HAL open science

Evaluation of the aerosol vertical distribution in global aerosol models through comparison against CALIOP measurements: AeroCom phase II results

Brigitte Koffi, Michael Schulz, Francois-Marie Breon, Frank Dentener, Birthe Marie Steensen, Jan Griesfeller, David Winker, Yves Balkanski, Susanne E. Bauer, Nicolas Bellouin, et al.

To cite this version:

Brigitte Koffi, Michael Schulz, Francois-Marie Breon, Frank Dentener, Birthe Marie Steensen, et al.. Evaluation of the aerosol vertical distribution in global aerosol models through comparison against CALIOP measurements: AeroCom phase II results. Journal of Geophysical Research: Atmospheres, 2016, 121 (12), pp.7254 - 7283. 10.1002/2015JD024639 . hal-01602797

HAL Id: hal-01602797

https://hal.science/hal-01602797v1

Submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2015JD024639

Key Points:
- The ability of 11 global models in simulating the aerosol vertical distribution is assessed
- Hypotheses for the models performance and evolution and for the inter-model diversity are discussed
- An analysis of CALIOP limitations and uncertainties contributing to the discrepancies is provided

Supporting Information:
- Supporting Information S1

Correspondence to:
B. Koffi,
brigitte.koffi-lefeivre@jrc.ec.europa.eu

Citation:
Koffi, B., et al. (2016), Evaluation of the aerosol vertical distribution in global aerosol models through comparison against CALIOP measurements: AeroCom phase II results, *J. Geophys. Res. Atmos.*, *121*, 7254–7283, doi:10.1002/2015JD024639.

Received 15 DEC 2015
Accepted 16 MAY 2016
Accepted article online 23 MAY 2016
Published online 30 JUN 2016

©2016. American Geophysical Union. All Rights Reserved.

Evaluation of the aerosol vertical distribution in global aerosol models through comparison against CALIOP measurements: AeroCom phase II results

Brigitte Koffi[1], Michael Schulz[2], François-Marie Bréon[3], Frank Dentener[1], Birthe Marie Steensen[2], Jan Griesfeller[2], David Winker[4], Yves Balkanski[3], Susanne E. Bauer[5,6], Nicolas Bellouin[7], Terje Berntsen[8,9], Huisheng Bian[10,11], Mian Chin[10], Thomas Diehl[1], Richard Easter[12], Steven Ghan[12], Didier A. Hauglustaine[3], Trond Iversen[2,13], Alf Kirkevåg[2], Xiaohong Liu[12,14], Ulrike Lohmann[15], Gunnar Myhre[9], Phil Rasch[10], Øyvind Seland[2], Ragnhild B. Skeie[9], Stephen D. Steenrod[10], Philip Stier[16], Jason Tackett[17], Toshihiko Takemura[18], Kostas Tsigaridis[5,6], Maria Raffaella Vuolo[3,19], Jinho Yoon[12,20], and Kai Zhang[12,21]

[1]European Commission, Joint Research Centre, Institute for Environment and Sustainability, Ispra, Italy, [2]Norwegian Meteorological Institute, Oslo, Norway, [3]Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France, [4]NASA Langley Research Center, MS/475, Hampton, Virginia, USA, [5]Center for Climate Systems Research, Columbia University, New York, New York, USA, [6]NASA Goddard Institute for Space Studies, New York, New York, USA, [7]Department of Meteorology, University of Reading, Reading, UK, [8]Department of Geosciences, University of Oslo, Oslo, Norway, [9]Center for International Climate and Environmental Research-Oslo (CICERO), Oslo, Norway, [10]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA, [11]Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore Country, Maryland, USA, [12]Pacific Northwest National Laboratory, Richland, Washington, USA, [13]Department of Geosciences, University of Oslo, Oslo, Norway, [14]Now at University of Wyoming, Laramie, Wyoming, USA, [15]ETH-Zentrum, Zürich, Switzerland, [16]Department of Physics, University of Oxford, Oxford, UK, [17]Science Systems and Applications, Inc., Hampton, Virginia, USA, [18]Research Institute for Applied Mechanics, Kyushu University, Fukuoka, Japan, [19]Now at National Institute for Agronomic Research, Thiverval-Grignon, France, [20]Now at Gwangju Institute of Science and Technology, Gwangju, Korea, [21]Max Planck Institute for Meteorology, Hamburg, Germany

Abstract The ability of 11 models in simulating the aerosol vertical distribution from regional to global scales, as part of the second phase of the AeroCom model intercomparison initiative (AeroCom II), is assessed and compared to results of the first phase. The evaluation is performed using a global monthly gridded data set of aerosol extinction profiles built for this purpose from the CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) Layer Product 3.01. Results over 12 subcontinental regions show that five models improved, whereas three degraded in reproducing the interregional variability in $Z_{\alpha 0\text{–}6\,\mathrm{km}}$, the mean extinction height diagnostic, as computed from the CALIOP aerosol profiles over the 0–6 km altitude range for each studied region and season. While the models' performance remains highly variable, the simulation of the timing of the $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ peak season has also improved for all but two models from AeroCom Phase I to Phase II. The biases in $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ are smaller in all regions except Central Atlantic, East Asia, and North and South Africa. Most of the models now underestimate $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ over land, notably in the dust and biomass burning regions in Asia and Africa. At global scale, the AeroCom II models better reproduce the $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ latitudinal variability over ocean than over land. Hypotheses for the performance and evolution of the individual models and for the intermodel diversity are discussed. We also provide an analysis of the CALIOP limitations and uncertainties contributing to the differences between the simulations and observations.

1. Introduction

Among the main sources of uncertainties in the impact of aerosols on the global radiation balance is the diversity in aerosol vertical distribution simulated by global models [*Textor et al.*, 2006, 2007; *Koffi et al.*, 2012; *Samset et al.*, 2013; *Kipling et al.*, 2013; *Myhre et al.*, 2013; *Kim et al.*, 2014; *Vuolo et al.*, 2014; *Hodnebrog et al.*, 2014; *Samset et al.*, 2014; *Kipling et al.*, 2015]. This diversity is due to the large uncertainty in the processes affecting the vertical distribution of aerosols [e.g., *Kipling et al.*, 2013, 2015], as well as to the lack of sufficient altitude-resolved information on their abundance and properties. However, significant progress in characterizing the tropospheric aerosol vertical distribution has been made in the recent years through the use of ground- and satellite-based lidar measurements, especially those from the Cloud–Aerosol Lidar and Infrared Pathfinder

Satellite Observations (CALIPSO) satellite that has been acquiring cloud and aerosol backscattering data since mid-June 2006 [*Winker et al.*, 2010]. The CALIOP Level 2 Layer 5 km product [*Liu et al.*, 2004, 2009; *Omar et al.*, 2009; *Vaughan et al.*, 2004, 2005] has been used in previous studies to assess the 3-D distribution of tropospheric aerosols from regional to global scales [e.g., *Yu et al.*, 2010; *Koffi et al.*, 2012; *Sheridan et al.*, 2012; *Di Pierro et al.*, 2013; *Winker et al.*, 2013; *Prijith et al.*, 2016]. Despite several limitations (see, for instance, *Koffi et al.* [2012], *Winker et al.*[2013], *Young et al.* [2013], and *Kim et al.* [2014]), the version 3.01 product [e.g., *Hu et al.*, 2009; *Liu et al.*, 2010; *Vaughan et al.*, 2010] was shown to provide a realistic and representative description of the mean regional vertical distributions and seasonal variations of aerosol extinctions, over source and downwind regions.

The AeroCom (Aerosol Comparison between Observations and Models; http://aerocom.met.no) project aims at documenting and evaluating the aerosol component modules of global models and proposes intercomparison experiments for such evaluation. Participating modelers are requested to follow a common data-protocol and provide detailed model outputs. In a first model comparison exercise ("AeroCom IA," hereafter), the inhomogeneous performance of different global models in reproducing aerosol properties and distribution showed the need for further model evaluation, as well as room for further model improvement. The unresolved questions have motivated the AeroCom participants to perform a new set of experiments ("AeroCom II," hereafter) with adapted diagnostics, together with further evaluation analyses [*Schulz et al.*, 2009].

The AeroCom models' aerosol properties have been widely documented, evaluated, and analyzed in many previous AeroCom I [e.g., *Kinne et al.*, 2006; *Schulz et al.*, 2006; *Textor et al.*, 2006, 2007; *Koch et al.*, 2009; *Huneeus et al.*, 2011; *Koffi et al.*, 2012] and AeroCom II [e.g., *Myhre et al.*, 2013; *Kim et al.*, 2014; *Mann et al.*, 2014; *Tsigaridis et al.*, 2014; *Kipling et al.*, 2015] studies. Our study is specifically dedicated to the evaluation of the aerosol vertical distribution that is assumed to play an important role in aerosol-climate forcing, which depends not only on the aerosols' concentration, composition, and optical depth but also on the altitude of the aerosol layers in the atmosphere. In a previous paper (*Koffi et al.*[2012]; referred to as "K12" hereafter), we used the CALIOP Layer product 3.01 to derive monthly regionally averaged aerosol extinction profiles over the period 2007–2009 for the evaluation of AeroCom IA "control" experiment for the year 2000 [*Dentener et al.*, 2006; *Textor et al.*, 2006; *Schulz et al.*, 2006]. The models' ability to simulate the mean annual vertical distribution of aerosol extinction, as well as its seasonal variations, was assessed over different subcontinental source and downwind regions representative of industrial, dust, and biomass burning pollution. Our presumption was that the CALIOP observations do provide, in a climatological sense, a consistent and robust depiction of the typical vertical distribution of aerosols and its seasonality over the 13 selected source and downwind regions, which was largely tested and documented in this first study. The evaluation of AeroCom IA simulations showed significant discrepancies compared to CALIOP observations, as well as large differences among models, due to a complex combination of different assumptions about emissions, transport, removal pathways, and optical properties. The present paper is an update and extension of K12 documentation and evaluation work, which aims at providing further spatial and temporal details on the individual AeroCom models' skills and at better understanding the reasons for the model discrepancies and diversity. As in K12, it is dedicated to the evaluation of the vertical distribution of the aerosol extinction. Therefore, the analysis of the models' performance focuses on normalized aerosol extinction profiles and mean extinction height diagnostics. Its specific objectives are (i) to determine to what extent and in what manner eight of these models have improved between the two AeroCom experiments, with respect to their skill in reproducing the shape of the 2007–2009 CALIOP-derived mean aerosol extinction profiles over the selected regions, (ii) to evaluate the performance of the three additional models, in the same regions and seasons, and (iii) to further evaluate the AeroCom II models results, as a function of latitude and of land/ocean cover type. To fulfill these objectives, we compiled a monthly, global $1° \times 1°$ gridded data set of aerosol extinction profiles, at 100 m vertical resolution using the K12 methodology, and performed an in-depth analysis of its limitations and uncertainties in order to better assess the uncertainties on models versus observation differences.

The CALIOP Layer product 3.01, the Aerosol Optical Depth (AOD) from MODIS (Moderate Resolution Imaging Spectroradiometer), and the AeroCom simulations analyzed in this study are described in section 2. The methodology used to build our in-house global product of CALIOP-derived aerosol profiles (the so-called "AeroCom-CALIOP" product, hereafter) and to evaluate the AeroCom models is presented and discussed in

Table 1a. AeroCom II Experiments Considered in the Present Study[a]

Model ID	Model (AeroCom Experiment Name)	Type	Operated By	δlon × δlat	$N_{lev0-10\,km}$ AeroCom I/AeroCom II	Degree of Revision Since K12	Model Reference
GISS-ModelE	Model E (GISS-modelE.A2.CTRL)	GCM	GISS	2.5° × 2.0°	9/19	high	*Koch et al.* [2011] and *Tsigaridis et al.* [2013]
GISS-MATRIX	MATRIX (GISS-MATRIX.A2.CTRL)	GCM	GISS	2.5° × 2.0°	-/19	-	*Bauer et al.* [2008]
GMI-MERRA-v3	GMI (GMI-MERRA-v3.A2.CTRL)	CTM	GSFC	2.5° × 2.0°	-/29	-	*Bian et al.* [2009]
GOCART	GOCARTv4 (GOCART-v4.A2.HCA-0)	CTM	GSFC	2.5° × 2.0°	17/12	low	*Chin et al.* [2009]
SPRINTARS	SPRINTARS (SPRINTARS-v384.A2.CTRL)	GCM	Kyushu University	1.1° × 1.1°	11/24	high	*Takemura et al.* [2000, 2005, 2009].
LSCE	LMDz-INCA (LSCEv2c.A2.CTRL)	GCM	LSCE	3.8° × 1.9°	10/10	low	*Hauglustaine et al.* [2004], *Schulz* [2007], and *Balkanski* [2011]
ECHAM-HAM	ECHAM5.5 (HAM2 MPIHAM_V2_KZ.A2.CTRL)	GCM	HAMMOZ Consortium	1.8° × 1.8°	20/20	high	*Zhang et al.* [2012]
HadGEM	HadGEM2-ES (HadGEM2-ES.A2.CTRL)	ESM	MetOffice	1.8° × 1.2°	-/23	-	*Bellouin et al.* [2011]
PNNL	CAM5.1 (CAM5.1-MAM3-PNNL.A2.CTRL)	GCM	PNNL	2.5° × 1.8°	16/16	high	*Liu et al.* [2012]
Oslo CTM2	CTM 2 (OsloCTM2-v2.A2.CTRL)	CTM	CICERO	2.8° × 2.8°	29/29	high	*Skeie et al.* [2011a]
CAM4-Oslo	CAM4-Oslo (CAM4-Oslo-Vcmip5.A2.CTRL)	GCM	met.no	2.5° × 1.8°	12/12	high	*Kirkevåg et al.* [2013]

[a]Models' description (in bold are highlighted the models also investigated in K12). $N_{lev0-10\,km}$ is the number of model levels below 10 km.

section 3. A qualitative and quantitative assessment of the ability of the individual models in reproducing the CALIOP-derived extinction profiles is provided in section 4. The (changes in the) performance of the individual models and the intermodel diversity, as well as the AeroCom-CALIOP product's limitations and uncertainties, are discussed in section 5 before concluding in section 6.

2. Model and Observational Data

2.1. AeroCom Simulations

Eleven global aerosol models that participated in the AeroCom Phase II present-day "control" experiment [*Schulz et al.*, 2009; *Myhre et al.*, 2013] are here examined. One global aerosol model is from Japan (SPRINTARS), five from Europe, and five from North America. Eight out of the 11 models (or their predecessors) were also part of the AeroCom IA "control" experiment analyzed in K12: CAM5.1 [*Liu et al.*, 2012], CAM4-Oslo [*Kirkevåg et al.*, 2013], GISS-modelE [*Koch et al.*, 2011], GOCART-v4 [*Chin et al.*, 2009], LMDz-INCA [*Hauglustaine et al.*, 2004; *Schulz*, 2007; *Balkanski*, 2011;], ECHAM5.5 [*Zhang et al.*, 2012], OsloCTM2 [*Skeie et al.*, 2011a], and SPRINTARS [*Takemura et al.*, 2005, 2009]. The model setup and configurations are model dependent, with a horizontal resolution from 1.1° longitude × 1.1° latitude (SPRINTARS) to 3.75° × 1.9° (LMDz-INCA). The number of vertical layers varies from 19 to 72, of which 10 (LMDz-INCA) to 29 (GMI-MERRA, OsloCTM2) are below 10 km altitude (Table 1a).

For consistency with K12, the phase II models CAM5.1, LMDz-INCA, ECHAM5.5, and GOCART-v4 models are, respectively, denoted PNNL, LSCE, ECHAM-HAM, and GOCART hereafter, as in phase I. While some models have not undergone major changes between the AeroCom phases, others have been so heavily revised (e.g., PNNL or CAM4-Oslo) that they could be viewed as different models (see "Degree of revision" column in Table 1a) as further discussed in section 5.2. The three additional phase II models studied here are GISS_MATRIX [*Bauer et al.*, 2008], GMI-MERRA [*Bian et al.*, 2009], and HadGEM [*Bellouin et al.*, 2011].

In AeroCom IA, the models used either their own choice of 2000 emission data set or the *Dentener et al.* [2006] 2000 1° × 1° unified data set. The fact that the modelers were allowed to use aerosol emissions of their choice complicated the interpre-

tation of the model diversity [*Kinne et al.*, 2006; *Textor et al.*, 2006]. Although the initial recommendation in AeroCom II was for the models to use 1°×1° HCA0 v1 2006 emissions [*Diehl et al.*, 2012], only four of them did, whereas the other models analyzed in this study applied the 0.5°×0.5° 2000 emissions from *Lamarque et al.* [2010] (Table 1b; see also http://aerocom.met.no/emissions.html). The three emission data sets include inventories of black carbon (BC), primary organic carbon (OC), and SO_2 emissions from land-based anthropogenic sources, ocean-going vessels, air traffic, and biomass burning (Table 1c). The temporal resolution varies from daily for volcanoes to monthly for biomass burning and aircraft emissions, and annual averages for land-based and ship emissions. There is generally more temporal variation within the anthropogenic emissions from *Dentener et al.* [2006] and *Diehl et al.* [2012] than from *Lamarque et al.* [2010], since this latter Coupled Model Intercomparison Project Phase 5 (CMIP5) inventory has no fluctuation within a decade. *Diehl et al.* [2012] emission data set captures the major trends of anthropogenic emissions since 1980 to present: the decrease over the US and Western Europe, the sharp decrease around 1990 over Eastern Europe and the former USSR, and the abrupt increase since 2000 over East and South Asia. It is also worth noting that *Dentener et al.* [2006] includes injection heights specifications, whereas none is given in *Lamarque et al.* [2010] and only suggestions are provided by *Diehl et al.* [2012]. To reproduce transport accounting for the 2006 meteorological conditions, models were nudged or driven by reanalysis fields (ECMWF, NCEP/NCAR, and others) for the year 2006, except in the case of CAM4-Oslo and PNNL. These two Global Climate Models provided climatological averages from 5 years of simulation, using identical emissions each year.

An overview of the models' configuration and updates between AeroCom phases I and II, as well as the emission and meteorological fields that are particularly relevant for understanding the simulation of the aerosol vertical distribution, is provided in Tables 1a, 1b, 1c. The aerosol components and microphysics schemes are further described in *Myhre et al.* [2013], *Mann et al.* [2014], and *Tsigaridis et al.* [2014], as well as on the AeroCom website (http://aerocom.met.no).

2.2. Satellite Retrievals

The CALIOP Level 2 Layer 5 km data set used to build the "AeroCom-CALIOP" global product (see section 3.1) is derived from the attenuated Level 1 data through the use of a complex algorithm system that detects cloud and aerosol features, estimates aerosol lidar ratio, and retrieves profiles of aerosol extinction coefficients [*Liu et al.*, 2004, 2009; *Omar et al.*, 2009; *Vaughan et al.*, 2004, 2005]. The extinction profiles are then integrated over height to derive estimates of the AOD of the cloud and aerosol detected layers at 532 nm and 1064 nm wavelengths. From the CALIOP Level 2 Layer 5 km data, we used the AOD, the bottom and top above sea level altitudes, and the cloud-aerosol discrimination score information of all individual aerosol and cloud detected layers. Only the data at 532 nm wavelength, with a better signal-to-noise ratio than at 1064 nm, and only the nighttime observations, which show significantly lower attenuated backscatter noise than daytime observations, are considered. All-sky data are used in order to ensure a higher representativeness of the aerosol regional climatology. In order to evaluate our AeroCom-CALIOP global product (section 3.1), we also analyzed AOD measurements from the Level-3 MODIS Atmosphere Monthly Global Product (Collection 5) over the same 2007 to 2009 period, using the MOD08_M3 "Optical Depth over Land and Ocean" data. Over desert areas, where MODIS algorithm fails at retrieving valid AOD, the "Deep Blue AOD at 550 nm" data are used.

