Multi-model evaluation of aerosol optical properties in the AeroCom phase III Control experiment, using ground and space based columnar observations from AERONET, MODIS, AATSR and a merged satellite product as well as 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[4,5], Ramiro Checa-Garcia[3], Mian Chin[5], Paul Ginoux[6], Jan J. Griesfeller[1], Andreas Heckel[7], Zak Kipling[9], Alf Kirkevåg[1], Harri Kokkola[10], Paolo Laj[11], Philippe Le Sager[12], 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. Olivié[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
[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), 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 simulated data representative for the year 2010, from 14 global aerosol models participating in the Phase III Control experiment. The model versions are close or equal to those used for CMIP6 and AerChemMIP and inform also on bias in state of the art ESMs. Modelled column optical depths (total, fine and coarse mode AOD) and Ångström Exponents (AE) were compared both with ground based observations from the Aerosol

Robotic Network (AERONET, version 3) as well as space based observations from AATSR-SU instruments. In addition, the modelled AODs were compared with MODIS (Aqua and Terra) data and a satellite AOD data-set (MERGED-FMI) merged from 12 different individual AOD products. Furthermore, for the first time, the modelled near surface scattering (under dry conditions) and absorption coefficients were evaluated against measurements made at low relative humidity at surface in-situ
10   GAW sites.

Statistics are based mainly on normalised mean biases and Pearson correlation coefficients from colocated model and observation data in monthly resolution. Hence, the results are mostly representative for the regions covered by each of the observation networks. Model biases established against satellite data yield insights into remote continental areas and oceans, where ground-based networks lack site coverage. The satellite data themselves are evaluated against AERONET observations,
15   to test our aggregation and re-gridding routines, suggesting relative AOD biases of -5%, -6%, +9% and +18% for AATSR-SU, MERGED-FMI, MODIS-aqua and MODIS-terra, respectively, with high correlations exceeding 0.8. Biases of fine and coarse AOD and AE in AATSR are found to be +2%, -16% and +14.7% respectively, at AERONET sites, with correlations of the order of 0.8.

The AeroCom MEDIAN and most of the participating models underestimate the optical properties investigated, relative
20   to remote sensing observations. AERONET AOD is underestimated by $21\% \pm 17\%$. Against satellite data, the model AOD biases range from -38% (MODIS-terra) to -17% (MERGED-FMI). Correlation coefficients of model AODs with AERONET, MERGED-FMI and AATSR-SU are high (0.8 - 0.9) and slightly lower against 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 AATSR-SU, respectively. The differences in bias against AERONET and AATSR-SU are in agree-
25   ment with the established satellite bias against AERONET. These results indicate that most of the AOD bias is due to missing coarse AOD in the regions covered by these observations.

Underestimates are also found when comparing the models against the surface GAW observations, showing AeroCom ME-DIAN mean bias and inter-model variation of $-44\% \pm 22\%$ and $-32\% \pm 34\%$ for scattering and absorption coefficients, respectively. Dry scattering shows higher underestimation than AOD at ambient relative humidity and is in agreement with recent
30   findings that suggest that models tend to overestimate scattering enhancement due to hygroscopic growth. Broadly consistent negative bias in AOD and scattering suggest a general underestimate in aerosol effects in current global aerosol models.

The large diversity in the surface absorption results suggests differences in the model treatment of light absorption by black carbon (BC), dust (DU) and to a minor degree, organic aerosol (OA). Considerable diversity is found among the models in the simulated near surface absorption coefficients, particularly in regions associated with dust (e.g. Sahara, Tibet), biomass burning
35   (e.g. Amazonia, Central Australia) and biogenic emissions (e.g. Amazonia). Regions associated with high anthropogenic BC emissions such as China and India exhibit comparatively good agreement for all models.

Evaluation of modelled column AEs shows an underestimation of $9\% \pm 24\%$ against AERONET and -21% against AATSR-SU. This suggests that overall, models tend to overestimate particle size, with implications for lifetime and radiative transfer calculations.

40   An investigation of modelled emissions, burdens and lifetimes, mass-extinction-coefficients (MECs) and optical depths (ODs) for each species and model reveals considerable diversity in most of these parameters. These are discussed in detail for each model individually. Inter-model spread of aerosol species lifetime appears to be similar to that of mass extinction coefficients, suggesting that AOD uncertainties are still associated to a broad spectrum of parameterised aerosol processes.

1   Introduction

45   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

50   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

55   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.

60   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 (e.g. Carslaw et al. (2013)). Marine dimethyl-sulfide (DMS) and volcanic emissions are responsible for approximately a third of the global anthropogenic

65   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 de-

70   pendent 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 $(n + i\kappa)$, which depends on aerosol type (chemical composition) and mixing. It is accounted for in models in different ways (e.g. volume mixing, Maxwell

75 Garnett, core-shell, e.g. Klingmüller et al. (2014)). The absorptive properties of dust aerosol, for instance, are dependent 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

80 absorptive aerosols such as dust and black carbon are generally considered to be hydrophobic (which can, to a minor degree be violated in aged aerosol due to mixing, e.g. Cappa et al. (2012)). For instance, between 0% and 40% relative humidity (RH) (a range that is often considered "dry" for the purposes of GAW in-situ measurements), the light scattering can be enhanced (up to about 20% due to hygroscopic growth for some types of aerosol (e.g. Zieger et al. (2013)). This is important when comparing models with in-situ observations, since the latter are often performed at low humidity (RH<40 %) but not at absolutely dry

85 conditions (Gaw Report 227 (2016)). Some models tend to overestimate the scattering enhancement factor at low RH (and high RH) and hence, the amount of light scattering (Burgos et al. (2020, submitted)).

AOD is the vertically integrated light extinction (absorption + scattering) due to an atmospheric column of aerosol and is a function of wavelength. AAOD is the corresponding equivalent for the absorptive power of an aerosol column and tends to be small compared to AOD (ca. 5-10% of AOD). Both AOD (mostly scattering) and AAOD (absorption) are of particular

90 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 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

95 is useful to also investigate other related optical 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 overestimation of the particle size. Like AE, fine and coarse mode AOD can also give insights into the particle size domains, which can help establish

100 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 14 aerosol models in the initial AeroCom experiments. They found that, on a global scale, aerosol optical depth (AOD) from different models compared well to each other and generally well to global annual averages involving trusted ground based references (model biases of the order

105 +10% to -20%). However, they also found considerable diversity in the aerosol speciation among the models, mainly related to differences in transport and water uptake. The diversity for carbonaceous aerosol remained small as similar approaches were

adopted in all models. They concluded that this diversity in component contribution adds (via differences to aerosol size and absorption), to uncertainties for associated aerosol direct radiative effects.

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.

This paper is structured as follows. The next section 2 introduces the observations (OBS), variables (VAR) and models (MOD) used, followed by a discussion of the analysis details for the model evaluation and a short section discussing the representativity of the results. The section ends with a brief discussion of results from a validation study investigating the performance of the satellites used against ground based AERONET data. Section 3 starts with an overview of globally averaged emissions, burdens, lifetimes, mass-extinction-coefficients (MECs) and optical depths (ODs) for each model and aerosol species, followed by a discussion of the results from the AeroCom MEDIAN model and regional model diversity in the optical parameters considered. The section ends with a discussion of the results from comparison of modelled optical properties with the different observation records used. These results are presented in the form of performance charts of retrieved normalised biases and correlation coefficients for each OBS / VAR / MOD combination. This is followed by a dedicated section 4, which discusses the results for each model individually in order to identify strengths and weaknesses of each model in comparison with the observations and the other models. The paper ends with our conclusions from this comprehensive inter-comparison study.

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 are 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 AATSR-SU for instance, shows a latitudinal southwards

140   decreasing gradient in remote ocean regions, indicating coarse(r) particle sizes, likely due 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 AATSR-SU. Finally, as can be seen in the lowermost panel of Fig. 1, in-situ sites from GAW show the highest density in Europe, followed by North America, while other regions are poorly represented. The differences in the spatial coverage for each observation data-set are important to keep in mind when interpreting the results presented in Sect. 3

145   (especially Figs. 10 and 11).

     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, well established remote sensing network based on sun photometer measurements of columnar optical properties. The network comprises several hundred measurement

150   sites around the globe (Fig. 1).

     In this paper, cloud screened and quality assured daily aggregates of AERONET AODs, $\mathrm{AOD} < 1\,\mu\mathrm{m}$, $\mathrm{AOD} > 1\,\mu\mathrm{m}$ 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. 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 short

155   term DRAGON campaigns (Holben et al. (2018)) was excluded in order to avoid putting too much weight on the associated regions (that show high density of measurement sites) with respect to the network averaged statistical parameters used in this study. No further site selection has been performed, since potential spatial representativity issues associated with some AERONET sites were found to be of minor relevance for this study (Sect. 2.4, Fig. A5). Fig. 1 shows the sites used for all variables, where colors indicate the 2010 mean values at each location. Table 1 includes relevant information about the data-set.

160   ### 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

165   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, submitted). 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 $10\,\mu\mathrm{m}$ 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*).

170   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≤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.

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. For scattering coefficients got excluded Granada; Phoenix; National Capitol - Central, Washington D.C; and for absorption coefficients 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. Due to the limited number of stations, and in order to increase temporal sampling coverage of the monthly aggregates used, a 2005-2015 climatology was used to compare with the 2010 model output (unlike for the other observations which solely used 2010 data, see Tab. 1). The climatology for each site was computed requiring at least 30 valid daily values over the 10 year period, for each of the months. Prior to that, daily values were computed from the hourly data applying a minimum 25% coverage constraint (i.e. at least 6 valid hourly values per day).

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 data-sets, see Hubanks (2017).

2.1.4 AATSR SU v4.3 data

The AATSR v4.3 SU data-set provides gridded AOD and associated parameters from the AATSR instrument series, developed by Swansea University (SU) under the ESA Aerosol Climate Change Initiative. The AATSR instrument on ENVISAT covers the period 2002-2012 and in this study, data from 2010 is used. The instrument's conical scan provides two near simultaneous views of the surface, at solar reflective wavelengths from 555 nm to 1.6 μm.

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
directly finds an optimal estimate of both the AOD at 550 nm, and size, parameterised as relative proportion of fine and coarse
mode aerosol. The local composition of fine and coarse mode is adopted from the MACv1 aerosol climatology (Kinne et al.
(2013)). The local coarse composition is defined by fraction of non-spherical dust and large spherical particles typical of sea
salt aerosol, while fine mode is defined by relative fractions of weak and strong absorbing aerosol. A full description of these
component models is given in (de Leeuw et al. (2015)). Further aerosol properties including AE and AAOD (not used in
this study) are determined from the retrieved AOD and composition. Aerosol is retrieved over all snow-free and cloud-free
surfaces. The most recent version SU AATSR 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 1x1 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. 2.5).

2.1.5 Merged satellite AOD data

The MERGED-FMI data-set (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
results of the evaluation of the individual satellite AOD products against AERONET. 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 AeroCom optics questionnaire (supplementary material).
An overview of all models is provided in table 2. More details about each of the models is provided in the corresponding
discussion section 4.

The AOD $> 1\,\mu$m fields were not directly submitted but were computed as difference: AOD - AOD $< 1\,\mu$m . For the comparison with the surface in-situ data, modellers were asked to provide fields of extinction (EC_{dry}) and absorption (AC_{dry}) at dry conditions (RH=0%). The dry scattering (SC_{dry}) was computed via SC_{dry} = EC_{dry} - AC_{dry}. Some of the models that provided these data, submitted dry EC but ambient AC, in which case that combination was used, consistent with the idea that absorbing aerosol tend to be hydrophobic. Note that for some models, not all required fields were available, which is indicated by gaps in the resulting heat-map plots shown in Figs. 3-7 and Figs. 10, 11). Some of the models reported the columnar optical properties based on clear-sky (CS) assumptions, while others assumed all-sky (AS) conditions to compute hygroscopic growth and extinction efficiencies. These choices are indicated in Table A1 and differences can be inferred from Fig. 7.

2.2.1 Computation of AeroCom ensemble mean and median

The monthly AeroCom ensemble mean and median fields were computed in $2° \times 3°$ resolution (Tab. 2). Model fields were all re-gridded to this resolution before the ensemble mean and median was computed. Only those models were considered that had submitted all required optical property variables used in this study (Tab. 1). Those used for constructing the ensemble model are indicated in Table A1. In this paper, the output from the median model is used (denoted MEDIAN below) if not otherwise indicated.

In addition local diversity δ fields were computed for each variable and are defined as follows:

– Diversity ensemble mean: δ_1=std. dev./mean (not so good / meaningful in case of outliers)

– Diversity ensemble median: δ_2= IQR/median (outlier resistant definition)

where IQR denotes the interquartile range (i.e. the difference between the 3rd and 1st quartile).

2.3 Data analysis

The analysis of the data was performed using the pyaerocom software (Appendix C). The ground and space based observations are colocated with the model simulations by matching with the closest model grid-point in the originally provided model resolution. In the case of ground-based observations (AERONET and GAW in-situ), the model grid point closest to each measurement site is used. For the satellite observations, both the model data and the (gridded) satellite product are re-gridded to a resolution of $5° \times 5°$ 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 (defined below).

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 averaged to obtain monthly mean values, prior to the analysis. For the higher resolution observations (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 month. 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
270   coverage constraints were invalidated.

Throughout this paper, the discussion of the results will use two statistical parameters to assess the model performance. The normalised mean bias (NMB) is defined as $\mathrm{NMB} = \frac{\sum_i^N (m_i - o_i)}{\sum_i^N o_i}$ and the Pearson correlation coefficient (R) to assess the model performance. These parameters were computed for each variable / obs. network / model combination and are presented in the form of heat-map performance charts (Figs. 10, 11).

275   The next section presents several sensitivity studies that were performed in order to investigate the spatio-temporal representativity of this analysis strategy, which is based on network-averaged, monthly aggregates, as representativity (or lack thereof) comprises a major source of uncertainty (e.g. Schutgens et al. (2016), Schutgens et al. (2017), Sayer and Knobelspiesse (2019)). The focus here was to assess how such potential representation errors affect the biases and correlation coefficients used to assess the model performance and comparison with other models.

