Multi-model evaluation of modelled aerosol optical properties in the AeroCom Phase III Control experiment using ground and space based columnar observations from AERONET, MODIS, and AATSR and surface in-situ observations from GAW sites

Jonas Gliß[1], Augustin Mortier[1], Michael Schulz[1], Elisabeth Andrews[2], Yves Balkanski[3], Susanne E. Bauer[20,19], Anna M. K. Benedictow[1], Huisheng Bian[4], Ramiro Checa-Garcia[3], Mian Chin ?[5], Paul Ginoux[6], Jan J. Griesfeller[1], Andreas Heckel[7], Brent N. Holben ?[5], Stefan Kinne ?[8], Zak Kipling[9], Alf Kirkevåg[1], Harri Kokkola[10], Paolo Laj[11], Philippe Le Sager[12], Robert Levy ?[5], Marianne Tronstad Lund[15], Cathrine Lund Myhre[13], Hitoshi Matsui[14], Gunnar Myhre[15], David Neubauer[16], Twan van Noije[12], Peter North[7], Dirk J. L. Olivie[1], Larisa Sogacheva[17], Toshihiko Takemura[18], Kostas Tsigaridis[19,20], and Svetlana G. Tsyro[1]

[1]Norwegian Meteorological Institute, Oslo, Norway
[2]Cooperative Institute for Research in Environmental Sciences,University of Colorado, Boulder, Colorado, USA
[3]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Gif sur Yvette Cedex, France
[4]Maryland Univ. Baltimore County (UMBC), Baltimore, MD, USA
[5]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[6]NOAA, Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA
[7]Dept. of Geography, Swansea University, Swansea, UK IUP Bremen ??
[8]Max Planck Institute for Meteorology, Hamburg, Germany
[9]European Centre for Medium-Range Weather Forecasts, Reading, UK
[10]Atmospheric Research Centre of Eastern Finland, Finnish Meteorological Institute, Kuopio, Finland
[11]Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, Institute for Geosciences and Environmental Research (IGE), 38000 Grenoble, France
[12]Royal Netherlands Meteorological Institute, De Bilt, the Netherlands
[13]NILU -Norwegian Institute for Air Research, Kjeller, Norway
[14]Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan
[15]CICERO Center for International Climate and Environmental Research, Oslo, Norway
[16]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[17]Finnish Meteorological institute, Climate Research Program, Helsinki, Finland
[18]Research Institute for Applied Mechanics, Kyushu University, 6-1 Kasuga-koen, Kasuga, Fukuoka, Japan
[19]Center for Climate Systems Research, Columbia University, New York, USA
[20]NASA Goddard Institute for Space Studies, New York, USA

Correspondence: Jonas Gliß(jonasg@met.no)

Abstract. Within the framework of the AeroCom (Aerosol Comparisons between Observations and Models) initiative, the present day modelling of aerosol optical properties has been assessed using 2010 simulated data from 14 global aerosol models participating in the Phase III Control experiment. Modelled column optical-depths (AOD, AOD $< 1\,\mu$m, AOD $> 1\,\mu$m) and Ångström Exponents (AE) were compared both with ground based observations from AERONET (version 3) as well as space based observations from ATSR-SU instruments. In addition, the total AODs were compared with MODIS (aqua and terra)

data and a satellite AOD data-set (MERGED-FMI) merged from 12 different instruments. Furthermore, for the first time, the modelled surface scattering (under dry conditions) and absorption coefficients were evaluated against measurements from in-situ GAW sites, which have highest site density in Europe and the US.

Prior to the assessment of the models, the individual satellites were evaluated against AERONET observations, suggesting relative AOD biases of -5%, -6%, +9% and +18% for ATSR, MERGED-FMI, MODIS-aqua and MODIS-terra, respectively with consistently high correlations exceeding 0.8. In addition, biases of fine and coarse AOD and AE in ATSR were found to be +2%, -16% and +14.7% respectively, at AERONET sites, with correlations of the order of 0.8. These (relative) biases in the satellite data are mostly reflected when evaluating the models against these satellites, since the satellites show similar spatial coverage.

The results of the AeroCom MEDIAN show that overall, the participating models underestimate all optical variables investigated. AOD is underestimated by -21% $\pm$ 17% against AERONET. Compared to satellites, the model AOD biases are ranging from -38% (MODIS-terra) to -17% (MERGED-FMI). Correlation coefficients of AOD are high against AERONET, MERGED-FMI and ATSR-SU (0.8 - 0.9) and slightly lower when comparing the models with the two MODIS data-sets (0.6 - 0.8).

Investigation of fine and coarse AODs from the MEDIAN model reveals biases of -10% $\pm$ 20% and -41% $\pm$ 29% against AERONET and -13% and -24% against ATSR-SU, respectively. The differences in bias between AERONET and ATSR-SU are in agreement with the established satellite biases. These results indicate that most of the AOD bias is due to missing coarse particles or due to underestimations in the extinction efficiencies of coarse particles.

Evaluation of modelled column AEs shows an underestimation of -9% $\pm$ 24% against AERONET and -21% against ATSR-SU. This suggests that overall, models tend to overestimate particle size. This may have important implications for lifetime and hence, transport of the aerosol.

Considerably high underestimations are also found when comparing the models against the surface GAW observations, showing MEDIAN biases of -44% $\pm$ 22% and -32% $\pm$ 34% for scattering and absorption, respectively. The fact that dry scattering shows higher underestimation than the AOD comparison (at ambient RH) is in agreement with recent findings that suggest that models tend to overestimate scattering enhancement due to hygroscopic growth. The large diversity in the surface absorption results suggests differences in the treatment of absorption optical properties of black carbon (BC), dust and organic aerosol (OA). An investigation of the modelled diversity of surface absorption indicates, that regions associated with dust (e.g. Sahara, Tibet), biomass burning (e.g. Amazonia, Central Australia) and biogenic emissions (e.g. Amazonia) are mostly accounting for this disagreement between the models, while models tend to agree in regions associated with high anthropogenic BC emissions such as China or India.

An investigation of modelled emissions, burdens and lifetimes, MECs and optical depths for each species and model, reveals considerably large diversity in most of these parameters. These are discussed in detail for each model individually and possible reasons are provided, that may explain the results from the optical properties inter-comparison.

1 Introduction

The global aerosol remains one of the largest uncertainties for the projection of future Earth's climate, in particular because of its impact on the radiation balance of the atmosphere (IPCC (2014)). Aerosol particles interact with radiation through scattering and absorption, thus directly altering the atmosphere's radiation budget (aerosol-radiation interactions, or ARI). Moreover, they serve as cloud condensation nuclei (CCN) and can thus, among other things, influence further climate relevant components such as clouds and their optical properties (e.g. cloud droplet number concentrations, cloud optical depth) and lifetime as well as cloud coverage and precipitation patterns (aerosol-cloud interactions, or ACI) (IPCC (2014)).

A challenging part of modelling the global aerosol is its comparatively high variability in space and time, as compared to well-mixed greenhouse gases such as carbon dioxide and methane. The radiative impact aerosols exert depends on the amount and the properties of the aerosol. Emissions and lifetime combined lead to different amounts of aerosol in transport models. The lifetime of aerosol particles in the atmosphere is of the order of one week and is, to first order, dependent on their size. Particles in the accumulation mode (particle diameter between 0.3–1 μm) show the longest residence times due to less effective atmospheric sink processes. The sources of aerosol are complicated since not all aerosol particles are directly emitted. Instead, particles can also be formed in the atmosphere (secondary aerosol) which is dealt with in various degrees of complexity in models (e.g. Tsigaridis et al. (2014)). Both natural and anthropogenic emissions are highly uncertain due to lack of measurements and information or documentation flow.

Natural aerosols constitute a large part of the atmospheric aerosol, being composed of sulphur and organic components, as well as sea salt and dust. Emissions of sea salt and dust are strongly dependent on local meteorology and surface properties and, thus, require sophisticated parameterisations in global models with comparatively coarse resolution. In models, these emissions are usually computed based on simulated winds and constitute a major source of uncertainty. Marine DMS and volcanic emissions are responsible for approximately a third of the global anthropogenic sulphur budget. Both eruptive and passively degassing volcanic sulphur emissions are highly uncertain, with estimates ranging between $1 - 50\,\mathrm{Tg}$ (e.g. Andres and Kasgnoc (1998), Halmer et al. (2002), Textor et al. (2004), Carn et al. (2017)). In addition, atmospheric aerosol particles undergo continuous alteration (e.g. growth, mixing) due to micro-physical processes that occur on lengths and timescales that cannot be resolved by global models, such as nucleation or gas-to-particle conversion.

The chemical and physical properties of aerosol particles determine how they interact with radiation. They are highly dependent on the aerosol type and state of mixing. Aerosol optical properties such as the aerosol scattering and absorption coefficients, the aerosol optical depth (AOD) and the Ångström exponent (AE) are closely linked with aerosol forcing estimates as they determine how aerosols interact with incoming and outgoing long and shortwave radiation. A key parameter that determines the efficiency of scattering and absorption of radiation is the complex refractive index ($n + ik$), which depends on aerosol type (chemical composition) and mixing. It is dealt with in models in different ways (e.g. volume mixing, Maxwell/Garnett, core shell, e.g. Klingmüller et al. (2014)). The absorptive properties of dust aerosol, for instance, are depen-

dent on the mineralogy of the dust particles, resulting in some dust types being more absorptive than others (e.g. Lafon et al. (2006)), which has direct implications for forcing estimates (e.g. Claquin et al. (1998)).

Scattering and absorption coefficients are derived from these extinction efficiencies and depend on particle size distribution and wavelength. In general, water uptake will enhance the light extinction efficiency. This is mostly relevant for scattering, since absorptive aerosols such as dust and black carbon are generally considered to be hydrophobic (which can, to a minor degree violated in aged aerosol due to mixing, e.g. Cappa et al. (2012)). For instance, between 0% and 40% relative humidity (a range that is often referred to as "dry"), the light scattering can be significantly enhanced due to hygroscopic growth. This is important when comparing models with in-situ observations, since the latter are often performed under low humidity but not at absolutely dry conditions. Some models tend to overestimate the enhancement factor at low RH (and high RH) and hence, the amount of light scattering (Burgos et al., in prep.).

The AOD comprises the vertically integrated light extinction (absorption + scattering) due to an atmospheric column of aerosol and is a function of wavelength. The AAOD is the corresponding equivalent for the absorptive power of an aerosol column and tends to be small compared to AOD (5-10% of AOD). Both AOD (mostly scattering) and AAOD (absorption) are of particular relevance for aerosol forcing assessments (e.g. Bond et al. (2013)). Major absorbing species are black carbon (BC), followed by dust (DU) and, to a certain degree, organic aerosols (OA) (e.g. Samset et al. (2018) and references therein).

Simulating the AOD (and AAOD) in a global model is hence, a challenging task as it requires many prerequisites to be correct, not only the assumptions on optics (e.g. shape and refractive index, atmospheric radiative transfer), but also the emissions, transport, ageing, sources and sinks of all aerosol species, which determine the aerosol composition in space and time. Therefore, it is useful to also investigate other optical (and related) parameters that can help to assess model performance. The AE, for instance, describes the wavelength dependency of aerosol extinction and is related to the size of the aerosol (i.e. larger particles exhibit less spectral dependence of scattering, resulting in smaller value of the AE). It can thus, provide a qualitative assessment of modelled particle size (e.g. Schuster et al. (2006)). For instance an underestimation of AE suggests an overestimate of the contribution of coarse particles. Like AE, fine and coarse mode AOD can also give insights into the particle size domains, which can help establish differences between natural and anthropogenic aerosols (since the major natural constituents, dust and sea salt, dominate the coarse mode AOD).

Kinne et al. (2006) provided a first analysis of modelled column aerosol optical properties of 20 aerosol models in the initial AeroCom experiments. They found that, on a global scale, modelled aerosol optical depth (AOD) compares well with observations (model biases of the order of **NUM**). However, they also found considerable diversity in the aerosol speciation among the models, particularly for absorption related aerosol species such as dust and carbonaceous aerosols. They concluded that this adds uncertainty related to estimates of the direct forcing effect.

This study investigates modelled aerosol optical properties of the most recent models participating in the AeroCom 2019 control experiment (in the following denoted CTRL, https://wiki.met.no/aerocom/phase3-experiments) on a global scale. Making use of the increasing amount of data which have become available during the past decade, we are able to extend the assessment of modelled optical properties beyond what was originally presented in Kinne et al. (2006). Here, we use observations of ground and space based observations of the above introduced columnar variables of total, fine and coarse AOD and AE as well

as, for the first time, surface in-situ measurements of scattering and absorption coefficients, primarily from surface observatories contributing to Global Atmospheric Watch (GAW), obtained from the World Data Centre for Aerosols (GAW-WDCA) archive.

PARAGRAPH THAT INTRODUCES THE STRUCTURE OF THE PAPER

2 Methods

2.1 Observations and variables

Several ground and space-based observations have been utilised in order to establish a comprehensive evaluation at all scales. Table 1 summarises all variables and observation networks that have been used. They will be introduced in more detail below.

Fig. 1 shows yearly average mean values of the observed AERONET AODs and column extinction Angstrom exponents. Dust dominated regions such as Northern Africa and Southwest Asia are clearly visible both in the coarse AOD and the AE, but also in the total AOD, indicating its importance for the global AOD signal due to dust. The displayed satellite fields of AOD (MERGED-FMI) and AE (ATSR-SU) are particularly useful in remote regions and over the oceans where ground based measurements are less common, and, thus, add substantially to the global picture when assessing models. For example, satellites capture the nearly constant ocean AOD background of around 0.1 (mostly arising from sea salt) which is not really measured by the land dominated, ground-based observation networks. The AE from ATSR-SU for instance, shows a latitudinal southwards gradient in remote ocean regions, indicating coarse(r) particle sizes, likely attributed to cleaner and, thus, more sea salt dominated regions. Transatlantic dust transport results in an increased particle size west of the Sahara (e.g. Kim et al. (2014)) as is captured by ATSR-SU. Finally, as can be seen in the lowermost panel of Fig. 1, in-situ sites from GAW show highest density in Europe, followed by North America, while other regions are not represented well. This is important to keep in mind for the assessment of the model evaluations against the different observation records.

The following subsections introduce briefly each of the observation data-sets used.

2.1.1 AERONET

The Aerosol Robotic Network (AERONET, Holben et al. (1998)) is a ground-based remote sensing network based on sun photometers and comprises a well established data-set that has been extensively tested and used over the last two decades.

In this paper, cloud screened and quality assured daily aggregates of AERONET AODs (total, fine, coarse) and AE from the version 3 (level 2) Sun and SDA products (e.g. O'Neill et al. (2003), Giles et al. (2019)), have been used to assess the model performance at AERONET locations. No further quality control measures have been applied due to the already high quality of the data.

For the analysis, the spectral AOD values were used to derive an AOD at 550 nm using the provided AE. Data from the DRAGON campaigns (Holben et al. (2018)) was excluded and no further site selection has been performed. However, a

sensitivity study was performed to investigate potential spatial representativity issues associated with some AERONET sites but the impact was found to be of minor relevance for this study (for details see Sect. 3.4 and Fig. A5).

2.1.2 Surface in-situ data

Surface in-situ measurements of the aerosol light scattering and absorption coefficients, were accessed through the GAW-WDCA database EBAS (http://ebas.nilu.no/). The EBAS database also includes various observations of atmospheric chemical composition and physical parameters, although those were not used here. For both scattering and absorption variables, only level 2 data from the EBAS database were used (i.e., quality controlled, hourly averaged, reported at STP). All data in EBAS have version control, and a detailed description of the quality assurance and quality control procedures for GAW aerosol in-situ data are available in Laj et al. (2020). Additionally, for this study data was only considered if it was associated with the EBAS categories *aerosol* or *pm10*. The *aerosol* category indicates the aerosol was sampled using a whole air inlet, while *pm10* indicates the aerosol was sampled after a 10μm aerodynamic diameter size cut. It was assumed whole air and *pm10* would provide the better comparison with model simulations than measurements with smaller cut size (e.g., *pm2.5* or *pm1*). Invalid measurements were removed based on values in the flag columns provided in the data files. Furthermore, outliers were identified and removed using value ranges of $\{-10, 1000\}\,\mathrm{Mm}^{-1}$ and $\{-1, 100\}\,\mathrm{Mm}^{-1}$ for scattering and absorption coefficients, respectively. The outliers were removed in the original 1h time resolution before averaging to monthly for comparison with the monthly model data. For most of the absorption data, the measurements are performed at wavelengths other than 550 nm. These were converted to 550 nm assuming an absorption Angstrom exponent (AAE) of 1 (e.g. Bond and Bergstrom (2006)). For the scattering coefficients, only measurements at RH$\leq$40% were considered. For the model evaluation, the 2010 monthly model data was converted to STP using the following formula:

$$X_{\mathrm{STP}} = X_{\mathrm{AMB}} \times \left(\frac{p_{\mathrm{STP}}}{p_{\mathrm{AMB}}}\right) \cdot \left(\frac{T_{\mathrm{AMB}}}{T_{\mathrm{STP}}}\right) \tag{1}$$

where p_{STP} and T_{STP} are standard IUPAC standard pressure and temperature, and p_{AMB} and T_{AMB} are air pressure and temperature at the corresponding site location. The correction was performed on a monthly basis using the station altitude to estimate the pressure and monthly near surface (2m) temperature from ERA5 (CITE).

A few urban sites were removed from consideration for the model analysis, as these sites are likely not representative on spatial scales of a typical model grid. These sites are:

Scat. coeff.: Granada; Phoenix; National Capitol - Central, Washington D.C

Abs. coeff.: Granada; Leipzig Mitte; Ústí n.L.-mesto

The biases of each model for individual in-situ sites are shown in Appendix Figs. A1 and A2 for scattering and absorption, respectively.

2.1.3 MODIS data

Daily gridded level 3 AOD data data from the Moderate Resolution Imaging Spectroradiometer (MODIS), has been used from both satellite platforms (Terra and Aqua) for evaluation of the models. The merged land and ocean global product (named Aerosol_Optical_Depth_Land_Ocean_Mean in the product files) of the recent collection 6.1 was used. This is an updated and improved version of collection 6 (e.g. Levy et al. (2013), Sayer et al. (2014). For changes between both datasets, see Hubanks (2017)).

2.1.4 (A)ATSR SU v4.3 data

The ATSR v4.3 SU dataset provides gridded AOD and associated parameters from the ATSR instrument series, developed by Swansea University (SU) under the ESA Aerosol Climate Change Initiative. The ATSR-2 instrument was hosted on the ERS-2 satellite, and provides a record from 1995-2003, with its successor instrument AATSR on ENVISAT covering the period 2002-2012. The instrument's conical scan provides two near simultaneous views of the surface, at solar reflective wavelengths from 555nm to 1.6 um.

Over land, the algorithm uses the dual-view capability of the instrument to allow estimation without a priori assumptions on surface spectral reflectance (North (2002), Bevan et al. (2012)). Over ocean, the algorithm uses a simple model of ocean surface reflectance including wind-speed and pigment dependency at both nadir and along-track view angles. The retrieval finds an optimal estimate of the AOD at 550 nm, and aerosol model. Further aerosol properties including AE and AAOD (not used in this study) are determined from this model. The aerosol model is found by optimal estimation of the fine and coarse mode fractions, where local composition of fine and coarse mode is given by a climatology (Kinne et al. (2013)). Aerosol is retrieved over all snow-free and cloud-free surfaces. The most recent version SU ATSR V4.3 (North and Heckel (2017)) advances on previous versions by improved surface modelling and shows reduced positive bias over bright surfaces. The output at L2 is total column AOD at 550 nm, at 10 km resolution, and associated aerosol properties. Retrieval uncertainty and comparison with sun photometer observations show highest accuracy retrieval over ocean and darker surfaces, with higher uncertainty over bright desert surfaces, and land surface at southern latitudes (Popp et al. (2016)). The level 3 output is re-gridded to daily and monthly 1 degree resolution, intended for climate model comparison.

