JGR Atmospheres

RESEARCH ARTICLE
10.1029/2021JD035483

Uncertainty in Aerosol Optical Depth From Modern Aerosol-Climate Models, Reanalyses, and Satellite Products

Annika Vogel[1,2,3], Ghazi Alessa[4], Robert Scheele[1], Lisa Weber[1,5], Oleg Dubovik[6], Peter North[7], and Stephanie Fiedler[1,4,5]

[1]Institute of Geophysics and Meteorology, University of Cologne, Cologne, Germany, [2]Rhenish Institute for Environmental Research at the University of Cologne, Cologne, Germany, [3]Now at Air Quality Research Division, Environment and Climate Change Canada, Dorval, QC, Canada, [4]Formerly at Max-Planck-Institute for Meteorology, Hamburg, Germany, [5]Hans-Ertel-Centre for Weather Research, Climate Monitoring and Diagnostics, Cologne/Bonn, Germany, [6]Laboratoire d'Optique Atmosphérique, CNRS, University Lille, Lille, France, [7]Department of Geography, Global Environmental Modelling and Earth Observation (GEMEO), Swansea University, Swansea, UK

Abstract Despite the implication of aerosols for the radiation budget, there are persistent differences in data for the aerosol optical depth (τ) for 1998–2019. This study presents a comprehensive evaluation of the large-scale spatio-temporal patterns of mid-visible τ from modern data sets. In total, we assessed 94 different global data sets from eight satellite retrievals, four aerosol-climate model ensembles, one operational ensemble product, two reanalyses, one climatology and one merged satellite product. We include the new satellite data SLSTR and aerosol-climate simulations from the Coupled Model Intercomparison Project Phase 6 (CMIP6) and the Aerosol Comparisons between Observations and Models Phase 3 (AeroCom-III). Our intercomparison highlights model differences and observational uncertainty. Spatial mean τ for 60°N – 60°S ranges from 0.124 to 0.164 for individual satellites, with a mean of 0.14. Averaged τ from aerosol-climate model ensembles fall within this satellite range, but individual models do not. Our assessment suggests no systematic improvement compared to CMIP5 and AeroCom-I. Although some regional biases have been reduced, τ from both CMIP6 and AeroCom-III are for instance substantially larger along extra-tropical storm tracks compared to the satellite products. The considerable uncertainty in observed τ implies that a model evaluation based on a single satellite product might draw biased conclusions. This underlines the need for continued efforts to improve both model and satellite estimates of τ, for example, through measurement campaigns in areas of particularly uncertain satellite estimates identified in this study, to facilitate a better understanding of aerosol effects in the Earth system.

Plain Language Summary Aerosols are known to affect atmospheric processes. For instance, particles emitted during dust storms, biomass burning and anthropogenic activities affect air quality and influence the climate through effects on solar radiation and clouds. Although many studies address such aerosol effects, there is a persistent difference in current estimates of the amount of aerosols in the atmosphere across observations and complex climate models. This study documents the data differences for aerosol amounts, including new estimates from climate-model simulations and satellite products. We quantify considerable differences across aerosol amount estimates as well as regional and seasonal variations of extended and new data. Further, this study addresses the question to what extent complex climate models have improved over the past decades in light of observational uncertainty.

Key Points:

- Present-day patterns in aerosol optical depth differ substantially between 94 modern global data sets
- The range in spatial means from individual satellites is −11% to +17% of the multi-satellite mean
- Spatial means from climate model intercomparison projects fall within the satellite range but strong regional differences are identified

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Vogel,
av@eurad.uni-koeln.de

Citation:

Vogel, A., Alessa, G., Scheele, R., Weber, L., Dubovik, O., North, P., & Fiedler, S. (2022). Uncertainty in aerosol optical depth from modern aerosol-climate models, reanalyses, and satellite products. *Journal of Geophysical Research: Atmospheres*, *127*, e2021JD035483. https://doi.org/10.1029/2021JD035483

Received 30 JUN 2021
Accepted 21 DEC 2021

Author Contributions:

Conceptualization: Stephanie Fiedler
Formal analysis: Annika Vogel, Ghazi Alessa, Robert Scheele, Lisa Weber
Supervision: Stephanie Fiedler
Visualization: Annika Vogel, Ghazi Alessa, Robert Scheele, Lisa Weber
Writing – original draft: Annika Vogel, Stephanie Fiedler
Writing – review & editing: Annika Vogel, Ghazi Alessa, Robert Scheele, Lisa Weber, Oleg Dubovik, Peter North, Stephanie Fiedler

1. Introduction

The mid-visible aerosol optical depth (τ) is important to determine the radiation budget and to understand climate. Given the number of studies calculating changes in the radiation budget, one could think that bulk aerosol quantities like τ are quantitatively well understood. There is, however, evidence that we have considerable uncertainty in our knowledge and capacity to simulate τ, despite several available observational data sets. In atmospheric models, uncertainty in τ partly stems from processes that determine aerosol emissions and their lifetime. In observations, uncertainty in the global mean τ was reported in the context of aerosol radiative effects (Bellouin et al., 2020) and in a comparison of satellite retrievals against ground-based measurements (Schutgens et al., 2020). Another example is the difference in historical trends of τ in aerosol-climate model simulations and observations (Cherian & Quaas, 2020; Mortier et al., 2020; Moseid et al., 2020; Zhang & Reid, 2010).

In light of seemingly persistent differences in changes of τ across satellite and model data, we address in this study the current quantitative knowledge of the spatio-temporal representation of τ in the present climate (1998–2019). We use a total of 94 different estimates based on five different methodological approaches. Our data intercomparison of τ addresses the following two questions.

1. How well do we know the spatio-temporal large-scale patterns of present-day τ from modern model simulations and observations?
2. Has the representation of τ improved in the new phases of aerosol-climate model intercomparison projects?

Our intercomparison study aims (a) to identify regions and months when τ is particularly uncertain in global observational data to guide the choice for future measurement campaigns and (b) to shed light on the aerosol-climate model performance for τ with a broader view on the observational uncertainty in τ patterns than previous studies.

We expect a spread in τ across the data sets for several reasons. These include different spatio-temporal coverage of satellite products due to the specific footprint of the instrument and the orbit of the satellite. Moreover, an accurate retrieval of aerosol information from the space-borne radiation measurements depends on many aspects, for example, cloud masking and surface conditions (e.g., Li et al., 2009; Witek et al., 2018). Uncertainties in satellite products are different from model spreads and their biases to observations. Global modeling offers a spatially and temporally complete coverage for information on τ. The model results stem from a multitude of interacting processes and depend therefore on their sufficient accuracy. A model spread is explained by differences in processes affecting the emission, transport, and deposition of aerosols, for example, sub-grid scale processes such as moist convection and precipitation (e.g., Fiedler et al., 2020; Marsham et al., 2013), as well as the chemical and physical properties of the aerosols, for example, their size distribution and composition (e.g., Kok et al., 2017; Samset et al., 2018). Not all influencing factors in model simulations are well represented, for example, dust-emitting winds in deserts (e.g., Heinold et al., 2013; Roberts et al., 2017; Shao, 2000). Model differences are thought to arise from a combination of different processes and may involve feedbacks, for example, via a precipitation-vegetation feedback influencing dust-aerosol emissions in semi-arid regions (e.g., Martin & Levine, 2012; Shao et al., 2011). Much work is dedicated to evaluating and improving satellite products and models on both regional and global scales that involves several data sources for τ (e.g., Cherian & Quaas, 2020; Evan et al., 2014; Huneeus et al., 2011; Johnson et al., 2011; Kahn et al., 2009; Mangla et al., 2020; Misra et al., 2016; Pérez et al., 2011; D. A. Ridley et al., 2016; Sai Suman et al., 2014; Tegen et al., 2018; Watson-Parris et al., 2019; Witek et al., 2016). One might therefore expect a subsequent improvement of our knowledge of present-day τ, for example, from aerosol-climate models participating in the coupled model inter-comparison project phase six (CMIP6, Eyring et al., 2016) compared to phase five (CMIP5, Taylor et al., 2012).

We build on the rich amount of existing works and present a first comprehensive intercomparison of global patterns in present-day τ from estimates of different modern techniques. The systematic data assessment in this study includes both new satellite data and a large number of CMIP6 simulations, in addition to previously presented data sets from aerosol-climate models, satellite retrievals and reanalyses. Taken together, we use τ from simulations of 73 different aerosol-climate models distributed over four aerosol-climate model intercomparison projects, from nine operational forecast systems combined to a multi-model mean, from eight satellite retrievals, from two global reanalysis products, from an established global aerosol climatology of the Max-Planck-Institute for Meteorology and from a new merged satellite product of the Finish Meteorological Institute. The large number of data sets allows us to calculate the spread in τ for different generations of aerosol-climate models and to quantify the uncertainty across observational products. The data and methods used in this study are introduced in Section 2, followed by the results in Section 3. We discuss the implication of the results and draw conclusions at the end.

2. Methods

2.1. Data

We use monthly means of the mid-visible aerosol optical depth of 94 different data sets, summarized in Table 1. The data sets are from different methodological approaches, namely four aerosol-climate model ensembles (Eyring et al., 2016; Gliβ et al., 2021; Schulz et al., 2006; Taylor et al., 2012), one operational aerosol multi-model

Table 1
Data Used in This Study

Name	Type	Time period	Horizontal res.	Reference
AeroCom-I	Aerosol-climate models	01/2000–12/2000	Various	Schulz et al. (2006)
AeroCom-III	Aerosol-climate models	01/2010–12/2010	Various	Gliβ et al. (2021)
CMIP5	Aerosol-climate models	01/1998–12/2005	Various	Taylor et al. (2012)
CMIP6	Aerosol-climate models	01/1998–12/2014	Various	Eyring et al. (2016)
ICAP	Operational ensemble	01/2015–12/2015	1° × 1°	Xian et al. (2019)
CAMS	Reanalysis	01/2003–12/2019	0.75° × 0.75°	Inness et al. (2019)
MERRA-2	Reanalysis	01/1998–12/2019	0.5° × 0.625°	Gelaro et al. (2017)
MAC-v2	Climatology	01/2005–12/2005	1° × 1°	Kinne (2019)
FMImerge	Merged satellites	01/1998–12/2017	Various	Sogacheva et al. (2020)
AATSR	Satellite	01/2002–12/2012	1° × 1°	P. R. North et al. (1999)
CALIPSO	Satellite	06/2006–05/2020	2° × 5°	Winker et al. (2010)
MISR	Satellite	01/2001–12/2019	0.5° × 0.5°	Kahn et al. (2005)
MODIS Aqua	Satellite	01/2003–12/2019	1° × 1°	R. C. Levy et al. (2007)
MODIS Terra	Satellite	01/2001–12/2019	1° × 1°	R. C. Levy et al. (2007)
POLDER	Satellite	01/2006–12/2011	1° × 1°	Chen et al. (2020)
SeaWiFS	Satellite	01/1998–12/2010	1° × 1°	Hsu et al. (2012)
SLSTR	Satellite	01/2018–12/2019	1° × 1°	P. North et al. (2021)

Note. Details on the CMIP models are given in Tables S1 and S2 in the Supporting Information S1.

ensemble (Xian et al., 2019), two global reanalyses (Gelaro et al., 2017; Inness et al., 2019), one global aerosol climatology (Kinne, 2019), as well as results from individual satellite observations (Chen et al., 2020; Hsu et al., 2012; Kahn et al., 2005; R. C. Levy et al., 2007; P. North et al., 2021; P. R. North et al., 1999; Winker et al., 2010) and a newly available merged satellite product (Sogacheva et al., 2020). Aerosol-climate models simulate τ using prescribed anthropogenic aerosol emissions and their precursors, paired with emissions of natural aerosols calculated online in the model simulations. Satellite retrievals determine τ by applying inverse methods to radiation measurements from instruments aboard satellites. And reanalyses combine observations with a model system through data assimilation.

2.1.1. Aerosol-Climate Models

Monthly mean output from aerosol-climate model simulations are taken from the Coupled Model Intercomparison Project phase five (CMIP5, Taylor et al., 2012) and six (CMIP6, Eyring et al., 2016), as well as simulations for the AEROsol Comparisons between Observations and Models (AeroCom) I (Kinne et al., 2006; Schulz et al., 2006) and III (Gliβ et al., 2021). The models calculate τ based on interactions of aerosols with atmospheric processes, that is, simulation of the emission, transport, and deposition of aerosols. Anthropogenic aerosol emissions are prescribed as input fields, whereas natural aerosol emissions, for example, sea-spray and desert-dust aerosols, are typically calculated online as function of other variables, for example, the near-surface wind speed simulated by the atmospheric model. For CMIP5 and CMIP6, we use one historical simulation per model which are free running simulations with year-to-year changes in atmospheric composition. These cover the time period from 1850 to different years in the 21st century (Table 1). We use 24 models from CMIP5 and 21 from CMIP6, with further details given in Table S1 and S2 in the Supporting Information S1. AeroCom-I and AeroCom-III are global aerosol-climate model simulations extending over the course of one year, namely 2000 and 2010. The participating models differ between AeroCom-I and AeroCom-III, for example, there are different models with more complex treatments of aerosols in AeroCom-III and the number of participating models reduced from 18 in AeroCom-I to 10 in AeroCom-III. Most of the contributing simulations in AeroCom are nudged to or driven by meteorological conditions from reanalysis of the same year. We calculate the multi-model average for each model intercomparison project using monthly output for the time periods listed in Table 1. Additionally, we use the multi-model monthly ensemble mean of the International Cooperative for Aerosol Prediction (ICAP, Xian

et al., 2019). The ICAP data is from operational and quasi-operational global forecast systems that interactively simulate aerosols for the period January 2015 to May 2019. The latest stage used here includes forecasts of τ for 550 nm from nine global aerosol models. Two of these models simulate only dust aerosols interactively. The other seven models additionally simulate sea-salt, black carbon, organic carbon and sulphate aerosols. Six of the models assimilate near real-time aerosol observations from MODerate resolution Imaging Spectroradiometer (MODIS), which contain filtered τ over oceans. We use here the publicly available multi-model ensemble mean, provided to us as monthly climatology representative for the year 2015 (Table 1). It weights all models equally consistent with our averaging approach for CMIP and AeroCom data (Section 2.2).

2.1.2. Global Reanalyses

Our assessment includes the global reanalysis data set of the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2, Gelaro et al., 2017), produced by NASA's Global Modeling and Assimilation Office (GMAO). The MERRA-2 atmospheric reanalysis system combines the GEOS-5 atmospheric model (Molod et al., 2015) with the 3D-variational data assimilation scheme GSI. Five prognostic aerosol species are simulated and assimilated as τ (Randles et al., 2017). It uses τ at 550 nm from AVHRR and MODIS onboard Terra and Aqua, and retrievals from Multiangle Imaging SpectroRadiometer (MISR) as well as ground-based measurements from the AErosol RObotic NETwork (AERONET, e.g., Holben et al., 1998). Note that AVHRR observations are only available until 2002 and the assimilation is mainly driven by MODIS and MISR data since then. We additionally include the global reanalysis of atmospheric composition from the Copernicus Atmosphere Monitoring Service (CAMS, Inness et al., 2019), produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). Based on the ECMWF's Integrated Forecasting System (IFS), CAMS simulates 12 prognostic aerosols (Inness et al., 2019). The total τ of the aerosol species is corrected by an incremental 4D-variational data assimilation system (Benedetti et al., 2009). Total τ of MODIS retrievals from Terra and Aqua satellites, and of AATSR onboard Envisat are assimilated into the system including a variational bias correction for both instruments. Note that AATSR is only available until March 2012 and τ is only corrected by MODIS from that time onward. The assimilation of satellite products does not necessarily align the reanalysis fields with the ones of the assimilated satellite. Instead, forecast fields are pushed toward observed values while they still rely on the underlying forecast model.

2.1.3. Climatology

We use two data sets that combine different types of information by means other than data assimilation, MAC-v2 and FMImerge. The Max-Planck-Institute for Meteorology (MPI-M) created the MPI-M Aerosol Climatology version two (MAC-v2, Kinne, 2019). The MAC-v2 climatology is based on several years of data to create the mean monthly climatology for the reference year 2005. It uses the spatial information from the AeroCom-I aerosol-climate models and constrains the magnitude with local observations. The observations of τ are taken from optical measurements with sun photometers at stations of AERONET (Holben et al., 1998) and opportunistic observations with hand-held sun photometers aboard research vessels in AERONET's Maritime Aerosol Network (MAN, Smirnov et al., 2009). The satellite-based climatology FMImerge was developed by the Finish Meteorological Institute (FMI) by merging different satellite products into one data set (Sogacheva et al., 2020). FMImerge provides monthly mean τ from 1998 to 2017. It uses 12 satellite products according to the available time period of the instruments AVHRR, SeaWiFS, AATSR, MODIS Terra and Aqua, MISR, POLDER and VIIRS (see also below). By combining different products, the FMI aimed at removing biases between individual products and increasing the spatio-temporal coverage. The FMImerge data was validated against AERONET.