280   2.4   Representativity of the results

As described in the previous Section 2.3, monthly aggregates of the models and observations were colocated in space and time. The resulting point cloud of monthly mean values from all sampling coordinates (sites / aggregated satellite pixels) was then used to compute the biases and correlation coefficients. These are then used to assess the performance of individual models and the ensemble mean, discussed in the following sections (Figs. 10 and 11). The comparison of the (often) temporally
285   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 observations but also other performance measures such as correlation coefficients. We consider this to be the major source of uncertainty for this study. Therefore, several sensitivity studies have been performed in order to investigate how potential spatio-temporal
290   representation errors affect the global monthly statistical parameters used in this study. Temporal representation uncertainties were investigated 1. for in-situ absorption coefficients using hourly TM5 data from the AeroCom INSITU experiment 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 representativity errors were investigated by colocating the ensemble mean AOD field both with observations from all AERONET sites (available in 2010) and with a selection of sites that are considered
295   representative on spatial scales covered by a typical model grid cell. The latter was selected based on Wang et al. (2018) using only sites that show an absolute spatial representation error smaller than 10% and the result of this comparison is shown in Fig. A5). The results of these 3 sensitivity studies are summarised in Tab. 3) and show that the overall differences are of the order of 10% and 0.2 for NMB and correlation, respectively. For the in-situ absorption inter-comparison, the results in monthly resolution show better performance in nearly all statistical parameters, compared to hourly (Fig. A4).

300   From these results, we conclude that differences in these network averaged statistics, arising from spatio-temporal representation errors, are small compared to the diversity in the results found among the different models participating in this study (shown in Figs. 10 and 11).

Based on these findings, and due to the fact that some model data was only available in monthly resolution, it was therefore decided that all model and observation comparisons in this study would be performed in monthly resolution. This was done

305 because we believe that it will make the inter-model results more consistent and hence, more suitable for *inter*-comparison, since they carry similar representation errors (which are introduced by the incompletely sampled observational records).

The small differences in bias and correlation that we find in our sensitivity tests (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) is of the order of $\pm 10\%$. For non-geostationary satellites, the absolute temporal representation

310 errors are likely larger due to the low sampling coverage, combined with cloud contamination in certain regions (e.g. South Pacific). A detailed investigation of these uncertainties is beyond the scope of this work. Nonetheless, a further simple sensitivity study was performed aiming to investigate, how our choice of resolution in the satellite/model comparison (i.e. based on $5° \times 5°$ resolution and monthly averages) would affect the results (NMB and R), as compared to an analysis that is performed in daily resolution and using the highest available horizontal resolution for each model / satellite (see Tab. 1 for an overview

315 of the satellites used). This was done for each model that provided daily (or higher resolution) data and for the variables AOD, $AOD < 1\,\mu m$, $AOD > 1\,\mu m$ and AE. The results are shown in Table A2. In most cases, the higher resolution data results in slightly less negative biases and differences can be up to +10% in NMB (e.g. AE SPRINTARS vs. AATSR-SU). However, in most cases the differences are marginal and are well below 5%.

Finally, we want to stress that the uncertainties established here and discussed above are not to be misinterpreted with

320 corresponding uncertainties over sub-domains or at specific locations and times, which can be significantly larger as shown in the various literature referred to above.

2.5 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

325 of the gridded satellite level 3 products was performed in the same manner as the evaluation of the models (see previous Sect. 2.3). Note that for this analysis the satellite data was used in the original $1° \times 1°$ resolution.

The results of this analysis are shown in Figure 2 and reveal generally high correlation with AERONET measurements ($R > 0.80$). In terms of NMB, AATSR-SU and the MERGED-FMI product show slight underestimations (NMB$\approx -5\%$) while MODIS Aqua and Terra yield slightly overestimated AODs of approximately +9% and +18%, respectively.

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

In case of the AATSR-SU data, the retrieval includes a conservative cloud mask utilising thermal channels in additional to optical, and thereby avoids retrieval near cloud edges. Evaluation under aerosol CCI of six data-sets showed AATSR and

335 SeaWifs exhibited the 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 an overview of annual averaged emissions, burdens, lifetimes, MECs and ODs for each aerosol species and model, followed by a discussion of the results from the ensemble model and regional model diversity. Finally, the results of the optical property evaluation are presented. This is
340 followed by a discussion section for the results from the individual models.

3.1 Modelled emissions, burdens, lifetimes, MECs and ODs

Figures 3, 4, 5, and 6 show the global annual average of emissions, lifetimes, burdens and MECs, for each aerosol species and for each model, respectively. The colors in the performance charts are applied row-wise in order to highlight differences between the models. Also included in each plot are mean, median and diversity (IQR) for each species. Note that the latter are
345 computed directly from the provided table values for each species and model, and not using the ensemble mean and median fields (which were used for the inter-comparison with the measurements shown in Figs. 10 and 11). In addition, Figure 7 shows corresponding averages for the individual optical depths (ODs) of each aerosol species (i.e. BC, DU, OA, NO_3) SO_4 and SS; and their sum) as well as H_2O OD, reported clear-sky (CS) and / or all-sky AOD (where provided).

In the following we briefly discuss the main results from this global perspective of the aerosol life-cycles. The focus in this
350 section will be on the discussion of the ensemble median results and the corresponding model-spread in percent, derived via the provided MEDIAN and δ_{IQR}. Results of individual models are discussed in Sect. 4.

The median emissions (Fig. 3) are highest for sea salt (5090 Tg/yr), followed by dust (1430 Tg/yr), sulphur species (192 Tg/yr), OA (78 Tg/yr, primary) and BC (10 Tg/yr). Models agree well in their BC emissions, which is expected since most models
355 used the CMIP6 BC emission inventories (see AeroCom optics questionnaire (supplementary material)). The highest diversity is found for OA (64%) followed by sea salt (51%) and DMS (42%) and dust (32%). These differences are not surprising, since the emissions of these species are typically computed online (fully or partly) in the models and hence, are highly dependent on the individual parameterisations applied (see AeroCom optics questionnaire (supplementary material)).

360 The lifetimes shown in Fig. 4 were computed using the provided burdens (Fig. 5) and total deposition for each variable (not shown). BC lifetime is around 5.5 days and, in contrast to the BC emissions, shows a rather high diversity of 42% between the models. The modelled NO_3 lifetimes show the largest diversity with values ranging between 2.7 days (GEOS) up to around 10 days (TM5 and EC-Earth). SO_4 and OA have lifetimes of around 5 and 6 days, respectively (and diversities of around 30%). The ensemble median lifetimes of dust and sea salt are around 0.6 and 3.7 days, respectively. However, the individual
365 models tend to show show high variability in these (globally dominant) species with diversities of around 100% and 52% for the residence times of dust and sea salt, respectively.

The modelled atmospheric burdens for each species are shown in Fig. 5). They mostly reflect the corresponding diversities that could be associated with their main sources (emissions) and sinks (deposition). Dust and sea salt burdens, for instance,
370 show considerable variability among the models, with median values of $15 \pm 8\,\text{Tg}$ and $9 \pm 3.4\,\text{Tg}$, respectively. The highest diversity is found for NO_3 (among the 8 models accounting for it) and burdens range between 0.08 Tg (OsloCTM3) and 0.93 Tg (GEOS) The modelled BC burdens also exhibit a considerable spread of around 65% with a median value of 0.16 Tg (Fig. 5). Since the BC emissions are relatively harmonised among the models, the variability in the BC burden is likely due to (ageing / mixing induced) differences in the BC deposition efficiencies, particularly in strong source regions such as China and India
375 (e.g. Riemer et al. (2009), Matsui et al. (2018)).

MECs are shown in Fig. 6 and were calculated by dividing the OD (Fig. 7) for each species by the corresponding burden (Fig. 5). The two ECHAM models (indicated with a star) were not considered to compute mean, median and IQR, since the OD_i values were diagnosed at dry conditions, while all other models reported ambient speciated ODs. This explains the
380 comparatively low MECs for these two models. Diversities of MECs are of the same order of magnitude as for the burdens, indicating different treatment related to the underlying assumptions that determine the extinction efficiencies for each species. The largest diversities in MEC are found for BC, DU and NO_3 (50%).

Ultimately, the model spread in the burdens, combined with the diversity in MECs results in a considerable large diversity
385 in the speciated ODs shown in Fig. 7. For most species, the model spread exceeds 50% (i.e. SS, NO_3, SO_4, OA) while, interestingly, AOD (both CS and AS where provided) indicate much better agreement between the models (11% in CS AOD and 17% in AS). Note that for these estimates of diversity, the two ECHAM models were excluded, since they reported dry speciated ODs. With values of around 0.13, the total AODs (CS and AS) and agree well with the values found in Kinne et al. (2006) (see Fig. 1 therein).

390 **3.2 Modelled annual global distributions of optical properties and their diversity**

Figure 8 shows global maps from the ensemble median (Sect. 2.2.1) for each variable (left) and corresponding diversities (right). Also included are annual average values of the corresponding ground observations at the individual site locations. The legends in each plot provide average values of the ensemble model (global and at obs. sites) and the observations (at obs. sites). The latter provide an indication of potential biases between model and observations and how representative the observation
395 locations are with respect to the whole globe. AODs, for instance, show an annual average value of 0.21 at the AERONET site locations, while the ensemble model shows a corresponding value of 0.16, indicating an absolute AOD bias of about 0.05 (or 24%).

The plotted diversity maps provide insights into the regional model-spread. These may be useful, for instance, to identify regions where models tend to disagree which ultimately may help to explain differences observed when comparing the models
400 with observations (which may be performed in different regions due to lack of spatial coverage). The overall highest diversity, for instance, 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 from biogenic emissions. Reasons for these differences may be a combination of the different treatments of SOA formation (and absorptive properties of OA), or potential differences in the emission altitudes (see AeroCom optics question-
405   naire (supplementary material)). The diversity in in-situ surface absorption is also high in the South Pacific / Antarctica and Australia, which is also affected by regular biomass burning events. Interestingly, models tend to agree in major source regions such as China and India (low diversity in surf. absorption).

The dust dominated Sahara region shows considerable diversity in surface absorption but little diversity in surface scattering.
410   This is an indication of differences in the treatment of dust absorption optical properties. The increased diversity in AE in this region suggests differences in dust size distribution, which may, to a certain degree, be linked with the increased diversity seen in AOD > $1\,\mu$m , which reflects the diversity between the models found for dust emissions, burdens, lifetimes and MECs (Figs. 3 and 5, 4, 6). Explaining these dust related differences in detail is beyond the scope of this work, and needs further investigation.
415   Another notable region is the (comparatively clean) South Pacific and Antarctica which shows a belt of high diversity in surface absorption (but not scattering) and AE and considerable diversity in AOD > $1\,\mu$m (over land). This behaviour may arise from a combination of differences in sources, lifetimes and long range transport of the aerosol (e.g. dust shows > 50% diversity in lifetime, see also Li et al. (2008)). It may also be due to differences in the absorption optical properties of OA (due to organic Ocean emissions), combined with potential differences in sea ice retreat. Most likely, it is a combination of all these
420   effects.

Furthermore, elevated and / or mountainous desert regions such as the Southern Peruvian and Northern Chilean Andes, Tibet show high diversity in AOD > $1\,\mu$m . These regions are however, associated with generally low AODs and thus such differences may not have a significant impact on the global radiation budget.

Unfortunately, most ground based observations (used in the following Sect. 3.3 to evaluate the individual models) provide
425   little or no coverage in these remote regions, where the models show high diversity.

Figure 9 shows annual mean biases retrieved when evaluating the ensemble AODs against the merged satellite product as well as biases established against AERONET AODs and the surface in-situ scattering measurements. The legend provides the network biases and correlation coefficients for each data-set.

South-east Asia appears to be a region where modelled AOD is low (by about -40%) both compared to MERGED-FMI
430   and to AERONET. It can also be clearly seen that the underestimated scattering (by 44% over all GAW stations) is mostly representative for Europe and the US, where the site density is highest. These regions also show underestimated AERONET AODs, but only by about -14% (as can be seen in web visualisation, see Appendix C).

Furthermore, models tend to underestimate scattering and AOD at the few available polar sites. This is also the case for surface absorption (e.g. Barrow, Alert, Tiksi and Neumeyer in Figs. A1, A2). However, models tend to yield rather diverse
435   results at some of these stations, showing over and underestimations (e.g. absorption at Barrow, scattering at Neumeyer).

3.3 Results from optical properties evaluation

Figures 10 and 11 show performance matrices of the normalised mean bias (NMB) and the Pearson correlation coefficient, respectively. These are displayed for each model, variable and observation data-set used. The results from the AeroCom ensemble mean and median (Sect. 2.2.1) are plotted in the rightmost column, suggesting that overall, models moderately underestimate

440   the selected optical properties, both when evaluated against ground and space-based remote sensing and in situ observations. In terms of bias (NMB) the mean model shows slightly better performance compared to the observations with up to +10% improvement (e.g. surface scattering and $AOD > 1\,\mu m$). In terms of correlation (Fig. 11) both median and mean show similar results. Relative biases between the different satellite AOD products mostly resemble the biases found when evaluating the satellites against AERONET (Fig. 2). However, compared to the ground based observations, the satellites can show sig-

445   nificantly different results as can be seen, for instance, in the $AOD > 1\,\mu m$ from NorESM2 vs. AATSR-SU and AERONET, respectively. This is because the satellites generally show higher spatial coverage and are thus, also sensitive to the oceans (Fig. 1). This demonstrates the usefulness of incorporating satellite data, even though these may carry larger uncertainties and representativity errors (Sect. 2.4). For instance, compared to AODs from the two MODIS instruments, models show the largest negative biases, which mostly reflects the results from the satellite evaluation (Sect. 2.5, Fig. 2, i.e. positive biases of +9% and

450   +18% for Aqua and Terra against AERONET).

The differences in NMB for $AOD > 1\,\mu m$ and AE between AERONET and AATSR-SU for the models mostly reflect the respective biases found in the satellite assessment (i.e. ca. -15% for $AOD > 1\,\mu m$ and +15% for AE).

The comparison with the surface in-situ data shows considerably large negative biases (and the lowest correlations) of -44% and -32%, for dry scattering and absorption, respectively at the GAW site locations (Fig. 1). In case of scattering, a

455   small fraction (but likely not more than 20%) of these biases may be due to the fact that models reported at RH=0% and the observations are being performed at RH between 0% - 40%.

Correlation coefficients (Fig. 11) are generally high for the median model (> 0.6) but can be as low as 0.12 for individual model assessments.

4   Discussion of results from individual models

460   In this section, the results shown in Figs. 3-7 and Figs. 10 and 11 are discussed for each model individually. This includes a small introduction into each of the models.

4.1   CAM5-ATRAS

The Community Atmosphere Model version 5 (CAM5) with the Aerosol Two-dimensional bin module for foRmation and Aging Simulation (ATRAS) (Matsui (2017); Matsui and Mahowald (2017)) calculates the following atmospheric aerosol and

465   chemistry processes: emissions, gas-phase chemistry, new particle formation, condensation of sulphate, nitrate, and organic aerosols, coagulation, cloud activation, aqueous-phase chemistry, dry and wet deposition, and aerosol-radiation and aerosol-

cloud interactions. Aerosol particles from 1 nm to 10000 nm in dry diameter are represented with a two-dimensional sectional representation with 12 size bins and 8 BC mixing state bins. Meteorological nudging was used for temperature and wind fields in the free troposphere (<800 hPa) using the MERRA2 data.

