Reply to the second round comments from the Reviewer 2

Reviewer Comments:
The authors reacted to all the points raised in the first review report and changed the manuscript accordingly. I am satisfied with their changes. Nevertheless, reading the revised manuscript I would like the authors to add / change some more things:

Reply:

We thank the reviewer for careful review of our revised manuscript and valuable comments to improve the quality of manuscript.

We have carefully addressed all the comments raised by the reviewer to further improve the presentation quality and organization of our paper. Please see below our point-by-point replies. All line numbers correspond to those in the revised version in the track mode.

1) Please add the update frequency of meteorological and chemical boundary conditions. From your replies to the first referee report I understood that the chemical boundary conditions are monthly averages. Nevertheless, I did not find any information about the update frequencies in the paper.

Reply:

The update frequency of meteorological and chemical boundary conditions. This information has actually already been included in the paper in Page 5, lines 112 – 113 as follows: "Considering the decadal applications of WRF/Chem in this work which is much longer than many past WRF/Chem applications, the simulations are reinitialized monthly…"

2) line 202/203: Note: this statement presumes that the cause for the biases stays the same in future.

Reply:

Yes, this is correct. This is an assumption that is made for future year simulations. To address the reviewer's comment, we've explicitly indicated this assumption, see lines 210-212, page 10.

3) The abbreviation NCDC is used in line 215, but introduced in line 339, please explain abbreviations at their first occurance (By the way NMB is used in the abstract without introduction)

Reply:

The introduction for NCDC has been moved to line 231 when it is first mentioned in the main text. NMB in the abstract is introduced in Page 1, line 23.

4) l. 250: NorathEast of what?

Reply:

This has been changed to northeastern U.S., se line 267-268, page 12.

5) I suggest to move the introduction of IOA (currently line 293-301) to Sect. 2.3 where IOA is named for the first time and the rest of evaluation procedure is explained (around line287). As it is, the explanation interrupts the interpretation of T2 behaviour.

Reply:

The introduction of IOA has been moved to Section 2.3, see lines 224-230, pages 10-11

6) l 291/292: I think it is the other way round. The red bars are longer in winter, thus T2 is overpredicted in cooler month.

Reply:

This sentence has been fixed, see lines 308-309, page 14.

7) According to my knowledge the "Code Availability Section" belongs to the main paper, thus the Acknowledgements should be placed behind the Code Availability section.

Reply:

The Acknowledgement section has been placed behind the Code Availability section.

Decadal Evaluation of Regional Climate, Air Quality, and Their Interactions over the Continental U.S. using WRF/Chem Version 3.6.1

Khairunnisa Yahya, Kai Wang, Patrick Campbell, Timothy Glotfelty, Jian He, and Yang Zhang*

Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, U.S.A.

Corresponding author: Yang Zhang, Email: yang_zhang@ncsu.edu

ABSTRACT

The Weather Research and Forecasting model with Chemistry (WRF/Chem) v3.6.1 with the Carbon Bond 2005 (CB05) gas-phase mechanism is evaluated for its first decadal application during 2001 - 2010 using the Representative Concentration Pathway (RCP 8.5) emissions to assess its capability and appropriateness for long-term climatological simulations. The initial and boundary conditions are downscaled from the modified Community Earth System Model/Community Atmosphere Model (CESM/CAM5) v1.2.2. The meteorological initial and boundary conditions are bias-corrected using the National Center for Environmental Protection's Final (FNL) Operational Global Analysis data. Climatological evaluations are carried out for meteorological, chemical, and aerosol-cloud-radiation variables against data from surface networks and satellite retrievals. The model performs very well for the 2-m temperature (T2) for the 10-year period with only a small cold bias of -0.3 °C. Biases in other meteorological variables including relative humidity at 2-m, wind speed at 10-m, and precipitation tend to be site- and season-specific; however, with the exception of T2, consistent annual biases exist for most of the years from 2001 to 2010. Ozone mixing ratios are slightly overpredicted at both urban and rural locations with a~~n~~ Normalized Mean Bias (NMB) of 9.7% but underpredicted at rural locations

with an NMB of -8.8%. $PM_{2.5}$ concentrations are moderately overpredicted with an NMB of 23.3% at rural sites, but slightly underpredicted with an NMB of -10.8% at urban/suburban sites. In general, the model performs relatively well for chemical and meteorological variables, and not as well for aerosol-cloud-radiation variables. Cloud-aerosol variables including aerosol optical depth, cloud water path, cloud optical thickness, and cloud droplet number concentration are generally underpredicted on average across the continental U.S. Overpredictions of several cloud variables over eastern U.S. result in underpredictions of radiation variables (such as GSW with an MB of -5.7 W m^{-2}) and overpredictions of shortwave and longwave cloud forcing (MBs of ~7 to 8 W m^{-2}) which are important climate variables. While the current performance is deemed to be acceptable, improvements to the bias-correction method for CESM downscaling and the model parameterizations of cloud dynamics and thermodynamics, as well as aerosol-cloud interactions can potentially improve model performance for long-term climate simulations.

