Value-added by high-resolution regional simulations of climate-relevant aerosol properties

P. Crippa¹, R. C. Sullivan², A. Thota³, S. C. Pryor²,³

¹COMET, School of Civil Engineering and Geosciences, Cassie Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
²Department of Earth and Atmospheric Sciences, Bradfield Hall, 306 Tower Road, Cornell University, Ithaca, NY 14853, USA
³Pervasive Technology Institute, Indiana University, Bloomington, IN 47405, USA

Correspondence to: P. Crippa (paola.crippa@ncl.ac.uk), School of Civil Engineering and Geosciences, Cassie Building, Room G15, Telephone: +44(0)191 208 5041, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
Despite recent advances in global Earth System Models (ESMs), the current global mean aerosol direct and indirect radiative effects remain uncertain, as does their future role in climate forcing and regional manifestations. Reasons for this uncertainty include the high spatio-temporal variability of aerosol populations. Thus, limited area (regional) models applied at higher resolution over specific regions of interest are generally expected to ‘add value’, i.e. improve the fidelity of the physical-dynamical-chemical processes that induce extreme events and dictate climate forcing, via more realistic representation of spatio-temporal variability. However, added value is not inevitable, and there remains a need to optimize use of numerical resources, and to quantify the impact on simulation fidelity that derives from increased resolution. Here we quantify the value added by enhanced spatial resolution in simulations of the drivers of aerosol direct radiative forcing by applying the Weather Research and Forecasting model with coupled Chemistry (WRF-Chem) over eastern North America at different resolutions. Using Brier Skill Scores and other statistical metrics it is shown that enhanced resolution (from 60 to 12 km) improves model performance for all of the meteorological parameters and gas phase concentrations considered, in addition to both mean and extreme Aerosol Optical Depth (AOD) in three wavelengths in the visible relative to satellite observations, principally via increase of potential skill. Some of the enhanced model performance for AOD appears to be attributable to improved simulation of specific humidity and the resulting impact on aerosol hygroscopic growth/hysteresis.

Keywords: added value, high-resolution WRF-Chem simulations, aerosol optical properties, extreme AOD
1 Motivation and Objectives

Aerosols alter Earth’s radiation balance primarily by scattering or absorbing incoming solar radiation (direct effect, dominated by accumulation mode (diameters ~ wavelength ($\lambda$)), where total extinction is often quantified using AOD), or regulating cloud formation/properties by acting as cloud condensation nuclei (CCN) (indirect effect, dominated by diameters $\geq 100$ nm, magnitude = $f$(composition)). Most aerosols (excluding black carbon) have a larger scattering cross-section than absorption cross-section, and act as CCN thus enhancing cloud albedo and lifetimes. Hence increased aerosol concentrations are generally (but not uniformly) associated with surface cooling (offsetting a fraction of greenhouse gas warming) (Boucher, 2013;Myhre et al., 2013b) to a degree that is principally dictated by the aerosol concentration, size and composition, in addition to the underlying surface and height of the aerosol layer (McComiskey et al., 2008). Despite major advances in measurement and modeling, both the current global mean aerosol direct effect (possible range: -0.77 to +0.23 W m$^{-2}$) and the indirect effect (possible range: -1.33 to -0.06 W m$^{-2}$) remain uncertain, as does their future role in climate forcing (Rockel et al., 2008) and regional manifestations (Myhre et al., 2013a). Specific to our current study region (eastern N. America), one analysis using the NASA GISS global model found that the “regional radiative forcing from US anthropogenic aerosols elicits a strong regional climate response, cooling the central and eastern US by 0.5–1.0 °C on average during 1970–1990, with the strongest effects on maximum daytime temperatures in summer and autumn. Aerosol cooling reflects comparable contributions from direct and indirect radiative effects” (Leibensperger et al., 2012). A recent comparison of multiple global models conducted under the AEROCOM-project indicated this is also a region that exhibits very large model-to-model variability in simulated AOD ($<\text{AOD}> \sim 0.5$, $\sigma$(AOD) $\sim 1$) (Myhre et al., 2013a).

Major reasons why aerosol radiative forcing on both the global and regional scales remains uncertain include short atmospheric residence times and high spatio-temporal variability of aerosol populations, and the complexity of the processes that dictate aerosol concentrations, composition and size distributions (Seinfeld and Pandis, 2016). Although aerosol processes and properties are increasingly being treated in the global Earth System Models (ESMs) (Long et al., 2015;Tilmes et al., 2015) being applied in Coupled Model Intercomparison Project Phase 6 (CMIP-6) (Meehl et al., 2014), the scales on which such models are applied remain much coarser than those on which aerosol population properties are known to vary (Anderson et al., 2003). Therefore, limited area atmospheric models (regional models) applied at higher resolution over specific regions of interest are expected to ‘add value’ (i.e. improve the fidelity)
of the physical-dynamical-chemical processes that induce extreme events and dictate climate forcing. However, debate remains regarding how to objectively evaluate model performance, quantify the value added by enhanced resolution (Di Luca et al., 2015; Rockel et al., 2008) and on possible limits to the improvement of climate representation in light of errors in the driving “imperfect lateral boundary conditions” (Diaconescu and Laprise, 2013). Nevertheless, although “it is unrealistic to expect a vast amount of added values since models already performs rather decently” (Di Luca et al., 2015) and global ESMs are now run at much higher resolution than in the past, it is generally assumed that high resolution regional models will add value via more realistic representation of spatio-temporal variability than global coarser-resolution simulations.