In this study, AE as well as total, fine and coarse AODs are used. Results (normalised biases and correlation coefficients) from an inter-comparison with AERONET measurements is shown in Fig. 2 (discussed in more detail in Sect. 2.4).

2.1.5 Merged satellite AOD data

The MERGED-FMI dataset (1995-2017), developed by the Finnish Meteorological Institute, includes gridded L3 monthly AOD products merged from 12 available satellite products (Sogacheva et al. (2019)). The merging method is based on the AOD evaluation results against AERONET for the individual satellite AOD products. Those results were utilised to infer a regional ranking which was then used to calculate a weighted AOD mean. Because it is combined from the individual products of different spatial and temporal resolution, the AOD merged product is characterised by the best possible coverage, compared

with other individual satellite products. The AOD merged product is at least as capable of representing monthly means as the individual products (Sogacheva et al. (2019)). Standard pixel-level uncertainties for the merged AOD product were estimated as the root mean squared sum of the deviations between that product and other eight merged AOD products calculated with different merging approaches applied for different aerosol types (Sogacheva et al. (2019)).

2.2 Models

This study uses output from 13 models that are participating in the AeroCom 2019 control experiment (https://wiki.met.no/aerocom/phase3-experiments, denoted in the following as CTRL). For this experiment, modellers were asked to submit simulations of at least the years 2010 and 1850, with 2010 meteorology and prescribed (observed) sea-surface temperature and sea ice concentrations. Modellers were asked to use CMIP6 emission inventories, when possible. Detailed information about the models and their treatment of aerosol optics are provided in the supplementary material Tab. LINK TO OPTICS QUESTIONNAIRE (PLEASE CHECK AND FILL IF YOU HAVE NOT DONE SO)currently available here: https://docs.google.com/spreadsheets/d/1VN_mG2r3bqQuzDVUajBwqohlbp_fHRsNe9ThfuHnHUs/edit#gid=786063029). An overview of all models is provided in table 2. More details about each of the models is provided in the corresponding discussion section in Sect. **??**.

2.2.1 AeroCom ensemble mean and median

The AeroCom ensemble mean and median were computed in monthly resolution, considering only those models where all required optical properties were available before 19.12.2019 (cf. Fig. 10). These models are also indicated in Tab. A1). Here we use output from the median model (denoted MEDIAN below) if not explicitly indicated differently.

For each of the variable fields, each of the included models was first re-gridded to a latitude / longitude resolution of $2^\circ \times 3^\circ$ and to monthly resolution (for the models that provided higher temporal resolution). From these harmonised models, both the arithmetic mean and the median model were computed as well as diversity fields for each variable. In case of the mean model, the diversity was computed as defined in Textor et al. (2006) (cf. Eq. 1 therein). For the median model the interquartile range (IQR) was used.

2.3 Data analysis

The analysis of the data was performed using the pyaerocom software (cf. Sect. C). The ground and space based observations are colocated with the model simulations by matching with the closest model grid-point in the provided model resolution. In the case of ground-based observations (AERONET and in-situ), the model grid point closest to each measurement station is used. For the satellite observations, both the model data and the (gridded) satellite product are re-gridded to a resolution of $5^\circ \times 5^\circ$ and the closest model grid point to each satellite pixel is used. The choice of this rather coarse resolution is a compromise mostly serving the purpose of increasing the temporal representativity (i.e. more data points per grid cell) in order to meet the time resampling constraints (cf. Sect. **??**) and also, to reduce the processing time and required data storage for the web visualisation of the results (https://aerocom-evaluation.met.no/overall.php?project=aerocom&exp=PIII-optics2019-

P). A sensitivity analysis was performed by comparing the monthly average statistics discussed below with statistics based on model colocation in daily resolution and in the original horizontal satellite resolution. The differences in bias and correlation coefficient were found to be small ($<$ **NUM**% in NMB and $<$ **NUM** in correlation) compared to the diversity found among the models (cf. Tab. **??**).

Since many model fields were only available in monthly resolution, the colocation of the data with the observations (and the computation of the statistical parameters used to compare the models) was performed in monthly resolution. Any model data provided in higher temporal resolution was resampled to monthly using the mean value, prior to the analysis. For the higher resolution observations (cf. 1), the computation of monthly means was done using a hierarchical resampling scheme, requiring at least 25% coverage. Practically this means that the daily AERONET data was resampled to monthly, requiring at least 7 daily values in each months. For the hourly in-situ data, first a daily mean was computed (requiring at least 6 valid hourly values) and from these daily means, monthly means were computed requiring at least 7 daily values. Data that did not match these coverage constraints were invalidated.

2.4 Evaluation of satellite products at AERONET stations

All satellite data-sets were evaluated against the ground based AERONET data in order to establish an estimate of the relative differences (biases, correlation coefficients) between the different data-sets when comparing with the models. The evaluation of the gridded satellite level 3 products was performed in the same manner as the evaluation of the models, as described in the previous section 2.3.

Figure 2 shows the results of this analysis. In terms of AOD, a high correlation is found for all satellite products with AERONET observations ($R > 0.80$). In terms of bias (normalized mean biasNMB), AATSR shows an underestimation of -4.5% while MODIS Aqua and Terra yield slightly overestimated biases of +9.3% and +18.1% respectively.

We remark that this analysis is biased by the uneven distribution of AERONET sites (highest density in Europe and North America, cf. Fig. 1) and that problematic regions in the satellite retrievals (e.g. Sahara) may not be well represented in this comparison.

The SU AATSR retrieval includes a conservative cloud mask utilising thermal channels in additional to optical, and avoids retrieval near cloud edges. Evaluation under aerosol CCI of six datasets showed AATSR and SeaWifs had lowest bias (with SeaWifs) with respect to ocean and coastal sun photometers (Popp et al. (2016)).

3 Results

In this section the results from the model evaluation are presented, starting with a presentation of the results from the ensemble model and the distributions of model diversity, followed by a brief discussion of annual averaged emissions, burdens, lifetimes, MECs and ODs for each modelled aerosol species. Finally, the results of the optical property evaluation are presented in ...

3.1 Annual global distributions of optical properties and their diversity

Figure 3 shows maps of yearly average values for each aerosol variable (left) including the mean values of each corresponding ground based network at the individual locations. The right panel shows corresponding diversity maps computed following the definition of diversity in Textor et al. (2006) (mean normalised standard deviation of results from individual models). Also provided are model mean values (global and at station locations) as well as the mean value of the observations (at stations).

The overall highest model diversity is found for the simulated surface in situ aerosol absorption coefficients and is particularly prominent in Amazonia, a region of substantial regular biomass burning events (peaking in early September in 2010) and also new particle formation (NPF) events due to biogenic emissions. The diversity may thus be a combination of the different treatments of SOA formation (and absorptive properties of OA) as well as differences in **NUM**. Diversity is also high in the South Pacific / Antarctica. Interestingly, models tend to agree in major source regions such as China and India (low diversity in surf. absorption).

Interestingly, coarse mode AOD shows relatively large global diversity (62 %) which seems to mostly arise from differences in elevated or mountainous desert regions such as the Southern Peruvian and Northern Chilean Andes, Tibet, but also Antarctica. These regions are however, associated with generally low AODs.

3.2 Modelled emissions, burdens, lifetimes and MECs

Lifetimes were calculated for each species by dividing global averages of the correspond load fields with the total depos

Figsures 5, 6, 7 and 8 show global annual average of emissions, loads, lifetimes and mass to extinction coefficients, for each aerosol species and for each model, respectively. Note that the colors in the heatmaps are applied row-wise in order to highlight differences between the models. Also included in each plot are mean, median and diversity for each species. The diversity is computed as the average width of the distribution inferred from 1st and 3d quantiles. Figure 9 shows corresponding averages for the individual optical depths of each species. Also shown are values of AOD due to water, reported clear sky (CS) and / or all sky AOD (lowermost rows) as well as the summed AOD of all non-water aerosol species (BC, DU, OA, (NO3) SO4 and SS) in the third last row. Models

3.3 Results from optical properties evaluation

Figures 10 and 11 show performance matrices of the normalised mean bias (NMB) as well and the Pearson correlation coefficient, respectively for each model, variable and observation dataset used. The ensemble model is plotted in the rightmost column. Overall, it appears from Fig. 10 that models moderately underestimate the selected optical properties, both when evaluated against ground-based remote sensing and in situ observations but also against the satellite datasets. The latter provide the highest spatial coverage (cf. Fig. 1) and are particularly sensitive to the oceans as they contribute the largest surface fraction. However, it is important to note that the temporal sampling of the satellites is limited to their local overpass time and cloud-free conditions, which may introduce sampling biases of the order of XX (Check Nick's paper(s), as compared with the fully sampled (24h) monthly model fields. To a certain degree, these differences are indicated in the evaluation of the satellite AODs

against AERONET (cf. Sect. 2.4, Fig. 2) and they propagate into the statistics established in the evaluation of the individual models. For instance, the ensemble model shows biases of $-8.1\,\%$, $-23.1\,\%$, $-8\,\%$

3.4 Representativity of the results

As described in Sect. 2.3, monthly averages of the models and observations (computed from higher temporal resolution observation data, requiring at least 25% sampling coverage as described in Sect. 2.3) were colocated in space (using nearest neighbour colocation) and time. This results in a point cloud of monthly mean values which is used to compute the statistical parameters (in this study: normalised biases and correlation coefficients). These statistical parameters are then used to assess the performance of individual models and the ensemble mean, discussed in the following sections (cf. Figs. 10 and 11). The comparison of the (often) temporally incomplete observational records, that are sampled at distinct locations, can introduce considerable representation errors both on spatial and on temporal scales (see e.g. Schutgens et al. (2016), Schutgens et al. (2017), Wang et al. (2018), Sayer and Knobelspiesse (2019) and references therein). These errors can affect established biases between model and observation but also other performance measures such as correlation coefficients.

Therefore, several sensitivity studies have been performed in order to investigate how these spatio-temporal representation errors affect the global monthly statistical parameters that are used in this study. Temporal representation uncertainties were investigated 1. for in-situ absorption coefficients using hourly TM5 data from the AeroCom INSITU experiment (CITE) evaluated against GAW measurements (Fig. A4) and 2. for columnar AOD using 3-hourly data from ECMWF-IFS, evaluated against AERONET AODs (Fig. A3). In addition, spatial uncertainty errors were investigated for the ensemble mean AOD using monthly means from all AERONET sites available in 2010 and a selection of AERONET sites with an absolute spatial representation error smaller than 10%, as established by Wang et al. (2018) (Fig. A5). The results are summarised in Tab. **??** and show that the overall differences are of the order of 10% and 0.2 for NMB and correlation, respectively. The monthly resolution colocation improved performance of the in-situ absorption inter-comparison in nearly all statistical parameters, relative to hourly colocation. For example the correlation improved by 0.2 (cf. Fig. A4). However, these differences arising from spatio-temporal representation errors are small compared to the diversity in the results among the different models participating in this study (shown in Figs. 10 and 11), which are discussed in the following sections.

Based on these results and due to the fact that some model data was only available in monthly resolution, it was decided that all model and observation comparisons in this study would be performed in monthly resolution. This will make the inter-model results more consistent and hence, more suitable for *inter*-comparison as they carry similar representation errors (which are introduced by the incompletely sampled observational records), with the spatial representation errors for each model being affected by its horizontal resolution (cf. 2, but to a minor degree. The small differences in bias and correlation that we find in our sensitivity tests (cf. Figs. A3, A4, A5) are important results that indicate that the magnitude of spatio-temporal representation uncertainties in statistical parameters derived from annual averages over whole networks (or satellite records) is of the order of $\pm 10\%$. However, we want to stress that these uncertainties are not to be misinterpreted with corresponding uncertainties over sub-domains or at specific locations and times, which can be much higher as shown in the various literature referred to above.

See also Fig. 2 and Tab. A2 for an assessment of potential biases and uncertainties related to the satellite data analysis, which was performed based on monthly means and based on re-gridded satellite and model data at a resolution of $5^\circ \times 5^\circ$ (for the reasons stated above).

4 Discussion

4.1 CAM5-ATRAS

The emissions and burden of OA (primary + secondary formation) are greater than the model ensemble by 90% and 50%, respectively (Figs. 5, 6), because the ATRAS model considers OA formation from semi-volatile and intermediate volatility organic compounds in addition to anthropogenic and biogenic VOCs based on the volatility basis set approach (Matsui et al., 2014a, 2014b; Matsui, 2017). NO3 burden is lower than the model ensemble and it is consistent with Matsui and Mahowald (2017). The burdens of BC, SO4, SS, and DU and the lifetimes of all aerosol species in CAM5-ATRAS are similar to those in the model ensemble (Figures 5-7). BC MEC is greater than the model ensemble by 40% (Figure 8) likely because the ATRAS model calculates the enhancement of absorption by BC aging processes explicitly by resolving BC mixing state with 8 bins (pure BC, BC-free, and 6 internally-mixed BC bins). The value of BC MEC in this study (9.5 m2 g-1) is close to that in Matsui et al. (2018) (10 m2 g-1). AOD biases are also similar to the model ensemble (Fig 10). The underestimation of AOD compared with MODIS and AERONET AODs is consistent with Matsui (2017).

4.2 EC-Earth3-AerChem and TM5

Two configurations of the atmospheric composition model TM5 (Tracer Model 5) are included in this study (van Noije et al. (2014)): a standalone version of TM5, and an atmosphere-only version of the CMIP6 climate model EC-Earth3-AerChem (van Noije (2019)). The standalone model is driven by meteorological and surface fields from the ERA-Interim reanalysis (Dee et al. (2011)), whereas in the climate model there is online interaction between TM5 and the atmospheric general circulation model, which is based on model cycle 36r4 of ECMWF's Integrated Forecasting System (IFS). The set of meteorological and surface variables that drive TM5 are the same in both configurations. In the EC-Earth simulations analyzed in this study, sea surface temperatures and sea ice concentrations were prescribed using AMIP forcing fields provided for CMIP6; in addition, vorticity, divergence and surface pressure fields were nudged to ERA-Interim, using a Newtonian relaxation scheme with a time constant of 8 h and 15 min in the whole atmosphere.

TM5 uses the aerosol scheme M7 (Vignati et al. (2004)), which represents sulfate, black carbon, organic aerosols, sea salt and mineral dust with seven lognormal size distributions or modes. Aerosol components are assumed to be internally mixed inside the modes. The formation of secondary organic aerosols in the atmosphere is described following Bergman et al., in prep). Ammonium-nitrate and methane sulfonic acid (MSA) are described by their total mass, and assumed to be present only in the soluble accumulation mode (see van Noije et al. (2014) for more details). TM5 has an interactive tropospheric chemistry scheme (Williams et al. (2017)), which also describes the aqueous-phase oxidation of dissolved sulfur dioxide in clouds.

When calculating the dust source, TM5 excludes the particles with dry diameter larger than 16 μm. This may explain why the mean emitted dust mass is smaller than in other models. Only if cut-off size applied in the other models are larger. What are they? The cut-off also affects the dust lifetime and MEC. Differences in 10 m wind speeds generally reduce the dust emissions from the main source regions in EC-Earth compared to TM5 (Fig. 5), leading to proportionally lower dust burdens. Sea-salt emissions, on the other hand, which depend on 10 m wind speeds and sea surface temperatures, are very similar in the two models. The mean OA lifetime in EC-Earth is 9% longer than in TM5, and in both models are longer than in the other models. This may be in part due to the use of interactive chemistry in TM5 (and EC-Earth), which may lead to a depletion of oxidants over regions with high biogenic VOC emissions, thereby increasing their lifetime (Sporre et al., in preparation). How does this lead to longer SOA lifetime? The aerosol optical properties in TM5 are calculated based on Mie theory, where the mixing rules of Bruggeman and Maxwell-Garnett are applied as approximations of the refractive index of the internally mixed modes. The contributions of the individual aerosol components are estimated by distributing the resulting total ambient extinction of each mode over the individual dry aerosol components, using volume weighting. In this way the extinction due to the presence of water is associated with the other aerosol components. This will enhance the MEC values for TM5 and EC-Earth compared to models in which the water contribution is excluded in the component MECs, such as ECHAM-HAM and ECHAM-SALSA (cf. Fig. **??**). Some general remarks highlighting the models' performance compared to the observations can be added.

4.3 ECHAM-HAM

The global aerosol-climate model ECHAM6.3-HAM2.3 (ECHAM-HAM in the following) is part of the fully coupled aerosol chemistry climate model ECHAM–HAMMOZ (Tegen et al. (2019), Schultz et al. (2018)). Aerosol microphysical processes in ECHAM-HAM are described with the modal M7 aerosol model (Vignati et al. (2004)) in contrast to ECHAM-SALSA which employs the sectional aerosol scheme SALSA (Kokkola et al. (2018)). The aerosol representation in ECHAM-HAM has been evaluated in Tegen et al. (2019) but using different aerosol emissions (different inventories for anthropogenic and biomass burning emissions as well as a different sea salt emission parameterisation). For the CTRL experiment the sea salt emission parameterisation from Guelle et al. (2001) was chosen, firstly because the one proposed by Long et al. (2011) and Sofiev et al. (2011) resulted in an underestimation of the sea salt concentrations (Tegen et al. (2019)) and secondly, to be consistent with the CTRL setup of ECHAM-SALSA (cf. Sect. 4.4. However, this comes at the price of larger sea salt particles (on average), resulting in a slightly decreased correlation against AERONET compared to Tegen et al. (2019). The latter, however, may to a certain degree also be affected by different representation errors as Tegen et al. (2019) use 6-hourly data to colocate in time, while this study relies on monthly means (cf. Sect. 3.4, particularly Tab. 3). AOD over land is lower than in AERONET or MODIS observations (Fig. 10) which may be due to several reasons, for instance because NO3 is missing, too low emissions of OA or a misrepresentation of SOA (the OA burden in ECHAM-HAM is lower than in most other models, see Fig. 6 and Tegen et al. (2019)). The fine mode AOD is overestimated over ocean and dusty regions which is indicated by the stronger overestimation compared to AATSR (dominated by ocean) than to AERONET (more representative of land). The coarse mode AOD on the other hand is underestimated over land (too low compared to AERONET, Fig. 10) but overestimated over subtropical ocean, leading to almost no bias compared to AATSR. Except for regions dominated by dust aerosol AE

is biased low. The underestimation of AE in dust dominated regions combined with the overestimation of fine mode AOD and the longer lifetime of dust particles compared to other models (Fig. 7) indicates a too small size of dust particles. The underestimation of AE compared to AERONET and AATSR is surprising since fine mode AOD is overestimated (Fig. 10), the aerosol size distribution of ECHAM-HAM agrees reasonably well with observations (Tegen et al., 2019) and Tegen et al. (2019) find a positive bias of AE compared to AERONET. This could be related to the different sea salt emission parameterization applied in CTRL or temporal sampling errors (Schutgens et al., 2016; Sayer and Knobelspiesse, 2019).

Comments:

- Region of strong anti-correlation in eastern Mediterranean in fine AOD and AOD
- Species AODs are reported as dry (!), also in ECHAM-SALSA: impacts MEC (!)
- Generally relatively low correlation against most variables compared to other models (cf. Fig. 11
- AOD underestimated by about 30% (Tegen et al. (2019), Fig. 5 find absolute biases of -0.03 to -0.05, our results correspond to roughly -0.06 in AOD units)
- ...

4.4 ECHAM-SALSA

SALSA is the sectional aerosol microphysics module within the ECHAM-HAMMOZ aerosol-chemistry-climate model (Kokkola et al., 2018) alongside the modal aerosol module M7 (Tegen et al., 2019). The implementation of SALSA to ECHAM-HAMMOZ and its evaluation against satellite retrievals, ground based remote sensing retrievals, and in situ observations has been described by Kokkola et al. (2018). One change in these model simulations compared to those in Kokkola et al. (2018) are, in addition to using anthropogenic emissions required for AEROCOM III simulations, is using sea salt emission parameterization of Guelle et al. (2001) for the reasons described in the previous section 4.3.

