Evaluating the skill of high resolution WRF-Chem simulations in describing drivers of aerosol direct climate forcing at the regional scale

P. Crippa[1], R. C. Sullivan[2], A. Thota[3], S. C. Pryor[2,3]

[1] COMET, School of Civil Engineering and Geosciences, Cassie Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

[2] Department of Earth and Atmospheric Sciences, Bradfield Hall, 306 Tower Road, Cornell University, Ithaca, NY 14853, USA

[3] Pervasive Technology Institute, Indiana University, Bloomington, IN 47405, USA

Correspondence to: P. Crippa (paola.crippa@ncl.ac.uk)

Abstract

Assessing the ability of global and regional models to describe aerosol optical properties is essential to reducing uncertainty in aerosol direct radiative forcing in the contemporary climate and to improving confidence in future projections. Here we evaluate the performance of high-resolution simulations conducted using the Weather Research and Forecasting model with coupled chemistry (WRF-Chem) in capturing spatio-temporal variability of aerosol optical depth (AOD) and Ångström exponent (AE) by comparison with ground- and space- based remotely sensed observations. WRF-Chem is run over eastern North America at a resolution of 12 km for a representative year (2008). A systematic positive bias in simulated AOD relative to observations is found (annual MFB=0.15 and 0.50 when comparing with MODIS and AERONET respectively), whereas the spatial variability is well captured during most months. The spatial correlation of observed and simulated AOD shows a clear seasonal cycle with highest correlation during summer months (r=0.5-0.7) when the aerosol loading is large and more observations are available. The model is biased towards simulation of coarse mode aerosols (annual MFB for AE = -0.10 relative to MODIS and -0.59 for AERONET), but the spatial correlation for AE with observations is 0.3-0.5 during most months, despite the fact that

AE is retrieved with higher uncertainty from the remote sensing observations. WRF-Chem also exhibits high skill in identifying areas of extreme and non-extreme aerosol loading, and its ability to correctly simulate the location and relative intensity of extreme aerosol events (i.e. AOD>75th percentile) varies between 30 and 70% during winter and summer months respectively.

1. Introduction and Objectives

Atmospheric aerosol particles (aerosols) play a major role in dictating Earth's climate by both directly interacting with solar radiation (direct effect) and acting as cloud condensation nuclei and thus changing cloud properties (indirect effect) (Boucher et al., 2013). The global mean aerosol direct effect is estimated to be -0.27 (possible range of -0.77 to +0.23) W m^{-2}, while the indirect effect is -0.55 (-1.33 to -0.06) W m^{-2} (Stocker et al., 2013). Therefore their combined radiative forcing is likely a significant fraction of the overall net anthropogenic climate forcing since pre-industrial times (i.e. 1.13-3.33 W m^{-2} (Stocker et al., 2013)) and a substantial source of uncertainty in quantifying anthropogenic radiative forcing.

Accurate quantification of direct aerosol radiative forcing is strongly dependent on aerosol precursor and primary aerosol emissions. Both have evolved over the past two decades in terms of their spatio-temporal distribution and absolute magnitude. Emissions have generally increased in emerging economies (Kurokawa et al., 2013), biogenic and anthropogenic emissions have altered in response to changing land use and land cover (Wu et al., 2012), and the implementation of pollution control strategies particularly in North America and Europe have resulted in declines in air pollutant emissions (Xing et al., 2015;Giannouli et al., 2011). Therefore there is evidence that aerosol burdens and thus direct climate forcing has varied markedly in the past and may change substantially in the future. Further, although best estimates of global anthropogenic radiative forcing from the aerosol direct and indirect effect are -0.27 and -0.55 W m^{-2} (Stocker et al., 2013) respectively, the short residence time and high spatio-temporal variability of aerosol populations mean their impact on regional climates can be much larger than the global mean but are even more uncertain.

Long-term measurements of aerosol properties are largely confined to aerosol mass (total, PM_{10} or $PM_{2.5}$) in the near-surface layer which may or may not be representative of either the total atmospheric burden (Ford and Heald, 2013;Alston et al., 2012), or radiation extinction and

hence climate forcing. Further, aerosol composition measurements are often a 24-hour integrated sample taken only 1 in 3 days and thus are subject to under sampling. Hence they provide an incomplete description of temporal variability and mean aerosol burdens for model performance evaluation. Columnar remote sensing measurements of aerosol optical properties are available from a range of ground-based and satellite-borne instrumentation, but have only a relatively short period of record, are subject to non-zero measurement uncertainty (and bias), and under-sample the range of atmospheric conditions due to cloud masking and infrequent satellite overpasses. Therefore, regional and global models are most commonly used to quantify historical and contemporary aerosol direct radiative forcing based on simulated properties such as the aerosol optical depth (AOD) and Ångström exponent (AE) (Boucher et al., 2013).

Most global models that include aerosol microphysics have been run at fairly coarse resolution (spatial resolution of the order of 1-2.5°) (Table 1) usually for periods of a few years. The resulting fields of AOD (and less frequently AE) have been evaluated relative to ground-based and satellite-borne remote sensing optical properties measurements (Table 1). However, aerosol populations (and dynamics) are known to exhibit higher spatial variability (and scales) than can be manifest in those models (Kovacs et al.,2006;Kulmala et al., 2011;Santese et al., 2007; Schutgens et al., 2013;Sinzuka and Redemann, 2011). Despite recent improvements in the sophistication of aerosol processes and properties within global models, there are still substantial regional and latitudinal discrepancies in both the magnitude of AOD and other aerosol properties which impact aerosol direct radiative forcing and the degree of model-to-model agreement (Myhre et al., 2013). Thus the skill of these models in reproducing the spatio-temporal variability in the aerosol size distribution, composition, concentration and radiative properties is incompletely characterized. Further large model-to-model variability both in the global mean direct aerosol forcing and in the spatial distribution thereof exists (Kulmala et al., 2011;Myhre et al., 2013) leading to high uncertainty in quantification of aerosol climate forcing. Although a direct comparison between the studies summarized in Table 1 is inherently very difficult due to the different performance metrics reported, and variations in both the model resolution and aerosol descriptions, there is a consistent finding of high spatial variability in model bias, both in sign and magnitude. Correlation coefficients of monthly and seasonal mean AOD from model simulations versus satellite-based measurements are typically in a range ~0.6-0.8 both in global (Colarco et al., 2010;Lee et al., 2015) and regional (Nabat et al., 2015) simulations. However, these correlations are largely reflective of the ability of the models to capture the seasonal cycle and columnar aerosol properties from remote sensing and thus ignore

substantial variability on the synoptic (Sullivan et al., 2015) and meso-scales (Anderson et al., 2003). A wider range of correlation coefficients are reported when comparisons are made to high frequency observations of AOD at the hourly/daily timescale both in global (Sič et al., 2015) and regional (Rea et al., 2015) simulations (r ~0.3-0.8). The largest range of correlation coefficients ([-0.99, 0.9]; Table 1) is reported when simulated AOD is compared with observations from the AErosol RObotic NETwork (AERONET), and appear to be function of temporal averaging, location of AERONET sites and model resolution. Correlations between time series of simulated AE versus AERONET observations are reported less frequently, and when conducted for monthly mean values range from ~0.4 (Li et al., 2015) to ~0.8 (Colarco et al., 2010).

At least some of the variability in model performance, as indicated by the mutual variability with observations described by correlation coefficients, and model-to-model agreement shown in AeroCom Phase II may be attributable to variations in model resolution, differences in gas and particle phase parameterizations and aerosol descriptions. However, there are also variations in the way in which model skill is evaluated and divergent opinions regarding prioritization of future research directions. The direct effect remains poorly quantified at the regional scale, due to uncertainty in aerosol loading, uncertainty and spatio-temporal variability in aerosol physical properties (Colarco et al., 2014) and a relative paucity of rigorous model verification and validation exercises. Confidence in projections of possible future aerosol radiative forcing requires detailed assessment of skill in the current climate, and the need for and benefits of regional downscaling and/or use of high-resolution global models requires careful quantification.

Regional models represent an opportunity to assess if running higher resolution simulations over specific regions of interest improves the characterization of aerosol optical properties of relevance to direct radiative forcing. Assessment of value added (or lack thereof) from high resolution regional versus global coarse resolution models has not been clearly quantified in previous studies (Table 1). Although high-resolution simulations, comparable to those presented herein, have been run, they are over a small temporal and spatial domain (e.g. (Tuccella et al., 2015)), or lack quantitative assessment of aerosol optical properties (e.g. (Tessum et al., 2014)). Thus, quantification of the skill of high-resolution modeling of aerosol optical properties is presented here along with a preliminary analysis of model performance as a function of spatial aggregation. Forthcoming work will include direct comparison to coarser

resolution simulations to quantify the value added (or lack thereof) from increased model resolution.

We evaluate the skill of state-of-the-art high-resolution regional model simulations of climate-relevant aerosol properties using a range of descriptive statistics and investigate possible sources of discrepancies with observations. The impact of aerosols on climate and human health are strengthened under conditions of enhanced aerosol concentrations, thus it is necessary to study and diagnose causes of 'extreme aerosol events' (Chu, 2004;Gkikas et al., 2012), and to evaluate the ability of numerical models to simulate their occurrence, intensity, spatial extent and location. Prior analyses of Level-3 (1° resolution) MODIS AOD over the eastern half of North America have indicated extreme AOD values (> local 90^{th} percentile) are coherent over regional scales (~ 150 km) (Sullivan et al., 2015). Thus, our evaluation exercise also includes an analysis of the spatio-temporal coherence of extreme events.

We applied the Weather Research and Forecasting model with coupled Chemistry (WRF-Chem version 3.6.1) at high resolution (12×12 km) over eastern North America during the year 2008, in the context of a pseudo type-2 downscaling exercise in which the high-resolution model is nested within reanalysis boundary conditions (Castro et al., 2005). The choice of this spatial resolution is taken in part to match the resolution of North American Mesoscale Model that is used for the meteorological lateral boundary conditions and to ensure we capture some mesoscale variability while remaining computationally feasible.

Our evaluation is designed to investigate spatio-temporal variability of aerosol optical properties (i.e. AOD and AE) in their mean and extreme values. Thus, we conduct our evaluation of the simulations using:

(i) High-frequency, disjunct time series data from point measurements at AERONET stations.

(ii) Relatively high-resolution spatial data from lower frequency (once daily or lower) data from polar orbiting satellites (i.e. MODIS and MISR).

We also include intercomparison with daily mean $PM_{2.5}$ concentrations from 1230 surface stations and near-surface $PM_{2.5}$ composition using data from 123 IMPROVE sites. The $PM_{2.5}$ concentrations data for 2008 were obtained from the US Environmental Protection Agency (EPA) AirData web site and represent all available outdoor near-surface 24-hour mean $PM_{2.5}$ measurements in the model domain. Most of these stations report values on a 1 day in 3

schedule. Daily average $PM_{2.5}$ chemical compositions are also available on 1 day in 3 and were accessed online through the IMPROVE data wizard. We further evaluate the WRF-Chem simulations of a key meteorological parameter – precipitation – relative to observations from the Delaware gridded dataset (Matsuura and Willmott, 2009). This data set includes monthly accumulated precipitation data on a 0.5×0.5° grid which is estimated by interpolating station observations from the Global Historical Climatology Network using the spherical version of Shepard's distance-weighting method (Shepard, 1968;Willmott et al., 1985).

This paper is structured as follows. We first describe the settings used in our WRF-Chem simulations and introduce the remote sensing and other data used for model evaluation (Sect. 2). A description of statistical metrics used for the evaluation is also provided. Section 3 presents results of the evaluation of simulated AOD and AE versus observations, as well as findings on extreme AOD values. In Section 4 we summarize our findings and draw conclusions.

2. Methods

2.1 WRF-Chem simulations

The Weather Research and Forecasting Model with coupled chemistry (WRF-Chem, version 3.6.1) (Grell et al., 2005;Fast et al., 2006) is used to simulate aerosol processes over eastern North America during the whole of 2008. The simulation domain comprises 300×300 grid points with 12 km resolution and is centered in southern Indiana (86°W, 39°N). The calendar year 2008 was selected because it is representative of average climate and aerosol conditions in the center of the model domain (near Indianapolis, IN). In 2008, mean T_{max}, T_{min}, precipitation, and wind speed as measured at the National Weather Service Automated Surface Observing Systems (NWS ASOS) station at Indianapolis International Airport are within ±0.25 standard deviations (σ) of the 2000-2013 seasonal means. Further, mean seasonal AOD from Level-3 MODIS retrievals is within ±0.2σ of 2000-2013 mean values. Additionally, choice of this year ensures availability of multiple sources of ground- and space-based measurements of aerosol properties for evaluation of the simulations.