Here we quantify the value added by enhanced resolution in the description of the drivers of aerosol direct radiative forcing using simulations from WRF-Chem over eastern North America. The primary performance evaluation focuses on AOD at the different wavelengths ($\lambda = 470, 550$ and 660 nm, where the AOD at different $\lambda$ is used as a proxy of the aerosol size distribution (Tomasi et al., 1983), see details in Sect. 2.1) and is measured relative to observations from satellite-borne instrumentation. Thus the term “value added” is used here to refer to an improvement of model performance in simulation of wavelength specific AOD as measured by the MODerate resolution Imaging Spectroradiometer (MODIS) instrument aboard the polar-orbiting Terra satellite. We begin by quantifying the performance of WRF-Chem when applied over eastern North America at a resolution of 60 km (WRF60) (~ finest resolution likely to be employed in CMIP-6 global simulations) and then compare the results to those from simulations conducted at 12 km (WRF12) (simulation details are given in Table S1). Quantification of model skill is undertaken by mapping the WRF12 output to the WRF60 grid (WRF12-remap) and computing Brier Skill Scores (BSS) using MODIS as the target, WRF60 as the reference forecast and WRF12-remap as the forecast to be evaluated. We also evaluate the impact of simulation resolution on extreme AOD values that are associated with enhanced impacts on climate and human health. This analysis uses both Accuracy and Hit Rate as the performance metrics and focuses on the co-occurrence of extreme values in space from the model output and MODIS.

Our final analysis focuses on evaluation of the value-added by enhanced resolution in terms of key meteorological and gas-phase drivers of aerosol concentrations and composition and is conducted relative to the MERRA-2 reanalysis product for the physical variables and columnar gas concentrations from satellite observations (see details of the precise data sets used given...
below). The meteorological parameters considered are air temperature at 2 m ($T_{2m}$), total monthly precipitation ($PPT$), planetary boundary-layer height ($PBLH$) and specific humidity in the boundary layer ($Q_{PBL}$). The gas phase concentrations considered are: sulfur dioxide ($SO_2$), ammonia ($NH_3$), nitrogen dioxide ($NO_2$) and formaldehyde ($HCHO$).

2 Materials and Methods

2.1 Spectral dependence of AOD

Three properties dictate the actual aerosol direct radiative forcing: AOD, single scattering albedo and asymmetry factor, all of which are a function of the wavelength ($\lambda$) of incident radiation. The first property is related to the total columnar mass loading, typically dominates the variability of direct aerosol effect (Chin et al., 2009) and is the focus of the current research. The relationship between the aerosol size distribution and spectral dependence of AOD is discussed in detail in (Tomasi et al., 1983) but can be understood by considering a simplified example:

$$\beta(\lambda) = \beta(\lambda = 1 \mu m) \times \lambda^{-\alpha}$$

(1)

where $\beta$ is the particle extinction coefficient, $\lambda$ is the wavelength and $\alpha$ is the Ångström exponent (Ångström, 1964) which describes the wavelength dependence of AOD (and is inversely proportional to the average aerosol Dp):

$$\alpha = -\frac{\ln AOD(\lambda_2)}{\ln AOD(\lambda_1)}$$

(2)

Using Mie theory for spherical particles with radius ($r$): 0.1-1 $\mu$m, if the aerosol size distribution is described by the Junge power law (Eq. 3) then $\alpha \sim \nu - 2$ (i.e. $\alpha \sim 1$):

$$\frac{dN}{d \ln(r)} = K \times r^{-\nu}$$

(3)

where $dN$ is the number of particles of size falling within the radius interval $d\ln(r)$, $K$ is a constant (function of particle total number concentration) and $\nu$ is the Junge parameter ($\nu$ is typically of the order of 2-3 for $r < 10$ $\mu$m and decreases with increasing proportion of coarse aerosols) (Tomasi et al., 1983). Thus, aerosol populations with a higher proportion of coarse mode aerosols will, on average, exhibit higher AOD in the longer wavelengths.
2.2 WRF-Chem simulations

WRF-Chem (version 3.6.1) simulations were performed for the year 2008 over eastern North America, in a domain centered over southern Indiana (86°W, 39°N) at two resolutions, one close to the finest resolution designed for CMIP-6 global model runs (i.e. 60 km, WRF60) and the other one at much higher resolution (12 km, WRF12). Simulation settings are identical for the two runs except for the time-step used for the physics (Table S1), and include use of a modal representation of the aerosol size distribution with three lognormal modes and fixed geometric standard deviations (\(\sigma_{g,Aitken}=1.6\) and \(\sigma_{g,accumulation}=2\), (Ackermann et al., 1998; Grell et al., 2005)) and telescoping vertical grid with 10 model layers from the surface to 800 hPa. Meteorological initial and boundary conditions from the North American Mesoscale Model at 12 km resolution are applied every 6 hours, while initial and chemical boundary conditions are taken from MOZART-4 (Model for Ozone and Related chemical Tracers, version 4) with meteorology from NCEP/NCAR-reanalysis (Emmons et al., 2010). Anthropogenic emissions are specified for both WRF60 and WRF12 from the US National Emission Inventory 2005 (NEI-05) (US-EPA, 2009) which provides hourly point and area emissions at 4 km on 19 vertical levels. The simulation settings and specifically the use of a modal representation of the aerosol size distribution were selected to retain computational tractability. Accordingly, the 60 km simulations for the year 2008 completed in 6.4 hours whereas the 12 km simulations completed 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.

Value added is quantified by degrading (averaging) hourly output from the 12 km resolution simulation to 60 km (hereafter WRF12-remap) as follows: the 12 km domain is resized excluding 2 grid cells at the border to exactly match the 60 km resolution domain. Each coarse grid cell thus includes 5\(\times\)5 12 km resolution cells and its value is the mean of all valid 12 km grid cells inside it if at least half of those cells contain valid AOD (i.e. no cloud cover), otherwise the whole coarse cell is treated as missing. In all comparisons only cells with simultaneous (i.e. model and MODIS) clear sky conditions are considered. A daily value is computed for the satellite overpass time, while a monthly mean is computed using values during the overpass hour and under clear sky conditions if there are at least five valid observations in the month.

