Response to review comments on acp-2016-453 from reviewer 1

The original comments are provided in black, our response is given below each comment in red.

Thank you for the careful reading of our manuscript and your review.

The study by Crippa et al. assesses possible improvements in high resolution simulations of aerosol by comparing aerosol optical depth (AOD) and aerosol precursor gases in two otherwise identical WRF-Chem simulations at 12 and 60 km horizontal resolution over eastern North America to MODIS for AOD and OMI/IASI for the precursor gases. The agreement of the simulations to observations in spatial patterns and extreme values are analyzed. This topic is well within the scope of Atmospheric Chemistry and Physics and the relatively long simulation period of one year could give insights whether improvements in high resolution simulations depend on season. Due to the large differences between the 12 km and the 60 km simulation, which are not aerosol related, the very low precipitation rates in the 60 km simulation and a problem with one of the analysis methods publication can only be recommend after major revisions.

Thank you for your positive assessment. We have addressed the general and specific comments below and modified the manuscript accordingly.

General comments:
1) Differences in meteorological variables, in particular relative humidity are identified in the paper as the main source of difference in the AOD simulation between 12 km and 60 km horizontal resolution. As the focus of the study is on improvements in the simulations of the aerosol at high resolution, the differences in meteorological variables would need to be as small as possible. Otherwise the quality of simulating meteorology is analysed rather than aerosol. Assessing AOD and precursor gases in cloud free scenes may prove useful if the differences in meteorological variables can be minimized.

Thanks for the comment. We agree that meteorological variables play a key role in dictating aerosol and gas properties, thus an accurate simulation of those variables naturally will help in reproducing satellite observations of aerosol properties. In response to your comments we expanded the literature review on the added value at line 74 (please refer to our specific answer below). E.g. We are aware of the work of Weigum et al (in review for ACP, and indeed we cite that work in our paper) and think that their attempts to decompose the performance are interesting. Our focus is slightly different - we aim to quantify the value added only by enhanced resolution to the meteorology, gas phase concentrations and the aerosol properties, we are not seeking to evaluate (per se) changes in physical parameterizations. Thus it is essential that we do not change parameterizations between the runs, and we have elected to use the parameterizations that prior research has demonstrated is appropriate at high resolutions (e.g. using a convective parameterization intended for near 'gray zone' resolution simulations) (Grell and Dévényi, 2002;Nasrollahi et al., 2012;Crippa et al., 2016). Our analysis indicates that the improved skill of the high-resolution simulations in reproducing AOD is driven by the skill in reproducing BOTH the meteorological and chemical fields via better representation of fine scale aerosol dynamics.

Naturally, there is a lot more work to be done. We are currently conducting a broader analysis to investigate meteorological, chemical and aerosol properties'

sensitivity to different parameterizations at different resolutions that will complement results presented in this work, but it is beyond the scope of this paper.

2) While the 12 km resolution simulation agrees fairly well with reanalysis data, the 60 km simulation shows large anomalies, in particular precipitation is very low. The annual mean precipitation in the studied region should be around 800 -1200 mm with a standard deviation of 180 - 260 mm (Groisman and Easterling, 1994). The precipitation of the 60 km simulation in Fig. 53 is significantly below these values in many areas. It needs to be checked if this is due to internal variability (e.g. by varying initial conditions), resolution dependent model parameters or whether one of the parameterizations used is not applicable for the resolutions used in the study.

We agree that the 12 km simulations perform better than WRF60 for most of the meteorological, chemical and aerosol components and that a big bias is present in the precipitation fields simulated by WRF60. The choice of the adopted parameterizations is based on our previous work and evaluation (Crippa et al., 2016), which showed good skill of WRF12 in reproducing aerosol optical properties. Therefore the current study aims to verify if the increased resolution (i.e. from 60 km to 12 km) played a role in a more accurate description of simulated properties relative to observations.

The reviewer is quite correct in identifying precipitation bias as a key challenge in regional modelling (both physical and coupled with chemistry). For example, the NARCCAP simulations with WRF at 50-km were also dry biased in the study domain. Although there have been a number of studies that have sought to evaluate different cumulus schemes over different regions at different resolutions, to our knowledge no conclusion (definitive recommendation) has been made regarding the dependence of model's skill on resolution and cumulus parameterization (Arakawa, 2004;Jankov et al., 2005;Nasrollahi et al., 2012). A strong sensitivity on the adopted cumulus scheme was found in (Li et al., 2014), where the Grell 3 scheme is responsible for a wet bias in the Southeast US (mostly in summer). In that study the model was run at 15 km resolution which the authors identified as the minimum resolution to be able to resolve the rainfall system with a 60-km spatial scale typical of the region. Further, the Grell 3D scheme was successfully applied at resolutions in the range of 1-36 km (e.g. (Grell and Dévényi, 2002;Lowrey and Yang, 2008;Nasrollahi et al., 2012;Sun et al., 2014;Zhang et al., 2016)), although further research is needed to identify the optimal cumulus scheme over North America at coarser resolution, which is part of our ongoing work.

Nevertheless, the reviewer's comments have prompted us to include a great deal more discussion of the possible sources of these discrepancies, linking to the adopted schemes and to the potential bias based on other sensitivity studies, and to the number of simulated cloud free grid cells at different resolutions. It would be very interesting to see the sensitivity of the model to varying initial conditions (e.g. using a different reanalysis product for initial conditions), but as the reviewer notes we are one of the first groups to attempt such long (computationally expensive) simulations and are not currently able to rerun the simulations with variable initial conditions.

3) In the computation of the Brier Skill Score (BSS) MODIS is used as the climatological mean and WRF60 as the current observation. This means if for example WRF60 would simulate unrealistic values, the ability of WRF12-remap is tested in this case to reproduce the unrealistic values, which is meaningless. Rather two BSS should be computed for each of the two simulations (WRF60 and WRF12-remap) where MODIS is used as the

current observation and seasonal or annual mean values of MODIS are used for the climatological mean.

An alternative definition for the BSS to the one reported in the manuscript (equation 4) is the following:

$$BSS = 1 - \frac{BS_F}{BS_{ref}}$$

where BS_F and BS_{ref} are the Brier Scores of the forecast (i.e. in our case WRF12-remap vs MODIS) and the reference (WRF60 vs MODIS).

The Brier Score can be computed as:

$$BS = \frac{1}{N} \sum_{i=1}^{N} \left(p_i - o_i \right)^2$$

where N is the sample size, o_i are the observations (i.e. MODIS) and p_i are the simulated values (i.e. either WRF60 or WRF12-remap). The whole derivation from the BSS reported here to the one in the manuscript can be found in (Murphy and Epstein, 1989). Given the BSS is based on the relative comparison of different simulations to the same reference (i.e. MODIS, as stated in the discussion manuscript from line 221), we believe it is an appropriate metric to quantify the improvement of using high versus coarse resolution.

For clarity, we rephrased at line 260 as follows:

"BSS measure how much a test simulation (i.e. WRF12-remap) more closely (or poorly) reproduces observations (from MODIS, MERRA-2 or other satellite products) relative to a control (WRF60) run."

4) The climatological relevance of the results is not shown although the study is motivated by the uncertainty in aerosol forcing. A better accuracy for simulating the regional distribution and extreme values of AOD is important for air quality. If the same is true for effects of aerosol on radiation, clouds or precipitation is not straightforward and it would be a valuable addition if this would be assessed.

The reviewer is quite correct, but we are clear (in the title and elsewhere) that we aim to quantify the value added by high resolution in simulating "climate-relevant aerosol properties" and not the added value in describing climate forcing due to aerosols. Therefore we decided to keep the original title (in response to the specific comment below) and devote further studies to investigate the possible reduction in aerosol climate forcing uncertainty due to the enhanced resolution.

Specific comments:
P4, L73: Other studies that quantify the impact of model resolution on AOD should be discussed here e. g. Qian et al. (2010), Gustafson et al. (2011). In parallel to this study also a paper by Weigum et al. appeared on ACPD for discussion.

Thanks for the useful references. We added the following discussion on them at line 74:

"There is empirical evidence to suggest strong resolution dependence in simulated aerosol particle properties. For example, WRF-Chem simulations with spatial resolution enhanced from 75 km to 3 km exhibited higher correlations and lower bias relative to observations of aerosol optical properties over Mexico likely due to more accurate description of emissions, meteorology and of the physicochemical processes that convert trace gases to particles (Gustafson et al., 2011;Qian et al., 2010). This improvement in the simulation of aerosol optical properties implies, a reduction of the uncertainty in associated aerosol radiative forcing (Gustafson et al., 2011). Further,

WRF-Chem run over the United Kingdom and Northern France at multiple resolutions in the range of 40-160 km, underestimated AOD by 10-16% and overestimated CCN by 18-36% relative to a high resolution run at 10 km, partly as a result of scale dependence of the gas-phase chemistry and differences in the aerosol uptake of water (Weigum et al., 2016).''

P4, L93: Table S1 gives relevant details of the simulations and should be moved into the main text. References for the parameterizations should be added in Table S 1.
Thanks, done.

P5, L 124-L 130: According to Tomasi et al. (1983) alpha is often not proportional to ny-2 in the atmosphere. Furthermore, the Junge power law used in Eq. (3) is mainly interesting for historical reasons (Schuster et al., 2006) and the atmospheric aerosol size distribution is rather described by four log-normal size distributions (modes), where not all modes are present all the time in the atmosphere. But this is not particularly relevant here and the information in this paragraph should rather be that fine mode particles have smaller AOD at shorter wavelengths (e.g. 440 nm) than at longer wavelengths (e.g. 865 nm) whereas for coarse mode particles AOD is similar at shorter and longer wavelengths. This is reflected in the Angstrom parameter and the Angstrom parameter can therefore be used as a proxy for the fine mode fraction or fine mode radius (depending on the definition, see Schuster al. 2006).
Thanks for this comment. We rephrased as follows and added the citation of Schuster et al., 2006.

The relationship between the aerosol size distribution and spectral dependence of AOD is described by a power law function:

$$\beta(\lambda_1) = \beta(\lambda_2) \times \left(\frac{\lambda_1}{\lambda_2}\right)^{-\alpha} \quad \textbf{(1)}$$

where β is the particle extinction coefficient at a specific wavelength λ and α is the Ångström exponent (Ångström, 1964) which describes the wavelength dependence of AOD (and is inversely proportional to the average aerosol diameter):

$$\alpha = \frac{\ln\frac{AOD(\lambda_1)}{AOD(\lambda_2)}}{\frac{\lambda_2}{\lambda_1}} \quad \textbf{(2)}$$

The aerosol volume distribution (and thus also its size distribution) usually conforms to a multi-lognormal function with n modes:

$$\frac{dV(r)}{d\ln r} = \sum_{i=1}^{n} \frac{C_i}{\sqrt{2\pi}\sigma_i} \exp\left[\frac{-(\ln r - \ln R_i)^2}{2\sigma_i^{2}}\right] \quad \textbf{(3)}$$

where C_i is the particle volume concentration in the mode i, R_i is the geometric mean radius and σ_i is the geometric standard deviation, thus we have:

$$AOD(\lambda) = \int \frac{3\beta(m,r,\lambda)}{4r} \frac{dV(r)}{d\ln r} d\ln r dZ \quad \textbf{(4)}$$

As indicated in (Schuster, 2006), "the spectral variability of extinction diminishes for particles larger than the incident wavelength", thus fine mode particles contribute more to AOD in the visible (λ~0.5 μm) than at longer wavelengths, whereas coarse mode particles provide a similar AOD both at short and long wavelengths. This is reflected in the Ångström parameter which can be thus used as a proxy for the fine mode fraction or fine mode radius (Schuster, 2006).

P6, L 144: For which year are the anthropogenic aerosol emissions, 2005, 2008, 2009?
If not 2008, why is 2008 simulated and not the year corresponding to the aerosol emissions?
Anthropogenic emissions are for the year 2005 since they are the closest in time to the year 2008. We are simulating the year 2008 for its climate representativeness, as assessed by other studies based on multiple sources of measurements over the area (e.g. (Crippa et al., 2016)) and for comparison with them.
We added the following comment from line 156:
"Physical and chemical parameterizations were chosen to match previous work using WRF-Chem at 12 km on the same region which showed good performance relative to observations and the year 2008 was selected because representative of average climate and aerosol conditions during 2000 - 2014 (Crippa et al., 2016)."

P6, L 152: Are the cells at the outer boarder of the domain excluded from the analysis?
In some Figures e.g. Fig. 4, Fig. 6, Figs. S1-S3 one can clearly see the effects of the boundary conditions.
Thanks for pointing this out. In the original manuscript, the outer cells of the domain were not excluded from the analyses. However we checked that removing either 3 or 5 cells from each side of the domain (i.e. ~180-300 km), does not significantly affect the BSS results (i.e. if present, changes in BSS occur after the fourth decimal digit). Therefore, we decided to keep the original analysis for a more clear comparison.

P6, L 157-160: This is not clear. Is a single, instantaneous value used at the time of the satellite overpass or are several time steps averaged around the time of the satellite overpass. If the latter: how many time steps, in which time period?
Thanks for pointing this out. We clarified at line 184 that daily values from WRF-Chem are for the hour nearest to the overpass time and that a monthly mean is computed from the daily values at the overpass time as follows:

"A daily value from WRF-Chem is computed as an instantaneous value for the hour nearest to the satellite overpass time. When the comparison is done on a monthly basis, a monthly mean value is computed from the daily values obtained under clear sky conditions, only if there are at least five valid observations in the month."

P7, L 172-175: Given the uncertainty of MODIS observations is there a minimum value for AOD used for the analysis? BSS incorporates the uncertainty in the observations but what about the other methods used?
The minimum value of AOD retrievals is -0.1, which are considered valid for near zero AOD conditions within the retrieval uncertainty; low AOD retrievals are physically representative of low aerosol concentrations (and thus removing them would bias the

analysis), and although low AOD may be degraded due to errors in land surface assumptions, we do not implement additional quality assurance constraints beyond those already implemented in the MODIS Level-2, Collection 6 product in order to increase the number of valid retrievals used in our analyses (Levy et al., 2013).

Random errors in the MODIS retrievals should not greatly impact the analyses, as any errors should decrease 'skill' equally in both WRF60 and WRF12-remap. Similarly, any systematic error in the MODIS product (e.g. due to assumptions about underlying land surface and/or predominant aerosol type (Levy et al., 2007)), should equally impact both WRF60 and WRF12-remap. As we have no a priori expectation that the different resolution simulations would have biases that coincide with that of the MODIS product, and the analysis methods used generally compare relative change in 'skill' between the different resolutions, we do not expect uncertainty in the MODIS product to significantly impact our finding.

PB, L 198: Different definitions are used in the literature for planetary boundary height (PBLH), which can result in large differences in PBLH (e. g. von Engeln and Teixeira, 2013). Are the definitions for PBLH in MERRA-2 and WRF-Chem the same?

MERRA PBLH is diagnosed as the level at which the heat diffusivity drops below a value of 1 m^2 s^{-1} (Jordan et al., 2010). The Mellor-Yamada-Janjich PBL scheme adopted here predicts the turbulent kinetic energy (TKE) at every model level and has a 2.5-order turbulent closure (Janjić, 2002). The PBLH is defined as the lowest model level where the turbulence approaches its prescribed lower bound (i.e. TKE ~ 0.2 m^2s^{-2}). Therefore some differences are present in the way PBLH is computed between MERRA-2 and WRF-Chem which may impact our results (von Engeln and Teixeira, 2013).

We have now rephrased from line 433 as follows:
"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. However, differences in PBL heights between WRF-Chem and MERRA-2 may also originate from the way they are computed (i.e. from heat diffusivity in MERRA-2 (Jordan et al., 2010) and from turbulent kinetic energy in WRF-Chem (Janjić, 2002)) (von Engeln and Teixeira, 2013). 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."

As indicated in previous studies, "over much of the United States and portions of the subtropical oceans, the MERRA PBL depths are within 25% of the estimates derived from CALIPSO…" although "over the arid and semiarid complex terrain of the Southwestern United States and the Rocky Mountain region, the CALIPSO retrievals estimate a relatively shallow PBL depth compared to reanalysis" (McGrath-Spangler and Denning, 2012).

P11, L314: No explanation is given why BSS is so small in September and October (Fig. 5). Also in Fig. 1 d)-f) the standard deviation of September and October of WRF12-remap is much larger than for the other months. What is the reason for this?

We have now added at line 370 the following explanation for the lower model performance in September and October and referred to our previous work in which we analyzed this aspect in more detail:

"Previous work with analogous WRF-Chem settings showed that the lower model skill during September and October can be partially attributable to a dry bias in precipitation from WRF-Chem relative to observations. As a result, a positive bias in simulated AOD and aerosol nitrate and sulfate concentrations is present over large regions of the domain (Crippa et al., 2016)."

P13, L370-375: How does AOD without AOD from aerosol water compare between WRF12-remap and WRF60?

This is a very interesting point. Unfortunately, we did not save aerosol water in the Aitken and accumulation mode in our output variables, but this will be certainly considered for future work.

P13, L377: What is the reason of the dry bias (also over the ocean) in WRF60?

As indicated by the new Figure 2, WRF60 simulates a higher number of cloud free grid cells than MODIS in all seasons and approximately twice the number of cloud free pixels of WRF12-remap, a factor that will be strongly associated with the detected dry bias. Although a dry bias is present in WRF60, we did not change parameterizations between the runs to be able to attribute differences in skills only to the enhanced resolution (please refer to our answers above).

P24, Fig. 3: Why are monthly values shown and not seasonal values as in the other Figures?

-, -: It should be mentioned clearly in the text that the analysis is conducted only over land and discussed why this is done.

Given this work seeks to investigate model's skill in describing MODIS AOD and given the high temporal frequency of the WRF-Chem output, all analyses (i.e. BSS, Taylor diagrams, extremes) are conducted on a monthly basis, thus also Figure 3 and 4 report differences in AOD spatial patterns and magnitude on a monthly basis. The figures on the meteorological and chemical variables in the supplementary materials are reported on a seasonal basis to allow the reader to better understand inter-seasonal changes in the spatial patterns looking at an aggregate information (reporting monthly data for the three datasets would have required a figure of 36 panels for each variable analyzed).