Table 2 provides details of the WRF-Chem simulations. In brief, we used 32 vertical levels up to 50 hPa with telescoping to allow for a good vertical resolution in the boundary layer (i.e. approximately 10 layers below 1 km for non-mountainous regions). Meteorological lateral boundary conditions are provided every 6 hours from the North American Mesoscale Model

(NAM) applied at 12 km resolution. The initial and boundary chemical conditions are based on output from the offline global chemical transport model MOZART-4 (Model for Ozone and Related chemical Tracers, version 4), driven by meteorology from NCEP/NCAR-reanalysis (Pfister et al., 2011;Emmons et al., 2010). Anthropogenic emissions are from the POET (Precursors of Ozone and their Effects in the Troposphere) and the EDGAR (Emissions Database for Global Atmospheric Research) databases. The land cover is specified based on the USGS 24-category data at 3.7 km resolution (Anderson et al., 1976). Anthropogenic point and area emissions at 4 km resolution are input hourly from the U.S. National Emissions Inventory (NEI-05) (US-EPA, 2009) and specified for 19 vertical levels (see Fig. 1 for an overview of the primary aerosol emissions). Biogenic emissions of isoprene, monoterpenes, other biogenic VOC (OVOC), and nitrogen gas emissions from the soil are described as a function of simulated temperature and photosynthetic active radiation (for isoprene) using the model of Guenther (Guenther et al., 1993;Guenther et al., 1994;Simpson et al., 1995). Aerosol and gas phase chemistry are described using the second generation Regional Acid Deposition Model (RADM2) chemical mechanism (Stockwell et al., 1990) and the Modal Aerosol Dynamics Model for Europe (MADE) which incorporates the Secondary Organic Aerosol Model (SORGAM) (Ackermann et al., 1998;Schell et al., 2001). The correct characterization of aerosol optical properties is dependent on model skill in describing particle composition and mixing state (Li et al., 2015;Curci et al., 2014). With this in mind, it is worthy of note that aerosol components are assumed to be internally mixed within each mode (although the composition differs by mode). The standard deviation on the log-normal Aitken and accumulation modes the standard deviations are fixed at 1.6 and 2, respectively. The choice of a modal representation of aerosol size distribution is dictated by the high computational demand of more sophisticated approaches (e.g. sectional description of the aerosol size distribution) for long-term simulations. With the current settings, the 1-year run was completed without restart in 9.5 days (230 hours) on the Cray XE6/XK7 supercomputer (Big Red II) owned by Indiana University using 256 processors distributed on 8 nodes, thus indicating feasibility of this configuration for climate scale simulations. Aerosol, and gas phase concentrations and meteorological properties are saved once hourly. AE from the WRF-Chem simulations is computed using:

$$AE = \frac{\ln \frac{AOD_{400nm}}{AOD_{600nm}}}{\ln \frac{600nm}{400nm}} \quad (1).$$

AOD at wavelengths (λ) of 500 and 550 nm for comparison with MODIS and MISR respectively, are derived using the Ångström power law:

$$AOD_{\lambda} = AOD_{300} \times \frac{\lambda}{300}^{(-AE)} \quad (2).$$

We investigated the wavelength dependence on AE calculation using λ at 300 nm and 1000 nm as proposed in (Kumar et al., 2014) and found that, although AOD estimates are independent on the wavelength range selected, $AE_{400\text{-}600nm}$ is systematically lower than $AE_{300\text{-}1000nm}$. Analyses of AE reported in this study are obtained using λ = 400 and 600 nm since they are closer to those used in AE satellite retrievals.

2.2 Remotely-sensed data

Consistent with previous research (Sect. 1 and Table 1), we evaluate the WRF-Chem simulations using four primary remote sensing products – three are drawn from instruments on the Aqua and Terra satellites, while the fourth is from ground-based radiometers operated as part of the AERONET network. The data sets are as follows:

1. The MODerate resolution Imaging Spectroradiometer (MODIS) instruments aboard the polar-orbiting Terra (~1030 overpass local solar time (LST)) and Aqua (~1330 LST) satellites. They have measured atmospheric aerosol optical properties since 2000 and 2002 respectively, with near-global daily coverage (Remer et al., 2005). Herein we use the Level 2 (L2; 10 km resolution) "dark-target" products of AOD at 550 nm and AE from 470 – 660 nm (Collection 5.1; (Levy et al., 2010)). The L2 AOD uncertainty is ± 0.05 ± 0.15×AOD over land relative to global sun photometer measurements from AERONET; even when no spatiotemporal averaging is used in the comparison (i.e. all combinations of MODIS retrievals within 30 km of an AERONET site and all AERONET retrievals within 30 min of the satellite overpass), 71% of MODIS retrievals fall within a ± 0.05 ± 0.2 ×AOD envelope relative to AERONET over E. CONUS (Hyer et al., 2011). AE is retrieved with higher uncertainty, and tends to exhibit a bi-modality in retrieved values (Levy et al., 2010;Remer et al., 2005) (see Fig. S1). For this reason where we compare WRF-Chem simulated AE with values from MODIS we treat AE as a binary variable, wherein AE<1 is taken as representing

coarse mode dominated aerosol populations and AE>1 indicates fine mode dominated populations (Pereira et al., 2011;Valenzuela et al., 2014).

2. The Multi-angle Imaging Spectroradiometer (MISR) instrument is also aboard the Terra satellite, and measures radiances at four wavelengths from 446 – 886 nm at nine viewing angles from nadir to 70.5°. MISR (L2, 17.6 km resolution) retrieves AOD with lower uncertainty than MODIS (± 0.05×AOD relative to AERONET), but with lower temporal resolution (global coverage in ~ one week) (Kahn et al., 2010;Kahn et al., 2005). Herein, we use the 0.5°×0.5° gridded Level 3 (Ver. 31) AOD (at 555 nm) and AE (calculated from AOD at 443 and 670 nm).
3. Ground-based sun-photometer measurements from 22 AErosol RObotic NETwork (AERONET) (Holben et al., 1998) stations are also used in this study (Fig. 1). This network is highly spatially inhomogeneous, but under cloud-free conditions the observations are available at multiple times during daylight hours. AOD is measured directly by the AERONET sun photometers at seven wavelengths (340, 380, 440, 500, 670, 870, and 1020 nm) with high accuracy (i.e. AOD uncertainty of < 0.01 for λ > 440 nm (Holben et al., 2001)). The Ångström Exponent (AE) is calculated for all available wavelengths within the AOD range. The AE 870-440 nm includes the 870, 670, 500 and 440 nm AOD data. Level-2 aerosol products from AERONET (i.e. cloud screened and quality assured) have been used extensively in satellite and model validation studies (including many of those summarized in Table 1) and are used herein.

To avoid the discontinuity in the MODIS retrieval algorithm due to different assumed aerosol types (Levy et al., 2007), we confine our analyses of model skill to longitudes east of 98°W. Only WRF grid cells with cloud fraction = 0 during the satellite over pass of each grid cell are used in comparison to MODIS/MISR observations, and only grid cells with at least 5 valid observations (both from MODIS/MISR and cloud-screened WRF) during a given month are included in the analyses presented herein. It is worth noting that setting a threshold of 10 observations does not significantly affect the results. For a uniform assessment, L2 MODIS and L3 MISR data have been interpolated from their native grids (and resolutions of 10 km and 0.5˚×0.5˚, respectively) to the WRF-Chem 12 km resolution grid by computing the mean of pixels with valid data within 0.1˚/0.3˚ for MODIS/MISR from the model centroids. The choice of averaging over a slightly larger area than model resolution is dictated by the sparsity of valid satellite retrievals. For AERONET vs. MODIS comparison, we only use the nearest MODIS

data (after regridding to WRF) to each site. Where hourly WRF-Chem output is compared with data from AERONET sites, a station is only included if there are at least 20 simultaneous estimates available, and each AERONET measurement is compared to the nearest WRF-Chem time step and to the grid cell containing the station.

2.3 Statistical methods used in the model evaluation

The primary error metric of overall model performance used herein is the Mean Fractional Bias (Boylan and Russell, 2006):

$$MFB = \frac{1}{N}\sum_{1}^{N}\frac{C_m - C_0}{\frac{C_m + C_0}{2}} \qquad (3).$$

MFB is a useful model performance indicator since it equally weights positive and negative biases. It varies between +2 and -2 and has a value of zero for an ideal model. Where MFB is reported for WRF-Chem versus MODIS/ MISR/AERONET, C_m is the monthly mean AOD or AE simulated by WRF-Chem at a specific location, C_0 refers to the same quantify from remote sensing data (Table 3) and N is the sample size.

The evaluation of WRF-Chem simulations of AOD and AE relative to satellite retrievals (MODIS and MISR) is also summarized using Taylor diagrams (Taylor, 2001) produced from the monthly means for the grid cells with simultaneous data availability. Taylor diagrams synthesize three aspects of model skill focused on evaluations of the spatial fields of the parameter of interest. The correlation coefficient of the modeled vs. observed field which is expressed by the azimuthal position, the root mean squared difference which is proportional to the distance between a point and the reference point on the x-axis (at 1, 0), and the ratio of simulated and observed spatial standard deviation which is proportional to the radial distance from the origin.

To investigate model performance at given locations through time, empirical quantile-quantile (EQQ) plots are constructed using high frequency realizations of AOD and AE at individual locations (AERONET sites) relative to WRF-Chem values simulated in the grid cell containing the measurement site. EQQ plots are thus generated for each of the AERONET stations using all hours when there are simultaneous estimates available from the direct observations and from the numerical simulations. The advantage of EQQ plots is that they make no assumptions

regarding the underlying form of the data, and can be readily used to determine which parts of the modeled distribution deviate from the observations (and thus fall away from a 1:1 line).

The validity of AE estimates is a function of both the absolute magnitude of AOD and the uncertainty in the wavelength dependent AOD. AE provides information regarding the relative abundance of fine to coarse particles. Thus, here we quantify the model skill in reproducing spatial patterns of fine and coarse mode particles observed by MODIS (Terra) by comparing the frequency distribution of AE lower and higher than 1 to distinguish populations dominated by coarse and fine aerosols respectively in WRF-Chem and MODIS (Valenzuela et al., 2014;Pereira et al., 2011). The choice of this threshold reflects the AE distribution. AE simulated by WRF-Chem generally conforms to a single normal distribution centered on 1 during January-April and on 1.3 from May-June to December; AERONET time series also tend to conform to a single mode, while MODIS estimates typically are bimodally distributed (see Fig. S1). We therefore consider the data in the form of a contingency table (Table 4) and compute a χ^2-test to assess if the frequency distribution of fine and coarse particles is the same between MODIS and WRF-Chem. The χ^2 statistic is applied with one degree of freedom and a 99% confidence limit.

As described above, the impact of aerosols on climate and human health are strengthened under conditions of enhanced aerosol concentrations, thus two analyses were undertaken to evaluate the ability of the WRF-Chem simulations to represent extreme AOD values:

1. Evaluation of the spatial patterns of extreme events. Using daily estimates of AOD in each grid cell and month we identified the 75th percentile value across space (i.e. *p75*) as threshold for extreme AOD for WRF-Chem and MODIS separately. Grid cells with AOD exceeding that threshold were classified as exhibiting extreme values. The consistency in the spatial distribution of extreme values as simulated by WRF-Chem relative to MODIS are quantified using three skill statistics: the Accuracy, Hit Rate (*HR*) and Threat Score (*TS*) defined in equations 4-6. In these equations, *WE*, *ME*, *WN* and *MN* correspond to frequency of extreme conditions in WRF-Chem (WE) or MODIS (ME) or not (WN or MN):

$$Accuracy = \frac{WE/ME + WN/MN}{WE/ME + WE/MN + WN/ME + WN/MN} \quad (4)$$

$$HR = \frac{WE/ME}{WE/ME + WN/ME} \quad (5)$$

$$TS = \frac{WE / ME}{WE / ME + WE / MN + WN / ME} \quad (6)$$

The Accuracy describes the fraction of grid cells co-identified as exceeding *p75* or not in MODIS and WRF-Chem, and thus equally weights event and non-event conditions. Since the Accuracy quantifies model skill in correctly identifying both extreme and non-extreme aerosol loadings, it is thus indicative of model performance in capturing the overall AOD spatial variability. In this application, where extreme is identified as the 75th percentile, a value of 0.5 would indicate none of the grid cells experiencing extreme events were reproduced by the model, while 1 would indicate perfect identification of events and non-events. The *HR* and *TS* metrics give 'credit' only those grid cells identified as 'extreme'. For these metrics, a value of 0 indicates no correct identification of grid cells with extreme values, while a perfect model would exhibit a value of 1.

2. Evaluation of the scales of coherence of extreme AOD. For each day during the overpass time and hours of clear sky conditions, we determine if AOD simulated at our reference location (i.e. the center of the domain, in Southern Indiana) is equal or larger than the local p75 for that grid cell and season and then identify all grid cells in the domain that also satisfy the condition of AOD ≥ local p75. The reference location represents the center of gravity of the domain and was previously used by Sullivan et al. (2015) for assessing scales of coherence. In that work they also found the spatial scales of coherence are not sensitive to the precise choice of reference location. For each season, we thus compute the probability of extreme AOD co-occurrence at our reference site and any other grid cell as the frequency of co-occurrence divided by the number of extreme occurrences at the reference location. The spatial scales of extreme AOD are then estimated by binning the radial distance of each grid cell centroid from the domain center into 100 km distance classes. An analogous procedure is applied to L2 MODIS data to compare with simulations.