2.1.4. Satellite Data

This study uses τ products from eight satellite-based instruments in total. These are (a) the Level-3 monthly aerosol product (MIL3MAE) version 4 from the MISR, a passive radiometer on Terra that provides images in nine different directions (Kahn et al., 2005, 2010), (b and c) the Level-3 monthly aerosol products (MOD08_3) collection-6 (C6; Hsu et al., 2013; R. Levy et al., 2013) of the MODIS on Aqua and Terra, a multi-spectral radiometer designed to retrieve aerosol microphysical and optical properties (Tanré et al., 1997; R. C. Levy et al., 2007), (d) the retrieval from images of the Sea-viewing Wide Field-of-view Sensor (SeaWiFS, Hsu et al., 2012), a multispectral ocean color sensor in the visible and near infrared (near infrared) spectrum (Jamet et al., 2004; Wang et al., 2005), and (e) the Level-3 daily mean aerosol properties product from the Advanced Along-Track Scanning Radiometer (AATSR), an instrument on board of the European Space Agency (ESA) satellite ENVISAT (P.

R. J. North, 2002; Bevan et al., 2012; Popp et al., 2016). We further include (f) the new Sea and Land Surface Temperature Radiometer (SLSTR) SU v1.12 satellite AOD product from both Sentinel-3 satellites, developed at Swansea University under the ESA Climate Change Initiative (P. North et al., 2021). While the SLSTR data set uses a similar algorithm to the AATSR retrieval, the SLSTR instrument differs, principally with the addition of two solar reflective channels, a wider viewing-swath, and change to the directionality of the dual view scanning. Moreover, we use (g) the PARASOL/GRASP/Models 2.1 product, produced with the Generalized Retrieval of Atmosphere and Surface Properties (GRASP) algorithm (Dubovik et al., 2014, 2021). The product is based on observations of the Polarization and Directionality of the Earth's Reflectances (POLDER) multi-angular polarimeter on board of the PARASOL satellite, which was developed by the French space agency Centre National D'Études Spatiales (CNES) and has been compared to station observations for aerosols (Chen et al., 2020). In addition to the passive instruments, we also use (h) the Level-3 monthly aerosol product from the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO). The active sensor CALIOP onboard CALIPSO is a two wavelength (532 and 1064 nm) polarization lidar (Winker et al., 2010).

2.2. Analysis Strategy

We use metrics common in climate research to compare characteristics of τ across the data sets, for example, averages, biases, standard deviations, and correlation coefficients for different temporal and spatial scales. We also use methods that are less often applied, but are informative to characterize spatio-temporal differences, for example, the calculation of the hemispheric asymmetry A:

$$A(\varphi) = \frac{\tau(\varphi) - \tau(-\varphi)}{2}, \tag{1}$$

where φ is the geographical latitude. Our spatial mean statistics are for multiple years for most of the Earth (60°N–60°S). Differences of individual datasets are calculated with respect to the satellite mean, excluding FMImerge. This does not imply that the satellite mean is necessarily the optimal estimate compared to station observations, but it serves as reference to determine differences of the data sets. We also compute the range across datasets as maximum minus minimum. Differences across satellites (models) are called observational uncertainty (model spread). Our metrics are calculated for all available model datasets, that is, we did not exclude outliers in the multi-model ensembles to account for the full model diversity.

All data sets are interpolated to the same horizontal grid of $1° \times 1°$ prior to the analysis. To this end, we apply second order conservative remapping in spherical coordinates (P. Jones, 1999). MODIS Aqua and Terra have a different time for the overpass, that is, cover different scenes, and have a mean difference of 0.015–0.02 (R. C. Levy et al., 2018). We therefore compute the combined mean across MODIS Aqua and Terra, and refer to it as MODIS hereafter. Monthly data in AATSR and SeaWiFS that had zero values globally at every gridpoint are due to missing data and were thus excluded from the analysis.

We use all available years from the data sets in the analyses. This choice is motivated by the small effect of using the entire time period compared to using an overlapping period of several years on our findings. We tested the comparison of the data for overlapping time periods, which did not change our conclusions. Take for instance the spatial mean statistics (Table 2) that show remarkably small differences when τ is averaged over the complete time periods or over the shorter time period, January 2007 to December 2010, when most products are available (compare Section 2.1). The shorter time period leads to lower temporal year-to-year variability in all datasets as one would expect from the number of measurements (Figure 1). Consequently, we compute temporal means over all monthly τ data within the available time period of each data set at each gridbox. Note that year-to-year variability has an effect on data sets covering a single year, particularly on regional scales.

3. Results

3.1. Spatial Mean

We compute the satellite mean as reference and the range in the means of individual satellite data as observational uncertainty. The satellite estimate for mean τ between 60°N and 60°S, referred to as spatial mean hereafter, is 0.140 ± 0.016 (mean $\pm$ standard deviation). τ from individual satellites ranges from around 0.124 for CALIPSO

Table 2
Mean τ and Standard Deviation w.r.t. Complete Available Time Period and w.r.t. Over Lapping Period From January 2007 Until December 2010

Time period	Complete available time period		January 2007–December 2010	
Data	$\bar{\tau} \pm \sigma_{ens}$	σ_{time} [10^{-3}]	$\bar{\tau} \pm \sigma_{ens}$	σ_{time} [10^{-3}]
AerCom-I[a]	0.135 ± 0.024 (18%)	/	/	/
AerCom-III[a]	0.159 ± 0.072 (45%)	/	/	/
CMIP5[a]	0.140 ± 0.048 (34%)	0.41 (0.3%)	/	/
CMIP6	0.160 ± 0.031 (19%)	2.39 (1.5%)	0.162 ± 0.031 (19%)	1.30 (0.8%)
CAMS	0.164	3.59 (2.2%)	0.164	2.82 (1.7%)
MERRA-2	0.146	6.28 (4.3%)	0.151	0.84 (0.6%)
MAC-v2[a]	0.132	/	/	/
ICAP[a]	0.152	/	/	/
Model mean	**0.149 ± 0.012** (8%)	**3.17** (2.1%)	**/**	**/**
FMImerge	0.159	4.59 (2.9%)	0.161	1.94 (1.2%)
AATSR	0.152	4.06 (2.7%)	0.151	2.22 (1.5%)
CALIPSO	0.124	2.90 (2.3%)	0.125	1.97 (1.6%)
MISR	0.124	3.41 (2.8%)	0.125	1.37 (1.1%)
MODIS	0.160	10.40 (6.5%)	0.164	0.69 (0.4%)
POLDER	0.128	2.61 (2.0%)	0.127	2.98 (2.3%)
SeaWiFS	0.130	3.75 (2.9%)	0.134	1.38 (0.8%)
SLSTR[a]	0.164	0.45 (0.3%)	/	/
Satellite mean	**0.140 ± 0.018** (13%)	**3.94** (2.8%)	**/**	**/**

Note. Averages are computed for 60° N to 60° S. The year-to-year (σ_{time}) and ensemble (σ_{ens}) standard deviations are shown as absolute values and relative to the annual mean in %, were absolute σ_{time} is given in 10^{-3}. Mean modeled and observed τ and their standard deviations w.r.t. the individual products are given as extra rows denoted as "model mean" and "satellite mean," respectively. FMImerge does not contribute to the model and satellite means. The model and satellite mean σ_{ens} is based on mean τ of the individual data sets, not their internal variability. The satellite mean σ_{time} is the arithmetic mean of σ_{time} from the individual satellite products.
[a]Data not available in the overlapping period.

and MISR to up to 0.16 for MODIS and SLSTR (Table 2). With a mean of 0.159, FMImerge lies at the upper end of the satellite range. SLSTR has a shorter time period than most satellite products and covers a time of intense forest fires that might increase the mean τ. Also other satellite data covering years since 2013 like MISR and MODIS indicate a relatively larger mean τ in 2019 and 2020 than in other years (Figure 1). Interestingly, the year-to-year standard deviation of mean τ is larger in more recent years than earlier, with a maximum of up to 7% from MODIS (Table 2). This is broadly consistent with the larger variability in the later half of the time series of annual mean τ in Figure 1 for both ocean and land.

We find a larger spatial mean τ in the multi-model mean from CMIP6 than CMIP5. CMIP5 has the same mean τ as the satellite mean ($\tau = 0.14$), whereas CMIP6 lies at the upper end of the satellite range. The multi-model mean of CMIP6, and also AeroCom-III, have a larger spatial mean τ by 0.02, which corresponds to +14%, relative to the satellite mean. The ICAP mean is closest to the averaged τ across all model data sets. AerCom-I and MAC-v2 show higher spatial mean τ compared to the satellite mean (Table 2 and Figure 2).

Comparing to a single satellite product changes a model evaluation result due to differences in the spatial mean τ across satellites. If we would only compare CMIP6 against MODIS, against CAMS, which mainly assimilates τ from MODIS (see Section 2), or against the combined satellite product FMImerge, we would conclude an improvement for CMIP6 compared to CMIP5. Using the satellite mean, however, suggests that the CMIP6 mean is larger and lies at the upper end of the range of individual satellite products for the spatial mean τ. Satellite-to-model differences in τ can in parts arise due to the sampling and retrieval algorithms of satellite products,

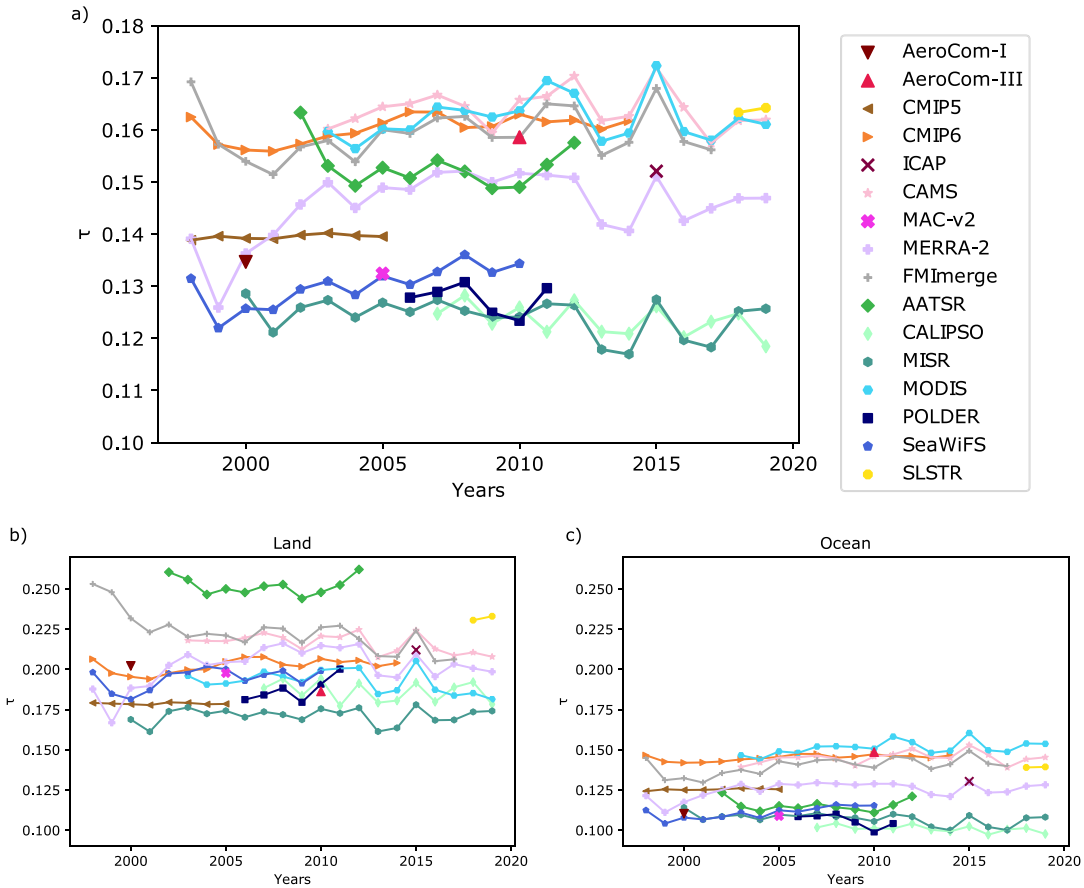

Figure 1. Time series of τ. Shown are annual spatial means of τ for (a) both land and ocean, (b) land only, and (c) ocean only. Data for the aerosol-climate model intercomparison projects are multi-model averages. AeroCom-I, AeroCom-III, ICAP and MAC-v2 are data for single years (see Table 1 for details).

discussed in Section 4.1. These results underline the need to use several observational products to evaluate a model and the need to better constraint observational estimates of τ.

Although the multi-model means for spatial mean τ fall within the observational range, not all individual models do. While the mean monthly ensemble spread differs considerably between the individual model intercomparison projects (σ_{ens} = 18%–45% relative to the mean, compare Table 2), it is generally higher than the variation of the model mean and the satellite mean. There is some improvement in the spread in spatial mean τ from models participating in aerosol-climate model intercomparisons over time, but the improvement is not seen for all intercomparison projects (Figure 2). The spread across the CMIP6 models has reduced compared to CMIP5, but the model spread in AeroCom-III is larger than for AeroCom-I. This might be associated with added complexity of models participating in AeroCom-III and poor observational references for validating all influencing factors, for example, sea-salt and desert-dust aerosols in remote regions (Witek et al., 2016). The distribution of spatial mean τ across individual models in AeroCom-III is asymmetric, indicated by differences between the mean and median (Figure 2). This is due to a few models with values well above 0.164, the maximum in spatial mean τ from satellite products, which substantially increase the spatial mean for AeroCom-III. In contrast to these high outliers, most AeroCom-III models have a spatial mean τ between 0.10 and 0.13, which is at the lower end of the satellite range or below.

The model mean for τ over ocean is larger compared to the satellite mean, while they are similar over land (Table 3). Taking the satellite mean as reference, the mean of CMIP6 over land improved compared to CMIP5, but not over ocean. At the same time the diversity of individual satellite datasets is larger over land than over ocean, consistent with a larger retrieval uncertainty over heterogeneous and brighter land surfaces. The variability in

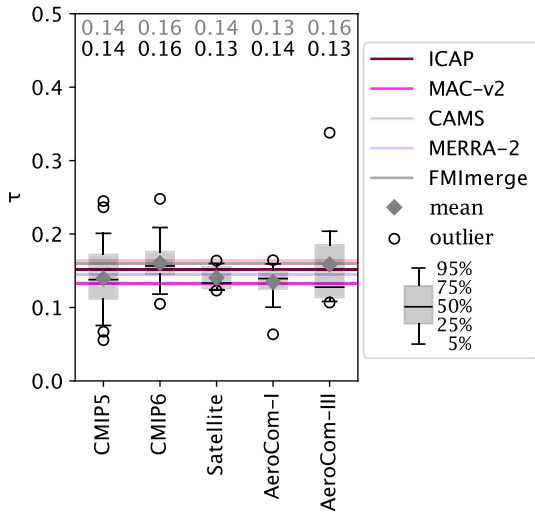

Figure 2. Spatial mean τ. Shown are the distributions of spatial mean τ in the aerosol-climate model intercomparison projects and across the satellite products. Mean values are indicated as gray diamonds and gray numbers at the top, medians are black horizontal lines and black numbers at the top, gray boxes are the 25%–75% percentiles, whiskers mark the 5%–95% percentiles, and circles are the extreme outliers. Color-coded lines are the spatial means from the reanalyses and the MAC-v2 climatology.

aerosol emissions over land, for example, fires and dust activity, explain the larger year-to-year standard deviation over land than over ocean in all data sets.

The land-to-ocean ratio in mean τ range from 1.25 to 2.23 (Table 3), again pointing to diversity in the spatial distribution of aerosols. Ratios larger than 1.8 are found for AeroCom-I, MAC-v2, AASTR and CALIPSO, with AATSR being the upper limit (compare Figure 1b). In contrast, ratios smaller than 1.4 are found for AeroCom-III, CMIP6 and MODIS, reflecting the fact that they have more τ over ocean than other data. The land-to-ocean ratios decrease as we go from AeroCom-I to AeroCom-III and from CMIP5 to CMIP6 and are therefore at the lower end of the satellite range. This model behavior may be due to more advection of aerosol-laden air from land to ocean, for example, dust-aerosol transport from North Africa, or due to stronger local emissions from sea spray, that we assess in Section 3.2.