3. Methods

3.1. The AeroCom-CALIOP Gridded Aerosol Product

The AeroCom-CALIOP product is a 3-D 1°×1° gridded global data set of aerosol extinction profiles that has been built from the CALIOP 5 km Level 2 aerosol layer data produced by the CALIPSO algorithm [*Omar et al.*, 2009; *Winker et al.*, 2009], after applying the same screening methods as described in more details in K12: First, two aerosol data screenings are applied, following *Yu et al.* [2010]: The first screening excludes the aerosol layers that have low cloud-aerosol discrimination scores, and the second screening excludes cases where the initial lidar ratio has been adjusted by the CALIPSO algorithm. Although these retrievals only rarely occur, they generally lead to anomalously high AOD [*Omar et al.*, 2009; *Winker et al.*, 2009; *Young and Vaughan*, 2009; *Kittaka et al.*, 2011]. We use *all-sky* data to ensure a higher representativeness of the aerosol regional climatology. Therefore, we also apply a vertical cloud mask (derived from the CALIOP 5 km cloud

Table 1b. Aerosol species and input emissions[a]

Model ID	Meteorology	Aerosol Species (Used to Calculate the AOD)	Anthropogenic (and Biomass Burning if Fifferent Data Set) Input Emissions	Injection Heights
GISS-ModelE	Nudged to NCEP winds[b]	SO_4, BC, OM, SS, DU, NH_4, NO_3	*Lamarque et al.* [2010] (GFED3)	Biomass burning: homogeneous from the surface to the top of the boundary layer
GISS-MATRIX	Nudged to NCEP winds	SO_4, BC, OM, SS, DU, NH_4, NO_3	*Lamarque et al.* [2010] (GFED3)	Biomass burning: homogeneous from the surface to the top of the boundary layer
GMI-MERRA-v3	GEOS-5 MERRA reanalysis, nudged	SO_4, BC, OM, SS, DU, NH_4, NO_3	*Lamarque et al.* [2010]	Biomass burning: at surface
GOCART	NASA GMAO GEOS-4 DAS reanalysis	SO_4, BC, OM, SS, DU	*Diehl et al.* [2012]: HCA0 v1	Biomass burning: homogeneous from the surface to the top of the boundary layer
SPRINTARS	NCEP/NCAR reanalysis, nudged	SO_4, BC, OM, SS, DU	*Diehl et al.* [2012]: HCA0 v1	Biomass burning: homogeneous from the surface to sigma level = 0.74. For other aerosols: Equation (A1) in *Takemura et al.* [2000]
LSCE	ECMWF reanalysis from the Integrated Forecast System (IFS)	SO_4, BC, OM, SS, DU	*Lamarque et al.* [2010]	Biomass burning takes place in the four lowest layers of the model: 50% is injected up to 80 m (meters), 25% between 80 and 200 m, and 12.5% from 200 to 520 m and 12.5% from 520 to 860 m
ECHAM-HAM	Nudged ECMWF analysis	SO_4, BC, POA, SOA, SS, DU	*Diehl et al.* [2012]: HCA0 v1	*Dentener et al.* [2006]
HadGEM	ERA Interim data, nudged	SO_4, BC (ff + bb), OC (ff + bb), SS (not transported), DU, NO_3, biogenic SOA	*Lamarque et al.* [2010]	Biomass burning: homogeneous from 800 m to 2.9 km; BC and OC (ff): 80 m; SO_2 energy sector at 500 m; SO2 degassing volcanoes depending on volcano altitude; SS: prescribed exponential vertical profile
PNNL	CAM5.1, Year 2000	SO_4, BC, OM, SS, DU	*Lamarque et al.* [2010]	*Dentener et al.* [2006] for biomass OM, BC and sulfate (up to 6 km height).
Oslo CTM2	ECMWF reanalysis from IFS	SO_4, BC, OM, SS, DU, NO_3	*Lamarque et al.* [2010]	Biomass burning: altitudes from EU-project RETRO (following *Schultz et al.*, 2007); SO_2, SO_4, NO and NO_2 anthropogenic emissions: distributed in the four lowest model layers (approximately 70 m) dependent on sector.
CAM4-Oslo	Online [c]	SO_4, BC, OM, SS, DU	*Diehl et al.* [2012] HCA0 v1 except DMS, mineral dust, SOA and volcanic SO_2 from *Dentener et al.* [2006]	*Dentener et al.* [2006]

[a]If meteorology was nudged or driven by reanalysis fields, the year 2006 meteorology was used.
[b]Horizontal winds are nudged to NCEP reanalysis [*Kalnay et al.*, 1996], with the rest of the climate parameters being calculated online.
[c]Five year mean of model's calculated meteorology.

Table 1c. Anthropic, Biomass Burning, and Volcanic Emissions from *Dentener et al.* [2006], *Lamarque et al.* [2010], and *Diehl et al.* [2012]

Emission Data Set	Emission Year	Resolution	Species	Sectors	Time Resolution	Injection Heights
Dentener et al. [2006]	2000	1° × 1°	BC, OC, SO_2, DU, SS, SOA, DMS	Eight anthropic sectors (including road traffic, off road, and shipping); + large-scale biomass burning (GFED 2000); + volcanoes (for SO_2)	Daily: DU, SS, DMS; monthly: biomass burning, SOA; yearly: others	Biomass burning (POM/BC/SO_2): six altitude regimes: 0–.1 km/.1–.5 km/.5–1 km / 1–2 km/2–3 km/3–6 km; SO_2 industry/power plant: from 100 to 300 m; SO_2 volcanic: continuous: 2/3 to 1/1 of volcano top; explosive: 0.5 to 1.5 km above volcanic top
Lamarque et al. [2010]	2000	0.5° × 0.5°	BC, OC, SO_2, CO, NH_3, NMVOC, NO_x,	Eight land-based anthropogenic sectors; + aircraft and ships; +2 biomass burning categories (ACCMIP-MACCity data set based on GFEDv2); no volcanic emissions	Monthly: biomass burning, soil NO_x, ship, and aircraft emissions; yearly: others (decadal intervals)	No vertical emission profile is provided
Diehl et al. [2012]	2006	1° × 1°	BC, OC, SO_2	Four land-based anthropic (residential, industry, power generation, and transport); + aircraft and ships; +2 biomass burning categories (based on GFEDv2); + volcanoes (for SO_2): explosive, effusive, and quiescent degassing events	Daily: volcanoes; monthly: biomass burning, aircraft; yearly: others; +recommended seasonality for SO_2 land-based emissions over Europe	Biomass burning: different suggestions as derived from literature SO_2 industry/power plant: from 0 to 500 m SO_2 volcanoes: volcanic plume heights from GVP database

layer product) that excludes cloudy layers, where aerosol layers cannot be detected by the lidar, as well as atmospheric levels below cloud and aerosol features that are optically thick (identified from the "layer opacity flag"). We also disregard vertical layers that are below the surface elevation at the lidar footprint. Elsewhere, at levels where the CALIOP 5 km Level 2 product reports neither cloud nor aerosol, the atmospheric level is assumed to have zero aerosol extinction. Under certain conditions, a "base extension" algorithm is applied during CALIOP Level 2 retrieval. In these cases, the aerosol retrieval stops 90 m above the surface in order to limit the contamination of the measured signal by the surface return [*Vaughan et al.*, 2004]. To reduce the impact of this feature on our mean aerosol vertical profiles, we set the aerosol layer base level at the surface whenever its reported elevation, in the CALIOP 5 km Level 2 product, is less than 10% of the aerosol layer depth. This procedure was shown to at least partly remove unrealistic low aerosol extinction at the surface K12. Each well-defined aerosol layer and aerosol-free layer is then split in 100 m height segments to allow for the calculation of the mean aerosol extinction profile.

In order to quantitatively assess the ability of the different models to reproduce the observed aerosol mean vertical distribution, the Z_α mean extinction height diagnostic (equation (1)) established by K12 is used:

$$Z_\alpha = \frac{\sum_{i=1}^{n} b_{\mathrm{ext},i} . Z_i}{\sum_{i=1}^{n} b_{\mathrm{ext},i}} \tag{1}$$

where n is the number of 100 m altitude layers, $b_{\mathrm{ext},i}$ is the aerosol extinction coefficient (km^{-1}) at level i, and Z_i is the altitude (km) above sea level of level i. It provides a useful and simple measure to gauge the performance of the models for different regions and seasons (at least in the case of monotonically decreasing aerosol vertical profiles), in terms of estimate of the typical mean height of the aerosol extinction. Because of CALIOP aerosol retrieval limitations, and notably to the fact that the retrieval is not sensitive to low aerosol loadings in the high troposphere [*Winker et al.*, 2009; *Winker et al.*, 2013], we use the $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ diagnostic only computed over the 0–6 km altitude range, as in K12.

While K12 analyzed mean aerosol profiles calculated over subcontinental regions, the averaging has been performed here at global scale onto the 1°×1° grid cells. Illustrations of the resulting AeroCom-CALIOP 3-D gridded product are provided in Figure 1. It shows the global distribution of the Z_α mean extinction height diagnostic computed (equation (1)) for the 2009 four seasons. Because Z_α is referenced to sea level, higher values are generally obtained over higher altitude regions. More interestingly, varying seasonal patterns of the aerosol mean extinction height show up, such as higher situated aerosol over the Northern Africa in Northern Hemisphere (NH)-summer (June–August (JJA)) and higher aerosol in the Southern Africa in Southern Hemisphere (SH)-summer (December-February (DJF)) season. The reliability of the AeroCom-CALIOP product was already assessed in K12. It was shown to provide a representative characterization of aerosol extinction typical heights at seasonal and subcontinental scales and a robust benchmark data set for the evaluation of the global atmospheric models. As mentioned previously, the CALIOP screening and averaging methods follow K12, except that the profiles are first aggregated and averaged onto a global 1°×1° latitude-longitude grid before the calculation of the regional mean. While the different averaging induces some differences in the magnitude of the mean aerosol extinction values over the selected regions shown in Figure 2 (see Figure S1a in the supporting information), it provides similar normalized profiles and $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ values (with differences $<5\%$) as shown in Figure S1b.

To further assess our CALIOP data processing and resulting global gridded product, we also compare the derived AOD and $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ values to the ones calculated from the NASA CALIOP Level 3 aerosol global data set (the so-called "NASA product," hereafter). In the latter study, different (e.g., in the low troposphere) or additional (e.g., based on quality flags) screenings are applied to the original CALIOP Layer data. For instance, to avoid underestimating the lowest part of the aerosol profile, when the base of the lowest detected aerosol layer is above the local surface but lower than 2.46 km, the NASA algorithm ignores the "clean air" parcel which is found between the surface and the base of the lowest aerosol layer [*Winker et al.*, 2013]. In contrast, we instead apply an extrapolation of the extinction from the aerosol layer base to the surface in cases of low

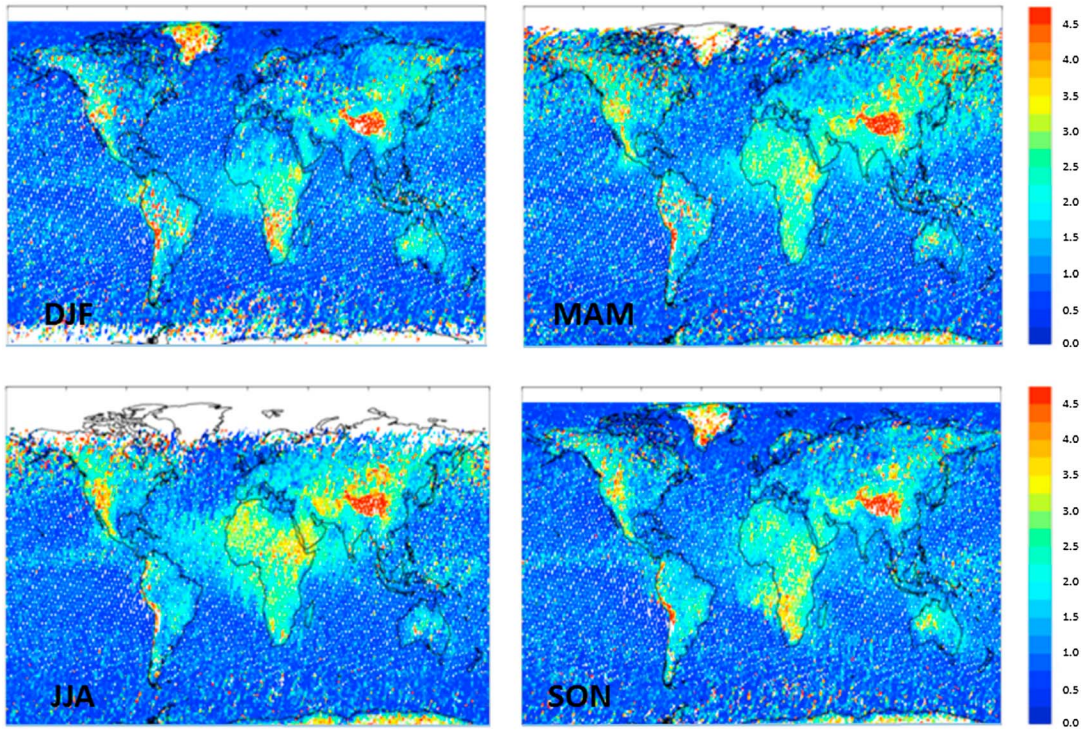

Figure 1. Illustration of the AeroCom-CALIOP 3-D global 1°×1° gridded product: 2009 $Z_{\alpha 0-6\,\mathrm{km}}$ mean seasonal aerosol extinction height (km) over the 0–6 km altitude range.

altitude aerosol layers (see K12). The AeroCom-CALIOP (1°×1° resolution) and NASA (2°×5° latitude-longitude resolution) global products show substantial differences in the shape of both the absolute and normalized mean regional extinction profiles (see Figures S2a and S2b in supporting information, respectively). Notably, and not surprisingly, higher aerosol extinction is obtained in the lowest kilometer (inducing lower $Z_{\alpha 0-6\,\mathrm{km}}$ values) when applying the NASA Level 3 processing, which can be at least partly explained the use of the CALIOP level 2 aerosol profile product rather than the level 2 aerosol layer product, which is used to build the AeroCom product. The level 2 aerosol profiles report retrieved aerosol extinction in the aerosol layer as a function of altitude, whereas the level 2 aerosol layer data report the average properties (e.g., AOD) for each aerosol layer detected. Since many aerosol extinction profiles increase near the surface, the aerosol

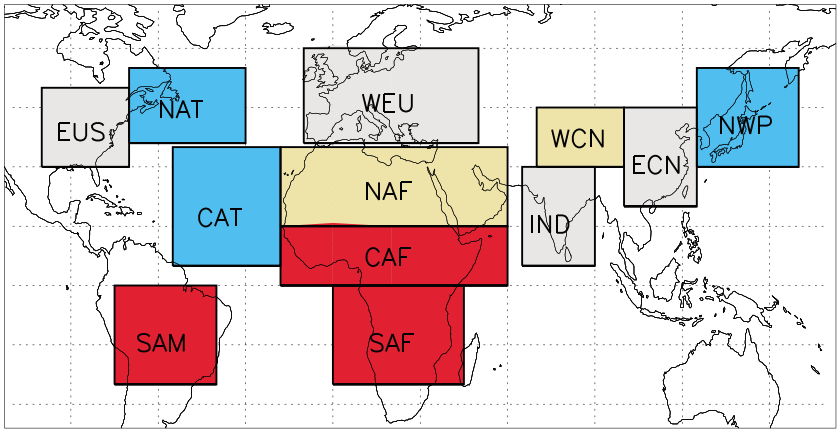

Figure 2. Location of the 12 studied regions. These regions have been selected because of the dominance of marine aerosols (blue), industrial (grey), dust (yellow), and biomass burning (red) aerosols. Adapted from *Koffi et al.* [2012].