As the atmospheric model is the same in ECHAM-HAM and ECHAM-SALSA, results between the two model configurations are quite similar. An overall view of the performance of SALSA is that the values fall within the spectrum of model ensemble values except for the burdens of BC and SU for which SALSA predicts highest values of all models (Fig. 6). The BC lifetime is highest among all models (9.6 days, Fig. 7) which explains the high burden. On the other hand, reasons for the high SO4 burden are not obvious and, since corresponding emissions and lifetimes are comparable with the other models. It may hence be related to the oxidation efficiency of sulphate from its precursors (DMS, SO2).

When comparing the total simulated (clear sky) AOD of SALSA to observation, values are higher than those from AATSR, MERGED-FMI, and ECMWF reanalysis and lower than those from Aeronet and MODIS (Aqua and Terra). Over most of the land area SALSA underestimates AOD while overestimating AOD over the oceans. The exceptions for the underestimation are Australia and North Africa where SALSA exhibit high values for the total AOD. This is likely due to the contribution of dust to the AOD and is also reflected in the coarse mode AOD. The coarse mode AOD of SALSA is significantly overestimated with normalized bias of 24%. Regions with high dust loads exhibit overestimation of coarse mode AOD. These features were

also seen in the study by Kokkola et al. (2018). In addition to dusty regions, coarse mode AOD has a large positive bias over the oceans. We expect this to be due to high simulated relative humidity in ECHAM over the oceans or too high hygroscopicity for SS aerosol. It is noteworthy that although coarse mode AOD is overestimated over regions where AOD is dominated by sea salt and dust, their emissions are not higher in SALSA and it is likely that the simulated size distribution of SALSA is such that SS and DU particles influence radiation effectively. For Aeronet sites, similar overestimation of coarse mode AOD is not seen since Aeronet sites cover the land area. On the other hand, over regions affected by DU, coarse mode AOD is overestimated in SALSA. For example, Aeronet sites north of Africa exhibit simulated values higher than those measured.

- From Fig. 9 it seems that both ECHAM models report dry species AODs (i.e. the sum of od550h2o and all the other species gives the total). This should be mentioned as all the other models report them at ambient conditions. This explains why MECs are lowest for the 2 ECHAM models.

- Can we explain, why surface scattering and absorption are underestimated by about -68% and -40%?

References: Bergman et al. (2012), Kokkola et al. (2018)

4.5 ECMWF-IFS

As part of the Copernicus Atmosphere Monitoring Service (CAMS; https://atmosphere.copernicus.eu/), ECMWF runs a version of IFS model that includes prognostic aerosol and tropospheric chemistry schemes to produce global forecasts of atmospheric composition. The underlying meteorological model is essentially identical to that used for operational medium-range weather forecasting and documented at https://www.ecmwf.int/en/forecasts/documentation-and-support, but at a lower resolution of 40 km to offset the cost of the extra schemes. The results presented here are from a "cycling forecast" configuration, i.e. a forecast with free-running aerosols and chemical species (no assimilation of atmospheric composition), with meteorology reinitialised at 00 UTC each day from operational ECMWF analyses.

The aerosol component is described in Rémy et al. (2019) and based on the earlier work of Morcrette et al. (2009). This is an externally-mixed hybrid bin/bulk scheme, consisting of three size bins each for desert dust (up to $20\mu m$ dry radius) and sea-salt (up to $20\mu m$ radius at 80% relative humidity), and bulk tracers for organic matter, black carbon and sulfate aerosol. For organic matter and black carbon, there are separate hydrophobic and hydrophilic tracers, with a fixed *ageing* timescale for conversion of the former to the latter. There is also an SO_2 precursor tracer driving the sulfate production via a latitude- and temperature-dependent conversion timescale. There is no separate DMS tracer, and no primary sulfate aerosol emission – all sulfate and precursor emissions are treated as SO_2. The tropospheric chemistry scheme is described in Flemming et al. (2015), but in the version described here this is not directly coupled to the aerosol scheme.

Compared to the other AP3 models, the total sea-salt emissions and burden are very large, as can be see in Figures 5 and 6. Emissions are three times larger than the ensemble mean, but due to a short lifetime (see Figure 7) the burden is only three times larger. However, the sea-salt contribution to AOD remains similar to other models because the large size distribution reduces the extinction per unit mass. These are known issues with the emission scheme in this version of the model (based on Grythe et al., (2014; https://doi.org/10.5194/acp-16-12081-2016), and the subject of ongoing development.

The model also has one of the smallest sulfate burdens, which appears to be the result of both relatively low total sulfur emissions and a short lifetime. Organic aerosol emissions are higher than most models, although the burden and lifetime are similar to other models. This is likely due to the fact that there is no secondary organic precursor scheme, and secondary organic production is included instead as if it were a primary emission.

Although correlation coefficients for AOD (Figure 11) for this model are amongst the higher ones, there is a significant low bias against all the AOD datasets (satellite and AERONET). This is likely related to the relatively short lifetimes of many species compared to other models, which can be seen in Figure 7. There is also a low bias against both AERONET and ATSR AE, suggesting that particles are on average too large; this may well be due at least in part to the unusually high sea-salt burden in the model noted above.

References:

4.6 EMEP MSC-W

The EMEP MSCW model is a chemical transport model, designed for policy related applications to combat acid deposition, eutrophication and health adverse air pollution (Simpson et al. (2012)). It calculates the mass concentrations of all main anthropogenic and natural aerosols, contributing to the health related indicators $\mathrm{PM}_{2.5}$ and PM_{10}. The results presented in the paper were obtained in a model run at $0.5 \times 0.5^{\circ}$ grid, driven by 3-hourly ECMWF-IFS meteorology and using ECLIPSE6b emissions (ECLIPSE6a for shipping), both for the year of 2010. The model includes aerosols with diameters up to 10 μg and calculates the mass concentrations aerosols in fine and coarse mode. Then, the extinction and absorption coefficients are calculated for the individual aerosol components using mass extinction/absorption coefficients and accounting for aerosol hygroscopic growth (aerosol effective radii, growth factors and specific extinction efficiencies are tabulated) (Schulz et al. (2012)).

The calculated all-sky AOD is -10% lower compared to globally averaged annual AOD from AERONET (correlation 0.76). Comparison with satellite AOD shows suggests underestimations between 34%-51%, and the relative differences here mostly reproduce the biases observed between the satellites (Fig. 2). These results indicate that EMEP underestimates AOD more over the oceans than over land. Evaluation results against those observations for different world regions are inconclusive in terms of model bias (inferred from web visualisation of the results, cf. Sect. **??**). Furthermore, fine AOD is overestimated by 20% compared with AERONET data and slightly (by only 11%) underestimated compared to ATSR-SU, whereas coarse AOD is considerably underestimated (by 68 and 70% respectively). Consistently with that, the AE is somewhat overestimated (by 36% and 44%), indicating a disproportion between the contributions to AOD from the fine and coarse aerosols and suggesting that either the EMEP model calculates too little of coarse particles or the applied mass extinction efficiencies are too low. One of the possible reasons for that is that fine sea salt and dust particles in the EMEP model are assumed to have diameters smaller than 2.5 μg, so that the extinction due to sea salt and dust aerosols with diameters between 1 and 2.5 μg contributes actually to the AOD<1 μg.

Regarding aerosol specific AODs (Fig. 9), the EMEP model calculates somewhat larger than the median AOD due to NO_3 and OA, which is in agreement with its relatively large loads for those components (the model calculates both fine ammonium

nitrate and coarse NO_3 on sea salt and dust; and EMEP OA include primary OA and anthropogenic and biogenic secondary OA). For the other aerosols, EMEP calculated aerosol loads and AODs are somewhat smaller than the mean/median values. The resulting MECs are in general agreement with the Aerocom median MECs, with the exception of the MEC for SO_4, which is one of the largest (probably due to too effective hygroscopic growth), which is compensated by the rather low SO_4 burden (SO_x emissions from ECLIPSE6b used by the EMEP model are smaller than from CMIP6).

The small discrepancy between Total AOD and the sum of the aerosol specific AODs is because the modelled BC AOD is only due to anthropogenic emissions (does not include forest fires) and DU AOD is only due to windblown dust (while some fugitive anthropogenic dust is also included in the Total AOD).

Absorption coefficient is diagnosed from BC and dust mass concentrations, using Mass Absorption Coefficients. Overall, EMEP Absorption coefficient for 2010 is 40% lower than ebas climatological observations, and the correlation is 0.66, which is a fair result given the crude simulation approach. The scattering coefficient is underestimated by 47% on the average by the model (Pearson R = 0.74). It should be noted that the absorption and scattering coefficients from the EMEP model are representative of dry aerosols.

Small text about burdens/lifetimes here...needs to be uncommented

References:

4.7 GEOS

The version of GEOS-5 Earth System Model with a GOCART aerosol module used for this study is Icarus-3_3_p2. The simulations run at a spatial resolution of 1.0° longitude x 1.0° latitude and 72 vertical levels from surface up to 0.01hPa (85km) with the "replay" mode, i.e., simulations driven by the reanalysis meteorological fields from the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA2) to assure that the weather and climate patterns are accurately represented for the simulated time. The GOCART module includes major aerosol types of black carbon (BC), organic carbon (OC), brown carbon (BRC), sulphate, nitrate, ammonium, dust, and sea salt (Chin et al., 2002, Colarco et al., 2010, Bian et al., 2019). The aerosol emissions used in this study follow the instruction of AeroCom Phase III History experiment. The major updates of this GOCART version include newly implemented nitrate, ammonium, anthropogenic SOA, and biomass burning SOA, as well as separate treatment of optical properties for brown carbon (from biomass burning source) and organic carbon (from all other sources) (Bian et al., 2017).

References:

Bian, H., Froyd, K., Murphy, D. M., Dibb, J., Darmenov, A., Chin, M., Colarco, P. R., da Silva, A., Kucsera, T. L., Schill, G., Yu, H., Bui, P., Dollner, M., Weinzierl, B., and Smirnov, A.: Observationally constrained analysis of sea salt aerosol in the marine atmosphere, Atmos. Chem. Phys., 19, 10773–10785, https://doi.org/10.5194/acp-19-10773-2019, 2019.

Bian, H., Chin, M., Hauglustaine, D. A., Schulz, M., Myhre, G., Bauer, S. E., Lund, M. T., Karydis, V. A., Kucsera, T. L., Pan, X., Pozzer, A., Skeie, R. B., Steenrod, S. D., Sudo, K., Tsigaridis, K., Tsimpidi, A. P., and Tsyro, S. G.: In-

vestigation of global particulate nitrate from the AeroCom phase III experiment, Atmos. Chem. Phys., 17, 12911–12940, https://doi.org/10.5194/acp-17-12911-2017, 2017.

Chin, M., P. Ginoux, S. Kinne, B. N. Holben, B. N. Duncan, R. V. Martin, J. A. Logan, A. Higurashi, and T. Nakajima, 2002: Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sun photometer measurements, J. Atmos. Sci. 59, 461-483.

Colarco, P., da Silva, A., Chin, M., and Diehl, T.: On-line simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground based aerosol optical depth, J. Geophys. Res., 115, D14207, https://doi.org/10.1029/2009JD0 2010.

PLEASE INSERT A SHORT PARAGRAPH THAT EXPLAINS / COMBINES THE RESULTS FROM YOUR MODEL SHOWN IN FIGS 10, 11 AND 5, 6, 7, 8. You may, e.g. see Sect. 4.11 as a guideline example. PLEASE ALSO COMMENT ON DUST SOURCE CUT OFF SIZE IF APPLICABLE (cf. 4.2). AND PLEASE DISCUSS BIAS IN Angstrom Exponent (AE), especially if AE is underestimated (i.e. suggesting too large particles) but comparisons of od550gt1aer (coarse AOD) and od550lt1aer (fine AOD) are suggesting differently (e.g. fine mode overestimated, coarse mode underestimated).

4.8 GFDL-AM4

The Geophysical Fluid Dynamics Laboratory Atmospheric Model version 4 has cubed-sphere topology with 96×96 grid boxes per cube face (approximately 100 km grid size) and 33 levels in the vertical, contains an aerosol bulk model that generates mass concentration of aerosol fields (sulfate, carbonaceous aerosols, sea salt and dust) from emissions and a "light" chemistry mechanism designed to support the aerosol model but with prescribed ozone and radicals Zhao et al. (2018). The model is driven by time-varying boundary conditions, and natural and anthropogenic forcings developed in support of CMIP6 Eyring et al. (2016), except for ship emission of SO_2 (BC ship emission is included), which has unintentionally not been included. The dust is emitted from constant sources with their erodibility, a function of surrounding orography Ginoux et al. (2001). The sea salt emissions are based on Mårtensson et al. (2003) and Monahan et al. (1986) for fine and coarse mode particles, respectively. Aerosols are externally mixed except for black carbon which is internally mixed with sulfate. The optical properties of the mixture are calculated by volume weighting of their refractive indices using a Mie code. In the present configuration, the model is run with observed sea surface temperatures (SSTs) and sea-ice distribution **?**. In addition, the wind components are nudged, with a 6-hour relaxation time, towards the NCEP-NCAR re-analysis provided on a T62 Gaussian grid with 192 longitude equally spaced and 94 latitude unequally spaced grid points ((Kalnay et al., 1996)). This resolution is lower than in GFDL-AM4, which may create a low bias of aerosol emission depending on surface winds.

In Fig. 5, aerosol emission from GFDL-AM4 are within 25% of the ensemble mean, except for SO_2 and SO_4, which are the lowest among all models essentially because ship emissions are missing in the simulations. The lower emissions of sulfur compounds does not translate in low atmospheric burden (Fig. 6) as their lifetime is among the highest between the models (Fig. 7), either because of weak oxidation or deposition. On the other hand, the other aerosols have a shorter lifetime than other models (Fig. 7) while their burdens are well within 25% the AP3 mean values (Fig. 6). The opposite bias between sulfur

compounds and the other aerosols suggest an issue with oxidation of SO_2 rather than wet or dry deposition. In Figure 8) the MEC values are within the diversity of the AP3 models except for sea salt which is lower by a third. This may be because of the cap at 97% relative humidity on the hygroscopicity or the emission parameterisation, as the scheme of Mårtensson et al. (2003) generates much less sea salt sub-micron particles than Monahan et al. (1986). The GFDL-AM4 AODs from individual species (Fig. 9) are within the AP3 model diversity except BC, which has the highest value most likely due to the treatment of its internal mixing with sulfate. This high bias will convert into high bias of fine mode AOD, as it appears in Figure 10 where the positive biases of fine mode AOD compare to AERONET and ATSR-SU are the largest among all models. Other normalized biases are relatively weak compared to other models (Figure 10). AOD bias is slightly negative against AERONET and the different satellites, and differences in these biases mostly represent the biases found for the different satellites at AERONET stations (cf. Fig. 2). However, it is important to note, that this model version reported all-sky AOD, while most other models report AOD at clear-sky, which would likely shift the biases towards increased underestimation of AOD (cf. e.g. Sect. 4.11, see also Tab. LINK TO OPTICS QUESTIONNAIRE (PLEASE CHECK AND FILL IF YOU HAVE NOT DONE SO)). Overall, optical properties are well correlated with observations with coefficients greater than 0.74 except for the scattering and absorption coefficients provided by the surface in-situ data with values at 0.49 and 0.57, respectively (Fig. 11). Concerning the Angstrom exponent, one set of value (AERONET) gives poor correlation (0.52) while another (ATSR-SU) provides reasonable correlation (0.74).

4.9 GISS-OMA

GISS-OMA is the short name of the GISS ModelE Earth system model (Kelley et al., will be submitted to JAMES before 12/31/19), coupled with the One-Moment Aerosol scheme (OMA; Bauer and Tsigaridis, submitted to JAMES). In OMA, all aerosols are externally mixed and tracked by their total mass only, except for sea salt and dust where 2 and 5 size-resolved sections are used, respectively. OMA tracks sulfate, nitrate, ammonium, carbonaceous aerosols (black and organic carbon), dust (up to 16um) and sea salt (up to 4um).

Relevant to this work, a random maximum cloud overlap is calculated in the column, which is then used to define a totally cloudy or totally cloud-free state in radiation, using a pseudo-random number generation. This is described in Hansen et al. (1983). For all-sky AOD calculations 100% relative humidity is used, while for clear-sky we use ambient. This applies to the whole atmospheric column, as dictated by the random maximum cloud overlap calculation. In GISS-OMA there is no calculation from AE. Instead, we calculate it from the AOD calculations in radiation, which are probably underestimating AOD at 870nm by about 10%.

- MEC of SS, SO4, NO3 and BC are very high
- Coarse mode AOD overestimated, particularly over the ocean (cf. ATSR comparison)
- Highly underestimated Angstrom Exponent (in agreement with overestimated coarse mode)
- ONLY MODEL THAT SHOWS OVERESTIMATED SURFACE SCATTERING (+20% NMB) AT GAW STATIONS.

4.10 INCA

The INCA (INteraction with Chemisty and Aerosols) and ORCHIDEE land surface modules has been coupled to LMDZ dynamical core to conform the LMDZORINCA model. It has been run with forced sea-surface temperatures, sea-ice concentrations and with nudged monthly wind-fields from ERA-Interim. The comparisons with the climatological simulations without nudged-winds shows slightly larger emissions of those aerosols driven interactively by wind at the surface Balkanski et al. (2004), Schulz et al. (2009). The aerosol modelling in INCA relies on a modal approach to represent the size distribution of DU, SS, BC, NO3, SO4, SO2 and OA with a combination of accumulation and coarse log-normal modes (both soluble and insoluble). Since these runs use a simplified chemistry scheme, DMS emissions are prescribed and not interactively calculated, and the secondary organic aerosols are not simulated. Hence the organic aerosols are underestimated by this model (cf. low burden in Fig. 6). The current version is modelling BC as internally mixed with sulphate (Wang et al. (2016)), where the refractive index is estimated using the Garnet-Maxwell method. This results in an increased and more accurate BC absorption. On the other hand, the dust refractive index is deduced from dedicated experiments Biagio et al. (2017, 2019) showing a marked impact on the longwave part of the spectrum. This results in a less absorbing dust aerosol. BC emissions are derived from inventories and are equally partitioned between surface and altitude.

The emissions of dust and sea-salt have values close to the ensemble mean. With LMDZORINCA the global emitted mineral dust is 1560 Tg/yr (cf. Fig. 5) is within the interval proposed by Kok et al. (2017) for fine and coarse modes. The simulations are based on a coarse insoluble mode (MMD=2.5 μm and σ). Meanwhile, an improved version with 4 modes (Albani et al, in prep) shows that including larger particles implies significant higher emissions, although burdens do not increase as substantially as emissions due to the small lifetime of larger particles (R.Checa-Garcia et al., in prep). Sea salt the emissions amount to 4030 Tg/yr and include accumulation and coarse soluble modes (the super-coarse mode is calculated but not included in this estimation). OA emissions (48.3 Tg/yr) are underestimated compared to other models (ensemble mean 98.2 Tg/yr) because SOA formation is not accounted for. This also explains the comparatively low burden of OA (0.79 Tg, cf. Fig. 6). All lifetimes are close to the ensemble central values but for dust and sea-salt the value given by the models depends on the cutoff applied to the size distribution. For dust our lifetime is estimated to be close to 4 days.

Our values of MEC are close to the ensemble mean, for those aerosol species modelled by a single mode (like dust) we expect less spatial variation of MEC that other models with several modes. Regarding optical properties, the AE has smaller values, due to an smaller dynamical response for wavelength in the visible with respect to other models. The estimations of optical depths has been done based on clear-sky relative humidity every 3 hours. The total aerosols optical depths indicates a slight overestimation over the multi-model central values, due to the overestimations of SO4 optical absorption partially compensated by the expected lower values of OA optical depths.