2.3 Observations

Model aerosol optical properties are evaluated relative to the MODIS Collection 6 dark-target
land aerosol product from aboard the Terra satellite (~1030 overpass local solar time (LST)) (Levy et al., 2013). Trace gas concentrations are evaluated relative to measurements from the Ozone Monitoring Instrument (OMI; version 3) (Chance, 2002) and the Infrared Atmospheric Sounding Interferometer (IASI; NN version 1) (Whitburn, 2016) aboard the Aura (~1345 LST) and MetOp satellites (~0930 LST), respectively. MODIS retrieves AOD at multiple \( \lambda \) including 470, 550, and 660 nm. The MODIS algorithm removes cloud-contaminated pixels prior to spatial averaging over 10 \( \times \) 10 km (at nadir). OMI and IASI have nadir resolutions of 13 \( \times \) 24 km and 12 km (circular footprint), respectively, and have been filtered to remove retrievals with cloud fractions \( > 0.3 \) (Fioletov et al., 2011; McLinden et al., 2014; Vinken et al., 2014) and OMI pixels affected by the row anomalies. Uncertainty in AOD from MODIS is spatially and temporally variable. It has been estimated as \( \pm 0.05 \pm 0.15 \times \) AOD over land (Levy et al., 2013), and prior research has reported 71% of MODIS Collection 5 retrievals fall within \( \pm 0.05 \pm 0.2 \times \) AOD relative to AERONET in the study domain (Hyer et al., 2011). The accuracy of OMI (“root sum of the square of all errors, including forward model, inverse model, and instrument errors” (Brinksma, 2003)) is 1.1 DU or 50% for \( \text{SO}_2 \), \( 2 \times 10^{14} \text{ cm}^{-2}/30\% \) for background/polluted \( \text{NO}_2 \) conditions, and 35% for HCHO. This uncertainty is typically reduced by spatial and temporal averaging, as described below (Fioletov et al., 2011; Krotkov et al., 2008). IASI \( \text{NH}_3 \) retrievals do not use an a priori assumption of emissions, vertical distribution, or lifetime of \( \text{NH}_3 \) (i.e. no averaging kernel); therefore, \( \text{NH}_3 \) accuracy is variable, and thus only retrievals with uncertainty lower than the retrieved concentrations are used (Whitburn, 2016).

For the model evaluation, MODIS AOD; OMI \( \text{SO}_2 \), \( \text{NO}_2 \), and HCHO; and IASI \( \text{NH}_3 \) for each day are regredded to the WRF-Chem domain by averaging all valid retrievals within 0.1° and 0.35°; 0.125° \( \times \) 0.18° and 0.365° \( \times \) 0.42°; and 0.12° and 0.36° of each WRF-Chem grid cell centroid, for the 12\( \times \)12 km and 60\( \times \)60 km resolutions, respectively. To avoid issues from under-sampling, we require at least 10 valid MODIS granules for the 60\( \times \)60 km daily average to be computed and at least 5 daily averages to compute a monthly average for each grid cell. Model evaluation of gaseous species is performed on standard scores (z-scores) computed relative to the spatial mean of each month, which allow comparing spatial patterns of satellite observations and model output in terms of standard deviation units from the mean.

The simulated meteorological properties are evaluated using Modern-Era Retrospective analysis for Research and Applications (MERRA-2) reanalysis data as the target. MERRA-2
is a homogenized and continuous in time description of atmospheric properties on a 3-dimensional global grid (horizontal resolution of 0.5° × 0.625°, L72), developed by NASA and was released in Fall 2015 (Molod et al., 2015). MERRA-2 provides hourly values of $T_{2m}$ and $PBLH$, and vertical profile of 3-dimensional variables every 3 hours on a large number of pressure levels. Here we compute the total specific humidity ($Q_{PBL}$) of the lowest 8 pressure levels (i.e. in the boundary-layer approximated as the layer from 1000 to 825 hPa) in MERRA-2, assuming an average air density in the PBL of 1.1 kg m$^{-3}$. For the evaluation of simulated precipitation, we use accumulated monthly total values.

2.4 Quantification of model performance and added-value

Taylor diagrams summarize three aspects of model performance relative to a reference: the spatial correlation coefficient (i.e. Pearson correlation of the fields, $r$), the ratio of spatial standard deviations of the two spatial fields ($\sigma_{wrf}/\sigma_{sat}$) and the root mean squared difference (Taylor, 2001). Here Taylor diagrams are presented for monthly mean AOD from WRF60, WRF12 and WRF12-remap relative to MODIS at different wavelengths (Fig. 1). Because AOD is not normally distributed, Spearman’s rank correlation coefficients ($\rho$) of the mean monthly AOD spatial fields are also computed to reduce the impact of a few outliers and the small sample size during cold months (Table 1). To assess the significance of $\rho$ while accounting for multiple testing, we apply a Bonferroni correction (Simes, 1986) in which for $m$ independent hypothesis tests, the null hypothesis is rejected if $p \leq \frac{\alpha}{m}$, where $p$ is the p-value and $\alpha$ is the confidence level (0.05 is used here).

We further quantify the value added (or lack of thereof) of the high-resolution simulations using the following metrics:

(i) Brier Skill Score

The primary metric used to quantify the added value of WRF12-remap versus WRF60 is the Brier Skill Score (BSS) (Murphy and Epstein, 1989):

$$BSS = \frac{r_{p,F}^2 - \left( \frac{\sigma_p}{\sigma_F} \right)^2 - \left( \frac{\langle P \rangle - \langle F \rangle}{\sigma_p} \right)^2 + \left( \frac{\langle P \rangle}{\sigma_p} \right)^2}{1 + \left( \frac{\langle P \rangle}{\sigma_p} \right)^2}$$

(4)

where $F$ is the “forecast” (i.e. the 12 km simulations mapped to 60 km, WRF12-remap); $P$ is
the “target” (i.e. MODIS at 60 km) and output from WRF60 are used as the reference forecast;  
$F'$ the difference between 12 km estimates regridded to 60 km and MODIS; $P'$ the difference  
between the 60 km simulation and MODIS.