3. Results

3.1 Evaluation of AOD

Overall WRF-Chem is positively biased relative to remotely-sensed AOD. The spatial MFB is 0.15 (0.14) when computed using all available MODIS measurements from Terra (Aqua) and 0.50 relative to data from the AERONET stations (Table 3). The sign of this bias is consistent across the entire simulation domain (Fig. 2). These results agree with findings from previous

366 regional studies that have also shown an overestimation of AOD by WRF-Chem over eastern
367 North America and Europe (i.e. regions dominated by sulfate aerosols), and underestimation in
368 western US and most of the rest of the globe (Zhang et al., 2012;Colarco et al., 2010;Curci et
369 al., 2014) (Table 1). Higher biases of WRF-Chem simulated annual mean AOD are found in
370 the southern portion of the domain (Fig. 2) where the model also exhibits a positive bias in
371 daily mean near-surface $PM_{2.5}$ relative to observations from 1230 US EPA sites (see Fig. 3 and
372 Fig. S2). We further investigated the bias in $PM_{2.5}$ by comparing WRF-Chem simulations with
373 ground-based measurements of particle composition at 123 IMPROVE sites over our domain.
374 We computed the MFB on a seasonal basis between sulfate and nitrate concentrations in fine
375 mode particles (i.e. Aitken and accumulation mode) versus observations (Fig. 4) and found
376 sulfate concentrations are underestimated almost over the entire domain during winter, whereas
377 a positive bias is present in the other seasons. Conversely, nitrates tend to be overestimated
378 during winter and fall at most sites, whereas they are underestimated during summer. Thus the
379 positive bias in AOD and $PM_{2.5}$ mass particularly during the summer appears to be associated
380 with excess sulfate concentrations.
381 The MFB of WRF-Chem relative to MODIS estimates of AOD is lower than the MFB relative
382 to most of the AERONET stations except for a few sites located along the coast, one polluted
383 site in the northeast and a few land sites in the North/North-West (Fig. 1 and 5a). This is
384 possibly a result of an inability of the model to capture variations in aerosol optical properties
385 occurring at a local scale (below the resolution of 12 km). However, the evaluation statistics
386 for WRF-Chem relative to AERONET did not vary consistently with the classification of
387 AERONET stations. Indeed, the mean MFB for AOD in coastal, polluted and land sites varies
388 between 0.26 (coastal) and 0.67 (land), whereas for AE it varies between -0.72 (coastal) and -
389 0.50 (land). When MODIS is compared to the 22 AERONET stations the MFB is -1.23
390 suggesting an underestimation of AOD from AERONET relative to MODIS. The large bias can
391 be explained noting that the number of co-samples between MODIS is quite small and that
392 MFB is strongly impacted by a few outliers. When we remove the three most biased sites (one
393 land site in the North and two sites along the East coast) the MFB decreases to -0.91.

394 Using very limited data, prior research indicated mesoscale variability (horizontal scales of 40–
395 400 km and temporal scales of 2–48 h) is a common and perhaps universal feature of lower-
396 tropospheric aerosol light extinction [Anderson et al., 2003]. However, we are not aware of
397 prior systematic attempts to quantify and test the universality of AOD scales of coherence over
398 the contiguous US. To test the sensitivity of the MFB in simulated AOD to spatial aggregation,

399 we excluded the first 12 cells to the left and to the top of the simulated domain and averaged
400 the remaining 12×12 km grid cells over the following scales: 24×24, 36×36, 48×48, 72×72,
401 96×96, 144×144, 192×192, 216×216, 288×288, 384×384, 432×432, 576×576, 864×864,
402 1152×1152, 1728×1728, 3456×3456 km. The last spatial average corresponds to a single grid
403 cell encompassing the entire domain (excluding the outer 12 cells located to the West and North
404 of the simulation domain). Each spatial average at a coarser resolution is computed as the mean
405 of all valid 12×12 km grid cells within the averaging area. We then computed the MFB for the
406 regridded WRF-Chem and MODIS data pair and found that, on a yearly basis, MFB is highest
407 at 12km (0.14 for Aqua and 0.15 for Terra) and reaches a first minimum at 72 km for Aqua
408 (MFB=0.13) and 384 km for Terra (MFB=0.13) (see Fig. 6). However, the MFB and hence
409 systematic error in AOD relative to MODIS exhibits only a weak dependence on the level of
410 spatial aggregation. Spatial patterns of monthly mean AOD show largest differences relative to
411 MODIS during winter months in the southern states and near the coastlines, which show MFB
412 up to 0.7, and lower spatial correlation (see Fig. 7a). This may be due to the larger uncertainty
413 in MODIS retrievals near the coast (Anderson et al., 2013), the smaller sample size in the
414 observations (particularly at high latitudes) during December to March or the lower overall
415 AOD. Conversely, the spatial correlation is maximized during the summer (r=0.5-0.7) for
416 MODIS and August for MISR, when most data are available. The spatial variability of monthly
417 mean AOD fields is also well simulated by WRF-Chem during the warm season (months May-
418 August), as indicated by the ratio of the spatial standard deviation which is close to 1. However,
419 σ(AOD) is usually higher in MODIS and/or MISR than in WRF-Chem. The RMSD is largest
420 and the spatial correlation is lowest during September and October, when MFB is also > 0.4 in
421 part because WRF-Chem simulates high AOD and aerosol nitrate and sulfate concentrations
422 over large regions in eastern North America (Fig. S3 and Fig. 4). The high positive bias in these
423 months is also reflected in the near-surface $PM_{2.5}$ concentrations and its composition (Fig. S2
424 and Fig. 4). A possible explanation for the relatively poor model performance during September
425 and October may derive from the simulation of precipitation. During the majority of calendar
426 months, domain averaged precipitation as simulated by WRF-Chem is slightly positively biased
427 relative to the gridded observational data. However, during September and October, the model
428 exhibits a negative bias (of 8-10% relative to observations) and substantial underestimation of
429 precipitation in regions of typically high AOD such as the Ohio River valley and along the east
430 coast (Fig. S4). We also examined the impact of spatial aggregation (at 12, 24, 36, 48, 72 and
431 96 km resolution) on the seasonality of model performance. For AOD the spatial correlations

are largest for most months when data are aggregated to a resolution of 24×24 km and the ratio of spatial standard deviation is also closer to 1 when AOD are spatially aggregated, possibly indicating that the spatial patterns simulated by WRF-Chem at a fine scale do not always match those observed by MODIS (Fig. 8). For AE both spatial correlations and ratio of standard deviations do not vary significantly when data are aggregated to a coarser resolution (Fig. S5). Empirical quantile-quantile plots of AOD at AERONET stations computed for both simultaneous MODIS observations and WRF-Chem with AERONET observations indicate that the positive bias in WRF-Chem simulated values of AOD is evident across much of the probability distribution (5^{th} to 95^{th} percentile values) at most AERONET stations. However, it is worthy of note that WRF-Chem comparisons with AERONET observations occupy much of the same observational range as simultaneous MODIS and AERONET at those sites (Fig. 9a), although the EQQ plot does not necessarily compare the same MODIS-AERONET and WRF-Chem-AERONET data pairs (i.e. the sample used to compare AERONET and MODIS may differ from that used to compare WRF-Chem and AERONET due to the cloud screening procedure). Thus, model simulations reproduce the range and probability of low-uncertainty AERONET measured AOD nearly as well as MODIS.

3.2 Evaluation of AE

Despite the low confidence in AE retrievals from MODIS, the comparison of WRF-Chem with the remote sensing estimates indicates some degree of agreement. The overall MFB of WRF-Chem vs MODIS Terra is -0.09 (-0.11 vs. Aqua) and the correlation between WRF-Chem and MODIS monthly mean AE seems to be independent of season and lies between 0.20 and 0.54 for all months except April, May and November when it is lower, whereas r is always < 0.14 when comparing with MISR (Fig. 7b). The AE RMSD relative to MODIS or MISR does not exhibit a clear seasonal pattern and the ratio of spatial standard deviations in the AE fields is always lower than 1, indicating more spatial variability in the satellite retrievals than in WRF-Chem. The degree to which these results are symptomatic of the difficulties in retrieving AE from the remote sensing observations is unclear. When the AE values are treated as binary samples (AE < 1 indicating coarse mode aerosols dominate, while AE > 1 indicating a dominance of the fine mode) and presented as a contingency table, WRF-Chem and MODIS simultaneously identify coarse mode dominance (i.e. AE < 1) in 18% of grid cells (Table 5). WRF-Chem simulates 31% of grid cells as exhibiting annual mean AE > 1, while MODIS indicates a larger fraction of grid cells with AE > 1 (80%, Table 5). Both WRF-Chem and MODIS indicate the highest prevalence of fine mode particles during the warm months with

highest agreement for co-identification (above 50%) during June-September. Co-identification of coarse mode particles is highest in the winter and spring months (above 20% during February-May and December, Table 5). However, when a χ^2 test is applied to the frequency of fine and coarse particles identified by WRF-Chem and MODIS, for all months except January and April, the p-value is <0.01, thus we reject the null hypothesis of equal distribution of fine and coarse mode particles identified by MODIS and WRF-Chem. The two data sets agree on 29% of the cases when trying to identify fine mode particles and approximately 53% of the cells are misclassified with MODIS usually identifying a high prevalence of fine aerosols than WRF-Chem. AE from WRF-Chem is also negatively biased relative to AERONET observations, with MFB = -0.59 indicating a greater prevalence of coarse mode aerosols in the simulations (Table 3, Fig. 2).

EQQ plots for all sites show good accord between WRF-Chem and AERONET observations, as indicated by the relatively consistent fractional error across the entire range of simulated and observed AE (Fig. 9b). Simulations from previous studies have also shown a systematic negative bias of simulated AE versus MODIS observations. AE is very difficult to derive from the MODIS measurements and the uncertainty in AE scales with AOD (AE is very uncertain at AOD < 0.2). Further, AE is derived from wavelength dependent AOD, thus the uncertainties on the measurements are certainly correlated. As indicated in Figure 5, for some AERONET sites there is evidence that positive bias in AOD is associated with high negative bias in AE, but this does not uniformly occur over eastern North America (e.g. for the site at 77.8W 55.3N WRF-Chem exhibits positive bias in AOD across the entire pdf while the simulated AE is negative biased, but the site at 84.28W 35.95N exhibits relative good accord for AOD but is negative biased in AE almost to the same amount as the northern station). Highest biases have been noted in regions dominated by dust aerosols or when the model overestimates the dust loading, since aerosol population mean diameter is inversely proportional to AE (Colarco et al., 2014;Balzarini et al., 2014). Sources of the biases in our study, include the simplified treatment of the size distribution, weaknesses in the emission inventory or uncertainties in meteorological variables affecting particle growth (e.g. temperature and relative humidity). Future work will focus on examining these sensitivities.

3.3 AOD Extremes

Averaged across the entire simulation period, WRF-Chem correctly identifies 70% of locations with extreme and non-extreme AOD in the MODIS observations (i.e. the Accuracy = 70%,

Table 6). The overall *TS* and *HR* also indicate the geographic location of extreme AOD is similar between the model and satellite retrievals. The annual mean *HR*, which is defined as the proportion of grid cells with extreme AOD co-identified by WRF-Chem and MODIS relative to MODIS extremes, is 41%. The annual mean *TS*, which also takes into account false alarms, is 27% (Table 6).

For each month, the *HR* is significantly higher than the probability of co-identification of extremes by random chance (i.e. $p_0 = 0.25^2=0.0625$), since the test statistic $\frac{HR - p_0^{\,2}}{\sqrt{\frac{p_0 \times (1 - p_0)}{N}}}$ is always larger than the critical value at 1% (i.e. 2.575). *HR* and *TS* vary seasonally, with highest skill during summer months (*HR* up to 70% and *TS* up to 54%), and lowest skill during winter and early spring (minimum *HR*=29% and minimum *TS*=17%) (Table 6 and Fig. 10). The relatively low skill in identifying the spatial occurrence of high AOD during winter and spring may reflect the relatively low AOD and low spatial variability during this season, which means 'extreme' AOD may differ only marginally from the 'non-extreme' areas (see Fig. S6 for monthly comparisons of extreme area identification).

The spatial distribution of extreme AOD also displays some seasonality with areas of AOD > *p75* concentrated over coastal regions and the southern states during summer months and smaller areas during winter and early spring (Fig. 10). Despite the relatively low simultaneous identification of extremes during cold seasons, the location of extremes moves from the coast to the Great Lakes region and Midwest states in both the model and MODIS (see Fig. S6). During winter and spring months WRF-Chem simulates more areas with extreme AOD over coastal regions, whereas MODIS shows more spatial variability and predicts higher AOD in the Great Lakes area and in the states west of Illinois. Conversely, WRF-Chem underestimates areas of extreme AOD relative to MODIS in the northern regions of the domain, possibly due to the underestimation of sulfate-aerosol. These two observations may be explained by noting that the mass fraction of aerosol nitrate in the accumulation and coarse mode predicted by WRF-Chem during most of fall and winter months dominates the sulfate fraction over virtually all of the domain (see Fig. S3), whereas point observations indicate aerosol nitrate mass fraction is dominant only over the Central Great Plains (Hand et al., 2012). This may be related to an overestimation of aerosol-nitrate in winter and fall (Fig. 4) as a result of the impact of air temperature and relative humidity on aerosol ammonium nitrate (NH_4NO_3) stability

(Aksoyoglu et al., 2011), as well as an underestimation of aerosol sulfate, mostly during winter (Fig. 4), likely due to underestimation of the rate of SO_2 gaseous and aqueous (missing) oxidation, or underestimation of the nighttime boundary layer height which impacts sulfate formation near the surface (Tuccella et al., 2012). Localized negative biases in the model over the coast may be associated with the higher uncertainties in MODIS retrievals at coastlines (Anderson et al., 2013).

Extreme AOD exhibits relatively large spatial scales of coherence in both the WRF-Chem simulations and MODIS L2 observations (Fig. 11). Consistent with prior analyses of L3 MODIS data (Sullivan et al., 2015), the largest scales of coherence are found in fall. In all seasons except winter the probability of co-occurrence of extremes at the domain center and any other grid cell in the simulation domain is > 0.5 up to a distance of 300 km. The simulated mean seasonal scales of extreme coherence are comparable to L2 MODIS AOD (Fig. 11), despite the larger variability in the MODIS data due to the limited retrievals with simultaneous extreme AOD at the reference location and each other grid cell. Thus, consistent with prior research this analysis indicates the occurrence of extreme AOD occurs on large spatial scales and therefore may significantly impact regional climate.