4.11 NorESM2

The atmosphere module in NorESM2 (NorESM2-MM, see Seland et al., in prep), CAM6-Nor (Olivie et al., in prep.), is an updated version of CAM5.3-Oslo, for which optical properties have been described and validated by Kirkevåg et al. (2018).

Seen in conjunction with these studies, the results presented here can be interpreted as follows. The dust burden is the lowest (5.7 Tg) among the AP3 models, and also low compared to the burden in the un-nudged NorESM2-LM simulation (9.9 Tg), and in CAM5.3-Oslo with fSST and nudged meteorology for year 2000 (16.3 Tg). The lifetime of dust is 1.9 days and is about the same in all these simulations. This is consistently also the lowest among the AP3 models. The large drop in burden from CAM5.3-Oslo and the un-nudged NorESM2 is to a large degree a result of tuned dust emissions, while the change between the un-nudged (1870 Tg/yr) and the nudged (1090 Tg/yr) NorESM simulations with fSST is consistent with the considerably lower U10 (especially over land) and dust emissions in nudged vs. free meteorology. While NorESM sea-salt emissions are among the lowest for AP3, the burden is mid-range, and with the highest MEC (4.1 m2/g), this model has the highest sea-salt AOD values, which is reflected in the positive coarse mode bias against AATSR satellite observations (cf. Fig 10). The relatively high MEC is likely due to SS particle sizes which are shifted towards the more optically efficient accumulation mode, compared to other AP3 models Sea-salt MEC was even higher in CAM5.3-Oslo (5.0 m2/g), but a change in assumed RH (from all-sky to clear-sky) for hygroscopic growth brought about a ca. 19% reduction. The excessive sea-salt AOD is a result of tuning of the CMIP6 control simulation for NorESM2 with respect to radiative balance at TOA. Compared with AERONET (mainly continental stations) AOD is underestimated, particularly by fine mode particles. One possible reason may be that nitrate aerosols and anthropogenic SOA are not taken into account in the model. Despite missing anthropogenic SOA, our OA burden is still among the highest compared to the other models. Due to the overestimated extinction by sea-salt, AE is more underestimated compared to satellite (ocean areas dominate) than to AERONET (mainly continental stations), but the over-all AE bias is close to the AeroCom AP3 mean. The large underestimate in surface scattering and absorption compared to EBAS is consistent with the underestimated AOD over the continents, but as for the majority of the models, the negative bias here is stronger than for the vertically integrated AOD values (compared to AERONET). The high negative bias in surface absorption is consistent with the low dust burden, resulting from the low emissions and short lifetimes (cf. Figs.), compared to the other models (cf. Figs. 5, 6, 7)

4.12 OsloCTM3

The OsloCTM3 is a global, offline CTM driven by 3-hourly meteorological data from the European Centre for Medium Range Weather Forecast (ECMWF) Integrated Forecast System (IFS) model, and is an updated version of the OsloCTM2 used in previous AeroCom phases (Søvde et al. (2012), Lund et al. (2018)). The model is run in a 2.25°x2.25° horizontal resolution, with 60 vertical levels (the uppermost centered at 0.1 hPa) using the Community Emission Data System (CEDS) (Hoesly et al. (2018), van Marle et al. (2017)) emission inventory. The treatment of transport and scavenging, as well as individual aerosol modules, is described in detail in Lund et al. (2018) and references therein. In OsloCTM3, the absorption properties have been updated, with BC mass absorption coefficient (MAC) following formula in Zanatta et al. (2016) and a weak absorption implemented for OA (Lund et al. (2018)). OsloCTM3 has a BC MAC value of 12 m^2/g and BC MEC is among the highest between the models (Fig. 8). The implementation of stronger absorption contributes to the high positive bias (+73%) in surface absorption compared to the surface in-situ observations and in contrast to the other models, which tend to underestimate surface absorption at the in-situ locations (Fig. 10). The burden of nitrate is low, and sulfate high compared to the other

models, whereas all other aerosol species in OsloCTM3 are close to model mean values. An evaluation of the burdens and AOD simulated by the OsloCTM3 for year 2010 CEDS emissions against in-situ and remote sensing observations is provided by Lund et al. (2018). The optical properties for aerosols emitted from biomass burning assume internally mixed aerosol and thus, the reported AOD from BC and OA includes only fossil fuel and biofuel emissions (cf. Fig. 9). This results in lower AOD from OA for OsloCTM3 compared to the other models. The combined BC+OA contribution to AOD amounts to 0.0086. Only all-sky (AS) AOD is provided from OsloCTM3 (cf. Tab. A1 for models that provided CS). This is done because a reliable sub-grid scale parameterisation for RH is unavailable, in order to avoid the AOD used in the radiative transfer calculations to be biased low or high. Compared with the observations, AOD is slightly underestimated, both at AERONET sites (-6%) and the satellite comparisons suggest slightly higher underestimations. The low bias (-20%) for AE is consistent between ground and satellite retrievals and is also reflected in the low bias for coarse and high bias for fine AOD (Fig. 10). In contrast to surface absorption, the surface scattering is biased low compared to observations, which would result in a strong low bias in single scattering albedo. Correlation with the observations is generally among the higher ones compared to the other models (Fig. 11).

4.13 SPRINTARS

SPRINTARS (Takemura et al. (2005, 2009)), coupled with a coupled atmosphere-ocean general circulation model (MIROC, Tatebe et al. (2019)), is used in this study although there is also a version coupled with a global cloud resolving model, NICAM (e.g., Sato et al. (2016)). The calculated dust and sea salt emissions with nudged wind field by meteorological reanalysis data are smaller than those without nudging because the emission amounts strongly depends on the wind speed near the surface (cf. Sect. 4.11), which are proportional to 3rd and 3.41th powers, respectively. The 6-hourly reanalysis data cannot represent the gust of wind. The difference between the case with and without nudging is larger in finer horizontal resolution. SPRINTARS has one of the finest resolutions among the participating models in this study. SPRINTARS estimates the global and annual total emissions of dust and sea salt to be 1390 Tg/yr and 3390 Tg/yr, respectively (cf. Fig. 5) with the horizontal resolution of T85 (approx. $1.4^{\circ} \times 1.4^{\circ}$). Both the lifetime of sea-salt and dust are short compared to the other models (Fig. 7), and in case of dust this may be attributed to strong wet deposition over the outflow regions. This, combined with the low emissions, explains the low burdens of these natural species (Fig. 6 and ultimately explains the fact that all extensive optical properties are underestimated against the various observations (cf. Fig. 10). On the other hand, the calculated AE by SPRINTARS is underestimated, suggesting an overestimation of particle size over the whole particle size distribution. However, for this model, this could be attributed to inappropriate computation of standard deviations of log-normal size distributions of SO4 and OA when calculating optical properties based on the Mie theory.

The diagnosed AE calculated from prognostic mass mixing ratio of each aerosol component from has been confirmed to be around 1.5 over the industrialized and biomass burning regions with the appropriate standard deviations of the size distributions. This revision results in the smaller bias of AOD, and the global annual mean value is 0.112 (as opposed to 0.072 found in this study, cf. Fig. 9).

- AE does not exceed 1, why???
- Underestimated AE suggests particles are too large, but this seems inconsistent with the fact that coarse mode AOD seems to show highest underestimation (around -80%).
- All optical parameters are underestimated agains the various observations.
- Fine mode looks best (-23% bias against AERONET)
- MECs appear to be in range of other models (OA high MEC of 7.3)
- Burden as well (like MEC)
- Seasalt and and SO4 lifetimes low. This may lead to reduced hygroscopic growth and hence underestimated optics.
- Missing natural aerosols?

5 Conclusions

TAKE HOME MESSAGES? ANYONE FEEL FREE TO ADD TO THE BULLET LIST

- AOD underestimated by about -20% on average and lowest spread among the models.
- AE underesimated by about -10%, indicating that particles are too large
- Mention consistency in bias over time (link with Augustins paper)
- Models don't get the size right (AE)
- Mention that the satellite comparisons are consistent with what we know about the models (e.g. high negative coarse bias in SPRINTARS (which misses SS and DU) and vice versa for NorESM, and that this would not be possible to identify easily using only ground based, mostly continental measurements.
- For the first time, columnar remote sensing and surface in-situ optical properties were evaluated.
- Surface in-insitu and coarse mode AOD show highest underestimation
- Missing natural aerosols (i.e. the models with lowest dust and seasalt emissions or shortes residence times showed highest negative bias in coarse AOD).
- More measurements needed for better spatial coverage.
- Mention that some models reported CS AOD, others AS
- ...

Code availability. TEXT

Data availability. TEXT

Code and data availability. TEXT

Appendix A: GAW site evaluation biases

INTRODUCE AND EXPLAIN RESULTS SHOWN IN Figs. A1 and A2.

And why this is shown (because these data has not been used so far for model evaluation and deserves more detailed / careful treatment.

In general, representativity of GAW sites tends to be less sensitive to grid size than AERONET (Wang et al. (2018)).

Appendix B: Sensitivity studies related to spatiotemporal representativity results

As introduced in Sect. 3.4 and summarised in Tab. 3, several tests have been performed in order to investigate the spatiotemporal representativity and associated uncertainties. The results of tests related to temporal representativity errors are shown in Figs. A3, A4, the former being an analysis of monthly vs. 3hourly AOD data vs. AERONET and the latter being an analysis of hourly vs. monthly using surface in-situ absorption data. Both tests do not indicate that the magnitude of these uncertainties in the network-averaged annual statistics exceed 10% in NMB or 0.15 in correlation. Particularly, the results from the in-situ test differ by only 2.4% in NMB which may be attributed to the fact that these data generally shows more continuous sampling coverage throughout the 24h of each day as these techniques do not rely on the availability of sunlight.

An investigation of spatial representativity errors was done for AERONET AODs, by choosing a subset of sites considered representative based on Wang et al. (2018). The result is shown in Fig. A5 and also does not show substantial differences in light of the diversity found in between the models (cf. Figs. 10 and 11).

Appendix C: Pyaerocom and web visualisation

Pyaerocom (Github: https://github.com/metno/pyaerocom, Website: https://pyaerocom.met.no/) is an open source python software and is being developed focussing on model evaluation within the AeroCom initative.

All results from the optical properties evaluation discussed in this paper are available online at:

https://aerocom-evaluation.met.no/overall.php?project=aerocom&exp=PIII-optics2019-P# (last access: 20.12.2019)

The website allows to explore the data from many angles and includes interactive visualisations of performance charts, scatter plots, bias maps and individual station timeseries data, for all models and observation variables, as well as barcharts summarising regional statistics.

BRIEFLY INTRODUCE WEB VISUALISATION

Author contributions. TEXT

Competing interests. TEXT

Disclaimer. TEXT

Acknowledgements. T. T. was supported by the supercomputer system of the National Institute for Environmental Studies, Japan, and JSPS KAKENHI Grant Number JP19H05669. TvN and PLS acknowledge funding from the European Union's Horizon 2020 research and innovation programme project CRESCENDO under grant agreement No 641816. DN acknowledges funding from the European Union's Horizon 2020 research and innovation programme project FORCeS under grant agreement No 821205. SEB and KT acknowledge funding from NASA's Atmospheric Composition Modeling and Analysis Program (ACMAP), contract number NNX15AE36G. They also thank Jingbo Yu for running the model. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. HK acknowledges the Academy of Finland Projects 317390 and 308292. The ECHAM-HAMMOZ model is developed by a consortium composed of ETH Zurich, Max Planck Institut für Meteorologie, Forschungszentrum Jülich, University of Oxford, the Finnish Meteorological Institute and the Leibniz Institute for Tropospheric Research, and managed by the Center for Climate Systems Modeling (C2SM) at ETH Zurich.

- Funding / projects
- Oskar, Hanna
- PI's from observations / data providers

References

Andres, R. J. and Kasgnoc, A. D.: A time-averaged inventory of subaerial volcanic sulfur emissions, Journal of Geophysical Research Atmospheres, https://doi.org/10.1029/98JD02091, 1998.

Balkanski, Y., Schulz, M., Claquin, T., Moulin, C., and Ginoux, P.: Global Emissions of Mineral Aerosol: Formulation and Validation using Satellite Imagery, in: Advances in Global Change Research, pp. 239–267, Springer Netherlands, https://doi.org/10.1007/978-1-4020-2167-1_6, https://doi.org/10.1007%2F978-1-4020-2167-1_6, 2004.

Bergman, T., Kerminen, V.-M., Korhonen, H., Lehtinen, K. J., Makkonen, R., Arola, A., Mielonen, T., Romakkaniemi, S., Kulmala, M., and Kokkola, H.: Evaluation of the sectional aerosol microphysics module SALSA implementation in ECHAM5-HAM aerosol-climate model, Geoscientific Model Development, 5, 845–868, https://doi.org/10.5194/gmd-5-845-2012, https://www.geosci-model-dev.net/5/845/2012/, 2012.

Bevan, S. L., North, P. R. J., Los, S. O., and Grey, W. M. F.: A global dataset of atmospheric aerosol optical depth and surface reflectance from AATSR, Remote Sensing of Environment, 116, 199–210, https://doi.org/https://doi.org/10.1016/j.rse.2011.05.024, http://www.sciencedirect.com/science/article/pii/S0034425711002239, 2012.

Biagio, C. D., Formenti, P., Balkanski, Y., Caponi, L., Cazaunau, M., Pangui, E., Journet, E., Nowak, S., Caquineau, S., Andreae, M. O., Kandler, K., Saeed, T., Piketh, S., Seibert, D., Williams, E., and Doussin, J.-F.: Global scale variability of the mineral dust long-wave refractive index: a new dataset of in situ measurements for climate modeling and remote sensing, Atmospheric Chemistry and Physics, 17, 1901–1929, https://doi.org/10.5194/acp-17-1901-2017, https://doi.org/10.5194%2Facp-17-1901-2017, 2017.

Biagio, C. D., Formenti, P., Balkanski, Y., Caponi, L., Cazaunau, M., Pangui, E., Journet, E., Nowak, S., Andreae, M. O., Kandler, K., Saeed, T., Piketh, S., Seibert, D., Williams, E., and Doussin, J.-F.: Complex refractive indices and single scattering albedo of global dust aerosols in the shortwave spectrum and relationship to iron content and size, https://doi.org/10.5194/acp-2019-145, https://doi.org/10.5194%2Facp-2019-145, 2019.

Bond, T. C. and Bergstrom, R. W.: Light Absorption by Carbonaceous Particles: An Investigative Review, Aerosol Science and Technology, 40, 27–67, https://doi.org/10.1080/02786820500421521, https://doi.org/10.1080/02786820500421521, 2006.

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, Journal of Geophysical Research: Atmospheres, 118, 5380–5552, https://doi.org/10.1002/jgrd.50171, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50171, 2013.

Cappa, C. D., Onasch, T. B., Massoli, P., Worsnop, D. R., Bates, T. S., Cross, E. S., Davidovits, P., Hakala, J., Hayden, K. L., Jobson, B. T., Kolesar, K. R., Lack, D. A., Lerner, B. M., Li, S.-M., Mellon, D., Nuaaman, I., Olfert, J. S., Petäjä, T., Quinn, P. K., Song, C., Subramanian, R., Williams, E. J., and Zaveri, R. A.: Radiative Absorption Enhancements Due to the Mixing State of Atmospheric Black Carbon, Science, 337, 1078–1081, https://doi.org/10.1126/science.1223447, https://science.sciencemag.org/content/337/6098/1078, 2012.

Carn, S. A., Fioletov, V. E., McLinden, C. A., Li, C., and Krotkov, N. A.: A decade of global volcanic SO2 emissions measured from space, Scientific Reports, 7, 44 095, https://doi.org/10.1038/srep44095, https://doi.org/10.1038/srep44095, 2017.

Claquin, T., Schulz, M., Balkanski, Y., and Boucher, O.: Uncertainties in assessing radiative forcing by mineral dust, Tellus B, 50, 491–505, https://doi.org/10.1034/j.1600-0889.1998.t01-2-00007.x, https://onlinelibrary.wiley.com/doi/abs/10.1034/j.1600-0889.1998.t01-2-00007.x, 1998.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.828, 2011.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, https://www.geosci-model-dev.net/9/1937/2016/, 2016.

Flemming, J., Huijnen, V., Arteta, J., Bechtold, P., Beljaars, A., Blechschmidt, A.-M., Diamantakis, M., Engelen, R. J., Gaudel, A., Inness, A., Jones, L., Josse, B., Katragkou, E., Marecal, V., Peuch, V.-H., Richter, A., Schultz, M. G., Stein, O., and Tsikerdekis, A.: Tropospheric chemistry in the Integrated Forecasting System of ECMWF, Geoscientific Model Development, 8, 975–1003, https://doi.org/10.5194/gmd-8-975-2015, https://www.geosci-model-dev.net/8/975/2015/, 2015.

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: Advancements in the Aerosol Robotic Network (AERONET) Version˜3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements, Atmospheric Measurement Techniques, 12, 169–209, https://doi.org/10.5194/amt-12-169-2019, https://www.atmos-meas-tech.net/12/169/2019/, 2019.

Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., and Lin, S.-J.: Sources and distributions of dust aerosols simulated with the GOCART model, Journal of Geophysical Research: Atmospheres, 106, 20 255–20 273, https://doi.org/10.1029/2000JD000053, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD000053, 2001.

Guelle, W., Schulz, M., Balkanski, Y., and Dentener, F.: Influence of the source formulation on modeling the atmospheric global distribution of sea salt aerosol, Journal of Geophysical Research: Atmospheres, 106, 27 509–27 524, https://doi.org/10.1029/2001JD900249, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001JD900249, 2001.

Halmer, M. M., Schmincke, H. U., and Graf, H. F.: The annual volcanic gas input into the atmosphere, in particular into the stratosphere: A global data set for the past 100 years, Journal of Volcanology and Geothermal Research, https://doi.org/10.1016/S0377-0273(01)00318-3, 2002.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geoscientific Model Development, 11, 369–408, https://doi.org/10.5194/gmd-11-369-2018, https://www.geosci-model-dev.net/11/369/2018/, 2018.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sensing of Environment, 66, 1–16, https://doi.org/https://doi.org/10.1016/S0034-4257(98)00031-5, http://www.sciencedirect.com/science/article/pii/S0034425798000315, 1998.

Holben, B. N., Kim, J., Sano, I., Mukai, S., Eck, T. F., Giles, D. M., Schafer, J. S., Sinyuk, A., Slutsker, I., Smirnov, A., Sorokin, M., Anderson, B. E., Che, H., Choi, M., Crawford, J. H., Ferrare, R. A., Garay, M. J., Jeong, U., Kim, M., Kim, W., Knox, N., Li, Z., Lim, H. S., Liu, Y., Maring, H., Nakata, M., Pickering, K. E., Piketh, S., Redemann, J., Reid, J. S., Salinas, S., Seo, S., Tan, F., Tripathi, S. N., Toon, O. B., and Xiao, Q.: An overview of mesoscale aerosol processes, comparisons, and validation studies from DRAGON networks, Atmospheric Chemistry and Physics, 18, 655–671, https://doi.org/10.5194/acp-18-655-2018, https://www.atmos-chem-phys.net/18/655/2018/, 2018.

Hubanks, P. A.: Collection 6.1Change Summary DocumentMODIS Atmosphere Level-3 Algorithmand Global Products, Tech. rep., https://atmosphere-imager.gsfc.nasa.gov/sites/default/files/ModAtmo/L3_C61_Changes_v2.pdf, 2017.

IPCC: Climate Change 2014, https://doi.org/10.1017/CBO9781107415324, 2014.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., and others: The NCEP/NCAR 40-year reanalysis project, Bulletin of the American meteorological Society, 77, 437–472, 1996.