470   The sources and burden of OA exceed the model ensemble by 90% and 50%, respectively (Figs. 3, 5). This is likely 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), Matsui et al. (2014b), Matsui (2017)). The NO_3 burden is lower than the model ensemble, consistent with Matsui and Mahowald (2017). The burdens of BC, SO_4, SS, and DU and the lifetimes of all aerosol species in CAM5-ATRAS are similar to those in the model ensemble

475   (Figs. 3-4). BC MEC is larger than the model ensemble by 40% (Fig. 6) likely because the ATRAS model calculates the enhancement of absorption by BC aging processes explicitly by resolving the BC mixing state with 8 bins (pure BC, BC-free, and 6 internally-mixed BC bins). The value of BC MEC in this study ($9.5\,m^2\,g^{-1}$) is close to that in Matsui et al. (2018) ($10\,m^2\,g^{-1}$).

The biases of aerosol optical properties in CAM5-ATRAS are similar to those in the model ensemble (Fig. 10). Model

480   simulations generally agree well with the observations for AOD (MERGED-FMI and AATSR-SU), coarse mode AOD (ATSR-SU), fine mode AOD (AERONET and AATSR-SU), and AE (AERONET) (Fig. 10) with correlation coefficients exceeding 0.6 (Fig. 11). Simulated AOD is underestimated by 21% compared with AERONET AOD and by 33 - 57% compared with MODIS AOD, which is consistent with Matsui and Mahowald (2017). Scattering and absorption coefficients are also underestimated by 24 - 40% compared with the GAW observations.

485   ## 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 et al. (2020, in preparation)). 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

490   general circulation model, which is based on model cycle 36r4 of ECMWF's Integrated Forecasting System (IFS). The sets of meteorological and surface variables that drive TM5 are similar 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.

495   TM5 uses the aerosol scheme M7 (Vignati et al. (2004)), which represents sulphate, 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 Bergmann et al. (in preparation)). Ammonium-nitrate and methane sulphonic 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

500   tropospheric chemistry scheme (Williams et al. (2017)), which also describes the aqueous-phase oxidation of dissolved sulphur

dioxide in clouds.

When calculating the dust source, TM5 does not include particles with dry diameter larger than 16 μm. This may explain why the mean emitted dust mass is smaller than in other models. Differences in 10 m wind speeds generally reduce the dust emissions from the main source regions in EC-Earth compared to TM5 (Fig. 3), leading to proportionally lower dust burdens.
505 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)). The aerosol optical properties in TM5 are calculated based on Mie theory, where the mixing rules of Bruggeman and Maxwell-
510 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 species AOD and MEC values for TM5 and EC-Earth compared to models in which the water contribution is not included, such as ECHAM-HAM and ECHAM-SALSA (Fig. 7).

515 Compared to the observations (Figs. 10 and 11), both TM5 and EC-Earth show similar performance and are generally in good agreement with observations in terms of bias (NMB), outperforming the ensemble values in all comparisons. Particularly, AOD and AOD $< 1\,\mu$m show good performance with biases smaller than 10% and high correlation (R ≤ 0.79), with AOD $< 1\,\mu$m being slightly overestimated and AOD being slightly underestimated. The latter is due to a slightly underestimated AOD $> 1\,\mu$m, both against AERONET and AATSR-SU, which is also reflected in the slightly positive AE bias. Comparison of the diagnosed
520 dry scattering with surface in-situ measurements (at RH<40%) results in biases of -15%. The corresponding comparison of dry absorption, indicates a slightly better performance in TM5 (-2%) than in EC-Earth (-7%), which may be due to the fact that the dust burden in TM5 is about 35% larger than in EC-Earth (and corresponding MECs are similar). The latter would also explain why AOD $> 1\,\mu$m biases are less negative (by about +10%) in TM5 compared to EC-Earth.

4.3 ECHAM-HAM

525 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
530 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 (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

535 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 (Sect. 2.4, 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 NO_3 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. 5 and Tegen et al. (2019)). AOD $< 1\,\mu m$ is overestimated over ocean and dusty regions which is indicated by the stronger overestimation

540 compared to AATSR-SU (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 the subtropical ocean (as can be seen in web visualisation of the results), leading to almost no bias compared to AATSR-SU. Except for regions dominated by dust aerosol AE is biased low. The overestimation 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. 4) indicates a too

545 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 parameterisation applied in CTRL or may be affected by temporal sampling errors (Schutgens et al. (2016), Sayer and Knobelspiesse (2019)).

550 **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.

555 (2018) are, in addition to using anthropogenic emissions required for AEROCOM III simulations, is using sea salt emission parameterisation 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. 5). The BC

560 lifetime is highest among all models (9.6 days, Fig. 4) which explains the high burden. On the other hand, reasons for the high SO_4 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 the observations (Fig. 10), values are biased low compared to AERONET as well as MODIS Aqua and Terra. The latter is likely due to the positive biases found for the MODIS in-

565 struments (Fig. 2) especially also because a positive AOD bias is found against the other two satellites (AATSR and MERGED-FMI). This indicates, that SALSA underestimates AOD over most of the land area while overestimating AOD over the oceans. 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. Compared to AATSR-

SU, the coarse mode AOD of SALSA is significantly overestimated with a normalized bias of +24%, while the AERONET
570   comparison indicates good agreement over land in $AOD > 1\,\mu m$ (NMB=-3%). On the other hand, over regions affected by
dust, coarse mode AOD is overestimated in SALSA. For example, AERONET sites north of Africa exhibit simulated values
higher than those measured. While the apparent high overestimation against AATSR-SU may be, to a certain degree, due to
low biased AATSR-SU data (Fig. 2), these results indicate that possible overestimates in $AOD > 1\,\mu m$ are likely due to ocean
regions. Regions with high dust loads also exhibit overestimation of coarse mode AOD. These is in agreement with the findings
575   of Kokkola et al. (2018) who find large positive biases in $AOD > 1\,\mu m$ over the oceans, in addition to dusty regions. This is
expected 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 (Fig. 3) and it is likely that the simulated size distribution of SALSA is such that SS
and DU particles influence radiation effectively.

4.5   ECMWF-IFS

As part of the Copernicus Atmosphere Monitoring Service (CAMS; https://atmosphere.copernicus.eu/), ECMWF runs a ver-
sion of the 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 is documented at https://www.ecmwf.int/en/forecasts/documentation-and-support, but at a lower
585   resolution of 40 km to offset the cost of the extra schemes. The results presented here are from a "cycling forecast" configu-
ration, that is, 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
590   sea salt (up to $20\mu m$ radius at 80% relative humidity), and bulk tracers for organic matter, black carbon and sulphate 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 sulphate production via a latitude-
and temperature-dependent conversion timescale. There is no separate DMS tracer, and no primary sulphate aerosol emission,
but all sulphate and precursor emissions are treated as SO_2 (resulting in a seemingly large contribution of SO_2 in Fig. 3). The
595   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 3 and 5.
Emissions are three times larger than the ensemble mean, but due to a short lifetime (see Figure 4) the burden is only three
times larger. However, the sea salt contribution to AOD remains similar to other models because the large size distribution
600   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)), and the subject of ongoing development.

The model also has one of the smallest sulphate burdens, which appears to be the result of both relatively low total sulphur emissions and a short lifetime (Fig. 4). 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 exhibit relatively high values, there is a significant low bias against all the AOD data-sets (satellite and AERONET, Fig. 10). This is likely related to the relatively short lifetimes of many species compared to other models, which can be seen in Figure 4. There is also a low bias against both AERONET and AATSR 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.

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 $PM_{2.5}$ and PM_{10}. The results presented in the paper were obtained in a model run at $0.5 \times 0.5°$ 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 μm 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, Appendix C). Furthermore, fine AOD is overestimated by 20% compared with AERONET data and slightly (by only 11%) underestimated compared to AATSR-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. This suggests that either the EMEP model calculates too few coarse particles or the applied MECs are too low (which may be the case for dust, Fig. 6). One of the possible reasons for that is that fine sea salt and dust particles are assumed to have diameters smaller than 2.5 μm, so that the extinction due to sea salt and dust aerosols with diameters between 1-2.5 μm contributes to the (overestimated) AOD $< 1\,\mu$m rather than the (underestimated) AOD $> 1\,\mu$m .

Aerosol specific ODs (Fig. 7) of NO_3 and OA are somewhat larger than the corresponding ensemble median values. This is in agreement with the relatively large loads for those components (Fig. 5) and may be due to the fact that the model calculates both fine ammonium nitrate and coarse NO_3 on sea salt and dust. Also, the OA burdens include both primary sources as well

as 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 ensemble, with the exception of SO_4, which is one of the largest (probably due to too effective hygroscopic growth). The latter, however, is compensated by the comparatively low SO_4 burden (SO_x emissions from ECLIPSE6b used by the EMEP model are smaller than from CMIP6).

640   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 (and 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. Compared to the climatological GAW observations (at RH<40%), the 2010 dry (RH=0%) modelled absorption coefficients are biased low

645   (by 40%) and the correlation is 0.66, which is a fair result given the crude simulation approach. The dry scattering coefficient is underestimated by 47% on average (R = 0.74).

4.7   GEOS

GEOS is a global Earth system model, containing components for atmospheric circulation and composition, ocean circulation and biogeochemistry, land surface processes, and data assimilation (Rienecker et al. (2008)). The version of GEOS Earth Sys-

650   tem 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° x 1.0° latitude and 72 vertical levels from surface up to 0.01hPa (85km) with "replay" mode, denoting simulations driven by the reanalysis meteorological fields from the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA2). This is done to assure that 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),

655   sulphate, nitrate, ammonium, dust, and sea salt (Chin et al. (2002), Colarco et al. (2010), Bian et al. (2019)). The emissions of dust, sea salt, DMS, and biogenic VOCs are model calculated time-varying fields. All other aerosol emissions used in this study follow the instruction of the AeroCom Phase III History experiment. The major updates on this GOCART version include newly implemented nitrate and ammonium (Bian et al. (2017)), anthropogenic and biomass burning SOAs, as well as separate treatment of optical properties for brown carbon (from biomass burning source) and organic carbon (from all other sources).

660   The emissions of aerosols and their precursors used in this GEOS study are similar to those of the ensemble median except OA, which is closer to the ensemble mean (Fig. 3). The simulated atmospheric burdens are within 30% of the ensemble median with the exception of dust and nitrate (Fig. 5). The higher dust burden given by GEOS can be explained by its long lifetime (with 9.7 days the longest among the models, Fig. 4). However, the higher nitrate burden cannot simply be explained with its lifetime (Fig. 5). According to the AeroCom Phase III nitrate experiment, the majority of nitrate formed in the atmosphere is

665   associated with atmospheric dust and sea salt in coarse mode (Bian et al. (2017)). A careful budget analysis for nitrate would need more information in its chemistry formation and particle size distribution, which is beyond the scope of this paper.

In general, aerosol optical fields (i.e. AE, AOD, Sc. coeff., and Abs. coeff.) simulated by GEOS agree well with various ground station and satellite observations (Fig. 10) and show comparatively high correlation (mostly >0.7, Fig. 11). The es-

tablished biases are close to the MEDIAN results, except for the AOD $> 1\,\mu$m , which is shows overestimations both against

670 AERONET and AATSR-SU (Fig. 10). This is consistent with the high dust and sea salt burdens (Fig. 5) as discussed above.

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 (sulphate, carbonaceous aerosols, sea salt and dust) from emissions and a "light" chemistry

675 mechanism designed to support the aerosol model but with prescribed ozone and radicals (Zhao et al., 2018a). 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 expressed as a function of surrounding topography (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

680 particles, respectively. Aerosols are externally mixed except for black carbon, which is internally mixed with sulphate. 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 (Taylor et al., 2000). 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

685 resolution is lower than in GFDL-AM4, which may create a low bias of aerosol emission depending on surface winds.

In Fig. 3, 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 sulphur compounds does not translate in low atmospheric burden (Fig. 5) as their lifetime is among the highest between the models (Fig. 4), either because of weak oxidation or deposition. On the other hand, the other aerosols have a shorter lifetime than other

690 models (Fig. 4) while their burdens are well within 25% the AP3 mean values (Fig. 5). The opposite bias between sulphur compounds and the other aerosols suggest an issue with oxidation of SO_2 rather than wet or dry deposition. In Figure 6 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 on hygroscopic growth at 97% relative humidity 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). An alternative explanation is that dry

695 deposition velocity is too strong. The GFDL-AM4 AODs from individual species (Fig. 7) are within the AP3 model diversity except BC, which has the highest value most likely due to the treatment of its internal mixing with sulphate. 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 AATSR-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. The differences in these

700 biases mostly represent the biases found for the different satellites at AERONET stations (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 (e.g. Sect. 4.11, see also AeroCom optics questionnaire (supplementary

material)). 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).

705 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. (in preparation)), coupled with the One-Moment Aerosol scheme (OMA; Bauer and Tsigaridis (submitted)). In OMA, all aerosols are externally mixed and tracked

710 by their total mass only, except for sea salt and dust where 2 and 5 size-resolved sections are used, respectively. OMA tracks sulphate, nitrate, ammonium, carbonaceous aerosols (black and organic carbon), dust (up to $16\,\mu m$) and sea salt (up to $4\,\mu m$).

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

715 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%.

The results from the evaluation of optical properties shown in Figs. 10 and 11 show a comparatively good agreement with the observations in terms of bias and correlation. The simulated CS AOD shows a bias of -26% against AERONET, which

720 is slightly lower than the ensemble median. In comparison with the satellites, biases of -14% and -19% are found against the MERGED-FMI data-set and AATSR-SU. Similar to the other models, and as explained above, the comparison with MODIS AODs indicates larger negative biases (and slightly decreased correlation) as these satellites show the overall highest AODs (Fig. 2). Considering these relative biases established for the satellites at AERONET sites, AE, AOD $> 1\,\mu$m and AOD $< 1\,\mu$m show similar results when compared with AERONET and AATSR-SU, with biases of the order of -20 to -40% for all three

725 variables.

A possible explanation for these underestimated AODs could be that burdens of SO_4 and sea salt are comparatively low (Fig. 5), which is also reflected by the fact that both AOD $< 1\,\mu$m and AOD $> 1\,\mu$m appear to be underestimated, both against AERONET and AATSR-SU. In case of sea salt, however, the comparatively low burden is likely due to low emissions (Fig. 3) and may, to a certain degree, be compensated by a relatively high SS MEC (+44% compared to median, Fig. 6). A comparatively

730 low burden for nitrate (-33%) is compensated by the largest MEC (ca. +166%). The increased dust emissions, together with an increased lifetime yield a comparatively high burden (Fig. 5) and the fact that the corresponding DU OD is close the the median (Fig. 7) may be due to the low dust MEC (6). In case of BC, the low burden (likely arising from short lifetime) is compensated by the highest BC MEC among the models.