3.2. Spatial Patterns

3.2.1. Summary Statistics

We measure the differences in spatial patterns of temporal mean τ across the data sets with the spatial correlation coefficients (r) separated by land and ocean only. We find higher r among different model intercomparison projects and among different satellite products, than between those two data

Table 3
Statistics Over Land and Ocean

	Land		Ocean		Ratio
Data	$\bar{\tau} \pm \sigma_{ens}$	σ_{time} [10^{-3}]	$\bar{\tau} \pm \sigma_{ens}$	σ_{time} [10^{-3}]	$\bar{\tau}_{land}/\bar{\tau}_{ocean}$
AerCom-I[a]	0.202 ± 0.045 (22%)	/	0.110 ± 0.025 (23%)	/	1.84
AerCom-III[a]	0.186 ± 0.057 (31%)	/	0.149 ± 0.079 (53%)	/	1.25
CMIP5[a]	0.197 ± 0.069 (35%)	0.53 (0.3%)	0.125 ± 0.043 (34%)	0.52 (0.4%)	1.58
CMIP6	0.202 ± 0.046 (23%)	4.14 (2.0%)	0.145 ± 0.030 (21%)	1.79 (1.2%)	1.39
CAMS	0.216	5.57 (2.6%)	0.145	3.51 (2.4%)	1.50
MERRA-2	0.202	11.53 (5.7%)	0.125	4.57 (3.7%)	1.62
MAC-v2[a]	0.198	/	0.109	/	1.81
ICAP[a]	0.212	/	0.130	/	1.63
Model mean	**0.200 ± 0.012** (6%)	**5.44** (2.7%)	**0.130 ± 0.016** (12%)	**2.60** (2.0%)	**1.58**
FMImerge	0.215	9.35 (4.3%)	0.140	4.69 (3.4%)	1.54
AATSR	0.257	4.74 (1.8%)	0.115	3.80 (3.3%)	2.23
CALIPSO	0.187	5.98 (3.2%)	0.101	1.98 (2.0%)	1.85
MISR	0.171	4.67 (2.7%)	0.107	3.52 (3.3%)	1.60
MODIS	0.190	11.69 (6.2%)	0.149	10.34 (6.9%)	1.28
POLDER	0.187	6.90 (3.7%)	0.106	3.74 (3.5%)	1.76
SeaWiFS	0.193	5.53 (2.9%)	0.112	3.39 (3.0%)	1.72
SLSTR[a]	0.232	1.22 (0.5%)	0.139	0.17 (0.1%)	1.67
Satellite mean	**0.202 ± 0.030** (15%)	**5.82** (2.8%)	**0.118 ± 0.018** (15%)	**3.85** (3.3%)	**1.73**

Note. Listed are the mean $\bar{\tau}$ and standard deviation σ_{time} for τ for the complete available time period separated into regions over land and over ocean. Conventions as in Table 2. The land-to-ocean ratio ($\bar{\tau}_{land}/\bar{\tau}_{ocean}$) of model and satellite mean is the arithmetic mean of the land-to-ocean ratio from the individual products.

[a]Data not available in the overlapping period.

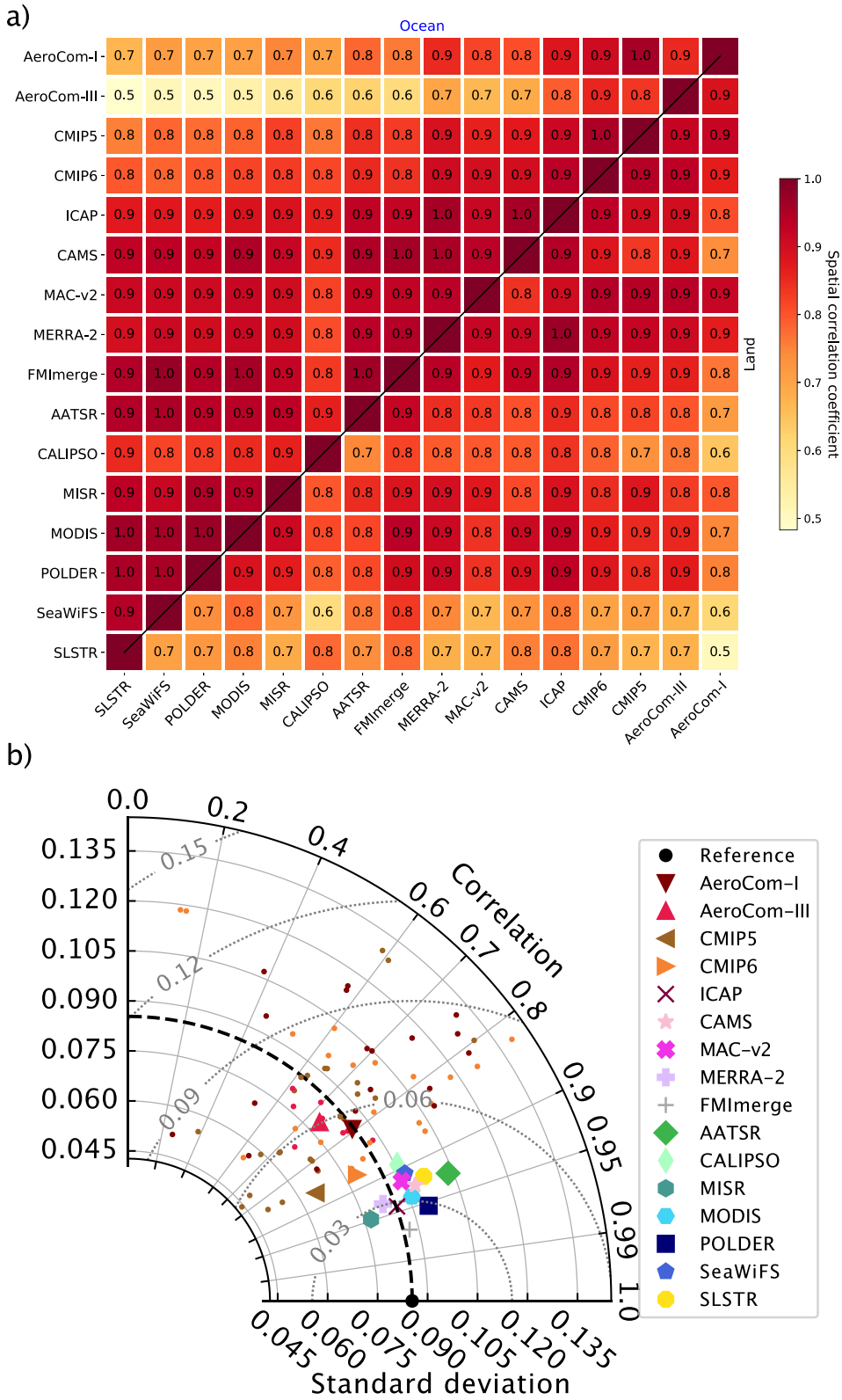
a)
Ocean
Land
AeroCom-I
AeroCom-III
CMIP5
CMIP6
ICAP
CAMS
MAC-v2
MERRA-2
FMImerge
AATSR
CALIPSO
MISR
MODIS
POLDER
SeaWiFS
SLSTR
Spatial correlation coefficient
b)
Correlation
Standard deviation
Reference
AeroCom-I
AeroCom-III
CMIP5
CMIP6
ICAP
CAMS
MAC-v2
MERRA-2
FMImerge
AATSR
CALIPSO
MISR
MODIS
POLDER
SeaWiFS
SLSTR

Figure 3.

sources over ocean and over land (Figure 3a). Take for instance AeroCom-III, which has a high $r = 0.9$ compared to AeroCom-I, but weaker correlation with satellite products. This points to no improvement in the mean spatial pattern across AeroCom phases. Compared to AeroCom-I and AeroCom-III, both CMIP5 and CMIP6 ensemble means have a higher r to satellite products and reanalyses. CMIP5 and CMIP6 have similar r to all other products, pointing to little changes across these CMIP phases. Both have lower r to satellite products than satellite products between each other, pointing to spatial differences. Satellite products and reanalyses tend to higher spatial standard deviations than the satellite mean, whereas multi-model ensemble means tend to lower spatial standard deviations (Figure 3b). The root mean square errors for individual satellite products and reanalyses against the satellite mean range from 0.031 to 0.044, which is smaller than for the multi-model means of CMIP and AeroCom. ICAP has a similar correlation as the satellite products consistent with the assimilation of MODIS data. Being a combination of satellite products, the FMImerge has the lowest root mean square deviation from the satellite mean across all datasets.

Spatial statistics of individual models with respect to the satellite mean differ substantially (Figure 3b). While correlations between the multi-model means and the satellite mean are typically 0.6–0.85, $r \simeq 0.2$ are found for some individual models in CMIP5 and CMIP6. Spatial standard deviations range from 0.045 to 0.140 for all data sets assessed. Compared to the spatial standard deviation from the satellite mean of 0.085, there are two extreme high outliers with 0.17 and 0.25 from the AeroCom-III ensemble (not shown). Spatial statistics of the MAC-v2 climatology and the reanalyses fall within the range of satellite products (Figure 3b). This is consistent with MACv2 being constrained by AERONET data, which are also the reference for satellite products. The reanalyses assimilate satellite data, such that they have similar summary statistics. The spatial statistics of MERRA-2 reanalyses are most similar to MODIS and MISR observations, due to their assimilation (Figures 3a and 3b). For the same reason, the spatial statistics for τ in CAMS are close to MODIS and AATSR, with the spatio-temporal statistics being closer to MODIS (compare also Figure 1 and Table 2). Note that assimilation of satellite products does not necessarily align the reanalysis fields with the ones of the assimilated satellites, as the are also rely on the physical forecast model as well as the assimilation setup.

We identify higher spatial correlations (r) over ocean than over land for all data sets except for AeroCom-III (Figure 3a). This indicates a larger similarity of the spatial patterns in models and satellites over ocean, although there is very low τ over some ocean regions. For instance MISR, MODIS, SLSTR, and POLDER have $r \geq 0.9$ to each other over ocean, but $r \leq 0.9$ to each other over land. This indicates generally higher uncertainty for τ from satellite retrievals over land than over ocean, consistent with earlier findings for MODIS and POLDER data (Chen et al., 2020). The AeroCom-III ensemble mean shows comparable r values over land to other data, but $r \leq 0.7$ against the satellite and reanalysis products over ocean. Although reduced correlations over the ocean may be induced by partly very small absolute values of τ, the results indicate a poorer representation of the observed spatial patterns over ocean in AeroCom-III than in AeroCom-I. This is not surprising, given few ground-based aerosol data over remote ocean regions that are needed to validate the models.

3.2.2. Hemispheric Asymmetry

We use the hemispheric asymmetries (A) to characterize the zonal distribution of aerosols. We find positive A at most geographical latitudes in most data sets (Figure 4), reflecting the fact that there is more zonal mean τ on the northern compared to southern hemisphere. Maximum A is found around 20°, which is largely explained by desert-dust aerosols from North Africa and Asia (also Figure S1 in the Supporting Information S1). A secondary weaker maximum in A is found around 40° geographical latitude, which is related to anthropogenic aerosol from anthropogenic sources in North America, Europe and eastern Asia. A converges toward zero at 0° and 60° geographical latitude. There are larger A over land than over ocean (Figure S2 in the Supporting Information S1), consistent with the hemispheric difference in land area where large amounts of aerosols are emitted.

Figure 3. Summary statistics for mean pattern of τ. Shown are (a) the spatial correlation coefficients of annual mean τ of the data sets over ocean regions (upper left triangle) and over land (lower right triangle), and (b) the Taylor diagram for the annual mean τ from individual data sets against the satellite mean. The Taylor diagram (Taylor, 2001) shows spatial standard deviations, correlation coefficients, and root mean square differences against the satellite mean as reference. Small points mark individual model results in the model intercomparison projects. Two AeroCom-III members fall outside of the displayed values (standard deviation of 0.17 and 0.25, and correlation of 0.47 and 0.11).

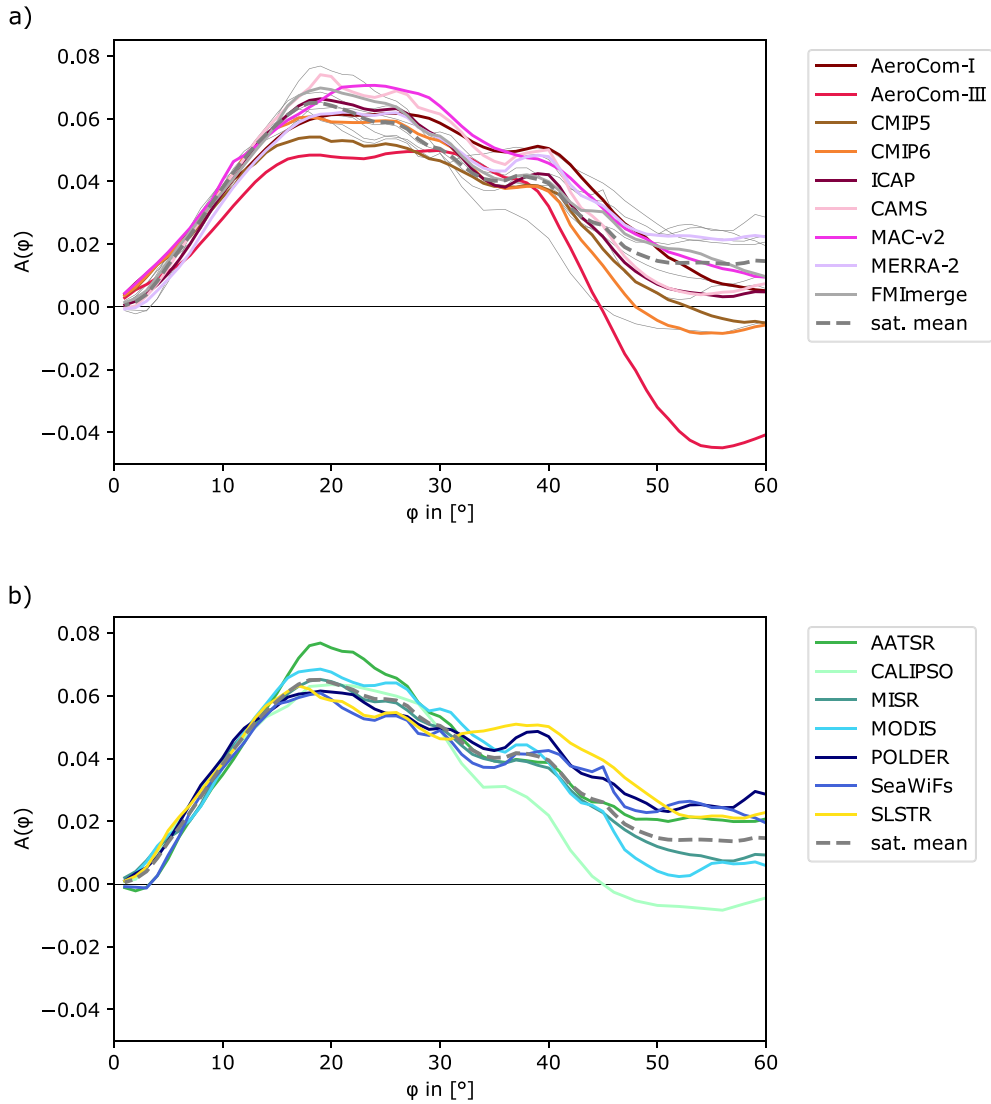

Figure 4. Hemispheric asymmetry in τ. Shown is A (Equation 1) for the color-coded (a) model data and (b) satellite products. For an easier comparison to the satellite range, the individual satellite products are also marked as thin gray lines in a. Sat. mean refers to the mean across all satellite products, excluding FMImerge.