Kim, D., Chin, M., Yu, H., Diehl, T., Tan, Q., Kahn, R. A., Tsigaridis, K., Bauer, S. E., Takemura, T., Pozzoli, L., Bellouin, N., Schulz, M., Peyridieu, S., Chédin, A., and Koffi, B.: Sources, sinks, and transatlantic transport of North African dust aerosol: A multimodel analysis and comparison with remote sensing data, Journal of Geophysical Research: Atmospheres, 119, 6259–6277, https://doi.org/10.1002/2013JD021099, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JD021099, 2014.

Kinne, S., Schulz, M., Textor, C., Guibert, S., Balkanski, Y., Bauer, S. E., Berntsen, T., Berglen, T. F., Boucher, O., Chin, M., Collins, W., Dentener, F., Diehl, T., Easter, R., Feichter, J., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Herzog, M., Horowitz, L., Isaksen, I., Iversen, T., Kirkevåg, A., Kloster, S., Koch, D., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Lesins, G., Liu, X., Lohmann, U., Montanaro, V., Myhre, G., Penner, J. E., Pitari, G., Reddy, S., Seland, O., Stier, P., Takemura, T., and Tie, X.: An AeroCom initial assessment - Optical properties in aerosol component modules of global models, Atmospheric Chemistry and Physics, https://doi.org/10.5194/acp-6-1815-2006, 2006.

Kinne, S., O'Donnel, D., Stier, P., Kloster, S., Zhang, K., Schmidt, H., Rast, S., Giorgetta, M., Eck, T. F., and Stevens, B.: MAC-v1: A new global aerosol climatology for climate studies, Journal of Advances in Modeling Earth Systems, 5, 704–740, https://doi.org/10.1002/jame.20035, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jame.20035, 2013.

Kirkevåg, A., Grini, A., Olivié, D., Seland, , Alterskjær, K., Hummel, M., Karset, I. H. H., Lewinschal, A., Liu, X., Makkonen, R., Bethke, I., Griesfeller, J., Schulz, M., and Iversen, T.: A production-tagged aerosol module for Earth system models, OsloAero5.3 – extensions and updates for CAM5.3-Oslo, Geoscientific Model Development, 11, 3945–3982, https://doi.org/10.5194/gmd-11-3945-2018, https://www.geosci-model-dev.net/11/3945/2018/, 2018.

Klingmüller, K., Steil, B., Brühl, C., Tost, H., and Lelieveld, J.: Sensitivity of aerosol radiative effects to different mixing assumptions in the AEROPT 1.0 submodel of the EMAC atmospheric-chemistry–climate model, Geoscientific Model Development, 7, 2503–2516, https://doi.org/10.5194/gmd-7-2503-2014, https://www.geosci-model-dev.net/7/2503/2014/, 2014.

Kok, J. F., Ridley, D. A., Zhou, Q., Miller, R. L., Zhao, C., Heald, C. L., Ward, D. S., Albani, S., and Haustein, K.: Smaller desert dust cooling effect estimated from analysis of dust size and abundance, Nature Geoscience, 10, 274–278, https://doi.org/10.1038/ngeo2912, https://doi.org/10.1038%2Fngeo2912, 2017.

Kokkola, H., Kühn, T., Laakso, A., Bergman, T., Lehtinen, K. E., Mielonen, T., Arola, A., Stadtler, S., Korhonen, H., Ferrachat, S., Lohmann, U., Neubauer, D., Tegen, I., Siegenthaler-Le Drian, C., Schultz, M. G., Bey, I., Stier, P., Daskalakis, N., Heald, C. L., and Romakkaniemi, S.: SALSA2.0: The sectional aerosol module of the aerosol-chemistry-climate model ECHAM6.3.0-HAM2.3-MOZ1.0, Geoscientific Model Development, https://doi.org/10.5194/gmd-11-3833-2018, 2018.

Lafon, S., Sokolik, I. N., Rajot, J. L., Caquineau, S., and Gaudichet, A.: Characterization of iron oxides in mineral dust aerosols: Implications for light absorption, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD007016, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD007016, 2006.

Laj, P., et, and Al: A global analysis of climate-relevant aerosol properties retrieved from the network of GAW near-surface observatories, 2020.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989–3034, https://doi.org/10.5194/amt-6-2989-2013, https://www.atmos-meas-tech.net/6/2989/2013/, 2013.

Long, M. S., Keene, W. C., Kieber, D. J., Erickson, D. J., and Maring, H.: A sea-state based source function for size- and composition-resolved marine aerosol production, Atmospheric Chemistry and Physics, 11, 1203–1216, https://doi.org/10.5194/acp-11-1203-2011, https://www.atmos-chem-phys.net/11/1203/2011/, 2011.

Lund, M. T., Myhre, G., Haslerud, A. S., Skeie, R. B., Griesfeller, J., Platt, S. M., Kumar, R., Myhre, C. L., and Schulz, M.: Concentrations and radiative forcing of anthropogenic aerosols from 1750 to 2014 simulated with the Oslo˜CTM3 and CEDS emission inventory, Geoscientific Model Development, 11, 4909–4931, https://doi.org/10.5194/gmd-11-4909-2018, https://www.geosci-model-dev.net/11/4909/2018/, 2018.

Mårtensson, E. M., Nilsson, E. D., de Leeuw, G., Cohen, L. H., and Hansson, H.-C.: Laboratory simulations and parameterization of the primary marine aerosol production, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD002263, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002JD002263, 2003.

Matsui, H.: Development of a global aerosol model using a two-dimensional sectional method: 1. Model design, Journal of Advances in Modeling Earth Systems, 9, 1921–1947, https://doi.org/10.1002/2017MS000936, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS000936, 2017.

Monahan, E. C., Spiel, D. E., and Davidson, K. L.: A Model of Marine Aerosol Generation Via Whitecaps and Wave Disruption BT - Oceanic Whitecaps: And Their Role in Air-Sea Exchange Processes, pp. 167–174, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-009-4668-2_16, https://doi.org/10.1007/978-94-009-4668-2_16, 1986.

Morcrette, J.-J., Boucher, O., Jones, L., Salmond, D., Bechtold, P., Beljaars, A., Benedetti, A., Bonet, A., Kaiser, J. W., Razinger, M., Schulz, M., Serrar, S., Simmons, A. J., Sofiev, M., Suttie, M., Tompkins, A. M., and Untch, A.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: Forward modeling, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2008JD011235, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD011235, 2009.

North, P. R. J.: Estimation of aerosol opacity and land surface bidirectional reflectance from ATSR-2 dual-angle imagery: Operational method and validation, Journal of Geophysical Research: Atmospheres, 107, 4–1, https://doi.org/10.1029/2000JD000207, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD000207, 2002.

North, P. R. J. and Heckel, A.: SU-ATSR Algorithm Theoretical Basis Document (ATBD) v4.3, Tech. rep., http://cci.esa.int/sites/default/files/Aerosol_cci2_ATBD_ATSR_SU_v4.3.pdf, 2017.

O'Neill, N. T., Eck, T. F., Smirnov, A., Holben, B. N., and Thulasiraman, S.: Spectral discrimination of coarse and fine mode optical depth, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD002975, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002JD002975, 2003.

Popp, T., de Leeuw, G., Bingen, C., Brühl, C., Capelle, V., Chedin, A., Clarisse, L., Dubovik, O., Grainger, R., Griesfeller, J., Heckel, A., Kinne, S., Klüser, L., Kosmale, M., Kolmonen, P., Lelli, L., Litvinov, P., Mei, L., North, P., Pinnock, S., Povey, A., Robert, C., Schulz,

M., Sogacheva, L., Stebel, K., Stein Zweers, D., Thomas, G., Tilstra, L., Vandenbussche, S., Veefkind, P., Vountas, M., and Xue, Y.: Development, Production and Evaluation of Aerosol Climate Data Records from European Satellite Observations (Aerosol_cci), Remote Sensing, 8, 421, https://doi.org/10.3390/rs8050421, http://www.mdpi.com/2072-4292/8/5/421, 2016.

Rémy, S., Kipling, Z., Flemming, J., Boucher, O., Nabat, P., Michou, M., Bozzo, A., Ades, M., Huijnen, V., Benedetti, A., Engelen, R., Peuch, V.-H., and Morcrette, J.-J.: Description and evaluation of the tropospheric aerosol scheme in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS-AER, cycle 45R1), Geoscientific Model Development, 12, 4627–4659, https://doi.org/10.5194/gmd-12-4627-2019, https://www.geosci-model-dev.net/12/4627/2019/, 2019.

Samset, B. H., Stjern, C. W., Andrews, E., Kahn, R. A., Myhre, G., Schulz, M., and Schuster, G. L.: Aerosol Absorption: Progress Towards Global and Regional Constraints, Current Climate Change Reports, 4, 65–83, https://doi.org/10.1007/s40641-018-0091-4, https://doi.org/10.1007/s40641-018-0091-4, 2018.

Sato, Y., Miura, H., Yashiro, H., Goto, D., Takemura, T., Tomita, H., and Nakajima, T.: Unrealistically pristine air in the Arctic produced by current global scale models, Scientific Reports, 6, 26 561, https://doi.org/10.1038/srep26561, https://doi.org/10.1038/srep26561, 2016.

Sayer, A. M. and Knobelspiesse, K. D.: How should we aggregate data? Methods accounting for the numerical distributions, with an assessment of aerosol optical depth, Atmospheric Chemistry and Physics, 19, 15 023–15 048, https://doi.org/10.5194/acp-19-15023-2019, https://www.atmos-chem-phys.net/19/15023/2019/, 2019.

Sayer, A. M., Munchak, L. A., Hsu, N. C., Levy, R. C., Bettenhausen, C., and Jeong, M.-J.: MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and "merged" data sets, and usage recommendations, Journal of Geophysical Research: Atmospheres, 119, 913–965, https://doi.org/10.1002/2014JD022453, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JD022453, 2014.

Schultz, M. G., Stadtler, S., Schröder, S., Taraborrelli, D., Franco, B., Krefting, J., Henrot, A., Ferrachat, S., Lohmann, U., Neubauer, D., Siegenthaler-Le Drian, C., Wahl, S., Kokkola, H., Kühn, T., Rast, S., Schmidt, H., Stier, P., Kinnison, D., Tyndall, G. S., Orlando, J. J., and Wespes, C.: The chemistry–climate model ECHAM6.3-HAM2.3-MOZ1.0, Geoscientific Model Development, 11, 1695–1723, https://doi.org/10.5194/gmd-11-1695-2018, https://www.geosci-model-dev.net/11/1695/2018/, 2018.

Schulz, M., Cozic, A., and Szopa, S.: LMDzT-INCA dust forecast model developments and associated validation efforts, IOP Conference Series: Earth and Environmental Science, 7, 12 014, https://doi.org/10.1088/1755-1307/7/1/012014, https://doi.org/10.1088%2F1755-1307%2F7%2F1%2F012014, 2009.

Schulz, M., Steensen, B. M., Tsyro, S., and Koffi, B.: First results of comparison modelled aerosol extinction profiles with CALIOP data, in: Transboundary acidification, eutrophication and ground level ozone in Europe in 2010. EMEP Status Report 1/2012, pp. 113–121, Norwegian Meteorological Institute, https://emep.int/publ/reports/2012/status_report_1_2012.pdf, 2012.

Schuster, G. L., Dubovik, O., and Holben, B. N.: Angstrom exponent and bimodal aerosol size distributions, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD006328, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006328, 2006.

Schutgens, N., Tsyro, S., Gryspeerdt, E., Goto, D., Weigum, N., Schulz, M., and Stier, P.: On the spatio-temporal representativeness of observations, Atmospheric Chemistry and Physics, 17, 9761–9780, https://doi.org/10.5194/acp-17-9761-2017, https://www.atmos-chem-phys.net/17/9761/2017/, 2017.

Schutgens, N. A. J., Partridge, D. G., and Stier, P.: The importance of temporal collocation for the evaluation of aerosol models with observations, Atmospheric Chemistry and Physics, 16, 1065–1079, https://doi.org/10.5194/acp-16-1065-2016, https://www.atmos-chem-phys.net/16/1065/2016/, 2016.

Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Ny\'\iri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, , and Wind, P.: The EMEP MSC-W chemical transport model – technical description, Atmospheric Chemistry and Physics, 12, 7825–7865, https://doi.org/10.5194/acp-12-7825-2012, https://www.atmos-chem-phys.net/12/7825/2012/, 2012.

Sofiev, M., Soares, J., Prank, M., de Leeuw, G., and Kukkonen, J.: A regional-to-global model of emission and transport of sea salt particles in the atmosphere, Journal of Geophysical Research: Atmospheres, 116, https://doi.org/10.1029/2010JD014713, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD014713, 2011.

Sogacheva, L., Popp, T., Sayer, A. M., Dubovik, O., Garay, M. J., Heckel, A., Hsu, N. C., Jethva, H., Kahn, R. A., Kolmonen, P., Kosmale, M., de Leeuw, G., Levy, R. C., Litvinov, P., Lyapustin, A., North, P., and Torres, O.: Merging regional and global AOD records from 15 available satellite products, Atmospheric Chemistry and Physics Discussions, 2019, 1–62, https://doi.org/10.5194/acp-2019-446, https://www.atmos-chem-phys-discuss.net/acp-2019-446/, 2019.

Søvde, O. A., Prather, M. J., Isaksen, I. S. A., Berntsen, T. K., Stordal, F., Zhu, X., Holmes, C. D., and Hsu, J.: The chemical transport model Oslo CTM3, Geoscientific Model Development, 5, 1441–1469, https://doi.org/10.5194/gmd-5-1441-2012, https://www.geosci-model-dev.net/5/1441/2012/, 2012.

Takemura, T., Nozawa, T., Emori, S., Nakajima, T. Y., and Nakajima, T.: Simulation of climate response to aerosol direct and indirect effects with aerosol transport-radiation model, Journal of Geophysical Research: Atmospheres, 110, https://doi.org/10.1029/2004JD005029, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JD005029, 2005.

Takemura, T., Egashira, M., Matsuzawa, K., Ichijo, H., O'ishi, R., and Abe-Ouchi, A.: A simulation of the global distribution and radiative forcing of soil dust aerosols at the Last Glacial Maximum, Atmospheric Chemistry and Physics, 9, 3061–3073, https://doi.org/10.5194/acp-9-3061-2009, https://www.atmos-chem-phys.net/9/3061/2009/, 2009.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O'ishi, R., Yamazaki, D., Suzuki, T., Kurogi, M., Kataoka, T., Watanabe, M., and Kimoto, M.: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6, Geoscientific Model Development, 12, 2727–2765, https://doi.org/10.5194/gmd-12-2727-2019, https://www.geosci-model-dev.net/12/2727/2019/, 2019.

Tegen, I., Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Bey, I., Schutgens, N., Stier, P., Watson-Parris, D., Stanelle, T., Schmidt, H., Rast, S., Kokkola, H., Schultz, M., Schroeder, S., Daskalakis, N., Barthel, S., Heinold, B., and Lohmann, U.: The global aerosol–climate model ECHAM6.3–HAM2.3 – Part 1: Aerosol evaluation, Geoscientific Model Development, 12, 1643–1677, https://doi.org/10.5194/gmd-12-1643-2019, https://www.geosci-model-dev.net/12/1643/2019/, 2019.

Textor, C., Graf, H.-F., Timmreck, C., and Robock, A.: Emissions from volcanoes, in: Emissions of Atmospheric Trace Compounds, edited by Granier, C., Artaxo, P., and Reeves, C. E., pp. 269–303, Springer Netherlands, Dordrecht, 2004.

Textor, C., Schulz, M., Guibert, S., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Easter, R., Feichter, H., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Horowitz, L., Huang, P., Isaksen, I., Iversen, I., Kloster, S., Koch, D., Kirkevåg, A., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Liu, X., Montanaro, V., Myhre, G., Penner, J., Pitari, G., Reddy, S., Seland, , Stier, P., Takemura, T., and Tie, X.: Analysis and quantification of the diversities of aerosol life cycles within AeroCom, Atmospheric Chemistry and Physics, 6, 1777–1813, https://doi.org/10.5194/acp-6-1777-2006, https://www.atmos-chem-phys.net/6/1777/2006/, 2006.

Tsigaridis, K., Daskalakis, N., Kanakidou, M., Adams, P. J., Artaxo, P., Bahadur, R., Balkanski, Y., Bauer, S. E., Bellouin, N., Benedetti, A., Bergman, T., Berntsen, T. K., Beukes, J. P., Bian, H., Carslaw, K. S., Chin, M., Curci, G., Diehl, T., Easter, R. C., Ghan, S. J., Gong, S. L., Hodzic, A., Hoyle, C. R., Iversen, T., Jathar, S., Jimenez, J. L., Kaiser, J. W., Kirkevåg, A., Koch, D., Kokkola, H., Lee, Y. H., Lin, G., Liu, X., Luo, G., Ma, X., Mann, G. W., Mihalopoulos, N., Morcrette, J.-J., Müller, J.-F., Myhre, G., Myriokefalitakis, S., Ng, N. L., O'Donnell, D., Penner, J. E., Pozzoli, L., Pringle, K. J., Russell, L. M., Schulz, M., Sciare, J., Seland, , Shindell, D. T., Sillman, S., Skeie, R. B., Spracklen, D., Stavrakou, T., Steenrod, S. D., Takemura, T., Tiitta, P., Tilmes, S., Tost, H., van Noije, T., van Zyl, P. G., von Salzen, K., Yu, F., Wang, Z., Wang, Z., Zaveri, R. A., Zhang, H., Zhang, K., Zhang, Q., and Zhang, X.: The AeroCom evaluation and intercomparison of organic aerosol in global models, Atmospheric Chemistry and Physics, 14, 10 845–10 895, https://doi.org/10.5194/acp-14-10845-2014, https://www.atmos-chem-phys.net/14/10845/2014/, 2014.

van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., Arneth, A., Forrest, M., Hantson, S., Kehrwald, N. M., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Yue, C., Kaiser, J. W., and van der Werf, G. R.: Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015), Geoscientific Model Development, 10, 3329–3357, https://doi.org/10.5194/gmd-10-3329-2017, https://www.geosci-model-dev.net/10/3329/2017/, 2017.

van Noije, T. P. C., Le Sager, P., Segers, A. J., van Velthoven, P. F. J., Krol, M. C., Hazeleger, W., Williams, A. G., and Chambers, S. D.: Simulation of tropospheric chemistry and aerosols with the climate model EC-Earth, Geoscientific Model Development, 7, 2435–2475, https://doi.org/10.5194/gmd-7-2435-2014, https://www.geosci-model-dev.net/7/2435/2014/, 2014.

van Noije, T. P. C. e. a.: EC-Earth3-AerChem, a global climate model with interactive aerosols and atmospheric chemistry for use in CMIP6, 2019.

Vignati, E., Wilson, J., and Stier, P.: M7: An efficient size-resolved aerosol microphysics module for large-scale aerosol transport models, Journal of Geophysical Research: Atmospheres, 109, https://doi.org/10.1029/2003JD004485, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003JD004485, 2004.

Wang, R., Balkanski, Y., Boucher, O., Ciais, P., Schuster, G. L., Chevallier, F., Samset, B. H., Liu, J., Piao, S., Valari, M., and Tao, S.: Estimation of global black carbon direct radiative forcing and its uncertainty constrained by observations, Journal of Geophysical Research: Atmospheres, 121, 5948–5971, https://doi.org/10.1002/2015jd024326, https://doi.org/10.1002%2F2015jd024326, 2016.

Wang, R., Andrews, E., Balkanski, Y., Boucher, O., Myhre, G., Samset, B. H., Schulz, M., Schuster, G. L., Valari, M., and Tao, S.: Spatial Representativeness Error in the Ground-Level Observation Networks for Black Carbon Radiation Absorption, Geophysical Research Letters, 45, 2106–2114, https://doi.org/10.1002/2017GL076817, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL076817, 2018.