4. Discussion and concluding remarks

Aerosol direct and indirect radiative forcing on the climate system are highly uncertain. A systematic assessment of the ability of global and regional models to reproduce aerosol optical properties in the contemporary climate is essential to increasing confidence in future projections. We contribute to this growing literature by presenting high resolution (12 km) simulations from WRF-Chem conducted over eastern North America during a year representative of average meteorological and aerosol conditions. We evaluate the simulations relative to daily MODIS and MISR observations, high frequency AERONET measurements of AOD and AE and near-surface $PM_{2.5}$ mass and composition measurements. Results from this study show:

- After grid cells with any cloud presence are removed and considering only overpass hours, the domain averaged simulated mean AOD is 0.22. Simulated AOD is positively biased relative to observations, with MFB=0.14 when comparing with MODIS-Aqua and 0.50 relative to AERONET (Fig. 1 and 2). A clear north-south gradient in AOD bias vs. MODIS is also observed. This positive bias is consistent across the entire probability distribution at most AERONET stations (Fig. 9), and is also evident in

559 comparison of modeled near-surface $PM_{2.5}$ mass relative to daily mean observations
560 distributed at 1230 stations across the domain (Fig. 3).
561 • Model skill in reproducing the spatial fields of monthly mean AOD as measured by the
562 spatial correlation and ratio of the spatial variability with MODIS is maximized during
563 the summer months ($r \sim 0.5$-0.7, and ratio of $\sigma \sim 0.8$ to 1.2). During this season observed
564 AOD is higher and more observations are available (Fig. 7). Lowest model-observations
565 agreement is found in September and October and is at least partially attributable to a
566 dry bias in precipitation from WRF-Chem (Fig. S4).
567 • In part because of the difficulties in retrieving robust estimates of AE, few previous
568 studies have evaluated model simulated AE values. We show that AE as simulated by
569 WRF-Chem over eastern North America is negatively biased relative to MODIS
570 (MFB=-0.10) and AERONET (MFB=-0.59). This bias indicates WRF-Chem simulates
571 a larger fraction of coarse mode particles than is evident in the remote sensing
572 observations (see Table 3 and 5). While some of the bias relative to MODIS may reflect
573 high observational uncertainty, the large bias relative to AERONET is consistent with
574 prior research (Table 1), and is symptomatic of substantial systematic error in the
575 aerosol size distribution.
576 • Causes of the model error may include insufficiently detailed treatment of size
577 distribution or inaccurate representation of aerosol composition and mixing state which
578 affect the simulated size distribution and thus AE (Li et al., 2015;Curci et al., 2014)).
579 Further, weaknesses in the emission inventory (e.g. size resolution of primary
580 emissions), as suggested by the systematic bias in simulated $PM_{2.5}$ concentrations
581 relative to ground-based observations, and/or biases in the representation of
582 meteorological conditions critical to determining aerosol nitrate concentrations may
583 also affect model performance. Currently it is not possible to fully attribute the relative
584 importance of these error sources.
585 • The majority of prior model evaluation exercises have tended to focus on mean AOD
586 values. However, the climate and health impacts of aerosols are greater under high
587 aerosol loadings. We demonstrate that WRF-Chem exhibits some skill in capturing the
588 spatial patterns of extreme aerosol loading, especially during summer months. During
589 this season, the Hit Rate for AOD > *p75* reaches 70%. Largest biases are found during
590 winter months and near the coastlines where AOD from MODIS also exhibits largest
591 retrieval uncertainty.

Despite the encouraging performance of WRF-Chem both in terms of simulation efficiency and in reproducing AOD (mean and extreme values) and the partial skill in reproducing AE over eastern North America, further investigations are needed to properly quantify the value added by running high-resolution simulations by direct comparison with analogous runs at coarser resolution. Future simulations will also involve assessment of accuracy of different aerosol schemes (i.e. sectional vs. modal approaches) to represent the size distribution. The inclusion of a direct description of new particle formation processes within WRF-Chem may also improve estimates of ultrafine particle concentrations and thus of simulated aerosol optical properties.

5. Acknowledgements

This research was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute, and in part by the Indiana METACyt Initiative. The Indiana METACyt Initiative at IU is also supported in part by Lilly Endowment, Inc. Additional support was provided by the L'Oréal-UNESCO UK and Ireland Fellowship For Women In Science (to PC), the Natural Environmental Research Council (NERC) through the LICS project (ref. NE/K010794/1), the US NSF (grants # 1102309 and 1517365 to SCP) and a NASA Earth and Space Science Fellowship Program - Grant "14-EARTH14F-0207" (to RCS). The data used in this study were acquired as part of the NASA's Earth-Sun System Division, and archived and distributed by the MODIS Level 1 and Atmosphere Archive and Distribution System (LAADS), and the Giovanni online data system, developed and maintained by the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC). We thank the PI investigators and their staff for establishing and maintaining the 22 AERONET sites used in this investigation. $PM_{2.5}$ surface concentrations from the United States Environmental Protection Agency were obtained from: http://www.epa.gov/airquality/airdata/ad_data_daily.html. Meteorological lateral boundary conditions from the North American Mesoscale model were obtained from the NOAA Operational Model Archive and Distribution System: ftp://nomads.ncdc.noaa.gov/NAM/analysis_only/.

6. References

Ackermann, I. J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F. S., and Shankar, U.: Modal aerosol dynamics model for Europe: development and first applications, Atmos. Environ., 32, 2981-2999, http://dx.doi.org/10.1016/S1352-2310(98)00006-5, 1998.

Aksoyoglu, S., Keller, J., Barmpadimos, I., Oderbolz, D., Lanz, V. A., Prévôt, A. S. H., and Baltensperger, U.: Aerosol modelling in Europe with a focus on Switzerland during summer and winter episodes, Atmos. Chem. Phys., 11, 7355-7373, 10.5194/acp-11-7355-2011, 2011.

Alston, E. J., Sokolik, I. N., and Kalashnikova, O. V.: Characterization of atmospheric aerosol in the US Southeast from ground- and space-based measurements over the past decade, Atmospheric Measurement Techniques, 5, 1667-1682, 10.5194/amt-5-1667-2012, 2012.

Anderson, J. C., Wang, J., Zeng, J., Leptoukh, G., Petrenko, M., Ichoku, C., and Hu, C.: Long-term statistical assessment of Aqua-MODIS aerosol optical depth over coastal regions: bias characteristics and uncertainty sources, Tellus Series B-Chemical and Physical Meteorology, 65, 10.3402/tellusb.v65i0.20805, 2013.

Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E.: A land use and land cover classification system for use with remote sensor data, Report 964, 1976.

Anderson, T. L., Charlson, R. J., Winker, D. M., Ogren, J. A., and Holmén, K.: Mesoscale Variations of Tropospheric Aerosols*, Journal of the Atmospheric Sciences, 60, 119-136, 10.1175/1520-0469(2003)060<0119:MVOTA>2.0.CO;2, 2003.

Balzarini, A., Pirovano, G., Honzak, L., Žabkar, R., Curci, G., Forkel, R., Hirtl, M., San José, R., Tuccella, P., and Grell, G. A.: WRF-Chem model sensitivity to chemical mechanisms choice in reconstructing aerosol optical properties, Atmos. Environ., http://dx.doi.org/10.1016/j.atmosenv.2014.12.033, 2014.

Boucher, O., D. Randall, and P. Artaxo, C. B., G. Feingold, P. Forster, V.-M. Kerminen, Y. Kondo, H. Liao, U. Lohmann, P. Rasch, S.K. Satheesh, S. Sherwood, B. Stevens and X.Y. Zhang: Clouds and Aerosols, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 33–115, 2013.

Boylan, J. W., and Russell, A. G.: PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models, Atmos. Environ., 40, 4946-4959, http://dx.doi.org/10.1016/j.atmosenv.2005.09.087, 2006.

Castro, C. L., Pielke, R. A., and Leoncini, G.: Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS), Journal of Geophysical Research: Atmospheres, 110, D05108, 10.1029/2004JD004721, 2005.

Chu, S. H.: PM2.5 episodes as observed in the speciation trends network, Atmos. Environ., 38, 5237-5246, 10.1016/j.atmosenv.2004.01.055, 2004.

Colarco, P., da Silva, A., Chin, M., and Diehl, T.: Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth, J. Geophys. Res.-Atmos., 115, 10.1029/2009jd012820, 2010.

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

Curci, G., Hogrefe, C., Bianconi, R., Im, U., Balzarini, A., Baró, R., Brunner, D., Forkel, R., Giordano, L., Hirtl, M., Honzak, L., Jiménez-Guerrero, P., Knote, C., Langer, M., Makar, P. A., Pirovano, G., Pérez, J. L., San José, R., Syrakov, D., Tuccella, P., Werhahn, J., Wolke, R., Žabkar, R., Zhang, J., and Galmarini, S.: Uncertainties of simulated aerosol optical properties induced by assumptions on aerosol physical and chemical properties: An AQMEII-2 perspective, Atmos. Environ., http://dx.doi.org/10.1016/j.atmosenv.2014.09.009, 2014.

de Meij, A., Pozzer, A., Pringle, K. J., Tost, H., and Lelieveld, J.: EMAC model evaluation and analysis of atmospheric aerosol properties and distribution with a focus on the Mediterranean region, Atmospheric Research, 114–115, 38-69, http://dx.doi.org/10.1016/j.atmosres.2012.05.014, 2012.

Drury, E., Jacob, D. J., Spurr, R. J. D., Wang, J., Shinozuka, Y., Anderson, B. E., Clarke, A. D., Dibb, J., McNaughton, C., and Weber, R.: Synthesis of satellite (MODIS), aircraft (ICARTT), and surface (IMPROVE, EPA-AQS, AERONET) aerosol observations over eastern North America to improve MODIS aerosol retrievals and constrain surface aerosol concentrations and sources, J. Geophys. Res.-Atmos., 115, 10.1029/2009jd012629, 2010.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J. F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geoscientific Model Development, 3, 43-67, 2010.

Fast, J. D., Gustafson, W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., Grell, G. A., and Peckham, S. E.: Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model, Journal of Geophysical Research: Atmospheres, 111, D21305, 10.1029/2005JD006721, 2006.

Ford, B., and Heald, C. L.: Aerosol loading in the Southeastern United States: reconciling surface and satellite observations, Atmospheric Chemistry and Physics, 13, 9269-9283, 10.5194/acp-13-9269-2013, 2013.

Giannouli, M., Kalognomou, E.-A., Mellios, G., Moussiopoulos, N., Samaras, Z., and Fiala, J.: Impact of European emission control strategies on urban and local air quality, Atmos. Environ., 45, 4753-4762, 10.1016/j.atmosenv.2010.03.016, 2011.

Gkikas, A., Houssos, E. E., Hatzianastassiou, N., Papadimas, C. D., and Bartzokas, A.: Synoptic conditions favouring the occurrence of aerosol episodes over the broader Mediterranean basin, Quarterly Journal of the Royal Meteorological Society, 138, 932-949, 10.1002/qj.978, 2012.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmos. Environ., 39, 6957-6975, 10.1016/j.atmosenv.2005.04.027, 2005.

Guenther, A., Zimmerman, P., and Wildermuth, M.: Natural volatile organic compound emission rate estimates for U.S. woodland landscapes, Atmos. Environ., 28, 1197-1210, 10.1016/1352-2310(94)90297-6, 1994.

Guenther, A. B., Zimmerman, P. R., Harley, P. C., Monson, R. K., and Fall, R.: Isoprene and monoterpene emission rate variability: model evaluations and sensitivity analyses, J. Geophys. Res.-Atmos., 98, 12609-12617, 10.1029/93jd00527, 1993.

Kovacs, T. (2006). "Comparing MODIS and AERONET aerosol optical depth at varying separation distances to assess ground-based validation strategies for spaceborne lidar." J. Geophys. Res-Atmos. 111(D24).

Hand, J. L., Schichtel, B. A., Pitchford, M., Malm, W. C., and Frank, N. H.: Seasonal composition of remote and urban fine particulate matter in the United States, Journal of Geophysical Research: Atmospheres, 117, D05209, 10.1029/2011JD017122, 2012.

Hyer, E. J., Reid, J. S., and Zhang, J.: An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals, Atmospheric Measurement Techniques, 4, 379-408, 10.5194/amt-4-379-2011, 2011.

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET - A federated instrument network and data archive for aerosol characterization, Remote Sensing of Environment, 66, 1-16, 10.1016/s0034-4257(98)00031-5, 1998.

Holben, B. N., Tanre, D., Smirnov, A., Eck, T. F., Slutsker, I., Abuhassan, N., Newcomb, W. W., Schafer, J. S., Chatenet, B., Lavenu, F., Kaufman, Y. J., Castle, J. V., Setzer, A., Markham, B., Clark, D., Frouin, R., Halthore, R., Karneli, A., O'Neill, N. T., Pietras, C., Pinker, R. T., Voss, K., and Zibordi, G.: An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET, J. Geophys. Res.-Atmos., 106, 12067-12097, 10.1029/2001jd900014, 2001.

Kahn, R. A., Gaitley, B. J., Martonchik, J. V., Diner, D. J., Crean, K. A., and Holben, B.: Multiangle Imaging Spectroradiometer (MISR) global aerosol optical depth validation based on 2 years of coincident Aerosol Robotic Network (AERONET) observations, J. Geophys. Res.-Atmos., 110, 10.1029/2004jd004706, 2005.

Kahn, R. A., Gaitley, B. J., Garay, M. J., Diner, D. J., Eck, T. F., Smirnov, A., and Holben, B. N.: Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network, J. Geophys. Res.-Atmos., 115, 10.1029/2010jd014601, 2010.