The analysis on AOD is conducted only over land since we are comparing relative to the MODIS Collection 6 dark-target land aerosol product. Retrievals of AOD over land and over ocean invoke different assumptions about surface and aerosol properties, and are thus retrieved with different uncertainty (Levy et al., 2013). Including the ocean product would have thus caused inconsistencies in the model skill assessment. We added the following at line 192:

 "To provide a consistent assessment of model skill, the evaluation of AOD is conducted only on land areas since the MODIS dark-target ocean aerosol product is based on a retrieval algorithm different from the one over land (Levy et al., 2013)."

Technical corrections:

P1, L 1: The relevance for climate of the results is unclear so the title should rather be "Value-added by high-resolution regional simulations of aerosol properties"
We believe our title is appropriate – and thus prefer to keep it as is.

P3, L51-52: References for the forcing estimates are missing.
Thanks for noting this. We have now added the following reference:
Stocker, T. F. a. Q., D. and Plattner, G.-K. and Alexander, L.V. and Allen, S.K. and Bindoff, N.L. and Bréon, F.-M. and Church, J.A. and Cubasch, U. and Emori, S. and Forster, P. and Friedlingstein, P. and Gillett, N. and Gregory, J.M. and Hartmann, D.L. and Jansen, E. and Kirtman, B. and Knutti, R. and Krishna Kumar, K. and Lemke, P. and Marotzke, J. and Masson-Delmotte, V. and Meehl, G.A. and Mokhov, I.I. and Piao, S. and Ramaswamy, V. and Randall, D. and Rhein, M. and Rojas, M. and Sabine, C. and Shindell, D. and Talley, L.D. and Vaughan, D.G. and Xie, S.-P. (2013), Summary for Policymakers, 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, pp. 33–115, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

P4, L76: Diaconescu and Laprise (2013) note that "the main added value of an RCM is provided by its small scales and its skill to simulate extreme events, particularly for precipitation:' As this is relevant for the current study it could be mentioned in the text.
We agree. We added the quote at line 94.
Further, "the main added value of a regional climate model is provided by its small scales and its skill to simulate extreme events, particularly for precipitation" (Diaconescu, 2013).

P5, L 118: Eq. (2) can be derived from Eq. (1) by integration over the atmospheric optical path. It would be clearer if lambda_ 1 and lambda_2 are also used in Eq. (1) instead of lambda and lambda=1 micrometer.
We agree. We modified the equation as follows:

"The relationship between the aerosol size distribution and spectral dependence of AOD is described by a power law function:

$$\beta(\lambda_1) = \beta(\lambda_2) \times \left(\frac{\lambda_1}{\lambda_2}\right)^{-\alpha} \quad (5)$$

where β is the particle extinction coefficient at a specific wavelength λ and α is the Ångström exponent (Ångström, 1964) which describes the wavelength dependence of AOD (and is inversely proportional to the average aerosol diameter)"

P5, L 121: Define Dp.
We removed Dp since it is not present in the new equations (please refer to our answer above).

P6, L 128: Which geometric standard deviation is used for the coarse mode?

The geometric standard deviation for the coarse mode is 2.5. We have now added this information in the manuscript.

P6, L 132-160: The model description should be expanded, in particular the part relevant for the aerosol simulation.

We have now expanded the section describing simulations settings by adding the following from line 154:

"Simulation settings are identical for the two runs except for the time-step used for the physics (Table 1). Physical and chemical parameterizations were chosen to match previous work using WRF-Chem at 12 km on the same region which showed good performance relative to observations and the year 2008 was selected because representative of average climate and aerosol conditions during 2000 - 2014 (Crippa et al., 2016). More specifically the simulations adopted the RADM2 chemical mechanism (Stockwell et al., 1990) and a modal representation of the aerosol size distribution (MADE/SORGAM, (Ackermann et al., 1998;Schell et al., 2001)) with three lognormal modes and fixed geometric standard deviations (i.e. 1.7, 2 and 2.5 for Aitken, accumulation and coarse mode, respectively (Ackermann et al., 1998;Grell et al., 2005)). Aerosol direct feedback was turned on and coupled to the Goddard shortwave scheme (Fast et al., 2006). A telescoping vertical grid with 32 model layers from the surface to 50 hPa and 10 layers up to 800 hPa was selected."

P6, L 139: The total number of layers should be mentioned here as well.
Added.

P7, L162-L183: Give more details about the satellite products used e.g. resolution, coverage etc.

We have explicitly stated the resolution of the satellite products in the discussion paper (lines 168-172), and have added a sentence regarding the temporal coverage of the satellite products. We have also already included details regarding overpass times, measurement uncertainty, and post-processing (e.g. cloud screening). We believe we have provided the information pertinent to our analyses, and as other papers have been dedicated to describing these products in further detail, we refer the readers to the product specific papers (e.g. reference given in section 2.3).

We have amended the text to include the spatial coverage of the satellite products:

"The MODIS algorithm removes cloud-contaminated pixels prior to spatial averaging over 10 × 10 km (at nadir). OMI and IASI have nadir resolutions of 13 × 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. MODIS, OMI, and IASI provide near daily global coverage, although the row anomalies render portions of the OMI viewing swath unusable. Uncertainty in AOD from MODIS is spatially and temporally variable. It has been estimated as ± (0.05 + 15%) for AOD over land (Levy et al., 2013), and prior research has reported 71% of MODIS Collection 5 retrievals fall within 0.05 ± 20% for AOD relative to AERONET in the study domain (Hyer et al., 2011)."

P7, L173-174: Give the right uncertainty values i.e. (+-0.05 +15%) and (+-0.05+15-20%).
Done. See comment above.

P7, L 184-L 187: Reformulate to explain better how the regridding is done.

We rephrased as follows:

"For the model evaluation, satellite observations for each day are regridded to the WRF-Chem domain by averaging all valid retrievals within: 0.1° and 0.35° for MODIS; 0.125° × 0.18° (along-track/latitudinal × cross-track/longitudinal) and 0.365° × 0.42° for OMI; 0.12° and 0.36° for IASI of each WRF-Chem grid cell centroid, for the 12×12 km and *60×60 km* resolutions, respectively."

P7, L 190: Standard scores could be shortly explained.

Done. We rephrased from line 222 as follows:

"Model evaluation of gaseous species is performed on a seasonal basis using standard scores (z-scores), which are computed as the difference between the seasonal mean within a grid cell and the seasonal spatial mean, divided by the seasonal spatial standard deviation. The use of standard scores allows comparing spatial patterns of satellite observations and model output in terms of standard deviation units from the mean."

P8, L206-207: The root mean square difference is not shown in Fig. 1 a)-c).

We agree. We now refer to Fig. 1 d-f.

P9, L225-239: This could be explained better. In Murphy and Epstein it is noted that the first term would be the skill if the second and third term were small. The second term is small if for all points F' is linear to P' (conditional bias). The third term gives the overall/mean bias. The fourth term is a correction and should be small.

We rephrased from line 263 as follows:

"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 bias (second term) and overall bias (third term) 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 and is small if for all points *F′* is linear to *P′*. 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."

P23, Fig. 2: It would be useful to add the number of cloud-free data points for each season and each of the three datasets (WRF12-remap, WRF60, MODIS).

We have modified the figure to include the number cloud free grid cells.
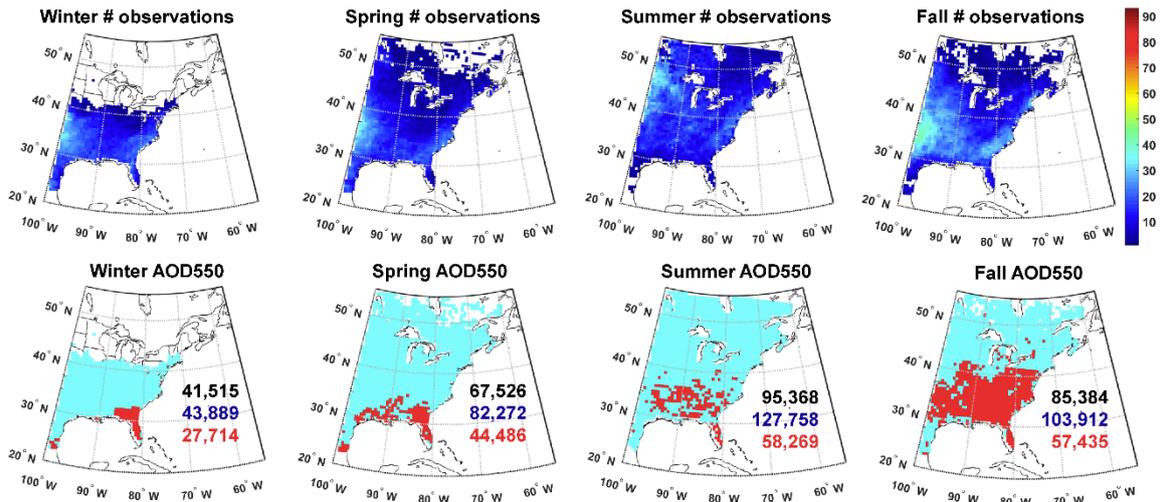
Figure 2. First line: Number of paired AOD observations at a wavelength (λ) of 550 nm (i.e. simultaneous values as output from WRF-Chem and observed by MODIS) used to perform a t-test designed to evaluate whether the difference computed for each grid cell as WRF60-MODIS differs from that computed as 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 α=0.10 are indicated in red and have been corrected for multiple testing using a False Discovery Rate approach. The number of observations of cloud-free conditions summed across all days in each season and all grid cells is also reported (black=MODIS, blue=WRF60, red=WRF12-remap).

References:

von Engeln and Teixeira, 2013, J. Climate, doi:10.1175/JCLl-D-12-00385.1

Groisman and Easterling, 1994, J. Climate, doi: 10.1175/1520-0442(1994)007<0184:VATOTP>2.0.C0;2

Gustafson et al. , 2011, J. Geophys. Res., doi: 10.1029/2010JD015480

Qian et al., 2010, Atmos. Chem. Phys., doi: 10.5194/acp-10-6917-2010

Schuster et al., 2006, J. Geophys. Res., doi: 10.1029/2005JD006328

Weigum et al., 2016, Atmos. Chem. Phys. Discuss., doi: 10.5194/acp-2016-360

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-453, 2016.

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.

Ångström, A.: The parameters of atmospheric turbidity, Tellus, 16, 64-75, 10.1111/j.2153-3490.1964.tb00144.x, 1964.

Arakawa, A.: The Cumulus Parameterization Problem: Past, Present, and Future, Journal of Climate, 17, 2493-2525, doi:10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2, 2004.

Crippa, P., Sullivan, R. C., Thota, A., and Pryor, S. C.: Evaluating the skill of high-resolution WRF-Chem simulations in describing drivers of aerosol direct climate forcing on the regional scale, Atmos. Chem. Phys., 16, 397-416, 10.5194/acp-16-397-2016, 2016.

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.

Fioletov, V. E., McLinden, C. A., Krotkov, N., Moran, M. D., and Yang, K.: Estimation of SO_2 emissions using OMI retrievals, Geophysical Research Letters, 38, L21811, 10.1029/2011GL049402, 2011.

Grell, G. A., and Dévényi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophysical Research Letters, 29, 38-31-38-34, 10.1029/2002GL015311, 2002.

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.

Gustafson, W. I., Qian, Y., and Fast, J. D.: Downscaling aerosols and the impact of neglected subgrid processes on direct aerosol radiative forcing for a representative global climate model grid spacing, Journal of Geophysical Research: Atmospheres, 116, D13303, 10.1029/2010JD015480, 2011.

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.

Janjić, Z. I.: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso model, NCEP office note, 437, 61, 2002.

Jankov, I., A. Gallus, J. W., Segal, M., Shaw, B., and E. Koch, S.: The Impact of Different WRF Model Physical Parameterizations and Their Interactions on Warm Season MCS Rainfall, Weather and Forecasting, 20, 1048-1060, doi:10.1175/WAF888.1, 2005.

Jordan, N. S., Hoff, R. M., and Bacmeister, J. T.: Validation of Goddard Earth Observing System-version 5 MERRA planetary boundary layer heights using CALIPSO, J. Geophys. Res.-Atmos., 115, 10.1029/2009jd013777, 2010.

Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., and Kaufman, Y. J.: Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance, J. Geophys. Res.-Atmos., 112, 21, 10.1029/2006jd007811, 2007.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989-3034, 10.5194/amt-6-2989-2013, 2013.

Li, L. F., Li, W. H., and Jin, J. M.: Improvements in WRF simulation skills of southeastern United States summer rainfall: physical parameterization and horizontal resolution, Clim Dyn, 43, 2077-2091, 10.1007/s00382-013-2031-2, 2014.

Lowrey, M. R. K., and Yang, Z. L.: Assessing the Capability of a Regional-Scale Weather Model to Simulate Extreme Precipitation Patterns and Flooding in Central Texas, Weather and Forecasting, 23, 1102-1126, 10.1175/2008waf2006082.1, 2008.

McGrath-Spangler, E. L., and Denning, A. S.: Estimates of North American summertime planetary boundary layer depths derived from space-borne lidar, J. Geophys. Res.-Atmos., 117, 10.1029/012jd017615, 2012.

McLinden, C. A., Fioletov, V., Boersma, K. F., Kharol, S. K., Krotkov, N., Lamsal, L., Makar, P. A., Martin, R. V., Veefkind, J. P., and Yang, K.: Improved satellite retrievals of NO_2 and SO_2 over the Canadian oil sands and comparisons with surface measurements, Atmos. Chem. Phys., 14, 3637-3656, 10.5194/acp-14-3637-2014, 2014.

Murphy, A. H., and Epstein, E. S.: Skill scores and correlation-coefficients in model verification, Monthly Weather Review, 117, 572-581, 10.1175/1520-0493(1989)117<0572:ssacci>2.0.co;2, 1989.

Nasrollahi, N., AghaKouchak, A., Li, J. L., Gao, X. G., Hsu, K. L., and Sorooshian, S.: Assessing the Impacts of Different WRF Precipitation Physics in Hurricane Simulations, Weather and Forecasting, 27, 1003-1016, 10.1175/waf-d-10-05000.1, 2012.

Qian, Y., Gustafson Jr, W. I., and Fast, J. D.: An investigation of the sub-grid variability of trace gases and aerosols for global climate modeling, Atmos. Chem. Phys., 10, 6917-6946, 10.5194/acp-10-6917-2010, 2010.

Quaas, J., Boucher, O., Bellouin, N., and Kinne, S.: Satellite-based estimate of the direct and indirect aerosol climate forcing, J. Geophys. Res.-Atmos., 113, D05204, 10.1029/2007JD008962, 2008.

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

Schuster, G. L., O. Dubovik, and B. N. Holben Angstrom exponent and bimodal aerosol size distributions, J. Geophys. Res.-Atmos., 111, D07207, doi:10.1029/2005JD006328., 2006.

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

Sun, Y., Yi, L., Zhong, Z., and Ha, Y.: Performance of a New Convective Parameterization Scheme on Model Convergence in Simulations of a Tropical Cyclone at Grey-Zone Resolutions, Journal of the Atmospheric Sciences, 71, 2078-2088, doi:10.1175/JAS-D-13-0285.1, 2014.

Vinken, G. C. M., Boersma, K. F., van Donkelaar, A., and Zhang, L.: Constraints on ship NO_x emissions in Europe using GEOS-Chem and OMI satellite NO_2 observations, Atmos. Chem. Phys., 14, 1353-1369, 10.5194/acp-14-1353-2014, 2014.

von Engeln, A., and Teixeira, J.: A Planetary Boundary Layer Height Climatology Derived from ECMWF Reanalysis Data, Journal of Climate, 26, 6575–6590, doi: 10.1175/JCLI-D-12-00385.1, 2013.

Weigum, N., Schutgens, N., and Stier, P.: Effect of aerosol sub-grid variability on aerosol optical depth and cloud condensation nuclei: Implications for global aerosol modelling, Atmos. Chem. Phys. Discuss., 2016, 1-36, 10.5194/acp-2016-360, 2016.

Zhang, Y., Dubey, M. K., Olsen, S. C., Zheng, J., and Zhang, R.: Comparisons of WRF/Chem simulations in Mexico City with ground-based RAMA measurements during the 2006-MILAGRO, Atmospheric Chemistry and Physics, 9, 3777-3798, doi:10.5194/acp-9-3777-2009, 2009.

Zhang, Y., He, J., Zhu, S., and Gantt, B.: Sensitivity of simulated chemical concentrations and aerosol-meteorology interactions to aerosol treatments and biogenic organic emissions in WRF/Chem, Journal of Geophysical Research: Atmospheres, 121, 6014-6048, 10.1002/2016JD024882, 2016.

Response to review comments on acp-2016-453 from reviewer 2

The original comments are provided in black, our response is given below each comment in red.

Thank you for the careful reading of our manuscript and your review.

Summary
The present work investigates how an increase in spatial (horizontal) resolution from 60 km to 12 km improves the ability of a mesoscale simulation with the WRF-Chem model to reproduce satellite observations of aerosol optical depth (AOD) and column concentrations of chemical species, and of key meteorological variables (relative to a reanalysis product). The WRF-Chem model is run for the year 2008 on a domain that covers the eastern USA, a region with strong emissions of natural and anthropogenic aerosol precursors.
The motivation of the work is established in the introduction. Simulations, observations, and reanalysis data are subsequently introduced with select relevant details, such as uncertainties in the satellite observations. The authors then provide an overview of the statistical methods used to evaluate model performance relative to observations and reanalysis in a succinct but effective way, which can serve as an introduction to novices to the subject. In the main part of the work the results are presented and analyzed, and the authors carefully quantify and discuss the improvements from an increase in resolution. It is found that a higher simulation resolution improves model fidelity in reproducing observed AOD as well as the ability of the model to identify extreme AOD values. The analysis reveals that the improved performance of the model at higher resolution can in part be attributed to improved agreement in meteorological quantities, in particular boundary layer specific humidity, which contributes to aerosol growth. The model is also shown to better reproduce satellite observations of chemical species at higher resolution. The authors do not fail to identify instances where a higher resolution does not result in an improvement: While model skill (measured by the Brier skill score) in reproducing AOD improves in seven out of twelve months, the model shows improvements in detecting extreme AOD values only in the warm season. In all this, statistical methods are used effectively to quantify model performance. The work addresses an interesting question of general importance in atmospheric modeling - what improvements can be expected from an increase in model resolution? The authors answer this question in exemplary fashion for aerosol properties, chemistry, and meteorology in a mesoscale model. The key insight is that increased spatial (horizontal) resolution clearly improves model performance but is not a panacea. Model aspects other than resolution need attention as well to improve model fidelity. The manuscript is thorough, clear, compelling, very well written, and presents the results with good figures and tables. I recommend publication after attending to the following detailed comments.
Thank you for your positive assessment. We have addressed your detailed comments below and modified the manuscript accordingly.