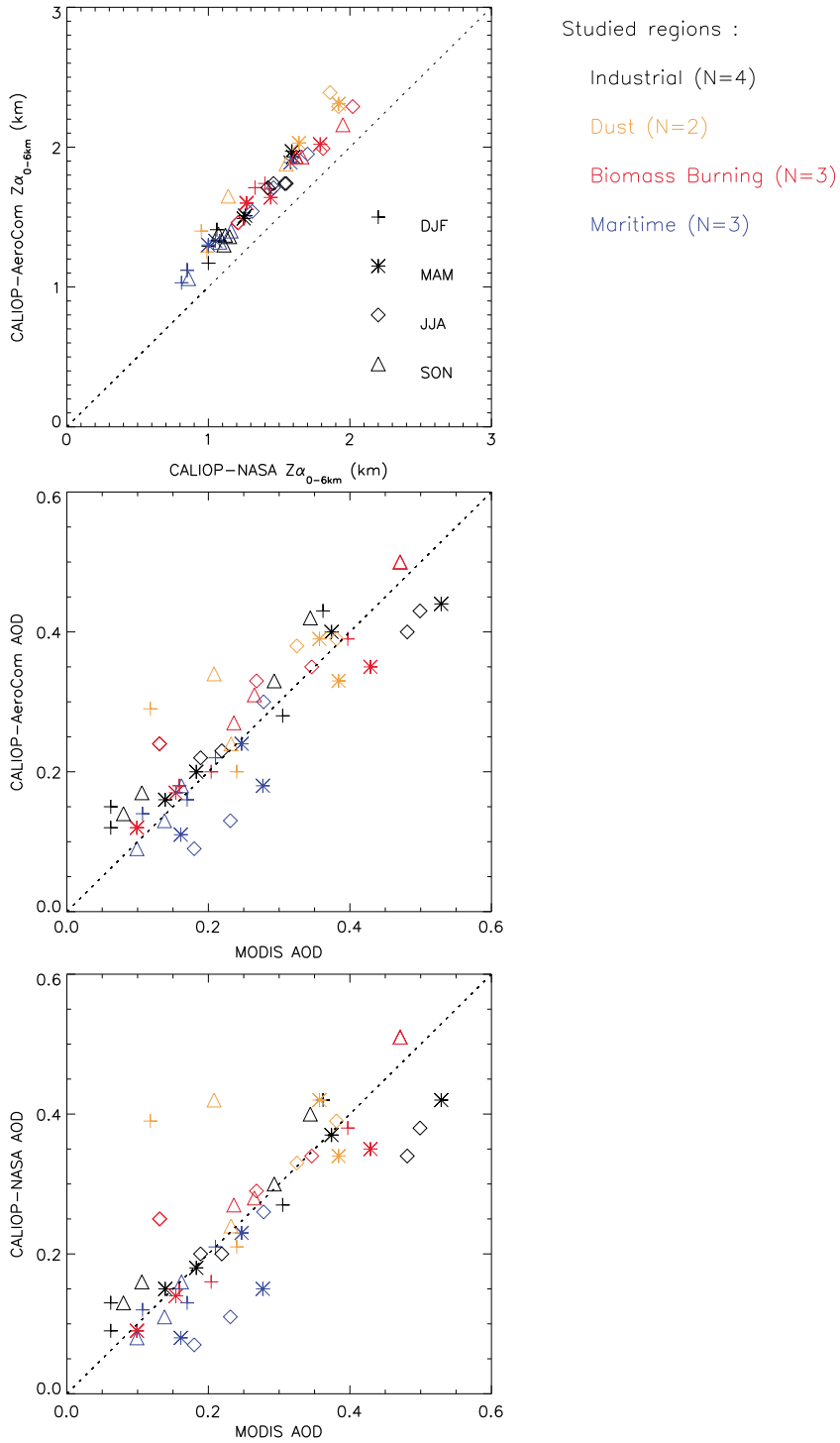

Figure 3. The 2007 $Z_{\alpha 0-6\,\mathrm{km}}$ mean seasonal extinction height over the 12 selected regions: (top) AeroCom (this study) versus NASA [*Winker et al.*, 2013] 2007 CALIOP gridded products. The 2007 mean seasonal AOD, over the 12 selected regions as derived from the (middle) AeroCom (this study) and (bottom) NASA [*Winker et al.*, 2013] CALIOP gridded products, compared to 2007 MODIS measurements.

extinction in the layer product is overestimated in the higher part of the aerosol layer and underestimated at lower altitudes. This feature, which is common to the averaged aerosol properties computed in the different model levels, can induce significant differences in the lowest part of the atmosphere compared to profiles derived from the CALIOP level 2 profile data. Nevertheless, despite a resulting $Z_{\alpha 0-6\,\mathrm{km}}$ difference of about

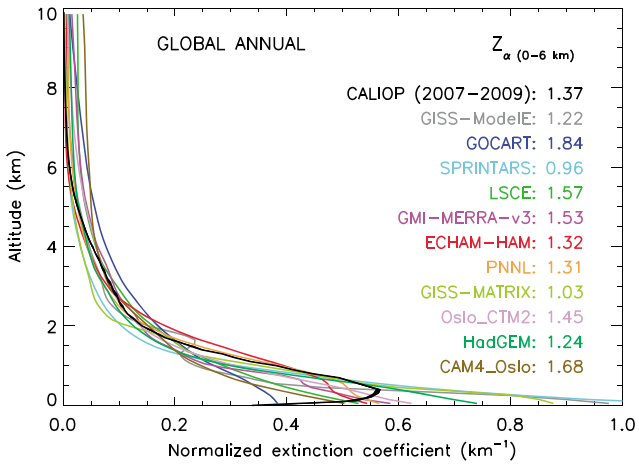

Figure 4. Models (2006) and CALIOP (2007–2009) global mean annual "normalized" extinction coefficient (km^{-1}) profiles (at 550 and 532 nm, respectively). The 2007 to 2009 range is shown for CALIOP (black). The simulated and observed $Z_{\alpha 0-6\ \mathrm{km}}$ mean extinction heights (km) are also reported. See Tables 1a, 1b, 1c for the definition of the models.

0.2 km between the two products (after interpolating the NASA profiles at the same 100 m vertical resolution than the AeroCom-CALIOP product), the interregional diversity appears to be very similar and strongly correlated (Figure 3, top). In Figure 3, we also compare the respective CALIOP-derived AOD to the MODIS measurements described in section 2.2. As K12, the present study focuses on subcontinental and seasonal mean aerosol distributions, so no exact colocation between MODIS and CALIOP overpasses is considered. Figure 3 (middle) shows a good agreement of the AeroCom-CALIOP product with the MODIS AOD over the four seasons, which is also true for the NASA product (Figure 3, bottom).

In the present study, we also evaluate the AeroCom models over land and ocean in Northern and Southern Hemispheres (10° latitude bands from 70°S to 70°N). Coherent and consistent seasonal AOD are generally obtained (see Figure S3 in the supporting information) from the AeroCom-CALIOP product as compared to MODIS measurements: As expected, for both data sets, a higher AOD latitudinal variability appears in all seasons over land than over ocean. Also expected is the larger south to north AOD gradient obtained during the March-April-May (MAM) and JJA seasons compared to the other seasons, both over land and ocean. Over ocean, the AOD derived from the AeroCom-CALIOP product is significantly correlated ($0.625 < r < 0.824$; $p < 0.05$) with the MODIS data in all seasons but tends to be underestimated in the Northern hemisphere during the MAM and JJA seasons. Over land, a particularly good agreement is obtained between the two data sets for the MAM and JJA seasons both in terms of correlation ($r > 0.900$) and accuracy, whereas no correlation with MODIS is obtained for the December-January-February (DJF) season. In addition to the present evaluation analysis that concerns the total aerosol, *Kim et al.* [2014] showed that the dust fraction of the aerosol, which is also part of the AeroCom-CALIOP product, generally agrees in magnitude, distribution and seasonality with that from MODIS, MISR, SeaWiFS, and AIRS observations over North Africa and the North Atlantic.

3.2. The Evaluation of the Aerosol Vertical Distribution in the AeroCom Models

The model-derived mean aerosol profiles and extinction height are calculated from the 3-D AOD fields in a way analogous to the CALIOP-derived products: horizontal averaging of the models' cells within the selected regions, after converting the pressure levels to a 100 m resolution altitude vertical grid. While the regional averaging does not allow an exact spatial and temporal match between the model cells and CALIOP overpasses, it provides a complete image of the simulated climatology of the tropospheric aerosol distribution over a given region. A sensitivity test conducted by K12 showed little change in the mean vertical profiles when sampling the model at the CALIPSO observation location versus a full regional average, indicating that the CALIOP coverage is adequate to evaluate the mean aerosol climatology on seasonal time scale over subcontinental areas. Other sensitivity analyses on our CALIOP/model data processing and comparison method were performed in K12 (e.g., nighttime versus 24 h data and all-sky versus cloud free data). In the present study, the impact of vertical resolution, cloud versus aerosol discrimination, and processing of the lower troposphere is further discussed in section 5.

As an illustration of the evaluation approach, the models (2000 or 2006) and CALIOP-derived (2007–2009) mean global annual aerosol extinction profiles are shown in Figure 4. Because our main purpose is to evaluate the shape of the profiles and the $Z_{\alpha 0-6\ \mathrm{km}}$ typical mean extinction heights (rather than the amplitudes of the extinction), the profiles are normalized to a common AOD (= 1) over the 0–10 km altitude range, which allows an easier visual comparison. A substantial intermodel diversity is observed, with significant negative

(e.g., GISS-ModelE, GISS-MATRIX, and SPRINTARS), and positive (GOCART and CAM4-Oslo) bias in the mean global $Z_{\alpha 0-6\,\mathrm{km}}$ diagnostic. As explained in K12 and shown in this figure, the extrapolation used in our CALIOP processed product for altitudes near the surface (see section 3.1) does reduce but does not totally remove the observed low mean extinction near the surface.

While the aerosol optical depth comparisons discussed in section 3.1 give further confidence in CALIOP-derived extinction profiles, it is worth noting that the AOD, which corresponds to the column integrated aerosol extinction, and the $Z_{\alpha 0-6\,\mathrm{km}}$ mean extinction height are two different aerosol diagnostics that are not necessarily linked. In the case of our 12 studied regions for instance, the scatterplots between the seasonal $Z_{\alpha 0-6\,\mathrm{km}}$ and AOD CALIOP-derived diagnostics (see Figure S4 in the supporting information) show significant correlations ($p < 0.05$) for the JJA ($r = 0.580$) and MAM seasons ($r = 0.763$), whereas the two variables are poorly to very poorly correlated in the September–November (SON) ($r = 0.546$) and DJF ($r = 0.320$) seasons, respectively. Because the main objective of this study is the evaluation of the shape of the vertical profile as derived from normalized aerosol extinction profile (see section 1), AeroCom model versus CALIOP AOD comparisons are mainly provided in the supporting information.

Most of the AeroCom II simulations used 2000 or 2006 emissions and 2006 meteorology (see Table 1b), whereas CALIOP measurements are only available from June 2006 onward. In order to check if 2006 can be considered as a "normal" year relative to the 2007–2009 observations, we also calculated 2006 CALIOP AOD and $Z_{\alpha 0-6\,\mathrm{km}}$ mean diagnostics for the JJA and SON seasons for 12 out of the 13 K12 subcontinental regions, i.e., excluding the so-called SEA (Southeastern Asian) region, which we now consider to be unsuited as region dominated by biomass burning emissions (Figure 2). The 2006 values are within the range ±10% (AOD) and ±5% ($Z_{\alpha 0-6\,\mathrm{km}}$) of the 2007–2009 values and show very similar seasonal and spatial patterns compared to the 2007–2009 CALIOP observations (supporting information Figure S5).

4. Results

4.1. AeroCom II Versus AeroCom I Performance Over Selected Regions

The shape of the mean annual AeroCom II extinction profiles are compared to AeroCom IA in section 4.1.1, whereas the quantitative assessment of the simulated annual and seasonal $Z_{\alpha 0-6\,\mathrm{km}}$ mean extinction heights is provided in section 4.1.2. First conclusions for the factors responsible for the models' biases and changes are further discussed in section 5.

4.1.1. Comparison of the Simulated and Observed Extinction Profiles

The main differences between the shapes of AeroCom IA (see Figure 5 of K12) and AeroCom II (Figure 5) mean annual regional aerosol extinction profiles are summarized hereafter:

1. *General shape*. Similar to AeroCom Phase I, most of the AeroCom II models reproduce the observed mean annual profiles, which are characterized by a decrease of the extinction coefficient magnitude from the surface to about 5 km. Different typical shapes are observed according to the type of the region, with a more abrupt decrease from the surface to the high troposphere for the marine (e.g., NWP and NAT) than for the biomass burning and dusty (e.g., NAF and WCN) regions. The profiles simulated by the GISS-ModelE show a bimodal shape. The local all-sky AOD maximum at the cloud level in GISS-modelE is caused by AOD from sulfate aerosols. In-cloud aerosol formation at these altitudes leads to enhanced AOD in that model. For an accurate evaluation between the model and a satellite retrieval, the clear-sky AOD needs to be used [*Kim et al.*, 2014], which was not available in three dimensions for this study.
2. *Dusty regions*. AeroCom IA models generally reproduced the general shape of the aerosol vertical distribution over the African (NAF) and Chinese (WCN) dust regions, as well as in the Indian (IND) industrial region. AeroCom II models still perform relatively well over India but generally overestimate the fraction of the aerosol extinction below 1.5 km altitude over the dust regions. This is particularly true for two models: the Oslo-CTM2 model, which shows a $Z_{\alpha 0-6\,\mathrm{km}}$ bias of −0.60 km over the WCN region and the CAM4_Oslo model, which shows a $Z_{\alpha 0-6\,\mathrm{km}}$ bias of −0.77 km over the NAF region.
3. *Upper troposphere*. The tendency of a few AeroCom IA models to overestimate the aerosol extinction above 4 to 6 km is confirmed in the case of some AeroCom II models (e.g., CAM4-Oslo, GOCART, LSCE, Oslo-CTM2, and GMI-MERRA), whereas others (SPRINTARS, GISS-MATRIX, and HadGEM) show a higher than observed aerosol fraction in the lowest part of the troposphere.

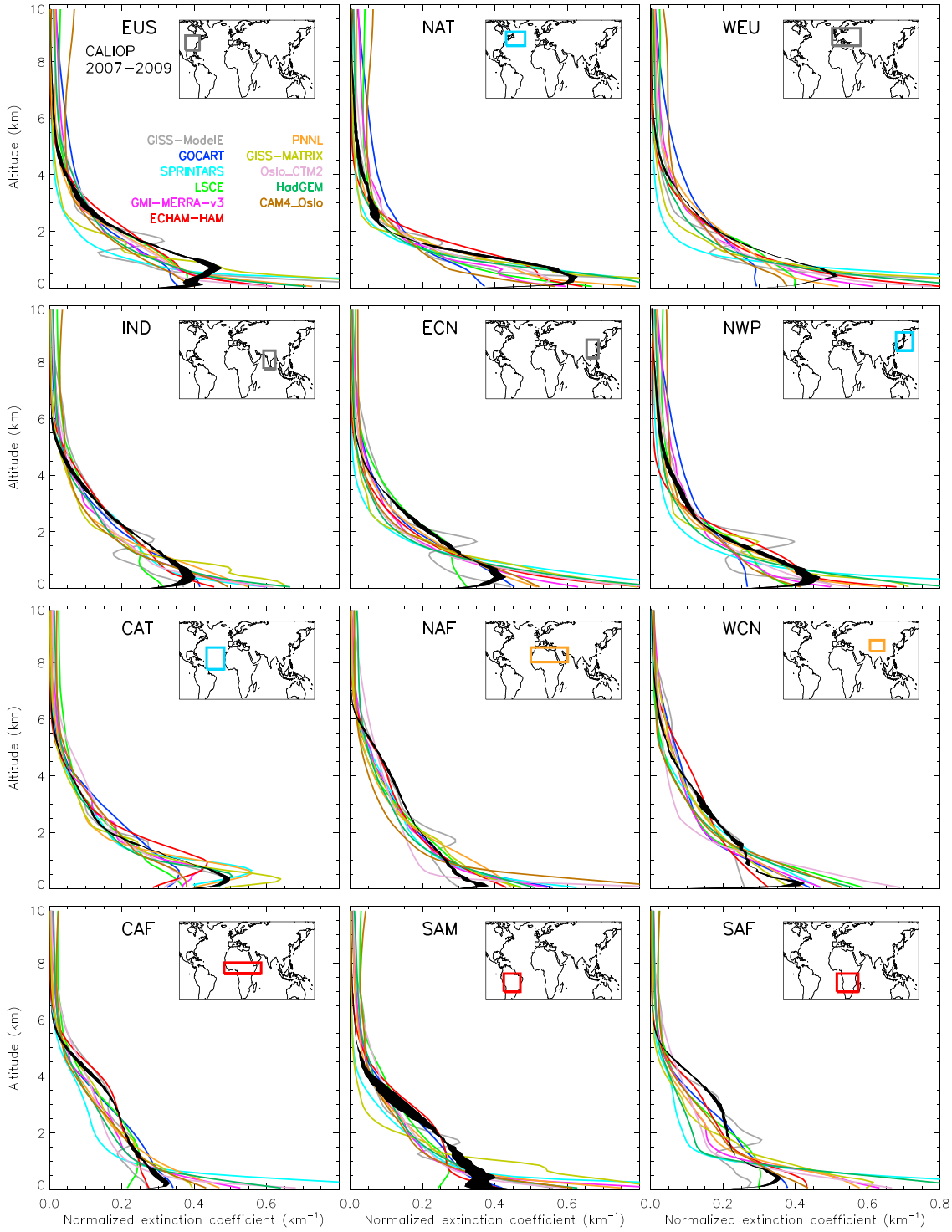

Figure 5. Mean annual normalized extinction coefficient (km^{-1}) profiles over the 12 selected regions from the AeroCom II model simulations (2000 and 2006) and the AeroCom CALIOP gridded product (2007–2009), at 550 and 532 nm, respectively. The 2007 to 2009 range (mean ± SD) is shown for CALIOP-derived profiles (black).

4. *Marine downwind regions*. Large discrepancies between observed and simulated aerosol extinction profiles over the Atlantic (NAT and CAT) and Pacific (NWP) downwind maritime regions were found during AeroCom-IA intercomparison exercise. In AeroCom II, these results were modestly improved in the NWP region and much improved in the CAT region, at least below 4 km of altitude.

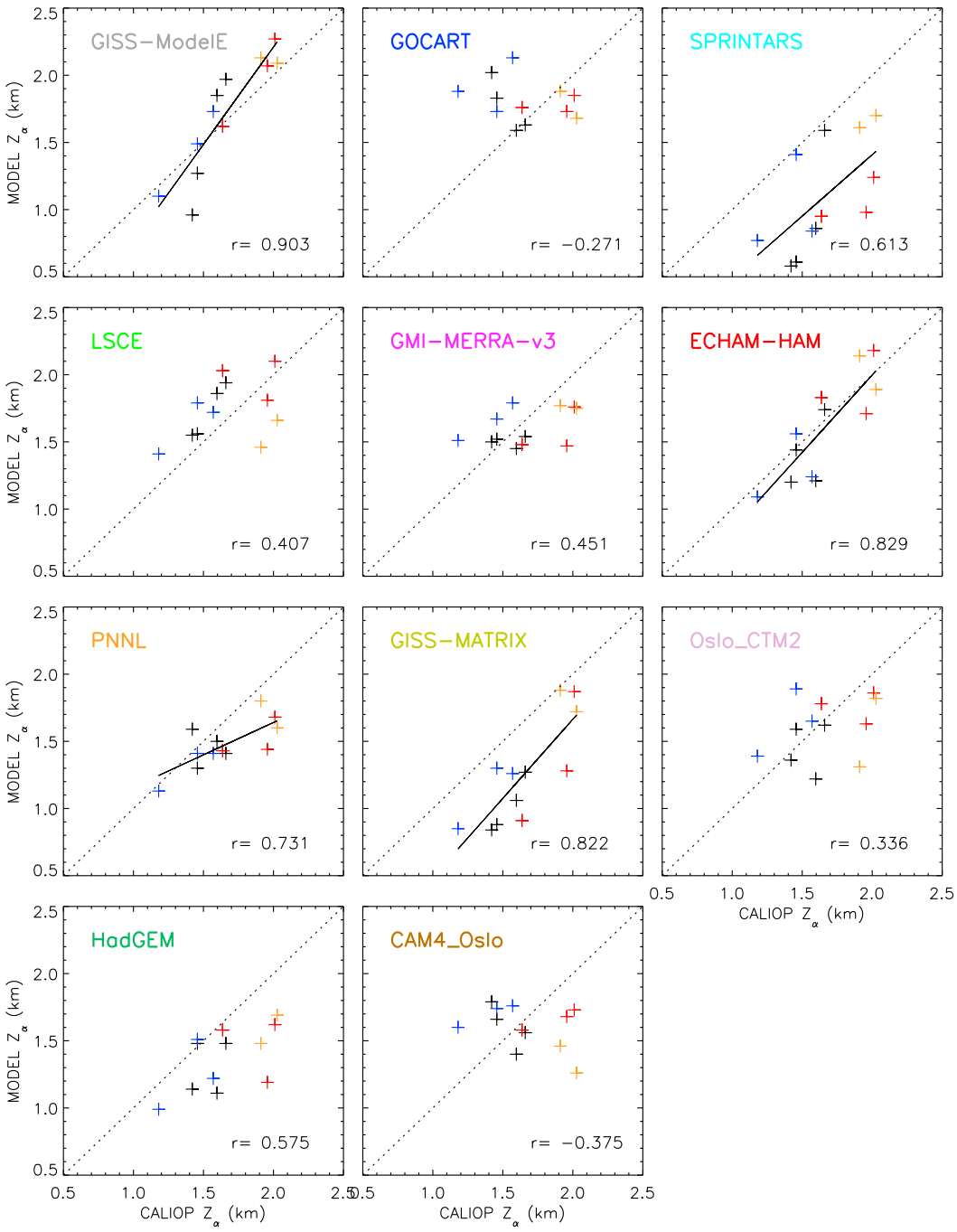

Figure 6. Models (2006) versus CALIOP (2007–2009) mean annual $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ aerosol extinction height over the 12 selected regions (see Figure 2 for the definition of the domains and the color legend). Significant ($p < 0.05$) regression lines are plotted.