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

Kulmala, M., Asmi, A., Lappalainen, H. K., Baltensperger, U., Brenguier, J. L., Facchini, M. C., Hansson, H. C., Hov, O., O'Dowd, C. D., Poeschl, U., Wiedensohler, A., Boers, R., Boucher, O., de Leeuw, G., van der Gon, H. A. C. D., Feichter, J., Krejci, R., Laj, P., Lihavainen, H., Lohmann, U., McFiggans, G., Mentel, T., Pilinis, C., Riipinen, I., Schulz, M., Stohl, A., Swietlicki, E., Vignati, E., Alves, C., Amann, M., Ammann, M., Arabas, S., Artaxo, P., Baars, H., Beddows, D. C. S., Bergstrom, R., Beukes, J. P., Bilde, M., Burkhart, J. F., Canonaco, F., Clegg, S. L., Coe, H., Crumeyrolle, S., D'Anna, B., Decesari, S., Gilardoni, S., Fischer, M., Fjaeraa, A. M., Fountoukis, C., George, C., Gomes, L., Halloran, P., Hamburger, T., Harrison, R. M., Herrmann, H., Hoffmann, T., Hoose, C., Hu, M., Hyvarinen, A., Horrak, U., Iinuma, Y., Iversen, T., Josipovic, M., Kanakidou, M., Kiendler-Scharr, A., Kirkevag, A., Kiss, G., Klimont, Z., Kolmonen, P., Komppula, M., Kristjansson, J. E., Laakso, L., Laaksonen, A., Labonnote, L., Lanz, V. A., Lehtinen, K. E. J., Rizzo, L. V., Makkonen, R., Manninen, H. E.,

McMeeking, G., Merikanto, J., Minikin, A., Mirme, S., Morgan, W. T., Nemitz, E., O'Donnell, D., Panwar, T. S., Pawlowska, H., Petzold, A., Pienaar, J. J., Pio, C., Plass-Duelmer, C., Prevot, A. S. H., Pryor, S., Reddington, C. L., Roberts, G., Rosenfeld, D., Schwarz, J., Seland, O., Sellegri, K., Shen, X. J., Shiraiwa, M., Siebert, H., Sierau, B., Simpson, D., Sun, J. Y., Topping, D., Tunved, P., Vaattovaara, P., Vakkari, V., Veefkind, J. P., Visschedijk, A., Vuollekoski, H., Vuolo, R., Wehner, B., Wildt, J., Woodward, S., Worsnop, D. R., van Zadelhoff, G. J., Zardini, A. A., Zhang, K., van Zyl, P. G., Kerminen, V. M., Carslaw, K. S., and Pandis, S. N.: General overview: European Integrated project on Aerosol Cloud Climate and Air Quality interactions (EUCAARI) - integrating aerosol research from nano to global scales, Atmospheric Chemistry and Physics, 11, 13061-13143, 10.5194/acp-11-13061-2011, 2011.

Kumar, R., Barth, M. C., Pfister, G. G., Naja, M., and Brasseur, G. P.: WRF-Chem simulations of a typical pre-monsoon dust storm in northern India: influences on aerosol optical properties and radiation budget, Atmospheric Chemistry and Physics, 14, 2431-2446, 10.5194/acp-14-2431-2014, 2014.

Kurokawa, J., Ohara, T., Morikawa, T., Hanayama, S., Janssens-Maenhout, G., Fukui, T., Kawashima, K., and Akimoto, H.: Emissions of air pollutants and greenhouse gases over Asian regions during 2000-2008: Regional Emission inventory in ASia (REAS) version 2, Atmospheric Chemistry and Physics, 13, 11019-11058, 10.5194/acp-13-11019-2013, 2013.

Lee, Y. H., Adams, P. J., and Shindell, D. T.: Evaluation of the global aerosol microphysical ModelE2-TOMAS model against satellite and ground-based observations, Geosci. Model Dev., 8, 631-667, 10.5194/gmd-8-631-2015, 2015.

Levy, R. C., Remer, L. A., and Dubovik, O.: Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land, J. Geophys. Res.-Atmos., 112, 15, 10.1029/2006jd007815, 2007.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., and Eck, T. F.: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, Atmospheric Chemistry and Physics, 10, 10399-10420, 10.5194/acp-10-10399-2010, 2010.

Li, S., Kahn, R., Chin, M., Garay, M. J., and Liu, Y.: Improving satellite-retrieved aerosol microphysical properties using GOCART data, Atmos. Meas. Tech., 8, 1157-1171, 10.5194/amt-8-1157-2015, 2015.

Matsuura, K., and Willmott, C. J.: Terrestrial precipitation: 1900–2008 gridded monthly time series, 2009.

Michou, M., Nabat, P., and Saint-Martin, D.: Development and basic evaluation of a prognostic aerosol scheme (v1) in the CNRM Climate Model CNRM-CM6, Geosci. Model Dev., 8, 501-531, 10.5194/gmd-8-501-2015, 2015.

Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., Bian, H., Bellouin, N., Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D., Iversen, T., Kinne, S., Kirkevag, A., Lamarque, J. F., Lin, G., Liu, X., Lund, M. T., Luo, G., Ma, X., van Noije, T., Penner, J. E., Rasch, P. J., Ruiz, A., Seland, O., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., Wang, P., Wang, Z., Xu, L., Yu, H., Yu, F., Yoon, J. H., Zhang, K., Zhang, H., and Zhou, C.: Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations, Atmospheric Chemistry and Physics, 13, 1853-1877, 10.5194/acp-13-1853-2013, 2013.

Nabat, P., Somot, S., Mallet, M., Michou, M., Sevault, F., Driouech, F., Meloni, D., di Sarra, A., Di Biagio, C., Formenti, P., Sicard, M., Léon, J. F., and Bouin, M. N.: Dust aerosol radiative

794 effects during summer 2012 simulated with a coupled regional aerosol–atmosphere–ocean
795 model over the Mediterranean, Atmos. Chem. Phys., 15, 3303-3326, 10.5194/acp-15-3303-
796 2015, 2015.

797 Nair, V. S., Solmon, F., Giorgi, F., Mariotti, L., Babu, S. S., and Moorthy, K. K.: Simulation of
798 South Asian aerosols for regional climate studies, Journal of Geophysical Research:
799 Atmospheres, 117, n/a-n/a, 10.1029/2011JD016711, 2012.

800 Pereira, S. N., Wagner, F., and Silva, A. M.: Seven years of measurements of aerosol scattering
801 properties, near the surface, in the southwestern Iberia Peninsula, Atmospheric Chemistry and
802 Physics, 11, 17-29, 10.5194/acp-11-17-2011, 2011.

803 Pfister, G. G., Parrish, D. D., Worden, H., Emmons, L. K., Edwards, D. P., Wiedinmyer, C.,
804 Diskin, G. S., Huey, G., Oltmans, S. J., Thouret, V., Weinheimer, A., and Wisthaler, A.:
805 Characterizing summertime chemical boundary conditions for airmasses entering the US West
806 Coast, Atmos. Chem. Phys., 11, 1769-1790, 10.5194/acp-11-1769-2011, 2011.

807 Rea, G., Turquety, S., Menut, L., Briant, R., Mailler, S., and Siour, G.: Source contributions to
808 2012 summertime aerosols in the Euro-Mediterranean region, Atmos. Chem. Phys. Discuss.,
809 15, 8191-8242, 10.5194/acpd-15-8191-2015, 2015.

810 Remer, L. A., Kaufman, Y. J., Tanre, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R. R.,
811 Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The
812 MODIS aerosol algorithm, products, and validation, Journal of the Atmospheric Sciences, 62,
813 947-973, 10.1175/jas3385.1, 2005.

814 Schell, B., Ackermann, I. J., Hass, H., Binkowski, F. S., and Ebel, A.: Modeling the formation
815 of secondary organic aerosol within a comprehensive air quality model system, J. Geophys.
816 Res.-Atmos., 106, 28275-28293, 10.1029/2001jd000384, 2001.

817 Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data, Proceedings
818 of the 1968 23rd ACM national conference, 1968.

819 Sič, B., El Amraoui, L., Marécal, V., Josse, B., Arteta, J., Guth, J., Joly, M., and Hamer, P. D.:
820 Modelling of primary aerosols in the chemical transport model MOCAGE: development and
821 evaluation of aerosol physical parameterizations, Geosci. Model Dev., 8, 381-408,
822 10.5194/gmd-8-381-2015, 2015.

823 Simpson, D., Guenther, A., Hewitt, C. N., and Steinbrecher, R.: Biogenic emissions in Europe.
824 1. estimates and uncertainties, J. Geophys. Res.-Atmos., 100, 22875-22890,
825 10.1029/95jd02368, 1995.

826 Spracklen, D. V., Carslaw, K. S., Merikanto, J., Mann, G. W., Reddington, C. L., Pickering, S.,
827 Ogren, J. A., Andrews, E., Baltensperger, U., Weingartner, E., Boy, M., Kulmala, M., Laakso,
828 L., Lihavainen, H., Kivekäs, N., Komppula, M., Mihalopoulos, N., Kouvarakis, G., Jennings,
829 S. G., O'Dowd, C., Birmili, W., Wiedensohler, A., Weller, R., Gras, J., Laj, P., Sellegri, K.,
830 Bonn, B., Krejci, R., Laaksonen, A., Hamed, A., Minikin, A., Harrison, R. M., Talbot, R., and
831 Sun, J.: Explaining global surface aerosol number concentrations in terms of primary emissions
832 and particle formation, Atmos. Chem. Phys., 10, 4775-4793, 10.5194/acp-10-4775-2010, 2010.

833 Stocker, T. F., Qin, D., and Plattner, G.-K., Alexander, L.V., Allen, S.K., Bindoff, N.L., Bréon,
834 F.-M., Church, J.A., Cubasch, U., Emori, S., Forster, P., Friedlingstein, P., Gillett, N., Gregory,
835 J.M., ,Hartmann, D.L., Jansen, E., Kirtman, B., Knutti, R., Krishna Kumar, K. and Lemke, P.
836 and Marotzke, J., Masson-Delmotte, V. , Meehl, G.A., Mokhov, I.I., Piao, S. , Ramaswamy,
837 V., Randall, D., Rhein, M., Rojas, M., Sabine, C., Shindell, D., Talley, L.D., Vaughan, D.G.,

838 Xie, S.-P.: Summary for Policymakers, in: Climate Change 2013: The Physical Science Basis.
839 Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental
840 Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New
841 York, NY, USA, 33–115, 2013.

842 Santese, M., F. De Tomasi and M. R. Perrone (2007). "AERONET versus MODIS aerosol
843 parameters at different spatial resolutions over southeast Italy." J. Geophys. Res-Atmos.
844 112(D10).

845 Schutgens, N. A. J., M. Nakata and T. Nakajima (2013). "Validation and empirical correction
846 of MODIS AOT and AE over ocean." Atmos. Meas. Techn. 6(9): 2455-2475.

847 Shinozuka, Y. and J. Redemann (2011). "Horizontal variability of aerosol optical depth
848 observed during the ARCTAS airborne experiment." Atmos. Chem. Phys. 11(16): 8489-8495.

849 Stockwell, W. R., Middleton, P., Chang, J. S., and Tang, X.: The second generation regional
850 acid deposition model chemical mechanism for regional air quality modeling, Journal of
851 Geophysical Research: Atmospheres, 95, 16343-16367, 10.1029/JD095iD10p16343, 1990.

852 Sullivan, R. C., Levy, R. C., and Pryor, S. C.: Spatiotemporal coherence of mean and extreme
853 aerosol particle events over eastern North America as observed from satellite, Atmos. Environ.,
854 112, 126-135, http://dx.doi.org/10.1016/j.atmosenv.2015.04.026, 2015.

855 Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J.
856 Geophys. Res.-Atmos., 106, 7183-7192, 10.1029/2000jd900719, 2001.

857 Tessum, C. W., Hill, J. D., and Marshall, J. D.: Twelve-month, 12 km resolution North
858 American WRF-Chem v3.4 air quality simulation: performance evaluation, Geosci. Model
859 Dev. Discuss., 7, 8433-8476, 10.5194/gmdd-7-8433-2014, 2014.

860 Tuccella, P., Curci, G., Visconti, G., Bessagnet, B., Menut, L., and Park, R. J.: Modeling of gas
861 and aerosol with WRF/Chem over Europe: Evaluation and sensitivity study, Journal of
862 Geophysical Research: Atmospheres, 117, D03303, 10.1029/2011JD016302, 2012.

863 Tuccella, P., Curci, G., Grell, G. A., Visconti, G., Crumeroylle, S., Schwarzenboeck, A., and
864 Mensah, A. A.: A new chemistry option in WRF/Chem v. 3.4 for the simulation of direct and
865 indirect aerosol effects using VBS: evaluation against IMPACT-EUCAARI data, Geosci.
866 Model Dev. Discuss., 8, 791-853, 10.5194/gmdd-8-791-2015, 2015.

867 US-EPA: 2005 National Emissions Inventory (NEI), US Environmental Protection Agency
868 available at: ftp://aftp.fsl.noaa.gov/divisions/taq/emissions_data_2005/, 2009.

869 Valenzuela, A., Olmo, F. J., Lyamani, H., Granados-Munoz, M. J., Anton, M., Guerrero-
870 Rascado, J. L., Quirantes, A., Toledano, C., Perez-Ramirez, D., and Alados-Arboledas, L.:
871 Aerosol transport over the western Mediterranean basin: Evidence of the contribution of fine
872 particles to desert dust plumes over Alboran Island, J. Geophys. Res.-Atmos., 119, 14028-
873 14044, 10.1002/2014jd022044, 2014.