Detailed comments
Line 51-52: Please give references for the possible range of values for the direct and indirect aerosol effects.
Thanks for noting this. We have now added the following reference:
Stocker, T. F. a. Q., D. and Plattner, G.-K. and Alexander, L.V. and Allen, S.K. and Bindoff, N.L. and Bréon, F.-M. and Church, J.A. and Cubasch, U. and Emori, S. and Forster, P. and Friedlingstein, P. and Gillett, N. and Gregory, J.M. and Hartmann, D.L. and Jansen, E. and Kirtman, B. and Knutti, R. and Krishna Kumar, K. and Lemke, P. and Marotzke, J. and Masson-Delmotte, V. and Meehl, G.A. and Mokhov, I.I. and Piao,

S. and Ramaswamy, V. and Randall, D. and Rhein, M. and Rojas, M. and Sabine, C. and Shindell, D. and Talley, L.D. and Vaughan, D.G. and Xie, S.-P. (2013), Summary for Policymakers, 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, pp. 33–115, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Section 2.1: Aerosol size distributions are generally log-normal, not power law functions. The upper tail of a log-normal distribution does, however, behave like a power law distribution. The discussion here is therefore not incorrect, however, it should be modified to assume a log-normal distribution (also because the aerosol scheme used in the WRF-Chem simulations assumes log-normal distributions).

Thanks for this comment. We rephrased the paragraph as follows:

The relationship between the aerosol size distribution and spectral dependence of AOD is described by a power law function:

$$\beta(\lambda) = \beta(\lambda_0) \times \lambda^{-\alpha} \quad \text{(1)}$$

where β is the particle extinction coefficient, λ is the wavelength ($\lambda_0 = 1\mu m$) and α is the Ångström exponent (Ångström, 1964) which describes the wavelength dependence of AOD (and is inversely proportional to the average aerosol diameter Dp):

$$\alpha = \frac{\ln \dfrac{AOD(\lambda_1)}{AOD(\lambda_2)}}{\dfrac{\lambda_2}{\lambda_1}} \quad \text{(2)}$$

The aerosol volume distribution (and thus also its size distribution) can be often described as a multi-lognormal function with *n* modes:

$$\frac{dV(r)}{d\ln r} = \sum_{i=1}^{n} \frac{C_i}{\sqrt{2\pi}\sigma_i} \exp\left[\frac{-(\ln r - \ln R_i)^2}{2\sigma_i^2}\right] \quad \text{(3)}$$

where C_i is the particle volume concentration in the mode *i*, R_i is the geometric mean radius and σ_i is the geometric standard deviation, thus we have:

$$AOD(\lambda) = \int \frac{3\beta(m, r, \lambda)}{4r} \frac{dV(r)}{d\ln r} d\ln r \, dZ \quad \text{(4)}$$

As indicated in (Schuster, 2006), "the spectral variability of extinction diminishes for particles larger than the incident wavelength", thus fine mode particles contribute more to AOD in the visible (λ~0.5 µm) than at longer wavelengths, whereas coarse mode particles provide a similar AOD both at short and long wavelengths. This is reflected in the Ångström parameter which can be thus used as a proxy for the fine mode fraction or fine mode radius (Schuster, 2006).

Line 417-418: "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)."

It is a stretch to write "high qualitative agreement" here. Comparing the OMI, WRF60, and WRF12-remap panels in Fig. S4-S7, my impression is that omitting "high" and leaving "qualitative agreement" is a more accurate assessment.

We agree. We modified the text accordingly and removed "high".

Table 1: A factor of 9 is placed in the denominator of the Bonferroni correction, but the data at the different wavelengths, resolutions, and remappings would not seem to be truly independent significance tests. For example, WRF12 and WRF12-remap would seem to be very dependent. Please carefully consider (and justify) whether the factor of 9 should be used or rather omitted.

We agree that some of the simulations are not really independent as also are the twelve months of the year. The Bonferroni correction can be applied to any set of experiments, either dependent or independent, and aims at providing the most conservative indication of the significance of a statistical test. Therefore we decided to keep the original significance assessment (i.e. with the factor 9) and clarified the explanation of the Bonferroni correction by removing the reference to independence at line 225 as follows:

"To assess the significance of ρ while accounting for multiple testing, we apply a Bonferroni correction (Simes, 1986) in which for *m* hypothesis tests, the null hypothesis is rejected if $p \leq \dfrac{\alpha}{m}$, where *p* is the p-value and α is the confidence level (0.05 is used here)."

Figure S5: OMI, WRF60, and WRF12-remap panels (or panel titles) are shuffled relative to the panel order in Figures S4, S6, and S7.

Thanks for pointing this. We have now reordered the panels to be consistent with the other figures on the gas phase evaluation.

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

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), School of Civil Engineering and
Geosciences, Cassie Building, Room G15, Telephone: +44 (0)191 208 5041, Newcastle
University, Newcastle upon Tyne, NE1 7RU, UK

16   **Abstract**

17   Despite recent advances in global Earth System Models (ESMs), the current global mean

18   aerosol direct and indirect radiative effects remain uncertain, as does their future role in

19   climate forcing and regional manifestations. Reasons for this uncertainty include the high

20   spatio-temporal variability of aerosol populations. Thus, limited area (regional) models

21   applied at higher resolution over specific regions of interest are generally expected to 'add

22   value', i.e. improve the fidelity of the physical-dynamical-chemical processes that induce

23   extreme events and dictate climate forcing, via more realistic representation of spatio-

24   temporal variability. However, added value is not inevitable, and there remains a need to

25   optimize use of numerical resources, and to quantify the impact on simulation fidelity that

26   derives from increased resolution. Here we quantify the value added by enhanced spatial

27   resolution in simulations of the drivers of aerosol direct radiative forcing by applying the

28   Weather Research and Forecasting model with coupled Chemistry (WRF-Chem) over eastern

29   North America at different resolutions. Using Brier Skill Scores and other statistical metrics it

30   is shown that enhanced resolution (from 60 to 12 km) improves model performance for all of

31   the meteorological parameters and gas phase concentrations considered, in addition to both

32   mean and extreme Aerosol Optical Depth (AOD) in three wavelengths in the visible relative

33   to satellite observations, principally via increase of potential skill. Some of the enhanced

34   model performance for AOD appears to be attributable to improved simulation of specific

35   humidity and the resulting impact on aerosol hygroscopic growth/hysteresis.

36

2

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 (λ), 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 (Stocker, 2013), 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 (<AOD> ~ 0.5, σ(AOD) ~ 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)

3

72 applied at higher resolution over specific regions of interest are expected to 'add value' (i.e.

73 improve the fidelity) of the physical-dynamical-chemical processes that induce extreme

74 events and dictate climate forcing. There is empirical evidence to suggest strong resolution

75 dependence in simulated aerosol particle properties. For example, WRF-Chem simulations

76 with spatial resolution enhanced from 75 km to 3 km provideexhibited higher correlations

77 and lower bias relative to observations of aerosol optical properties over Mexico likely due to

78 more accurate description of emissions, meteorology and of the physicochemical processes

79 that convert trace gases to particles (Gustafson et al., 2011;Qian et al., 2010). As a resultThis

80 improvement in the simulation of aerosol optical properties implies, a reduction of the

81 uncertainty in associated aerosol radiative forcing will be also achieved (Gustafson et al.,

82 2011). Further, when WRF-Chem is run over the United Kingdom and Northern France at

83 multiple resolutions in the range of 40-160 km, it underestimateds AOD by 10-16% and

84 overestimateds CCN by 18-36% relative to a high resolution run at 10 km, partly as a result

85 of scale dependence of the gas-phase chemistry and different gas-phase chemistry

86 anddifferences in the aerosol uptake of water (Weigum et al., 2016).

87 However, debate remains regarding how to objectively evaluate model performance, quantify

88 the value added by enhanced resolution (Di Luca et al., 2015;Rockel et al., 2008) and on

89 possible limits to the improvement of climate representation in light of errors in the driving

90 "imperfect lateral boundary conditions" (Diaconescu and Laprise, 2013). Nevertheless,

91 although "it is unrealistic to expect a vast amount of added values since models already

92 performs rather decently" (Di Luca et al., 2015) and global ESMs are now run at much higher

93 resolution than in the past, it is generally assumed that high resolution regional models will

94 add value via more realistic representation of spatio-temporal variability than global coarser-

95 resolution simulations. Further, "the main added value of a regional climate model is

96 provided by its small scales and its skill to simulate extreme events, particularly for

97 precipitation" (Diaconescu and Laprise, 2013). (Qian et al., 2010)(Gustafson et al., 2011)

98 Here we quantify the value added by enhanced resolution in the description of the drivers of

99 aerosol direct radiative forcing using year-long simulations from WRF-Chem over eastern

100 North America. The primary performance evaluation focuses on AOD at the different

101 wavelengths (λ = 470, 550 and 660 nm, where the AOD at different λ is used as a proxy of

102 the aerosol size distribution (Tomasi et al., 1983), see details in Sect. 2.1) and is measured

103 relative to observations from satellite-borne instrumentation. Thus the term "value added" is

104 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 1 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 (λ) 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 described by a power law function: in detail in (Tomasi et al., 1983) but can be understood by considering a simplified example:

5

$$\beta(\lambda_1) = \beta(\lambda_2) \times \frac{\lambda_1}{\lambda_2}^{-\alpha} \quad (1)$$

134

where β is the particle extinction coefficient at a specific wavelength, λ ~~is the wavelength~~

135

and α is the Ångström exponent (Ångström, 1964) which describes the wavelength

136

dependence of AOD (and is inversely proportional to the average aerosol diameter ~~Dp~~):

137

$$\alpha = \frac{\ln \frac{AOD(\lambda_1)}{AOD(\lambda_2)}}{\frac{\lambda_2}{\lambda_1}} \quad (2)$$

138

The aerosol volume distribution (and thus also its size distribution) usually conforms to a

139

multi-lognormal function with n modes:

140

$$\frac{dV(r)}{d\ln r} = \sum_{i=1}^{n} \frac{C_i}{\sqrt{2\pi\sigma_i}} \exp\left[\frac{-(\ln r - \ln R_i)^2}{2\sigma_i^2} \right] \quad (3)$$

141

where C_i is the particle volume concentration in the mode i, R_i is the geometric mean radius

142

and σ_i is the geometric standard deviation, thus we have:

143

144

$$AOD(\lambda) = \int \frac{3\beta(m,r,\lambda)}{4r} \frac{dV(r)}{d\ln r} d\ln r dZ \quad (4)$$

145

As indicated in (Schuster, 2006), "the spectral variability of extinction diminishes for

146

particles larger than the incident wavelength", thus fine mode particles contribute more to

147

AOD in the visible (λ~0.5 µm) than at longer wavelengths, whereas coarse mode particles

148

provide a similar AOD both at short and long wavelengths. This is reflected in the Ångström

149

parameter which can be thus used as a proxy for the fine mode fraction or fine mode radius

150

(Schuster, 2006)~~{Ackermann, 1998 #26}.Using Mie theory for spherical particles with radius~~

151

~~(r): 0.1-1 µm, if the aerosol size distribution is described by the Junge power law (Eq. 3) then~~

152

~~$\alpha \sim v$-2 (i.e. α~1):~~

153

$$\frac{dN}{d\ln(r)} = K \times r^{-v} \quad (3)$$

154

~~where dN is the number of particles of size falling within the radius interval dln(r), K is a~~

155

~~constant (function of particle total number concentration) and v is the Junge parameter (v is~~

156

2.2 WRF-Chem simulations

162 WRF-Chem (version 3.6.1) simulations were performed for the calendar year 2008 over

163 eastern North America, in a domain centered over southern Indiana (86°W, 39°N) at two

164 resolutions, one close to the finest resolution designed for CMIP-6 global model runs (i.e. 60

165 km, WRF60) and the other one at much higher resolution (12 km, WRF12). Simulation

166 settings are identical for the two runs except for the time-step used for the physics (Table

167 1)Table S1),. Physical and chemical parameterizations were chosen to match previous work

168 using WRF-Chem at 12 km on the same region which showed good performance relative to

169 observations and the year 2008 was selected because representative of average climate and

170 aerosol conditions during 2000 - 2014 (Crippa et al., 2016). More specifically the simulations

171 adopted the RADM2 chemical mechanism (Stockwell et al., 1990) and include use of a

172 modal representation of the aerosol size distribution (MADE/SORGAM, (Ackermann et al.,

173 1998;Schell et al., 2001)) with three lognormal modes and fixed geometric standard

174 deviations ($\sigma_{g,Aitken}$=i.e. 1.7, 6 and $\sigma_{g,accumulation}$=2 and, 2.5 for Aitken, accumulation and

175 coarse mode, respectively –(Ackermann et al., 1998;Grell et al., 2005)). Aerosol direct

176 feedback was turned on and coupled to the Goddard shortwave scheme (Fast et al., 2006).

177 Aand telescoping vertical grid with 32 model layers from the surface to 50 hPa and 10 model

178 layers up from the surface to 800 hPa was selected. Meteorological initial and boundary

179 conditions from the North American Mesoscale Model at 12 km resolution are applied every

180 6 hours, while initial and chemical boundary conditions are taken from MOZART-4 (Model

181 for Ozone and Related chemical Tracers, version 4) with meteorology from NCEP/NCAR-

182 reanalysis (Emmons et al., 2010). Anthropogenic emissions are specified for both WRF60

183 and WRF12 from the US National Emission Inventory 2005 (NEI-05) (US-EPA, 2009)

184 which provides hourly point and area emissions at 4 km on 19 vertical levels. The simulation

185 settings and specifically the use of a modal representation of the aerosol size distribution

186 were selected to retain computational tractability. Accordingly, the 60 km simulations for the

187 year 2008 completed in 6.4 hours whereas the 12 km simulations completed in 9.5 days (230

188 hours) on the Cray XE6/XK7 supercomputer (Big Red II) owned by Indiana University,

189 using 256 processors distributed on 8 nodes.

190 Value added is quantified by degrading (averaging) hourly output from the 12 km resolution

191 simulation to 60 km (hereafter WRF12-remap) as follows: the 12 km domain is resized

192 excluding 2 grid cells at the border to exactly match the 60 km resolution domain. Each

193 coarse grid cell thus includes 5×5 12 km resolution cells and its value is the mean of all valid

194 12 km grid cells inside it if at least half of those cells contain valid AOD (i.e. no cloud cover),

195 otherwise the whole coarse cell is treated as missing. In all comparisons only cells with

196 simultaneous (i.e. model and MODIS) clear sky conditions are considered. A daily value

197 from WRF-Chem is computed as an instantaneous value for the hour nearest to the satellite

198 overpass time. When the comparison is done on a monthly basis, a monthly mean value is

199 computed from the daily values obtained~~and~~ under clear sky conditions, only if there are at

200 least five valid observations in the month. ~~A daily value is computed for the satellite overpass~~

201 ~~time, while a monthly mean is computed using values during the overpass hour and under~~

202 ~~clear sky conditions if there are at least five valid observations in the month.~~

203 **2.3 Observations**

204 Model aerosol optical properties are evaluated relative to the MODIS Collection 6 dark-target

205 land aerosol product from aboard the Terra satellite (~1030 overpass local solar time (LST))

206 (Levy et al., 2013). To provide a consistent assessment of model skill, the evaluation of AOD

207 is conducted only on land areas since the MODIS dark-target ocean aerosol product is based

208 on a retrieval algorithm different from the one over land (Levy et al., 2013). Trace gas

209 concentrations are evaluated relative to measurements from the Ozone Monitoring Instrument

210 (OMI; version 3) (Chance, 2002) and the Infrared Atmospheric Sounding Interferometer

211 (IASI; NN version 1) (Whitburn, 2016) aboard the Aura (~1345 LST) and MetOp satellites

212 (~0930 LST), respectively. MODIS retrieves AOD at multiple λ including 470, 550, and 660

213 nm, and t~~. T~~he MODIS algorithm removes cloud-contaminated pixels prior to spatial

214 averaging over 10×10 km (at nadir). OMI and IASI have nadir resolutions of 13×24 km

215 and 12 km (circular footprint), respectively, and have been filtered to remove retrievals with

216 cloud fractions > 0.3 (Fioletov et al., 2011;McLinden et al., 2014;Vinken et al., 2014) and

217 OMI pixels affected by the row anomalies. MODIS, OMI, and IASI provide near daily global

218 coverage, although the row anomalies render portions of the OMI viewing swath unusable.

219 Uncertainty in AOD from MODIS is spatially and temporally variable. It has been estimated

220 as $\pm$ (0.05 + $\pm$ 0.15%) × for AOD over land (Levy et al., 2013), and prior research has

221 reported 71% of MODIS Collection 5 retrievals fall within $\pm$0.05 $\pm$ $\pm$ 0.20% for × AOD

222 relative to AERONET in the study domain (Hyer et al., 2011). The accuracy of OMI ("root

223 sum of the square of all errors, including forward model, inverse model, and instrument
224 errors" (Brinksma, 2003)) is 1.1 DU or 50% for SO_2, 2×10^{14} cm^{-2}/30% for
225 background/polluted NO_2 conditions, and 35% for HCHO. This uncertainty is typically
226 reduced by spatial and temporal averaging, as ~~described below~~employed herein (Fioletov et
227 al., 2011;Krotkov et al., 2008). IASI NH_3 retrievals do not use an a priori assumption of
228 emissions, vertical distribution, or lifetime of NH_3 (i.e. no averaging kernel); therefore, NH_3
229 accuracy is variable, and thus only retrievals with uncertainty lower than the retrieved
230 concentrations are used (Whitburn, 2016).