5. *Biomass burning dominated regions.* K12 showed the difficulty encountered by the previous version of the models to reproduce the specific shape of the aerosol vertical distribution over the SAM, SAF, and CAF regions. With the exception of the SPRINTARS model, the AeroCom II models' performance over the CAF biomass burning region seems qualitatively similar, in annual mean, to the performance of AeroCom IA. Results are worse over the SAF region where the AeroCom II models either still do not reproduce or underestimate even more, the convex character of the observed mean profile. Conversely, the performance is slightly improved and the intermodel range reduced compared to AeroCom IA over South America.

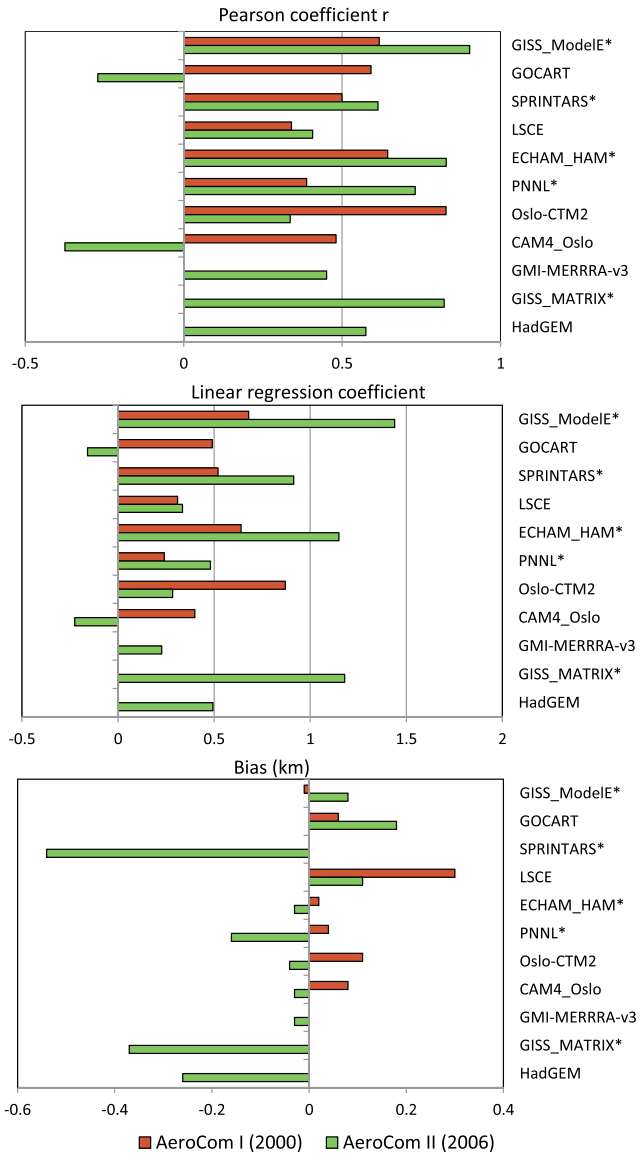

Figure 7. Comparison of AeroCom I (2000) and AeroCom II (2006) models performance in reproducing CALIOP (2007–2009) mean annual $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ extinction height over the 12 selected regions: (a) Pearson coefficient; (b) linear regression coefficient, and (c) bias (km). Asterisks indicate models showing a significant positive correlation with CALIOP ($p < 0.05$).

4.1.2. Comparison of the Simulated and Observed $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ Diagnostics

4.1.2.1. Annual Mean Distributions

Figure 6 compares the AeroCom II and CALIOP mean annual $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ diagnostics over the 12 regions. A significant correlation ($p < 0.05$) with CALIOP observed interregional variability in $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ is only obtained for 5 (GISS-ModelE, SPRINTARS, ECHAM-HAM, PNNL, and GISS-MATRIX) out of the 11 models. The simulation of $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ interregional variability is improved for five models (GISS-ModelE, SPRINTARS, LSCE, ECHAM-HAM, and PNNL), whereas it is deteriorated for three of them (GOCART, Oslo-CTM2, and CAM4-Oslo) compared to AeroCom IA experiment (Figure 7). The best correlation with observations is obtained by the GISS-ModelE ($r = 0.903$) and ECHAM-HAM (0.829) models, which were already among the best performing models in the AeroCom IA experiment. Despite the unusual shape of the aerosol profiles simulated by the GISS-ModelE, the $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ bias (+80 m) and mean absolute error (<200 m) are still rather low. On the other hand, while good ($r = 0.591$) and very good ($r = 0.828$) correlations were obtained in AeroCom IA by the GOCART and Oslo-CTM2 models, respectively, their phase II model versions and simulated $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ diagnostics are not anymore correlated with CALIOP interregional variability. In the case of CAM4_Oslo, the correlation with observations is still poor, as for the first phase of AeroCom.

It is worth noting that the $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ mean annual bias decreased and changed from +75 m to −53 m in model mean between the two studies (eight common models). This result highlights a tendency that models now simulate higher than observed aerosol fraction in the lowest part of the troposphere (see section 4.1.1). This is particularly true for the SPRINTARS model that now simulates $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ values about 500 m below the observed ones. A similar but less pronounced negative bias (≈370 m) is also obtained for the GISS-MATRIX model. The two other newly analyzed models (HadGEM and GMI-MERRA) have little capacity to reproduce the $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ interregional variability. The HadGEM model also has a tendency to underestimate the mean extinction height over most of the studied regions (Figure 6), which results in a $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ mean absolute error of about 300 m (Figure 7). While the uncertainties in CALIOP-derived $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ values (such as the 200 m difference between the AeroCom and NASA-derived $Z_{\alpha 0\text{–}6\,\mathrm{km}}$ diagnostics due to differences in the CALIOP data postprocessing) might partly contribute to some of the models biases, they cannot explain missing regional variability in $Z_{\alpha 0\text{–}6\,\mathrm{km}}$.

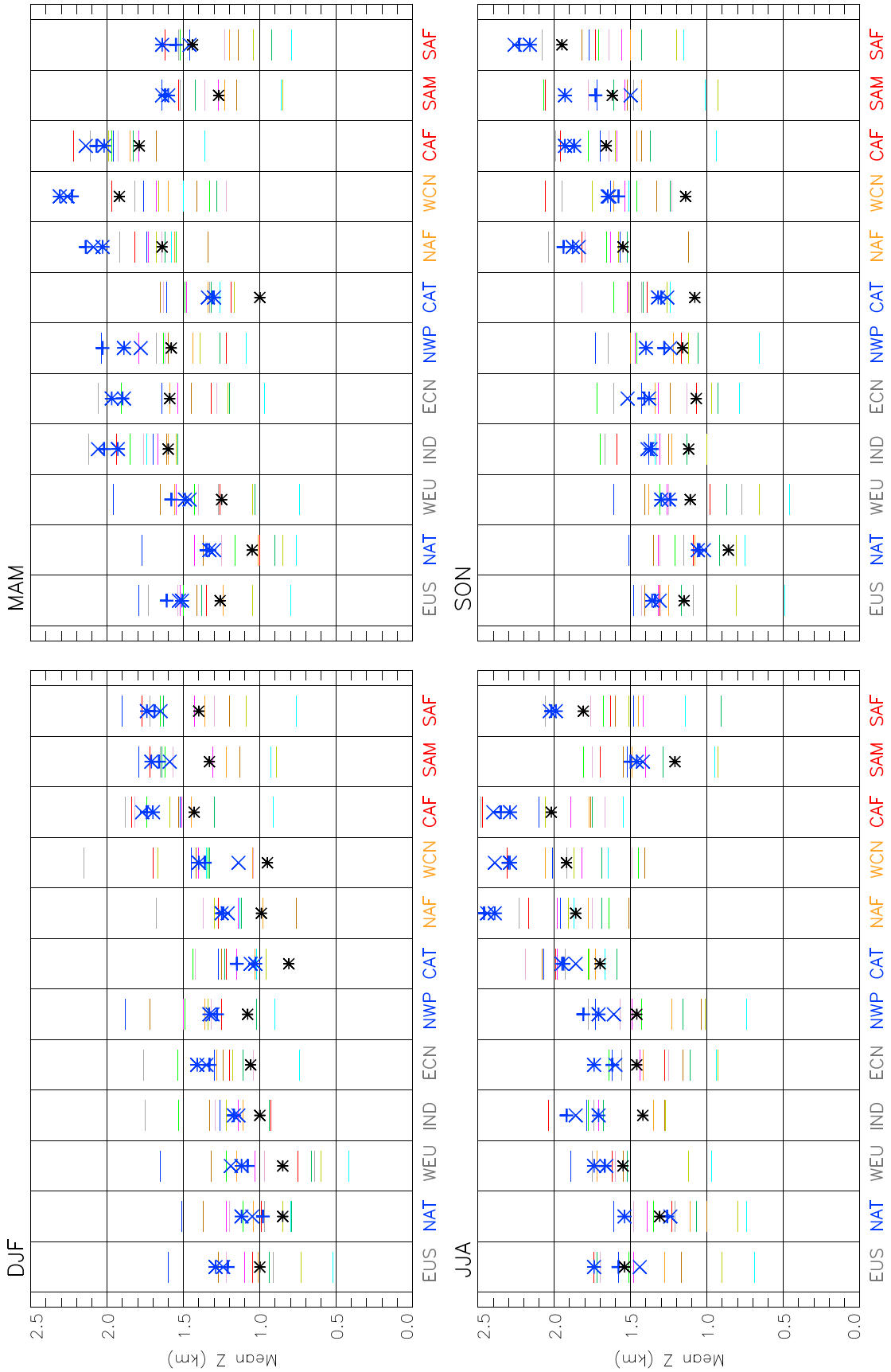

Figure 8. Models (2006) and CALIOP (2007–2009) mean seasonal $Z_{\alpha 0-6\ \mathrm{km}}$ extinction height over the 12 selected regions. CALIOP $Z_{\alpha 0-6\ \mathrm{km}}$ is indicated by an asterisk.

Table 2. AeroCom I (I) and AeroCom II (II) mean model (8 common models) bias (%) in the seasonal $Z_{\alpha\ 0\text{-}6\ km}$ mean height diagnostic compared to 2007-2009 CALIOP observations. The AeroCom Phase (I or II) with lowest bias in the model mean is shown in bold

	MAM		JJA		SON		DJF		*Mean*	
AeroCom	I	II	I	II	I	II	I	II	*I*	*II*
NAT	+10.9	**-2.0**	+18.3	**+6.6**	+22.3	**+21.3**	+26.8	**+9.2**	*+20.5*	***+8.8***
CAT	**-8.0**	+9.6	-16.6	**+8.2**	**+2.3**	+21.7	+31.4	**+16.8**	***+2.3***	*+14.1*
NWP	-15.6	**-5.3**	**+0.9**	-9.4	**+8.0**	+16.2	+27.2	**+12.6**	*+5.1*	***+3.5***
EUS	+18.1	**+0.6**	+12.0	**+7.5**	-11.1	**-3.1**	+27.8	**-9.6**	*+11.4*	***-1.2***
WEU	+20.2	**+0.4**	+18.7	**+2.0**	+22.9	**-7.6**	**-8.2**	-16.4	*+13.4*	***-5.4***
IND	**+5.5**	-10.8	+29.7	**-6.0**	+18.5	**+3.5**	+21.6	**+13.6**	*+18.8*	***+0.1***
ECN	**-9.3**	-18.0	**+0.9**	-13.0	+45.3	**-15.9**	**+3.2**	-6.4	***+10.0***	*-13.3*
SAM	+21.3	**-15.4**	+10.0	**+5.7**	**+3.7**	+12.1	+21.4	**-4.1**	*+14.1*	***-0.4***
CAF	-13.8	**-7.5**	-21.8	**-12.9**	-39.0	**-10.7**	**-4.7**	-8.9	*-19.8*	***-10.0***
SAF	+20.5	**-11.4**	**-6.8**	-23.4	**-3.5**	-24.4	+31.1	**-7.4**	***+10.3***	*-16.7*
NAF	-21.3	**-17.3**	**+4.6**	-18.3	**-4.2**	-7.3	+11.6	**+2.0**	***-2.3***	*-10.2*
WCN	**-6.5**	-26.2	**-4.0**	-18.1	+18.2	**+0.8**	**+28.2**	+31.7	*+9.0*	***-3.0***
Mean	***+1.8***	*-8.6*	***+3.8***	*-5.9*	*+6.95*	***+0.6***	*+18.4*	***+2.8***	*+9.4*	***-2.8***

4.1.2.2. Seasonal Mean Distributions

The 2007–2009 CALIOP benchmark data set reveals a moderate (e.g., industrial pollution source regions) to very pronounced (e.g., Atlantic Ocean downwind the Sahara) seasonality of the aerosol vertical distribution (see K12). Figure 8 presents the model spread in simulating the $Z_{\alpha 0\text{–}6\ km}$ seasonal diagnostic over the 12 selected regions. The $Z_{\alpha 0\text{–}6\ km}$ seasonal values derived from the AeroCom-CALIOP product for the years 2007, 2008, and 2009 are reported. It shows that the model bias toward observations depends highly on the season, with the most systematic biases observed during the MAM and JJA seasons. During these seasons, models tend to underestimate the mean aerosol height over industrial (ECN) and maritime (NWP) regions, as well as over dust (NAF and WCN) and biomass burning (CAF and SAF) source regions. For the dusty regions: While all (for WCN) but one (for NAF) of the models capture the $Z_{\alpha 0\text{–}6\ km}$ seasonal peak in the JJA or MAM seasons, its magnitude is underestimated by all the models compared to the AeroCom-CALIOP diagnostic. The important $Z_{\alpha 0\text{–}6\ km}$ peak occurring during the biomass burning season in the CAF (JJA) and SAF (SON) African regions is also underestimated by all but one model. Very similar but less pronounced patterns were obtained for AeroCom IA simulations over these regions (see Figure 9 of K12). As for AeroCom IA, for almost all the models, there is no negative bias of $Z_{\alpha 0\text{–}6\ km}$ over the downwind Atlantic region (CAT) despite the negative bias over Northern Africa. This unexpected result was explained in K12 by model underestimation of the aerosol extinction magnitude within the maritime boundary layer, which compensates for the too low transport height of African dust. The 2007 seasonal $Z_{\alpha 0\text{–}6\ km}$ mean extinction heights computed from the

Table 3. AeroCom I (I) and AeroCom II (II) Mean Model (Eight Common Models) Bias (%) in the Seasonal AOD Compared to 2007–2009 CALIOP Observations[a]

	MAM		JJA		SON		DJF		Mean	
AeroCom	I	II	I	II	I	II	I	II	I	II
NAT	**+39.8**	+56.8	**−22.2**	+132.1	**+47.1**	+61.1	**+20.8**	+32.1	***+21.4***	*+70.5*
CAT	−12.5	**+4.6**	+8.1	**−1.0**	−39.1	**−15.3**	+55.4	**+29.2**	***+3.0***	*+4.4*
NWP	**+1.8**	+78.7	**−31.2**	+172.9	**+66.3**	+73.1	**−10.0**	+17.6	***+6.7***	*+85.6*
EUS	**−35.6**	+42.9	**−34.1**	+67.9	**−10.6**	+122.7	**+20.8**	+80.8	***−14.9***	*+78.6*
WEU	+96.9	**+80.6**	**+29.5**	+66.4	**−5.7**	+140.0	**+59.1**	+180.4	***+44.9***	*+116.8*
IND	+52.3	**−32.7**	+98.2	**−9.7**	+61.1	**−15.1**	−45.1	**−31.5**	*+41.6*	***−22.2***
ECN	**−47.4**	+89.7	**+55.8**	+119.4	**+67.2**	+127.6	**−16.1**	+85.6	***+14.9***	*+105.6*
SAM	−36.5	**+25.0**	**+11.5**	+49.0	−31.8	**−21.6**	−44.2	**−8.1**	*−25.2*	***+11.1***
CAF	+40.0	**+17.8**	−14.4	**+9.6**	+48.9	**−0.4**	**+18.9**	+40.2	*+23.3*	***+16.8***
SAF	**+8.8**	−10.8	−24.7	**−23.9**	**+20.3**	−37.1	**+5.8**	+22.7	***+2.5***	*−12.3*
NAF	**−49.6**	+69.3	+115.0	**+48.8**	**−3.6**	+32.8	**+42.4**	+67.1	***+26.0***	*+54.5*
WCN	**+37.5**	+104.3	**+8.9**	+194.3	**−34.4**	+76.2	−70.0	**+0.8**	***−14.5***	*+93.9*
Mean	**+8.0**	+43.9	**+16.7**	+68.8	**+15.5**	+45.3	**+3.2**	+43.1	***+10.8***	*+50.3*

[a]The AeroCom Phase (I or II) with lowest bias in the model mean is shown in bold.

Table 4a. AeroCom I (12 Models)[a]

	CALIOP	**GISS**[b]	**GOCART**[b]	**SPRINTARS**[b]	**LSCE**[b]	MATCH	**ECHAM-HAM**[b]	**PNNL**[b]	MOZGN	**Oslo-CTM2**[b]	TM5	**CAM4-Oslo**[b]	ARQM
EUS	JJA/MAM	**MAM**	DJF	**JJA**	**JJA**	**JJA**	**JJA**	**MAM**	**MAM**	**MAM**	**JJA**	**JJA**	**JJA**
NAT	JJA/MAM	**JJA**	**MAM**	**JJA**	SON	**JJA**	**JJA**	**JJA**	**MAM**	**JJA**	**JJA**	DJF	*DJF*
WEU	JJA	**JJA**	MAM	MAM	**JJA**	**JJA**	**JJA**	MAM	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**
IND	MAM	**MAM**	**MAM**	**MAM**	*JJA*	JJA	JJA	JJA	**MAM**	**MAM**	JJA	SON	JJA
ECN	MAM	**MAM**	**MAM**	**MAM**	*JJA*	JJA	JJA	JJA	**MAM**	**MAM**	JJA	*SON*	**MAM**
NWP	MAM	**MAM**	DJF	**MAM**	*SON*	*JJA*	*JJA*	**MAM**	**MAM**	**MAM**	SON	*SON*	*DJF*
CAT	JJA	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	*DJF*	JJA
NAF	JJA	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	MAM	JJA
WCN	JJA	*MAM*	**JJA**	MAM	*DJF*	**JJA**	**JJA**	**JJA**	**JJA**	MAM	**JJA**	MAM	*DJF*
CAF	JJA	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	MAM	**JJA**	MAM	**JJA**	**JJA**	**JJA**
SAM	MAM/SON	DJF	**SON**	**SON**	JJA	DJF	**SON**	**SON**	**SON**	**SON**	DJF	**SON**	**SON**
SAF	SON	*DJF*	**SON**	**SON**	*DJF*	**SON**	DJF	**SON**	**SON**	JJA	**SON**	**SON**	**SON**

[a]Performance of AeroCom I Model at Capturing the Peak Season of CALIOP $Z_{\alpha 0-6\,\mathrm{km}}$ Over the 12 Regions Shown in Figure 2. Observed CALIOP peak season is indicated in the second column. If the model agrees with the peak season, then its entry is highlighted in bold; if the model simulates CALIOP peak season as the second highest, model entry is in normal font; else, if the model simulates the peak season as the third or fourth highest, the entry is italics.
[b]The eight models that participated in both AeroCom I and II are highlighted in bold.