Differences in A between the model means and the satellite data increase from low to high latitudes (Figure 4). Satellite retrievals at high latitudes are affected by difficulties to retrieve τ over bright snow-covered surfaces and for low sun elevation angles, in addition to a relatively sparse ground-based observation network for validation. It explains the large differences in A across satellite products, particularly over land (Figure S2 in the Supporting Information S1). For example, AASTR has the largest A between 15 and 25° and the smallest A between 35 and 50° over land, but is close to the satellite mean over ocean. At least in parts due to the observational uncertainty, we also see a considerably spread in A of multi-model means, particularly poleward of 40° geographical latitude. This is primarily explained by model biases in τ over the Southern Ocean, found at most geographical longitudes (Figures S2a and S2b in the Supporting Information S1, compare Figure 5). It suggest excessive emissions of aerosols from sea-spray linking for instance to the simulated near-surface winds. This behavior is particularly pronounced in the AeroCom-III mean followed by the CMIP6 mean, so that A becomes negative at high latitudes in contrast to all satellite products indicating positive A at all geographical latitudes except CALIPSO (Figure 4). For CALIPSO, larger τ over the Southern Ocean paired with a tendency to lower τ at the same latitudes in the northern hemisphere relative to the satellite mean explains the negative A over the ocean (Figures S2c and S2d in the Supporting Information S1, compare Figure 6). Some of the different behavior of CALIPSO might be due to

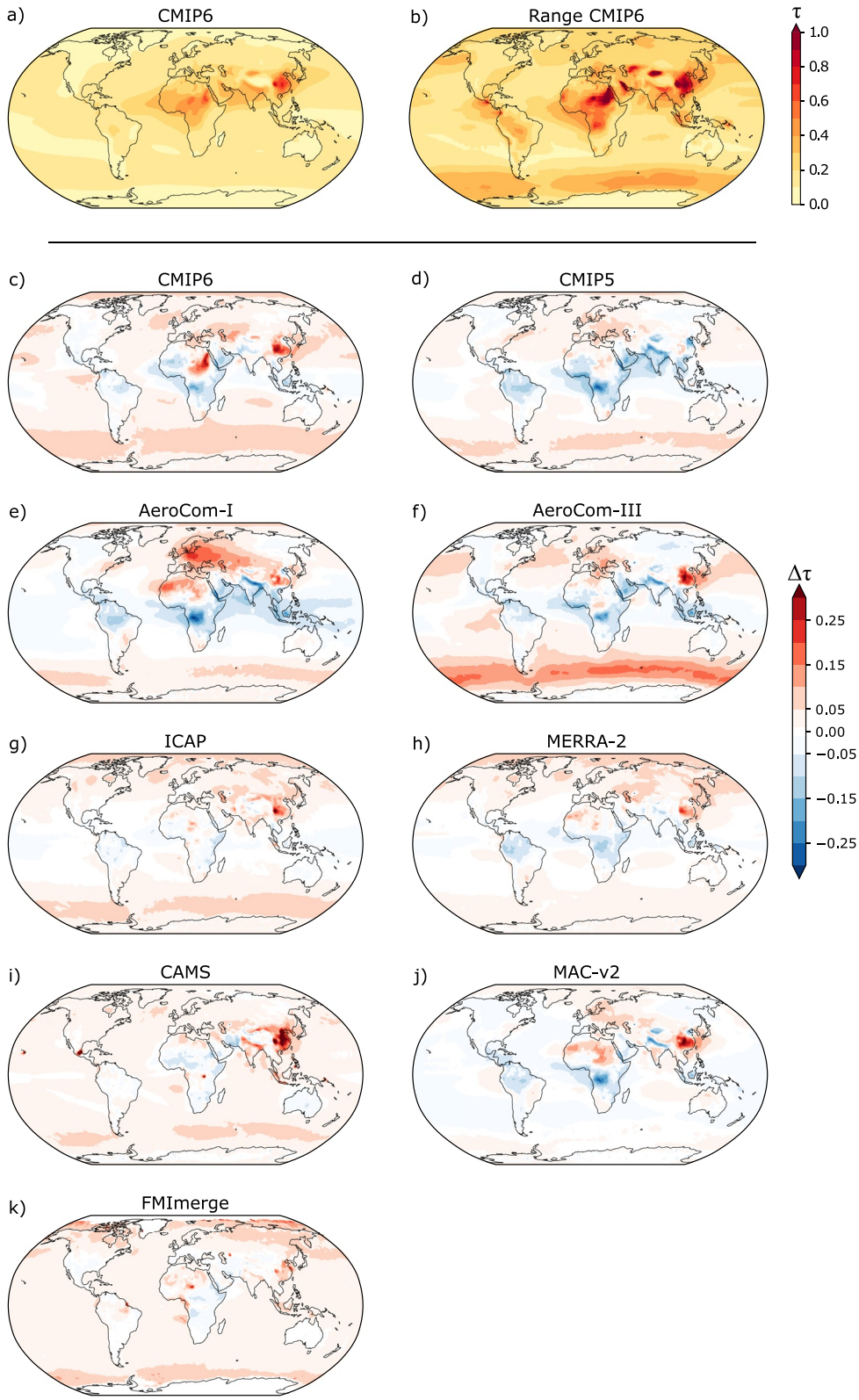

Figure 5. Simulated spatial pattern of τ. Shown are (a) the temporal mean pattern of τ averaged across all CMIP6 simulations, (b) the range in the mean patterns of τ in the CMIP6 model ensemble, and (c–k) the difference of model means against the satellite mean (compare against Figure 6a). Values in (c–f) are based on the ensemble mean of the aerosol-climate model intercomparison project. White shading marks differences of ±0.01.

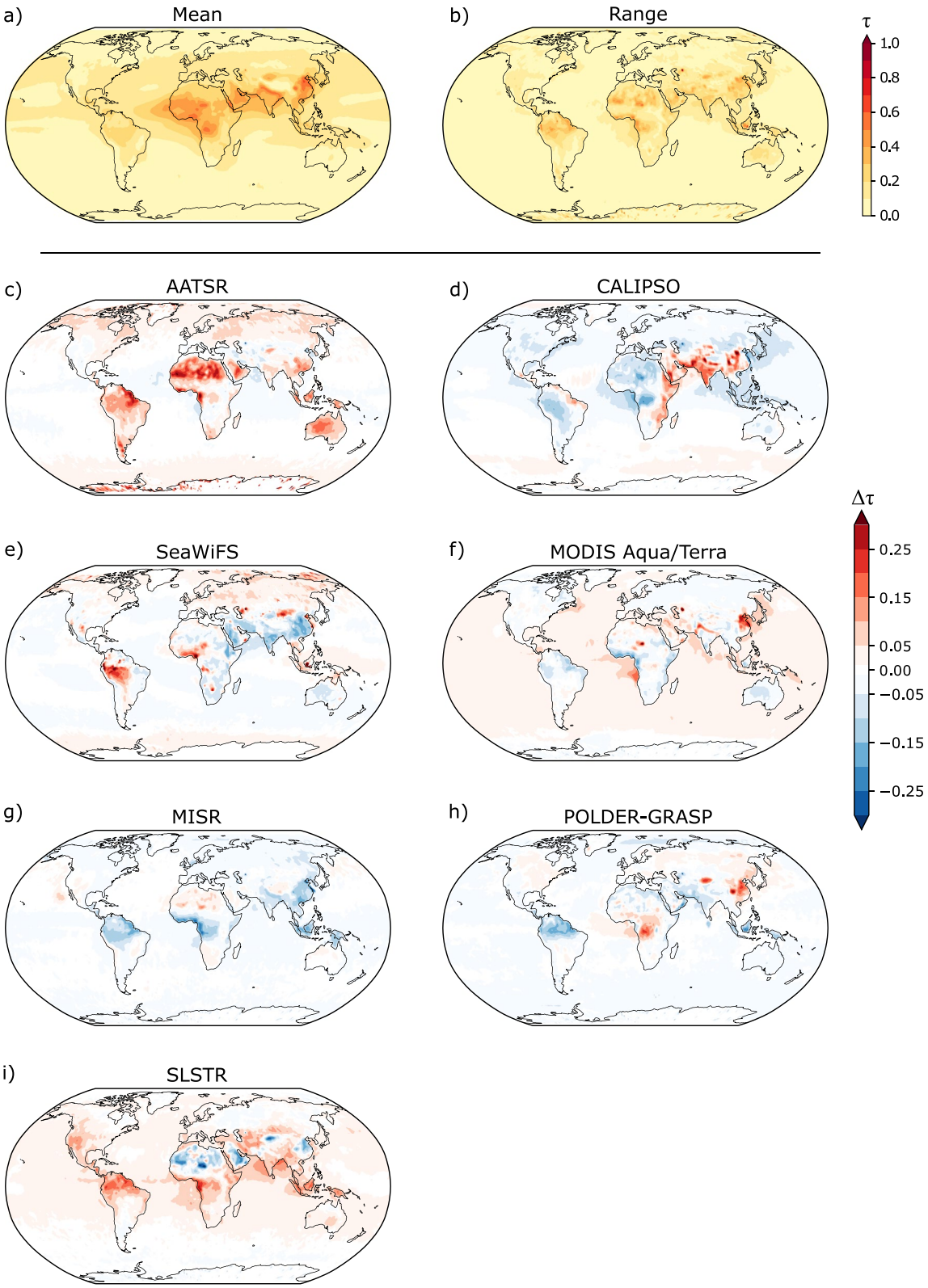

Figure 6. Observed spatial pattern of τ. Shown are (a) the temporal mean pattern of τ averaged across all satellite products, (b) the range in the mean patterns of τ across the satellite products, and (c–i) the difference of individual satellite products compared to the satellite mean (compare against (a)). White shading marks differences of ± 0.01.

the different measurement technique compared to passive sensors used in the other satellite products for τ. CALIPSO uses active remote sensing with a lidar beam in a narrow viewing-swath and requires a different algorithm to obtain τ than passive sensors. An advantage is that CALIPSO allows a retrieval of τ above clouds, even if thin cirrus occurs. This is not possible for retrievals from passive sensors that are restricted to cloud-free conditions. Large τ over the Southern Ocean and the negative A at high latitudes for both CMIP phases could be evaluated as correct if they were only compared to CALIPSO observations. It is another example for the need of using more than a single observational data set to evaluate models. The diversity in A across the data sets is explained by differences in the spatial patterns of τ, assessed next.

3.2.3. Regional Distribution

Although relative differences between satellite products are the same for land and ocean (compare Section 3.1), regional absolute differences are larger over land (Figure 5). This is due to difficulties in retrieving aerosol information over heterogeneous and brighter land surfaces (e.g., Chen et al., 2020; Schutgens et al., 2020). Over land regions, we find deviations of individual satellite estimates from the satellite mean with varying sign, magnitude and spatial extension (Figure 6). For instance, AATSR has a tendency to larger regional means in τ compared to the satellite mean, consistent with the largest maximum in A with 0.07 near 20° geographical latitude among the satellite products (compare Figures 4 and 6). The locations point to more desert-dust aerosols and biomass burning aerosols than in other data. Such behavior is in contrast to the tendency toward slightly smaller values relative to the satellite mean for the trusted MISR data in many regions except in North Africa. MISR performs well against AERONET observations (Kahn et al., 2005). The small differences in τ between MISR and the satellite mean over most regions suggests that also a simple satellite mean across several products can be useful as a reference for the evaluation of other data sets. CALIPSO for instance exceeds the satellite mean of τ in parts of eastern Africa, India, and southern China, and tends to lower τ elsewhere, for example, in South-America and North-Africa. In contrast, lower values than in the satellite mean are indicated by SeaWiFS in West-Africa, India, and East-China. Interestingly, none of the multi-model means have larger τ than the satellite mean over South-America as it is seen by AATSR, SeaWiFS, and SLSTR (Figures 5 and 6). MODIS and POLDER also indicate higher τ than the satellite mean in regions of southern Asia and Africa. For MODIS, the larger τ in these areas paired with the widespread larger τ over all ocean regions are responsible for the maximum in spatial mean τ across the satellite data sets (compare Section 3.1).

Reanalyses and the ICAP operational ensemble are typically closer to satellite data than the aerosol-climate model means. This is because they use satellite data in their assimilation systems to produce the reanalyses. For instance, both reanalyses match the satellite mean over India, while CAMS fits the satellite mean τ better over South-America, South-Africa and Southeast-Asia. The MAC-v2 climatology typically falls within the uncertainty of satellite products except for South-Africa. While spatial patterns of differences to the satellite mean are often similar to AeroCom-I, the magnitude of the differences is through the use of AERONET data typically smaller for MAC-v2, except in eastern China (Figures 5 and 7). FMImerge tends to higher τ over oceans than the satellite mean, consistent with the relatively high global mean of the merged product. Regional τ across individual satellite products differ by less than 0.1 for most ocean regions, which is consistent with the high spatial correlation among the satellite products over ocean (Section 3.1).

All models show differences in regional maxima in τ relative to the satellite means without a clear overall improvement across the phases of the model intercomparison projects (Figure 5). The largest regional maxima in τ over East-China, and also India and North Africa are found for all models, but with different magnitudes. For instance, the multi-model ensemble mean of CMIP5 underestimates τ in India, Southeast-Asia, and also South-Africa compared to the satellite mean and reanalysis data (Figure 7). In these regions, the ensemble mean and median of CMIP6 is closer to the satellite observations than CMIP5, indicating an improvement across the CMIP phases here. All multi-model ensemble means simulate lower τ than the satellite mean over the Kongo area in central Africa. It is difficult to judge whether this is due to the models or the satellite, since there is no AEROCOM station which could constrain τ in this area. Although the range of CMIP6 is large in this area, the ensemble-mean bias against the satellite mean is reduced in the more recent phases of CMIP and AeroCom. An example for a regional improvement in τ from models is Europe, were CMIP5 and AerCom-I show a positive bias of up to 0.2 (Figure 5). For AeroCom-I, this regional difference remains when comparing only data of 2000, indicating that the regional bias in τ is not due to an anomalous year or a trend. AeroCom-III and CMIP6 reduce the bias over Europe indicating a regional improvement across model intercomparison project phases. Also the

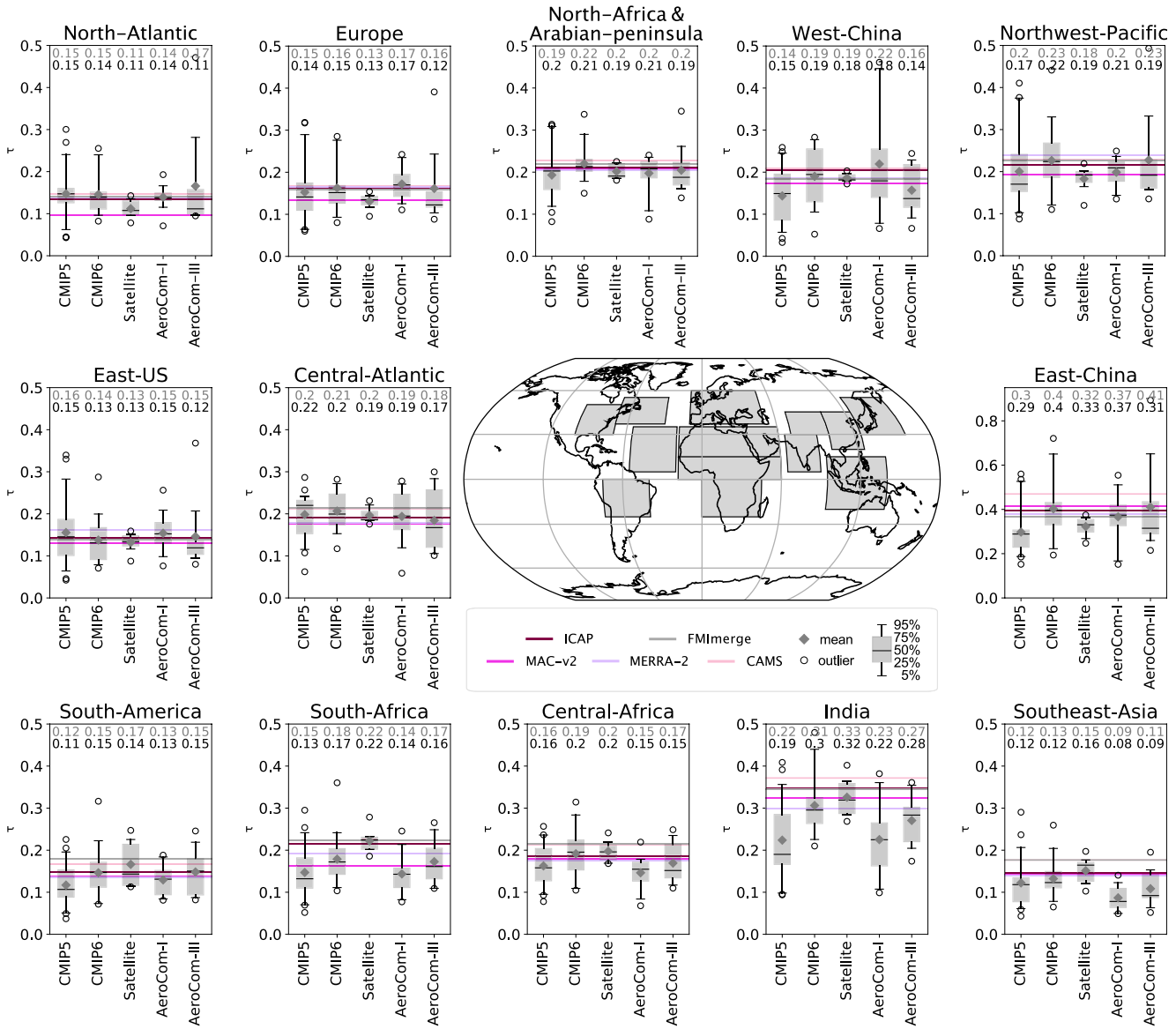

Figure 7. As Figure 2, but for regional means of areas marked as gray boxes in the map. The selection of regions is taken from Koffi et al. (2012) which is based on Yu et al. (2010). Note the different scale of the *y*-axis for East-China.

bias in AerCom-I over central Asia and North Africa is reduced in AerCom-III, seen in regional averages of τ (Figure 7).