231 ~~For the model evaluation,~~ For the model evaluation, satellite observations for each day are
232 regridded to the WRF-Chem ~~domain~~discretization. This is done by averaging all valid
233 retrievals within: 0.1° and 0.35° of the WRF-Chem grid-cell center for the 12×12 km and
234 60×60 km resolutions, respectively for MODIS; 0.125° × 0.18° (along-track/latitudinal ×
235 cross-track/longitudinal) and 0.365° × 0.42° for OMI; 0.12° and 0.36° for IASI ~~of each WRF-~~
236 ~~Chem grid cell centroid, for the 12×12 km and 60×60 km resolutions, respectively.~~MODIS
237 ~~AOD; OMI SO$_2$, NO$_2$, and HCHO; and IASI NH$_3$ for each day are regridded to the WRF-~~
238 ~~Chem domain by averaging all valid retrievals within 0.1° and 0.35°; 0.125° × 0.18° and~~
239 ~~0.365° × 0.42°; and 0.12° and 0.36° of each WRF-Chem grid cell centroid, for the 12×12 km~~
240 ~~and 60×60 km resolutions, respectively.~~ To avoid issues from under-sampling, we require at
241 least 10 valid MODIS granules for the 60×60 km daily average to be computed and at least 5
242 daily averages to compute a monthly average for each grid cell. Model evaluation of gaseous
243 species is performed on a seasonal basis using standard scores (z-scores), which are
244 computed~~computed~~ as the difference between ~~relative to~~the seasonal mean within a grid cell
245 and the seasonal spatial mean ~~of each month~~, divided by the seasonal spatial standard
246 deviation. The use of standard scores, ~~which a~~ allows ~~comparing~~ comparison of the spatial
247 patterns of satellite observations and model output in terms of standard deviation units from
248 the mean.

249 The simulated meteorological properties are evaluated using Modern-Era Retrospective
250 analysis for Research and Applications (MERRA-2) reanalysis data as the target. MERRA-2
251 is a homogenized and continuous in time description of atmospheric properties on a 3-
252 dimensional global grid (horizontal resolution of 0.5°×0.625°, L72), developed by NASA and
253 was released in Fall 2015 (Molod et al., 2015). MERRA-2 provides hourly values of T_{2m} and
254 $PBLH$, and vertical profile of 3-dimensional variables every 3 hours on a large number of

255   pressure levels. Here we compute the total specific humidity (Q_{PBL}) of the lowest 8 pressure

256   levels (i.e. in the boundary-layer approximated as the layer from 1000 to 825 hPa) in

257   MERRA-2, assuming an average air density in the PBL of 1.1 kg m^{-3}. For the evaluation of

258   simulated precipitation, we use accumulated monthly total values.

2.4 Quantification of model performance and added-value

260   Taylor diagrams summarize three aspects of model performance relative to a reference: the

261   spatial correlation coefficient (i.e. Pearson correlation of the fields, r), the ratio of spatial

262   standard deviations of the two spatial fields ($\sigma_{wrf}/\sigma_{sat}$) and the root mean squared difference

263   (Taylor, 2001). Here Taylor diagrams are presented for monthly mean AOD from WRF60,

264   WRF12 and WRF12-remap relative to MODIS at different wavelengths (Fig. 1 d-f). Because

265   AOD is not normally distributed, Spearman's rank correlation coefficients (ρ) of the mean

266   monthly AOD spatial fields are also computed to reduce the impact of a few outliers and the

267   small sample size during cold months (Table 2~~Table 2Table 1~~). To assess the significance of

268   ρ while accounting for multiple testing, we apply a Bonferroni correction (Simes, 1986) in

269   which for m ~~independent~~ hypothesis tests, the null hypothesis is rejected if $p \le \dfrac{\alpha}{m}$, where p

270   is the p-value and α is the confidence level (0.05 is used here).

271   We further quantify the value added (or lack of thereof) of the high-resolution simulations

272   using the following metrics:

(i) Brier Skill Score

274   The primary metric used to quantify the added value of WRF12-remap versus WRF60 is the

275   Brier Skill Score (BSS) (Murphy and Epstein, 1989):

276
$$BSS = \frac{r_{F'P'}^2 - \left(r_{F'P'} - \dfrac{\sigma_{F'}}{\sigma_{P'}}\right)^2 - \left(\dfrac{\langle P'\rangle - \langle F'\rangle}{\sigma_{P'}}\right)^2 + \left(\dfrac{\langle P'\rangle}{\sigma_{P'}}\right)^2}{1 + \left(\dfrac{\langle P'\rangle}{\sigma_{P'}}\right)^2} \quad \textbf{(5)}$$

277   where F is the "forecast" (i.e. the 12 km simulations mapped to 60 km, WRF12-remap); P is

278   the "target" (i.e. MODIS at 60 km) and output from WRF60 are used as the reference

279   forecast; F' the difference between 12 km estimates regridded to 60 km and MODIS; P' the

280   difference between the 60 km simulation and MODIS.

281   BSS measure ~~the improvement in the accuracy with which~~how much a test simulation (i.e.
282   WRF12-remap) more closely (or poorly) reproduces observations (from MODIS, MERRA-2
283   or other satellite products) relative to a control (WRF60) run~~over output from WRF60~~. A
284   BSS>0 indicates WRF12, even when regridded to 60 km, does add value. The first term in
285   (~~4~~5) ranges from 0 to 1, is described as the potential skill, and is the square of the spatial
286   correlation coefficient between forecast and reference anomalies to MODIS. It is the skill
287   score achievable if both the conditional bias (second term) and overall bias (third term) were
288   zero, and for most of the variables considered herein (particularly AOD) it contributes to a
289   positive BSS in most calendar months (and seasons). The second term (the conditional bias, >
290   0), is the square of the difference between the anomaly correlation coefficient and the ratio of
291   standard deviation of the anomalies and is small if for all points F' is linear to P'.~~.~~ The third
292   term is referred to as the forecast anomaly bias, and is the ratio of the difference between the
293   mean anomalies of WRF12-remap and the observations relative to WRF60 and the standard
294   deviation of WRF60 anomaly relative to observed values. The fourth term is the degree of
295   agreement and appears in both the numerator and denominator. It is computed as the square
296   of the ratio of the mean anomaly between WRF60 and observations and the standard
297   deviation of the anomalies.

298   **(ii) Pooled paired t-test**

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

311   $$p_j \le \frac{j}{m}\alpha \quad \textbf{(6)}$$

312   is identified. Here all p-values satisfying Eq. ~~5~~6 with α=0.1 are considered significant.

11

(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 (ρ) with MODIS observations generally occur during months with highest mean AOD (i.e. during summer, Fig. 1 d-f and Fig. 3), and reach a maximum in August ($\rho = 0.60$,~ Table 2Table 2Table 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 d-f and Fig. 3, Table S21). High spatial correlations ($\rho > 0.40$) are also observed in March, April and November (Table 2Table 2Table 1), when the ratio of spatial standard deviations is closer to 1 (Fig. 1 d-f, Table S2S1). 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 S2S1), 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 2Table 2Table 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 ρ with MODIS for all λ 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~~2 and S~~2~~1).

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 2~~Table 2Table 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 2~~Table 2Table 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~~that the eastern-half of North America was also identified as a region of high discrepancy in global ESM (Myhre et al., 2013a).

To further investigate differences in the simulation output 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

377   for September and October, and largest BSS (> 0.5) for AOD (all λ) is found during most

378   months between December and July (Fig. 5). This indicates that running WRF-Chem at 12

379   km resolution adds value relative to WRF60, even when the WRF12 output is remapped to 60

380   km. BSS do not strongly depend on λ, indicating the added value from enhanced resolution

381   similarly affects particles of different sizes. Inspecting the terms defining the BSS provides

382   information about the origin of the added value (Fig. 5). The positive BSS derives principally

383   from the potential skill (first term in Eq. ~~4~~5), which demonstrates a reduction in bias and/or

384   more accurate representation of the spatial gradients in WRF12-remap. This term exhibits ~~a~~

385   weak seasonality with values below 0.5 only during August and fall months. The second and

386   third terms are close to zero during most months, although bigger biases are found during

387   August-October. The substantial conditional bias during late summer and early fall is the

388   result of the large ratio of standard deviations (> 1, i.e. the spatial variability of the anomaly

389   relative to MODIS is larger for WRF12-remap than WRF60, Table ~~S2~~S1). It thus contributes

390   to the negative BSS found in September and October, which are also identified as outlier

391   months in WRF12-remap from the Taylor diagram analysis (Fig. 1). Output for these months

392   show modest spatial correlations with MODIS and higher ratio of standard deviations than in

393   WRF60-MODIS comparisons (Fig. 1, Table ~~S2~~S1). Previous work ~~with analogous WRF-~~

394   ~~Chem settings~~ showed that the lower model skill (in WRF12) during September and October

395   ~~can~~may be partially attributable to a dry bias in precipitation from WRF-Chem relative to

396   observations. As a result, ~~a positive bias in~~ simulated AOD and near-surface aerosol nitrate

397   and sulfate concentrations are positively biased ~~is present~~ over large ~~region~~parts of the

398   domain (Crippa et al., 2016).

399   Model resolution also affects the *Accuracy* and *Hit Rate* (*HR*) for identification of areas of

400   extreme AOD (AOD>75th percentile). Highest coherence in the identification of extreme

401   AOD in space identified in WRF12-remap (and WRF12) relative to MODIS is found during

402   May-August (*HR* = 53-77%) vs. WRF60 (*HR* = 17-54%, Table 3~~Table 2~~). Conversely highest

403   *HR* are found for WRF60 and MODIS during winter and early spring, and indeed exceed

404   those for WRF12 and WRF12-remap (Table 3~~Table 2~~, e.g. Feb: *HR* = 0.78 for WRF60, and

405   0.67 and 0.68 for WRF12 and WRF12-remap, respectively). These differences are consistent

406   with the observation that WRF12-remap overestimates the scales of AOD coherence and

407   AOD magnitude during the cold season along coastlines and over much of the domain in

408   April (Fig. 3).

409   The synthesis of these analyses is thus that the higher resolution simulation increases the

410   overall spatial correlation, decreases overall bias in AOD close to the peak of the solar

411   spectrum relative to MODIS observations and therefore the higher-resolution simulations

412   better represent aerosol direct climate forcing. However, WRF12-remap exhibits little

413   improvement over WRF60 in terms of reproducing the spatial variability of AOD ~~at these~~in

414   the visible wavelengths and further that WRF12-remap tends to be more strongly positively

415   biased in terms of mean monthly AOD outside of the summer months (Fig. 2 and Fig. 3).

416   Also the improvement in detection of areas of extreme AOD in the higher resolution

417   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

419   As documented above, WRF-Chem applied at either 60 or 12 km resolution over eastern

420   North America exhibits some skill in reproducing observed spatial fields of AOD and the

421   occurrence of extreme AOD values. However, marked discrepancies both in space and time

422   are found, and at least some of them show a significant dependence on model resolution.

423   Thus, we investigated a range of physical conditions and gas phase concentrations known to

424   be strongly determinant of aerosol dynamics in terms of the BSS as a function of model

425   resolution and also in terms of the mean monthly spatial patterns.

426   WRF12 even when remapped to 60 km provides more accurate description of key

427   meteorological variables such as specific humidity (Q) within the boundary layer, *PBLH*,

428   surface temperature and precipitation (Fig. 6, S1, S2 and S3) when ~~comparing~~ compared to

429   MERRA-2, as indicated by the positive BSS during almost all months (Fig. 7a). Good

430   qualitative agreement is observed for the spatial patterns and absolute magnitude of T_{2m} in

431   both WRF60 and WRF12-remap relative to MERRA-2 for all seasons (Fig. S1) leading to

432   only modest magnitude of BSS (i.e. value added by the higher resolution simulations (Fig.

433   7a)). The aerosol size distribution and therefore wavelength specific AOD exhibits a strong

434   sensitivity to Q (Santarpia et al., 2005) due to the presence of hygroscopic components in

435   atmospheric aerosols and thus the role of water uptake in determining aerosol diameter,

436   refractivity and extinction coefficient (Zieger et al., 2013). For example, the hygroscopic

437   growth factor, which indicates the change of aerosol diameter due to water uptake, is ~ 1.4

438   for pure ammonium sulfate with dry diameter of 532 nm at relative humidity of 80%, thus

439   biases in representation atmospheric humidity may lead to big errors in simulated aerosol size

440   and AOD (Flores et al., 2012). Our previous analyses of the 12 km resolution simulations

441   indicated overestimation of sulfate aerosols (a highly hygroscopic aerosol component, and

15

442 one which in many chemical forms exhibits strong hysteresis (Martin et al., 2004)) relative to

443 observed near-surface $PM_{2.5}$ concentrations during all seasons except for winter (Crippa et al.,

444 2016), leading to the hypothesis that simulated AOD and discrepancies therein may exhibit a

445 strong dependence on Q. Consistent with that postulate, Q_{PBL} from WRF12-remap exhibits a

446 ~~wet~~ moist bias in cloud-free grid cells mostly during warm months, whereas WRF60 is

447 characterized by a dry bias during all seasons (Fig. 6). Despite the positive bias, WRF12-

448 remap better captures the seasonal spatial patterns of Q_{PBL} in MERRA-2, leading to positive

449 BSS in all calendar months. Thus, there is added value by higher-resolution simulations in

450 representation of one of the key parameters dictating particle growth and optical properties.

451 Spatial patterns of differences in Q_{PBL} from WRF60 and WRF12-remap relative to MERRA-

452 2 (Fig. 6) exhibit similarities to differences in AOD (Fig. 4). WRF60 is dry-biased relative to

453 WRF12 particularly during the summer (and fall) and underestimates Q_{PBL} relative to

454 MERRA-2 during all seasons over the southern states and over most of continental US during

455 summer and fall. Conversely, WRF12-remap overestimates Q_{PBL} over most of continental US

456 during summer and fall relative to MERRA-2.

457 *PBLH* is a key variable for dictating near-surface aerosol concentrations but is highly

458 sensitive to the physical schemes applied, and biases appear to be domain and resolution

459 dependent. However, this parameter is comparatively difficult to assess because differences

460 in *PBLH* ~~PBL heights between~~from WRF-Chem and MERRA-2 may also originate from the

461 way they are computed (i.e. from heat diffusivity in MERRA-2 (Jordan et al., 2010) and from

462 turbulent kinetic energy in WRF-Chem (Janjić, 2002;von Engeln and Teixeira, 2013)~~)~~).

463 Nevertheless, t~~TFor example, t~~he Mellor-Yamada-Janjich *PBL* scheme combined with the

464 Noah Land Surface Model applied in this work was found to produce lower *PBL* heights

465 (Zhang et al., 2009) than other parameterizations. ~~(Jordan et al., 2010)(Janjić, 2002)(von~~

466 ~~Engeln and Teixeira, 2013)~~ Thus, the positive bias in simulated AOD and surface $PM_{2.5}$

467 concentrations (reported previously in (Crippa et al., 2016)) may be linked to the systematic

468 underestimation of *PBLH* simulated by WRF12-remap over continental US relative to

469 MERRA-2 during all seasons (except winter) with greatest bias over regions of complex

470 topography (Fig. S2). ~~However, a~~A positive bias (of several hundred meters) in terms of

471 *PBLH* for WRF simulations using the MYJ parameterization was previously reported for

472 high-resolution simulations over complex terrain (Rissman et al., 2013), and a positive bias in

473 *PBLH* is also observed in the 60 km simulations presented herein (Fig. S2). This may provide

474 a part~~ially~~ explanation for the strong negative bias in AOD in WRF60 during summer (Fig.

475   3). In general, the BSS indicate improvement in the simulation of *PBLH* in WRF12-remap
476   than in WRF60 (Fig. 7a).

477   Consistent with the dry bias in Q_{PBL} in WRF60, total accumulated precipitation is also
478   underestimated in WRF60, while WRF12-remap captures the absolute magnitudes and the
479   spatial patterns therein (Fig. S3). Analysis of hourly precipitation rates also showed higher
480   skill of WRF12-remap than WRF60 in correctly simulating precipitation occurrence (*HR*)
481   relative to MERRA-2 (Table S~~3~~2). More specifically WRF12-remap correctly predicts
482   between 40% and 70 % of precipitation events in MERRA-2 with highest skill during winter
483   months, whereas WRF60 output exhibits lower HR (~6% during summer and 30% during
484   winter). This result thus confirms our expectation of a strong sensitivity of model
485   performance to resolution due to the inherent scale dependence in the cumulus scheme.

486   Gas phase concentrations (transformed into z-scores) from WRF12-remap show higher
487   agreement with satellite observations during almost all months, as indicated by the positive
488   BSS (Fig. 7b). However given the limited availability of valid satellite observations
489   (especially during months with low radiation intensity), the BSS are likely only robust for the
490   summer months for all species. Nevertheless, with the exception of NH_3 during June, BSS for
491   all months are above or close to zero indicating <u>that</u> on average, the enhanced resolution
492   simulations do improve the quality of the simulation of the gas phase species even when
493   remapped to 60 km resolution. Further, the seasonal average spatial patterns of the total
494   columnar concentrations, expressed in terms of z-scores, also exhibit ~~high~~ qualitative
495   agreement with the satellite observations (Fig. S4-S7).

4 Concluding remarks

497   This analysis is one of the first to quantify the impact of model spatial resolution on the
498   spatio-temporal variability and magnitude of AOD<u>, and does so using simulations for a full</u>
499   <u>calendar year</u>. Application of WRF-Chem at two different resolutions (60 km and 12 km)
500   over eastern North America for a representative year (2008) leads to the following
501   conclusions:

502   -   Higher resolution simulations add value (i.e. enhance the fidelity of AOD at and near
503   to the peak in the solar spectrum) relative to a coarser run, although the improvement
504   in model performance is not uniform in space and time. Brier Skill Scores for the
505   remapped simulations (i.e. output from simulations conducted at 12 km (WRF12)
506   then averaged to 60 km, WRF12-remap) are positive for ten of twelve calendar moths,

17

507   and for AOD(λ=550 nm) exceed 0.5 for seven of twelve months.

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

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

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

521 - Higher-resolution simulations also add value in the representation of other key
522   meteorological variables such as temperature, boundary layer height and precipitation.
523   Both spatial patterns and precipitation occurrence are better captured by WRF12-
524   remap.

525 - More accurate representation of spatial patterns and magnitude of gaseous species that
526   playing a key role in particle formation and growth is also achieved by running WRF-
527   Chem at high resolution.