Williams, J. E., Boersma, K. F., Le Sager, P., and Verstraeten, W. W.: The high-resolution version of TM5-MP for optimized satellite retrievals: description and validation, Geoscientific Model Development, 10, 721–750, https://doi.org/10.5194/gmd-10-721-2017, https://www.geosci-model-dev.net/10/721/2017/, 2017.

Zanatta, M., Gysel, M., Bukowiecki, N., Müller, T., Weingartner, E., Areskoug, H., Fiebig, M., Yttri, K. E., Mihalopoulos, N., Kouvarakis, G., Beddows, D., Harrison, R. M., Cavalli, F., Putaud, J. P., Spindler, G., Wiedensohler, A., Alastuey, A., Pandolfi, M., Sellegri, K., Swietlicki, E., Jaffrezo, J. L., Baltensperger, U., and Laj, P.: A European aerosol phenomenology-5: Climatology of black carbon optical properties at 9 regional background sites across Europe, Atmospheric Environment, 145, 346–364, https://doi.org/https://doi.org/10.1016/j.atmosenv.2016.09.035, http://www.sciencedirect.com/science/article/pii/S135223101630735X, 2016.

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., Chen, J.-H., Chen, X., Donner, L. J., Dunne, J. P., Dunne, K., Durachta, J., Fan, S.-M., Freidenreich, S. M., Garner, S. T., Ginoux, P., Harris, L. M., Horowitz, L. W., Krasting, J. P., Langenhorst, A. R., Liang, Z.,

Lin, P., Lin, S.-J., Malyshev, S. L., Mason, E., Milly, P. C. D., Ming, Y., Naik, V., Paulot, F., Paynter, D., Phillipps, P., Radhakrishnan, A., Ramaswamy, V., Robinson, T., Schwarzkopf, D., Seman, C. J., Shevliakova, E., Shen, Z., Shin, H., Silvers, L. G., Wilson, J. R., Winton, M., Wittenberg, A. T., Wyman, B., and Xiang, B.: The GFDL Global Atmosphere and Land Model AM4.0/LM4.0: 1. Simulation Characteristics With Prescribed SSTs, Journal of Advances in Modeling Earth Systems, 10, 691–734, https://doi.org/10.1002/2017MS001208, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001208, 2018.

Table 1. Observations and optical property variables used in this study, including other relevant meta information.

ID Date	Source	Var.	Ver.	Lev.	Freq.	Clim.	#st.
GAW 2019/12/18	EBAS	Abs. coeff.		3	hourly	Y	39
GAW 2019/12/18	EBAS	Sc. coeff.		3	hourly	Y	37
AERONET 2019/09/20	Sun	AE	3	2	daily	N	250
AERONET 2019/09/20	Sun	AOD	3	2	daily	N	240
AERONET 2019/09/20	SDA	AOD<1μm	3	2	daily	N	226
AERONET 2019/09/20	SDA	AOD>1μm	3	2	daily	N	226
MODISt 2019/11/22	terra	AOD	6.1	3	daily	N	2235
MODISa 2019/11/25	aqua	AOD	6.1	3	daily	N	2241
ATSR-SU 2016/09/30	Swansea	AOD	4.3	3	daily	N	2055
ATSR-SU 2016/09/30	Swansea	AE	4.3	3	daily	N	2055
ATSR-SU 2016/09/30	Swansea	AOD<1μm	4.3	3	daily	N	2055
ATSR-SU 2016/09/30	Swansea	AOD>1μm	4.3	3	daily	N	2055
MERGED-FMI 2019/10/21	FMI	AOD			monthly	N	2080

Table 2. Models used in this study including horizontal grid resolution, number of levels and references. Insert more references and link them properly; Remove AEROCOM-MEAN here

Name	Res (Lat., Lon.)	Levels	References
CAM5-ATRAS	1.9 x 2.5	30	
EC-Earth	2.0 x 3.0	34	Van Noije, T.P.C., et al. (Geosci. Model Dev., 7, 2435-2475, 2014); Van Noije, T.P.C., et al., (manuscript in preparation)
TM5	2.0 x 3.0	34	Van Noije, T.P.C., et al. (Geosci. Model Dev., 7, 2435-2475, 2014); Bergman, T., et al., (manuscript in preparation)
ECHAM-HAM	1.9 x 1.9	47	
ECHAM-SALSA	1.9 x 1.9	47	
ECMWF-IFS	0.4 x 0.4		
EMEP	0.5 x 0.5		
GEOS	1.0 x 1.0	72	
GFDL-AM4	1.0 x 1.2	33	
GISS-OMA	2.0 x 2.5	40	Koch et al., 2006; 2007; Tsigaridis et al., 2013
INCA	1.3 x 2.5	79	
NorESM2	0.9 x 1.2	32	
OsloCTM3	2.2 x 2.2	60	Myhre et al. 2009 (ACP, 9, 1365-1392); Lund et al., 2018 (GMD, 11, 4909-4931)
SPRINTARS	0.6 x 0.6	56	Takemura et al. (J. Geophys. Res., 2000, 17853-17873) (J. Climate, 2002, 333-352) (J. Geophys. Res., 2005, 2004JD005029) (Atmos. Chem. Phys., 9, 3061-3073)

Table 3. Results from sensitivity studies related to spatio-temporal representation errors. AERONET* indicates that two different site selection schemes where used (cf. text and Fig. A5). See also Tab. A2 for an assessment of satellite resampling sensitivities.

Test type	Var.	Model	Freq.	Obs	$\Delta_{\mathrm{NMB}}\,[\%]$	Δ_R	Fig.
Temporal	Abs. coeff.	TM5 (INSITU)	hourly	In-situ (GAW)	-2.3	+0.20	A4
Temporal	AOD	ECMWF-IFS	3-hourly	AERONET	+6.9	-0.10	A3
Spatial	AOD	ENSEMBLE	monthly	AERONET*	-3.6	-0.04	A5

Table A1. Model names and corresponding AeroCom IDs. Also indicated is whether the AODs were diagnosed as clear-sky (CS) or not and which models were included in the ensemble.

Name	AeroCom ID	CS AOD	Ensemble
CAM5-ATRAS	CAM5-ATRAS_AP3-CTRL	Y	Y
EC-Earth	EC-Earth3-AerChem-met2010_AP3-CTRL2019	Y	Y
TM5	TM5-met2010_AP3-CTRL2019	Y	Y
ECHAM-HAM	ECHAM6.3-HAM2.3-met2010_AP3-CTRL	Y	Y
ECHAM-SALSA	ECHAM6.3-SALSA2.0-met2010_AP3-CTRL	Y	Y
ECMWF-IFS	ECMWF-IFS-CY45R1-CAMS-CTRL-met2010_AP3-CTRL	Y	N
EMEP	EMEP_rv4_33_Glob-CTRL	N	Y
GEOS	GEOS-i33p2-met2010_AP3-CTRL	N	Y
GFDL-AM4	GFDL-AM4-met2010_AP3-CTRL	N	Y
GISS-OMA	GISS-ModelE2p1p1-OMA_AP3-fSST	Y	Y
INCA	INCA_AP3-CTRL	Y	N
NorESM2	NorESM2-met2010_AP3-CTRL	Y	Y
OsloCTM3	OsloCTM3v1.01-met2010_AP3-CTRL	N	Y
SPRINTARS	MIROC-SPRINTARS_AP3-CTRL	Y	Y

Table A2. Comparison of statistical parameters (NMB and R) retrieved when colocating models with satellite data 1. in daily resolution and in highest available horizontal resolution from both data-sets (*High* and 2. when colocating in monthly resolution and $5 \times 5^{\circ}$ horizontally with requirement of at least 7 daily values to compute a monthly mean, as done in this study (*Low*).

		Stat	NMB [%]		R	
		Res	Low	High	Low	High
Mod	Sat	Var				
CAM5-ATRAS	AATSR4.3-SU	AOD	-1.75	-2.14	0.67	0.51
	MODIS6.1-aqua	AOD	-25.88	-20.40	0.58	0.36
	MODIS6.1-terra	AOD	-33.28	-28.21	0.58	0.36
ECMWF-IFS	AATSR4.3-SU	AE	-47.25	-36.86	0.74	0.65
		AOD	-19.20	-22.78	0.79	0.70
	MODIS6.1-aqua	AOD	-35.47	-24.37	0.64	0.51
	MODIS6.1-terra	AOD	-41.92	-31.49	0.62	0.51
EMEP	AATSR4.3-SU	AE	36.57	42.61	0.67	0.50
		AOD	-34.40	-30.38	0.73	0.58
		AOD<1μm	-10.88	-2.56	0.74	0.57
		AOD>1μm	-69.81	-69.25	0.64	0.54
	MODIS6.1-aqua	AOD	-45.39	-40.26	0.66	0.48
	MODIS6.1-terra	AOD	-50.77	-45.66	0.66	0.48
OsloCTM3	AATSR4.3-SU	AOD	-12.35	-13.43	0.83	0.69
	MODIS6.1-aqua	AOD	-27.38	-28.25	0.72	0.52
	MODIS6.1-terra	AOD	-34.56	-35.21	0.72	0.51
SPRINTARS	AATSR4.3-SU	AE	-51.16	-41.39	0.59	0.52
TM5	AATSR4.3-SU	AE	2.92	8.71	0.74	0.62
		AOD	-1.84	-3.93	0.75	0.55
		AOD<1μm	3.42	3.42	0.81	0.66
		AOD>1μm	-9.76	-14.64	0.64	0.41
	MODIS6.1-aqua	AOD	-19.93	-18.23	0.73	0.53
	MODIS6.1-terra	AOD	-27.80	-25.95	0.72	0.52

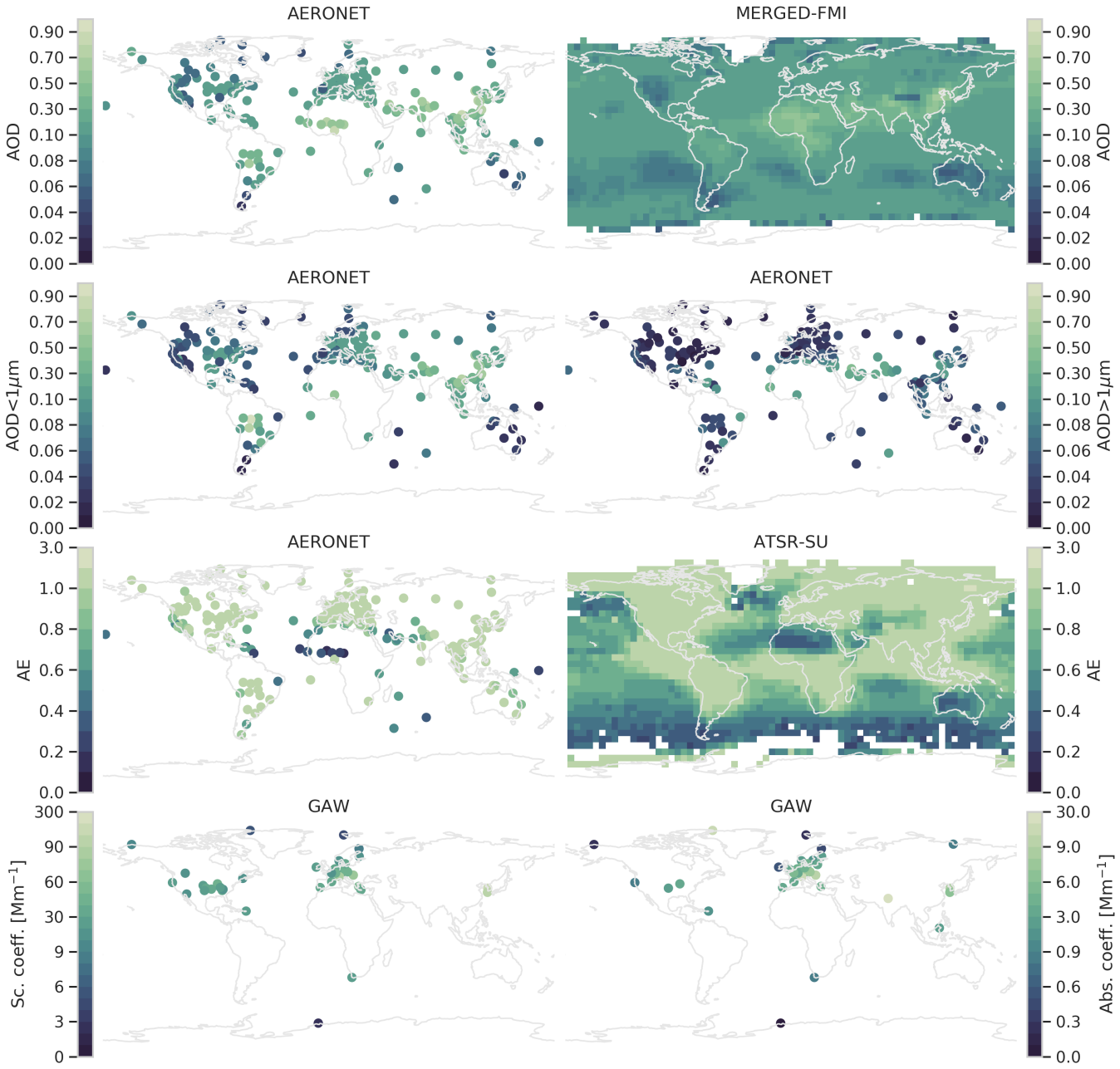

Figure 1. Yearly averages of AODs from AERONET and merged satellite dataset (top panel), fine and coarse AOD from AERONET (2nd panel), AE from AERONET and ATSR (3rd panel) as well as surface in-situ observations of scattering and absorption coefficients.

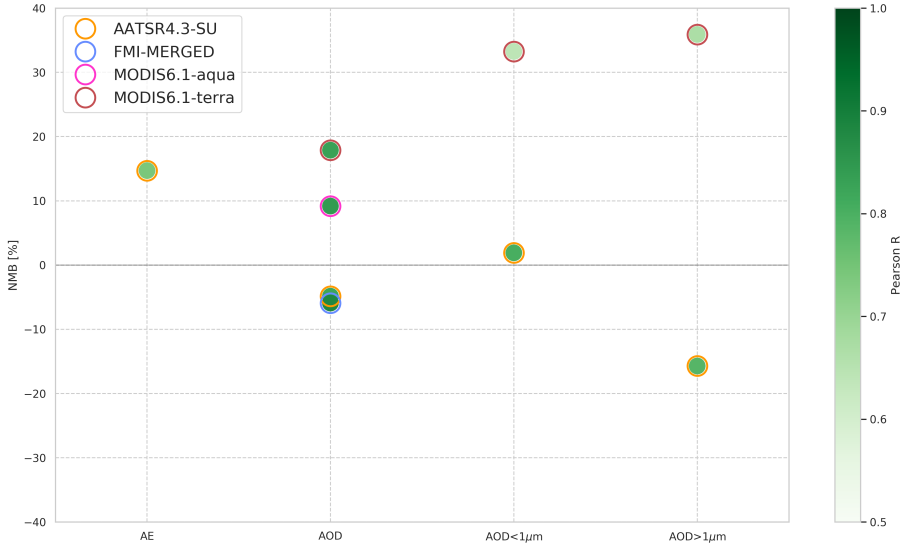

Figure 2. NMBs from satellite evaluation against AERONET for different variables. Also plotted are the corresponding correlation coefficients in green colors. Note that fine and coarse AOD from MODIS terra is not further used in this study.

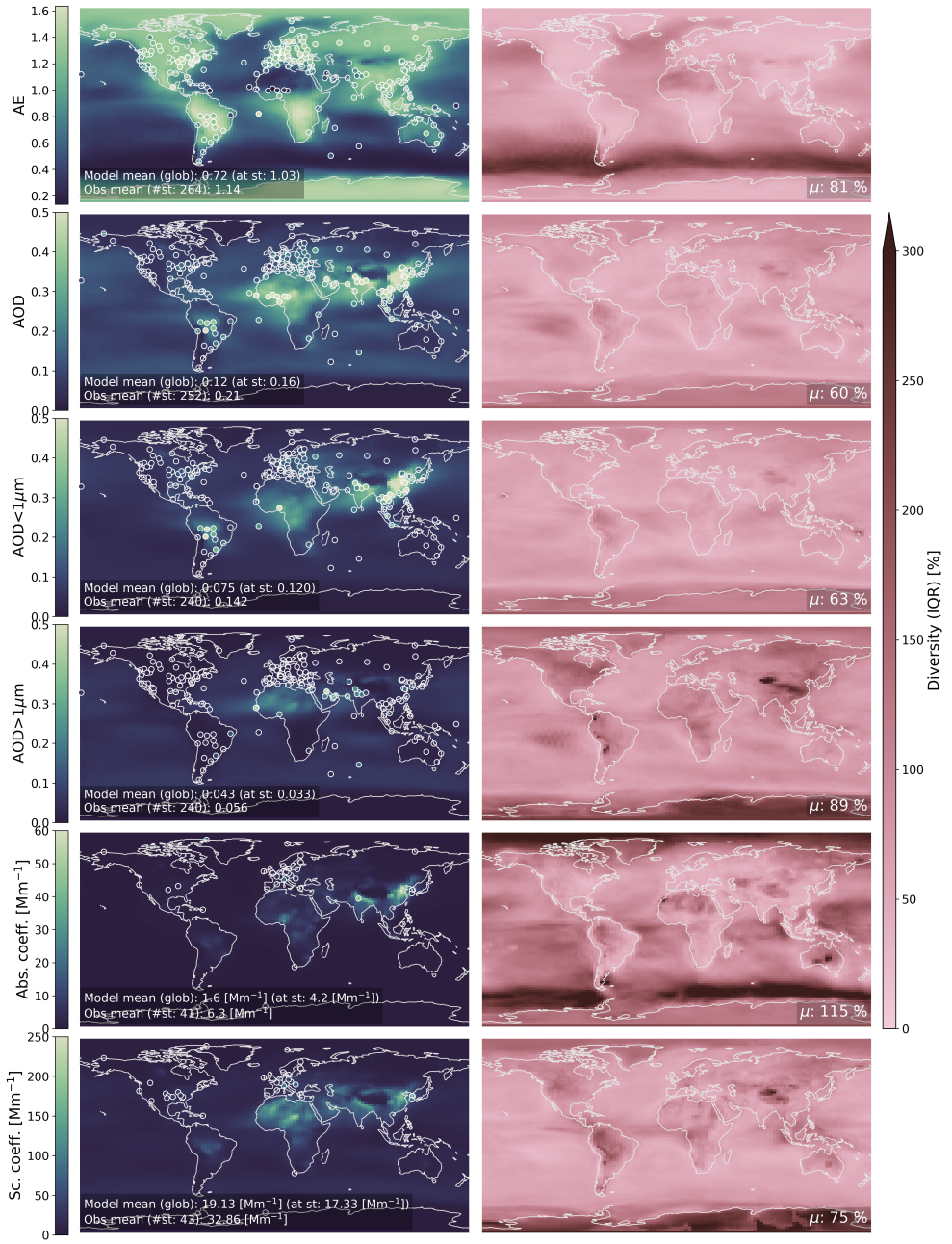

Figure 3. Left: maps showing yearly averages of relevant variables from the ensemble model as well as mean values from corresponding ground-based network used (circles). Also shown are the yearly mean values from model (both global and at obs. stations) as well as the observation mean from all stations. Right: diversity fields of ensemble mean calculated using standard deviation of the individual results normalised by the mean (cf. Textor et al. (2006))

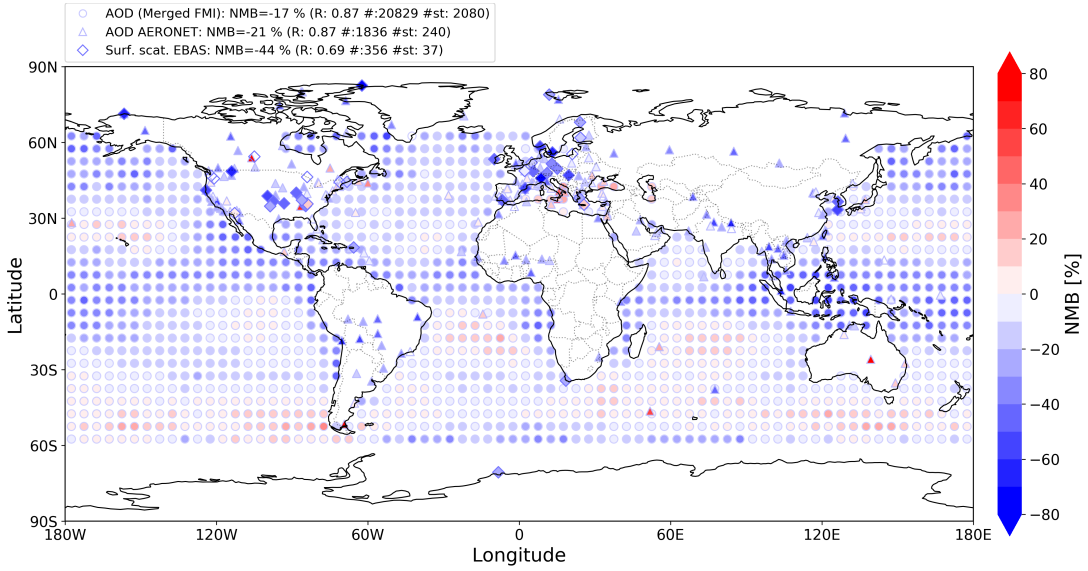

Figure 4. Figure showing NMB in percent of the ensemble median AOD against the merged satellite AOD dataset (circles, only ocean locations are displayed) and AERONET (triangles) as well as surface scattering coefficient against the in-situ sites (diamond). The edge colors of the markers correspond to the respective global average NMB, which is also indicated in the legend as well as Pearson correlation and total number of data points and stations / coordinates.