BSS measure the improvement in the accuracy with which WRF12-remap reproduces  
observations (from MODIS, MERRA-2 or other satellite products) over output from WRF60.  
A BSS>0 indicates WRF12, even when regridded to 60 km, does add value. The first term in  
(4) ranges from 0 to 1, is described as the potential skill, and is the square of the spatial  
correlation coefficient between forecast and reference anomalies to MODIS. It is the skill score  
achievable if both the conditional and overall bias were zero, and for most of the variables  
considered herein (particularly AOD) it contributes to a positive BSS in most calendar months  
(and seasons). The second term (the conditional bias, > 0), is the square of the difference  
between the anomaly correlation coefficient and the ratio of standard deviation of the  
anomalies. The third term is referred to as the forecast anomaly bias, and is the ratio of the  
difference between the mean anomalies of WRF12-remap and the observations relative to  
WRF60 and the standard deviation of WRF60 anomaly relative to observed values. The fourth  
term is the degree of agreement and appears in both the numerator and denominator. It is  
computed as the square of the ratio of the mean anomaly between WRF60 and observations  
and the standard deviation of the anomalies.

(ii) Pooled paired t-test

To identify which areas in space contribute most to the added value, we compare daily mean  
AOD fields from WRF-Chem at different resolutions and MODIS. We perform a pooled paired  
t-test to evaluate the null hypothesis that those differences come from normal distributions with  
equal means and equal but unknown variances (the test statistic has Student's t distribution with  
df = n + m - 2, and the sample standard deviation is the pooled standard deviation, where n  
and m are the two sample sizes). The test is conducted by climatological season (e.g. winter =  
DJF) since there are fewer than 20 valid AOD observations in most 60 km grid cells for each  
calendar month (Fig. 2). Given the large number of hypothesis tests performed (i.e. one for  
each 60 km grid cell), we adjust the p-values using the False Discovery Rate (FDR) approach  
(Benjamini and Hochberg, 1995). In this approach, p-values from the t-tests are ranked from  
low to high ($p_1,p_2,…,p_m$), then the test with the highest rank, $j$, satisfying:

$$p_j \leq \frac{j}{m} \alpha$$ (5)
is identified. Here all p-values satisfying Eq. 5 with $\alpha=0.1$ are considered significant.

(iii) Accuracy and Hit Rate in identification of extremes

For each month we identify grid cells in which the wavelength specific AOD exceeds the 75th percentile value computed from all grid cells and define that as an extreme. Thus grid cells with extreme AOD are independently determined for MODIS and WRF-Chem at different resolutions. The spatial coherence in identification of extremes in the fields is quantified using two metrics: the Accuracy and the Hit Rate (HR). The Accuracy indicates the overall spatial coherence and is computed as the number of grid cells co-identified as extreme and non-extreme between WRF-Chem and MODIS relative to the total number of cells with valid data. The HR weights only correct identification of extremes in MODIS by WRF-Chem.

3 Results

3.1 Quantifying the value added of increased spatial resolution

When WRF-Chem is applied at 60 km resolution the degree of association of the resulting spatial fields of mean monthly AOD at the three wavelengths with MODIS varies seasonally. Smallest RMSD and highest Spearman spatial correlations ($\rho$) with MODIS observations generally occur during months with highest mean AOD (i.e. during summer, Fig. 1 and 3), and reach a maximum in August ($\rho = 0.60$, Table 1). However, while the patterns of relative AOD variability are well captured, the absolute magnitudes and spatial gradients of AOD during the summer are underestimated by WRF60 (Fig. 1 and Fig. 3, Table S2). High spatial correlations ($\rho > 0.40$) are also observed in March, April and November (Table 1), when the ratio of spatial standard deviations is closer to 1 (Fig. 1, Table S2). Only a weak wavelength dependence is observed in the performance metrics as described on Taylor diagrams. The spatial variability is generally more negatively biased for AOD at 660 nm (Table S2), indicating that WRF60 simulations tend to produce larger diameter aerosols homogeneously distributed over the domain, whereas MODIS observations indicate more spatial variability.

The performance of WRF60 simulations relative to MODIS contrasts with analyses of WRF12 and WRF12-remap. WRF12 and WRF12-remap indicate highest spatial correlations with MODIS observations throughout the summer months ($\rho = 0.5-0.7$, Table 1), although the bias towards simulation of more coarse aerosols than are observed is consistent across the two simulations and with prior research (see details provided in (Crippa et al., 2016)). However, simulations at 12 km (WRF12) show positive $\rho$ with MODIS for all $\lambda$ in all calendar months, while mean monthly spatial fields of AOD from WRF60 show low and/or negative correlations.
with MODIS during May, June, September, October and December, indicating substantial
differences in the degree of correspondence with MODIS AOD in the two simulations, and
higher fidelity of the enhanced resolution runs (Tables 1 and S2).

Monthly mean spatial fields of AOD(λ) as simulated by WRF12 or WRF12-remap exhibit
positive Spearman correlation coefficients (ρ) with MODIS observations for all calendar
months and range from ~ 0.25 for WRF12-remap (0.20 for WRF12) during winter to ~ 0.70
and 0.64, respectively during summer (Table 1). Spearman’s ρ are uniformly higher in WRF12-
remap than WRF12 indicating a mismatch in space in the high-resolution simulation (i.e. that
grid cells with high AOD are slightly displaced in the 12 km simulations possibly due to the
presence of sub-grid scale aerosol plumes (Rissman et al., 2013)). Mean monthly fields of AOD
(all λ) from both WRF12 and WRF12-remap exhibit lower ρ with MODIS in February-April
and November than the 60 km runs (Table 1). These discrepancies appear to be driven by
conditions in the south of the domain. For example, differences between WRF60/WRF12-
remap vs. MODIS during all seasons are significant according to the paired t-test over Florida
and along most of the southern coastlines (Fig. 2). This region of significant differences extends
up to ~ 40°N during summer and fall, reflecting the stronger north-south gradient in AOD from
MODIS and WRF12-remap that is not captured by WRF60 (see example for λ = 550 nm, Fig.
3). These enhancements in the latitudinal gradients from WRF12-remap are also manifest in
the physical variables (particularly specific humidity as discussed further below).