874 Willmott, C. J., Rowe, C. M., and Philpot, W. D.: Small-Scale Climate Maps: A Sensitivity
875 Analysis of Some Common Assumptions Associated with Grid-Point Interpolation and
876 Contouring, The American Cartographer, 12, 5-16, 10.1559/152304085783914686, 1985.

877 Wu, S., Mickley, L. J., Kaplan, J. O., and Jacob, D. J.: Impacts of changes in land use and land
878 cover on atmospheric chemistry and air quality over the 21st century, Atmospheric Chemistry
879 and Physics, 12, 1597-1609, 10.5194/acp-12-1597-2012, 2012.

880 Xing, J., Mathur, R., Pleim, J., Hogrefe, C., Gan, C. M., Wong, D. C., Wei, C., Gilliam, R., and
881 Pouliot, G.: Observations and modeling of air quality trends over 1990-2010 across the
882 Northern Hemisphere: China, the United States and Europe, Atmospheric Chemistry and
883 Physics, 15, 2723-2747, 10.5194/acp-15-2723-2015, 2015.

884 Zhang, Y., Chen, Y., Sarwar, G., and Schere, K.: Impact of gas-phase mechanisms on Weather
885 Research Forecasting Model with Chemistry (WRF/Chem) predictions: Mechanism
886 implementation and comparative evaluation, J. Geophys. Res.-Atmos., 117,
887 10.1029/2011jd015775, 2012.

Tables

Table 1. Synthesis of some recent prior studies comparing simulated aerosol optical properties from global or regional model simulations with remote sensing products. The first column summarizes the model used, the second the domain and the time period simulated and the third shows the model resolution and summarizes the description of the aerosol size distribution. Columns 4 to 9 summarize the evaluation statistics in terms of the overall correlation coefficient (R), bias (as described using the mean fractional error (MFE)) and root mean square error (RMSE) or mean absolute error (MAE) relative to satellite or AERONET observations as reported in the references shown in column 10.

Model	Domain, Time	Resolution, Aerosol Approach	R AOD vs. Satellite	bias AOD vs. Satellite	R AOD vs. AERONET	bias AOD vs. AERONET	R AE vs. AERONET	RMSE, MAE AE vs. AERONET	Ref
TOMAS in GISS	Global, 2000-2003	2°x2.5°, Sectional: 15 bins from 3 nm-10 μm	0.63 (average of monthly from 2004-2006, MODIS), 0.73 average of monthly from 2004-2006, MISR)	MFE: -29% (average of monthly from 2004-2006, MODIS), -34% (average of monthly from 2004-2006, MISR)	-0.7-0.99 (monthly, 28)	-77-72% (monthly, 28)	N/A	N/A	(Lee et al., 2015)
GOCART with GEOS DAS	CONUS, 2006-2009	1°x1.25°, not specified	N/A	N/A	0.5 (2 hr. average at MISR overpass, 32)	N/A	0.43 (2 hr. average at MISR overpass, 32)	N/A	(Li et al., 2015)

GEMS/MACC aerosol module in CNRM-GAME and CERFACS	Global, 1993-2012	1.4°, Sectional, 12 bins	N/A	Mean relative bias -41-(-52)% (monthly, MISR)	<0-0.9 (monthly, 166)	N/A	N/A	N/A	(Michou et al., 2015)
CNRM-RCSM5	Mediter., Summer 2012	50 km, Sectional, 12 bins	0.64 (seasonal, MODIS), 0.77 (seasonal, MISR), 0.65 (seasonal SEVIRI)	N/A	0.7 (daily, 30)	RMSE~1.75 (daily, 30)	N/A	N/A	(Nabat et al., 2015)
CHIMERE chemical transport model with WRF meteorology	Europe, Mediter. -10°-40°E, 30°-55°N, Summer 2012	50 km, Sectional: 5 bins 40 nm-40 μm	0.35-0.75 (hourly, MODIS)	RMSE: 0.04-0.1 (hourly, MODIS)	0.44-0.73 (hourly, 65)	RMSE: 0.8-0.11 (hourly, 65)	N/A	N/A	(Rea et al., 2015)
MOCAGE	Global, 2007	2°x2°, Sectional: 6 bins per species	0.322 (daily MODIS)	normalized mean bias 0.098 (daily MODIS)	N/A	N/A	N/A	N/A	(Sič et al., 2015)
WRF-Chem	0°-10°E, 50°-55°N; -10°-15°E, 46°-57°N; -15°-30°E,	nested 2 - 30 km, modal	N/A	0.38±0.12 and 0.42±0.10 domain average AOD from MODIS and model respectively	N/A	N/A	N/A	N/A	(Tuccella et al., 2015)

	36°-62°N, 14-30 May 2008								
GOCART in GEOS	Global, 2000-2006	1°x1.25°, dust (8 bins 0.1-10 μm), sea salt (5 bins 0.03-10 μm), carbonaceous/sulfate (modal)	0.747, 0.72 E.US (monthly, MODIS)	N/A	0.707 (monthly, 53)	rms: 0.133 (monthly, 53)	0.81 (monthly, 53)	rms: 0.285 (monthly, 53)	(Colarco et al., 2010)
EMAC	Global, Year 2006	1.1°x1.1°, modal	N/A	Negative (North America)	0.27-0.60 (North America)	RMSE=0.1-0.2	>0.5 (Europe)	N/A	(de Meij et al., 2012)
GEOS-Chem	N. America, 06 July - 14 Aug 2004	2°x2.5°, modal	N/A	N/A	0.87 (study period mean, 24)	N/A	N/A	N/A	(Drury et al., 2010)
WRF-Chem	Europe and N. Africa, Year 2010	23 km, Modal and sectional (4 bins: 0.04-10 μm)	N/A	N/A	0.52 (mod) 0.51 (sect)	NMB=-0.06(mod) NMB=−0.21 (sect) (daily, 12 stations)	N/A	N/A	(Balzarini et al., 2014)
RegCM4	South Asia,	50 km,	N/A	N/A	0.47-0.71	N/A	N/A	N/A	(Nair et al., 2012)

	2005-2007	Sectional (4 bins: 0.01-20 μm)			Monthly, 6				

894

Table 2. Physical and chemical schemes adopted in the WRF-Chem simulations presented herein.

Simulation settings	**Values**
Domain size	300 × 300 cells
Horizontal resolution	12 km
Vertical resolution	32 levels up to 50 hPa
Timestep for physics	72 s
Timestep for chemistry	5 s
Physics option	**Adopted scheme**
Microphysics	WRF Single-Moment 5-class
Longwave Radiation	Rapid Radiative Transfer Model (RRTM)
Shortwave Radiation	Goddard
Surface layer	Monin Obhukov similarity
Land Surface	Noah Land Surface Model
Planetary boundary layer	Mellor-Yamada-Janjich
Cumulus parameterizations	Grell 3
Chemistry option	**Adopted scheme**
Photolysis	Fast J
Gas-phase chemistry	RADM2
Aerosols	MADE/SORGAM
Anthropogenic emissions	NEI (2005)
Biogenic emissions	Guenther, from USGS land use classification

Table 3. Spatial Mean Fractional Bias (MFB) over the entire year. Recall $MFB = \frac{1}{N}\sum_{1}^{N}\frac{C_m - C_0}{\frac{C_m + C_0}{2}}$, where C_m is the monthly mean AOD or AE simulated by WRF-Chem at a specific location and C_0 refers to the same quantity from MODIS/MISR/AERONET. Thus a negative value indicates the model is negatively biased relative to the observations. The total sample size N is 358,048 and 359,633 when comparing WRF-Chem with MODIS onboard Terra and Aqua respectively. The comparison between MODIS and AERONET is affected by a few outlier sites, so in parenthesis is the MFB when the three most biased sites are removed. The mean domain averaged AOD and AE from WRF-Chem (after applying the cloud screen and selecting only MODIS overpass hours) are 0.222 and 1.089, respectively.

Comparisons	MFB AOD	MFB AE
WRF-MODIS (Terra)	0.15	-0.09
WRF-MODIS (Aqua)	0.14	-0.11
WRF-MISR (Terra)	0.16	-0.11
WRF-AERONET	0.50	-0.59
MODIS (Terra)-AERONET	-1.23 (-0.91)	-0.13 (-0.11)

911 Table 4. Contingency table used to compare the fraction of grid cells classified as fine (AE >
912 1) and coarse (AE < 1) by MODIS and WRF-Chem.

		MODIS	
		Fine	Coarse
WRF-Chem	Fine	WF/MF	WF/MC
	Coarse	WC/MF	WC/MC

913

914

Table 5. Contingency table showing the fraction of grid cells simultaneously identified as fine (WF/MF) or coarse (WC/MC) mode particles by WRF-Chem and MODIS, as well as cells with different classification (columns 4 and 5). Recall a threshold of AE = 1 is used to define fine (AE>1) and coarse mode (AE<1) dominance. Months in bold indicate the distribution of observed and simulated fine/coarse mode fractions are significantly different (p-value < 0.01) according to the χ^2 test described in Sect. 2.3.

Month	WF/MF	WC/MC	WF/MC	WC/MF
1	0.025	0.176	0.007	0.792
2	0.030	0.241	0.004	0.725
3	0.005	0.297	0.001	0.697
4	0.013	0.230	0.004	0.753
5	0.141	0.204	0.028	0.628
6	0.541	0.122	0.055	0.283
7	0.623	0.094	0.030	0.252
8	0.520	0.061	0.017	0.402
9	0.561	0.118	0.032	0.288
10	0.486	0.145	0.088	0.281
11	0.321	0.179	0.058	0.442
12	0.164	0.248	0.015	0.573
mean	0.286	0.176	0.028	0.510

Table 6. Synthesis of the skill with which WRF-Chem identifies the spatial distribution and location of extreme AOD values. Cells with extreme AOD are identified as exceeding the 75th percentile computed on a monthly basis across space from monthly averaged daily means. The second column reports the Accuracy, which indicates the spatial coherence of extremes and non-extremes between WRF-Chem and MODIS. The Accuracy metric is computed as the sum of cells co-identified as exceeding the 75th percentile and not exceeding that threshold by WRF-Chem and MODIS (Terra) relative to the total number of cells with valid data (fifth column, *N*). The third column reports the Threat Score (*TS*) which indicates the probability of correctly forecasting extreme AOD conditional upon either forecasting or observing extremes. The fourth column shows the Hit Rate (*HR*) (i.e. probability of correct forecast), which is the proportion of cells correctly identified as extremes by WRF-Chem relative to MODIS extremes. Values in parenthesis refer to the same metrics when comparing WRF-Chem and MODIS onboard the Aqua satellite.

Month	Accuracy	TS	HR	N
Jan	0.664 (0.651)	0.196 (0.178)	0.328 (0.302)	14899 (15051)
Feb	0.654 (0.583)	0.182 (0.091)	0.308 (0.167)	13721 (13643)
Mar	0.656 (0.647)	0.185 (0.173)	0.312 (0.295)	16641 (16541)
Apr	0.645 (0.680)	0.169 (0.219)	0.289 (0.360)	25265 (24974)
May	0.664 (0.699)	0.196 (0.248)	0.327 (0.397)	32770 (31239)
Jun	0.796 (0.800)	0.420 (0.428)	0.592 (0.600)	36148 (34654)
Jul	0.850 (0.823)	0.538 (0.477)	0.700 (0.646)	36055 (35480)
Aug	0.834 (0.832)	0.500 (0.496)	0.667 (0.663)	39173 (39130)
Sep	0.667 (0.665)	0.200 (0.197)	0.333 (0.329)	35883 (35081)
Oct	0.656 (0.665)	0.185 (0.198)	0.311 (0.330)	29662 (26456)
Nov	0.703 (0.696)	0.254 (0.245)	0.405 (0.393)	21630 (19538)
Dec	0.648 (0.653)	0.173 (0.181)	0.295 (0.306)	14914 (14527)
Mean	0.703 (0.699)	0.266 (0.261)	0.406 (0.399)	26397 (25526)

Figures

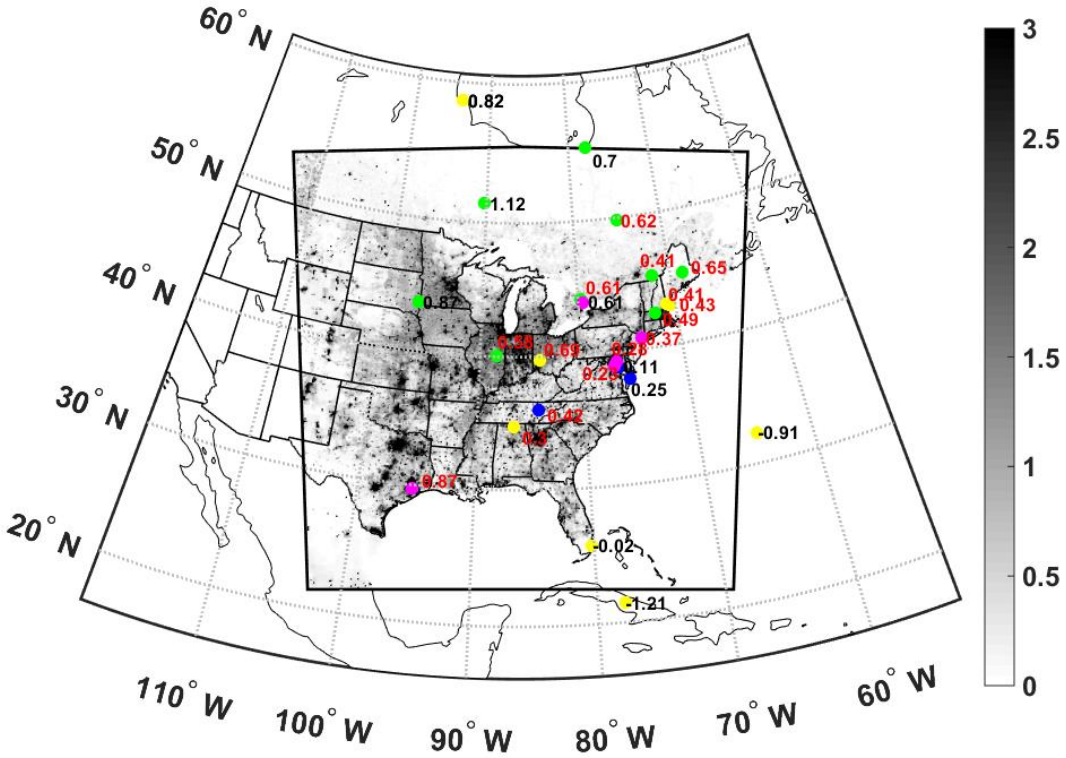


Figure 1. Location of the AERONET stations (colored dots) used in this study and mean daily $PM_{2.5}$ emissions [mg m^{-2} day^{-1}] during 2008 (gray shading). Colors indicate the AERONET site classification based on (Kinne et al., 2013): polluted (magenta), land (green), coastal (blue), unclassified (yellow). The numbers are MFB for WRF-Chem vs. AERONET stations (red numbers indicate WRF-Chem vs. AERONET has a larger MFB than WRF-Chem vs. MODIS whereas black numbers indicate a lower bias in the comparison with AERONET).