Other regional differences newly occurred or remained as we go from the previous to the current model intercomparison phases. One example is East-China, where the means of CMIP6 and AeroCom-III are not better than CMIP5 and AeroCom-I compared to the satellite mean. CMIP5 for instance did not overestimate mean τ over eastern China, but all other multi-model means do (CMIP6, AeroCom-I, AeroCom-III). Another example of no improvement across intercomparison phases is the increased τ in AerCom-III and CMIP6 over the extra-tropical storm-track regions in the North-Atlantic, North-Pacific, and particularly the Southern Ocean, pointing to excessive emissions of aerosols from sea spray in the model simulations. In the Southern Ocean, the AerCom-III ensemble mean is larger by up to 0.2 compared to the satellite mean, consistent with particularly negative A at high latitudes (Figures 4 and S1 in the Supporting Information S1). The summary statistics indicate that this behavior might be explained by a few participating models that should be addressed in future work (compare Figure 3). Such regional biases in τ have an implication for the degree to which aerosols influence the regional climate.

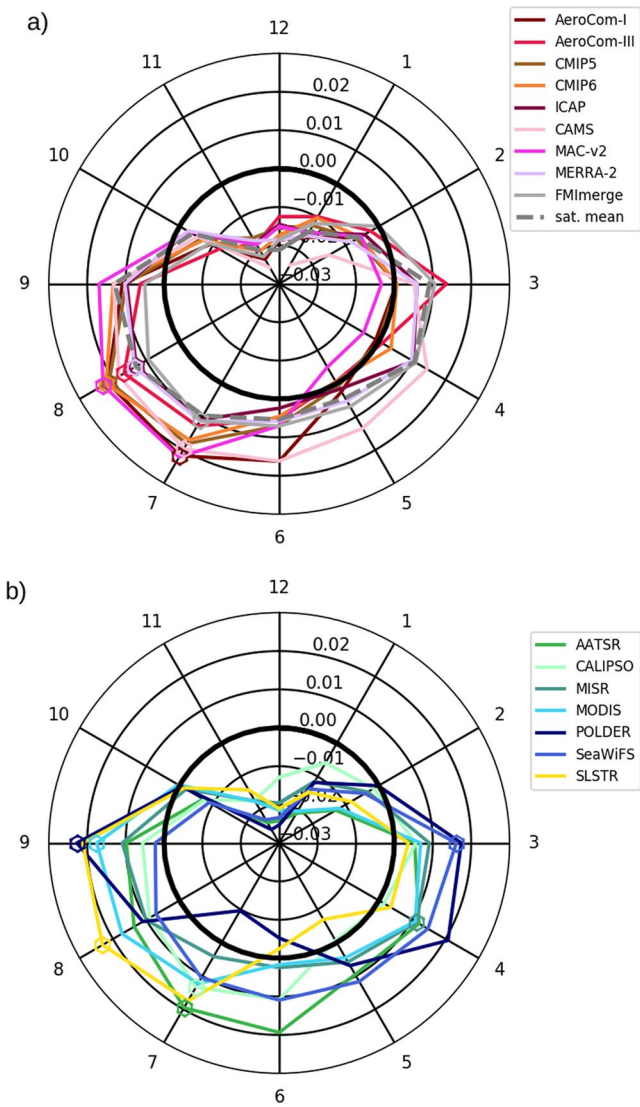

Figure 8. Seasonal cycle of τ. Shown are monthly anomalies in the spatial mean τ from (a) models and (b) satellites. Anomalies are calculated as the difference in monthly means against the annual mean of the data set. Circles mark the month of the maximum. "Sat. mean" is arithmetic mean of the satellite products.

Individual models show a large spread in regionally averaged τ that can fall outside of the range of satellite means (Figure 7). We find no systematic reduction in the model spreads for regional means of τ as we go from CMIP5 to CMIP6 and from AeroCom-I to AeroCom-III. The magnitude of the range of the CMIP6 models exceeds for instance the magnitude of the ensemble mean at most locations and is larger than the range in satellite estimates (Figures 5a, 5b, and 6b). Each aerosol-climate model ensemble has minimal ensemble spread in different regions. While the spread of AeroCom-I is particular small over the North Atlantic, it is smallest over East US for AeroCom-III, over East China for CMIP5, and over North Africa for CMIP6. The range across individual models in CMIP6 is particularly large over regions in South America, Africa, Asia, and also the Southern Ocean. Studies toward better representing regional patterns of τ in aerosol-climate models, and therefore improving our capacity to simulate regional climate change, should focus on these regions. In some of these regions, the uncertainty in the observational data sets of τ is also particularly large pointing to the need of better constraining regional τ with ground-based reference data (Figure 6). This might be in parts due to year-to-year variability in natural aerosol emissions from deserts and biomass burning, posing a challenge for constraining regional τ.

3.3. Seasonal Cycle

3.3.1. Spatial Mean

The monthly contributions to spatial annual mean τ depend on the time of the year and have a surprisingly large diversity across the satellite products (Figure 8). The largest monthly contributions to the spatial mean τ occurs between July and September in most satellite and model data, coinciding for instance with the biomass burning season. A striking feature is the larger diversity in the timing of the monthly maximum across satellite products than across model-ensemble means. It reflects more variability in biomass burning emissions and dust storms across individual satellite products than is represented in multi-model means. Two satellite products even have a much earlier maximum in the monthly contribution to annual mean τ, namely in March and April (Figure 8b). This absolute maximum in northern hemispheric spring for SeaWiFS and MISR is often identified by other data as a secondary maximum with varying magnitude. Interestingly, the timing of the maximum during northern hemispheric spring in SeaWiFS and MISR is due to reduced τ between July and September, rather than an increased τ in March or April, relative to other satellite data. For MISR, the spatial pattern of differences in τ suggests that this behavior might be due to less τ associated with biomass burning over tropical land (compare Figure 6). SeaWiFS has the lowest τ over the Arabian Peninsula, India and East Asia across all satellites, pointing to differences in desert-dust and anthropogenic aerosols (compare Figure 6). The smallest contributions to the spatial mean τ is found for November and December across all models and satellite products. This is consistent with reduced zonal mean τ at all latitudes compared to other months visible for both model-ensemble means and satellite products (compare Figures 9a and 10a).

The differences in the temporal occurrence of maxima is not exclusively driven by differences in regional maxima in τ. This is seen in the diversity of zonal means in τ per month across the data sets (Figures 9 and 10). Maxima in the zonal mean of around $\tau \simeq 0.3$ are found between 10°N and 20°N during northern hemispheric summer (Figure 10). During other months, the zonal mean τ in the northern hemispheric tropics also dominate the pattern, but is less pronounced compared to other latitudes (also Figures S3 and S4 in the Supporting Information S1). Individual satellite and model-ensemble means are often qualitatively similar to the satellite mean, but there are marked quantitative differences and diversity in representing details of the patterns, for example, the location and

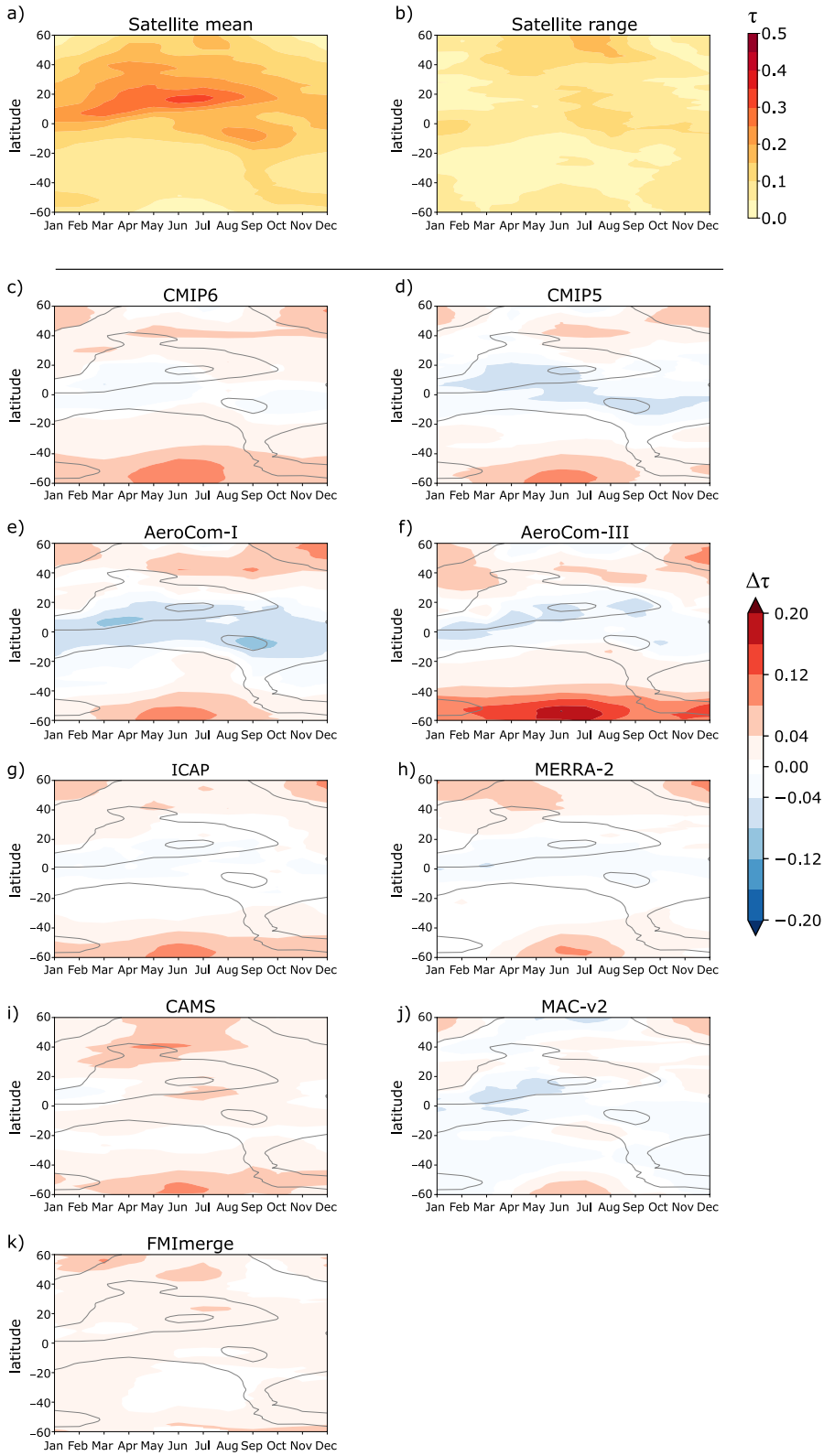

Figure 9. Simulated seasonal cycle of zonal means in τ. Shown are the (a) zonal means for each month averaged over all CMIP6 models, (b) the range in the zonal means per month in the CMIP6 model ensemble, and (c–k) the difference of model means compared to the satellite mean (gray isolines, compare against Figure 10a). Values in (c–f) are based on ensemble averages of the aerosol-climate model intercomparison projects. Absolute values for (c–i) are shown in Figure S4 in the Supporting Information S1.

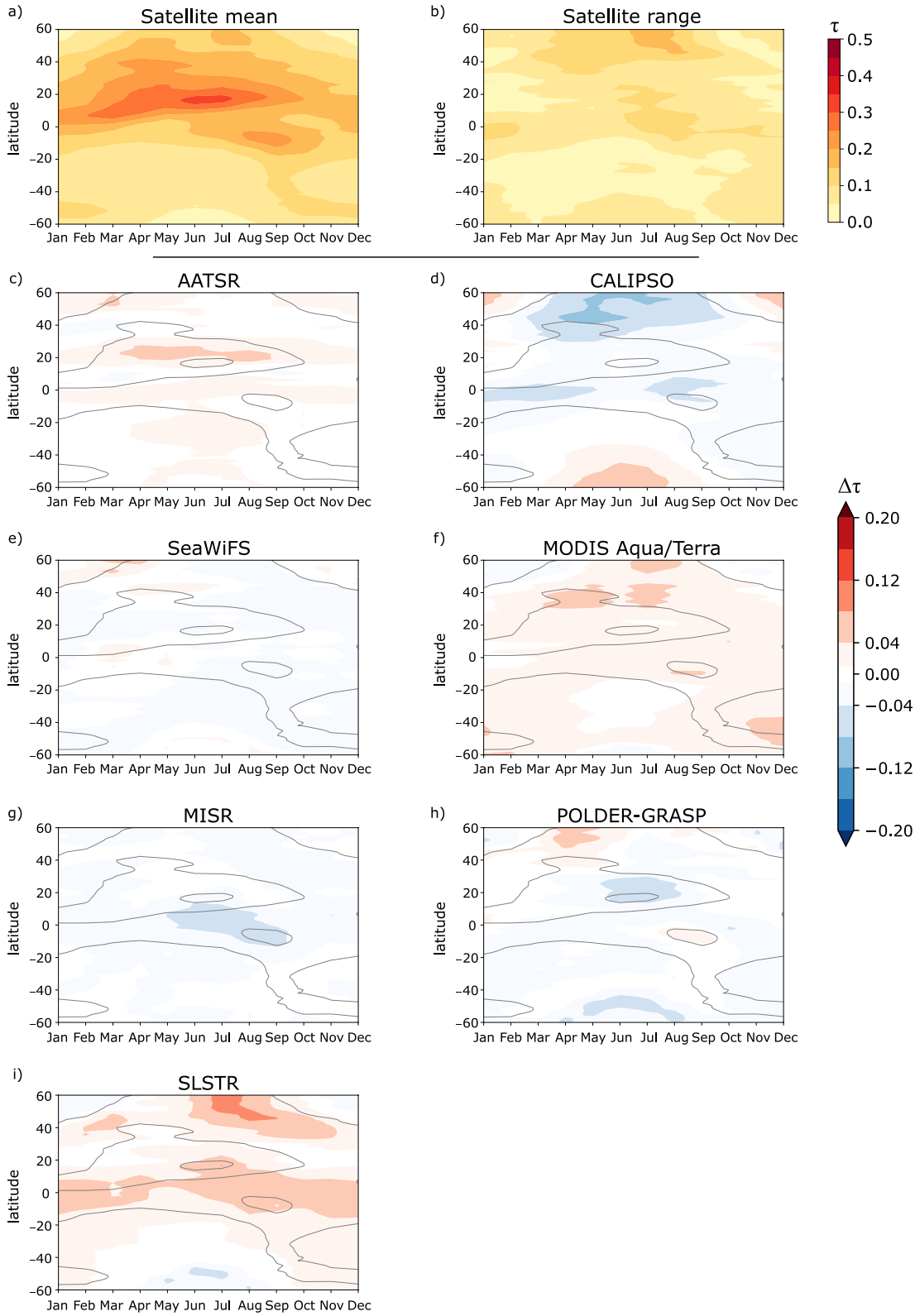

Figure 10. Observed seasonal cycle of zonal means in τ. Shown are the (a) zonal means for each month averaged across all satellite products, (b) the range in the zonal means across the satellite products, and (c–i) the difference of individual observational estimates compared to the satellite mean (gray isolines, compare against a). Absolute values for (c–i) are shown in Figure S5 in the Supporting Information S1, and differences between the satellite products and CMIP6 in Figure S8 in the Supporting Information S1.