The differences in the absolute values of mean monthly AOD deriving from differences in the
resolution at which WRF-Chem was applied are of sufficient magnitude (a difference of up to
0.2 in regions with a mean AOD value of 0.4), particularly in the summer months (Fig. 4), to
raise concerns. However, detailed investigation of the simulations settings and repetition of the
60 km simulation resulted in virtually identical results indicating no fault can be found in the
analysis. Further, we note this is a region of high discrepancy in global ESM (Myhre et al.,
2013a).

To further investigate differences due to spatial discretization we computed Brier Skill Scores
(BSS, Eq. 4). In this analysis AOD for each λ from WRF12-remap are used as the ‘forecast’,
output from WRF60 are used as the reference forecast and MODIS observations at 60 km are
used as the target. BSS exceed 0 during all months except for September and October, and
largest BSS (> 0.5) for AOD (all λ) is found during most months between December and July
(Fig. 5). This indicates that running WRF-Chem at 12 km resolution adds value relative to
WRF60, even when the WRF12 output is remapped to 60 km. BSS do not strongly depend on \( \lambda \), indicating the added value from enhanced resolution similarly affects particles of different sizes. Inspecting the terms defining the BSS provides information about the origin of the added value (Fig. 5). The positive BSS derives principally from the potential skill (first term in Eq. 4), which demonstrates a reduction in bias and/or more accurate representation of the spatial gradients in WRF12-remap. This term exhibits a weak seasonality with values below 0.5 only during August and fall months. The second and third terms are close to zero during most months, although bigger biases are found during August-October. The substantial conditional bias during late summer and early fall is the result of the large ratio of standard deviations (\( > 1 \)), i.e. the spatial variability of the anomaly relative to MODIS is larger for WRF12-remap than WRF60, Table S2). It thus contributes to the negative BSS found in September and October, which are also identified as outlier months in WRF12-remap from the Taylor diagram analysis (Fig. 1). Output for these months show modest spatial correlations with MODIS and higher ratio of standard deviations than in WRF60-MODIS comparisons (Fig. 1, Table S2).

Model resolution also affects the Accuracy and Hit Rate (HR) for identification of areas of extreme AOD (AOD > 75\(^{th}\) percentile). Highest coherence in the identification of extreme AOD in space identified in WRF12-remap (and WRF12) relative to MODIS is found during May-August (HR = 53-77\%) vs. WRF60 (HR = 17-54\%, Table 2). Conversely highest HR are found for WRF60 and MODIS during winter and early spring, and indeed exceed those for WRF12 and WRF12-remap (Table 2, e.g. Feb: HR = 0.78 for WRF60, and 0.67 and 0.68 for WRF12 and WRF12-remap, respectively). These differences are consistent with the observation that WRF12-remap overestimates the scales of AOD coherence and AOD magnitude during the cold season along coastlines and over much of the domain in April (Fig. 3).

The synthesis of these analyses is thus that the higher resolution simulation increases the overall spatial correlation, decreases overall bias in AOD close to the peak of the solar spectrum relative to MODIS observations and therefore the higher-resolution simulations better represent aerosol direct climate forcing. However, WRF12-remap exhibits little improvement over WRF60 in terms of reproducing the spatial variability of AOD at these wavelengths and further that WRF12-remap tends to be more strongly positively biased in terms of mean monthly AOD outside of the summer months (Fig. 2 and Fig. 3). Also the improvement in detection of areas of extreme AOD in the higher resolution simulations (WRF12-remap) is manifest only during the warm season.
3.2 Investigating the origin of the added value and sources of error in simulated AOD

As documented above, WRF-Chem applied at either 60 or 12 km resolution over eastern North America exhibits some skill in reproducing observed spatial fields of AOD and the occurrence of extreme AOD values. However, marked discrepancies both in space and time are found, and at least some of them show a significant dependence on model resolution. Thus, we investigated a range of physical conditions and gas phase concentrations known to be strongly determinant of aerosol dynamics in terms of the BSS as a function of model resolution and also in terms of the mean monthly spatial patterns.

WRF12 even when remapped to 60 km provides more accurate description of key meteorological variables such as specific humidity ($Q$), PBLH, surface temperature and precipitation (Fig. 6, S1, S2 and S3) when comparing to MERRA-2, as indicated by the positive BSS during almost all months (Fig. 7a). Good qualitative agreement is observed for the spatial patterns and absolute magnitude of $T_{2m}$ in both WRF60 and WRF12-remap relative to MERRA-2 for all seasons (Fig. S1) leading to only modest magnitude of BSS (i.e. value added by the higher resolution simulations (Fig. 7a). The aerosol size distribution and therefore wavelength specific AOD exhibits a strong sensitivity to $Q$ (Santarpia et al., 2005) due to the presence of hygroscopic components in atmospheric aerosols and thus the role of water uptake in determining aerosol diameter, refractivity and extinction coefficient (Zieger et al., 2013). For example, the hygroscopic growth factor, which indicates the change of aerosol diameter due to water uptake, is ~ 1.4 for pure ammonium sulfate with dry diameter of 532 nm at relative humidity of 80%, thus biases in representation atmospheric humidity may lead to big errors in simulated aerosol size and AOD (Flores et al., 2012). Our previous analyses of the 12 km resolution simulations indicated overestimation of sulfate aerosols (a highly hygroscopic aerosol component, and one which in many chemical forms exhibits strong hysteresis (Martin et al., 2004)) relative to observed near-surface PM$_{2.5}$ concentrations during all seasons except for winter (Crippa et al., 2016), leading to the hypothesis that simulated AOD and discrepancies therein may exhibit a strong dependence on $Q$. Consistent with that postulate, $Q_{PBL}$ from WRF12-remap exhibits a wet bias in cloud-free grid cells mostly during warm months, whereas WRF60 is characterized by a dry bias during all seasons (Fig. 6). Despite the positive bias, WRF12-remap better captures the seasonal spatial patterns of $Q_{PBL}$ in MERRA-2, leading to positive BSS in all calendar months. Thus, there is added value by higher-resolution simulations in representation of one of the key parameters dictating particle growth and optical properties. Spatial patterns of differences in $Q_{PBL}$ from WRF60 and WRF12-remap relative to
MERRA-2 (Fig. 6) exhibit similarities to differences in AOD (Fig. 4). WRF60 is dry-biased relative to WRF12 particularly during the summer (and fall) and underestimates $Q_{PBL}$ relative to MERRA-2 during all seasons over the southern states and over most of continental US during summer and fall. Conversely, WRF12-remap overestimates $Q_{PBL}$ over most of continental US during summer and fall relative to MERRA-2.