NASA product (Figure 8) show a very similar interregional and interseasonal diversity but lower values than the ones derived from the CALIOP-AeroCom product, especially in the dust regions. These differences, which are due to the differences of the aerosol profiles in the lowest troposphere (see section 3.1 and Figure S2), do not affect the main results and above conclusions. Furthermore, according to our preliminary analysis, the recent corrections and updates in the NASA product tend to reduce the $Z_{\alpha 0-6\,\mathrm{km}}$ differences with the AeroCom-CALIOP product (see section 5.3).

The results in model ensemble mean bias (eight common models) between the two phases of AeroCom as compared to the AeroCom-CALIOP product are improved in all seasons and more particularly during the SON and DJF seasons (Table 2). The highest $Z_{\alpha 0-6\,\mathrm{km}}$ relative biases are now obtained for the MAM and JJA seasons. The largest improvement occurs over the industrial EUS and its downwind NAT regions. In these two regions, the $Z_{\alpha 0-6\,\mathrm{km}}$ seasonal bias is reduced in all seasons, resulting in a Normalized Mean Absolute Error (NMAE) of less than 8% on average. Results are also improved compared to AeroCom IA K12 for the IND (NMAE ~ 9%) and WEU (NMAE ~ 11%) industrial regions and to a lower extent for the CAT and NWP maritime ones. In the other regions (ECN, SAF, NAF, and WCN), the NMAE is increased (by 6% for ECN to 39% WCN), due to the underestimation of $Z_{\alpha 0-6\,\mathrm{km}}$ in all seasons (except in DJF for WCN). Despite the fact that the $Z_{\alpha 0-6\,\mathrm{km}}$ and AOD aerosol properties are somewhat intercorrelated (section 3.2), no clear relation is found between the $Z_{\alpha 0-6\,\mathrm{km}}$ (Table 2) and AOD (Table 3) mean model biases, which increased by a factor of about 3 and decreased by a factor of 5 in average over the 12 selected regions, respectively. Similarly, no correlation is found between $Z_{\alpha 0-6\,\mathrm{km}}$ and AOD individual models' biases (see Table S1 in supporting information), neither

Table 4b. AeroCom II (11 Models)

	CALIOP	**GISS**[b]	**GOCART**[b]	**SPRINTARS**[b]	**LSCE**[b]	GMI-MERRA	**ECHAM-HAM**[b]	**PNNL**[b]	GISS-MATRIX	**Oslo-CTM2**[b]	HadGEM	**CAM4-Oslo**[b]
EUS	JJA/MAM	**JJA**	**MAM**	**MAM**	**JJA**	**MAM**	**JJA**	**JJA**	**MAM**	**JJA**	**JJA**	**MAM**
NAT	JJA/MAM	**JJA**	**MAM**	DJF	**JJA**	**MAM**	**JJA**	**MAM**	DJF	**JJA**	**JJA**	DJF
WEU	JJA	**JJA**	MAM	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	MAM
IND	MAM	**MAM**	JJA	**MAM**	**MAM**	JJA	JJA	**MAM**	**MAM**	**MAM**	JJA	**MAM**
ECN	MAM	**MAM**	**MAM**	**MAM**	**MAM**	**MAM**	**MAM**	**MAM**	**MAM**	**MAM**	**MAM**	**MAM**
NWP	MAM	JJA	**MAM**	**MAM**	**MAM**	**MAM**	*DJF*	**MAM**	**MAM**	**MAM**	**MAM**	DJF
CAT	JJA	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**
NAF	JJA	**JJA**	**JJA**	**JJA**	SON	**JJA**	**JJA**	**JJA**	**JJA**	SON	**JJA**	**JJA**
WCN	JJA	*DJF*	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	*DJF*
CAF	JJA	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	**JJA**	MAM	**JJA**	*MAM*	MAM	**JJA**
SAM	MAM/SON	DJF	DJF	**SON**	**SON**	**SON**	**SON**	JJA	JJA	**SON**	DJF	JJA
SAF	SON	**SON**	DJF	**SON**	**SON**	**SON**	DJF	**SON**	JJA	JJA	DJF	**SON**

[a]Performance of AeroCom II Model at Capturing the Peak Season of CALIOP $Z_{\alpha 0-6\,\mathrm{km}}$ Over the 12 Regions Shown in Figure 2. Observed CALIOP peak season is indicated in the second column. If the model agrees with the peak season, then its entry is highlighted in bold; if the model simulates CALIOP peak season as the second highest, model entry is in normal font; else, if the model simulates the peak season as the third or fourth highest, the entry is italics.
[b]The eight models that participated in both AeroCom I and II are highlighted in bold.

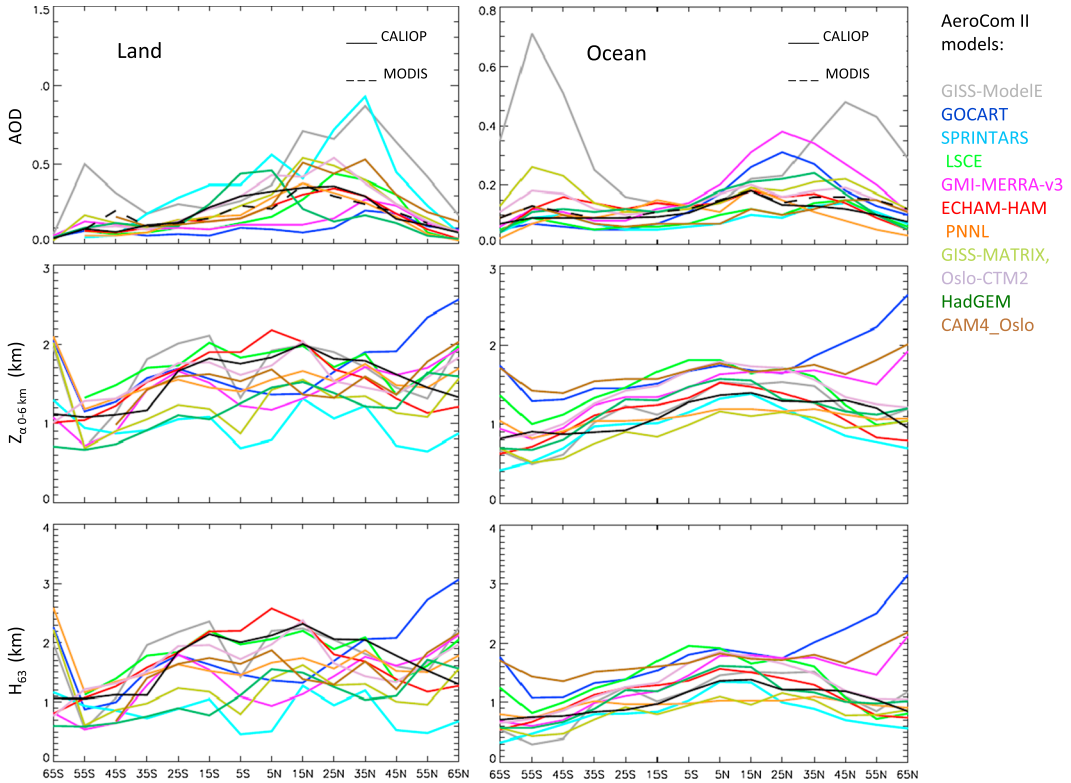

Figure 9. Models (2006) and CALIOP (2007–2009) mean annual latitudinal gradient of the (first row) AOD, (second row) $Z_{\alpha\,0\text{–}6\,km}$, and (third row) H_{63} aerosol diagnostics over the 0–6 km altitude range (14 latitude bands), over (left column) land and (right column) ocean. MODIS AOD is also reported.

for AeroCom I (see also Figures 7 and 8 in K12) and AeroCom II simulations nor for the change in performance (see also Figures S6 and S7 in the supporting information). These expected results reflect the changes in the models' input and the different ways different schemes affect the column integrated aerosol extinction and the shape of the aerosol vertical profile. This is due to (i) the fact that the selected regions are a combination of (different types of aerosol) source and downwind regions, with different predominant aerosol altitudes and different $Z_{\alpha 0\text{–}6\,km}$ versus AOD patterns (see, for instance, K12 and references inside) [*Zhang et al.*, 2013; *Prijith et al.*, 2016] and to (ii) a different sensitivity of the two diagnostics to the model capability in simulating the emission of aerosol particles and the subsequent transport/redistribution processes over and downwind from a given region.

The performance of the individual models in reproducing the timing of the seasonal maximum of $Z_{\alpha 0\text{–}6\,km}$ is summarized in Tables 4a and 4b. The ability of the models to simulate the $Z_{\alpha 0\text{–}6\,km}$ seasonal peak (Table 4b) has generally increased compared to AeroCom IA (Table 4a). In 83% of the cases, the AeroCom II models agree with the CALIOP peak season (instead of 69% for AeroCom IA). In the other cases, the season with CALIOP $Z_{\alpha 0\text{–}6\,km}$ maximum is ranked second (15%) or worse (2%). Like in AeroCom IA, all the models are able to reproduce the right timing in most of the regions, with five of them (SPRINTARS, LSCE, GMI-MERRA, GISS-MATRIX, and CAM4-Oslo) performing better than the others (i.e., failing in one region only). The JJA maximum over the Atlantic Ocean (CAT), due to the transport of Saharan dust from Africa is now captured by all models. Agreement is also obtained for all the models over the U.S. (EUS), Chinese (ECN), and Indian (IND) industrial regions. Consistently, the simulated seasonality over the North Atlantic (NAT) and West Pacific (NWP) downwind regions is also improved, i.e., only one and two models fail to capture the right season, respectively. Over the main dust regions, the models' performance in simulating Z_α seasonal peak is similar (NAF) or improved (WCN) compared to AeroCom IA. It is also somewhat improved over the SAM and SAF biomass burning regions (SAF and SAM) for some models (GISS-ModelE and LSCE) but deteriorated for others (GOCART and PNNL).

It is worth mentioning that three out of the four models that used *Diehl et al.* [2012] (CAM4-Oslo, SPRINTARS, and ECHAM-HAM) or *Lamarque et al.* [2010] (GISS-ModelE, LSCE, and PNNL) emissions improved in their capabilities in reproducing the seasonality of the Z_α aerosol height. Therefore, the differences in the seasonality,

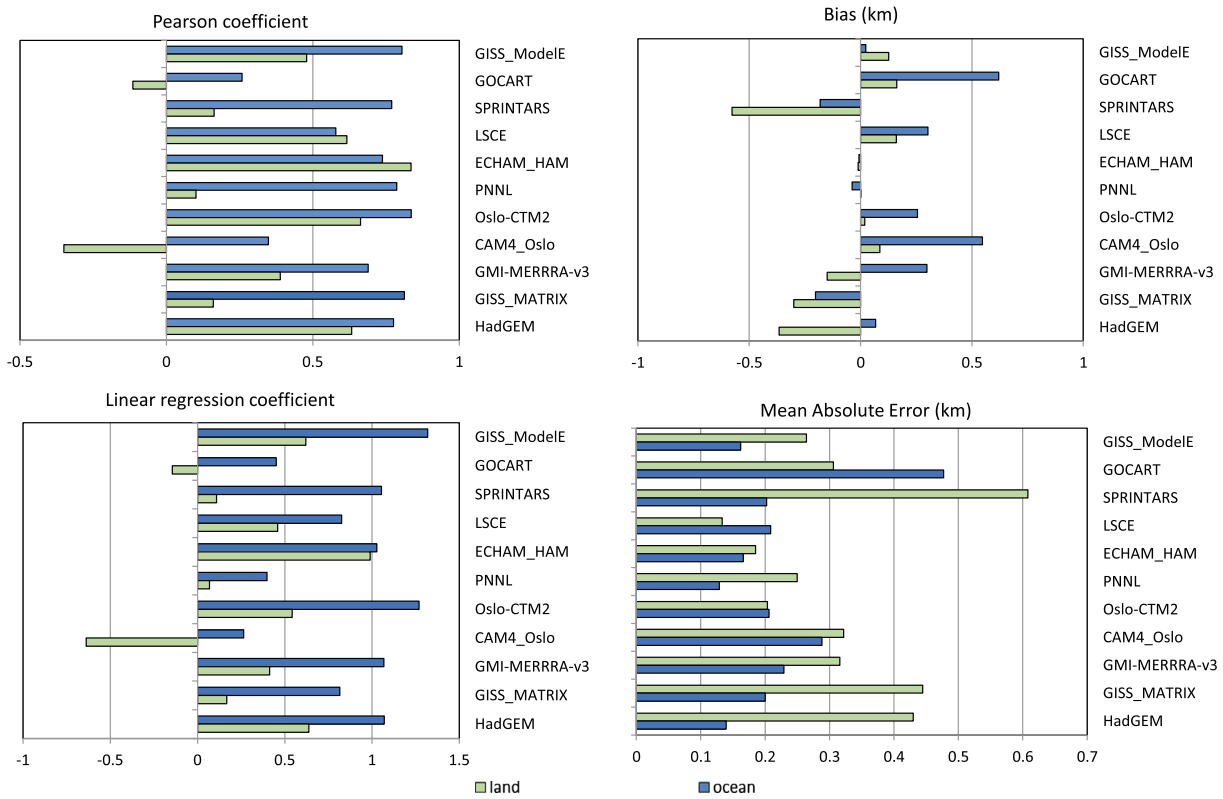

Figure 10. Comparison of AeroCom II (2006) models performance in reproducing CALIOP (2007–2009) mean annual $Z_{\alpha 0-6\,\mathrm{km}}$ aerosol extinction height over land (green) and ocean (blue): (top left) Pearson coefficient (threshold value for $p = 0.05$ is 0.532); (bottom left) linear regression coefficient, and (bottom right) mean absolute error (km) and (top right) bias (km).

and implicit variations in injection height, of both data sets compared to *Dentener et al.* [2006] might be at least partly responsible for the improvement of the AeroCom II models in this respect. In addition to input emissions, changes in the explicit injection heights, the meteorology and the models' schemes all contribute to the differences. This is also true for the shift from positive to negative of the $Z_{\alpha 0-6\,\mathrm{km}}$ model mean bias. The impact of the region's type (land and ocean) and of the changes in input and model schemes on the models' performance is further documented and discussed in sections 4 and 5, respectively.

4.2. AeroCom Phase II Models Performance in Reproducing Aerosol Latitudinal Variation Over Land and Ocean

In the previous section, we analyzed the performance (and evolution) of 11 (8) AeroCom II models in reproducing the aerosol mean vertical distribution and $Z_{\alpha 0-6\,\mathrm{km}}$ mean extinction height diagnostic over 12 subcontinental regions representative of industrial, dust and biomass burning pollution, or marine regions downwind from these continental source regions. In this section, we assess the ability of the same 11 AeroCom II models to simulate the latitudinal variability (14 latitudinal bands from 70°S to 70°N) of $Z_{\alpha 0-6\,\mathrm{km}}$, over land and ocean.

4.2.1. Annual Mean Distributions

The zonal annual mean $Z_{\alpha 0-6\,\mathrm{km}}$ values over land and ocean versus latitude are shown in Figure 9. To further assess the AeroCom-CALIOP product (see also section 3.1) and to facilitate the interpretation and the comparison with other studies, the AOD and the H63 metric (the altitude below which 63% of the AOD is found) are also reported.

Over both land and ocean, the mean annual CALIOP-derived AOD shows a maximum around 15°N, which is consistent with MODIS (Figure 9) and with SeaWiFS (see, for instance, *Hsu et al.* [2012]) retrievals. The (MODIS and) CALIOP-derived AOD values over land and ocean are within the AeroCom II intermodel range. Disregarding GISS-ModelE, the model diversity in AOD magnitude and AOD latitudinal variability is larger, and the model performance is lower, over the land than over the oceanic regions. The same is found for the

$Z_{\alpha 0-6\,\mathrm{km}}$ and H63 diagnostics, with a lower intermodel diversity and a better agreement with observations over the ocean compared to land. The performance of the 11 AeroCom II models, with respect to the latitudinal variation in $Z_{\alpha 0-6\,\mathrm{km}}$ as observed by CALIOP over land and ocean (see Figure 10) is summarized hereafter:

1. *Ocean*. All but two models are able to reproduce the $Z_{\alpha 0-6\,\mathrm{km}}$ latitudinal variations over ocean ($r > 0.532$; $p < 5\%$), with three them (GISS-ModelE, ECHAM-HAM, and HadGEM) performing particularly well (the linear regression slope coefficient is between 0.7 and 1.3 and the error is less than 100 m). Nevertheless, significant (>200 m) positive biases are found for five models (GOCART, CAM4-Oslo, LSCE, GMI-MERRA-v3, and Oslo-CTM2), whereas two models (SPRINTARS and GISS-Matrix) simulate lower than observed aerosol extinction height, resulting in a multimodel mean bias of +15%.
2. *Land*. A poorer performance is obtained for the AeroCom II models over land compared to ocean. Only 4 (ECHAM-HAM, HadGEM, LSCE, and Oslo-CTM2) out of the 11 models do reproduce the $Z_{\alpha 0-6\,\mathrm{km}}$ mean annual latitudinal gradient, with the best agreement obtained for the ECHAM-HAM model ($r = 0.989$; bias = 11 m). While the majority of the models show larger mean absolute errors over land than over ocean, the multimodel mean bias over land (−7%) is smaller, due to a higher compensation between positive and negative biases. The negative biases (HadGEM, GISS_MATRIX, and SPRINTARS) obtained over land (Figure 10) are very similar to those obtained over the 12 selected regions (Figure 7), which is simply explained by the fact that nine of them represent land regions. This previous analysis (see also Figure 6) also suggests that the $Z_{\alpha 0-6\,\mathrm{km}}$ negative bias over land is mainly due to an underestimation of the aerosol height over the dust and biomass burning source regions (whereas the models generally perform better over the industrial regions).
3. Overall, the $Z_{\alpha 0-6\,\mathrm{km}}$ and AOD results at global scale of the individual models over land and ocean (see Figures S8 and S9, respectively) are consistent with our findings and conclusions for the selected regions (see section 4.2 and Figure 7).