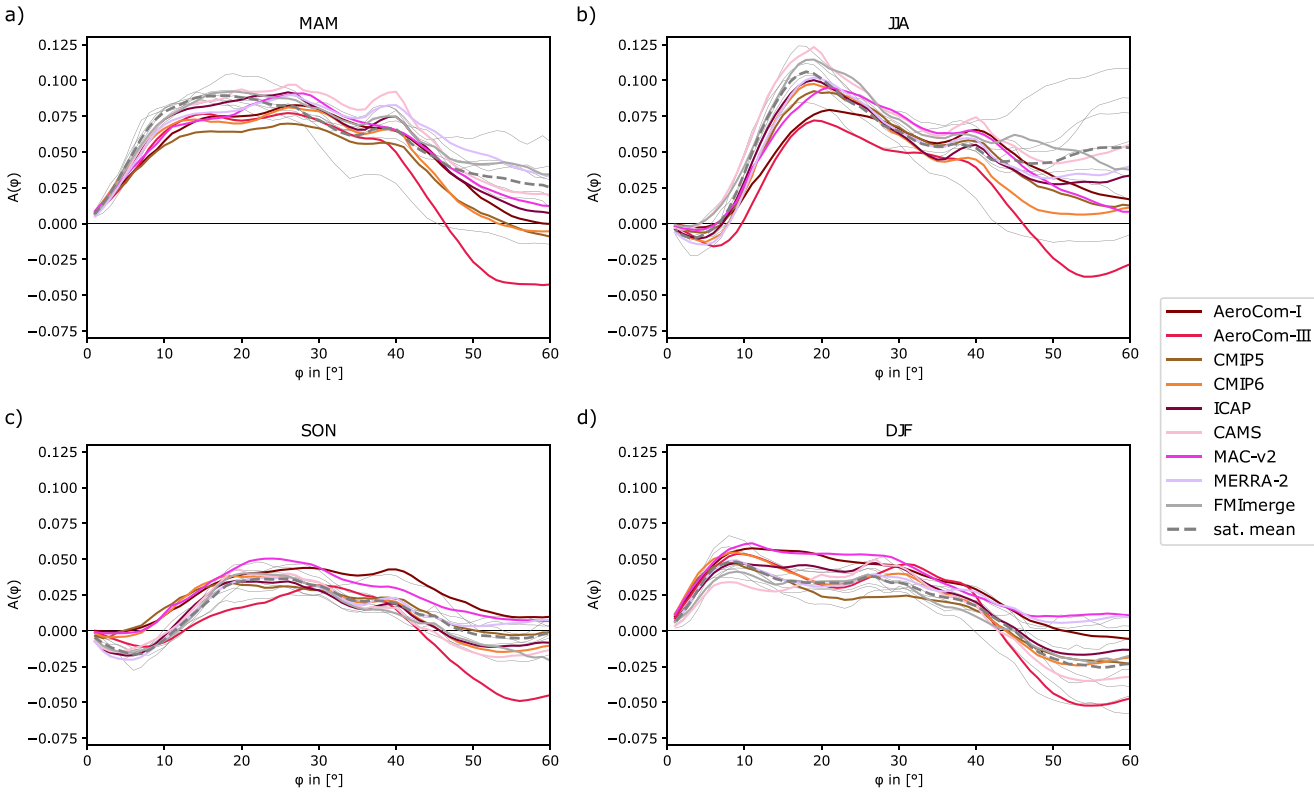

Figure 11. Seasonal hemispheric asymmetry of τ. As Figure 4a, but for the seasonal means of (a) March–May, (b) June–August, (c) September–November, and (d) December–February. Individual satellite products are marked in Figure S7 in the Supporting Information S1. "Sat. mean" is arithmetic mean of the satellite products.

spatial extent of month-to-month changes (Figure 9, also Figure S5 in the Supporting Information S1). Take for instance the tendency to underestimate (overestimate) zonal means in τ around the equator (at higher latitudes) seen for all model-ensemble means and MERRA-2 (Figures 9c–9g). Such opposing regional biases can compensate in the spatial mean, for example, for MERRA-2 that is closest to the spatial mean in the monthly contributions from the satellite mean (compare Figure 8a), but this is not the case for all data sets.

3.3.2. Zonal Distribution

We analyze the zonal means per month to assess the differences in the monthly contributions to the spatial mean τ (Figures 9 and 10). The identified secondary maximum of spatial mean τ in March to April is explained by increased τ in the northern hemisphere (compare Figure 8). Zonal mean τ from the satellite mean during northern hemispheric spring is largest between 0–20°N (Figure 10a). Additionally, τ at 20–50°N is increased around April compared to other months. This secondary maximum can be identified in many data sets, including the reanalyses as well as in the AerCom-III and CMIP6 means (Figure S5 in the Supporting Information S1). At the same time, the same zones in the southern hemisphere do not show substantial changes in the first half of the year. Taken together, the zonal distribution cause a relatively broad maximum in the hemispheric asymmetries ranging from 10° to 40° geographical latitude for March–May (Figures 11a).

Individual satellites indicate the secondary maximum in regional τ in the northern hemisphere in April, also seen in the monthly contribution to the spatial mean τ (Figure 8a), but with large differences across the satellites. The satellite range here exceeds the one from most other regions and months. One exception, when the satellite uncertainty is larger, is July and August in the northern hemispheric high-latitudes (Figure 10b). The zonal mean at 40°N in April differs by more than $\tau = 0.1$, with data from CALIPSO and MODIS being the furthest apart from each other (Figures 10 and S6 in the Supporting Information S1). The strong secondary maximum in April by MODIS is also seen as too large zonal means in τ and A in CAMS and MERRA-2, due to the assimilation of MODIS data. The overestimation of τ in CAMS compared to the satellite mean and the FMImerge is, however,

strongest in the storm tracks paired with a broad tendency of too much τ (Figure 9), explaining the relatively large spatial mean τ in CAMS, compared to the satellite mean and other reanalysis data.

The typically largest monthly contributions to the spatial mean τ between June and November is largely explained by increased τ in the tropics, that is, equatorwards of 20° geographical latitude. Particularly large τ is found for June to August at 0–20°N and for August to October additionally at 0–20°S (Figures 9 and 10), primarily over South-America and Central-Africa (compare Section 3.2). Here, the satellite mean shows zonal mean τ of up to 0.25. This pattern is explained by the seasonal biomass burning in Africa, South America, and Southeast Asia paired with desert-dust aerosols from North Africa.

The southern tropical maximum in τ is qualitatively well represented by the model-ensemble means, for example, indicated by the CMIP6 mean. The magnitudes thereof, however, again substantially differ across the model data sets. Here, the patterns of increased regional τ in CMIP5 are similar to reduced patterns of CALIPSO and MISR compared to the satellite mean, highlighting again the need for using several data sets for a model evaluation. In contrast, MODIS and SLSTR show higher peak values in the southern tropical maximum. The southern tropical maximum in τ seen from June to November is consistent with the negative A near the equator for the seasonal means of northern hemispheric summer and autumn, in satellite products, reanalyses, and AeroCom-III (Figures 11b and 11c, and S7 in the Supporting Information S1). The North-South tropical dipole of regional τ between June and November is consistent with the seasonal mean A being larger in northern hemispheric summer than autumn.

We find no overall improvement in zonal mean τ from CMIP5 to CMIP6 and from AeroCom-I to AeroCom-III. The tropical biases in τ, which was present throughout the year in CMIP5 and AeroCom-I, reduced in CMIP6 and AeroCom-III, suggesting a mean model improvement for the tropics (Figure 9). However, the biases in the northern hemispheric extra-tropical storm track poleward of 40°N remained largely unchanged paired with increased biases in the Southern Ocean in both CMIP6 and AeroCom-III.

The model spread for τ in the northern hemispheric extra-tropics is only partly consistent with a larger uncertainty in satellite retrieved τ than for other regions with a better observational constraint. The spatial patterns indicate that differences between individual models considerably increase poleward of 40° (compare Section 3.2). The larger model means in τ compared to the satellite mean is evident throughout the entire year, but stronger in the southern than the northern hemisphere poleward of 40° (Figure 9). On average, the models have larger τ than the satellite mean during winter than summer of the hemisphere. This implies that the seasonal variation as seen in the satellite mean is smaller at high latitudes for all multi-model means (Figure 10a), for example, at latitudes north of 50°N. Interestingly also the MERRA-2 reanalysis, CALIPSO, and MAC-v2 climatology underestimate the seasonal change at higher latitudes, which is consistent with smaller A compared to the satellite mean (Figure 11). In contrast, the CAMS reanalysis shows reduced seasonal variation in τ at high northern latitudes, primarily during northern hemispheric summer. This behavior compensates the larger southern hemispheric τ in A, resulting in a good agreement of A relative to the satellite mean (Figure 11). It indicates a seemingly correct representation of A in CAMS compared to the satellite mean, but it is due to compensating regional differences poleward of 40° geographical latitude.

The range of regional τ from individual CMIP6 simulations is larger than the CMIP6 multi-model mean of τ during most of the year and at most latitudes (Figure 9b). Particularly large diversity in regional τ is found in the extra-tropical storm tracks followed by the northern hemispheric tropics throughout the year. Differences in the magnitude of the CMIP6 models are particularly striking in comparison to the satellite data. For instance, the range in satellite estimates for regional τ does not exceed 0.2 and is therefore substantially smaller than the model range in the aforementioned regions and of comparable magnitude for 20–40°S (Figure 9). Single CMIP6 models show a large spread in the southern hemisphere poleward of 40°S, where the magnitude of the CMIP6 range exceeds the magnitude of both, the CMIP6 mean and the satellite range by more than a factor of two. While most differences of the CMIP6 mean to individual satellite products vary in sign and magnitude, all satellite products have lower τ poleward of 40°S with a common maximal deviation in southern hemispheric winter (Figure S8 in the Supporting Information S1). This points to considerable uncertainties in regional τ from individual CMIP6 simulations with implication for the regional radiation budget for most regions throughout the year.

4. Discussion

4.1. Observational Uncertainty

Retrievals of aerosol optical depth from satellite-based radiation measurements face several challenges. All aerosol products have to determine aerosol of different characteristics over partly very bright surfaces and mask aerosols beneath clouds, which are treated differently in the individual retrieval algorithms. For passive sensors, a heterogeneous land surface and the presence of clouds limit the number of quality-ensured retrievals. Cloudy conditions are also associated with higher humidity leading through aerosol hygroscopic growth to a larger τ compared to dry air (e.g., Chand et al., 2012). Satellite retrievals of τ from measurements of passive instruments are restricted to cloud-free conditions. That fact alone could imply an underestimation of τ from such products, but it is difficult to ascertain in light of other sources of differences including potential cloud contamination in pixels used for retrieving τ. This study builds on the work of many validating the used satellite products against observations to continuously improve the data quality. Nevertheless there are uncertainties, especially on short time scales and for products with a narrow viewing-swath (e.g., Colarco et al., 2014; Schutgens et al., 2020). Different overpass times of the satellites in addition to different temporal coverage paired with variability in aerosol emissions cause further sampling differences, which is problematic when looking at short temporal and spatial scales, for example, spanning days to a couple of years over a small region. It is therefore no surprise that past studies document considerable uncertainty for aerosol data from collocated satellite data against AERONET stations (e.g., Schutgens et al., 2020). Such data is nevertheless useful for model evaluation, for example, in the framework of AeroCom (Schutgens et al., 2021). One way could be to evaluate models against a merged satellite product such as FMImerge, although merging satellite data must not necessarily imply an unbiased product compared to AERONET measurements. We find that FMImerge has the best performance with respect to the root mean square deviation from individual satellite products, pointing to it's skillful combination of different observations into one product. We here use averaging over long time periods and many satellite products to compile a reference for our data intercomparison. This allows us to also calculate the range across satellite products and use it as an estimate of the observational uncertainty in the data intercomparison.

Our results indicate an observational uncertainty for the spatial mean τ of -11% to $+17\%$ about the satellite mean of 0.14 (60°N to 60°S). Despite the spatial restriction to regions equatorward of 60° geographical latitude in our work, our estimate is similar to the global range of 0.13–0.17 based on three satellite products Bellouin et al. (2020), but we have a slightly lower bound (0.124) due to MISR and CALIPSO included here. The satellite range for the spatial mean τ (0.04) is an order of magnitude larger than the spatial mean bias of satellite products against ground-based measurements of τ from AERONET (0.0024–0.0031, Kinne et al., 2013; Schutgens et al., 2020). The intercomparison also shows considerable differences in the satellite products for τ for land and ocean. At least in parts this is explained by diversity in satellite estimates of τ in regions without AERONET observations, for example, in deserts and over the ocean. We identify the largest uncertainty across the satellite data in the northern hemispheric extra-tropics between April and September, and in the tropics from July to August and from January to February, partly due to difficulties in retrieving aerosols in cloudy scenes. Here future field campaigns would be particularly useful to better constrain τ.

The results further underline the need for using several observational data products for model evaluation due to the uncertainty in observations. Past assessments often used single data sets for evaluating the model performance for τ. Such practices do not account for the measurement and retrieval uncertainty in the observational products, which is even larger at sub-monthly time scales (Schutgens et al., 2020). If the observational uncertainty is taken into account, the model evaluation is not biased toward a single product. Due to sampling differences across individual observational data sets, modeled τ should be compared to a larger set of individual observations. This can be especially critical for those regions and times when the uncertainty in the observational estimates of τ is large. This might be the case, when ground-based observations for validation are rare and when the surface is highly reflective, for example, over deserts, snow, and ice. Further difficulties in retrieving aerosols from satellite measurements arise in cloudy scenes and for low solar elevation angles. Using the quantified observational uncertainty indicates that the multi-model mean τ from contemporary intercomparison projects of AeroCom-III and CMIP6 are at the upper end of the range in satellite retrievals, with substantial regional and seasonal differences compared to the satellite mean. This implies in the mean a stronger influence of the modeled than the retrieved τ in calculations for the atmospheric radiation transfer, all else being equal. Our results further show that the regional τ from model means often fall outside of the range from satellite products, for example, over North Africa,

Southern Ocean and East Asia. Especially the former two regions point to natural aerosols from deserts and sea spray not being accurately represented. Despite the regional rather large observational uncertainty, the existing satellite data might here already be an opportunity to constrain the regional τ in modern aerosol-climate models.

4.2. Model Performance

Our results indicate no overall model improvement for representing all characteristics of τ over time, here measured by the differences between CMIP5 and CMIP6, and between AeroCom-I and AeroCom-III. Our assessment is based on metrics for a detailed investigation of model performances with respect to different spatio-temporal characteristics of τ. While indeed some improvement is seen, for example, for the spatial distributions of AerCom-III and CMIP6 over land compared to AerCom-I and CMIP5, the spatial mean τ did not improve in light of the satellite uncertainty. Also both CMIP5 and CMIP6 have broadly similar mean patterns for τ, but differ in their magnitudes that dependent on the region and season. Some of the differences in AeroCom-III might be explained by just a few models in the ensemble.

Also among reanalyses and climatologies, we find differences making an overall ranking of the best data sets for a present-day climatology of τ difficult, although these differences are typically smaller than for the aerosol-climate model means. For example, ICAP performs well in terms of the spatio-temporal mean τ over land compared to the satellite mean. This is due to the assimilation of satellite data into the different systems to produce the ICAP data. Further examples are the spatial mean seasonal cycle from the satellite mean that is best reproduced by MERRA-2 and the observed hemispheric asymmetries that is rather reproduced by the CAMS reanalysis. The latter is, however, explained by compensating differences poleward of 40° geographical latitude. This is at least in parts due to the assimilated data from MODIS and AATSR resulting in similar spatio-temporal characteristics for τ in CAMS. Consequently, CAMS has larger global and regional τ compared to the satellite mean, consistent with another evaluation by Inness et al. (2019). Our results indicate that the spatial mean τ of CAMS is at the upper end of the satellite range. It is an example of the dependence of a reanalysis on the assimilated satellite data. Assimilating other datasets can lead to different results (Cheng et al., 2019).