$PBLH$ is a key variable for dictating near-surface aerosol concentrations but is highly sensitive to the physical schemes applied, and biases appear to be domain and resolution dependent. For example, the Mellor-Yamada-Janjich PBL scheme combined with the Noah Land Surface Model applied in this work was found to produce lower $PBL$ heights (Zhang et al., 2009) than other parameterizations. Thus, the positive bias in simulated AOD and surface PM$_{2.5}$ concentrations (reported previously in (Crippa et al., 2016)) may be linked to the systematic underestimation of $PBLH$ simulated by WRF12-remap over continental US relative to MERRA-2 during all seasons (except winter) with greatest bias over regions of complex topography (Fig. S2). However, a positive bias (of several hundred meters) in terms of $PBLH$ for WRF simulations using the MYJ parameterization was previously reported for high-resolution simulations over complex terrain (Rissman et al., 2013), and a positive bias in $PBLH$ is also observed in the 60 km simulations presented herein (Fig. S2). This may provide a partly explanation for the strong negative bias in AOD in WRF60 during summer (Fig. 3). In general, the BSS indicate improvement in the simulation of $PBLH$ in WRF12-remap than in WRF60 (Fig. 7a). Consistent with the dry bias in $Q_{PBL}$ in WRF60, total accumulated precipitation is also underestimated in WRF60, while WRF12-remap captures the absolute magnitudes and the spatial patterns therein (Fig. S3). Analysis of hourly precipitation rates also showed higher skill of WRF12-remap than WRF60 in correctly simulating precipitation occurrence ($HR$) relative to MERRA-2 (Table S3). More specifically WRF12-remap correctly predicts between 40% and 70% of precipitation events in MERRA-2 with highest skill during winter months, whereas WRF60 output exhibits lower $HR$ (~6% during summer and 30% during winter). This result thus confirms our expectation of a strong sensitivity of model performance to resolution due to the inherent scale dependence in the cumulus scheme.

Gas phase concentrations (transformed into $z$-scores) from WRF12-remap show higher agreement with satellite observations during almost all months, as indicated by the positive BSS (Fig. 7b). However given the limited availability of valid satellite observations (especially during months with low radiation intensity), the BSS are likely only robust for the summer months for all species. Nevertheless, with the exception of NH$_3$ during June, BSS for all months...
are above or close to zero indicating on average, the enhanced resolution simulations do improve the quality of the simulation of the gas phase species even when remapped to 60 km resolution. Further, the seasonal average spatial patterns of the total columnar concentrations, expressed in terms of $z$-scores, also exhibit high qualitative agreement with the satellite observations (Fig. S4-S7).

4 Concluding remarks

This analysis is one of the first to quantify the impact of model spatial resolution on the spatio-temporal variability and magnitude of AOD. Application of WRF-Chem at two different resolutions (60 km and 12 km) over eastern North America for a representative year (2008) leads to the following conclusions:

- Higher resolution simulations add value (i.e. enhance the fidelity of AOD at and near to the peak in the solar spectrum) relative to a coarser run, although the improvement in model performance is not uniform in space and time. Brier Skill Scores for the remapped simulations (i.e. output from simulations conducted at 12 km (WRF12) then averaged to 60 km, WRF12-remap) are positive for ten of twelve calendar months, and for AOD(\(\lambda=550\) nm) exceed 0.5 for seven of twelve months.

- Spatial correlations of output from WRF12 and WRF12-remap with observations from MODIS are higher than output from a simulation conducted at 60 km during most months. For example, in contrast to WRF-Chem simulations at 60 km (WRF60), simulations conducted at 12 km (WRF12) show positive spatial correlations with MODIS for all \(\lambda\) in all calendar months, and particularly during summer (\(\rho = 0.5-0.7\)).

- Output from WRF12 and WRF12-remap exhibit highest accord with MODIS observations in capturing the frequency, magnitude and location of extreme AOD values during summer when AOD is typically highest. During May-August WRF12-remap has Hit Rates for identification of extreme AOD of 53-78%.

- At least some of the improvement in the accuracy with which AOD is reproduced in the higher resolution simulations may be due to improved fidelity of specific humidity and thus more accurate representation of hygroscopic growth of some aerosol components.

- Higher-resolution simulations add value in the representation of key meteorological variables such as temperature, boundary layer height and precipitation. Both spatial patterns and precipitation occurrence are better captured by WRF12-remap.
- More accurate representation of spatial patterns and magnitude of gaseous species playing a key role in particle formation and growth is achieved by running WRF-Chem at high resolution.