Compared to the in-situ measurements, GISS-OMA shows good agreement (NMB=1%) and comparatively high correlation

735 with surface scattering, and fairly good performance also for the surface absorption coefficients (NMB=-24%), with comparatively low correlation (R=0.52).

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, SO_4, 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 (low burden in Fig. 5). 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 (Fig. 3) 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= $2.5 \mu m$ and σ). Meanwhile, an improved version with 4 modes (Albani et al. (in preparation)) 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 (Checa-Garcia et al. (in preparation)). 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, Fig. 5). All lifetimes are close to the ensemble central values but for sea salt which has a lifetime of 3.1 days.

Our values of MEC are close to the ensemble mean. For those species modelled by a single mode (like dust) we expect less spatial variation of MEC compared to other models with several modes. Regarding optical properties, the AE is highly underestimated both against AERONET and AATSR-SU (ca. -65%). This is due to a smaller dynamical response for wavelength in the visible with respect to other models. The total AOD indicates a slight overestimation compared to the multi-model central values, which is likely due to the overestimations of SO_4 and dust contributions to optical depth, which may partially be compensated by the expected lower values of OA optical depths (Fig. 7).

4.11 NorESM2

The atmosphere module in NorESM2 (NorESM2-MM, see Seland et al. (in preparation)), CAM6-Nor (Olivie et al. (in preparation)), 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

770   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) NorESM2 simulations with fSST is consistent with

775   the considerably lower U10 (especially over land) and dust emissions in nudged vs. free meteorology. While NorESM2 sea salt emissions are among the lowest for AP3, the burden is mid-range, and with the highest MEC (4.1 m^2/g), this model has the highest sea salt AOD values, which is reflected in the positive coarse mode bias against AATSR satellite observations (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 m^2/g), but a change in assumed

780   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

785   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, compared to the other

790   models (Figs. 3-5)

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,

795   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

800   between the models (Fig. 6). 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 sulphate 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

805 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 (Fig. 7). 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 (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

810 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 (ca. -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.

815 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

820 are smaller than those without nudging since the emission amounts strongly depend on the wind speed near the surface (see also 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 (Fig. 3) with the horizontal resolution of T85

825 (approx. $1.4° \times 1.4°$). Both the lifetime of sea salt and dust are short compared to the other models (Fig. 4), 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. 5 which is consistent with the high underestimation of the AOD $> 1\,\mu m$ (Fig. 10). On the other hand, the calculated AE by SPRINTARS is underestimated, which would rather suggest an overestimation of particle size. However, for this model, this could be attributed to an inappropriate computation of standard deviations of log-normal

830 size distributions of SO_4 and OA, when calculating optical properties (based on the Mie theory). An internal investigation has confirmed that the diagnosed AE calculated from prognostic mass mixing ratio of each aerosol component is around 1.5 over the industrialized and biomass burning regions, with the appropriate standard deviations of the size distributions. This revision (which is not shown in this article) results in a better AOD performance, with an global annual mean of 0.112, as opposed to 0.072 found in this study (Fig. 7).

835 Overall, the underestimated dust and sea salt sources result in an underestimation and low correlation in all optical properties that have been investigated in this study (Figs. 10 and 11). Consistently, the largest negative biases are found in the evaluation of the coarse AOD, both for AERONET and AATSR (Fig. 4).

5 Conclusions

In this study a comprehensive inter-comparison of 14 models from the Phase III AeroCom Control experiment has been
840   performed. The focus was on the assessment of the modelled column integrated aerosol optical properties AOD, $AOD < 1\,\mu m$
, $AOD > 1\,\mu m$, and AE, as well as, for the first time, surface (dry) scattering and absorption coefficients. The columnar data
was compared to ground based observations from AERONET as well as to several space based observations. In addition to
the model evaluation, the performance of the satellite products - in the resolution as aggregated and used for this study - was
investigated by comparison with AERONET observations. This was done in order to establish potential relative biases when
845   evaluating the models using satellite observations (Fig. 2). From this analysis, AATSR-SU and MERGED-FMI showed slight
underestimations of AOD (ca. -5%) and MODIS Aqua and Terra showed overestimations of about 10% and 20%, respectively
at AERONET sites. AE from AATSR-SU was found to be biased high by about 15% against AERONET, while $AOD > 1\,\mu m$
was found to be underestimated by about 15%. $AOD < 1\,\mu m$ from AATSR-SU showed good agreement with AERONET. All
satellite products showed high correlation against AERONET.

850   The results of the model evaluation against all ground based observations are summarised in Fig. 12. It shows results of
the AeroCom MEDIAN and MEAN (triangles) and corresponding uncertainties estimated from the results of the individual
models (plotted as circles). The AE is underestimated by about -9% and shows considerable spread between the models. This
suggests that, on average, the simulated particle size is overestimated. This may imply a too short aerosol lifetime or too large
fraction of coarse particles present in the models. It may also impact the atmospheric radiation budget due to shifts in the
855   wavelength dependency of aerosol scattering. While the underestimated AE suggests too coarse particles in the models, the
analysis of the $AOD > 1\,\mu m$ reveals an underestimation by -40%, with a considerable inter-model spread. The average AOD
bias amounts to -20% and shows highest consistency (lowest spread) between the models. The AOD bias primarily appears
to arise from the low $AOD > 1\,\mu m$, since $AOD < 1\,\mu m$ shows a smaller bias (-10%, i.e. smallest underestimation) against
the respective observations, with a similar spread as for AOD. Compared to Kinne et al. (2006), our AOD bias indicates a
860   slightly larger underestimation in the more recent model versions (AP3 relative to earlier AeroCom phases). This may partly
be attributed to the fact that in this study, 10 out of 14 models reported clear-sky (CS) AODs (see Tab. A1 and AeroCom
optics questionnaire (supplementary material)), while the AOD diagnostics used by Kinne et al. (2006) were likely based on
more models reporting AOD under all-sky (AS) assumptions. This hypothesis is underpinned by a +20% increase in NMB
in NorESM2, when using AS instead of CS (results available here, leftmost simulation: https://aerocom-evaluation.met.no/
865   overall.php?project=aerocom&exp=hygro#).
The recent findings from the trends analysis by Mortier et al. (2020, submitted) indicate, that observed model biases remain
mostly constant over time and are not particular for the investigated year 2010.

Surface dry scattering and absorption coefficients are underestimated by about -40% and -30%, respectively (at GAW sites,
Fig. 1). Both variables show considerable spread between the models, similar to the results for $AOD > 1\,\mu m$. In case of the
870   dry scattering comparison, the large negative bias may be attributed to differences in the scattering enhancement due to hygro-
scopic growth. Models used here reported scattering at RH=0%, while in the observations scattering is measured at RH that can

be somewhere between 0% and 40%. Thus, on average, the measurements should show larger scattering due to hygroscopic growth. However, the models overestimate the scattering enhancement factor due to hygroscopic growth, as found by Burgos et al. (2020, submitted) (Fig. 5 therein). From a qualitative perspective, a potential overestimation of the scattering enhancement
875 factor in the models agrees well with our finding that models underestimate (ambient) AOD less than dry scattering (by about a factor of 2).

Altogether it is noteworthy that most models underestimate consistently several of the different extensive aerosol optical properties (AOD, fine and coarse mode AOD, scattering and absorption coefficients), both derived from in-situ and remote
880 sensing sensors. This suggests that aerosol loads might be underestimated in the models for the year 2010. Such underestimates are partly compensated by different aerosol optical models and, for instance, higher mass exctinction coefficients.

In future studies the biases found in this study should be investigated, for instance, by incorporating different aspects into the analysis, such as model resolution (particularly vertical), profile extinction data (to investigate "where" the mass is located) and
885 column water content (to assess hygroscopic growth). In addition, a comparison with surface mass concentration measurements could provide valuable insights related to the question, whether the models are missing mass or whether assumptions about optical properties are causing the underestimated scattering coefficients and optical depth. Such an analysis would certainly benefit also from a better global coverage of surface measurement sites, since the analysis performed in this study is mostly representative for Europe and the US, where the density of GAW sites is highest (Fig. 1).

890 *Code and data availability.* Most of the data analysis was performed using the open source software pyaerocom (version 0.9.0, release upcoming). All additional analysis scripts are stored in a private GitHub repository and can be provided upon request. All data used in this study is stored on servers of the Norwegian Meteorological Institute and can be provided upon request.

Appendix A: GAW site evaluation biases

Figs. A1 and A2 show the established biases (NMB) of near surface scattering and absorption coefficients, at all GAW sites
895 (Sect. 2.1.2) used in this paper.

Appendix B: Sensitivity studies related to spatiotemporal representativity results

As introduced in Sect. 2.4 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
900 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.

905 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 (Figs. 10 and 11).

Appendix C: Pyaerocom and web visualisation

Most of the analysis in this study was performed with Pyaerocom (Github: https://github.com/metno/pyaerocom, Website: https://pyaerocom.met.no/). It is an open source python software project and is being developed and maintained at the Norwe-
910 gian Meteorological Institute, focussing on model evaluation for aerosol models and the AeroCom initative.

A dedicated website is associated to this study and 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. All results from the optical properties evaluation discussed in
915 this paper are available online at: https://aerocom-evaluation.met.no/overall.php?project=aerocom&exp=PIII-optics2019-P# (last access: 20.12.2019)

Author contributions. JG, MS and AM designed the study. JG did most of the analysis and wrote the manuscript. AM contributed the interactive website associated to this study. EA, AH, PL, PS, PN, CM and LS helped analysing and interpreting the observational data used. AB and JG helped organising the AeroCom database and AeroCom infrastructure needed. YV, SB, ZK, AK, HK, ML, HM, GM, DN, TN,
920 DO, TT, KT and ST provided model data and model interpretations. All co-authors gave feedback to the manuscript.

Competing interests. No competing interests.

Acknowledgements. Data providers from all observation networks and satellite products and all people involved in the production of these products are highly acknowledged for sharing and submitting their data and for support. JG wishes to thank O. Landgren and H. Svennevik for providing data and code used in this study. TT was supported by the supercomputer system of the National Institute for Environmental Stud-
925 ies, 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 GISS model. Resources supporting this work were provided by the NASA High-End Computing (HEC)

930   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. PN and AH are supported under
      the ESA Aerosol Climate Change Initiative (Aerosol CCI). HM acknowledges funding from the Ministry of Education, Culture, Sports,
935   Science, and Technology and the Japan Society for the Promotion of Science (MEXT/JSPS) KAKENHI Grant Number JP17H04709. AK.,
      DO., and MS. acknowledge support from the Research Council of Norway funded projects EVA (229771), INES (270061), and KeyClim
      (295046), and the Horizon 2020 project CRESCENDO (Coordinated Research in Earth Systems and Climate: Experiments, Knowledge, Dis-
      semination and Outreach, no. 641816). High performance computing and storage resources were provided by the Norwegian infrastructure
      for computational science (through projects NN2345K, NN9560K, NS2345K, and NS9560K)

940 **References**

Albani et al.: TITLE, in preparation.

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
945   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.

Bauer and Tsigaridis: TITLE, JAMES, will be submitted to JAMES before 12/31/19, submitted.

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,
950   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.

Bergmann et al.: TITLE, in preparation.

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:
955   //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.

960 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.

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,
965   A., Skeie, R. B., Steenrod, S. D., Sudo, K., Tsigaridis, K., Tsimpidi, A. P., and Tsyro, S. G.: Investigation of global particulate nitrate from the AeroCom phase~III experiment, Atmospheric Chemistry and Physics, 17, 12 911–12 940, https://doi.org/10.5194/acp-17-12911-2017, https://www.atmos-chem-phys.net/17/12911/2017/, 2017.

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,
970   Atmospheric Chemistry and Physics, 19, 10 773–10 785, https://doi.org/10.5194/acp-19-10773-2019, https://www.atmos-chem-phys.net/19/10773/2019/, 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,
975   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.

980 Burgos et al.: A global model-measurement evaluation of particle light scattering coefficients at elevated relative humidity, Atmospheric Chemistry and Physics, 2020, submitted.

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, 985 337, 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.

Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., Mann, G. W., Spracklen, D. V., Woodhouse, M. T., Regayre, L. A., and Pierce, J. R.: Large contribution of natural aerosols to uncertainty in indirect forcing, Nature, 503, 67–71, 990 https://doi.org/10.1038/nature12674, https://doi.org/10.1038/nature12674, 2013.

Checa-Garcia et al.: TITLE, in preparation.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A., and Nakajima, T.: Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun Photometer Measurements, Journal of the Atmospheric Sciences, 59, 461–483, https://doi.org/10.1175/1520-0469(2002)059<0461:TAOTFT>2.0.CO;2, https://doi. 995 org/10.1175/1520-0469(2002)059%3C0461:TAOTFT%3E2.0.COhttp://0.0.0.2, 2002.

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.

Colarco, P., da Silva, A., Chin, M., and Diehl, T.: Online simulations of global aerosol distributions in the NASA GEOS-4 1000 model and comparisons to satellite and ground-based aerosol optical depth, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/10.1029/2009JD012820, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JD012820, 2010.

de Leeuw, G., Holzer-Popp, T., Bevan, S., Davies, W. H., Descloitres, J., Grainger, R. G., Griesfeller, J., Heckel, A., Kinne, S., Klüser, L., Kolmonen, P., Litvinov, P., Martynenko, D., North, P., Ovigneur, B., Pascal, N., Poulsen, C., Ramon, D., Schulz, M., Siddans, R., Sogacheva, L., Tanré, D., Thomas, G. E., Virtanen, T. H., von Hoyningen Huene, W., Vountas, M., and Pinnock, S.: Evaluation 1005 of seven European aerosol optical depth retrieval algorithms for climate analysis, Remote Sensing of Environment, 162, 295–315, https://doi.org/https://doi.org/10.1016/j.rse.2013.04.023, http://www.sciencedirect.com/science/article/pii/S0034425713003507, 2015.

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, 1010 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.

ERA5: Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), https://cds.climate.copernicus.eu/cdsapp#!/home, accessed: 2019-12-27.

1015   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
1020   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.

        Gaw Report 227: WMO/GAW Aerosol Measurement Procedures, Guidelines and Recommendations, https://library.wmo.int/doc_num.php?
            explnum_id=3073, accessed: 2019-12-27, 2016.

        Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Camp-
1025   bell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: Advancements in the Aerosol Robotic Network (AERONET) Ver-
            sion~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
1030   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.

        Grythe, H., Ström, J., Krejčí, R., Quinn, P. K., and Stohl, A.: A review of sea-spray aerosol source functions using a large global set of sea
            salt aerosol concentration measurements, 2014.