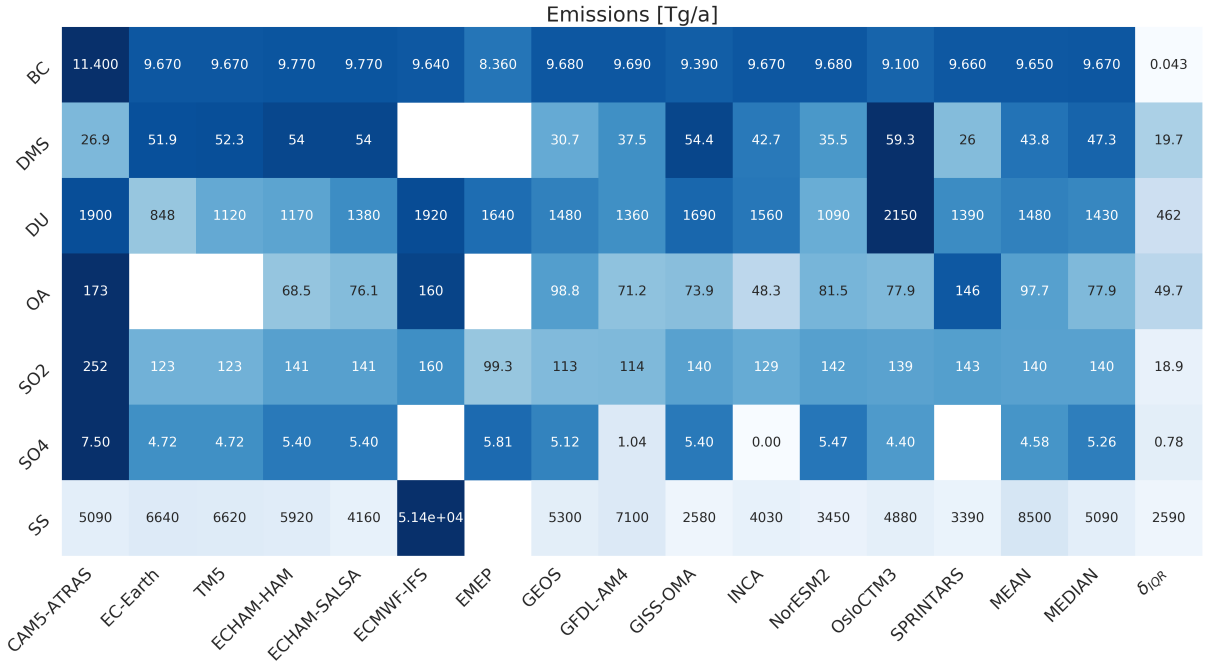

Figure 5. Global emissions of major aerosol species and precursors. Units are full molecular weight and for OA, the total organic weight is used. Note that only major species are included and that other potentially provided species (e.g. NO_x or NH_3, VOCs) are not shown. The rightmost columns show mean, median and spread of the results from the individual models, the latter being computed as the half difference between 1st and 3rd quantiles. Note that displayed precision varies for each row and colors are applied row-wise in order to highlight differences between the models.

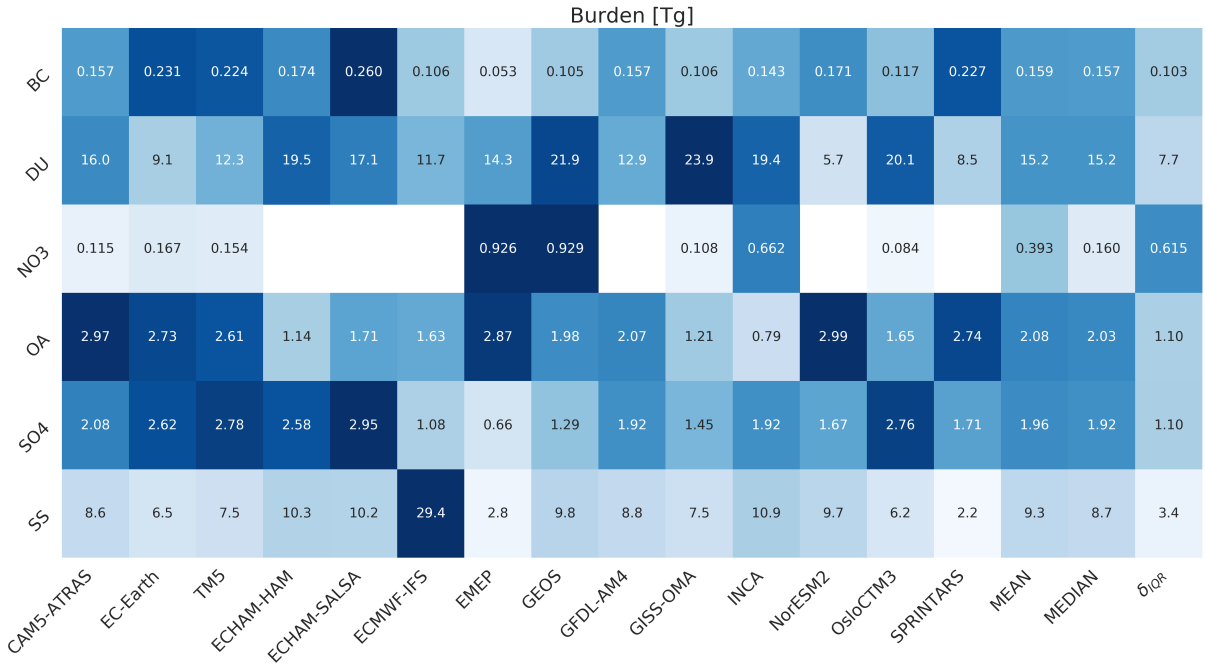

Figure 6. Global annual burdens of major aerosol species in units of Tg. A more detailed description of this plot type is provided in Fig. 5.

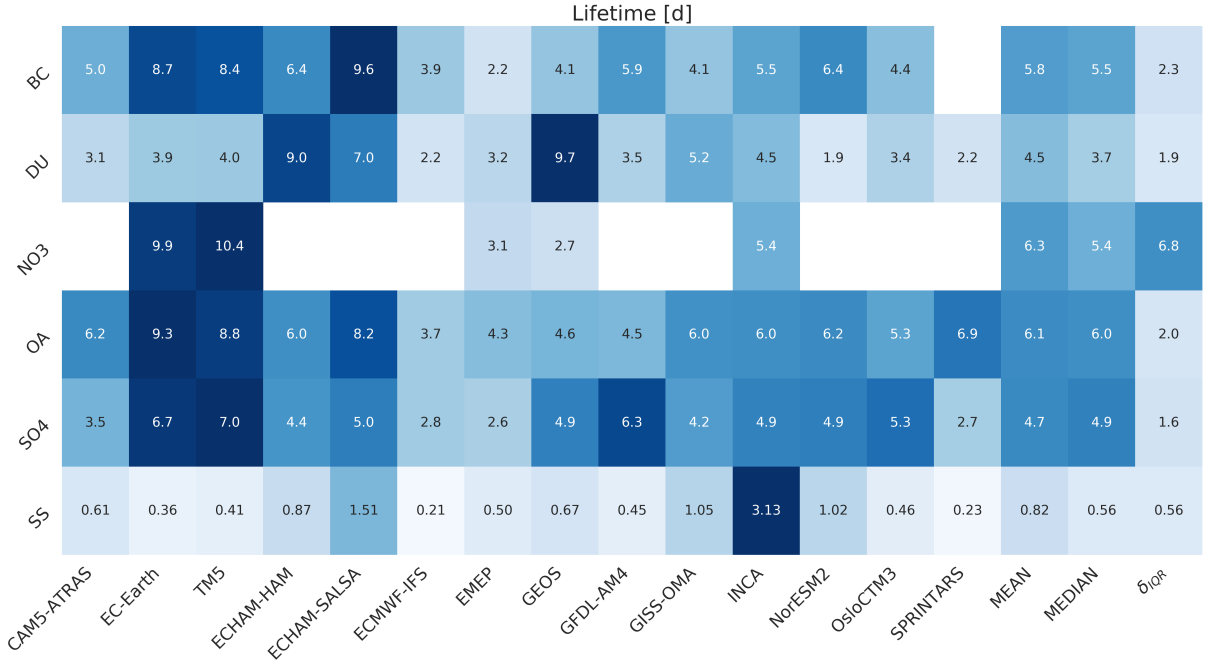

Figure 7. Global lifetimes in days of aerosol species computed from burdens (Fig. 6 and total deposition (wet + dry). A more detailed description of this plot type is provided in Fig. 5.

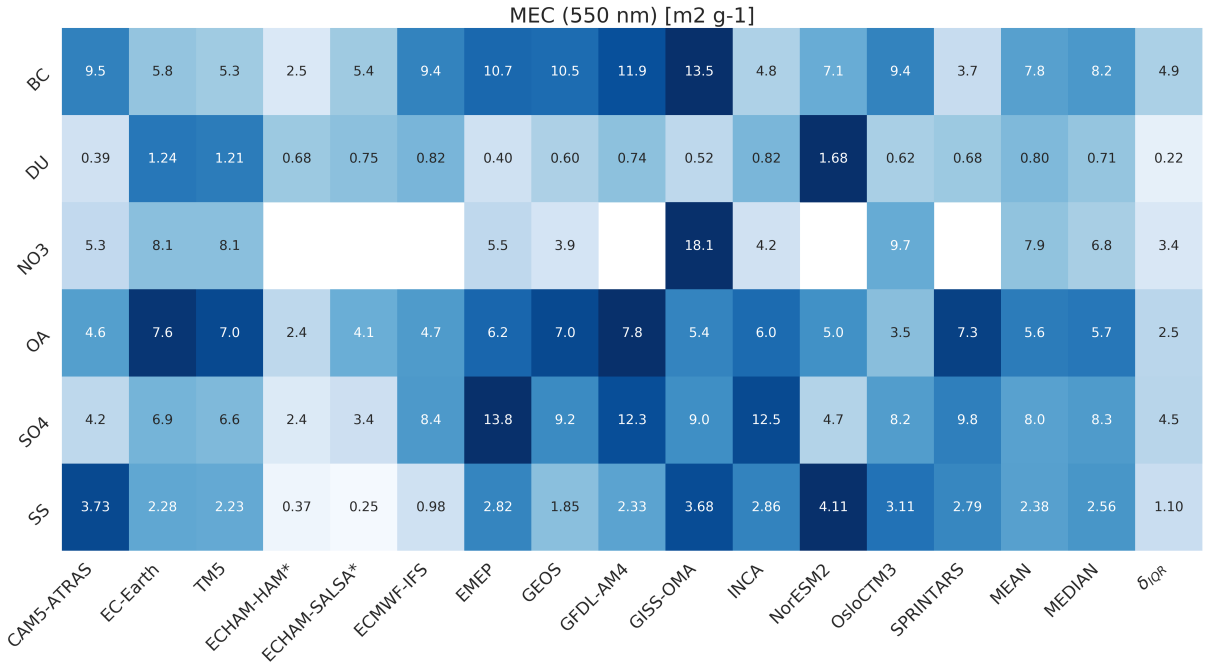

Figure 8. Globally averaged columnar MECs of models for all major aerosol species. The MECs for each species i were computed via $OD_i/LOAD_i$ (cf. Figs. 6, 9). Note that the two ECHAM models reported the OD_i fields at dry conditions (Fig. 9) and show hence, comparatively small MECs for the hygrophilic species. A more detailed description of this plot type is provided in Fig. 5.

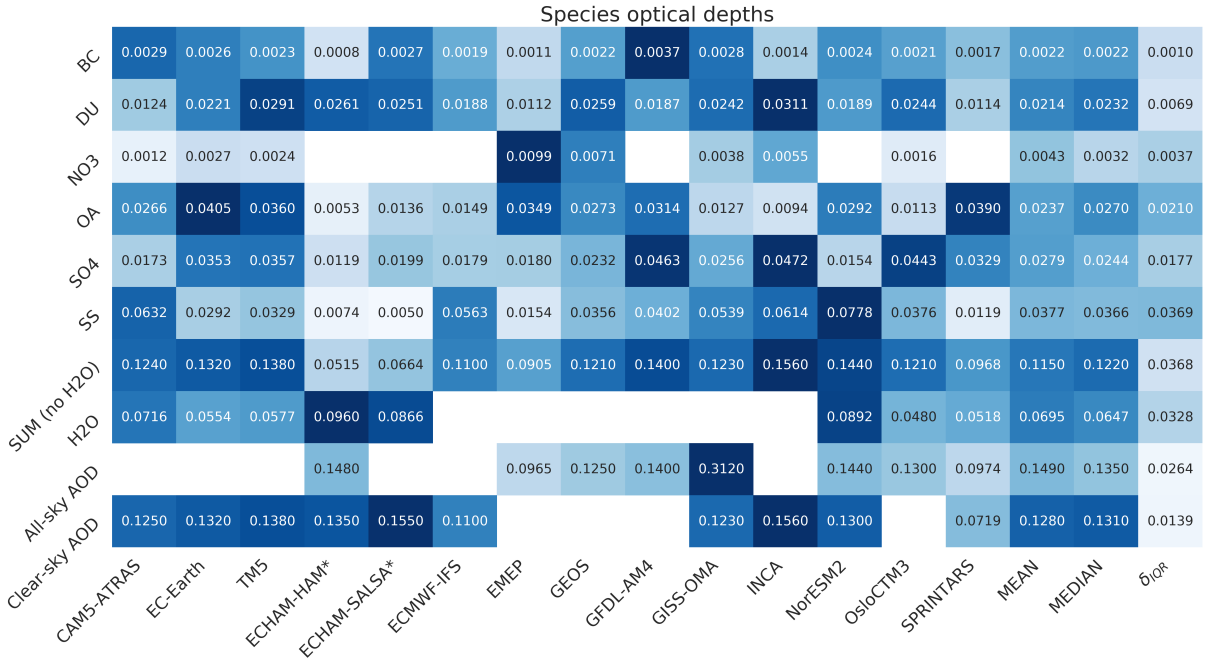

Figure 9. AODs from individual species as well as the sum and, dependent on availability clear-sky and all-sky AOD. Please note that for OsloCTM3 an additional OD of 0.0086 due to biomass burning was reported (combination of OA and BC) which is not included here. A more detailed description of this plot type is provided in Fig. 5.

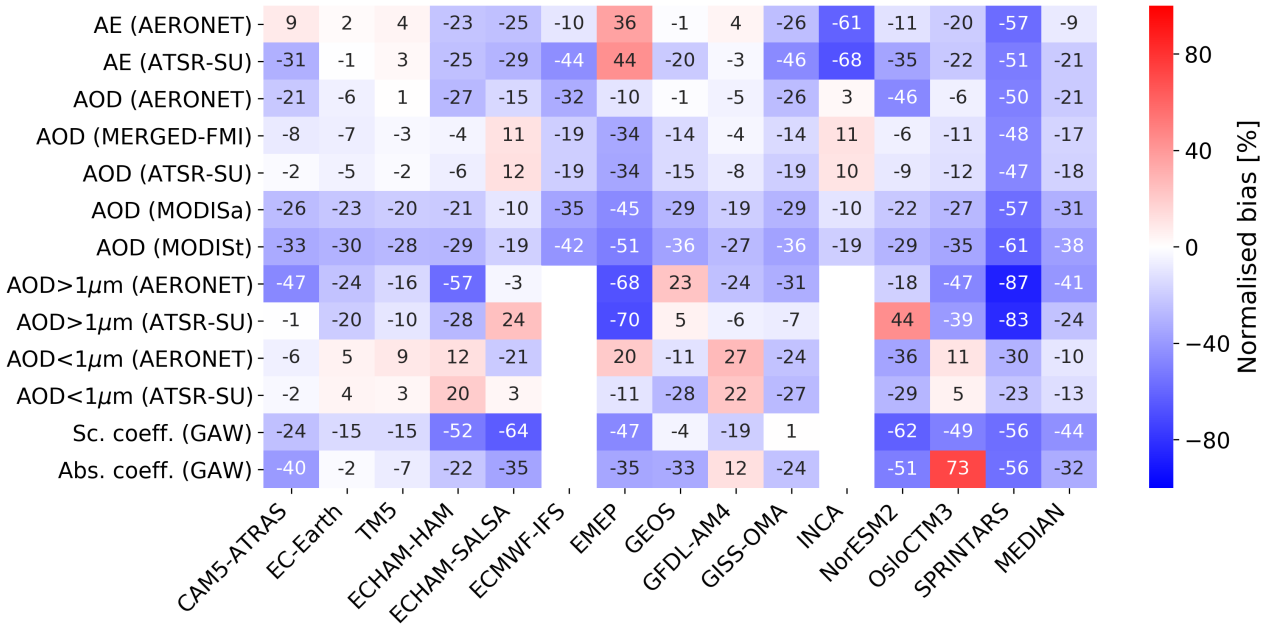

Figure 10. Normalised mean bias (NMB) computed from the monthly collocated data for each model (columns) and observation / variable combination (rows). For the $5^\circ \times 5^\circ$ satellite products, area weights were applied to compute the average bias. Please note that the biases do not represent global averages but the site / sampling locations of each data-set with more weight given to regions with higher spatial density (see e.g.Fig. 1). Please also note potential offsets in the absolute biases arising from uncertainties in the observation retrievals, particularly for the satellite products (cf. Sect. 2.4 and Fig. 2).