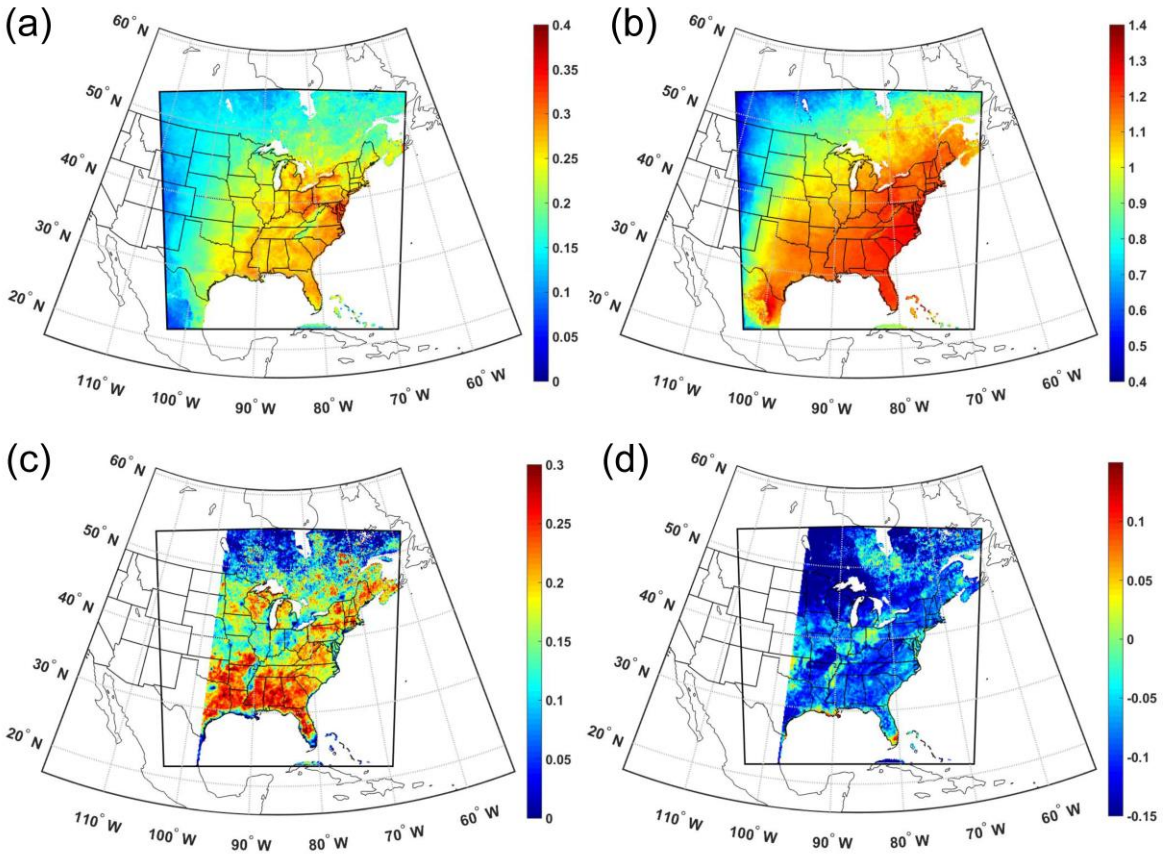


Figure 2. Mean (a) AOD and (b) AE simulated by WRF-Chem during the year 2008. The mean values are computed after applying a cloud mask and are for the Terra overpass time. Mean Fractional Bias (MFB) for (c) AOD and (d) AE for WRF-Chem relative to MODIS (Terra) (similar results are found for Aqua). The inner black frame indicates the entire model domain, while as stated in the text model evaluation is only undertaken for longitudes east of 98°W.

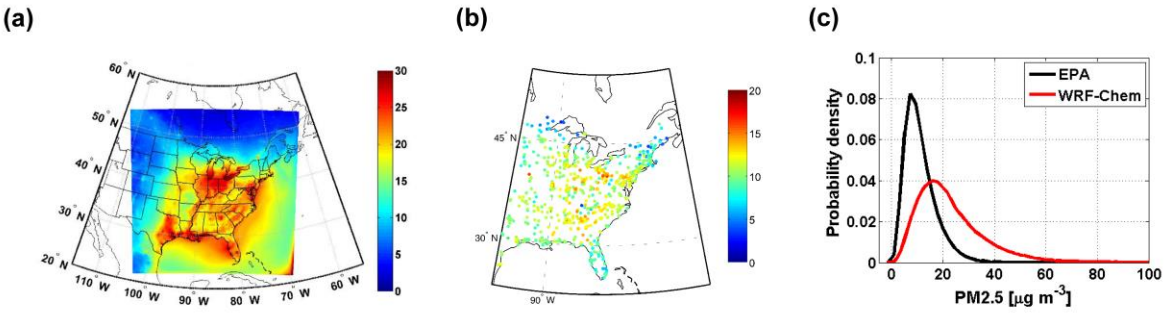
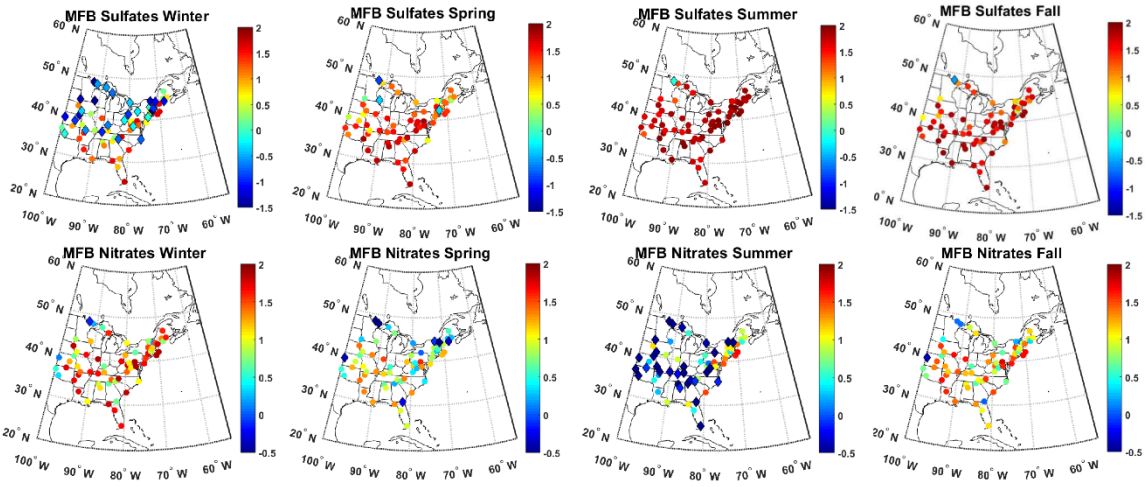


Figure 3. Mean daily $PM_{2.5}$ concentrations [μg m^{-3}] during 2008 as (a) simulated by WRF-Chem in the layer closest to the surface and (b) observed at 1230 EPA sites (note the different colorbar). Panel (c) shows the probability distribution of daily mean $PM_{2.5}$ concentrations observed (black line) and simulated (red line) at the measurement stations.



968

969 Figure 4. Mean fraction bias (MFB) of near-surface daily mean sulfate (first line) and nitrate
970 (second line) concentrations in fine aerosol particles as simulated by WRF-Chem and observed
971 in $PM_{2.5}$ measurements at 123 IMPROVE sites in different seasons. A positive MFB indicates
972 WRF-Chem overestimates the observations. Note the scales differ between the frames shown
973 for sulfate and nitrate MFB and dots/diamonds refer to positive/negative MFB.

974

975

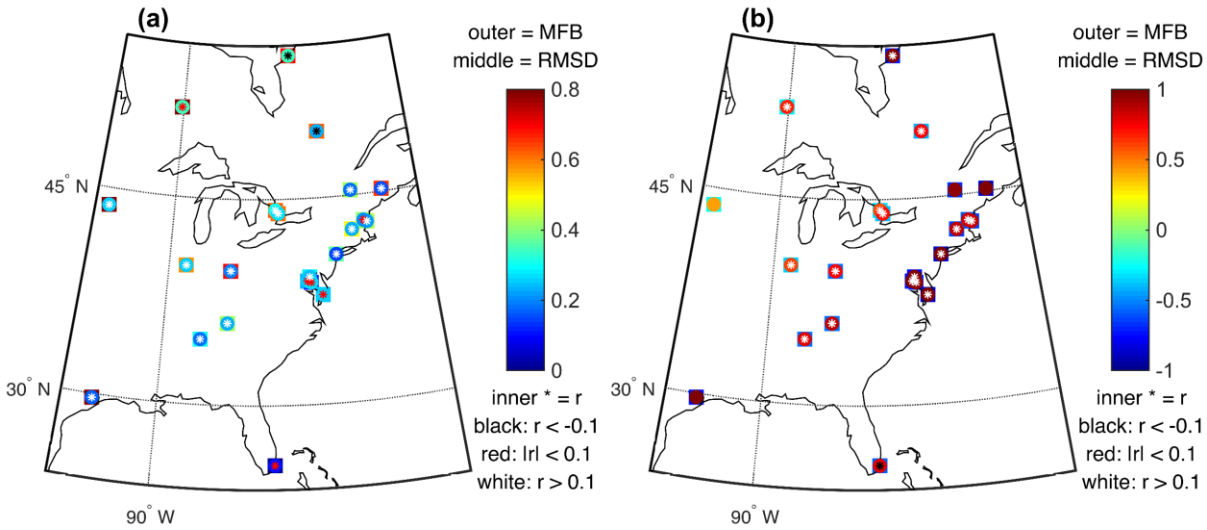


Figure 5. Summary statistics of comparisons of WRF-Chem simulations of (a) AOD and (b) AE relative to simultaneous observations at the AERONET sites. For a location to be included in this analysis at least 20 coincident observations and simulations must be available. The symbols at each AERONET station report MFB (outer square), root mean squared difference (RMSD, inner circle) and correlation coefficient (r, inner *). Note the different colorbar for MFB and RMSD between the two frames. The correlation coefficient is displayed with different colors according with 3 classes: r <-0.1 (black), $|r|$ <0.1 (red) and r>0.1 (white).

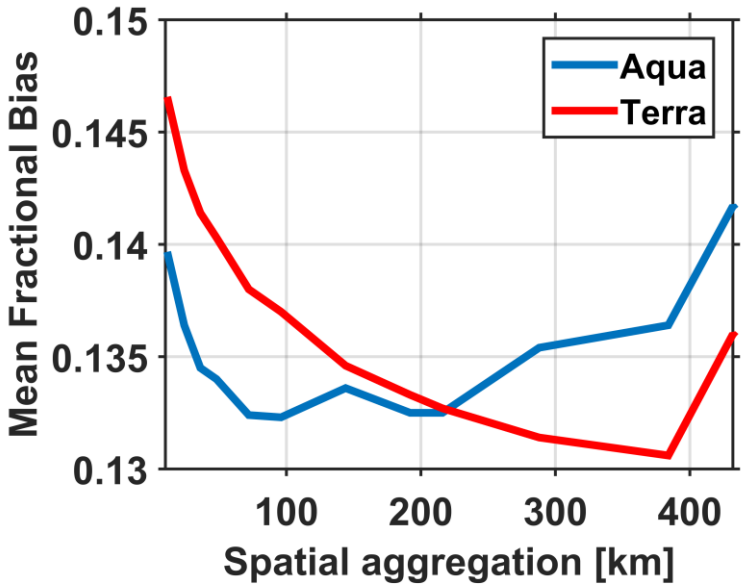


Figure 6. Mean Fractional Bias (MFB) on AOD from WRF-Chem as a function of spatial aggregation relative to observations from Terra (red line) and Aqua (blue line).