4.2.2. Seasonal Mean Distributions

The seasonal latitudinal results (not shown) show similar patterns compared with the annual results, as well as with AeroCom II results at subcontinental scale (section 4.2). As for the annual mean, a better correlation with CALIOP $Z_{\alpha 0-6\,\mathrm{km}}$ is obtained over ocean than over land during each of the four seasons, as well as a model mean negative bias over land and a positive one over ocean. This main result (better performance over ocean) is consistent with AeroCom II simulations over the 12 selected regions that show a good (improved) agreement with CALIOP observations over the maritime regions located downwind of the continental source areas. Both simulated and observed subcontinental patterns show a larger mean aerosol extinction height during Northern Hemisphere summer (JJA) than during NH winter (DJF), over both ocean and land. The largest $Z_{\alpha 0-6\,\mathrm{km}}$ seasonal bias and mean absolute error are calculated for the JJA (−107 m and 24% in model mean, respectively) and the MAM (−74 m and 24% in model mean) seasons over land, whereas they are largest for the SON season over ocean (+177 m and 29% in model mean, respectively), which is also consistent with the simulated subcontinental patterns (Table 2).

The main difference compared to the conclusions of the subcontinental regions analysis is that better correlations with CALIOP observations are generally obtained for the JJA season than for the DJF season, during which only two models (against five for the JJA season) reproduce the CALIOP $Z_{\alpha 0-6\,\mathrm{km}}$ latitudinal gradient over land. But this result could be also partly explained by CALIOP limitations in capturing the aerosol optical properties over land during this season (see section 4.1 and Figure S3a).

5. Discussion

AeroCom II versus AeroCom IA results show that the intermodel diversity did not decrease between the two AeroCom phases. While some models do better, others (e.g., Oslo-CTM2, SPRINTARS, and GOCART) are worse in terms of correlation and/or bias than previously. Our results also show that the performance in simulating the $Z_{\alpha 0-6\,\mathrm{km}}$, extinction height diagnostic over the 0–6 km altitude range generally decreased in model mean over North America's and Europe's industrial and downwind maritime regions, whereas it increased in the dust and biomass burning source regions of Asia and Africa. AeroCom II results also reveal that the vertical aerosol profiles over land are more difficult to capture than over ocean, especially above the main source regions of dust and biomass burning aerosol. Hypotheses for the model diversity and general changes or

Table 5. Change in AeroCom Models' Performance at Capturing the CALIOP Mean Annual $Z_{\alpha 0-6\,km}$ and MODIS AOD Over the 12 Regions from AeroCom I to AeroCom II Experiments[a]

Model	AeroCom II Emissions	AeroCom II Meteorology	AeroCom II Injection Heights: BB/Anthropogenic	$Z_{\alpha 0-6\,km}$ Simulation's Performance Compared to AeroCom I			AOD Simulation's Performance Compared to AeroCom I		Respective Importance in the Change in $Z_{\alpha 0-6\,km}$ Simulation's Performance; little (+) to very (+++) Important; not important (−); n.a: Not Assessed		
	See Tables 1a, 1b, 1c for Details			Peak Timing	Bias	*R*	Bias	*R*	Emissions	Meteorology	Model schemes
GOCART	2006 [*Diehl et al.*, 2012]	2006	Homogenous/n.d.	-	-	-	+	-	++	n.a.	++
SPRINTARS		2006	Homogenous/*Takemura et al.* [2000]	+	-	+	-	-	+	n.a.	+++
ECHAM_HAM		2006	*Dentener et al.* [2006] / n.d.	+	~	+	+	+	++	n.a.	++
CAM4-Oslo		5 years' mean	*Dentener et al.* [2006]/n.d.	+	+	-	-	~	++	-	++
GISS-ModelE	2000 [*Lamarque et al.*, 2010]	2006	Homogenous/n.d.	+	-	-	-	-	+	n.a.	+++
LSCE		2006	Layer specific/n.d.	+	+	+	-	-	++	n.a.	+
PNNL		2000	*Dentener et al.* [2006]/four lowest layers	+	-	+	~	+	+	-	+++
Oslo_CTM2		2006	Layer specific/n.d.	-	+	-	-	+	+/++	n.a.	++

[a]The + and − symbols mean an improved and a degraded performance in $Z_{\alpha 0-6\,km}$ simulation, respectively (i.e., "+" in the "Bias" column indicates a decrease in the absolute bias). n.d.: not defined.

patterns are suggested in section 5.1. It is not possible to identify all the reasons for the diversity in the (changes in the) individual model performances, as it varies widely according to region, season, and model and would require many dedicated sensitivity analyses to isolate causes. Only first conclusions, part of which were proposed in section 4, are further completed and discussed in section 5.2. Being the provider of the AeroCom-CALIOP product, we also found it helpful to the modelers to perform an in-depth analysis of the CALIOP limitations and uncertainties that can potentially contribute to the models versus observation differences (section 5.3)

5.1. Model Diversity and Discrepancies

The uncertainties in emissions (including the injection heights), the complexity of the convective transport processes and of the secondary aerosol formation, and the poor availability of ground-based observations to constrain the models altogether account to a large extent for the remaining discrepancies in the simulation by global models of the aerosol vertical distribution over land, especially in the dust and biomass burning regions. Other studies corroborate our results in many aspects and allow for identifying possible explanations for them. For instance, *Textor et al.* [2006] showed that the AeroCom models best agree on the emission mass fluxes of "anthropogenic" compared with "natural" matter and that differences in (natural sea salt and) dust aerosols are due to differences in the simulation of the size spectra, in the parameterizations of source strength as a function of the wind speed (and soil properties for dust aerosol), and in the atmospheric dynamical fields. In *Kim et al.* [2014], five of the AeroCom models examined in our study have also been evaluated in their ability in simulating the dust fraction of the aerosol over North Africa and the North Atlantic. The authors found remarkable differences among the models in dust emission, removal, and mass extinction efficiency, which all contribute to the large intermodel diversity and large differences compared to observations. *Mann et al.* [2014] more specifically compared and evaluated 12 AeroCom II models (five of which are examined in the present study) whose aerosol schemes explicitly simulate the microphysical processes that determine the particle size distribution. They found that the vertical profile of particle concentrations in marine regions (evaluated against limited in situ aircraft observations) is captured well by the models, with relatively small intermodel diversity, while higher

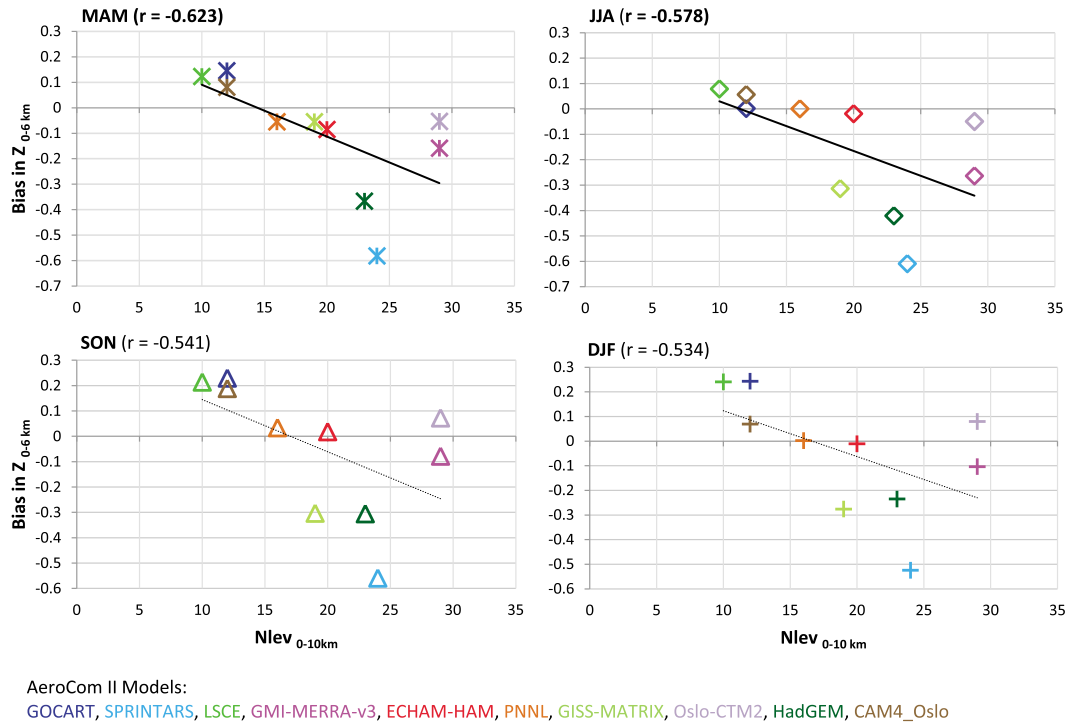

Figure 11. Bias in $Z_{\alpha 0-6\ \mathrm{km}}$ over land (70°N to 70°S latitude bands of 10°) as a function of the number of model layers below 10 km. Significant correlation (solid lines) are obtained for the MAM ($r = -0.623$) and JJA ($r = -0.578$) seasons ($p < 0.05$; $N = 10$ models).

model discrepancies and diversity occur over source regions. The *Tsigaridis et al.* [2014] study provides an intercomparison of the AeroCom II annual mean vertical profiles of organic aerosol (OC) for northern midlatitudes, South America, and the tropics (see their Figure 8), which shows a high diversity in the shape and in the magnitude of the simulated vertical distributions of the OC concentrations, notably over South America. *Kipling et al.* [2015] investigated the effects of switching off individual processes in one particular model (HadGEM3-UKCA) and compared the resulting diversity of aerosol mass mixing ratio vertical profiles with the intermodel diversity from the AeroCom A2 experiment. They showed that (in this model at least) the vertical profile is controlled by a relatively small number of processes (convective transport, in-cloud scavenging, aqueous oxidation, aging, and the vertical extent of biomass burning emissions) and that sufficiently coarse variations in these processes can produce a diversity similar to that between the different models, in some respects. On the other hand, results also showed that some model features cannot be reproduced, suggesting the influence of further structural differences between models.

In the second phase of AeroCom, four models have used 2006 emissions from *Diehl et al.* [2012], whereas the others have used 2000 emissions from *Lamarque et al.* [2010] (see also http://aerocom.met.no/emissions.html). For the biomass burning emissions, different data sets (GFED2 [*van der Werf et al.*, 2006] and GFED3 [*van der Werf et al.*, 2010] Global Fire Emission Data Sets and CMIP5 [*Lamarque et al.*, 2010]) and injection heights were applied. Regarding the 2006 meteorology, different reanalysis fields were also used. All these differences (Tables 1a, 1b, 1c) complicate the interpretation of the model diversity and the changes in the individual models' performance (summarized in Table 5) for which few hypotheses can be proposed hereafter (to be further investigated from emission data comparison and sensitivity analyses by the modelers):

1. The fact that the timing of the $Z_{\alpha 0-6\ \mathrm{km}}$ peak season is improved compared to AeroCom I (see section 4.2. and Tables 4a and 4b) might be at least partly due to differences in the seasonality of the emissions from *Diehl et al.* [2012] and *Lamarque et al.* [2010] compared to *Dentener et al.* [2006], in particular if injection height assumptions are coupled to emissions.
2. The strong year to year variation in fire emissions, the fact that GFED 2000 emissions [*van der Werf et al.*, 2003] used in AeroCom I were below average because of wet La Niña conditions [*Dentener et al.*, 2006] and that GFED2 and GFED3 data used in AeroCom II are known to be biased low (see, for instance, *Kaiser et al.*

[2012] and *Liu et al.* [2013]), could partly explain the underestimation of the mean aerosol height over biomass burning regions in both experiments.

3. We can also suspect that the systematic $Z_{\alpha 0-6\,\mathrm{km}}$ underestimation over of the biomass burning and dust regions during the MAM and JJA peak season are not only due to the input emissions and/or meteorology. Such discrepancies probably also reflect the models' common difficulty in accurately representing the vertical redistribution processes associated to these emissions. The simulation of the uplift of emissions highly depends on the planetary boundary layer (PBL) height and the vertical diffusion parameterization. Each model has a different way of acknowledging and accounting for the PBL, which is not investigated in details in the present study but only slightly evoked for some of the models in the next section. As a preliminary analysis, we simply looked at the seasonal models bias as a function of the vertical resolution (Figure 11) over the 14 land latitudinal bands (excluding the GISS-ModelE model for the reason mentioned in section 4.1.1). While one could expect that increasing the vertical resolution would extend the model's capability to represent the aerosol vertical distribution within the PBL, Figure 11 shows negative slopes with significant correlations ($p < 0.05$) between the bias in $Z_{\alpha 0-6\,\mathrm{km}}$ and the number of model layers below 10 km altitude ($N_{\mathrm{lev0-10\,km}}$), for the MAM ($r = -0.623$) and JJA ($r = -0.578$) seasons ($p < 0.05$; $N = 10$ models). In other words, a higher vertical resolution of the troposphere may also lead—at least in the dust and biomass burning regions—to an underestimation of the mean aerosol extinction height, e.g., due to the simulation of more (too) shallow boundary layer heights and reduced vertical mixing and upward transport, whereas coarser resolution models may potentially overestimate the aerosol uplift and vertical transport and consequently the $Z_{\alpha 0-6\,\mathrm{km}}$ diagnostic. Such results suggest that, despite the relatively low vertical resolution of the global climate models, there is still room for improving their capability in simulating the vertical distribution of the aerosol. As an additional illustration, the impacts of the revisions of the SPRINTARS model are discussed in this respect in the following section.

5.2. Changes in Individual Models' Performance

Not only the input but also the models' schemes have experienced changes between the two AeroCom phases, laying 5 years apart (see Tables 1a, 1b, 1c). While it is beyond the scope of this study to perform sensitivity analyses for the different models, hypotheses for individual changes (Table 5) are investigated hereafter, based on further investigation and discussions with the modelers.

1. The revisions in the *PNNL* model with respect to the Aerocom IA version [*Ghan and Easter*, 2006] have been so important that it could be viewed as a different model: They include a newer version of the host GCM with major changes to its cloud parameterizations, changes to convective wet removal, single versus two coarse modes, different secondary organic aerosol treatment (burden much higher), and different emissions and injection heights. The overall performance of the PNNL model in simulating the mean aerosol height over the selected regions has increased (Figure 7). While it is now ranked between the five highest performing AeroCom II models over these regions, its ability in reproducing the latitudinal variability of the mean aerosol height at global scale over land is poor (Figure 10). The fact that it used the 2000 meteorology and *Lamarque et al.* [2010] 2000 monthly (decadal) emissions (with seasonality only given for the biomass burning emissions) could partly account for the model discrepancies.
2. The *Oslo-CTM2* has been substantially revised. Among the changes that can affect its performance are the changes in the BC aging that now depends on the latitude and season [*Skeie et al.*, 2011b], whereas it was constant in AeroCom IA, the reduced scavenging from ice precipitation for some of the components (BC, NO_3, SO_4, OC), the inclusion of schemes for the formation of secondary organic aerosol (SOA) and nitrates, the shift from 40 to 60 vertical levels, as well as differences in the relative humidity, and the enhanced absorption due to hydrophilic BC. Also, the large-scale precipitation is now removed every hour, rather than every third hour in the previous version, which has a large effect on the vertical distribution of the species. Despite (or because of) of these changes, the model that was among AeroCom I highest performers now shows a significant decrease in its ability in reproducing the shape of the CALIOP-derived aerosol profiles (Figure 5) as well as the magnitude (Figures 6 and 7) and seasonality (Tables 4a and 4b) of the $Z_{\alpha 0-6\,\mathrm{km}}$, extinction height diagnostic over some source and downwind regions. Nevertheless, the $Z_{\alpha 0-6\,\mathrm{km}}$, mean bias over the 12 regions has decreased, and the results at global scale in terms of the latitudinal variability of the mean aerosol height are within the higher part of AeroCom II models' performance (Figure 10).

3. The *SPRINTARS* model also experienced significant developments between the two AeroCom phases. Of particular importance for the simulation of the vertical distribution of the aerosol is the vertical resolution that has been enhanced, from 20 to 56 vertical levels. The emission schemes of soil dust and sea salt have also been revised (See Appendixes A and B in *Takemura et al.* [2009]). Its ability in reproducing the vertical aerosol profiles over the selected regions is still among the best of the AeroCom models. However, a systematically too low aerosol height is now simulated (Figure 6). This decrease in the simulated mean aerosol height between the two AeroCom phases could be due to the version of the MIROC-SPRINTARS boundary scheme version used both in AeroCom phase I and phase II, in which the uplift level of emissions (constrained by equation A1 in *Takemura et al.* [2000]), may depend on the vertical resolution. It has been revised in the latest version of MIROC-SPRINTARS that is used in the AeroCom Phase III (T. Takemura, personal communication, 2015).
4. The main relevant changes in the *ECHAM-HAM* model that affect the vertical distribution of aerosols are the new cloud microphysics parameterization ([*Lohmann et al*., 2007] longer aerosol lifetime in that model version due to slower autoconversion rate) and the new SOA treatment ([*O'Donnell et al*., 2011] chemical production in the free troposphere, more POM in upper troposphere). As a result, the model which was already among the AeroCom I best performers in its ability in simulating the $Z_{\alpha 0-6\,\mathrm{km}}$ mean extinction over the selected regions but showed a marked systematic negative bias in the aerosol optical depth K12 has improved in both respects (see Figure 7 as well as S5 and S7 in the supporting information). It shows the highest AeroCom II models' ability in reproducing the CALIOP-derived aerosol vertical distribution. This increase in accuracy, notably in the dust and biomass burning regions, is at least partly due to the recent ECHAM-HAM improvements in process representations, which nature and impacts are further discussed in *Zhang et al.* [2012].
5. The *LSCE* model was not changed significantly between the two AeroCom phases. While no significant correlation ($p < 0.05$) with CALIOP $Z_{\alpha 0-6\,\mathrm{km}}$, extinction height diagnostic is yet obtained over the selected regions (Figure 7), the model shows a lower positive bias compared to AeroCom I. This might be partly due to the AeroCom I versus II different modeling setup and input as suggested by the overall decrease in the mean extinction height simulated by most of the models between the two AeroCom phases. The same is true for the significant improvement in the simulation of the seasonality in the $Z_{\alpha 0-6\,\mathrm{km}}$ extinction height diagnostic.
6. Unlike *CAM4-Oslo*, its early predecessor CCM-Oslo denoted UIO_GCM in K12 used prescribed mineral and sea salt aerosol concentrations [*Kirkevåg et al*., 2005], thus underestimating the seasonal variability in regions that are dominated by sea salt and dust. This probably explains to a large extent the improvement obtained in the simulation of the $Z_{\alpha 0-6\,\mathrm{km}}$, mean extinction height seasonality by CAM4-Oslo compared to CCM-Oslo (Tables 4a and 4b). The model also experienced changes in the treatment of convective transport and mixing between convective downdraft and updraft plumes. In CCM-Oslo (AeroCom I) vertical transport of aerosols and precursors in deep convective clouds was omitted altogether, probably underestimating concentration in the free troposphere, while it is included in CAM4-Oslo [*Kirkevåg et al*., 2013] and in the intermediate model version CAM-Oslo [*Seland et al*., 2008]. In CAM-Oslo, it was assumed that shear-generated turbulence fully mixes the constituents between the updrafts and downdrafts. This special assumption was not ported to CAM4-Oslo. As a result, CAM4-Oslo was shown [*Kirkevåg et al*., 2013] to provide increased (and relatively high) mass mixing ratios at high altitudes compared to the previous versions. Although giving a smaller bias, the model ability to reproduce the mean annual patterns of the aerosol vertical distribution over the selected regions has become poorer in the AeroCom II model version (Figure 7).
7. The case of the *GISS-ModelE* is particular in the sense that its improvement in simulated mean extinction height over the 12 selected regions is associated to a deterioration of the simulated shape of the vertical profiles, which now show a quite unrealistic bimodal shape (Figure 5) due to in-cloud aerosol formation of sulfate aerosol (see section 4.1.1). This also at least partly explains why the AOD simulated over the sea and the land latitudinal bands are also not consistent with MODIS and CALIOP observations (Figure 9 and S9). The improved performance of the model in simulating CALIOP $Z_{\alpha 0-6\,\mathrm{km}}$, over the 12 regions might be partly due to the (increased) vertical resolution, which we showed to be significantly correlated to the magnitude of the simulated mean extinction height over land (see Figures 11).
8. Finally, while a low bias and a good correlation ($p < 5\%$) with CALIOP $Z_{\alpha 0-6\,\mathrm{km}}$, extinction height diagnostic was obtained in AeroCom IA by the *GOCART* model (Figure 7), the most recent model's version shows a particularly low performance in simulating the mean aerosol height over the selected regions and the

latitudinal bands, despite the fact that it has not been highly revised since K12 (see Tables 1a, 1b, 1c). The model performance in simulating $Z_{\alpha 0-6\,\mathrm{km}}$ could be affected by the lower vertical resolution in AeroCom II simulations ($N_{\mathrm{lev0-10\,km}} = 12$) compared to AeroCom I ($N_{\mathrm{lev0-10\,km}} = 17$) simulations. This hypothesis, which is supported by the negative correlation obtained between the change in $Z_{\alpha 0-6\,\mathrm{km}}$ mean bias over the 12 selected regions and the changes in $N_{\mathrm{lev0-10\,km}}$ (see Table S1 and Figure S7), would require, as for SPRINTARS and GISS-ModelE simulations, further model sensitivity studies to be confirmed.