5. Conclusions

We assessed the current knowledge of present-day τ and the degree to which aerosol-climate model simulations of τ improved in the new generations of models. To this end, we have analyzed 94 data sets from different techniques to estimate τ, including new data from CMIP6, AeroCom-III, and satellite products. We quantified the uncertainty in observational estimates of the large-scale patterns of τ using the mean and range across modern satellite products. The conclusions from the analysis of the comprehensive data are:

1. The range in the spatial mean τ from the eight satellite products is 0.124–0.164 with a mean of 0.14, averaged for 60°N–60°S and 1998–2019. This range translates to an observational uncertainty of −11% to +17% about the satellite mean. Regional and seasonal differences across the satellite products can be larger.
2. A clear and sustainable improvement in the simulated spatio-temporal mean τ across generations of aerosol-climate models is not seen in the data assessment. Although multi-model mean τ over space and time in CMIP5, CMIP6, AeroCom-I, and AeroCom-III fall all within the satellite uncertainty, individual aerosol-climate models often do not. A reduction in the model spread is identified in CMIP6 compared to CMIP5, but we see a clear shift of the spatial mean of τ from CMIP6 and AeroCom-III toward the upper end of the satellite range. The model spread for spatial mean τ in AeroCom-III has significantly increased compared to AeroCom-I, although it might be explained by few models.
3. Our results also indicate no systematic improvement over time for how aerosol-climate models represent the spatio-temporal patterns of τ. Some regional model spreads have been reduced, for example, over China from AeroCom-I to AeroCom-III, but the global spatial patterns of τ differ across the four multi-model means. We identified a larger regional difference for the multi-model mean patterns of τ in CMIP6 and AeroCom-III over the Southern Ocean. Being remote and therefore pristine, the Southern Ocean is perceived as a useful environment for assessing aerosol effects on climate with models, but our results suggest that this is prone to errors.
4. The observational uncertainty in τ implies that evaluation studies using too few observation data sets might draw biased conclusions, even for a global mean. For instance, if we compare the spatial mean τ from CMIP6 only against MODIS or CAMS, we would conclude an improvement for CMIP6 compared to CMIP5.

Comparing against the satellite mean or another satellite product like MISR, however, indicates that CMIP6 has larger τ and lies at the upper end of the observational range for the spatial mean of τ, whereas CMIP5 is in agreement with the satellite mean.

Taken together, we found considerable differences in the spatio-temporal representation of τ across contemporary data sets. The performance of individual data sets varies with respect to the different spatio-temporal metrics assessed, making an overall ranking of the data sets for a present-day climatology of τ difficult. This is at least in parts due to a shortage of ground-based observational data to constrain biomass burning and natural aerosols in satellite data and model simulations. The largest uncertainties occur between April and September and in the Southern Ocean, North Africa, and East Asia, that should be considered for the planning of future measurement campaigns.

The growing number of observational data sets for the global patterns of present-day τ is an opportunity to more extensively use observational uncertainty for studying aerosols. Combining different data sets via an assimilation system is an opportunity to better constrain the spatio-temporal patterns of τ. Our analysis suggests that an assimilation of multiple satellite products for τ would be beneficial to account for observational uncertainty. Paired with an increased spatial resolution and a separate treatment of different aerosol species, it appears to be a promising path to improve our knowledge of aerosol distributions and ultimately their effects on climate.

Data Availability Statement

Data was provided by the Earth System Grid Federation for CMIP (https://esgf-data.dkrz.de), Met Norway for AeroCom (https://aerocom.met.no/data), MPI-M for MAC-v2, the ECMWF for CAMS (https://www.ecmwf.int/en/research/climate-reanalysis/cams-reanalysis), the Finnish Meteorological Institute for FMImerge (http://nsdc.fmi.fi/data/data_aod), the French Space Agency CNES, Laboratoire d'Optique Atmosphérique and GRASP SAS for POLDER-GRASP (https://www.grasp-open.com), the ESA Climate Change Initiative for AATSR (http://www.esa-aerosol-cci.org), and the Copernicus Climate Data Store for SLSTR (https://cds.climate.copernicus.eu). Data from MODIS, MISR, SeaWiFS and MERRA-2 is available through the Giovanni online data system, developed and maintained by the NASA GES DISC (https://giovanni.gsfc.nasa.gov/giovanni/). The CALIPSO data was obtained from the NASA Langley Research Center Atmospheric Science Data Center (https://asdc.larc.nasa.gov/project/CALIPSO/CAL_LID_L3_Tropospheric_APro_AllSky-Standard-V4-20_V4-20).

Acknowledgments
We thank MPI-M and the German Climate Computing Centre for acquiring the data, and the data creators for producing and providing it: The Earth System Grid Federation for CMIP, Met Norway for AeroCom, MPI-M for MAC-v2, ECMWF for CAMS, the Finnish Meteorological Institute for FMImerge, the French Space Agency CNES, Laboratoire d'Optique Atmosphrique and GRASP SAS for POLDER-GRASP, the ESA Climate Change Initiative for AATSR, the Copernicus Climate Data Store for SLSTR, NASA GES DISC for providing MODIS, MISR, SeaWiFS and MERRA-2, and the NASA Langley Research Center Atmospheric Science Data Center for providing CALIPSO data. We thank P. Xian for providing ICAP data. We thank Y. Copin for the program to create the Taylor diagram. We thank the four reviewers and the editor L. Russel for the thoughtful comments and the appraisal of our manuscript. G. Alessa and S. Fiedler acknowledge the funding from the Max-Planck-Institute for Meteorology (MPI-M). S. Fiedler acknowledges funding for the research area "Climate Monitoring and Diagnostic" in the Hans-Ertel-Center for Weather Research (ID: BMVI/DWD 4818DWDP5A) and for the Collaborative Research Centre 1211 (ID: DFG 268 236 062). Open access funding enabled and organized by Projekt DEAL.

References

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., et al. (2020). Bounding global aerosol radiative forcing of climate change. *Reviews of Geophysics*, *58*(1), e2019RG000660. https://doi.org/10.1029/2019RG000660

Benedetti, A., Morcrette, J.-J., Boucher, O., Dethof, A., Engelen, R. J., Fisher, M., & Suttie, M. (2009). Aerosol analysis and forecast in the European Centre for medium-range weather forecasts integrated forecast system: 2. Data assimilation. *Journal of Geophysical Research: Atmospheres*, *114*(D13). https://doi.org/10.1029/2008jd011115

Bevan, S. L., North, P. R., Los, S. O., & Grey, W. M. (2012). A global dataset of atmospheric aerosol optical depth and surface reflectance from AATSR. *Remote Sensing of Environment*, *116*, 199–210. https://doi.org/10.1016/j.rse.2011.05.024

Chand, D., Wood, R., Ghan, S. J., Wang, M., Ovchinnikov, M., Rasch, P. J., & Moore, T. (2012). Aerosol optical depth increase in partly cloudy conditions. *Journal of Geophysical Research*, *117*(D17). https://doi.org/10.1029/2012jd017894

Chen, C., Dubovik, O., Fuertes, D., Litvinov, P., Lapyonok, T., Lopatin, A., et al. (2020). Validation of GRASP algorithm product from POLDER/PARASOL data and assessment of multi-angular polarimetry potential for aerosol monitoring. *Earth System Science Data*, *12*(4), 3573–3620. https://doi.org/10.5194/essd-12-3573-2020

Cheng, Y., Dai, T., Goto, D., Schutgens, N. A. J., Shi, G., & Nakajima, T. (2019). Investigating the assimilation of CALIPSO global aerosol vertical observations using a four-dimensional ensemble Kalman filter. *Atmospheric Chemistry and Physics*, *19*(21), 13445–13467. https://doi.org/10.5194/acp-19-13445-2019

Cherian, R., & Quaas, J. (2020). Trends in AOD, clouds, and cloud radiative effects in satellite data and CMIP5 and CMIP6 model simulations over aerosol source regions. *Geophysical Research Letters*, *47*(9), e2020GL087132. https://doi.org/10.1029/2020GL087132

Colarco, P. R., Kahn, R. A., Remer, L. A., & Levy, R. C. (2014). Impact of satellite viewing-swath width on global and regional aerosol optical thickness statistics and trends. *Atmospheric Measurement Techniques*, *7*(7), 2313–2335. https://doi.org/10.5194/amt-7-2313-2014

Dubovik, O., Fuertes, D., Litvinov, P., Lopatin, A., Lapyonok, T., Doubovik, I., & Federspiel, C. (2021). Multi-term LSM for applying multiple a priori constraints in problems of atmospheric remote sensing: GRASP algorithm—Concept and applications. *Frontiers in Remote Sensing*.

Dubovik, O., Lapyonok, T., Litvinov, P., Herman, M., Fuertes, D., Ducos, F., et al. (2014). *GRASP: A versatile algorithm for characterizing the atmosphere*. SPIE: Newsroom. Retrieved from http://spie.org/x109993.xml

Evan, A. T., Flamant, C., Fiedler, S., & Doherty, O. (2014). An analysis of Aeolian dust in climate models. *Geophysical Research Letters*, *41*(16), 5996–6001. https://doi.org/10.1002/2014GL060545

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fiedler, S., Crueger, T., D'Agostino, R., Peters, K., Becker, T., Leutwyler, D., et al. (2020). Simulated tropical precipitation assessed across three major phases of the Coupled Model Intercomparison Project (CMIP). *Monthly Weather Review*, *148*(9), 3653–3680. https://doi.org/10.1175/MWR-D-19-0404.1

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/jcli-d-16-0758.1

Gliß, J., Mortier, A., Schulz, M., Andrews, E., Balkanski, Y., Bauer, S. E., et al. (2021). AeroCom phase III multi-model evaluation of the aerosol life cycle and optical properties using ground- and space-based remote sensing as well as surface in situ observations. *Atmospheric Chemistry and Physics*, *21*(1), 87–128. https://doi.org/10.5194/acp-21-87-2021

Heinold, B., Knippertz, P., Marsham, J. H., Fiedler, S., Dixon, N. S., Schepanski, K., & Tegen, I. (2013). The role of deep convection and nocturnal low-level jets for dust emission in summertime West Africa: Estimates from convection-permitting simulations. *Journal of Geophysical Research: Atmospheres*, *118*(10), 4385–4400. https://doi.org/10.1002/jgrd.50402

Holben, B., Eck, T., Slutsker, I., Tanré, D., Buis, J., Setzer, A., et al. (1998). AERONET- A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, *66*(1), 1–16. https://doi.org/10.1016/S0034-4257(98)00031-5

Hsu, N., Gautam, R., Sayer, A., Bettenhausen, C., Li, C., Jeong, M., & Holben, B. (2012). Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010. *Atmospheric Chemistry and Physics*, *12*(17), 8037–8053.

Hsu, N., Jeong, M.-J., Bettenhausen, C., Sayer, A., Hansell, R., Seftor, C., & Tsay, S.-C. (2013). Enhanced deep blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research: Atmospheres*, *118*(16), 9296–9315. https://doi.org/10.1002/jgrd.50712

Huneeus, N., Schulz, M., Balkanski, Y., Griesfeller, J., Prospero, J., Kinne, S., et al. (2011). Global dust model intercomparison in AeroCom phase I. *Atmospheric Chemistry and Physics*, *11*(15), 7781–7816. https://doi.org/10.5194/acp-11-7781-2011

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., et al. (2019). The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, *19*(6), 3515–3556. https://doi.org/10.5194/acp-19-3515-2019

Jamet, C., Moulin, C., & Thiria, S. (2004). Monitoring aerosol optical properties over the Mediterranean from SeaWiFS images using a neural network inversion. *Geophysical Research Letters*, *31*(13). https://doi.org/10.1029/2004gl019951

Johnson, B. T., Brooks, M. E., Walters, D., Woodward, S., Christopher, S., & Schepanski, K. (2011). Assessment of the Met Office dust forecast model using observations from the GERBILS campaign. *Quarterly Journal of the Royal Meteorological Society*, *137*(658), 1131–1148. https://doi.org/10.1002/qj.736

Jones, P. (1999). First- and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review*, *127*(9), 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:fasocr>2.0.co;2

Kahn, R. A., Gaitley, B. J., Garay, M. J., Diner, D. J., Eck, T. F., Smirnov, A., & Holben, B. N. (2010). Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the aerosol robotic network. *Journal of Geophysical Research*, *115*(D23). https://doi.org/10.1029/2010jd014601

Kahn, R. A., Gaitley, B. J., Martonchik, J. V., Diner, D. J., Crean, K. A., & Holben, B. (2005). Multiangle Imaging Spectroradiometer (MISR) global aerosol optical depth validation based on 2 years of coincident Aerosol Robotic Network (AERONET) observations. *Journal of Geophysical Research*, *110*(D10). https://doi.org/10.1029/2004jd004706

Kahn, R. A., Nelson, D. L., Garay, M. J., Levy, R. C., Bull, M. A., Diner, D. J., et al. (2009). Misr aerosol product attributes and statistical comparisons with MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(12), 4095–4114. https://doi.org/10.1109/tgrs.2009.2023115

Kinne, S. (2019). The MACv2 aerosol climatology. *Tellus B: Chemical and Physical Meteorology*, *71*(1), 1–21. https://doi.org/10.1080/16000889.2019.1623639

Kinne, S., O'Donnel, D., Stier, P., Kloster, S., Zhang, K., Schmidt, H., et al. (2013). MAC-v1: A new global aerosol climatology for climate studies. *Journal of Advances in Modeling Earth Systems*, *5*(4), 704–740. https://doi.org/10.1002/jame.20035

Kinne, S., Schulz, M., Textor, C., Guibert, S., Balkanski, Y., Bauer, S. E., et al. (2006). An AeroCom initial assessment–Optical properties in aerosol component modules of global models. *Atmospheric Chemistry and Physics*, *6*(7), 1815–1834. https://doi.org/10.5194/acp-6-1815-2006

Koffi, B., Schulz, M., Bréon, F.-M., Griesfeller, J., Winker, D., Balkanski, Y., et al. (2012). Application of the CALIOP layer product to evaluate the vertical distribution of aerosols estimated by global models: AeroCom phase I results. *Journal of Geophysical Research*, *117*(D10). https://doi.org/10.1029/2011JD016858

Kok, J. F., Ridley, D. A., Zhou, Q., Miller, R. L., Zhao, C., Heald, C. L., et al. (2017). Smaller desert dust cooling effect estimated from analysis of dust size and abundance. *Nature Geoscience*, *10*(4), 274–278. https://doi.org/10.1038/ngeo2912

Krasting, J. P., John, J. G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., et al. (2018). *NOAA-GFDL GFDL-ESM4 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from https://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.NOAA-GFDL.GFDL-ESM4.historical

Levy, R., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., & Hsu, N. (2013). The Collection 6 MODIS aerosol products over land and ocean. *Atmospheric Measurement Techniques*, *6*(11), 2989. https://doi.org/10.5194/amt-6-2989-2013

Levy, R. C., Mattoo, S., Sawyer, V., Shi, Y., Colarco, P. R., Lyapustin, A. I., & Remer, L. A. (2018). Exploring systematic offsets between aerosol products from the two MODIS sensors. *Atmospheric Measurement Techniques*, *11*(7), 4073–4092. https://doi.org/10.5194/amt-11-4073-2018

Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., & Kaufman, Y. J. (2007). Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. *Journal of Geophysical Research*, *112*(D13). https://doi.org/10.1029/2006jd007811

Li, Z., Zhao, X., Kahn, R., Mishchenko, M., Remer, L., Lee, K.-H., et al. (2009). Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: A review and perspective. *Annales Geophysicae*, *27*(7), 2755–2770. https://doi.org/10.5194/angeo-27-2755-2009

Mangla, R., Indu, J., & Chakra, S. S. (2020). Inter-comparison of multi-satellites and aeronet AOD over Indian region. *Atmospheric Research*, *240*, 104950. https://doi.org/10.1016/j.atmosres.2020.104950

Marsham, J. H., Dixon, N. S., Garcia-Carreras, L., Lister, G. M. S., Parker, D. J., Knippertz, P., & Birch, C. E. (2013). The role of moist convection in the West African monsoon system: Insights from continental-scale convection-permitting simulations. *Geophysical Research Letters*, *40*(9), 1843–1849. https://doi.org/10.1002/grl.50347

Martin, G. M., & Levine, R. C. (2012). The influence of dynamic vegetation on the present-day simulation and future projections of the south Asian summer monsoon in the HadGEM2 family. *Earth System Dynamics*, *3*(2), 245–261. https://doi.org/10.5194/esd-3-245-2012

Misra, A., Kanawade, V. P., & Tripathi, S. N. (2016). Quantitative assessment of AOD from 17 CMIP5 models based on satellite-derived AOD over India. *Annales Geophysicae*, *34*(8), 657–671. https://doi.org/10.5194/angeo-34-657-2016

Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geoscientific Model Development*, *8*(5), 1339–1356. https://doi.org/10.5194/gmd-8-1339-2015