It is worthy of note that even the 12 km resolution WRF-Chem simulations exhibit substantial differences in AOD relative to MODIS over eastern North America, and the agreement varies only slightly with wavelength. This may be partially attributable to use of the modal approach to represent the aerosol size distribution in order to enhance computational tractability. In this application each mode has a fixed geometric standard deviation ($\sigma_g$), which can lead to biases in simulated AOD in the visible wavelengths by up to 25% (Brock et al., 2016) (with the model overestimating observations if the prescribed $\sigma_g$ is larger than the observed one). Setting $\sigma_g = 2$ for the accumulation mode (the default in WRF-Chem) may lead to an overestimation of the number of particles at the end of the accumulation mode tail, and there is evidence that a value of $\sigma_{g,acc}=1.40$ leads to higher agreement with observations (Mann et al., 2012). Further possible sources of the AOD biases reported herein derive from selection of the physical schemes (e.g. planetary boundary layer (PBL) schemes and land-surface model (Misenis and Zhang, 2010; Zhang et al., 2009)). Further, it is worth mentioning that NEI emissions are specified based on an average summertime weekday, so higher model performance might be achieved if seasonally varying emissions would be available. Future work will include a systematic sensitivity analysis of these effects.

Acknowledgments

This research was supported in part by a 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), grants to SCP from US NSF (grants # 1102309 and 1517365) and NASA (under NRA NNH14ZDA001N), and a NASA Earth and Space Science Fellowship Program - Grant “14-EARTH14F-0207” (to RCS). Further support was provided by the Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute and the Indiana METACyt Initiative. We gratefully acknowledge the NASA scientists responsible for MERRA-2 and MODIS products, the developers of WRF-Chem, and Lieven Clarisse, Simon Whitburn, and Martin Van Damme for producing and sharing the NH3 retrievals.
References


3-D global chemical transport model, Atmospheric Chemistry and Physics, 12, 4449-4476, 10.5194/acp-12-4449-2012, 2012.


Figure 1. Probability density function of once daily AOD at a wavelength (\( \lambda \)) of 550 nm for (a) MODIS, (b) WRF60 and (c) WRF12 and WRF12-remap during the year 2008. (d-f) Taylor diagrams of mean monthly AOD at wavelengths (\( \lambda \)) of (d) 470, (e) 550 and (f) 660 nm as simulated by WRF-Chem at different resolutions (black diamonds=WRF60 and red dots=WRF12-remap) relative to MODIS observations. The numbers by each symbol denote the calendar month (e.g. 1=January).
Figure 2. First line: Number of paired AOD observations at a wavelength ($\lambda$) of 550 nm (i.e. simultaneous values as output from WRF-Chem and observed by MODIS) used to perform a t-test designed to evaluate whether the difference computed for each grid cell as WRF60-MODIS differs from that computed as WRF12-remap-MODIS on a seasonal basis (columns show Winter (DJF), Spring (MAM), Summer (JJA) and Fall (SON)).

Second line: Results of the t-test. Pixels that have p-values that are significantly different at $\alpha=0.10$ are indicated in red and have been corrected for multiple testing using a False Discovery Rate approach.
Figure 3. Monthly mean AOD at a wavelength (\(\lambda\)) of 550 nm from MODIS (first line) and WRF-Chem at different resolutions (WRF60 and WRF12-remap, second and third line) during a representative month in each climatological season (columns). Note that a different color scale is applied for different months. For a monthly mean value for a grid cell to be shown, there must be at least 5-simultaneous daily values (for the time of the satellite overpass) available.
Figure 4. Difference in monthly mean AOD at a wavelength ($\lambda$) of 550 nm between WRF-Chem simulations conducted at 60 km resolution (WRF60) and output from WRF-Chem simulations conducted with a resolution of 12 km but remapped to 60 km (WRF12-remap). Differences are computed as WRF60 minus WRF12-remap. Similar spatial patterns and magnitudes of differences are found for $\lambda$ of 470 and 660 nm. The calendar months of 2008 are shown in the titles of each panel.
Figure 5. Brier Skill Scores (BSS, black dots) for monthly mean AOD by calendar month (1=January) for AOD at 470, 550 and 660 nm. In this analysis of model skill WRF12 output is mapped to the WRF60 grid (WRF12-remap) and BSS are computed using MODIS as the target, WRF60 as the reference forecast and WRF12-remap as the forecast. Also shown by the color lines are the contributions of different terms to BSS.
Figure 6. Seasonal mean specific humidity \([\text{kg} \text{m}^{-2}]\) integrated from the surface to 825 hPa \((Q_{\text{PBL}})\) from MERRA-2 (first row) assuming an average air density in the PBL of 1.1 kg \text{m}^{-3}, WRF60 (second row), and WRF12-remap (third row). The data are 3-hourly and show only cloud-free hours in all three data sets.
Figure 7. Brier Skill Scores (BSS) for key (a) meteorological and (b) chemical variables. BSS are computed using hourly data of T at 2m ($T_{2m}$) and $PBLH$, 3-hourly estimates of specific humidity in the boundary layer ($Q_{PBL}$), and $z$-scores of monthly total precipitation ($PPT$), and of monthly mean columnar gas phase concentrations.
Table 1. Spearman correlation coefficients (ρ) between AOD at wavelengths (λ) of 470, 550 and 660 nm from MODIS observations averaged over 12 or 60 km and WRF-Chem simulations conducted at 60 km (WRF60, shown in the table as -60), at 12 km (WRF12, shown in the table as -12), and from WRF-Chem simulations at 12 km but remapped to 60 km (WRF12-remap, shown in the table as -remap). Given WRF12-remap is obtained by averaging WRF12 when at least half of the 5×5 12 km resolution cells contain valid data, ρ from WRF60 and WRF12-remap may be computed on slightly different observations and sample size. The bold text denotes correlation coefficients that are significant at α=0.05 after a Bonferroni correction is applied (i.e. \( \rho \leq \frac{0.05}{9 \times 12} = 4.63 \times 10^{-4} \)) is significant). The yellow shading is a visual guide that shows for each month and λ the model output that has highest ρ with MODIS.