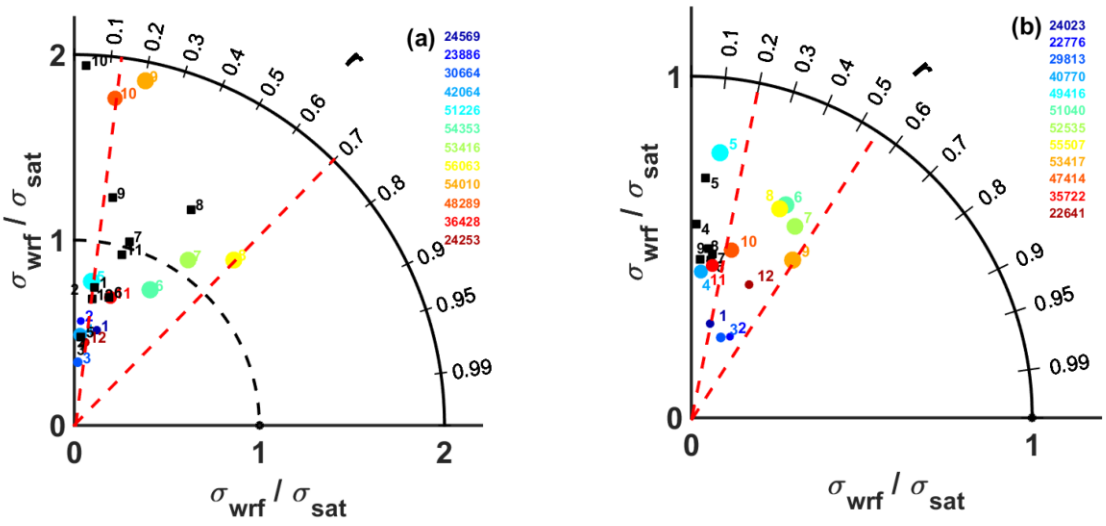


Figure 7. Taylor diagrams comparing the spatial fields of monthly mean (a) AOD and (b) AE from WRF-Chem vs MODIS-Terra (color dots) or MISR (black squares). The numbers shown in the frames denote the month (e.g. 1 = Jan). The numbers shown in the legend indicate the sample size of WRF-Chem data used for computing the monthly mean and the scale of the dots is proportional to the sample size. Note the change in scale for the ratio of standard deviations between the frames. The red dashed lines define the sector with Pearson correlation coefficient between (a) 0.12-0.70 for AOD and (b) 0.20-0.54 for AE which comprise at least two thirds of the months. Each dot/square summarizes the statistics (i.e. RMSD, ratio of standard deviations and correlation coefficient) of the WRF-Chem vs MODIS/MISR comparison for a single month.

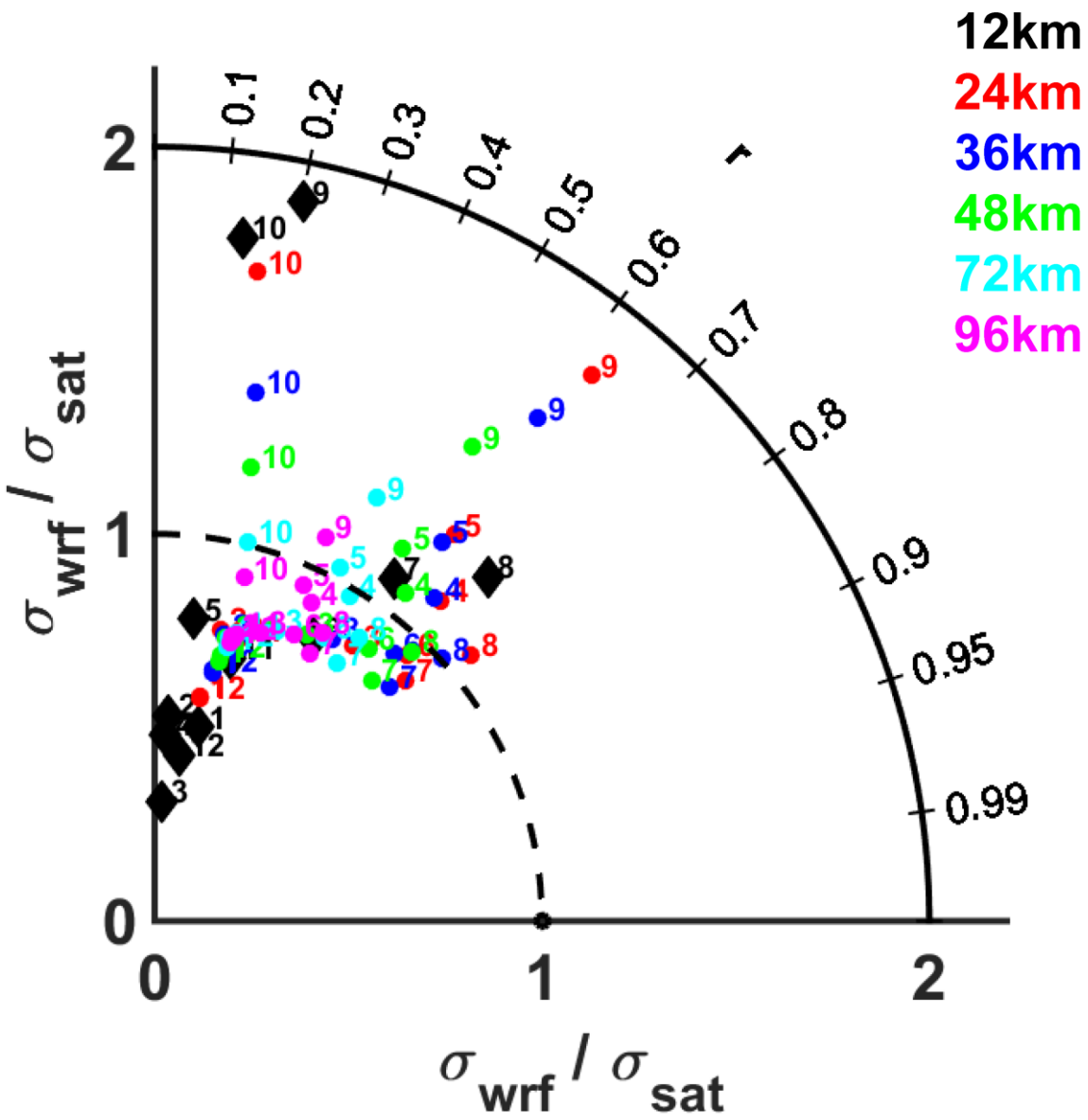
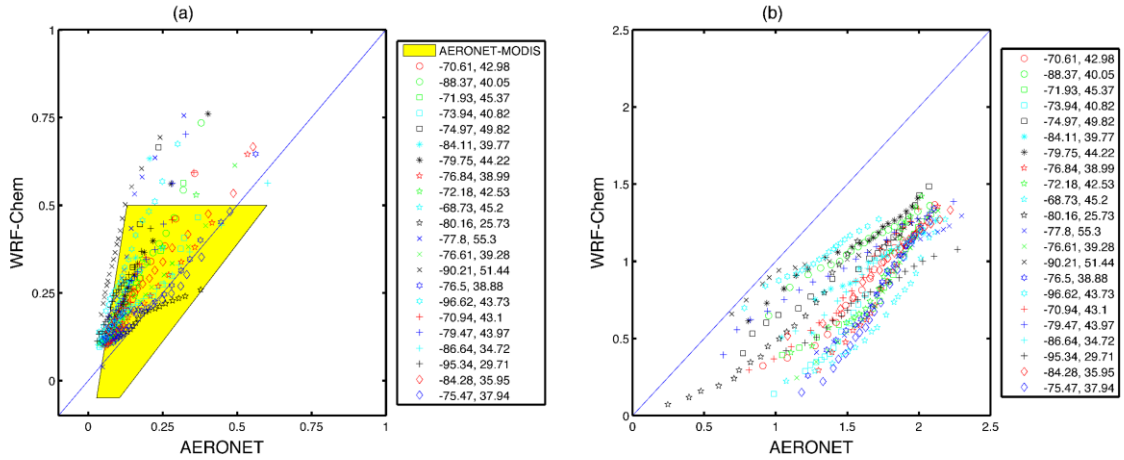


Figure 8. Taylor diagrams for AOD when MODIS observations and WRF-Chem simulations at 12 km are spatially aggregated to 24, 36, 48, 72 and 96 km. Numbers next to the colored dots/diamonds indicate different months (e.g. 1 = Jan).



1022

1023 Figure 9. Empirical quantile-quantile (EQQ) plots of (a) AOD and (b) AE of the 5th to 95th
1024 percentile as simulated by WRF-Chem relative to 22 AERONET stations (their longitude (E)
1025 and latitude (N) is reported in the legend). The yellow shading shows the data envelope for
1026 EQQ plots of AERONET and MODIS. For inclusion in the analysis a location must have at
1027 least 20 coincident observations and simulations in the grid cell containing the AERONET
1028 station. Note MODIS uncertainty in the retrievals (±0.05) in near zero AOD conditions may
1029 lead to negative AOD values which are considered valid. The parameter space for MODIS-
1030 AERONET comparisons of AE are not shown because AE from the MODIS L2 data product
1031 are strongly bimodal (see examples given in Fig. 1 in the Supplementary Materials).

1032

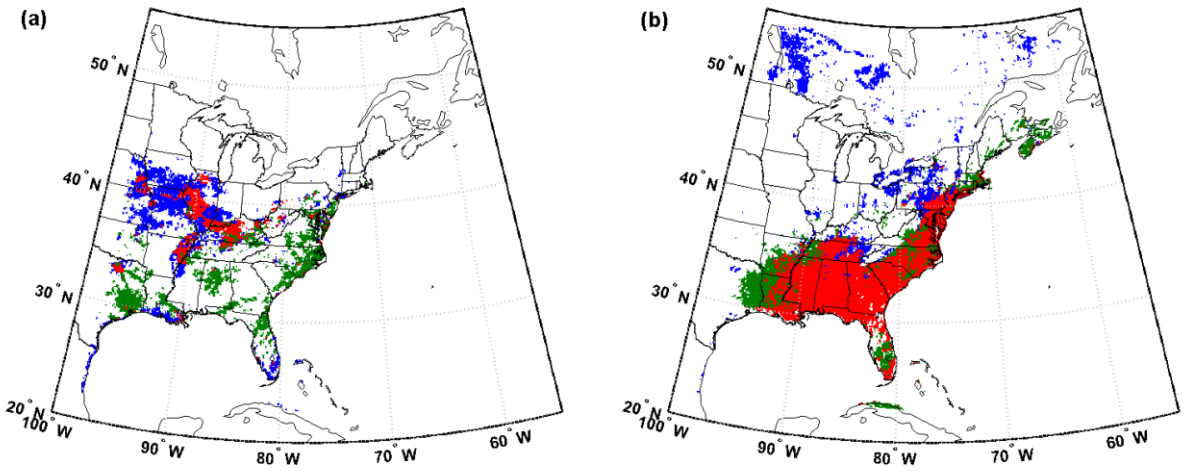


Figure 10. Spatial coherence in extreme AOD (i.e. the occurrence of AOD above the 75th percentile value) from WRF-Chem and MODIS Terra during (a) March (03/2008) and (b) July (07/2008). Green areas denote grid cells defined as experiencing extreme AOD only in the WRF-Chem simulations, blue pixels indicate extreme values as diagnosed using MODIS, while red pixels indicate areas where the occurrence of extreme values is indicated by both the WRF-Chem simulations and the MODIS observations.

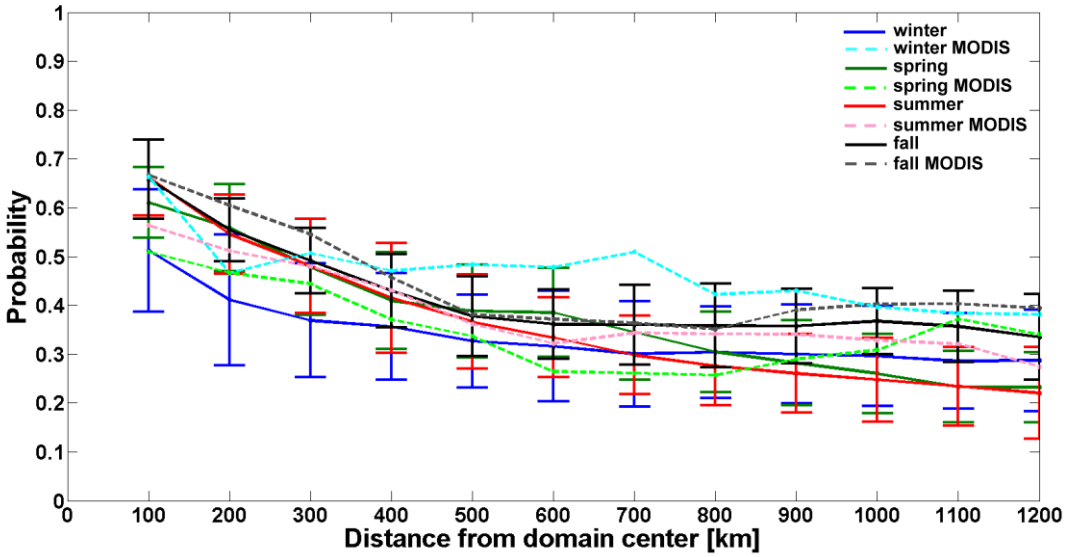


Figure 11. Mean and error bars (± 1 standard deviation from the mean) of the probability of co-occurrence of extreme AOD (i.e. AOD > 75th percentile) at the reference location (i.e. domain center) and any other simulated grid cell during different seasons. The distance between the reference point and each grid cell centroid was binned using 100 km distance classes. Solid lines indicate mean seasonal spatial scales simulated by WRF-Chem, whereas dashed lines are observed means from L2 MODIS data (only the mean of the coherence ratios is plotted for the MODIS data).