        Guelle, W., Schulz, M., Balkanski, Y., and Dentener, F.: Influence of the source formulation on modeling the atmospheric global distribution
1035   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.

1040   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,
1045   F., Jankowiak, I., and Smirnov, A.: AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, Re-
            mote 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.,
1050   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.

1055 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.

Kelley et al.: TITLE, JAMES, will be submitted to JAMES before 12/31/19, in preparation.

Kim, D., Chin, M., Yu, H., Diehl, T., Tan, Q., Kahn, R. A., Tsigaridis, K., Bauer, S. E., Takemura, T., Pozzoli, L., Bellouin, N.,
1060 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,
1065 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.:
1070 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.
1075 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.

Koch, D., Schmidt, G. A., and Field, C. V.: Sulfur, sea salt, and radionuclide aerosols in GISS ModelE, Journal of Geophysical Research: At-
1080 mospheres, 111, https://doi.org/10.1029/2004JD005550, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JD005550, 2006.

Koch, D., Bond, T. C., Streets, D., Unger, N., and van der Werf, G. R.: Global impacts of aerosols from particular source regions and sectors, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2005JD007024, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD007024, 2007.

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
1085 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
1090 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 et al.: A global analysis of climate-relevant aerosol properties retrieved from the network of GAW near-surface observatories, Atmos. Meas. Tech., 2020, submitted.

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.

Li, F., Ginoux, P., and Ramaswamy, V.: Distribution, transport, and deposition of mineral dust in the Southern Ocean and Antarctica: Contribution of major sources, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD009190, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009190, 2008.

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.

Matsui, H. and Mahowald, N.: Development of a global aerosol model using a two-dimensional sectional method: 2. Evaluation and sensitivity simulations, Journal of Advances in Modeling Earth Systems, 9, 1887–1920, https://doi.org/10.1002/2017MS000937, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS000937, 2017.

Matsui, H., Koike, M., Kondo, Y., Fast, J. D., and Takigawa, M.: Development of an aerosol microphysical module: Aerosol Two-dimensional bin module for foRmation and Aging Simulation (ATRAS), Atmospheric Chemistry and Physics, 14, 10 315–10 331, https://doi.org/10.5194/acp-14-10315-2014, https://www.atmos-chem-phys.net/14/10315/2014/, 2014a.

Matsui, H., Koike, M., Kondo, Y., Takami, A., Fast, J. D., Kanaya, Y., and Takigawa, M.: Volatility basis-set approach simulation of organic aerosol formation in East Asia: implications for anthropogenic–biogenic interaction and controllable amounts, Atmospheric Chemistry and Physics, 14, 9513–9535, https://doi.org/10.5194/acp-14-9513-2014, https://www.atmos-chem-phys.net/14/9513/2014/, 2014b.

Matsui, H., Hamilton, D. S., and Mahowald, N. M.: Black carbon radiative effects highly sensitive to emitted particle size when resolving mixing-state diversity, Nature Communications, 9, 3446, https://doi.org/10.1038/s41467-018-05635-1, https://doi.org/10.1038/s41467-018-05635-1, 2018.

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.

1130 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.

Mortier et al.: Evaluation of climate model aerosol trends with ground-based observations over the last two decades - an AeroCom and
1135 CMIP6 analysis, 2020, submitted.

MYHRE, G., BELLOUIN, N., BERGLEN, T. F., BERNTSEN, T. K., BOUCHER, O., GRINI, A. L. F., ISAKSEN, I. S. A., JOHNSRUD, M., MISHCHENKO, M. I., STORDAL, F., and TANRÉ, D.: Comparison of the radiative properties and direct radiative effect of aerosols from a global aerosol model and remote sensing data over ocean, Tellus B, 59, 115–129, https://doi.org/10.1111/j.1600-0889.2006.00226.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0889.2006.00226.x, 2007.

1140 Myhre, G., Berglen, T. F., Johnsrud, M., Hoyle, C. R., Berntsen, T. K., Christopher, S. A., Fahey, D. W., Isaksen, I. S. A., Jones, T. A., Kahn, R. A., Loeb, N., Quinn, P., Remer, L., Schwarz, J. P., and Yttri, K. E.: Modelled radiative forcing of the direct aerosol effect with multi-observation evaluation, Atmospheric Chemistry and Physics, 9, 1365–1392, https://doi.org/10.5194/acp-9-1365-2009, https://www.atmos-chem-phys.net/9/1365/2009/, 2009.

North, P. R. J.: Estimation of aerosol opacity and land surface bidirectional reflectance from ATSR-2 dual-angle imagery: Operational
1145 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.

Olivie et al.: TITLE, in preparation.

1150 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,
1155 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
1160 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.

Riemer, N., West, M., Zaveri, R. A., and Easter, R. C.: Simulating the evolution of soot mixing state with a particle-resolved aerosol model, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2008JD011073, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD011073, 2009.

1165 Rienecker, M. M., Suarez, M., Todling, R., Bacmeister, J., Takacs, L., Liu, H., Gu, W., Sienkiewicz, M., Koster, R., Gelaro, R., et al.: The GEOS-5 Data Assimilation System: Documentation of Versions 5.0. 1, 5.1. 0, and 5.2. 0, 2008.

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.

1170 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.

1175 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,
1180 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.

1195 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.

Seland et al.: The Norwegian Earth System Model, NorESM2 - Evaluation of theCMIP6 DECK and historical simulations, in preparation.

Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M.,
1200 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.

Sporre et al.: TITLE, in preparation.

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.

Taylor, K. E., Williamson, D., and Zwiers, F.: The sea surface temperature and sea-ice concentration boundary conditions for AMIP II simulations, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore˜. . . , 2000.

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., Koch, D., and Menon, S.: Uncertainties and importance of sea spray composition on aerosol direct and indirect effects, Journal
of Geophysical Research: Atmospheres, 118, 220–235, https://doi.org/10.1029/2012JD018165, https://agupubs.onlinelibrary.wiley.com/
doi/abs/10.1029/2012JD018165, 2013.

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 et al.: EC-Earth3-AerChem, a global climate model with interactive aerosols and atmospheric chemistry for use in CMIP6, Geosci.
Model Dev., 2020, in preparation.

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.: Esti-
mation 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 Rep-
resentativeness 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., Pan-
dolfi, 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 Environ-

ment, 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, 2018a.

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: 2. Model Description, Sensitivity Studies, and Tuning Strategies, Journal of Advances in Modeling Earth Systems, 10, 735–769, https://doi.org/10.1002/2017MS001209, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001209, 2018b.

Zieger, P., Fierz-Schmidhauser, R., Weingartner, E., and Baltensperger, U.: Effects of relative humidity on aerosol light scattering: results from different European sites, Atmospheric Chemistry and Physics, 13, 10609–10631, https://doi.org/10.5194/acp-13-10609-2013, https://www.atmos-chem-phys.net/13/10609/2013/, 2013.

Table 1. Observations used in this study, including relevant meta data information (Ver: Data Version; Lev: Data level; Freq: Original frequency of data used to derive monthly means; Res: Resolution of gridded data product; Clim: Use of a multi-annual climatology or not; #st: Number of stations / coordinates, with observations used; Date: Retrieval date from respective data base).

Data ID and Source	Variable	Ver.	Lev.	Freq.	Res.	Clim.	#st.	Date
GAW EBAS	Abs. coeff.		3	hourly		Y	39	2019/12/18
GAW EBAS	Sc. coeff.		3	hourly		Y	37	2019/12/18
AERONET Sun	AE	3	2	daily		N	250	2019/09/20
AERONET Sun	AOD	3	2	daily		N	240	2019/09/20
AERONET SDA	AOD<1um	3	2	daily		N	226	2019/09/20
AERONET SDA	AOD>1um	3	2	daily		N	226	2019/09/20
MODISt Terra	AOD	6.1	3	daily	1x1	N	2235	2019/11/22
MODISa Aqua	AOD	6.1	3	daily	1x1	N	2241	2019/11/25
AATSR-SU Swansea	AOD	4.3	3	daily	1x1	N	2055	2016/09/30
AATSR-SU Swansea	AE	4.3	3	daily	1x1	N	2055	2016/09/30
AATSR-SU Swansea	AOD<1um	4.3	3	daily	1x1	N	2055	2016/09/30
AATSR-SU Swansea	AOD>1um	4.3	3	daily	1x1	N	2055	2016/09/30
MERGED-FMI	AOD			monthly	1x1	N	2080	2019/10/21

Table 2. Models used in this study, along with horizontal grid resolution, number of vertical levels (Levs) and key references.

Name	Lat. / Lon.	Levs.	References
CAM5-ATRAS	1.9 x 2.5	30	Matsui (2017), Matsui and Mahowald (2017)
EC-Earth3-AerChem	2.0 x 3.0	34	van Noije et al. (2014), van Noije et al. (2020, in preparation)
TM5	2.0 x 3.0	34	van Noije et al. (2014), Bergmann et al. (in preparation)
ECHAM-HAM	1.9 x 1.9	47	Tegen et al. (2019)
ECHAM-SALSA	1.9 x 1.9	47	Bergman et al. (2012), Kokkola et al. (2018)
ECMWF-IFS	0.4 x 0.4		Rémy et al. (2019)
EMEP	0.5 x 0.5	20	Simpson et al. (2012), Schulz et al. (2012)
GEOS	1.0 x 1.0	72	Colarco et al. (2010),
GFDL-AM4	1.0 x 1.2	33	Zhao et al. (2018b)
GISS-OMA	2.0 x 2.5	40	Koch et al. (2006, 2007), Tsigaridis et al. (2013)
INCA	1.3 x 2.5	79	Balkanski et al. (2004), Schulz et al. (2009)
NorESM2	0.9 x 1.2	32	Kirkevåg et al. (2018), Olivie et al. (in preparation), Seland et al. (in preparation)
OsloCTM3	2.2 x 2.2	60	MYHRE et al. (2007); Myhre et al. (2009)
SPRINTARS	0.6 x 0.6	56	Takemura et al. (2005)

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

Test type	Var.	Model	Freq.	Obs	Δ_{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 database IDs. Also indicated is whether models diagnosed a clear-sky AOD (CS) or not and which models were included in the AeroCom mean and median ensemble.

Name	AeroCom ID	CS AOD	Ensemble
CAM5-ATRAS	CAM5-ATRAS_AP3-CTRL	Y	Y
EC-Earth3-AerChem	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 statistics (NMB and R) retrieved when co-locating models with satellite data a) in monthly resolution and $5° \times 5°$ horizontally with requirement of at least 7 daily values to compute a monthly mean, as done in this study (*Low*) and b) in daily resolution and in highest available horizontal resolution from both data-sets (*High*).