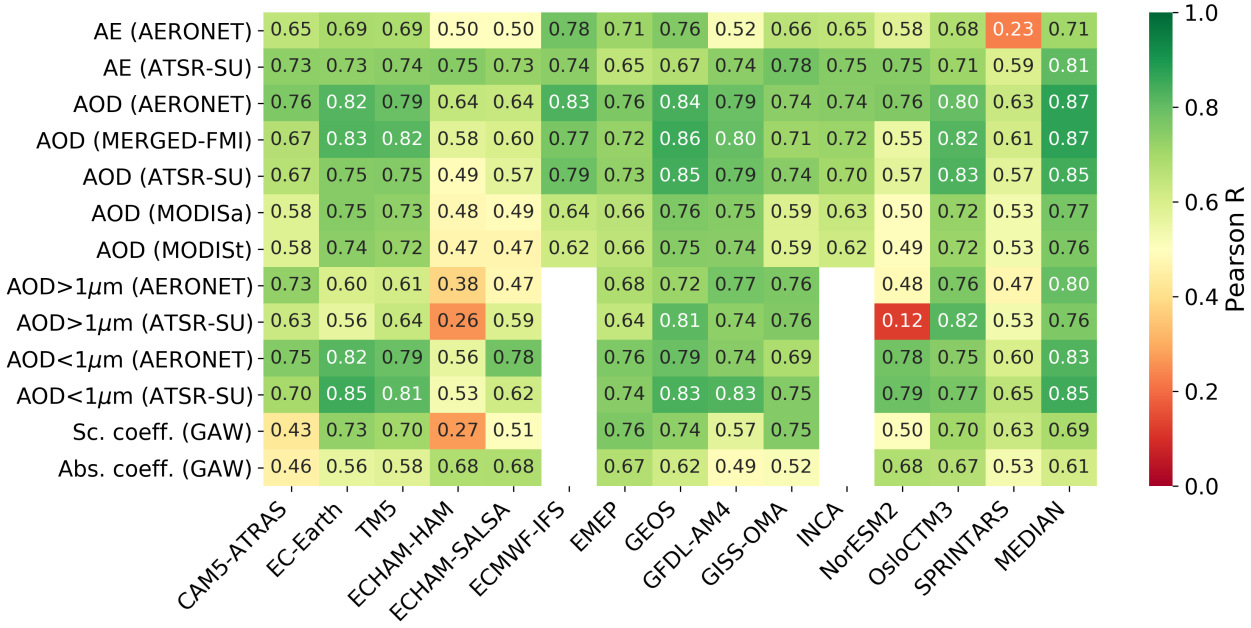

Figure 11. Pearson correlation coefficients (R) computed from the monthly colocated data for each model (columns) and observation / variable combination (rows). For the $5^\circ \times 5^\circ$ satellite products, area weights were applied to the monthly values. Please note further remarks on representativity in Fig. 10.

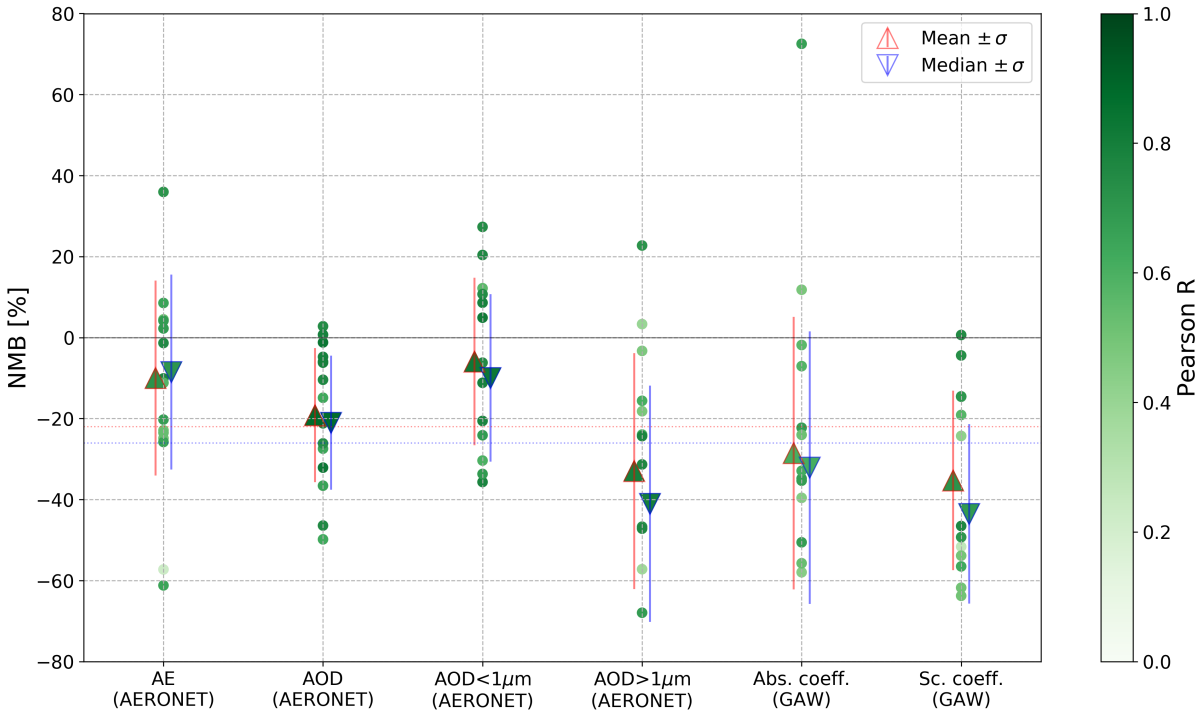

Figure 12. Results from optical properties evaluation DESCRIPTION FOLLOWS BUT ESSENTIALLY A COMBINED VERSION OF THE PREV. 2 HEATMAPS HERE INCLUDING MEAN; MEDIAN AND STD

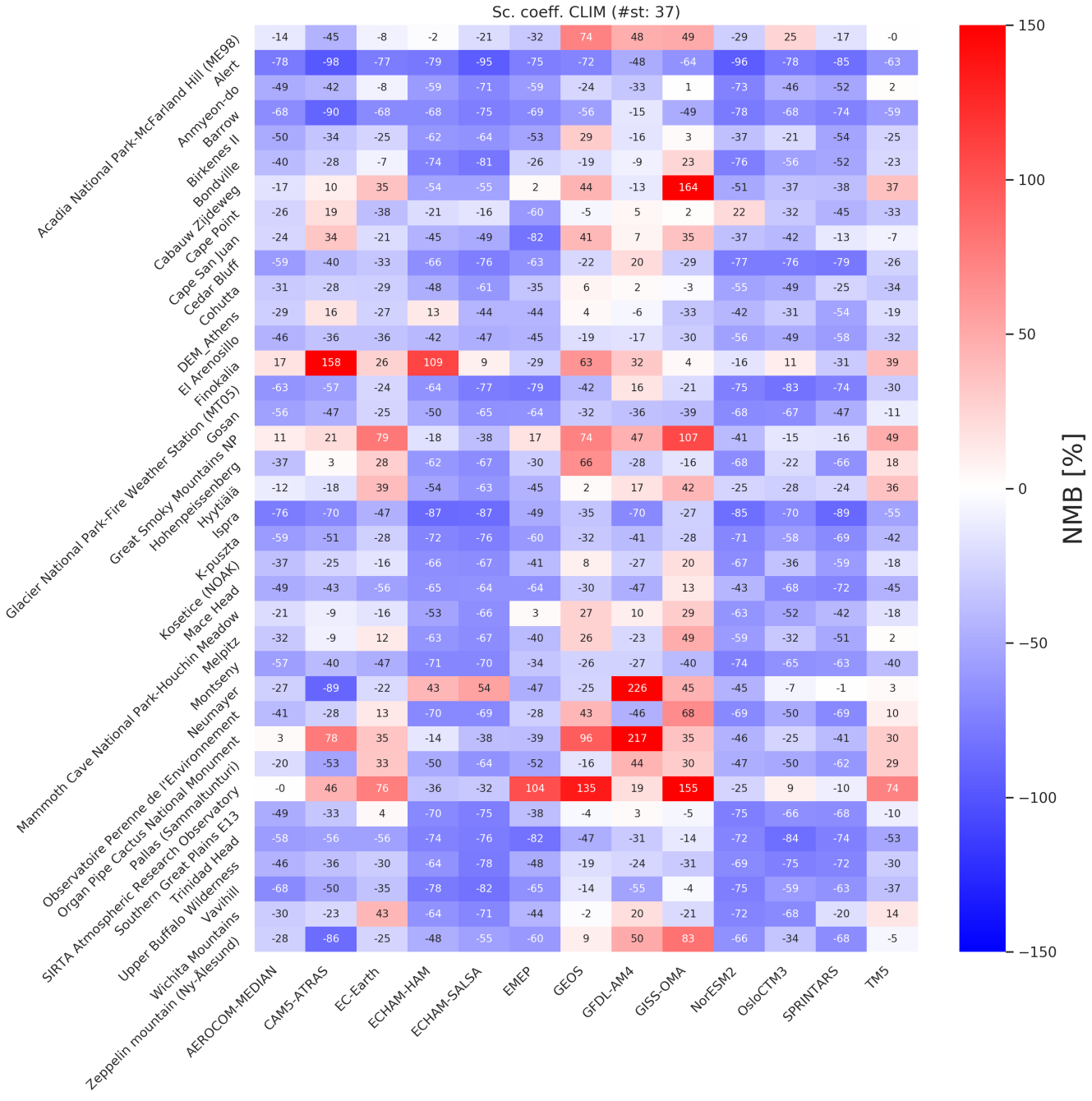

Figure A1. Model biases of surface dry scattering at all in-situ sites that had sufficient temporal coverage to compute monthly climatology.

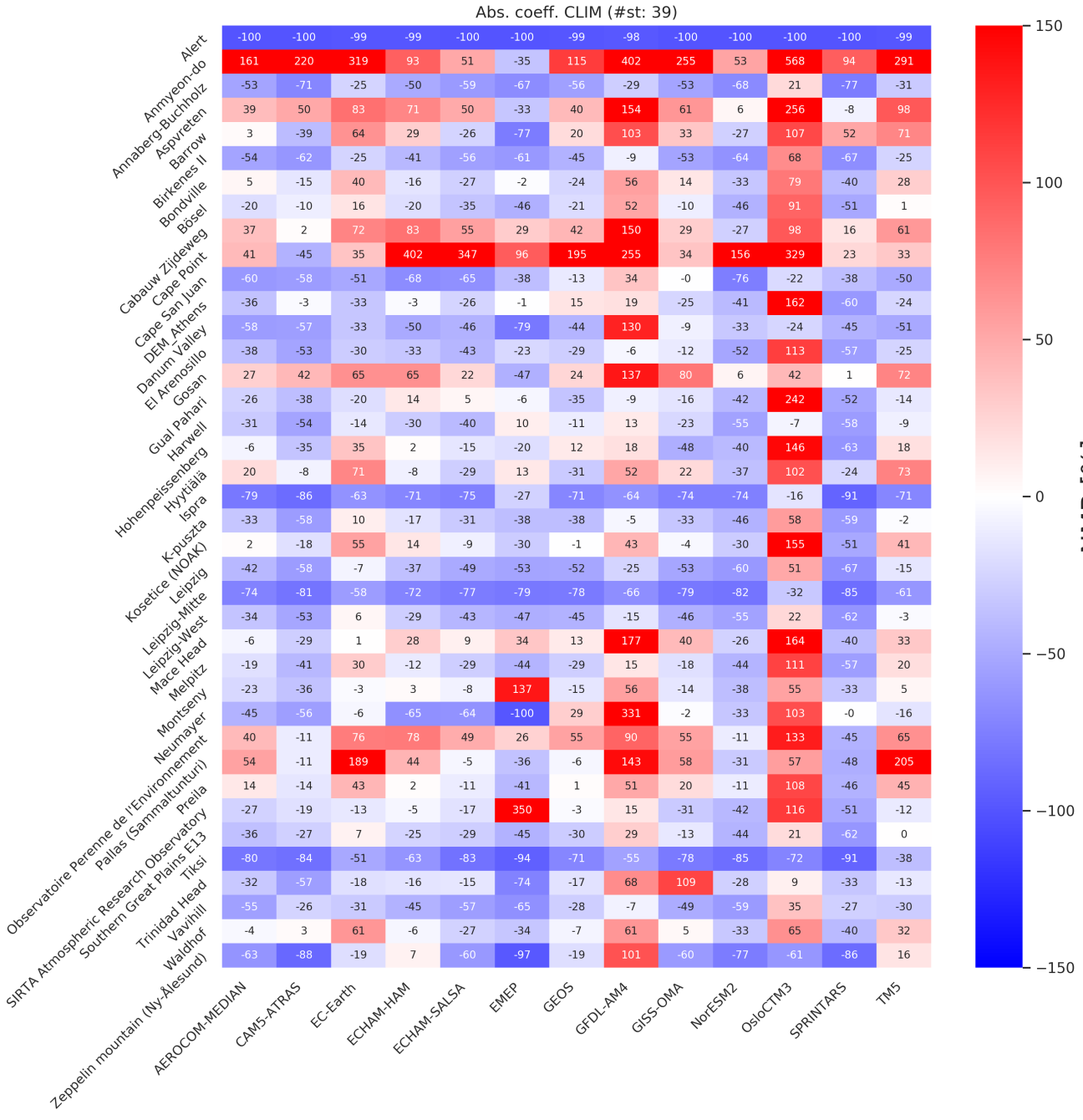

Figure A2. Model biases of surface absorption coefficient at all in-situ sites that had sufficient temporal coverage to compute monthly climatology.

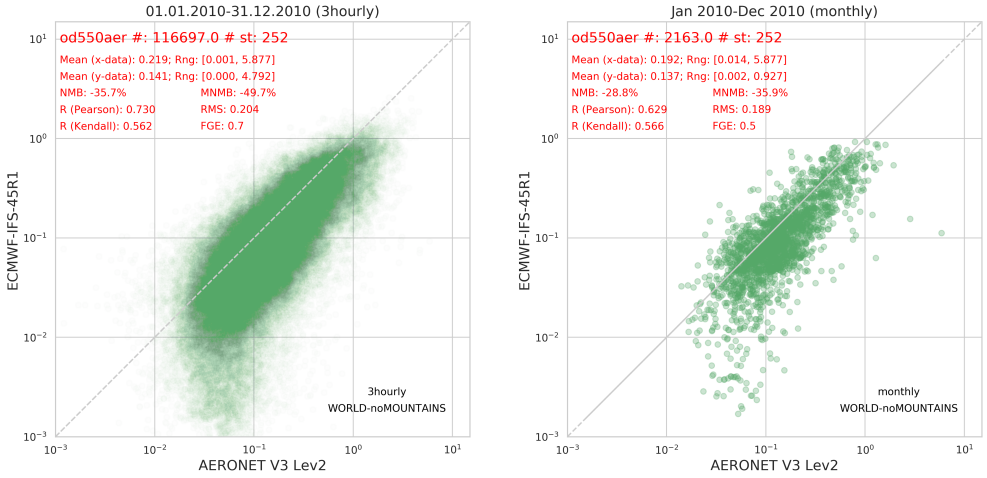

Figure A3. Scatter plot showing results of 3-hourly (left) vs. monthly (right) colocation of AOD from ECMWF-IFS model against AERONET all points data. Also included are statistical results.

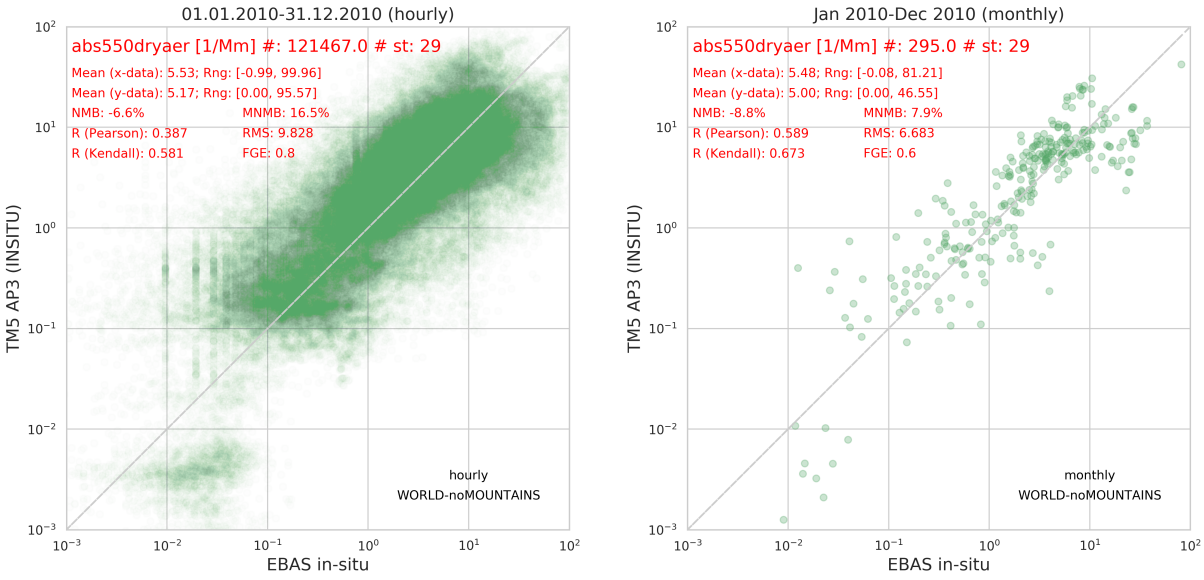

Figure A4. Scatter plot showing results of hourly (left) vs. monthly (right) colocation of in-situ surface absorption from TM5 model (from AeroCom INSITU experiment, i.e. different version than the one used in this study) evaluated at GAW stations. Also included are statistical results.

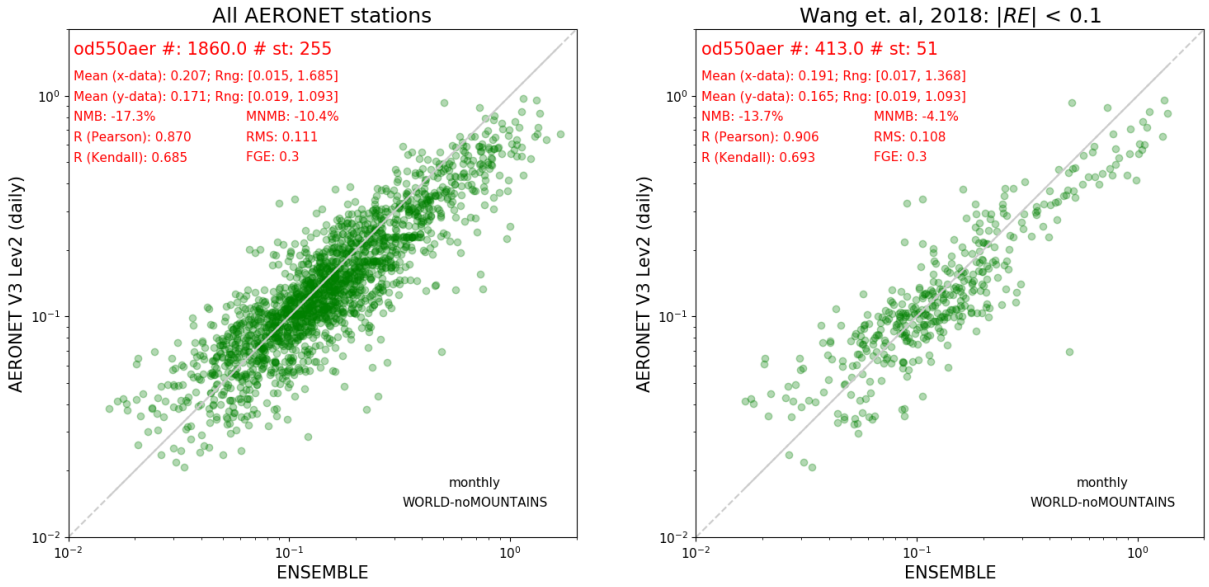

Figure A5. Scatter plot showing colocation results of the ENSEMBLE model AOD evaluated at all available AERONET stations (left) and evaluated only at stations with small spatial representativity errors, selected based on the results from Wang et al. (2018)