Despite the complex combination of differences in emission, meteorology, and model schemes between AeroCom I and II simulations, the results of K12 and of the present study, together with the modelers' understanding of their model's skills, limitations, and sensitivity, allowed to put forward first conclusions on the main factors responsible for the changes in the models' performance. These analyses and conclusions also allow us to propose a first guess of the respective weight of changes in emissions (meteorology) and model schemes (see Table 5). For most of the studied models, the changes in (the vertical resolution and other aspects of) the model schemes are suspected to have significantly affected the simulated mean aerosol vertical profiles over the selected regions, showing the need and the room for further sensitivity analyses and model improvement.

5.3. CALIOP Limitations and Uncertainties

In the first model evaluation K12, several sensitivity analyses demonstrated that the AeroCom-CALIOP product provides a robust and representative signal of the seasonal aerosol typical height, over selected subcontinental regions. Among those, the effect of removing daytime data was shown to only have a relatively small impact on the shape of the aerosol vertical distributions and mean aerosol height over the selected regions (i.e., with $Z_{\alpha 0-6\,\mathrm{km}}$ differences $<10\%$ for nine regions and $<17\%$ for the four other ones, as compared to the 24 h mean annual profiles).

The present study reveals that the AODs derived from the AeroCom-CALIOP extinction product also generally agree with MODIS measurements, over land and ocean latitude bands, with some discrepancies obtained during the DJF season over land, and during the MAM and JJA seasons over ocean (see sections 4.1 and 4.2.1, Figure 9). The particularly good agreement obtained between CALIOP and MODIS over land during the MAM and JJA seasons suggests that the high model biases obtained in these seasons in the dust and biomass burning regions are more likely due to modeling than to CALIOP product limitation. More generally, the consistency of the AeroCom-CALIOP-derived AOD with the MODIS retrievals, and the similarity of the mean aerosol extinction heights derived from the AeroCom and NASA aerosol products (see section 4.1) provides good confidence in the use of either data set for the evaluation of the model simulations. Nevertheless, further analysis of uncertainties in the CALIOP-derived extinction profiles and the differences between the AeroCom and NASA aerosol products is still needed.

The AeroCom and NASA CALIOP global products show substantial differences within the 0–500 m altitude range (Figure S2), which we suspect to be to a large extent due to the use of the CALIOP Level 2 layer data versus CALIOP Level 2 profile data, respectively (sections 3 and 4.1.). Moreover, as also discussed in section 3, some of the screening and averaging procedures used in the present study are different than in *Winker et al.* [2013]. In particular, the fact that the NASA method ignores more "clear" air near the surface during averaging tends to increase the magnitude of extinction near the surface compared to our methodology. Another difference is that we use a more restrictive range of the cloud-aerosol discrimination (CAD) score compared to *Winker et al.* [2013]. In addition to the different processing choices, further investigation highlighted potential sources of bias in the respective algorithms of CALIOP Layer data processing. Both the biases and the algorithm differences, and their impact on the retrieved aerosol profiles and mean aerosol height diagnostics are discussed hereafter.

5.3.1. Sources of Bias in the Processing

Sources of bias due to the data processing have been identified in both the NASA (D. Winker, personal communication, 2015) and the AeroCom CALIOP-derived global product. In our case, the bias is due to how aerosol layers are identified and processed within the CALIOP Layer algorithm. Indeed, the CALIOP Layer 5 km product reports layers detected at multiple averaging resolutions (5 km, 20 km, and 80 km). Because of this multistep approach, aerosol layers may appear to overlap in the vertical dimension. Such multiaerosol layers at a given altitude have been averaged when building the AeroCom-CALIOP product, whereas the layers detected at higher spatial resolutions should have been overwritten, rather than overlapped, the ones detected at coarser spatial resolutions. Figure S2 shows the impact of this misrepresentation on the shape

of the normalized mean annual profiles, as calculated from a simple regional averaging as in K12. The impact appears to be very small, inducing differences in the $Z_{\alpha 0-6\,\mathrm{km}}$ annual values (up to 2%) that are inferior to the ones due to the horizontal averaging (Figure S1). The impact on the seasonal $Z_{\alpha 0-6\,\mathrm{km}}$ values is also very limited, to the point that it would be hard to distinguish between the corrected (not reported) and noncorrected $Z_{\alpha 0-6\,\mathrm{km}}$ values in Figure 8. In conclusion, this source of bias—to be corrected in the next version of the AeroCom-CALIOP product—neither affects the present and K12 studies' results and conclusions nor explains the differences between the NASA and AeroCom mean aerosol profiles.

5.3.2. Effects of Different Algorithm Choices

In the CALIOP Layer product, the base altitudes of optically thick aerosol layers are sometimes biased high due to lidar signal attenuation or perturbation, which causes an underestimate in the aerosol extinction at low levels. To reduce this anomaly, we assume the lowest aerosol layer to extend to the surface whenever its height above the surface is less than 10% of the layer depth. *Winker et al.* [2013] ignore the "clear air" samples between the surface and the lowest aerosol layer when its base is above the local surface but lower than 2.46 km. Second, similarly to *Yu et al.* [2010] and K12, we exclude the aerosol layers that have *low* CAD scores (i.e., lying between −50 and 0), in order to include only well-defined aerosol layers, whereas *Winker et al.* [2013] used a less restrictive threshold (i.e., −20 instead of −50). In order to examine and quantify the effects of different algorithm choices, we recalculated the mean aerosol profiles over our 12 selected regions, applying the K12 methodology but using a CAD threshold of −20 and *Winker et al.* [2013] processing below 2.46 km. Because of the horizontal resolution of the NASA Level 3 aerosol product (2°×5° latitude-longitude grid), the domains of the regions are slightly different compared to Figure 2. As expected, an increase in the magnitude of extinction is obtained compared to K12 but to an extent that is not sufficient to reach the extinction values derived from the NASA product in the lowest kilometer (Figure S2a). Moreover, the modification of the CALIOP Layer data processing does not substantially impact the normalized profiles and the mean regional $Z_{\alpha 0-6\,\mathrm{km}}$ annual extinction values (Figure S2b). As for the incorrect treatment of the overlapping layers, the change in $Z_{\alpha 0-6\,\mathrm{km}}$ is (up to 2.6%) less than the one due to the horizontal resolution (up to 5%).

The processing of the "overlapping" aerosol layers, the different treatment of the clean air below the lowest aerosol layer, and the use of different CAD thresholds and vertical resolutions have small effects on the profile shape and do not allow to explain the 0.2 km $Z_{\alpha 0-6\,\mathrm{km}}$ difference between the two CALIOP-derived products shown in Figure 3. While additional sensitivity tests would be required in order to quantify the impact of other differences (e.g., in the cloud masking), we conclude that the higher aerosol extinction obtained in the first kilometer in the NASA product is very likely predominantly due to the use of the CALIOP Level 2 aerosol profile data instead of the Level 2 aerosol layer data, which are used to build the AeroCom-CALIOP product. It is also worth noting that the NASA CALIOP Level 3 product has been recently updated. The new version 3.00 includes many corrections, modifications, and improvements including refined sky conditions reduced biases in single-species averages and corrected mean aerosol optical depth calculations. Among the modifications, the altitude threshold under which the "clean air" parcel between the surface and the base of the lowest aerosol layer is ignored (see section 3) has been decreased in the new NASA algorithm (from 2.46 km) to 250 m. Figure S10 compares our $Z_{\alpha 0-6\,\mathrm{km}}$ mean annual extinction heights over the 12 selected regions to the ones calculated from the two versions of the NASA CALIOP Level 3 product. It shows that the recent updates in the NASA product tend to reduce the $Z_{\alpha 0-6\,\mathrm{km}}$ differences with the AeroCom-CALIOP product for 10 out of the 12 studied regions. As a result, the $Z_{\alpha 0-6\,\mathrm{km}}$ mean bias between the AeroCom-CALIOP and NASA products decreases from 18% to 14%.

6. Summary and Conclusions

This paper is an update and extension of the K12 study. Its primary objective was to determine to what extent state-of-the-art AeroCom II global aerosol models reproduce CALIOP-derived aerosol vertical distribution over 12 selected subcontinental regions and in what aspect models have improved since the AeroCom Phase I. It also aimed at further evaluating the models' performance at global scale, through the analysis of the latitudinal gradients over land and ocean. To fulfill these objectives, an in-house AeroCom-CALIOP global product has been built, using the methodology applied in K12 to produce the CALIOP-derived mean subcontinental benchmark data set. Further sensitivity tests were also performed to better assess the data set's limitations and uncertainties.

The models' ability in reproducing CALIOP-derived mean aerosol extinction profiles was shown to still depend highly on the model, the season, and the region. Nevertheless, agreements and discrepancies with observations were identified that should provide guidance to the modelers in identifying and investigating the factors responsible for the changes in the model performance and remaining biases. From the analysis of the AeroCom II models' performance on subcontinental regions, we found that about half of the models capture the $Z_{\alpha 0-6\,\mathrm{km}}$ interregional diversity and the regional seasonality well. Five models improved, whereas three degraded in reproducing the mean annual $Z_{\alpha 0-6\,\mathrm{km}}$ values compared to AeroCom IA and the multimodel mean absolute error decreased from 16% to 11%. The seasonal biases in $Z_{\alpha 0-6\,\mathrm{km}}$ have generally decreased in the U.S. and European industrial and downwind maritime regions, whereas the timing of the $Z_{\alpha 0-6\,\mathrm{km}}$ maximum over the different regions has improved for all but two models. This improvement may at least partly be explained by the use of higher time resolution or better-resolved seasonality of the input emission data and/or of the models' emissions schemes (e.g., for dust and sea salt aerosols). On the other hand, most of the models now underestimate the mean aerosol extinction height over land, which is particularly pronounced during the season of peak dust and biomass burning emissions in Asia and Africa. This result, also found but to a lower extent in AeroCom Phase I simulations, is confirmed at global scale: All but two models reproduce the $Z_{\alpha 0-6\,\mathrm{km}}$ latitudinal gradient over ocean, whereas only four out of the 11 AeroCom II models do over land, where they also generally show higher mean absolute errors than over ocean. The negative bias of the aerosol mean extinction height over land is mainly due to a $Z_{\alpha 0-6\,\mathrm{km}}$ underestimation in the dust and biomass burning source regions, which does not seem to be improved—and might even be worsened—from an enhanced vertical resolution within the PBL. Our results suggest that a higher vertical resolution of the troposphere may lead—at least in the dusty and biomass burning regions—to an underestimation of the mean aerosol extinction height, due to the too shallow PBL height and reduced vertical mixing and upward transport, whereas coarser resolution models might overestimate the uplift and vertical transport of the aerosol. It is also worth mentioning that the GFED2 and GFED3 input emissions may partly contribute to the simulated low-biased aerosol height in the fire regions. The use in the future of the GFED4 biomass burning emission inventory, which includes more burned area in most regions of the globe and shows significant differences compared to GFED3 (http://www.globalfiredata.org/), might substantially modify these figures.

As for the AeroCom-CALIOP global monthly product, we carried out sensitivity tests that provided further information on its limitations and uncertainties, and their impact on the models versus observations differences. Its comparison to the NASA gridded product for the year 2007 revealed significant differences in the lowest troposphere, due notably to the different vertical resolution the processed CALIOP Level 2 Layer and profile data, which however do not strongly affect the main results and conclusions reported in section 4. Furthermore, according to our preliminary analysis of the new version of the NASA product, these differences are reduced compared to the ones with *Winker et al.* [2013] product. These results confirm the robustness and potential of the AeroCom-CALIOP data set to assist aerosol models' evaluation. Together with the analysis of the (AeroCom I to) AeroCom II (changes in the) model results, they also raise the need for further work in assessing the impact of the models' original vertical resolution on the shape of the interpolated model extinction profiles, notably within the PBL, and on the resulting $Z_{\alpha 0-6\,\mathrm{m}}$ biases.

Our methodological approach is currently being refined and the AeroCom-CALIOP product extended to the year 2011. It will serve for the evaluation the AeroCom global and AQMEII (Air Quality Model Evaluation International Initiative; http://ensemble2.jrc.ec.europa.eu/aqmeii/) regional models, which are participating in the second multimodel exercise of the Task Force on Hemispheric Transport on Air Pollution (HTAP) that focuses on the years 2008 and 2010 and requires the use of common anthropogenic emissions data by the global models [*Janssens-Maenhout et al.*, 2015]. One of the goals of the HTAP Task Force is to further understand aerosol intercontinental transport and to assess its impacts on air quality, climate, and ecosystems (http://www.htap.org/). This requires evaluating and/or constraining models with measurements having adequate temporal frequency and spatial-temporal coverage on a global scale (see, for instance, *Yu et al.* [2013]). In this respect, our preliminary sensitivity analyses on the impact of the input emissions (Figure S4 of K12), of the sampling of the models' outputs on the time of CALIOP overpasses (Figure S2 of K12) and of the horizontal (Figure S1, present study) and vertical (Figures 11, S2, and S7) resolutions of the CALIOP and model-derived profiles need to be extended if we want to further assess, notably at regional scale, the models' discrepancies, and the limits of the AeroCom-CALIOP product. Use of satellite aerosol

retrievals to constrain aerosol models could also be further improved if aerosol instrument simulators [*Bodas-Salcedo et al.*, 2011] are to be applied to the simulated aerosols to overcome issues of biases due to attenuation and contamination. While it is difficult to find independent data with which to evaluate the accuracy of CALIOP monthly mean aerosol, one could make use of the large data set of colocated high spectral resolution lidar measurements of 532 nm aerosol extinction that is available over North America and the eastern Caribbean [*Hair et al.*, 2008; *Rogers et al.*, 2011; *Winker et al.*, 2013], and the aerosol-type-dependent extinction data derived from European Aerosol Research Lidar Network ground-based measurements over Europe [*Pappalardo et al.*, 2014]. Finally, while aerosol extinction is an important constraint on the simulated aerosol, the vertical distribution of aerosol absorption is also important to constrain. Although vertical profiles of aerosol absorption cannot be measured by current satellites, a new generation of lidars have such capability, at least under sufficiently polluted conditions. Deployment of such lidars in space will open new opportunities for constraining simulations of aerosol and their impacts on climate.

Acknowledgments
This work was supported by the European Commission under the project IS-ENES (Infrastructure for the European Network for Earth System Modelling) and the Administrative Arrangement AMITO (070307/ENV/2012/636596/C3). We thank NASA teams and the ICARE Data and Services Center for providing access to the CALIOP CNES/NASA data used in this study and for providing continuous computing access and support. We are also very grateful to three reviewers for their valuable comments and suggestions that allowed improving the quality of the manuscript and reinforcing some of our findings. T. Iversen, A. Kirkevåg, and Ø. Seland (and B. Koffi) were (also) supported by the Research Council of Norway through the EarthClim (207711/E10), EVA (229771), and NOTUR/NorStore projects, CRAICC, and through the EU projects PEGASOS and ACCESS. M. Schulz and A. Kirkevåg also received funding from the Norwegian Space Center through the PM-VRAE and PM-MACS projects. S. Ghan, R. Easter, P. Rasch, J. Yoon and K. Zhang were funded by the US Department of Energy, Office of Science, Scientific Discovery through Advanced Computing (SciDAC) program. The Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC06-76RLO 1830. S.E. Bauer and K. Tsigaridis acknowledge resources supporting this work by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center and support by the NASA MAP program Modeling, Analysis, and Prediction Climate Variability and Change (NNH08ZDA001N-MAP). M. Schulz, Jan Griesfeller, R.B. Skeie, T. Berntsen and G. Myhre were supported by the Research Council of Norway, through the grants SLAC, AEROCOM-P3 and ClimSense. K. Zhang acknowledges the German Climate Computing Center (Deutsches Klimarechenzentrum GmbH, DKRZ) for making the computational resources available for ECHAM5.5-HAM2 simulations. The CALIOP and AeroCom data and tools used to produce the aerosol extinction profiles analyzed in this paper are available on the AeroCom Database and User Server (aerocom-users.met.no). They are accessible upon request, following the AeroCom Policy and access conditions described under http://aerocom.met.no/data.html.

References

Balkanski, Y. (2011), L'Influence des Aérosols sur le Climat, Thèse d'Habilitation à Diriger des Recherches, Université de Versailles Saint-Quentin, Saint-Quentin-en-Yvelines.

Bauer, S. E., D. Wright, D. Koch, E. R. Lewis, R. McGraw, L. -S. Chang, S. E. Schwartz, and R. Ruedy (2008), MATRIX (Multiconfiguration Aerosol TRacker of mIXing state): An aerosol microphysical module for global atmospheric models, *Atmos. Chem. Phys.*, *8*, 6003–6035, doi:10.5194/acp-8-6003-2008.

Bellouin, N., J. Rae, A. Jones, C. Johnson, J. Haywood, and O. Boucher (2011), Aerosol forcing in the climate model intercomparison project (CMIP5) simulations by HADGEM2-ES and the role of ammonium nitrate, *J. Geophys. Res.*, *116*, D20206, doi:10.1029/2011JD016074.