Mortier, A., Gliß, J., Schulz, M., Aas, W., Andrews, E., Bian, H., & Tilmes, S. (2020). Evaluation of climate model aerosol trends with ground-based observations over the last 2 decades—An AeroCom and CMIP6 analysis. *Atmospheric Chemistry and Physics*, *20*(21), 13355–13378. https://doi.org/10.5194/acp-20-13355-2020

Moseid, K. O., Schulz, M., Storelvmo, T., Julsrud, I. R., Olivié, D., Nabat, P., & Gastineau, G. (2020). Bias in CMIP6 models as compared to observed regional dimming and brightening. *Atmospheric Chemistry and Physics*, *20*(24), 16023–16040. https://doi.org/10.5194/acp-20-16023-2020

North, P., Heckel, A., & Pearson, K. (2021). *SU-SLSTR ATBD, C3S Ref: C3S_D312a_Lot5_202103_SU_SLSTR_ATBD_v1.12*. Copernicus Climate Data Store. Retrieved from cds.climate.copernicus.eu

North, P. R., Briggs, S. A., Plummer, S. E., & Settle, J. J. (1999). Retrieval of land surface bidirectional reflectance and aerosol opacity from ATSR-2 multiangle imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(1), 526–537. https://doi.org/10.1109/36.739106

North, P. R. J. (2002). Estimation of aerosol opacity and land surface bidirectional reflectance from ATSR-2 dual-angle imagery: Operational method and validation. *Journal of Geophysical Research*, *107*(D12). https://doi.org/10.1029/2000JD000207

Pérez, C., Haustein, K., Janjic, Z., Jorba, O., Huneeus, N., Baldasano, J. M., & Thomson, M. (2011). Atmospheric dust modeling from MESO to global scales with the online NMMB/BSC-dust model – Part 1: Model description, annual simulations and evaluation. *Atmospheric Chemistry and Physics*, *11*(24), 13001–13027. https://doi.org/10.5194/acp-11-13001-2011

Popp, T., De Leeuw, G., Bingen, C., Brühl, C., Capelle, V., Chedin, A., & Xue, Y. (2016). Development, production and evaluation of aerosol climate data records from European satellite observations (aerosol_CCI). *Remote Sensing*, *8*(5). https://doi.org/10.3390/rs8050421

Randles, C. A., da Silva, A. M., Buchard, V., Colarco, P. R., Darmenov, A., Govindaraju, R., & Flynn, C. J. (2017). The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. *Journal of Climate*, *30*(17), 6823–6850. https://doi.org/10.1175/jcli-d-16-0609.110.1175/jcli-d-16-0609.1

Ridley, D. A., Heald, C. L., Kok, J. F., & Zhao, C. (2016). An observationally constrained estimate of global dust aerosol optical depth. *Atmospheric Chemistry and Physics*, *16*(23), 15097–15117. https://doi.org/10.5194/acp-16-15097-2016

Roberts, A. J., Marsham, J. H., Knippertz, P., Parker, D. J., Bart, M., Garcia-Carreras, L., & Walker, D. (2017). New Saharan wind observations reveal substantial biases in analysed dust-generating winds. *Atmospheric Science Letters*, *18*(9), 366–372. https://doi.org/10.1002/asl.765

Sai Suman, M. N., Gadhavi, H., Ravi Kiran, V., Jayaraman, A., & Rao, S. V. B. (2014). Role of coarse and fine mode aerosols in MODIS AOD retrieval: A case study over southern India. *Atmospheric Measurement Techniques*, *7*(4), 907–917. https://doi.org/10.5194/amt-7-907-2014

Samset, B. H., Stjern, C. W., Andrews, E., Kahn, R. A., Myhre, G., Schulz, M., & Schuster, G. L. (2018). Aerosol absorption: Progress towards global and regional constraints. *Current Climate Change Reports*, *4*(2), 65–83. https://doi.org/10.1007/s40641-018-0091-4

Schulz, M., Textor, C., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., et al. (2006). *Radiative forcing by aerosols as derived from the AeroCom present-day and pre-industrial simulations*. https://doi.org/10.5194/acp-6-5225-2006

Schutgens, N., Dubovik, O., Hasekamp, O., Torres, O., Jethva, H., Leonard, P. J. T., & Stier, P. (2021). AeroCom and aerosat AAOD and ssa study – Part 1: Evaluation and intercomparison of satellite measurements. *Atmospheric Chemistry and Physics*, *21*(9), 6895–6917. https://doi.org/10.5194/acp-21-6895-2021

Schutgens, N., Sayer, A. M., Heckel, A., Hsu, C., Jethva, H., de Leeuw, G., & Stier, P. (2020). An AeroCom–AeroSat study: Intercomparison of satellite AOD datasets for aerosol model evaluation. *Atmospheric Chemistry and Physics*, *20*(21), 12431–12457. https://doi.org/10.5194/acp-20-12431-2020

Shao, Y. (2000). *Physics and modelling of wind erosion*. Springer Netherlands. Retrieved from https://books.google.ca/books?id=THNRNImhgaAC

Shao, Y., Wyrwoll, K.-H., Chappell, A., Huang, J., Lin, Z., McTainsh, G. H., & Yoon, S. (2011). Dust cycle: An emerging core theme in earth system science. *Aeolian Research*, *2*(4), 181–204. https://doi.org/10.1016/j.aeolia.2011.02.001

Smirnov, A., Holben, B., Slutsker, I., Giles, D., Mcclain, C., Eck, T., & Jourdin, F. (2009). Maritime aerosol network as a component of aerosol robotic network. *Journal of Geophysical Research*, *114*. https://doi.org/10.1029/2008JD011257

Sogacheva, L., Popp, T., Sayer, A. M., Dubovik, O., Garay, M. J., Heckel, A., & Arola, A. (2020). Merging regional and global aerosol optical depth records from major available satellite products. *Atmospheric Chemistry and Physics*, *20*(4), 2031–2056. https://doi.org/10.5194/acp-20-2031-2020

Tanré, D., Kaufman, Y., Herman, M., & Mattoo, S. (1997). Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances. *Journal of Geophysical Research*, *102*(D14), 16971–16988. https://doi.org/10.1029/96jd03437

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. https://doi.org/10.1029/2000JD900719

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/bams-d-11-00094.1

Tegen, I., Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Bey, I., Schutgens, N., & Lohmann, U. (2018). The aerosol-climate model ECHAM6.3-HAM2.3: Aerosol evaluation. In *2018 Geoscientific model development discussions* (pp. 1–54). https://doi.org/10.5194/gmd-2018-235

Wang, M., Knobelspiesse, K. D., & McClain, C. R. (2005). Study of the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) aerosol optical property data over ocean in combination with the ocean color products. *Journal of Geophysical Research*, *110*(D10). https://doi.org/10.1029/2004jd004950

Watson-Parris, D., Schutgens, N., Reddington, C., Pringle, K. J., Liu, D., Allan, J. D., & Stier, P. (2019). In situ constraints on the vertical distribution of global aerosol. *Atmospheric Chemistry and Physics*, *19*(18), 11765–11790. https://doi.org/10.5194/acp-19-11765-2019

Winker, D., Pelon, J., Coakley, J., Jr., Ackerman, S., Charlson, R., Colarco, P., et al. (2010). The CALIPSO mission: A global 3D view of aerosols and clouds. *Bulletin of the American Meteorological Society*, *91*(9), 1211–1230. https://doi.org/10.1175/2010bams3009.1

Witek, M. L., Diner, D. J., & Garay, M. J. (2016). Satellite assessment of sea spray aerosol productivity: Southern ocean case study. *Journal of Geophysical Research: Atmospheres*, *121*(2), 872–894. https://doi.org/10.1002/2015JD023726

Witek, M. L., Garay, M. J., Diner, D. J., Bull, M. A., & Seidel, F. C. (2018). New approach to the retrieval of AOD and its uncertainty from MISR observations over dark water. *Atmospheric Measurement Techniques*, *11*(1), 429–439. https://doi.org/10.5194/amt-11-429-2018

Xian, P., Reid, J. S., Hyer, E. J., Sampson, C. R., Rubin, J. I., Ades, M., & Zhang, J. (2019). Current state of the global operational aerosol multi-model ensemble: An update from the international cooperative for aerosol prediction (ICAP). *Quarterly Journal of the Royal Meteorological Society*, *145*(S1), 176–209. https://doi.org/10.1002/qj.3497

Yu, H., Chin, M., Winker, D. M., Omar, A. H., Liu, Z., Kittaka, C., & Diehl, T. (2010). Global view of aerosol vertical distributions from CALIPSO lidar measurements and GOCART simulations: Regional and seasonal variations. *Journal of Geophysical Research*, *115*(D4). https://doi.org/10.1029/2009JD013364

Zhang, J., & Reid, J. S. (2010). A decadal regional and global trend analysis of the aerosol optical depth using a data-assimilation grade over-water MODIS and level 2 MISR aerosol products. *Atmospheric Chemistry and Physics*, *10*(22), 10949–10963. https://doi.org/10.5194/acp-10-10949-2010

References From the Supporting Information

Adachi, Y., Yukimoto, S., Deushi, M., Obata, A., Nakano, H., Tanaka, T., et al. (2015). *MRI-ESM1 model output prepared for CMIP5 historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=MRM1hi

Atmosphere and Ocean Research Institute, The University Of Tokyo (AORI), National Institute for Environmental Studies (NIES), & Japan Agency for Marine-Earth Science and Technology (JAMSTEC). (2015a). *MIROC4h model output prepared for CMIP5 historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=MIMHhi

Atmosphere and Ocean Research Institute, The University Of Tokyo (AORI), National Institute for Environmental Studies (NIES), & Japan Agency for Marine-Earth Science and Technology (JAMSTEC). (2015b). *MIROC5 model output prepared for CMIP5 historical, served by ESGF. World Data Center for Climate (WDCC) at DKRZ*. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=MIM5hi

Bader, D. C., Leung, R., Taylor, M., & McCoy, R. B. (2019a). *E3SM-Project E3SM1.0 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/834meta/CMIP6/CMIP6.CMIP.E3SM-Project.E3SM-1-0.historical

Bader, D. C., Leung, R., Taylor, M., & McCoy, R. B. (2019b). *E3SM-Project E3SM1.1 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.E3SM-Project.E3SM-1-1.historical

Bentsen, M., Bethke, I., Debernard, J., Drange, H., Heinze, C., Iversen, T., et al. (2012). *cmip5 output1 NCC NorESM1-M historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=NCCNMhi

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Uotila, P., Hirst, T., et al. (2016a). *ACCESS1-0 model output prepared for CMIP5 historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=CSA0hi

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Uotila, P., Hirst, T., et al. (2016b). ACCESS1-3 model output prepared for CMIP5 historical, served by ESGF. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=CSA3hi

Byun, Y.-H., Lim, Y.-J., Sung, H. M., Kim, J., Sun, M., & Kim, B.-H. (2019). NIMS-KMA KACE1.0-G model output prepared for CMIP6 CMIP historical. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.NIMS-KMA.KACE-1-0-G.historical

Danabasoglu, G. (2019a). *NCAR CESM2-FV2 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.NCAR.CESM2-FV2.historical

Danabasoglu, G. (2019b). *NCAR CESM2 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.NCAR.CESM2.historical

Danabasoglu, G. (2019c). *NCAR CESM2-WACCM model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.NCAR.CESM2-WACCM.historical

Danek, C., Shi, X., Stepanek, C., Yang, H., Barbi, D., Hegewald, J., & Lohmann, G. (2020). *AWI AWI-ESM1.1LR model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.AWI.AWI-ESM-1-1-LR.historical

Denvil, S., Foujols, M. A., Caubel, A., Marti, O., Dufresne, J.-L., Bopp, L., et al. (2016). *IPSL-CM5A-LR model output prepared for CMIP5 historical experiment, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=IPILhi

Dunne, J. P., John, J. G., Adcroft, A., Hallberg, R., Shevliakova, E., Stouffer, R., et al. (2014). *NOAA GFDL GFDL-ESM2G, historical experiment output for CMIP5 AR5, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=NGEGhi

Dunne, J. P., John, J. G., Shevliakova, E., Stouffer, R., Griffies, S. M., Malyshev, S., et al. (2014). *NOAA GFDL GFDL-ESM2M, historical experiment output for CMIP5 AR5, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=NGEMhi

Fairhead, L., Denvil, S., Foujols, M. A., Caubel, A., Marti, O., Dufresne, J.-L., et al. (2016). *IPSL-CM5B-LR model output prepared for CMIP5 historical experiment, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=IPIBhi

Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Dufresne, J.-L., Bopp, L., et al. (2016). *IPSL-CM5A-MR model output prepared for CMIP5 historical experiment, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=IPIMhi

Hajima, T., Abe, M., Arakawa, O., Suzuki, T., Komuro, Y., Ogura, T., et al. (2019). *MIROC MIROC-ES2L model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.MIROC.MIROC-ES2L.historical

Hardiman, S., Butchart, N., Hinton, T., Osprey, S., Gray, L., Jones, C., & Hughes, J. (2014). *HadGEM2-CC model output prepared for CMIP5 historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=MOGChi

Horowitz, L. W., Donner, L. J., Levy, H., Naik, V., Zhao, M., Winton, M., et al. (2014). *NOAA GFDL GFDL-CM3, historical experiment output for CMIP5 AR5, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=NGG3hi

Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Atmosphere and Ocean Research Institute, The University Of Tokyo (AORI), & National Institute for Environmental Studies (NIES). (2015a). *MIROC-ESM-CHEM model output prepared for CMIP5 historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=MIM7hi

Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Atmosphere and Ocean Research Institute, The University Of Tokyo (AORI), & National Institute for Environmental Studies (NIES). (2015b). *MIROC-ESM model output prepared for CMIP5 historical,*

served by ESGF. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=MIMEhi

Jeffrey, S., Rotstayn, L., Collier, M., Dravitzki, S., Hamalainen, C., Moeseneder, C., et al. (2016). *CSIRO-Mk3-6-0 model output prepared for CMIP5 historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=CQMKhi

Ji, D., Wang, L., Feng, J., Wu, Q., & Cheng, H. (2015). *BNU-ESM model output prepared for CMIP5 historical experiment, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=BUBUhi

Jones, C., Hughes, J., Jones, G., Christidis, N., Lott, F., Sellar, A., et al. (2014). *HadGEM2-ES model output prepared for CMIP5 historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=MOGEhi

LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences (IAP-LASG). (2015). *FGOALS-g2 model output prepared for CMIP5 historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=LSF2hi

Lee, W.-L., & Liang, H.-C. (2020). *AS-RCEC TaiESM1.0 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.AS-RCEC.TaiESM1.historical

NASA Goddard Institute for Space Studies (NASA/GISS). (2014b). *NASA-GISS: GISS-E2-R model output prepared for CMIP5 historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=GIGRhi

Neale, R. (2013). CESM1-CAM5 model output prepared for CMIP5 historical, served by ESGF. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=NFCChi

Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Stoll, J., Folini, D. S., Tegen, I., et al. (2019). *HAMMOZ-Consortium MPI-ESM1.2-HAM model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.HAMMOZ-Consortium.MPI-ESM-1-2-HAM.historical

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., & Andrews, T. (2019a). *MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.MOHC.HadGEM3-GC31-LL.historical

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., & Andrews, T. (2019b). *MOHC HadGEM3-GC31-MM model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.MOHC.HadGEM3-GC31-MM.historical

Seferian, R. (2018). *CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.CNRM-CERFACS.CNRM-ESM2-1.historical

Seland, Ø., Bentsen, M., Oliviè, D. J. L., Toniazzo, T., Gjermundsen, A., Graff, L. S., et al. (2019). *NCC NorESM2-LM model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.NCC.NorESM2-LM.historical

Tjiputra, J., Roelandt, C., Bentsen, M., Lawrence, D., Lorentzen, T., Schwinger, J., et al. (2012). *CMIP5 output1 NCC NorESM1-ME historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=NCCNEhi

Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., et al. (2019a). *INM INM-CM4-8 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.INM.INM-CM4-8.historical

Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., et al. (2019b). *INM INM-CM5-0 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.INM.INM-CM5-0.historical

Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., et al. (2015). *MRI-CGCM3 model output prepared for CMIP5 historical, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. Retrieved from http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=MRMChi

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). *MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.MRI.MRI-ESM2-0.historical

Zhang, J., Wu, T., Shi, X., Zhang, F., Li, J., Chu, M., et al. (2018). *BCC BCC-ESM1 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.BCC.BCC-ESM1.historical

Ziehn, T., Chamberlain, M., Lenton, A., Law, R., Bodman, R., Dix, M., et al. (2019). *CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. Retrieved from http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.CSIRO.ACCESS-ESM1-5.historical