528 It is worthy of note that even the 12 km resolution WRF-Chem simulations exhibit substantial
529 differences in AOD relative to MODIS over eastern North America, and the agreement varies
530 only slightly with wavelength. This may be partially attributable to use of the modal approach
531 to represent the aerosol size distribution in order to enhance computational tractability. In this
532 application each mode has a fixed geometric standard deviation (σ_g), which can lead to biases
533 in simulated AOD in the visible wavelengths by up to 25% (Brock et al., 2016) (with the
534 model overestimating observations if the prescribed σ_g is larger than the observed one).
535 Setting σ_g = 2 for the accumulation mode (the default in WRF-Chem) may lead to an
536 overestimation of the number of particles at the end of the accumulation mode tail, and there
537 is evidence that a value of $\sigma_{g,acc}$=1.40 leads to higher agreement with observations (Mann et
538 al., 2012). Further possible sources of the AOD biases reported herein derive from selection
539 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~~ enhanced model performance might be achieved if seasonally varying emissions ~~would be~~were available. Future work will include a systematic sensitivity analysis of these effects.

Acknowledgments

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.

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, doi: http://dx.doi.org/10.1175/1520-0469(2003)060<0119:MVOTA>2.0.CO;2, 2003.

Ångström, A.: The parameters of atmospheric turbidity, Tellus, 16, 64-75, 10.1111/j.2153-3490.1964.tb00144.x, 1964.

Benjamini, Y., and Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society. Series B (Methodological), 57, 289-300, 1995.

Boucher, O., D. Randall, P. Artaxo, C. Bretherton, 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.

574   Brinksma, E. J., K. F. Boersma, P. F. Levelt, and R. D. McPeters OMI validation
575   requirements document, Version 1, Rep. RS-OMIE-KNMI-345, 66, 2003.

576   Brock, C. A., Wagner, N. L., Anderson, B. E., Attwood, A. R., Beyersdorf, A., Campuzano-
577   Jost, P., Carlton, A. G., Day, D. A., Diskin, G. S., Gordon, T. D., Jimenez, J. L., Lack, D. A.,
578   Liao, J., Markovic, M. Z., Middlebrook, A. M., Ng, N. L., Perring, A. E., Richardson, M. S.,
579   Schwarz, J. P., Washenfelder, R. A., Welti, A., Xu, L., Ziemba, L. D., and Murphy, D. M.:
580   Aerosol optical properties in the southeastern United States in summer – Part 1: Hygroscopic
581   growth, Atmos. Chem. Phys., 16, 25695-25738, doi:10.5194/acp-16-5009-2016, 2016.

582   Chance, K.: OMI algorithm theoretical basis document, volume IV: OMI trace gas
583   algorithms, 2002.

584   Chen, F., and Dudhia, J.: Coupling an advanced land surface–hydrology model with the Penn
585   State–NCAR MM5 modeling system. Part I: model implementation and sensitivity, Monthly
586   Weather Review, 129, 569-585, doi:10.1175/1520-
587   0493(2001)129<0569:CAALSH>2.0.CO;2, 2001.

588   Chin, M., Kahn, R. A., and Schwartz , S. E.: Atmospheric Aerosols Properties and Climate
589   Impacts. A Report by the U.S. Climate Change Science Program and the Subcommittee on
590   Global Change Research, in, National Aeronautics and Space Administration, Washington,
591   D.C., USA, 128, 2009.

592   Crippa, P., Sullivan, R. C., Thota, A., and Pryor, S. C.: Evaluating the skill of high-resolution
593   WRF-Chem simulations in describing drivers of aerosol direct climate forcing on the regional
594   scale, Atmos. Chem. Phys., 16, 397-416, 10.5194/acp-16-397-2016, 2016.

595   Di Luca, A., de Elía, R., and Laprise, R.: Challenges in the Quest for Added Value of
596   Regional Climate Dynamical Downscaling, Curr Clim Change Rep, 1, 10-21,
597   10.1007/s40641-015-0003-9, 2015.

598   Diaconescu, E., and Laprise, R.: Can added value be expected in RCM-simulated large
599   scales?, Clim Dyn, 41, 1769-1800, 10.1007/s00382-012-1649-9, 2013.

600   Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J. F., Pfister, G. G., Fillmore, D.,
601   Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G.,
602   Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model
603   for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geoscientific Model
604   Development, 3, 43-67, doi:10.5194/gmd-3-43-2010, 2010.

605   Fast, J. D., Gustafson, W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G.,
606   Grell, G. A., and Peckham, S. E.: Evolution of ozone, particulates, and aerosol direct
607   radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-
608   aerosol model, Journal of Geophysical Research: Atmospheres, 111, D21305,
609   10.1029/2005JD006721, 2006.

610   Fioletov, V. E., McLinden, C. A., Krotkov, N., Moran, M. D., and Yang, K.: Estimation of
611   SO_2 emissions using OMI retrievals, Geophysical Research Letters, 38, L21811,
612   10.1029/2011GL049402, 2011.

613   Flores, J. M., Bar-Or, R. Z., Bluvshtein, N., Abo-Riziq, A., Kostinski, A., Borrmann, S.,
614   Koren, I., Koren, I., and Rudich, Y.: Absorbing aerosols at high relative humidity: linking

615 hygroscopic growth to optical properties, Atmospheric Chemistry and Physics, 12, 5511-
616 5521, 10.5194/acp-12-5511-2012, 2012.

617 Grell, G. A., and Dévényi, D.: A generalized approach to parameterizing convection
618 combining ensemble and data assimilation techniques, Geophysical Research Letters, 29, 38-
619 31-38-34, 10.1029/2002GL015311, 2002.

620 Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and
621 Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmos. Environ., 39,
622 6957-6975, 10.1016/j.atmosenv.2005.04.027, 2005.

623 Guenther, A., Zimmerman, P., and Wildermuth, M.: Natural volatile organic compound
624 emission rate estimates for U.S. woodland landscapes, Atmos. Environ., 28, 1197-1210,
625 10.1016/1352-2310(94)90297-6, 1994.

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

629 Gustafson, W. I., Qian, Y., and Fast, J. D.: Downscaling aerosols and the impact of neglected
630 subgrid processes on direct aerosol radiative forcing for a representative global climate model
631 grid spacing, Journal of Geophysical Research: Atmospheres, 116, D13303,
632 10.1029/2010JD015480, 2011.

633 Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A Revised Approach to Ice Microphysical
634 Processes for the Bulk Parameterization of Clouds and Precipitation, Monthly Weather
635 Review, 132, 103-120, doi:10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2, 2004.

636 Hyer, E. J., Reid, J. S., and Zhang, J.: An over-land aerosol optical depth data set for data
637 assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth
638 retrievals, Atmospheric Measurement Techniques, 4, 379-408, 10.5194/amt-4-379-2011,
639 2011.

640 Janjić, Z. I.: The Step-Mountain Eta Coordinate Model: Further Developments of the
641 Convection, Viscous Sublayer, and Turbulence Closure Schemes, Monthly Weather Review,
642 122, 927-945, doi:10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2, 1994.

643 Janjić, Z. I.: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the
644 NCEP Meso model, NCEP office note, 437, 61, 2002.

645 Jordan, N. S., Hoff, R. M., and Bacmeister, J. T.: Validation of Goddard Earth Observing
646 System-version 5 MERRA planetary boundary layer heights using CALIPSO, J. Geophys.
647 Res.-Atmos., 115, 10.1029/2009jd013777, 2010.

648 Krotkov, N. A., McClure, B., Dickerson, R. R., Carn, S. A., Li, C., Bhartia, P. K., Yang, K.,
649 Krueger, A. J., Li, Z., Levelt, P. F., Chen, H., Wang, P., and Lu, D.: Validation of SO_2
650 retrievals from the Ozone Monitoring Instrument over NE China, Journal of Geophysical
651 Research: Atmospheres, 113, D16S40, 10.1029/2007JD008818, 2008.

652 Leibensperger, E., Mickley, L. J., Jacob, D. J., Chen, W.-T., Seinfeld, J., Nenes, A., Adams,
653 P., Streets, D., Kumar, N., and Rind, D.: Climatic effects of 1950–2050 changes in US

654 anthropogenic aerosols–Part 1: Aerosol trends and radiative forcing, Atmospheric Chemistry
655 and Physics, 12, 3333-3348, doi:10.5194/acp-12-3333-2012, 2012.

656 Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N.
657 C.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric
658 Measurement Techniques, 6, 2989-3034, 10.5194/amt-6-2989-2013, 2013.

659 Long, M., Yantosca, R., Nielsen, J., Keller, C., da Silva, A., Sulprizio, M., Pawson, S., and
660 Jacob, D.: Development of a grid-independent GEOS-Chem chemical transport model (v9-
661 02) as an atmospheric chemistry module for Earth system models, Geoscientific Model
662 Development, 8, 595-602, doi:10.5194/gmd-8-595-2015, 2015.

663 Mann, G. W., Carslaw, K. S., Ridley, D. A., Spracklen, D. V., Pringle, K. J., Merikanto, J.,
664 Korhonen, H., Schwarz, J. P., Lee, L. A., Manktelow, P. T., Woodhouse, M. T., Schmidt, A.,
665 Breider, T. J., Emmerson, K. M., Reddington, C. L., Chipperfield, M. P., and Pickering, S. J.:
666 Intercomparison of modal and sectional aerosol microphysics representations within the same
667 3-D global chemical transport model, Atmospheric Chemistry and Physics, 12, 4449-4476,
668 10.5194/acp-12-4449-2012, 2012.

669 Martin, S. T., Hung, H. M., Park, R. J., Jacob, D. J., Spurr, R. J. D., Chance, K. V., and Chin,
670 M.: Effects of the physical state of tropospheric ammonium-sulfate-nitrate particles on global
671 aerosol direct radiative forcing, Atmospheric Chemistry and Physics, 4, 183-214,
672 doi:10.5194/acp-4-183-2004, 2004.

673 McComiskey, A., Schwartz, S. E., Schmid, B., Guan, H., Lewis, E. R., Ricchiazzi, P., and
674 Ogren, J. A.: Direct aerosol forcing: Calculation from observables and sensitivities to inputs,
675 Journal of Geophysical Research: Atmospheres, 113, D09202, 10.1029/2007JD009170, 2008.

676 McLinden, C. A., Fioletov, V., Boersma, K. F., Kharol, S. K., Krotkov, N., Lamsal, L.,
677 Makar, P. A., Martin, R. V., Veefkind, J. P., and Yang, K.: Improved satellite retrievals of
678 NO_2 and SO_2 over the Canadian oil sands and comparisons with surface measurements,
679 Atmos. Chem. Phys., 14, 3637-3656, 10.5194/acp-14-3637-2014, 2014.

680 Meehl, G. A., Moss, R., Taylor, K. A., Eyring, V., Stouffer, R. J., Sandrine, B., and Stevens,
681 B.: Climate model intercomparisons: preparing for the next phase, Eos, Transaction,
682 American Geophysical Union, 95, 77-84, doi:10.1002/2014EO09, 2014.

683 Misenis, C., and Zhang, Y.: An examination of sensitivity of WRF/Chem predictions to
684 physical parameterizations, horizontal grid spacing, and nesting options, Atmospheric
685 Research, 97, 315-334, 10.1016/j.atmosres.2010.04.005, 2010.

686 Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative
687 transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the
688 longwave, Journal of Geophysical Research: Atmospheres, 102, 16663-16682,
689 10.1029/97JD00237, 1997.

690 Molod, A., Takacs, L., Suarez, M., and Bacmeister, J.: Development of the GEOS-5
691 atmospheric general circulation model: evolution from MERRA to MERRA2, Geoscientific
692 Model Development, 8, 1339-1356, 10.5194/gmd-8-1339-2015, 2015.

693   Murphy, A. H., and Epstein, E. S.: Skill scores and correlation-coefficients in model
694   verification, Monthly Weather Review, 117, 572-581, 10.1175/1520-
695   0493(1989)117<0572:ssacci>2.0.co;2, 1989.

696   Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., Bian, H.,
697   Bellouin, N., Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D.,
698   Iversen, T., Kinne, S., Kirkevag, A., Lamarque, J. F., Lin, G., Liu, X., Lund, M. T., Luo, G.,
699   Ma, X., van Noije, T., Penner, J. E., Rasch, P. J., Ruiz, A., Seland, O., Skeie, R. B., Stier, P.,
700   Takemura, T., Tsigaridis, K., Wang, P., Wang, Z., Xu, L., Yu, H., Yu, F., Yoon, J. H., Zhang,
701   K., Zhang, H., and Zhou, C.: Radiative forcing of the direct aerosol effect from AeroCom
702   Phase II simulations, Atmospheric Chemistry and Physics, 13, 1853-1877, 10.5194/acp-13-
703   1853-2013, 2013a.

704   Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D.,
705   Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura,
706   T., and Zhang, H.: Anthropogenic and Natural Radiative Forcing, in: Climate Change 2013:
707   The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report
708   of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D.,
709   Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and
710   Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York,
711   NY, USA, 659–740, 2013b.

712   Qian, Y., Gustafson Jr, W. I., and Fast, J. D.: An investigation of the sub-grid variability of
713   trace gases and aerosols for global climate modeling, Atmos. Chem. Phys., 10, 6917-6946,
714   10.5194/acp-10-6917-2010, 2010.

715   Rissman, J., Arunachalam, S., Woody, M., West, J. J., BenDor, T., and Binkowski, F. S.: A
716   plume-in-grid approach to characterize air quality impacts of aircraft emissions at the
717   Hartsfield–Jackson Atlanta International Airport, Atmos. Chem. Phys., 13, 9285-9302,
718   10.5194/acp-13-9285-2013, 2013.

719   Rockel, B., Castro, C. L., Pielke, R. A., von Storch, H., and Leoncini, G.: Dynamical
720   downscaling: Assessment of model system dependent retained and added variability for two
721   different regional climate models, Journal of Geophysical Research: Atmospheres, 113,
722   D21107, 10.1029/2007JD009461, 2008.

723   Santarpia, J. L., Gasparini, R., Li, R. J., and Collins, D. R.: Diurnal variations in the
724   hygroscopic growth cycles of ambient aerosol populations, J. Geophys. Res.-Atmos., 110,
725   10.1029/2004jd005279, 2005.

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

729   Schuster, G. L., O. Dubovik, and B. N. Holben Angstrom exponent and bimodal aerosol size
730   distributions, J. Geophys. Res.-Atmos., 111, D07207, doi:10.1029/2005JD006328., 2006.

731   Seinfeld, J. H., and Pandis, S. N.: Atmospheric chemistry and physics: from air pollution to
732   climate change, John Wiley & Sons, 1152 pp., 2016.

733   Simes, R. J.: An improved Bonferroni procedure for multiple tests of significance,
734   Biometrika, 73, 751-754, 10.2307/2336545, 1986.

735   Simpson, D., Guenther, A., Hewitt, C. N., and Steinbrecher, R.: Biogenic emissions in
736   Europe. 1. estimates and uncertainties, J. Geophys. Res.-Atmos., 100, 22875-22890,
737   10.1029/95jd02368, 1995.

738   Stocker, T. F. a. Q., D. and Plattner, G.-K. and Alexander, L.V. and Allen, S.K. and Bindoff,
739   N.L. and Bréon, F.-M. and Church, J.A. and Cubasch, U. and Emori, S. and Forster, P. and
740   Friedlingstein, P. and Gillett, N. and Gregory, J.M. and Hartmann, D.L. and Jansen, E. and
741   Kirtman, B. and Knutti, R. and Krishna Kumar, K. and Lemke, P. and Marotzke, J. and
742   Masson-Delmotte, V. and Meehl, G.A. and Mokhov, I.I. and Piao, S. and Ramaswamy, V.
743   and Randall, D. and Rhein, M. and Rojas, M. and Sabine, C. and Shindell, D. and Talley,
744   L.D. and Vaughan, D.G. and Xie, S.-P.: Summary for Policymakers, in: Climate Change
745   2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment
746   Report of the Intergovernmental Panel on Climate Change, Cambridge University Press,
747   Cambridge, United Kingdom and New York, NY, USA, 33–115, 2013.

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

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

753   Tilmes, S., Lamarque, J.-F., Emmons, L., Kinnison, D., Ma, P.-L., Liu, X., Ghan, S.,
754   Bardeen, C., Arnold, S., and Deeter, M.: Description and evaluation of tropospheric
755   chemistry and aerosols in the Community Earth System Model (CESM1. 2), Geoscientific
756   Model Development, 8, 1395-1426, doi:10.5194/gmd-8-1395-2015, 2015.

757   Tomasi, C., Caroli, E., and Vitale, V.: Study of the Relationship between Ångström's
758   Wavelength Exponent and Junge Particle Size Distribution Exponent, Journal of Climate and
759   Applied Meteorology, 22, 1707-1716, 10.1175/1520-
760   0450(1983)022<1707:SOTRBW>2.0.CO;2, 1983.

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

763   Vinken, G. C. M., Boersma, K. F., van Donkelaar, A., and Zhang, L.: Constraints on ship
764   NO_x emissions in Europe using GEOS-Chem and OMI satellite NO_2 observations, Atmos.
765   Chem. Phys., 14, 1353-1369, 10.5194/acp-14-1353-2014, 2014.

766   von Engeln, A., and Teixeira, J.: A Planetary Boundary Layer Height Climatology Derived
767   from ECMWF Reanalysis Data, Journal of Climate, 26, 6575–6590, doi: 10.1175/JCLI-D-12-
768   00385.1, 2013.

769   Weigum, N., Schutgens, N., and Stier, P.: Effect of aerosol sub-grid variability on aerosol
770   optical depth and cloud condensation nuclei: Implications for global aerosol modelling,
771   Atmos. Chem. Phys. Discuss., 2016, 1-36, 10.5194/acp-2016-360, 2016.

772   Whitburn, S., Van Damme, M., Clarisse, L., Bauduin, S., Heald, C., Hadji-Lazaro, J.,
773   Hurtmans, D., Zondlo, M. A., Clerbaux, C., Coheur, P.-F. : A flexible and robust neural

774   network IASI-NH$_3$ retrieval algorithm, J. Geophys. Res. Atmos, In Press,
775   10.1002/2016JD024828, 2016.

776   Wild, O., Zhu, X., and Prather, M. J.: Fast-J: Accurate Simulation of In- and Below-Cloud
777   Photolysis in Tropospheric Chemical Models, Journal of Atmospheric Chemistry, 37, 245-
778   282, 10.1023/a:1006415919030, 2000.