		Statistics:	NMB [%]		R	
		Resolution:	Low	High	Low	High
Model	Satellite	Variable				
CAM5-ATRAS	AATSR4.3-SU	AOD	-1.8	-2.1	0.67	0.51
	MODIS6.1-aqua	AOD	-25.9	-20.4	0.58	0.36
	MODIS6.1-terra	AOD	-33.3	-28.2	0.58	0.36
ECMWF-IFS	AATSR4.3-SU	AE	-47.3	-36.9	0.74	0.65
		AOD	-19.2	-22.8	0.79	0.70
	MODIS6.1-aqua	AOD	-35.5	-24.4	0.64	0.51
	MODIS6.1-terra	AOD	-41.9	-31.5	0.62	0.51
EMEP	AATSR4.3-SU	AE	36.6	42.6	0.67	0.50
		AOD	-34.4	-30.4	0.73	0.58
		AOD<1μm	-10.9	-2.6	0.74	0.57
		AOD>1μm	-69.8	-69.3	0.64	0.54
	MODIS6.1-aqua	AOD	-45.4	-40.3	0.66	0.48
	MODIS6.1-terra	AOD	-50.8	-45.7	0.66	0.48
OsloCTM3	AATSR4.3-SU	AOD	-12.4	-13.4	0.83	0.69
	MODIS6.1-aqua	AOD	-27.4	-28.3	0.72	0.52
	MODIS6.1-terra	AOD	-34.6	-35.2	0.72	0.51
SPRINTARS	AATSR4.3-SU	AE	-51.1	-41.4	0.59	0.52
TM5	AATSR4.3-SU	AE	2.9	8.7	0.74	0.62
		AOD	-1.8	-3.9	0.75	0.55
		AOD<1μm	3.4	3.4	0.81	0.66
		AOD>1μm	-9.8	-14.6	0.64	0.41
	MODIS6.1-aqua	AOD	-19.9	-18.2	0.73	0.53
	MODIS6.1-terra	AOD	-27.8	-25.9	0.72	0.52
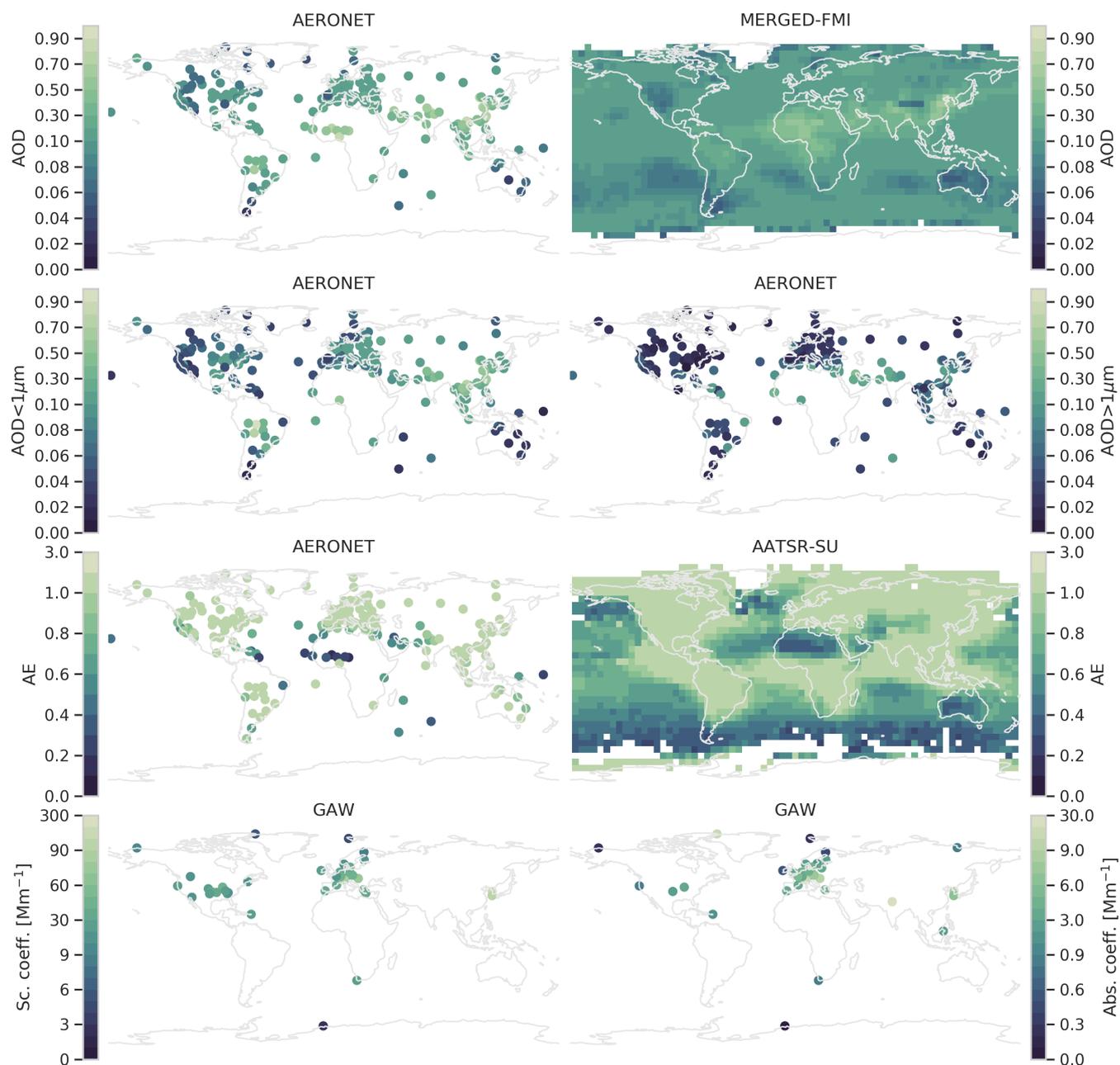
Figure 1. Yearly averages of AODs from AERONET and merged satellite data-set (top panel), fine and coarse AOD from AERONET (2nd panel), AE from AERONET and AATSR (3rd panel) as well as surface in-situ observations of scattering and absorption coefficients.
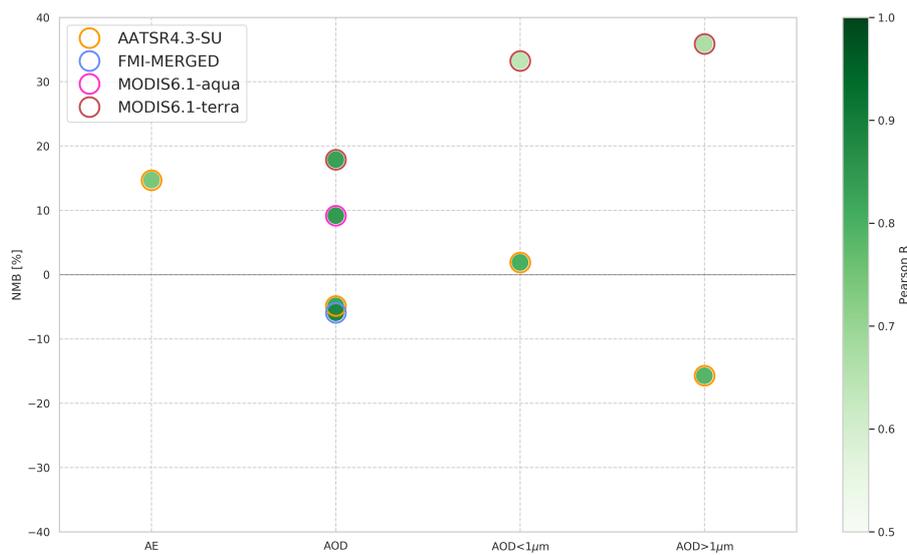
Figure 2. NMBs from satellite evaluation against AERONET for different variables. Also plotted are the corresponding correlation coefficients in green colors. Note that fine and coarse AOD from MODIS terra is not further used in this study.
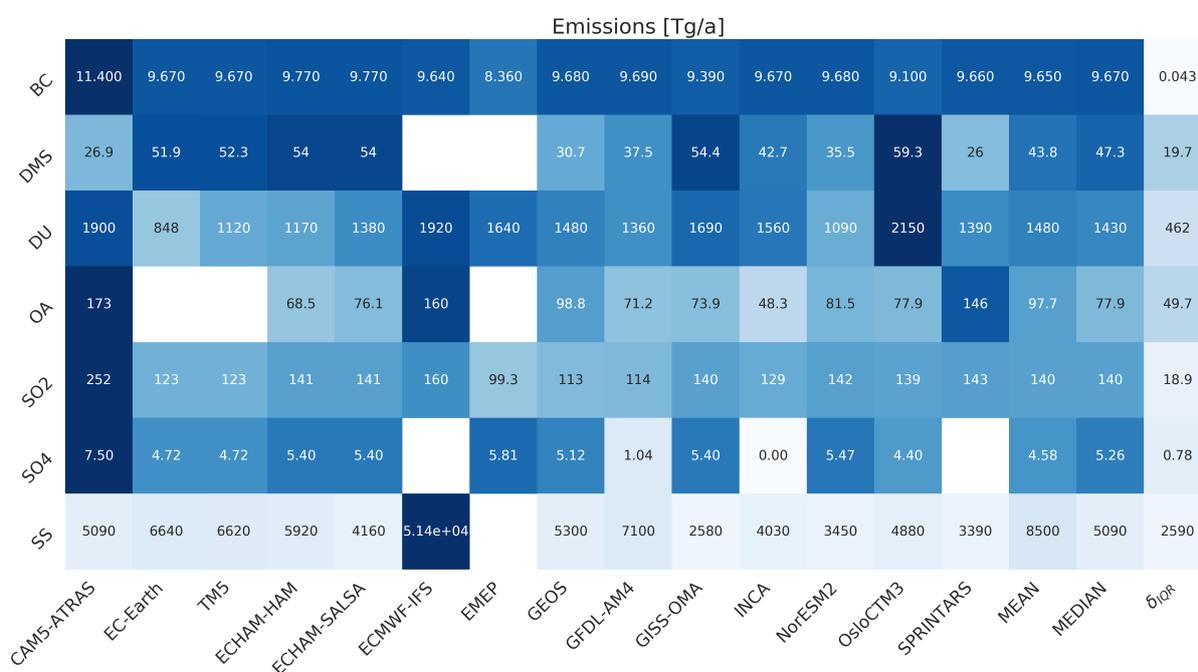
Figure 3. Global emissions of major aerosol species and precursors. Units are full molecular weight and for OA, the total organic weight is used. Note that only major species are included and that other potentially provided species (e.g. NO$_x$ or NH$_3$, VOCs) are not shown. The rightmost columns show mean, median and spread of the results from the individual models, the latter being computed as the half difference between 1st and 3rd quantiles. Colors are scaled to min and max row-wise in order to highlight differences between the models.
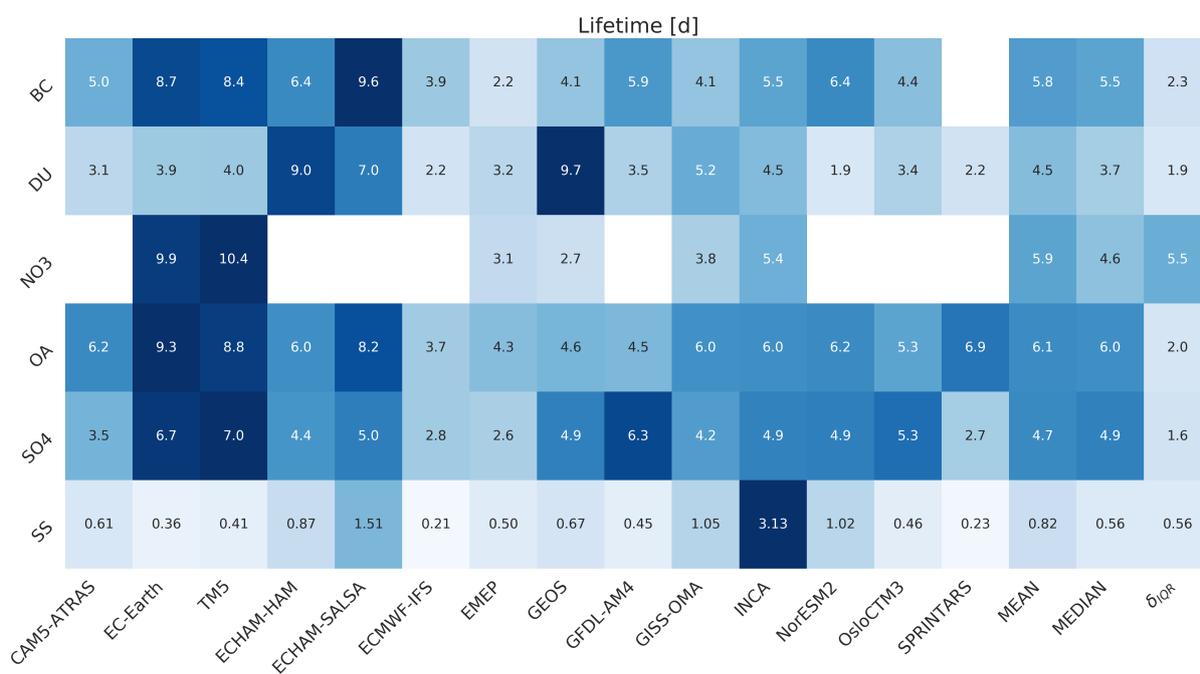
Figure 4. Global lifetimes in days for all major aerosol species, computed from burdens (Fig. 5) and total deposition (wet + dry). A more detailed description of this plot type is provided in Fig. 3.
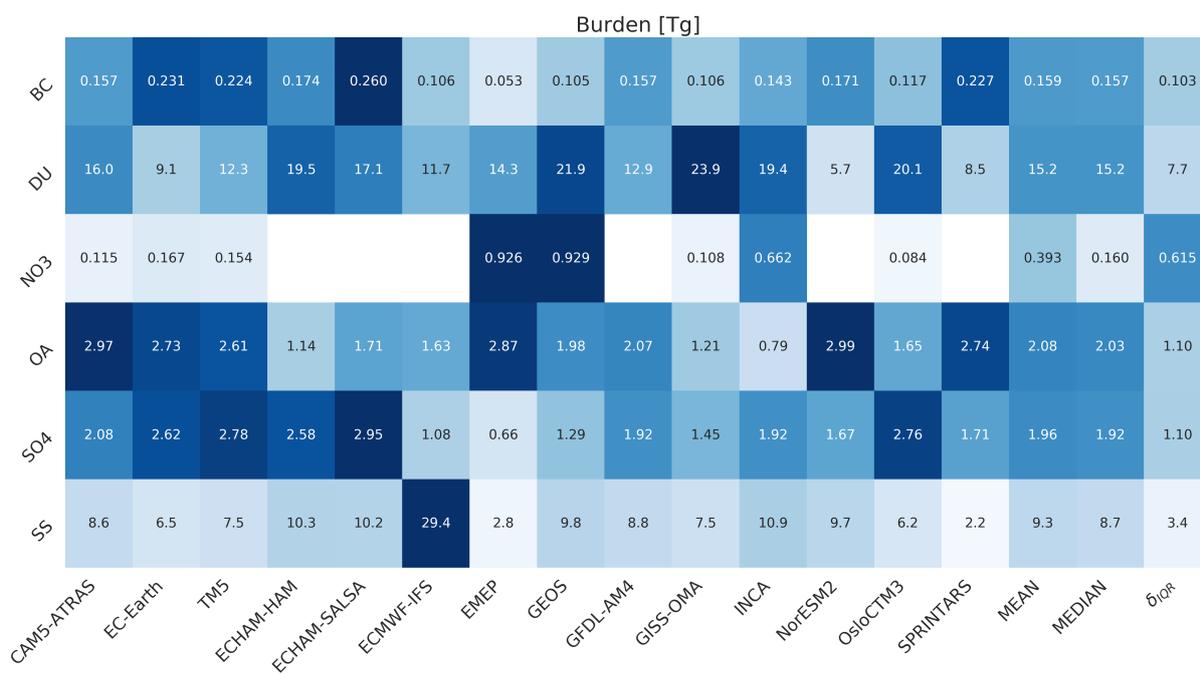
Figure 5. Global annual burdens of major aerosol species in units of Tg. A more detailed description of this plot type is provided in Fig. 3.
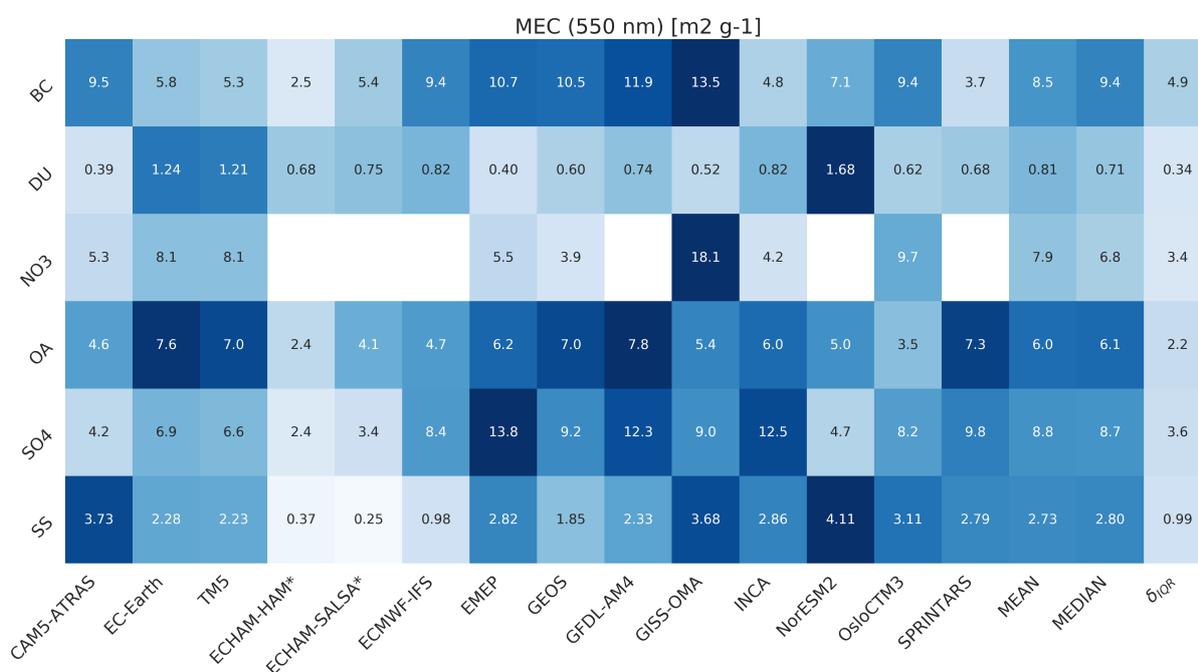
Figure 6. Globally averaged columnar MECs of models for all major aerosol species. The MEC for each species i is computed from $OD_i/LOAD_i$ (Figs. 5, 7). Note that the two ECHAM models reported the OD_i fields at dry conditions (indicated with a star, Fig. 7) and thus, show comparatively small MECs for the hydrophilic species. Therefore, they were excluded for the computation of mean, median and diversity shown in the rightmost columns. Hence, A more detailed description of this plot type is provided in Fig. 3.