Bian, H., M. Chin, J. M. Rodriguez, H. Yu, J. E. Penner, and S. Strahan (2009), Sensitivity of aerosol optical thickness and aerosol direct radiative effect to relative humidity, *Atmos. Chem. Phys.*, *9*, 2375–2386, doi:10.5194/acp-9-2375-2009.

Bodas-Salcedo, A., et al. (2011), COSP: Satellite simulation software for model assessment, *Bull. Am. Meteorol. Soc.*, *92*, 1023–1043.

Chin, M., T. Diehl, O. Dubovik, T. F. Eck, B. N. Holben, A. Sinyuk, and D. G. Streets (2009), Light absorption by pollution, dust, and biomass burning aerosols: A global model study and evaluation with aeronet measurements, *Ann. Geophys.*, *27*, 3439–3464, doi:10.5194/angeo-27-3439-2009.

Dentener, F., et al. (2006), Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom, *Atmos. Chem. Phys.*, *6*, 4321–4344, doi:10.5194/acp-6-4321-2006.

Di Pierro, M., L. Jaeglé, E. W. Eloranta, and S. Sharma (2013), Spatial and seasonal distribution of Arctic aerosols observed by CALIOP (2006–2012), *Atmos. Chem. Phys. Discuss.*, *13*, 4863–4915, doi:10.5194/acpd-13-4863-2013.

Diehl, T., A. Heil, M. Chin, X. Pan, D. Streets, M. Schultz, and S. Kinne (2012), Anthropogenic, biomass burning, and volcanic emissions of black carbon, organic carbon, and SO_2 from 1980 to 2010 for hindcast model experiments, *Atmos. Chem. Phys. Discuss.*, *12*, 24,895–24,954, doi:10.5194/acpd-12-24895-2012.

Ghan, S., and R. Easter (2006), Impact of cloud-borne aerosol representation on aerosol direct and indirect effects, *Atmos. Chem. Phys.*, *6*, 4163–4174.

Hair, J. W., C. A. Hostetler, A. L. Cook, D. B. Harper, R. A. Ferrare, T. L. Mack, W. Welch, L. R. Izquierdo, and F. E. Hovis (2008), Airborne High Spectral Resolution Lidar for profiling aerosol optical properties, *Appl. Opt.*, *47*, 6734–6752.

Hauglustaine, D. A., F. Hourdin, S. Walters, L. Jourdain, M.-A. Filiberti, J.-F. Larmarque, and E. A. Holland (2004), Interactive chemistry in the Laboratoire de Météorologie Dynamique general circulation model: Description and background tropospheric chemistry evaluation, *J. Geophys. Res.*, *109*, D04314, doi:10.1029/2003JD003957.

Hodnebrog, Ø., G. Myhre, and B. H. Samset (2014), How shorter black carbon lifetime alters its climate effect, *Nat. Commun.*, *5*, 5065.

Hsu, N. C., R. Gautam, A. M. Sayer, C. Bettenhausen, C. Li, M. J. Jeong, S.-C. Tsay, and B. N. Holben (2012), Global and regional trends of aerosol optical depth over land and ocean using SeaWIFS measurements from 1997 and 2010, *Atmos. Chem. Phys.*, *12*, 8037–8053.

Hu, Y., et al. (2009), CALIPSO/CALIOP cloud phase discrimination algorithm, *J. Atmos. Oceanic Technol.*, *26*, 2293–2309, doi:10.1175/2009JTECHA1280.1.

Huneeus, N., et al. (2011), Global dust model intercomparison in AeroCom phase I, *Atmos. Chem. Phys.*, doi:10.5194/acp-11-7781-2011.

Janssens-Maenhout, G., et al. (2015), HTAP_v2: A mosaic of regional and global emission gridmaps for 2008 and 2010 to study hemispheric transport of air pollution, *Atmos. Chem. Phys.*, 11,411–11,432, doi:10.5194/acp-15-11411-2015.

Kaiser, J. W., et al. (2012), Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, *Biogeosciences*, *9*, 527–554, doi:10.5194/bg-9-527-2012.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-Year, Reanalysis Project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471.

Kim, D., et al. (2014), Sources, sinks, and transatlantic transport of North African dust aerosol: A multimodel analysis and comparison with remote sensing data, *J. Geophys. Res. Atmos.*, *119*, 6259–6277, doi:10.1002/2013JD021099.

Kinne, S., et al. (2006), An AeroCom initial assessment—Optical properties in aerosol component modules of global models, *Atmos. Chem. Phys.*, *6*, 1815–1834, doi:10.5194/acp-6-1815-2006.

Kipling, Z., P. Stier, J. P. Schwarz, A. E. Perring, J. R. Spackman, G. W. Mann, C. E. Johnson, and P. J. Telford (2013), Constraints on aerosol processes in climate models from vertically-resolved aircraft observations of black carbon, *Atmos. Chem. Phys.*, *13*, 5969–5986, doi:10.5194/acp-13-5969-2013.

Kipling, Z., et al. (2015), What controls the vertical distribution of aerosol? Relationships between process sensitivity in HadGEM3–UKCA and inter-model variation from AeroCom Phase II, *Atmos. Chem. Phys. Discuss.*, *15*, 25,933–25,980.

Kirkevåg, A., T. Iversen, Ø. Seland, and J. E. Kristjánsson (2005), Revised schemes for optical parameters and cloud condensation nuclei in CCM-Oslo, Institute Report Series, No. 128, Section for Meteorology and Oceanography, Department of Geosciences, Oslo, Norway.

Kirkevåg, A., et al. (2013), Aerosol-climate interactions in the Norwegian Earth System Model—NorESM1-M, *Geosci. Model Dev.*, *6*, 207–244, doi:10.5194/gmd-6-207-2013.

Kittaka, C., D. M. Winker, M. A. Vaughan, A. Omar, and L. A. Remer (2011), Intercomparison of column aerosol optical depths from CALIPSO and MODIS-Aqua, *Atmos. Meas. Tech.*, *4*, 131–141, doi:10.5194/amt-4-131-2011.

Koch, D., et al. (2009), Evaluation of black carbon estimations in global aerosol models, *Atmos. Chem. Phys.*, *9*, 9001–9026, doi:10.5194/acp-9-9001-2009.

Koch, D., et al. (2011), Coupled aerosol-chemistry-climate twentieth-century transient model investigation: Trends in short lived species and climate responses, *J. Clim.*, *24*, 2693–2714, doi:10.1175/2011jcli3582.1.

Koffi, B., et al. (2012), Application of the CALIOP layer product to evaluate the vertical distribution of aerosols estimated by global models: Aerocom phase I results, *J. Geophys. Res.*, *117*, D10201, doi:10.1029/2011JD016858.

Lamarque, J., et al. (2010), Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application, *Atmos. Chem. Phys.*, *10*, 7017–7039, doi:10.5194/acp-10-7017-2010.

Liu, J., J. A. Logan, L. T. Murray, H. C. Pumphrey, M. J. Schwartz, and I. A. Megretskaia (2013), Transport analysis and source attribution of seasonal and interannual variability of CO in the tropical upper troposphere and lower stratosphere, *Atmos. Chem. Phys.*, *13*, 129–146.

Liu, X., et al. (2012), Toward a minimal representation of aerosols in climate models: Description and evaluation in the Community Atmosphere Model CAM5, *Geosci. Model Dev.*, *5*, 709–739, doi:10.5194/gmd-5-709-2012.

Liu, Z., M. Vaughan, D. Winker, C. A. Hostetler, L. R. Poole, D. L. Hlavka, W. D. Hart, and M. J. McGill (2004), Use of probability distribution functions for discriminating between cloud and aerosol in lidar backscatter data, *J. Geophys. Res.*, *109*, D15202, doi:10.1029/2004JD004732.

Liu, Z., M. A. Vaughan, D. M. Winker, C. Kittaka, R. E. Kuehn, B. J. Getzewich, C. R. Trepte, and C. A. Hostetler (2009), The CALIPSO lidar cloud and aerosol discrimination: Version 2 algorithm and initial assessment of performance, *J. Atmos. Oceanic Technol.*, *26*, 1198–1213, doi:10.1175/2009JTECHA1229.1.

Liu, Z., R. Kuehn, M. Vaughan, D. Winker, A. Omar, K. Powell, C. Trepte, Y. Hu, and C. Hostetler (2010), The CALIPSO cloud and aerosol discrimination: Version 3 algorithm and test results, paper presented at 25th International Laser Radar Conference, Int. Assoc. of Meteorol. and Atmos. Phys., St. Petersburg, Russia.

Lohmann, U., P. Stier, C. Hoose, S. Ferrachat, S. Kloster, E. Roeckner, and J. Zhang (2007), Cloud microphysics and aerosol indirect effects in the global climate model ECHAM5-HAM, *Atmos. Chem. Phys.*, *7*, 3425–3446.

Mann, G. W., et al. (2014), Intercomparison and evaluation of global aerosol microphysical properties among AeroCom models of a range of complexity, *Atmos. Chem. Phys.*, *14*, 4679–4713, doi:10.5194/acp-14-4679-2014.

Myhre, G., et al. (2013), Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations, *Atmos. Chem. Phys.*, *13*, 1853–1877, doi:10.5194/acp-13-1853-2013.

O'Donnell, D., K. Tsigaridis, and J. Feichter (2011), Estimating the direct and indirect effects of secondary organic aerosols using ECHAM5–HAM, *Atmos. Chem. Phys.*, *11*, 8635–8659.

Omar, A. H., et al. (2009), The CALIPSO automated aerosol classification and lidar ratio selection algorithm, *J. Atmos. Oceanic Technol.*, *26*, 1994–2014, doi:10.1175/2009JTECHA1231.1.

Pappalardo, G., et al. (2014), towards an advanced sustainable European aerosol lidar network, *Atmos. Meas. Tech.*, *7*, 2389–2409, doi:10.5194/amt-7-23892014.

Prijith, S. S., S. Suresh Babu, N. B. Lakshmi, S. K. Satheesh, and K. Krishna Moorthy (2016), Meridional gradients in aerosol vertical distribution over Indian Mainland: Observations and model simulations, *Atmos. Environ.*, *125*, 337–345.

Rogers, R. R., et al. (2011), Assessment of the CALIPSO Lidar 532 nm attenuated backscatter calibration using the NASA LaRC airborne High Spectral Resolution Lidar, *Atmos. Chem. Phys.*, *11*, 1295–1311, doi:10.5194/acp-11-1295-2011.

Samset, B. H., et al. (2013), Black carbon vertical profiles strongly affect its radiative forcing uncertainty, *Atmos. Chem. Phys.*, *13*, 2423–2434, doi:10.5194/acp-13-2423-2013.

Samset, B. H., et al. (2014), Modelled black carbon radiative forcing and atmospheric lifetime in AeroCom Phase II constrained by aircraft observations, *Atmos. Chem. Phys.*, *14*(22), 12,465–12,477.

Schultz M., et al. (2007), RETRO report on emission data sets and methodologies for estimating emissions, Workpackage 1, Deliverable D1–6, EU-Contract No. EVK2-CT-2002–00170.

Schulz, M. (2007), Constraining model estimates of the aerosol radiative forcing, Thèse d'Habilitation à Diriger des Recherches, Université Pierre et Marie Curie, Paris VI.

Schulz, M., et al. (2006), Radiative forcing by aerosols as derived from the AeroCom present-day and pre-industrial simulations, *Atmos. Chem. Phys.*, *6*, 5225–5246, doi:10.5194/acp-6-5225-2006.

Schulz, M., M. Chin, and S. Kinne (2009), The aerosol model comparison project, AeroCom, phase II: Clearing up diversity, *IGAC Newslett.*, *41*, 2–11.

Seland, Ø., T. Iversen, A. Kirkevåg, and T. Storelvmo (2008), Aerosol-climate interactions in the CAM-Oslo atmospheric GCM and investigations of associated shortcomings, *Tellus A*, *60*, 459–491.

Sheridan, P. J., E. Andrews, J. A. Ogren, J. L. Tackett, and D. M. Winker (2012), Vertical profiles of aerosol optical properties over central Illinois and comparison with surface and satellite measurements, *Atmos. Chem. Phys.*, *12*, 11,695–11,721, doi:10.5194/acp-12-11695-2012.

Skeie, R. B., T. K. Berntsen, G. Myhre, K. Tanaka, M. M. Kvalevåg, and C. R. Hoyle (2011a), Anthropogenic radiative forcing time series from pre-industrial times until 2010, *Atmos. Chem. Phys.*, *11*, 11,827–11,857, doi:10.5194/acp-11-11827-2011.

Skeie, R. B., T. K. Berntsen, G. Myhre, C. A. Pedersen, J. Ström, S. Gerland, and J. A. Ogren (2011b), Black carbon in the atmosphere and snow, from pre-industrial times until present, *Atmos. Chem. Phys.*, *11*, 6809–6836, doi:10.5194/acp-11-6809-2011.

Takemura, T., H. Okamoto, Y. Maruyama, A. Numaguti, A. Higurashi, and T. Nakajima (2000), Global three-dimensional simulation of aerosol optical thickness distribution of various origins, *J. Geophys. Res.*, *105*, 17,853–17,873, doi:10.1029/2000JD900265.

Takemura, T., T. Nozawa, S. Emori, T. Y. Nakajima, and T. Nakajima (2005), Simulation of climate response to aerosol direct and indirect effects with aerosol transport-radiation model, *J. Geophys. Res.*, *110*, D02202, doi:10.1029/2004JD005029.

Takemura, T., M. Egashira, K. Matsuzawa, H. Ichijo, R. O'ishi, and A. Abe-Ouchi (2009), A simulation of the global distribution and radiative forcing of soil dust aerosols at the Last Glacial Maximum, *Atmos. Chem. Phys.*, *9*, 3061–3073, doi:10.5194/acp-9-3061-2009.

Textor, C., et al. (2006), Analysis and quantification of the diversities of aerosol life cycles within AeroCom, *Atmos. Chem. Phys.*, *6*, 1777–1813, doi:10.5194/acp-6-1777-2006.

Textor, C., et al. (2007), The effect of harmonized emissions on aerosol properties in global models—An AeroCom experiment, *Atmos. Chem. Phys.*, *7*, 4489–4501, doi:10.5194/acp-7-4489-2007.

Tsigaridis, K., D. Koch, and S. Menon (2013), Uncertainties and importance of sea spray composition on aerosol direct and indirect effects, *J. Geophys. Res. Atmos.*, *118*, 220–235, doi:10.1029/2012JD018165.

Tsigaridis, K., et al. (2014), The AeroCom evaluation and intercomparison of organic aerosol in global models, *Atmos. Chem. Phys.*, *14*, 10,845–10,895, doi:10.5194/acp-14-10845-2014.

Van der Werf, G. R., J. T. Randerson, G. J. Collatz, and L. Giglio (2003), Carbon emissions from fires in tropical and subtropical ecosystems, *Global Change Biol.*, *9*, 547–562, doi:10.1046/j.1365-2486.2003.00604.x.

Van der Werf, G. R., J. T. Randerson, L. Giglio, G. J. Collatz, P. S. Kasibhatla, and A. F. Arellano Jr. (2006), Interannual variability in global biomass burning emissions from 1997 to 2004, *Atmos. Chem. Phys.*, *6*, 3423–3441, doi:10.5194/acp-6-3423-2006.

Van der Werf, G. R., J. T. Randerson, L. Giglio, G. J. Collatz, M. Mu, P. S. Kasibhatla, D. C. Morton, R. S. DeFries, Y. Jin, and T. T. van Leeuwen (2010), Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), *Atmos. Chem. Phys.*, *10*, 11,707–11,735, doi:10.5194/acp-10-11707-2010.

Vaughan, M., S. Young, D. Winker, K. Powell, A. Omar, Z. Liu, Y. Hu, and C. Hostetler (2004), Fully automated analysis of space-based lidar data: An overview of the CALIPSO retrieval algorithms and data products, *Proc. SPIE Int. Soc. Opt. Eng.*, *5575*, 16–30.

Vaughan, M., D. Winker, and K. Powell (2005), CALIOP algorithm theoretical basis document. Part 2: Feature detection and layer properties algorithms, PC-SCI-202.01, NASA Langley Res. Cent., Hampton, Va. [Available at http://www-calipso.larc.nasa.gov/resources/project_documentation.php.]

Vaughan, M., et al. (2010), Strategies for improved CALIPSO aerosol optical depth estimates, *25th International Laser Radar Conference (ILRC)*, St. Petersburg, Russia.

Vuolo, M. R., M. Schulz, Y. Balkanski, and T. Takemura (2014), A new method for evaluating the impact of vertical distribution on aerosol radiative forcing in general circulation models, *Atmos. Chem. Phys.*, *14*(2), 877–897.

Winker, D. M., M. A. Vaughan, A. Omar, Y. Hu, K. A. Powell, Z. Liu, W. H. Hunt, and S. A. Young (2009), Overview of the CALIPSO mission and CALIOP data processing algorithms, *J. Atmos. Oceanic Technol.*, *26*, 2310–2323, doi:10.1175/2009JTECHA1281.1.

Winker, D. M., et al. (2010), The CALIPSO mission: A global 3D view of aerosols and clouds, *Bull. Am. Meteorol. Soc.*, *91*, 1211–1229.

Winker, D. M., J. L. Tackett, B. J. Getzewich, Z. Liu, M. A. Vaughan, and R. R. Rogers (2013), The global 3-D distribution of tropospheric aerosols as characterized by CALIOP, *Atmos. Chem. Phys.*, *13*, 3345–3361, doi:10.5194/acp-13-3345-2013.

Young, S. A., and M. A. Vaughan (2009), The retrieval of profiles of particulate extinction from Cloud Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) data: Algorithm description, *J. Atmos. Oceanic Technol.*, *26*, 1105–1119, doi:10.1175/2008JTECHA1221.1.

Young, S. A., M. A. Vaughan, R. E. Kuehn, and D. M. Winker (2013), The retrieval of profiles of particulate extinction from Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) data: Uncertainty and error sensitivity analyses, *J. Atmos. Oceanic Technol.*, *30*, 395–428.

Yu, H., M. Chin, D. M. Winker, A. H. Omar, Z. Liu, C. Kittaka, and T. Diehl (2010), Global view of aerosol vertical distributions from CALIPSO lidar measurements and GOCART simulations: Regional and seasonal variations, *J. Geophys. Res.*, *115*, D00H30, doi:10.1029/2009JD013364.

Yu, H., L. A. Remer, R. A. Kahn, M. Chin, and Y. Zhang (2013), Satellite perspective of aerosol intercontinental transport: From qualitative tracking to quantitative characterization, *Atmos. Res.*, *124*, 73–100.

Zhang, K., et al. (2012), The global aerosol-climate model ECHAM-HAM, version 2: Sensitivity to improvements in process representations, *Atmos. Chem. Phys.*, *12*, 8911–8949, doi:10.5194/acp-12-8911-2012.

Zhang, L., Q. Li, Y. Gu, K. N. Liou, and B. Meland (2013), Dust vertical profile impact on global radiative forcing estimation using a coupled chemical-transport-radiative-transfer model, *Atmos. Chem. Phys.*, *13*, 7097–7114, doi:10.5194/acp-13-7097-2013.