the system through a series of benchmarking metrics, covering both the quality of the data, the performance of the system, and the quality of the service.	Y
REQ-AER-GEN-13	The service shall deliver CDRs that meet the target requirements of the ITT (at minimum single-sensor data products, in addition, integrated-multi-sensor high-level products to maximize information content, accuracy, consistency and length of the record).	Y – ensemble dataset
REQ-AER-GEN-14	The CDR shall be continually updated with new data, preferably on a monthly basis, or on an annual basis as an absolute minimum, to enable climate monitoring.	Y - Update on 6-month basis
REQ-AER-GEN-15	Some re-processing of the whole data set shall be foreseen on a routine basis to ensure integration of the latest information and scientific knowledge.	Y - Reprocessing at project end
REQ-AER-GEN-16	The service shall maximize use of satellite data while complementing what is undertaken under other initiatives.	Y
REQ-AER-GEN-17	The service shall deliver gridded high-level products to the CDS in widely accepted standard formats. A suitable adaptor shall be provided.	Y

4. Summary and outlook

This report provides the seventh “routine” evaluation of the satellite aerosol products in the portfolio of the Copernicus Climate Change Service (the first three under the initial contract C3S_312a_Lot5 and four under the operational contract C3S_312b_Lot2). Its goal is not a complete and detailed evaluation, but an overview assessment, which can be applied on a routine basis to identify any major issues in the datasets and its future increments. By combining earlier analysis (scatter plots and stability analysis) of the basic CDR records processed in 9/2018 and new analysis of ICDR extensions processed in 2/2019, 8/2019 and 2/2020 and 8/2020 and the CDRs processed in 11/2020 (also including ATSR-2/AATSR, MERIS, POLDER and GOMOS), the report provides a comprehensive overview of the aerosol product quality available from C3S.

The first evaluation report had still several aspects, where (due to weak reference datasets or due to limited information content) further discussion with the providers is needed to consolidate the validation methodology (for the next report):

- IASI layer height / exact definition which CALIOP reference variable shall be used and then conduct statistical inter-comparison; EARLINET ground-based reference data quality needs to be understood
- POLDER SSA: since neither reference data nor satellite data have much information content at low AOD a filtering of reliable values (AOD > 0.4?) needs to be agreed and either implemented in the products (but then not all pixels do contain SSA values) or during validation (then a new feature needs to be defined in the AEROCOM tools)
- Ångström exponent: since neither reference data nor satellite data have much information content at low AOD a filtering of reliable values (AOD > 0.2 over land, AOD > 0.1 over ocean) needs to be implemented in the products (but then not all pixels do contain AE values) or during validation (then a new feature needs to be defined in the AEROCOM tools)

The second version of the product assessment included following additional aspects:

- Use of stability assessment ATSR and IASI
- Assessment of first 6 month SLSTR datasets
- Upgrades to most recent product versions

The third version of this assessment report (draft) added following analysis

- Addition of recent dataset versions and periods (e.g. newest MAPIR version)
- GOMOS stratosphere evaluation (with reference data available)

The fourth version of this assessment report added following analysis

- Covering all 2018 for the ICDR extensions processed under C3S_312b_Lot2 for SLSTR, IASI and (first time) OLCI datasets
- Using AERONET v3 SDA datasets as reference for dust AOD

The fifth version of this assessment report adds following analysis

- Update to latest versions (SDV v2.00, SLSTR_ensemble v1.2)

- Covering all 2019 for the ICDR extensions processed under C3S_312b_Lot2 for SLSTR (3 algorithms + ensemble), IASI (4 algorithms) and OLCI (for the first time 2 algorithms) datasets
- Using AERONET latest update (April 2020) of Version3 Level2.0 datasets as reference for total AOD.

The sixth version of this report updates the analysis (except the ensembles which will be analysed after their full consolidation / reprocessing with latest input dataset versions in May 2021)

- Updates using AERONET v3
- Updates of the stability analysis (ATSR, IASI).

The seventh version of this report updates the analysis:

- Include the ensemble datasets in the analysis
- Update the GOMOS dataset to version 5.00
- Update the SLSTR SDV dataset to version 2.1
- Update the OLCI S4O dataset to version 2.0