<table>
<thead>
<tr>
<th>Month–variable</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>470-12</td>
<td>0.238</td>
<td>0.150</td>
<td>0.137</td>
<td>0.147</td>
<td>0.377</td>
<td>0.581</td>
<td>0.610</td>
<td>0.723</td>
<td>0.352</td>
<td>0.306</td>
<td>0.259</td>
<td>0.212</td>
</tr>
<tr>
<td>470-60</td>
<td>0.156</td>
<td>0.226</td>
<td>0.438</td>
<td>0.412</td>
<td>-0.219</td>
<td>-0.146</td>
<td>0.379</td>
<td>0.601</td>
<td>0.087</td>
<td>-0.051</td>
<td>0.500</td>
<td>-0.059</td>
</tr>
<tr>
<td>470-remap</td>
<td>0.295</td>
<td>0.197</td>
<td>0.250</td>
<td>0.182</td>
<td>0.516</td>
<td>0.637</td>
<td>0.675</td>
<td>0.771</td>
<td>0.368</td>
<td>0.441</td>
<td>0.315</td>
<td>0.274</td>
</tr>
<tr>
<td>550-12</td>
<td>0.223</td>
<td>0.124</td>
<td>0.142</td>
<td>0.146</td>
<td>0.349</td>
<td>0.541</td>
<td>0.580</td>
<td>0.689</td>
<td>0.275</td>
<td>0.301</td>
<td>0.280</td>
<td>0.215</td>
</tr>
<tr>
<td>550-60</td>
<td>0.179</td>
<td>0.244</td>
<td>0.429</td>
<td>0.332</td>
<td>-0.288</td>
<td>-0.188</td>
<td>0.324</td>
<td>0.567</td>
<td>0.073</td>
<td>-0.077</td>
<td>0.491</td>
<td>0.002</td>
</tr>
<tr>
<td>550-remap</td>
<td>0.297</td>
<td>0.164</td>
<td>0.261</td>
<td>0.199</td>
<td>0.493</td>
<td>0.605</td>
<td>0.651</td>
<td>0.747</td>
<td>0.286</td>
<td>0.437</td>
<td>0.352</td>
<td>0.309</td>
</tr>
<tr>
<td>660-12</td>
<td>0.217</td>
<td>0.136</td>
<td>0.165</td>
<td>0.152</td>
<td>0.324</td>
<td>0.476</td>
<td>0.540</td>
<td>0.644</td>
<td>0.183</td>
<td>0.290</td>
<td>0.292</td>
<td>0.221</td>
</tr>
<tr>
<td>660-60</td>
<td>0.191</td>
<td>0.230</td>
<td>0.437</td>
<td>0.402</td>
<td>-0.305</td>
<td>-0.189</td>
<td>0.389</td>
<td>0.616</td>
<td>0.099</td>
<td>-0.137</td>
<td>0.536</td>
<td>0.049</td>
</tr>
<tr>
<td>660-remap</td>
<td>0.356</td>
<td>0.211</td>
<td>0.289</td>
<td>0.208</td>
<td>0.480</td>
<td>0.624</td>
<td>0.669</td>
<td>0.772</td>
<td>0.371</td>
<td>0.432</td>
<td>0.393</td>
<td>0.368</td>
</tr>
</tbody>
</table>
Table 2. Spatial coherence in the identification of extreme AOD values (i.e. areas with AOD>75th percentile over space for each month) between WRF-Chem at different resolutions relative to MODIS. No significant wavelength dependence is found for model skill in identifying extreme AOD so results are only shown for $\lambda = 550$ nm. The different model output is denoted by -60 for simulations at 60 km, -12 for simulations at 12 km resolution, and as –remap for simulations at 12 km but with the output remapped to 60 km. The Accuracy (Acc) indicates the fraction of grid cells co-identified as extremes and non-extremes between WRF-Chem and MODIS relative to the total number of cells with valid data. The Hit Rate (HR) is the probability of correct forecast and is the proportion of cells correctly identified as extremes by both WRF-Chem and MODIS. The yellow shading indicates the model resolution with highest skill in each month for AOD at 550 nm.

<table>
<thead>
<tr>
<th>Month</th>
<th>Metric</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc-12</td>
<td></td>
<td>0.673</td>
<td>0.665</td>
<td>0.659</td>
<td>0.638</td>
<td>0.710</td>
<td>0.800</td>
<td>0.855</td>
<td>0.839</td>
<td>0.666</td>
<td>0.679</td>
<td>0.723</td>
<td>0.661</td>
</tr>
<tr>
<td>Acc-60</td>
<td></td>
<td>0.707</td>
<td>0.778</td>
<td>0.735</td>
<td>0.730</td>
<td>0.600</td>
<td>0.587</td>
<td>0.658</td>
<td>0.769</td>
<td>0.661</td>
<td>0.637</td>
<td>0.729</td>
<td>0.681</td>
</tr>
<tr>
<td>Acc-remap</td>
<td></td>
<td>0.674</td>
<td>0.680</td>
<td>0.694</td>
<td>0.640</td>
<td>0.766</td>
<td>0.824</td>
<td>0.887</td>
<td>0.837</td>
<td>0.667</td>
<td>0.699</td>
<td>0.767</td>
<td>0.641</td>
</tr>
<tr>
<td>HR-12</td>
<td></td>
<td>0.346</td>
<td>0.331</td>
<td>0.319</td>
<td>0.275</td>
<td>0.421</td>
<td>0.599</td>
<td>0.711</td>
<td>0.678</td>
<td>0.333</td>
<td>0.358</td>
<td>0.447</td>
<td>0.323</td>
</tr>
<tr>
<td>HR-60</td>
<td></td>
<td>0.417</td>
<td>0.558</td>
<td>0.471</td>
<td>0.460</td>
<td>0.200</td>
<td>0.173</td>
<td>0.315</td>
<td>0.538</td>
<td>0.321</td>
<td>0.274</td>
<td>0.458</td>
<td>0.364</td>
</tr>
<tr>
<td>HR-remap</td>
<td></td>
<td>0.350</td>
<td>0.361</td>
<td>0.387</td>
<td>0.281</td>
<td>0.532</td>
<td>0.649</td>
<td>0.775</td>
<td>0.674</td>
<td>0.333</td>
<td>0.399</td>
<td>0.535</td>
<td>0.284</td>
</tr>
</tbody>
</table>