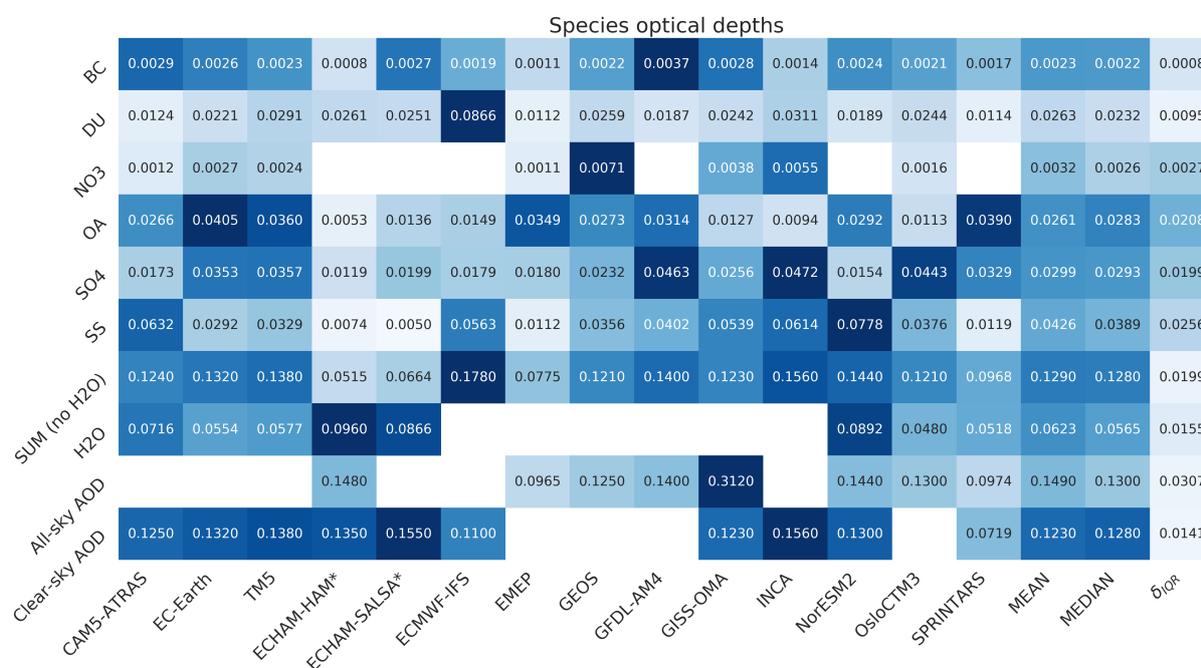
Figure 7. ODs from individual species as well as the sum and, dependent on availability clear-sky and all-sky AOD. Please note that for OsloCTM3 an additional OD of 0.0086 due to biomass burning was reported (combination of OA and BC) which is not included here. Like in Fig. 6, the two ECHAM models were excluded for the computation of mean, median and diversity, since dry speciated ODs were reported. A more detailed description of this plot type is provided in Fig. 3.
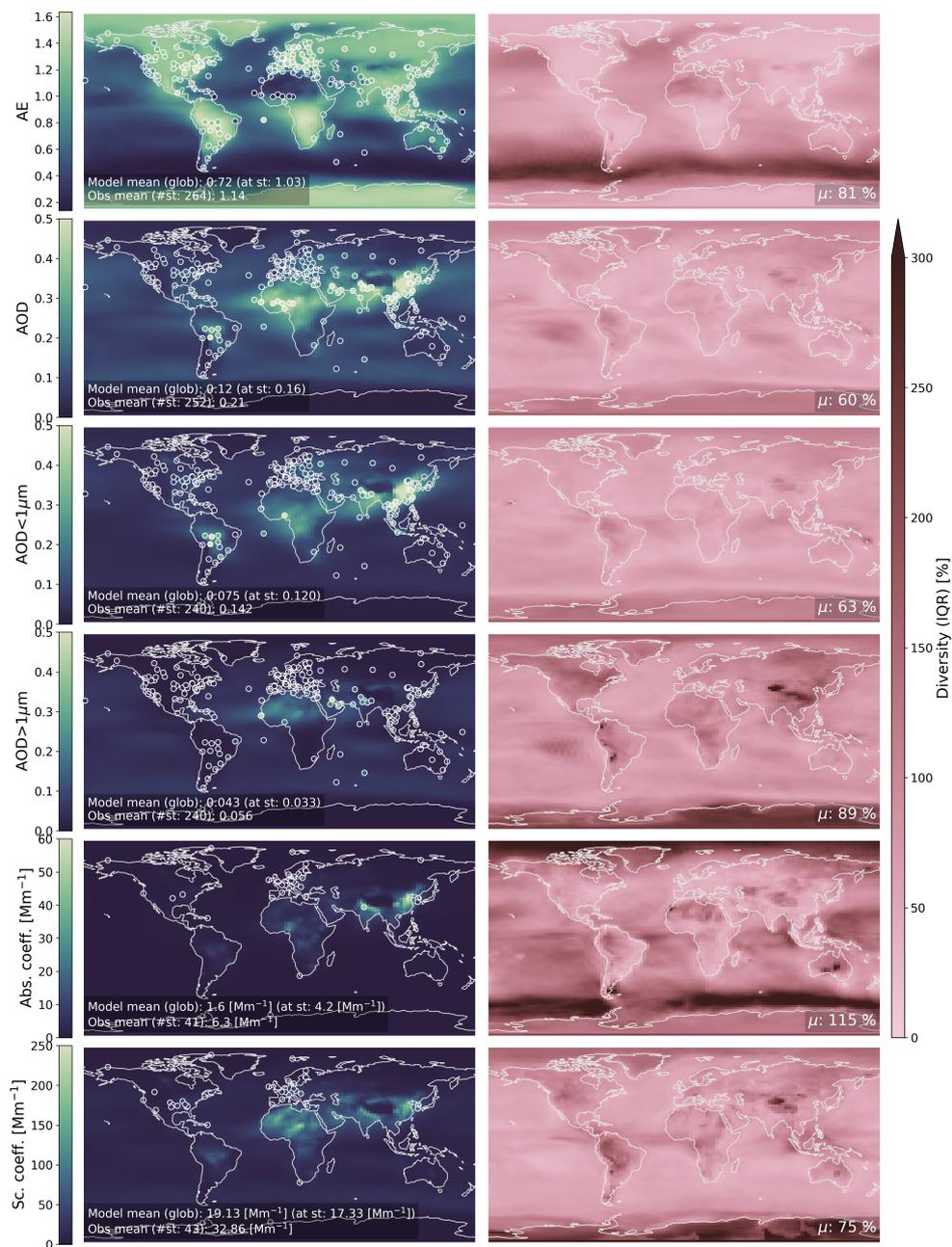
Figure 8. Left: maps showing yearly averages of relevant variables from the ensemble model as well as mean values from corresponding ground-based network used (circles). Also shown are the yearly mean values from model (both global and at obs. stations) as well as the observation mean from all stations. Right: diversity fields of ensemble mean calculated using standard deviation of the individual results normalised by the mean (Textor et al. (2006))
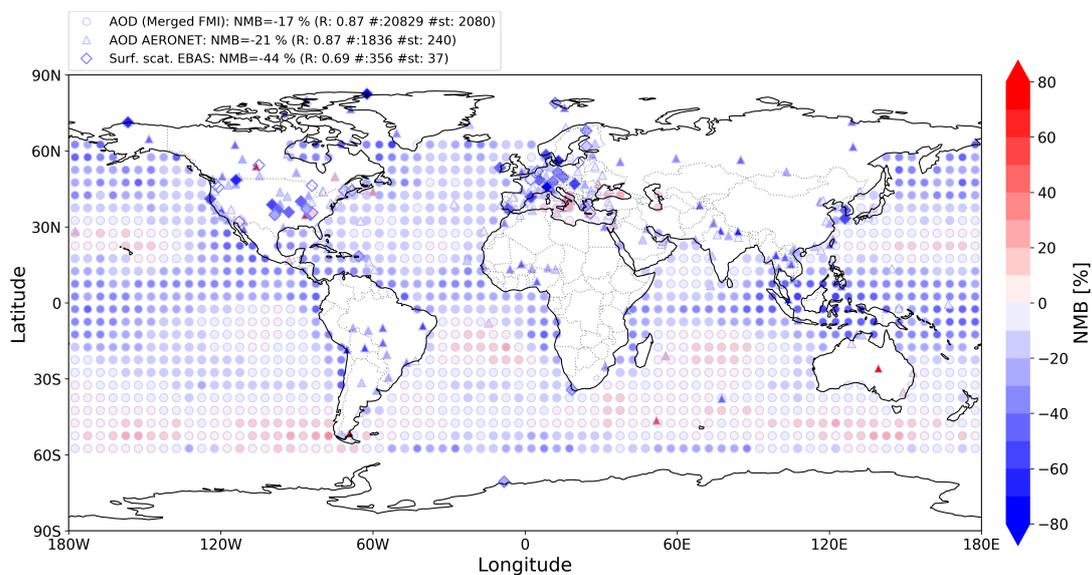
Figure 9. Figure showing NMB in percent of the ensemble median AOD against the merged satellite AOD data-set (circles, only ocean locations are displayed) and AERONET (triangles) as well as surface scattering coefficient against the in-situ sites (diamond). The edge colors of the markers correspond to the respective global average NMB, which is also indicated in the legend as well as Pearson correlation coefficient and total number of monthly data points and number of stations respective grid points for the Merged FMI product.
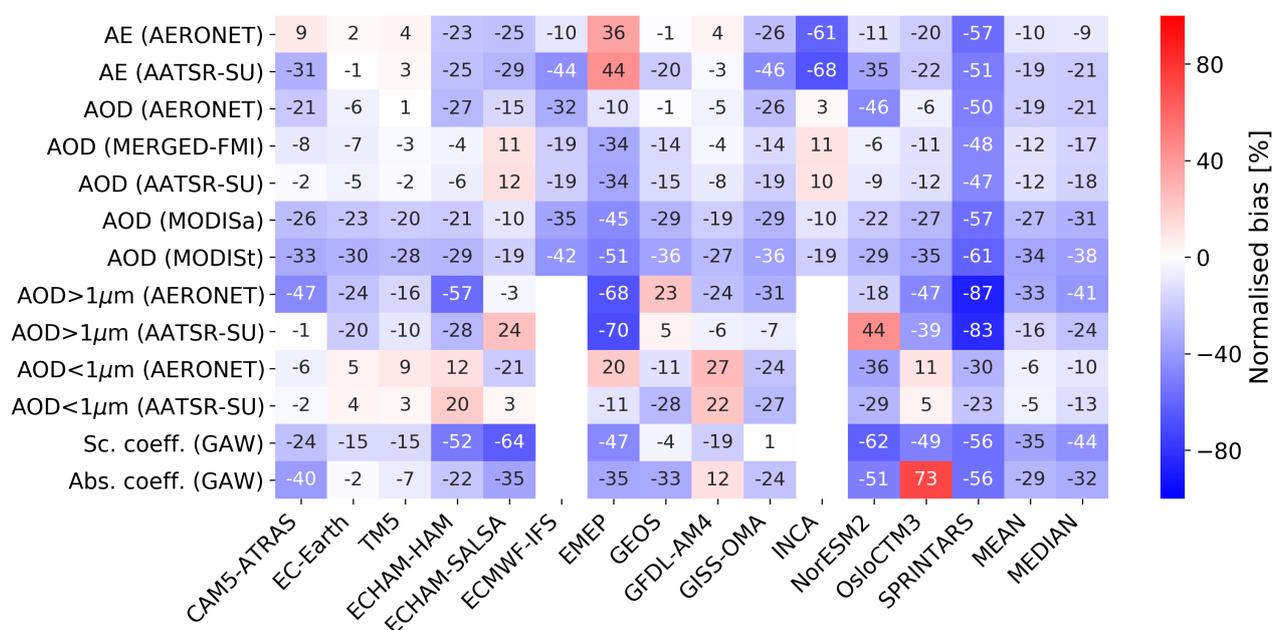
Figure 10. Normalised mean bias (NMB) computed from the monthly collocated data for each model (columns) and observation / variable combination (rows). For the $5° \times 5°$ satellite products, area weights were applied to compute the average bias. Please note that the biases do not represent global 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 (Sect. 2.5 and Fig. 2).
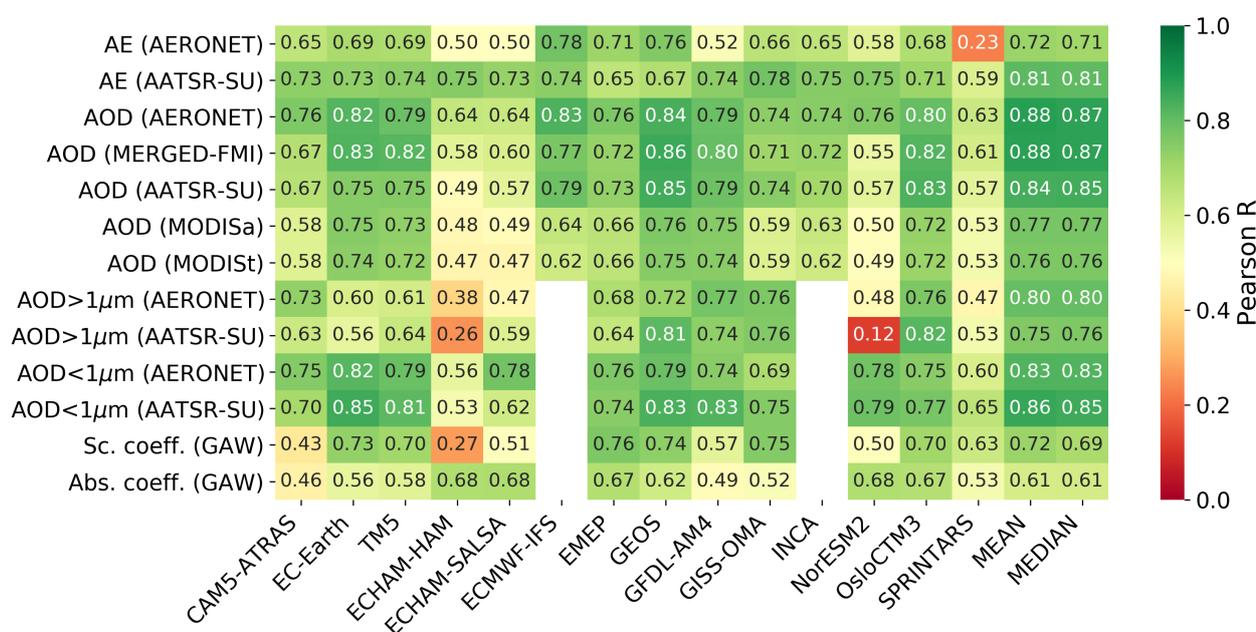
Figure 11. Pearson correlation coefficients (R) computed from the monthly colocated data for each model (columns) and observation / variable combination (rows). For the $5° \times 5°$ satellite products, area weights were applied to the monthly values. Please note further remarks on representativity in Fig. 10.
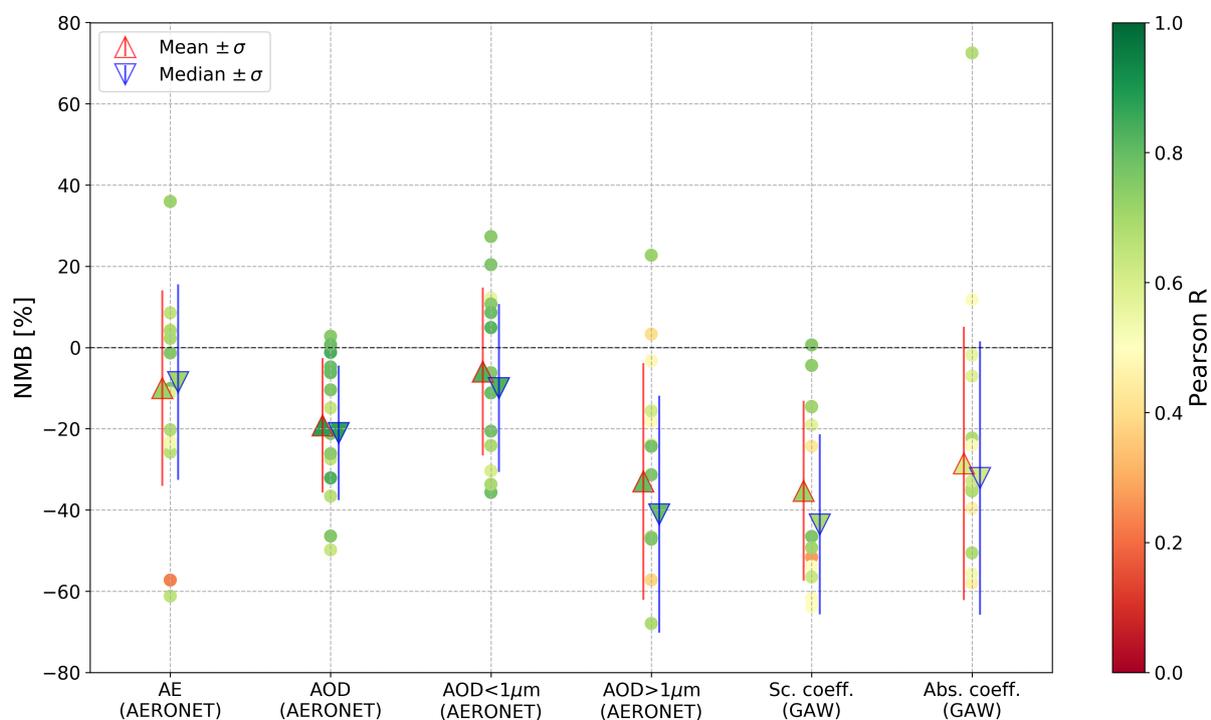
Figure 12. Overall results of optical properties evaluation for all models and the AeroCom ensemble model. NMB Biases of all variables for AeroCom median (blue triangles) and mean (red triangles) as well as those from individual models (circles). Pearson correlation coefficients are plotted in red-yellow-green colors (same as in Fig. 11). Also included is the standard deviation of NMBs from the models for each variable (red and blue error bars).
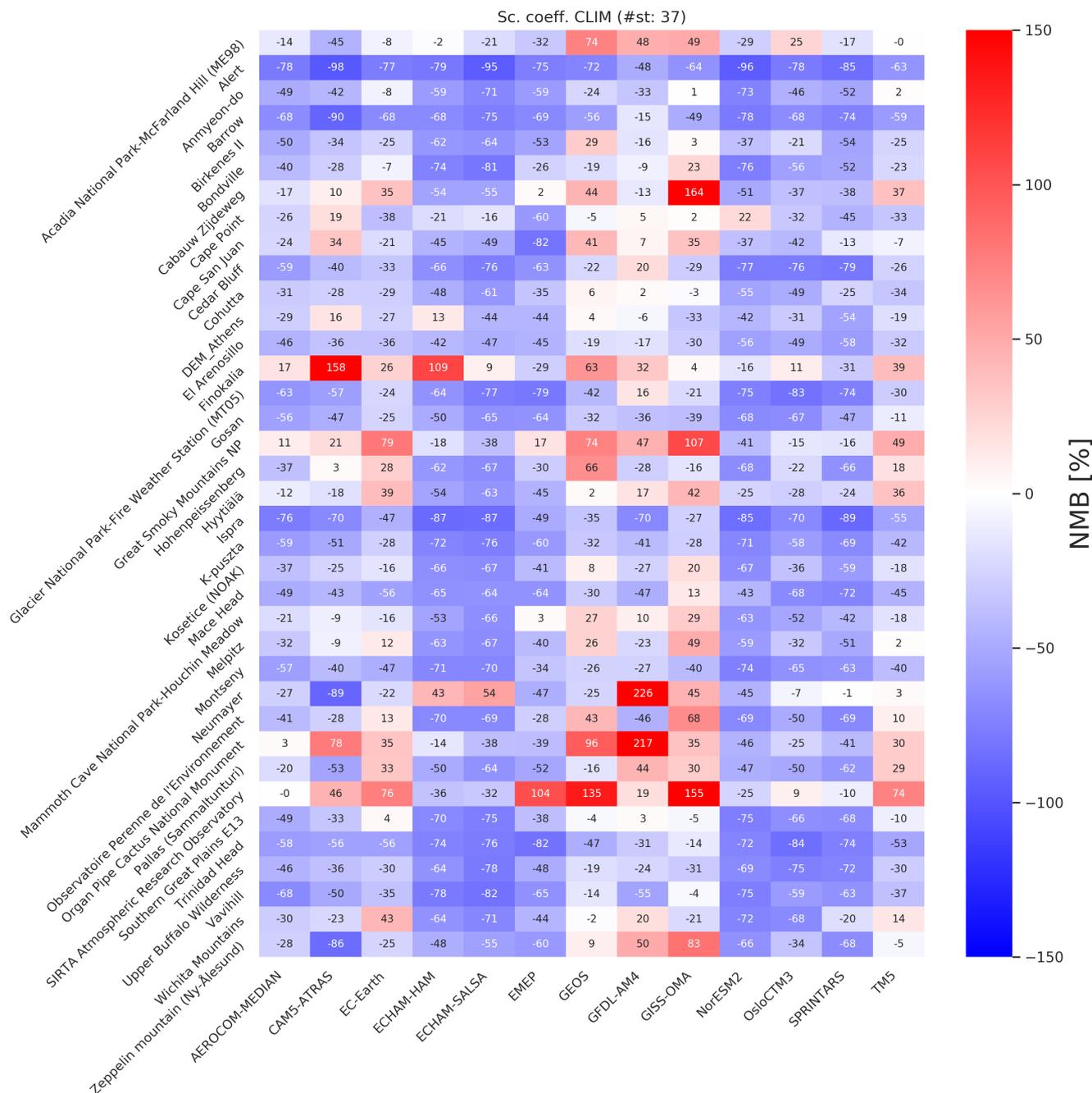
Figure A1. Model biases of surface dry scattering at all GAW in-situ sites that had sufficient temporal coverage to compute monthly climatology.