779   Zhang, Y., Dubey, M. K., Olsen, S. C., Zheng, J., and Zhang, R.: Comparisons of
780   WRF/Chem simulations in Mexico City with ground-based RAMA measurements during the
781   2006-MILAGRO, Atmospheric Chemistry and Physics, 9, 3777-3798, doi:10.5194/acp-9-
782   3777-2009, 2009.

783   Zieger, P., Fierz-Schmidhauser, R., Weingartner, E., and Baltensperger, U.: Effects of
784   relative humidity on aerosol light scattering: results from different European sites,
785   Atmospheric Chemistry and Physics, 13, 10609-10631, 10.5194/acp-13-10609-2013, 2013.

786

787
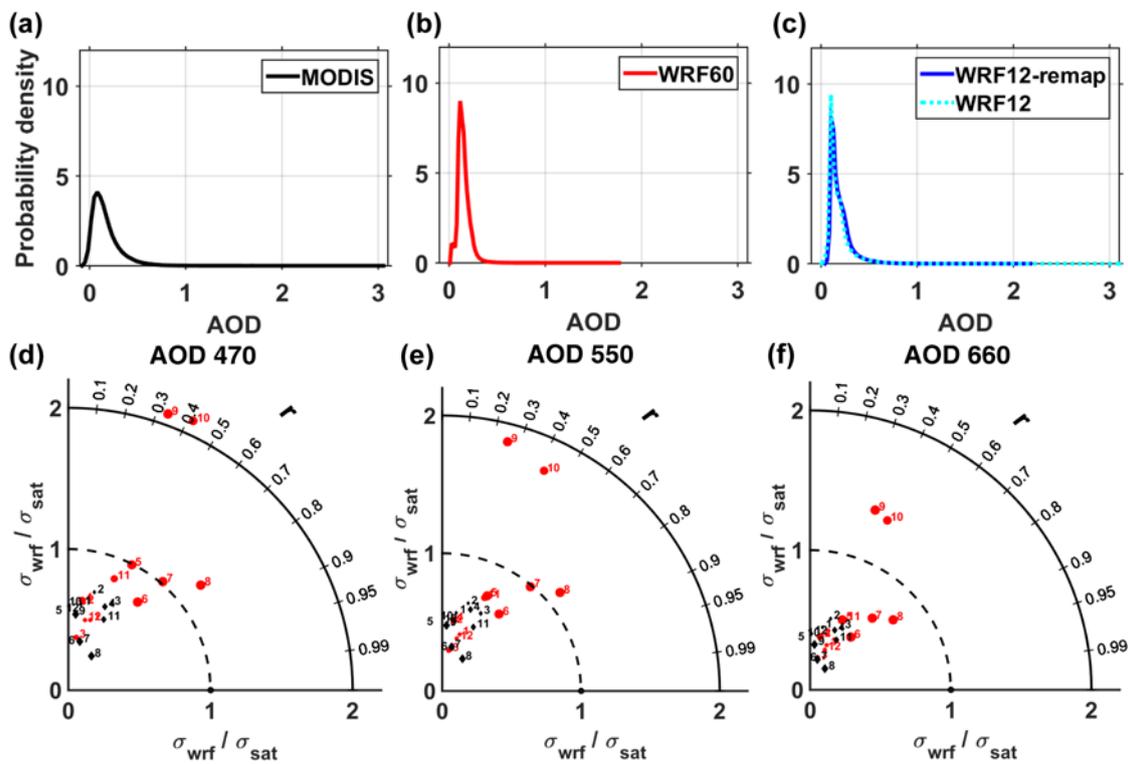
Figures

Figure 1. Probability density function of once daily AOD at a wavelength (λ) 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 (λ) 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
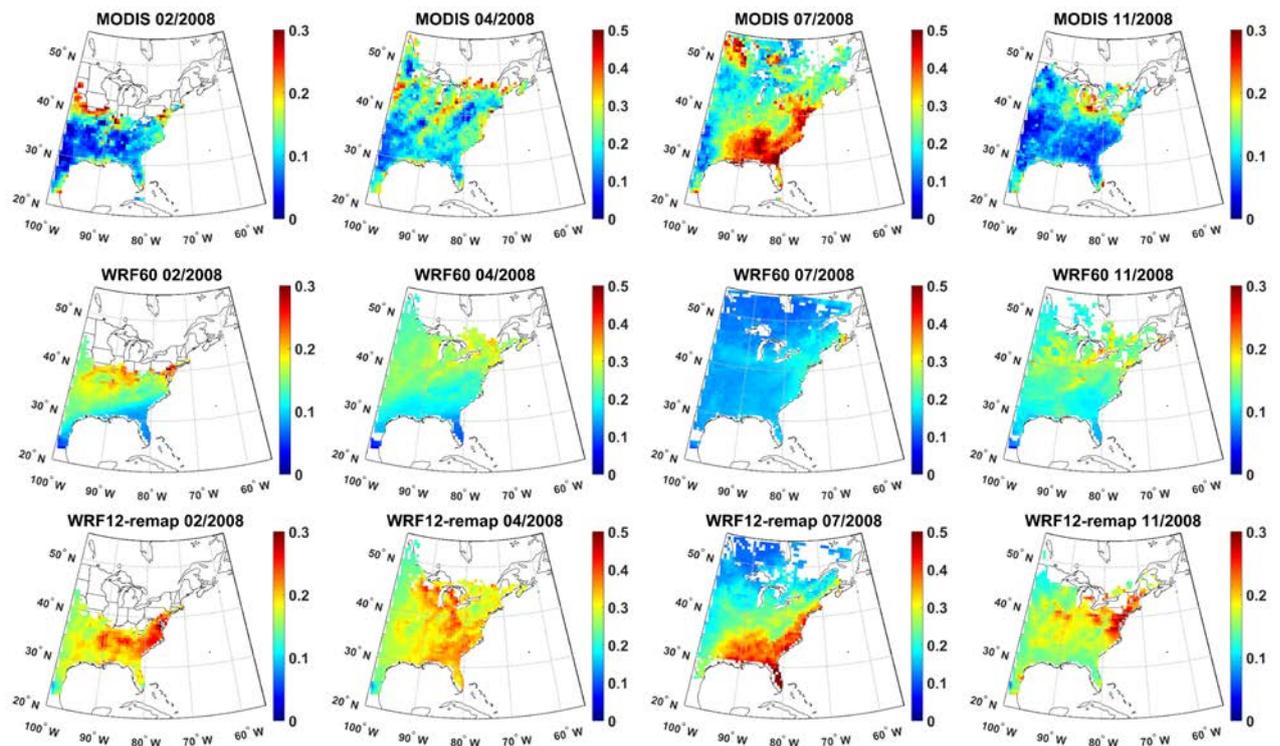
numbers by each symbol denote the calendar month (e.g. 1=January).

798

799

Figure 2. First line: Number of paired AOD observations at a wavelength (λ) 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 α=0.10 are indicated in red and have been corrected for multiple testing using a False Discovery Rate approach. <u>The number of observations of cloud-free conditions summed across all days in each season and all grid cells is also reported (black=MODIS, blue=WRF60, red=WRF12-remap).</u>

810

811

Figure 3. Monthly mean AOD at a wavelength (λ) 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.**

818
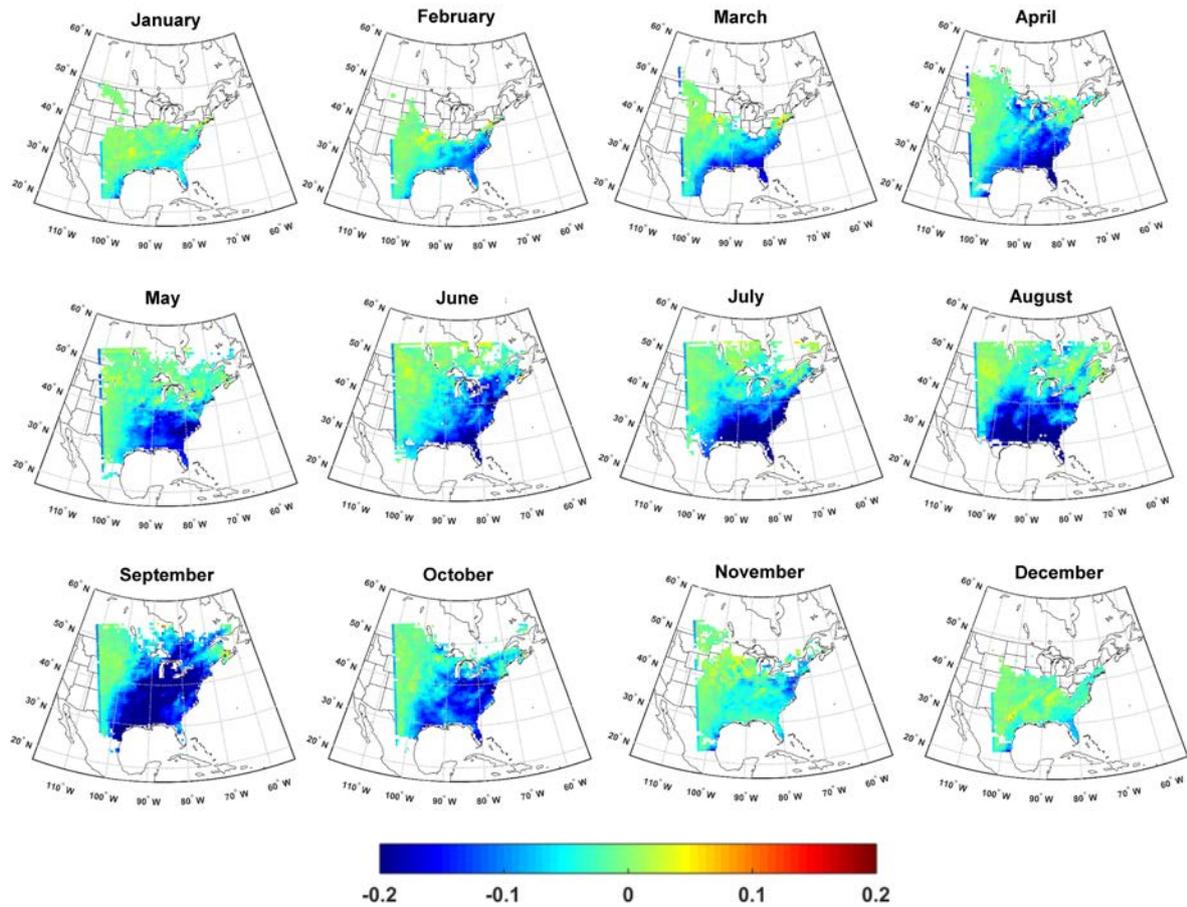
Figure 4. Difference in monthly mean AOD at a wavelength (λ) of 550 nm between WRF-Chem simulations conducted at 60 km resolution (WRF60) and output from WRF-Chem simulations conducted with a resolution of 12 km but remapped to 60 km (WRF12-remap). Differences are computed as WRF60 minus WRF12-remap. Similar spatial patterns and magnitudes of differences are found for λ of 470 and 660 nm. The calendar months of 2008 are shown in the titles of each panel.

827
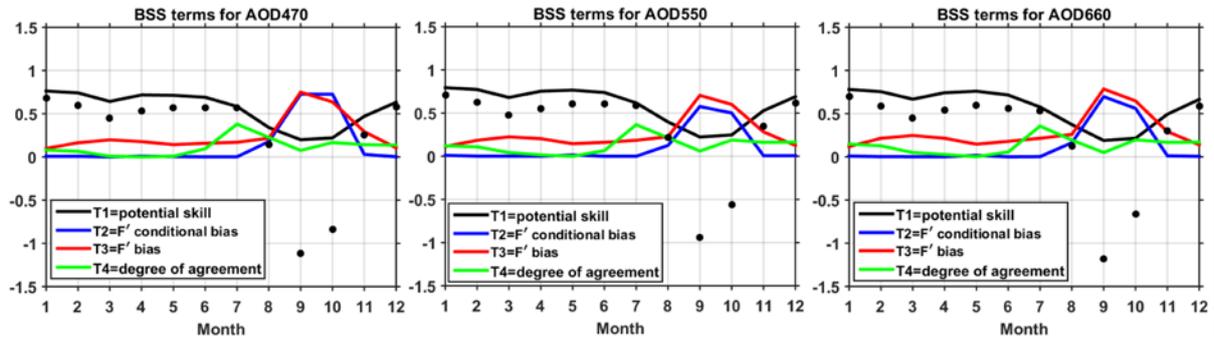
Figure 5. Brier Skill Scores (BSS, black dots) for monthly mean AOD by calendar month (1=January) for AOD at 470, 550 and 660 nm. In this analysis of model skill WRF12 output is mapped to the WRF60 grid (WRF12-remap) and BSS are computed using MODIS as the target, WRF60 as the reference forecast and WRF12-remap as the forecast. Also shown by the color lines are the contributions of different terms to BSS.
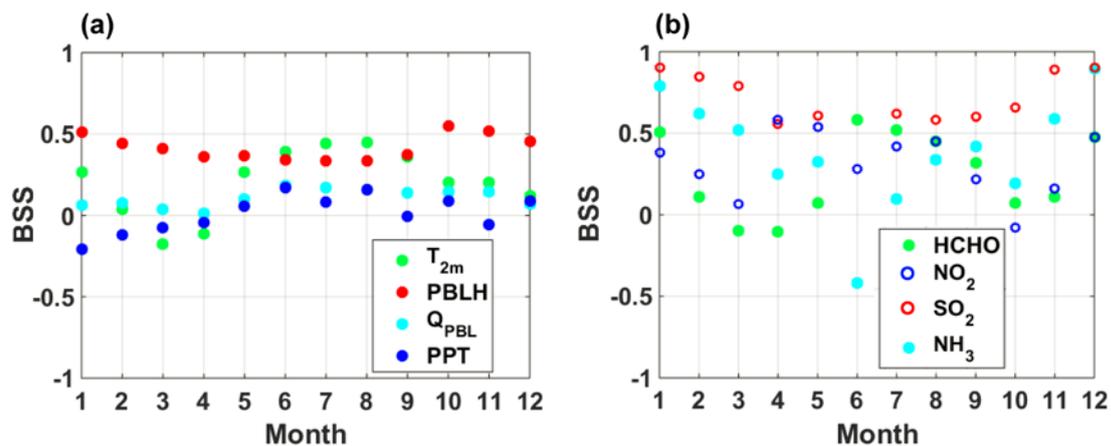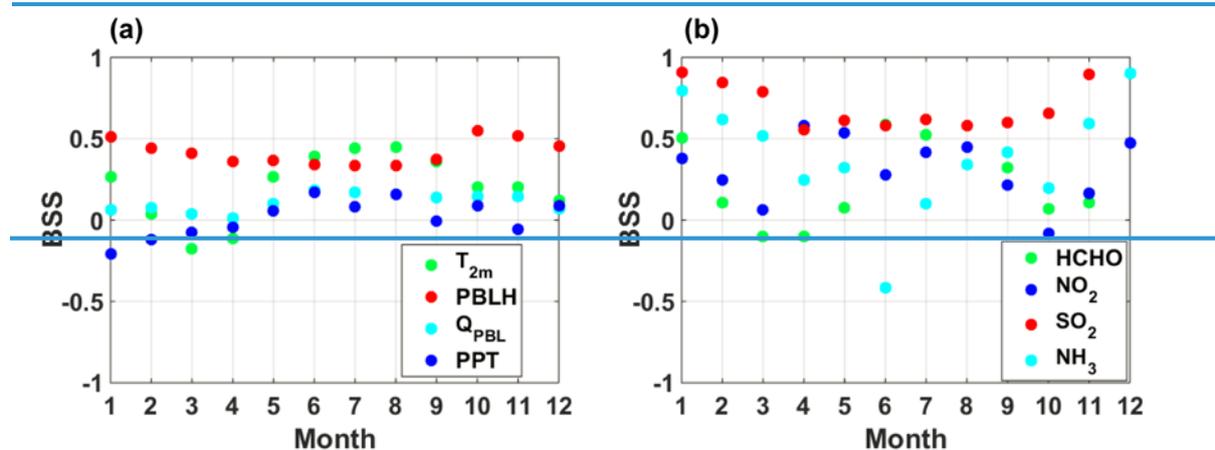
833

834
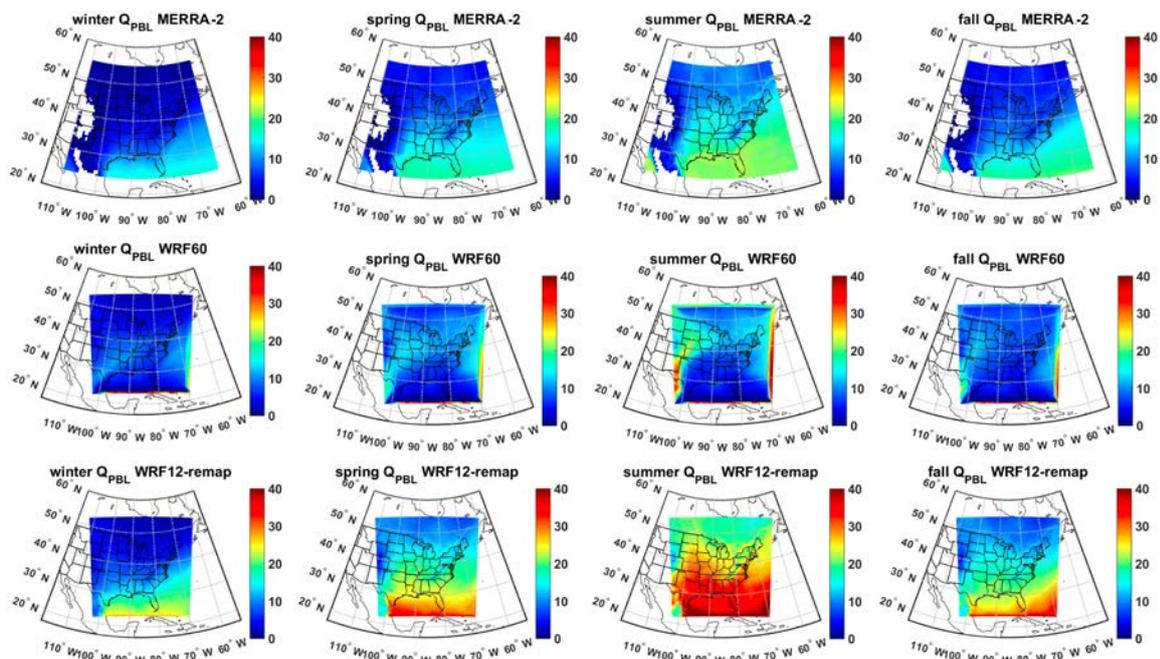
Figure 6. Seasonal mean specific humidity [kg m⁻²] integrated from the surface to 825 hPa (Q_{PBL}) from MERRA-2 (first row) assuming an average air density in the *PBL* of 1.1 kg m⁻³, WRF60 (second row), and WRF12-remap (third row). The data are 3-hourly and show only cloud-free hours in all three data sets.

839

840

841

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.

848 **Tables**

849 **Table 1. Physical and chemical schemes adopted in the WRF-Chem simulations**
850 **presented herein.**

Simulation settings	Values
Domain size	300 × 300 (60 × 60) grid points
Horizontal resolution	12 km (60 km)
Vertical resolution	32 levels up to 50 hPa
Timestep for physics	72 s (300 s)
Timestep for chemistry	5 s
Physics option	**Adopted scheme**
Microphysics	WRF Single-Moment 5-class (Hong et al., 2004)
Longwave Radiation	Rapid Radiative Transfer Model (RRTM) (Mlawer et al., 1997)
Shortwave Radiation	Goddard (Fast et al., 2006)(Chou, 1994){Fast, 2006 #191}
Surface layer	Monin Obhukov similarity (Janjić, 2002;Janjić, 1994)
Land Surface	Noah Land Surface Model (Chen and Dudhia, 2001)
Planetary boundary layer	Mellor-Yamada-Janjich (Janjić, 1994)
Cumulus parameterizations	Grell 3 (Grell and Dévényi, 2002)
Chemistry option	**Adopted scheme**
Photolysis	Fast J (Wild et al., 2000)
Gas-phase chemistry	RADM2 (Stockwell et al., 1990)
Aerosols	MADE/SORGAM (Ackermann et al., 1998;Schell et al., 2001)
Anthropogenic emissions	NEI (2005) (US-EPA, 2009)
Biogenic emissions	Guenther, from USGS land use classification (Guenther et al., 1994;Guenther et al., 1993;Simpson et al., 1995)

851

852

Table 2. 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. $p \leq \dfrac{0.05}{9 \times 12} = 4.63 \times 10^{-4}$ 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.

Month→/ Variable↓	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
470-12	0.238	0.150	0.137	0.147	0.377	0.581	0.610	0.723	0.352	0.306	0.259	0.212
470-60	0.156	0.226	0.438	0.412	-0.219	-0.146	0.379	0.601	0.087	-0.051	0.500	-0.059
470-remap	0.295	0.197	0.250	0.182	0.516	0.637	0.675	0.777	0.368	0.441	0.315	0.274
550-12	0.223	0.124	0.142	0.146	0.349	0.541	0.580	0.689	0.275	0.301	0.280	0.215
550-60	0.179	0.244	0.429	0.332	-0.288	-0.188	0.324	0.567	0.073	-0.077	0.491	0.002
550-remap	0.297	0.164	0.261	0.199	0.493	0.605	0.651	0.747	0.286	0.437	0.352	0.309
660-12	0.217	0.136	0.165	0.152	0.324	0.476	0.540	0.644	0.183	0.290	0.292	0.221
660-60	0.191	0.230	0.437	0.402	-0.305	-0.189	0.389	0.616	0.099	-0.137	0.536	0.049
660-remap	0.356	0.211	0.289	0.208	0.480	0.624	0.669	0.772	0.371	0.432	0.393	0.368

866 **Table 3. ~~Table 2.~~ Spatial coherence in the identification of extreme AOD values (i.e.**
867 **areas with AOD>75th percentile over space for each month) between WRF-Chem at**
868 **different resolutions relative to MODIS. No significant wavelength dependence is found**
869 **for model skill in identifying extreme AOD so results are only shown for λ = 550 nm.**
870 **The different model output is denoted by -60 for simulations at 60 km, -12 for**
871 **simulations at 12 km resolution, and as –remap for simulations at 12 km but with the**
872 **output remapped to 60 km. The *Accuracy* (Acc) indicates the fraction of grid cells co-**
873 **identified as extremes and non-extremes between WRF-Chem and MODIS relative to**
874 **the total number of cells with valid data. The *Hit Rate* (*HR*) is the probability of correct**
875 **forecast and is the proportion of cells correctly identified as extremes by both WRF-**
876 **Chem and MODIS. The yellow shading indicates the model resolution with highest skill**
877 **in each month for AOD at 550 nm.**

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

878

879

880

881

882

883

884

885

Supplementary Materials for the manuscript:

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

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

Table S1. Physical and chemical schemes adopted in the WRF-Chem simulations presented herein.

Simulation settings	Values
Domain size	300 × 300 (60 × 60) grid points
Horizontal resolution	12 km (60 km)
Vertical resolution	32 levels up to 50 hPa
Timestep for physics	72 s (300 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 S12. Ratio of spatial variability (i.e. the standard deviation of AOD computed across all grid cells) between AOD at wavelengths (λ) of 470, 550 and 660 nm from MODIS observations mapped at 60 km and WRF-Chem simulations conducted at 60 km resolution (WRF60, shown in the table as -60), at 12 km resolution (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, the ratio of standard deviations from WRF60 and WRF12-remap may be computed on slightly different observations and sample size. The yellow shading shows for each month and λ the model with ratio of standard deviations closer to 1.

Month→/ Variable↓	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
470-12	0.489	0.581	0.382	0.595	0.806	0.802	1.033	1.20	1.935	1.698	0.766	0.457
470-60	0.615	0.717	0.682	0.648	0.556	0.331	0.353	0.291	0.541	0.605	0.562	0.564
470-remap	0.522	0.630	0.380	0.644	0.993	0.791	1.018	1.194	2.079	2.099	0.853	0.512
550-12	0.406	0.475	0.307	0.480	0.630	0.690	0.996	1.106	1.709	1.401	0.663	0.370
550-60	0.578	0.663	0.629	0.624	0.502	0.302	0.327	0.274	0.480	0.525	0.518	0.505
550-remap	0.431	0.503	0.299	0.524	0.764	0.693	0.990	1.110	1.872	1.758	0.745	0.396
660-12	0.401	0.454	0.283	0.462	0.571	0.671	1.004	1.114	1.684	1.343	0.665	0.351
660-60	0.458	0.531	0.497	0.462	0.378	0.214	0.225	0.184	0.328	0.391	0.402	0.405
660-remap	0.342	0.393	0.235	0.391	0.553	0.474	0.676	0.777	1.369	1.331	0.557	0.307

Table S3S2. Spatial coherence in the identification of hourly precipitation between WRF-Chem at different resolutions relative to MERRA-2. The Hit Rate (*HR*) indicates the probability of correct forecast and is the proportion of cells correctly identified as with precipitation by both WRF-Chem and MERRA-2. The Mean Fractional Bias (MFB) in space is also reported for each month and computed from the hourly precipitation rates. The yellow shading indicates the model resolution with highest HR and lower absolute MFB in each month for precipitation.

Month→/ Metric↓	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
HR-60	0.344	0.298	0.228	0.122	0.083	0.072	0.057	0.059	0.067	0.078	0.154	0.218
HR-remap	0.698	0.715	0.680	0.539	0.402	0.440	0.479	0.438	0.438	0.454	0.581	0.666
MFB-60	-0.340	-0.347	-0.384	-0.442	-0.462	-0.468	-0.475	-0.474	-0.469	-0.459	-0.423	-0.385
MFB-12	-0.095	-0.068	-0.065	-0.168	-0.273	-0.269	-0.260	-0.274	-0.281	-0.261	-0.170	-0.119