Figure A2. Model biases of surface absorption coefficient at all GAW in-situ sites that had sufficient temporal coverage to compute monthly climatology.
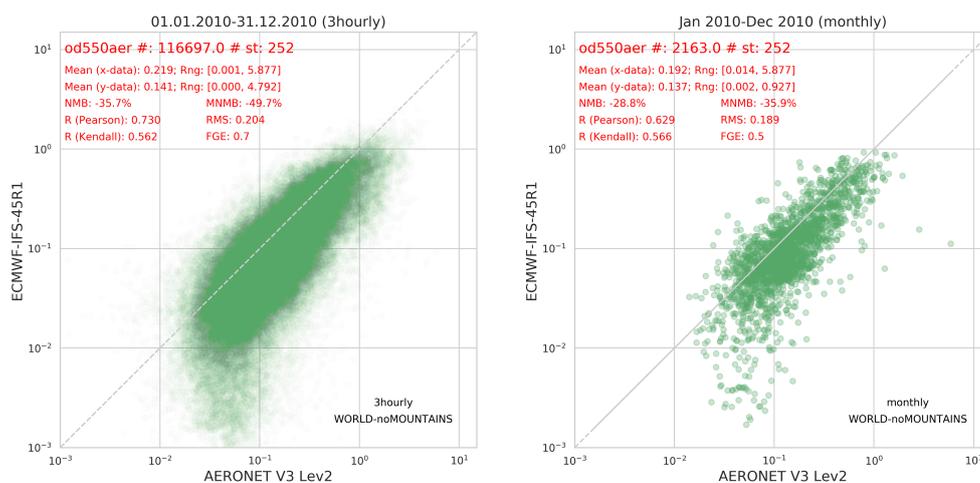
Figure A3. Scatter plot showing results of 3-hourly (left) vs. monthly (right) colocation of AOD from ECMWF-IFS model against AERONET all points data. Also included are statistical results.
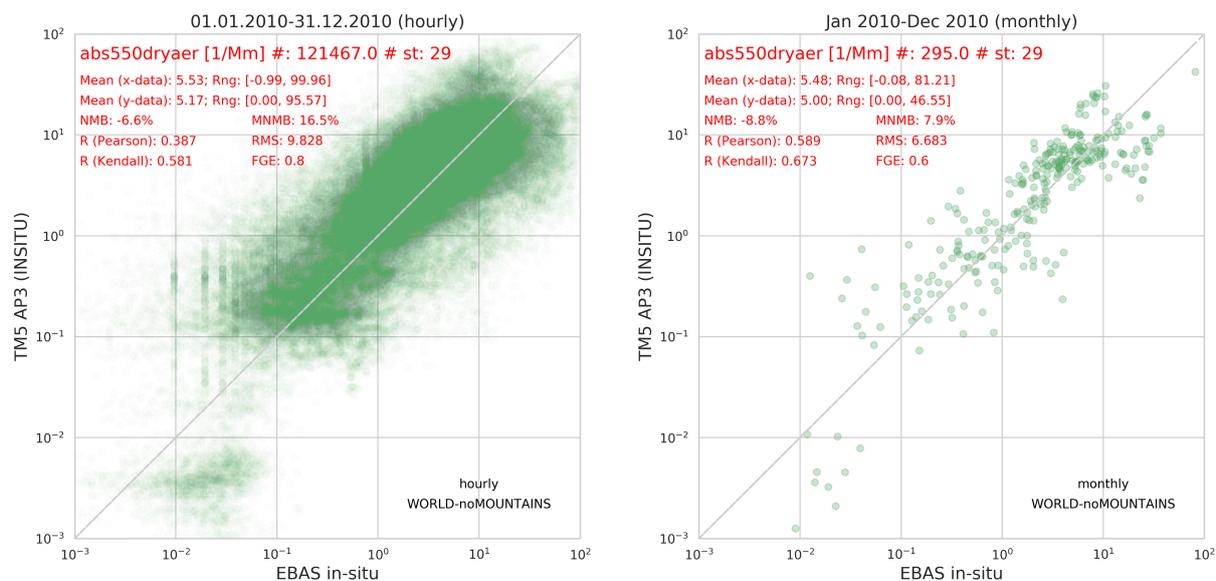
Figure A4. Scatter plot showing results of hourly (left) vs. monthly (right) colocation of in-situ surface absorption from TM5 model (from AeroCom INSITU experiment, i.e. different version than the one used in this study) evaluated at GAW stations. Also included are statistical results.
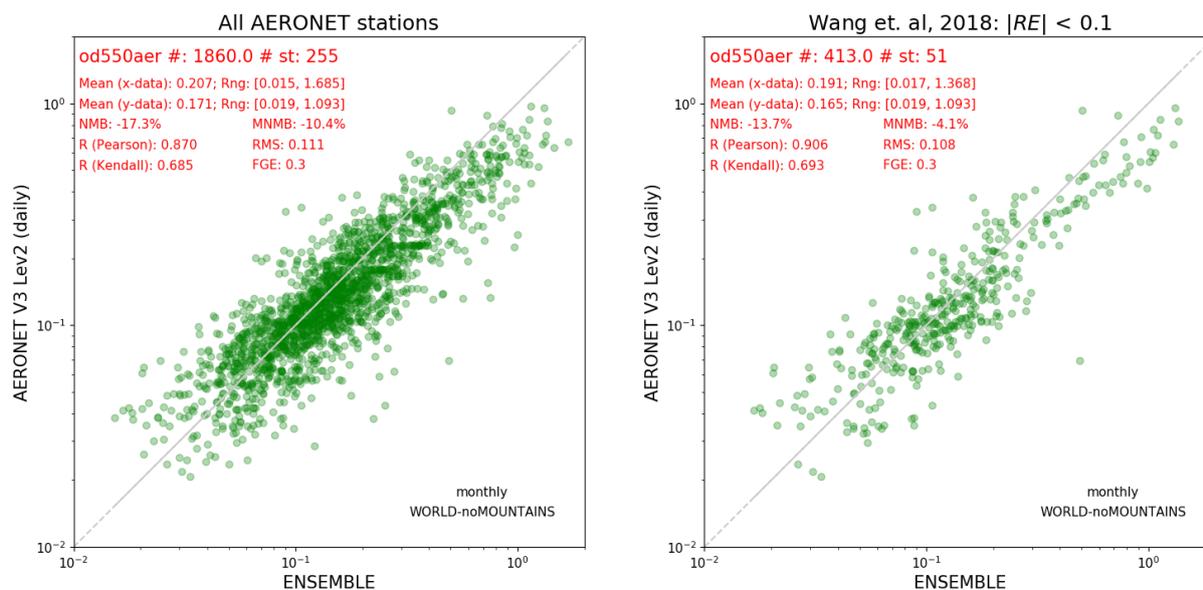
Figure A5. Scatter plot showing colocation results of the ENSEMBLE model AOD evaluated at all available AERONET stations (left) and evaluated only at stations with small spatial representativity errors, selected based on the results from Wang et al. (2018)