Figure S1. Seasonal mean of hourly temperature at 2 meters [K] from MERRA-2 (first row), WRF60 (second row), and WRF12-remap (third row), for simultaneous data from all three datasets.
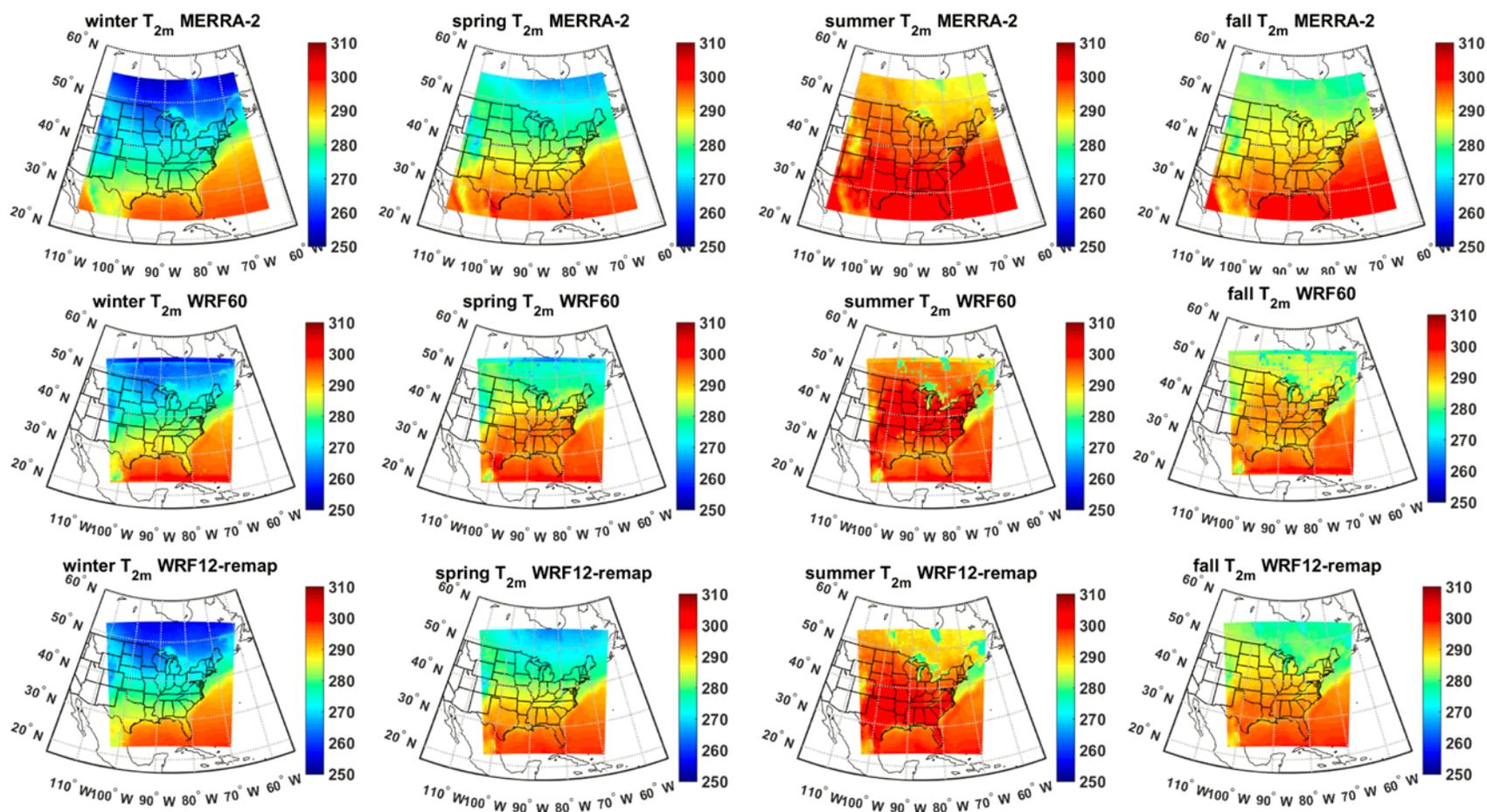
Figure S2. Seasonal average of hourly Planetary Boundary Layer Height, *PBLH* [m] from MERRA-2 (first row), WRF60 (second row), and WRF12-remap (third row), for simultaneous hours of the three datasets.
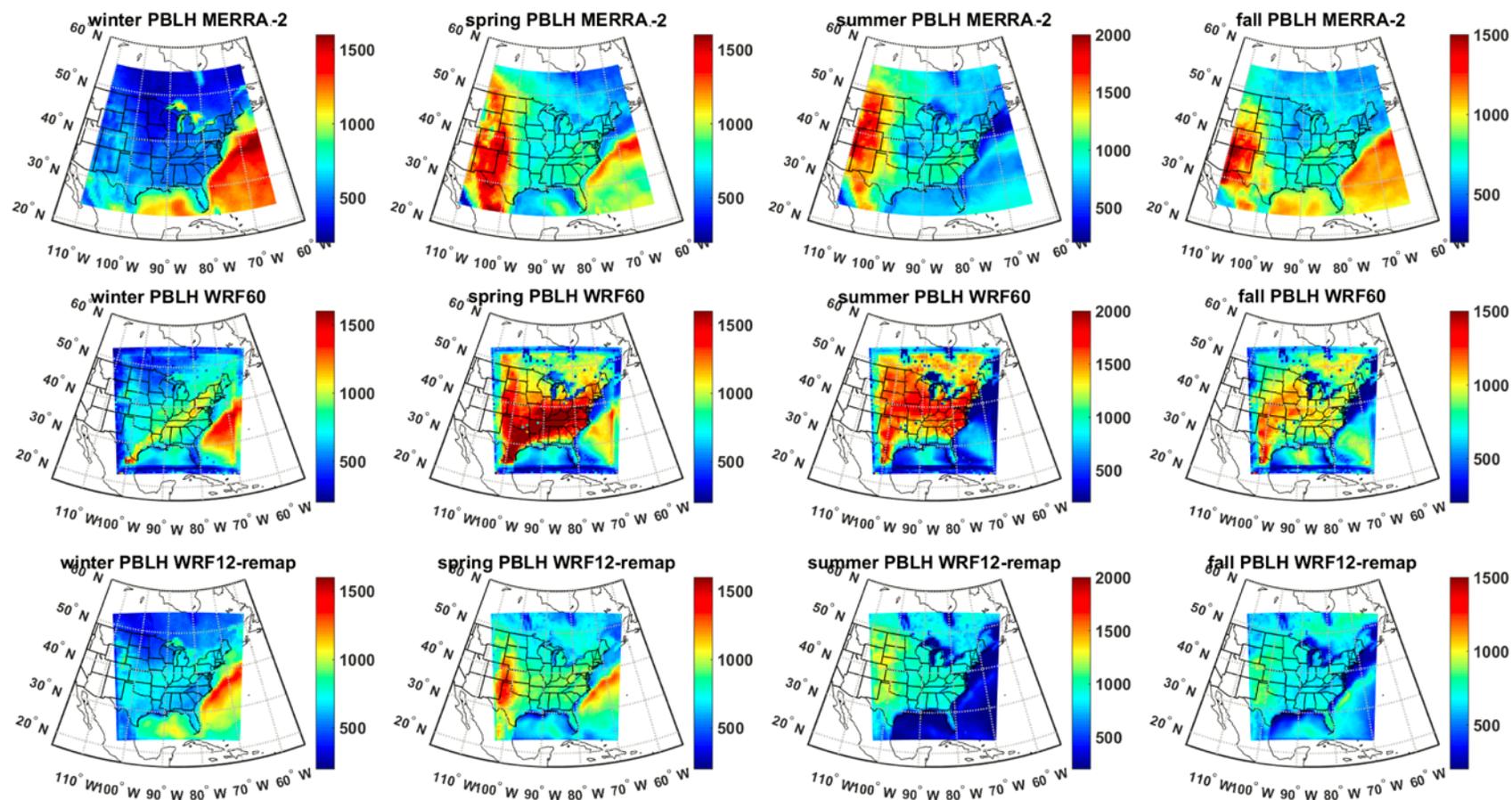
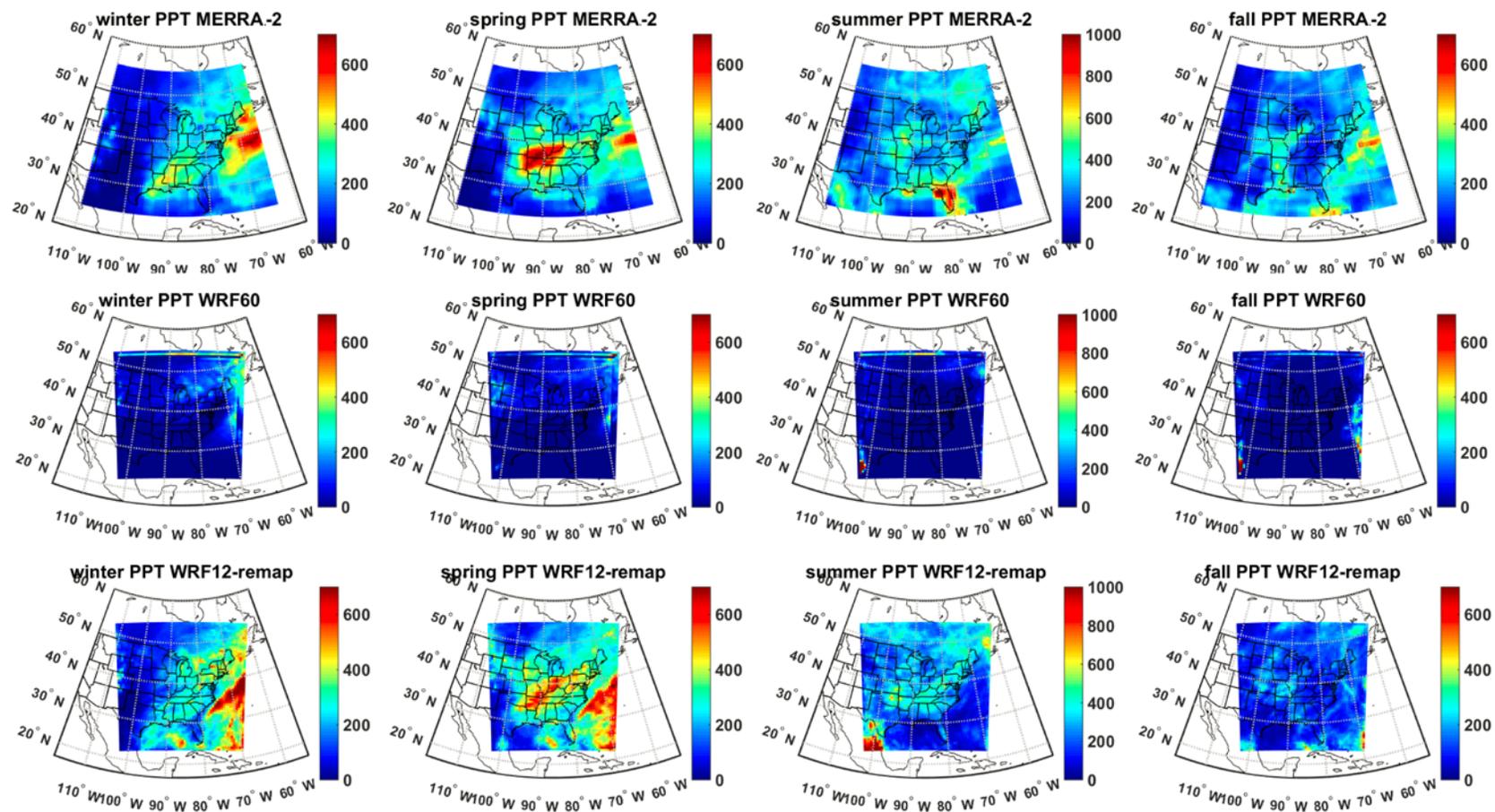Figure S3. Seasonal total precipitation (mm) from MERRA-2 (first row), WRF60 (second row), and WRF12-remap (third row).
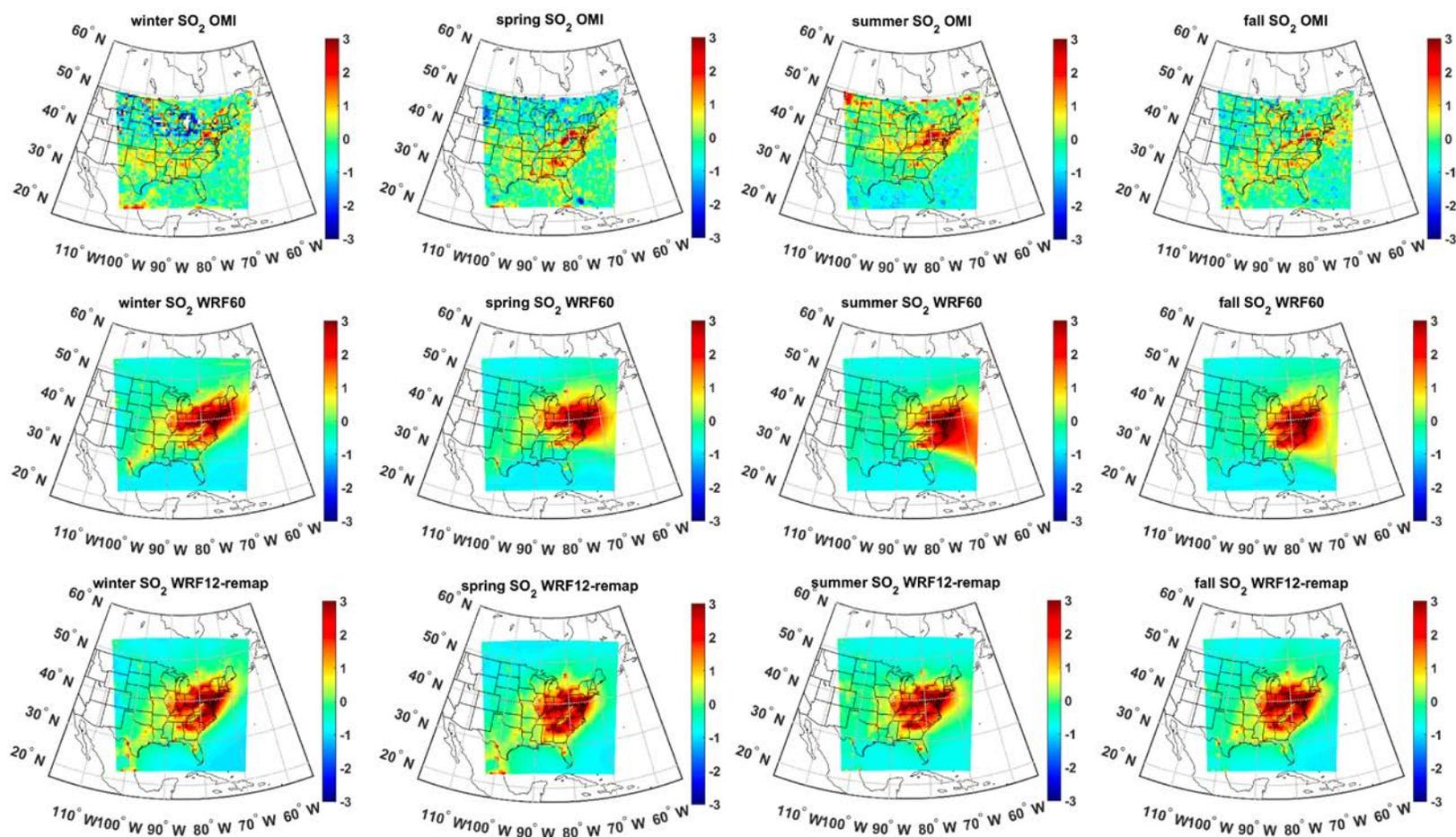
Figure S4. Seasonal total column SO$_2$ z-scores from OMI (first row), WRF60 (second row), and WRF12-remap (third row). z-scores are computed relative to the spatial seasonal mean of each dataset and indicate the distance from the mean in terms of standard deviation units. A cloud screen of 0.3 is applied to both satellite observations and simulated values. Only grid cells with at least 5 valid observations in a month are used to compute a mean value, otherwise the grid cell is shown as white.
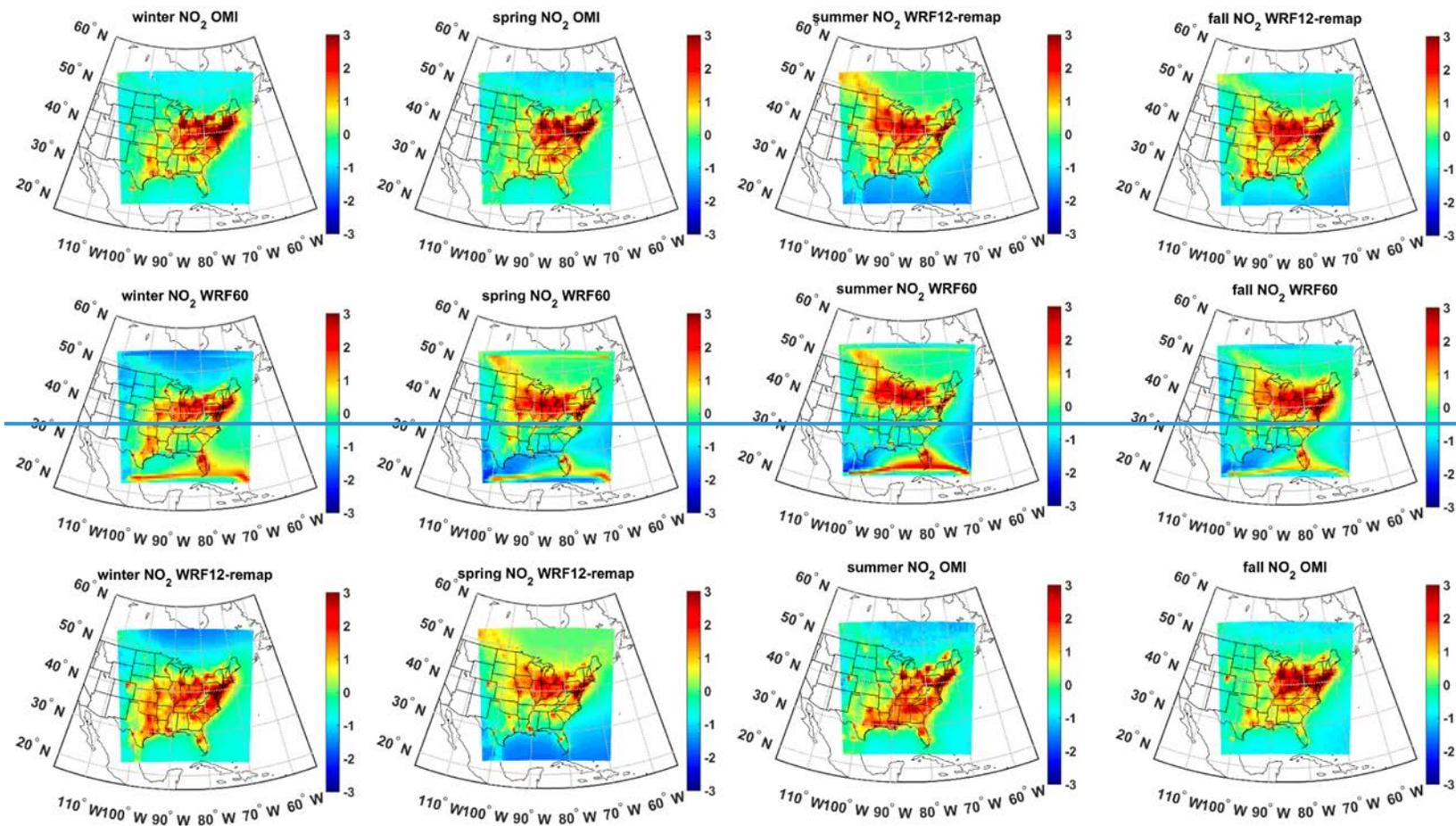
Figure S5. Seasonal total column NO_2 z-scores from OMI (first row), WRF60 (second row), and WRF12-remap (third row). z-scores are computed relative to the spatial seasonal mean of each dataset and indicate the distance from the mean in terms of standard deviation units. A cloud screen of 0.3 is applied to both satellite observations and simulated values. Only grid cells with at least 5 valid observations in a month are used to compute a mean value, otherwise the grid cell is shown as white.
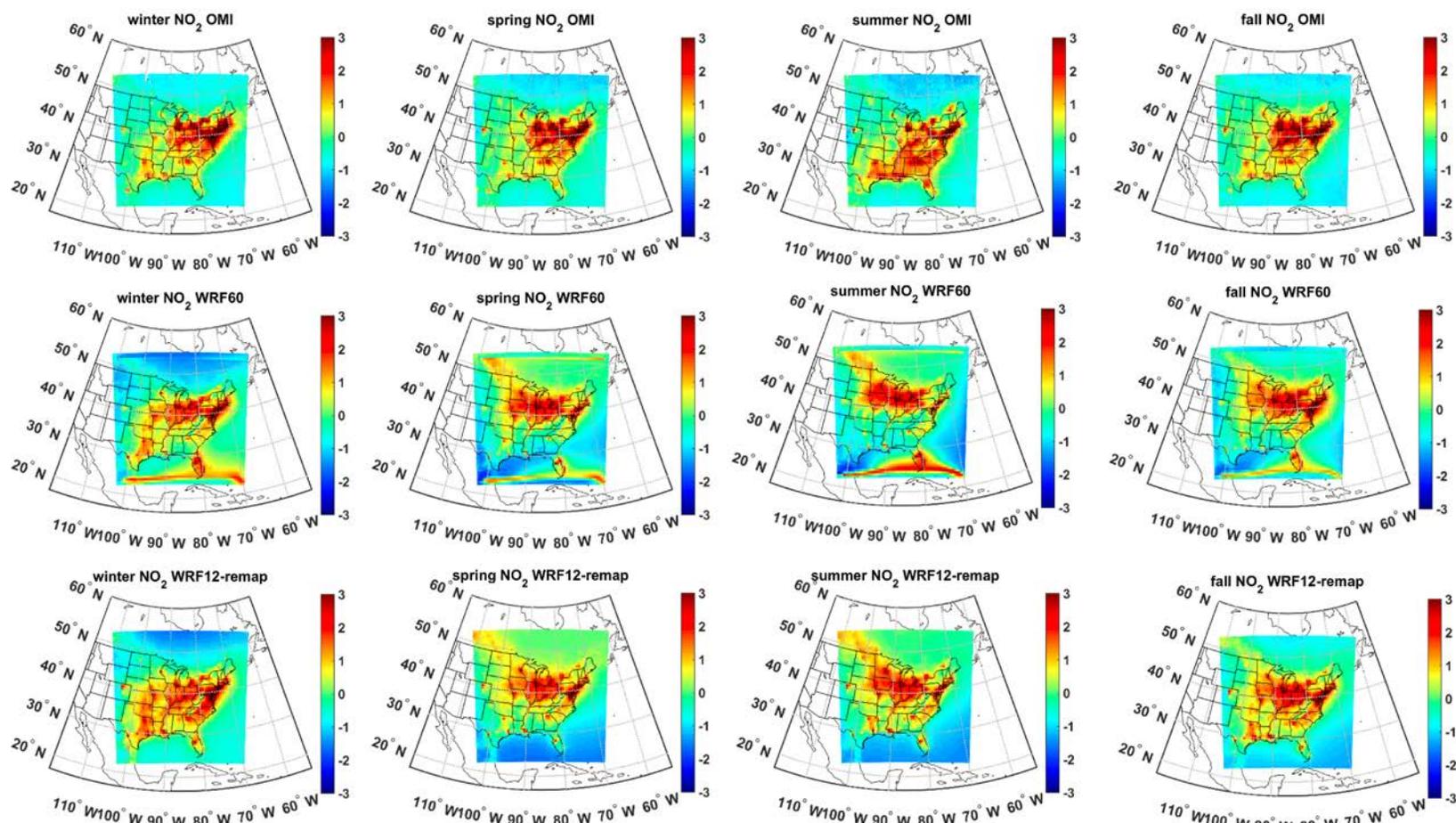
Figure S6. Seasonal total column NH$_3$ z-scores from OMI (first row), WRF60 (second row), and WRF12-remap (third row). z-scores are computed relative to the spatial seasonal mean of each dataset and indicate the distance from the mean in terms of standard deviation units. A cloud screen of 0.3 is applied to both satellite observations and simulated values. Only grid cells with at least 5 valid observations in a month are used to compute a mean value, otherwise the grid cell is shown as white.
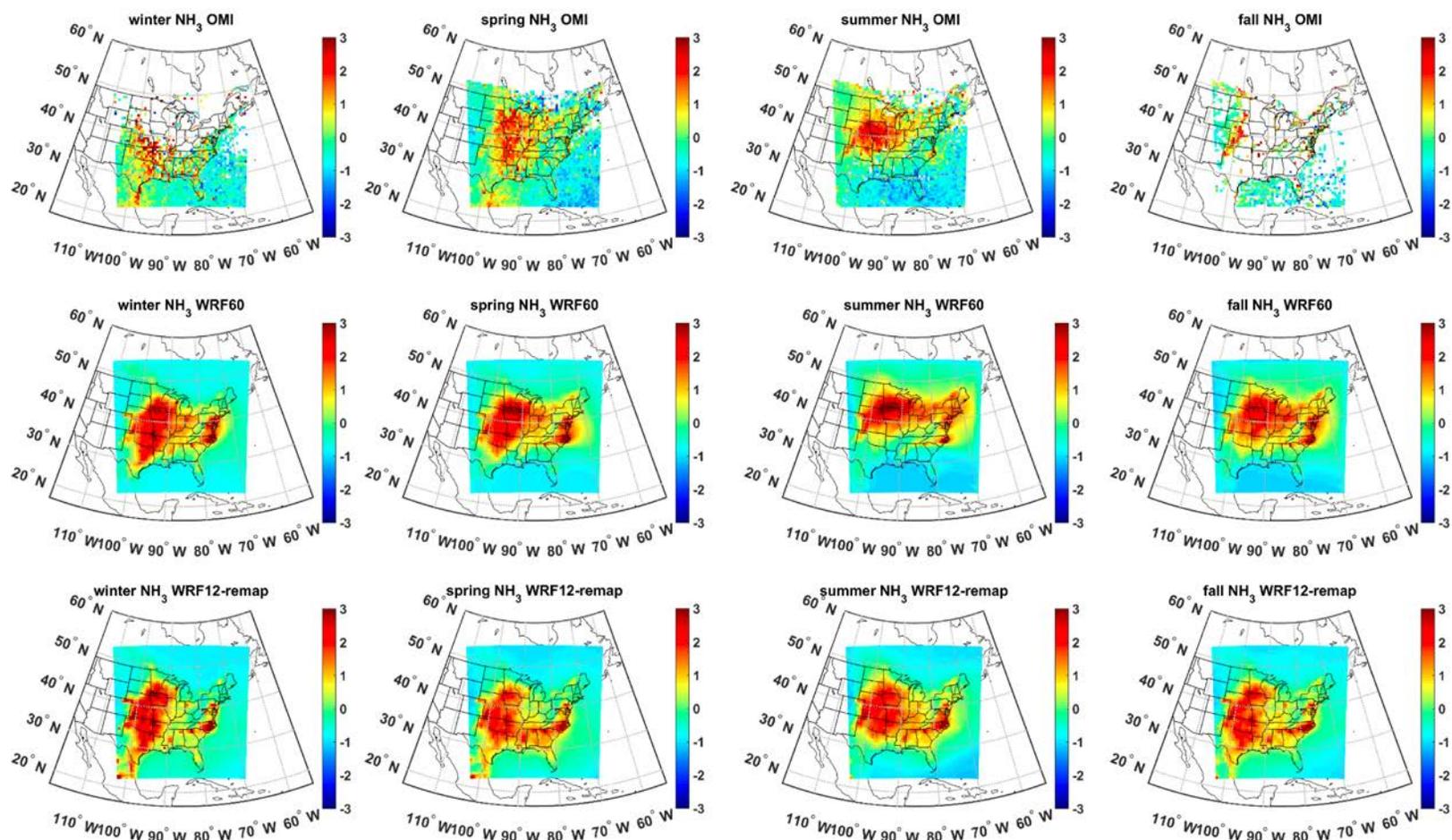
Figure S7. Seasonal total column HCHO z-scores from OMI (first row), WRF60 (second row), and WRF12-remap (third row). z-scores are computed relative to the spatial seasonal mean of each dataset and indicate the distance from the mean in terms of standard deviation units. A cloud screen of 0.3 is applied to both satellite observations and simulated values. Only grid cells with at least 5 valid observations in a month are used to compute a mean value, otherwise the grid cell is shown as white.