KEYWORDS: Online-Coupled WRF/Chem; Climate, Air Quality, the Representative Concentration Pathway Scenarios, Climatological Evaluation; Chemistry-Climate Interactions

1. Introduction

Regional atmospheric models have been developed and applied for high resolution climate, meteorology, and air quality modeling in the past few decades. Comparing to global models with a coarser domain resolution (Leung et al., 2003) those regional models have advantages over global models because they can more accurately represent mesoscale variability (Feser et al., 2011), and also better predict the local variability of concentrations of specific species such as black carbon and sulfate (Petikainen et al., 2012). General circulation models (GCMs) and global chemical transport models (GCTMs) are usually downscaled to regional meteorological models such as the Weather Research and Forecasting model (WRF) (Caldwell et al., 2009; Gao et al.,

2012), regional climate models such as REMO-HAM (Petikainen et al., 2012), the regional modeling system known as Providing Regional Climates for Impacts Studies (PRECIS) (Jones et al., 2004; Fan et al., 2014), and a number of European models described in Jacob et al. (2007), as well as regional CTMs such as the Community Multiscale Air Quality Model (CMAQ) (Penrod et al., 2014; Xing et al., 2015). These regional models are used for climate/meteorology or air quality simulations. Some are applied for more than ten years (Caldwell et al., 2009; Warrach-Sagi et al., 2013; Xing et al., 2015). However these regional models either lack the detailed treatment of chemistry (e.g., in WRF), or use prescribed chemical concentrations (e.g., REMO-HAM uses monthly mean oxidant fields for several chemical species), or do not have online-coupled meteorology and chemistry (e.g., in CMAQ). In addition, the past regional model simulations and analyses have mainly focused on meteorological parameters such as surface temperature and precipitation, cloud variables such as net radiative cloud forcing, and chemical constituents such as ozone. Regional climate model simulations tend to focus on significant climatic events such as extreme temperatures (very cold or very hot) (Dasari et al., 2014), heat waves, heavy precipitation, drought, and storms (Beniston et al., 2007), rather than the important air quality and climate interactions. In addition, the impacts of complex chemistry-aerosol-cloud-radiation-climate feedbacks on future climate change remain uncertain, and these feedbacks are most accurately represented using online-coupled meteorology and chemistry models (Zhang, 2010; IPCC, 2013). An online-coupled meteorology and chemistry model, however, is more computationally expensive compared to an offline-coupled model (Grell et al., 2004), and thus requires significant computing resources for their long-term (a decade or longer) applications. With rapid increases in the availability of high performance computing resources on the petaflop scale, however, long term simulations using online-coupled models have become possible in recent years. For example,

recently, the WRF model has been coupled online to the CMAQ model with the inclusion of aerosol indirect effects to study chemistry and climate interactions (Yu et al., 2014).

The online-coupled WRF model with Chemistry (WRF/Chem) has been updated with a suite of physical parameterizations from the Community Atmosphere Model version 5 (CAM5) (Neale et al., 2010) so that the physics in the global CAM5 model is consistent with the regional model for downscaling purposes (Ma et al., 2014). There are also limited applications of dynamical downscaling (Gao et al., 2013) under the new Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report's Representative Concentration Pathway (RCP) scenarios (van Vuuren et al., 2011). Gao et al. (2013) applied dynamic downscaling to link the global-climate-chemistry model CAM-Chem with WRF and CMAQ using RCP 8.5 and RCP 4.5 emissions to study the impacts of climate change and emissions on ozone (O_3). Molders et al. (2014) downscaled the Community Earth System Model (CESM) (Hurrell et al., 2013) to drive the online-coupled WRF/Chem model over Southeast Alaska using RCP 4.5 emissions; however, their study did not address the feedback processes between chemistry and meteorology. This study evaluates the online-coupled regional WRF/Chem model, which takes into account gas and aerosol-phase chemistry, as well as aerosol direct and indirect effects. WRF/Chem is used to simulate the "current" climate scenario for 10 years, from 2001 to 2010 using the RCP 8.5 emissions and boundary conditions from an updated version of CESM with advanced chemistry and aerosol treatments over continental U.S. (CONUS) (He at al., 2015; Glotfelty et al., 2015) with a focus on air-quality and climate interactions. Both CESM and WRF/Chem include similar gas-phase chemistry and aerosol treatments. To our best knowledge, this study is the first to report the WRF/Chem simulation, evaluation, and analyses over a period of 10 years (i.e., 2001-2010) to assess if the model is able to accurately simulate decadal long air quality and climatology by taking

into account feedback processes between chemistry and meteorology. This study also assesses whether the RCP8.5 emissions for the 10-year period are robust enough to produce satisfactory performance against observations with WRF/Chem.

2. Model Set-up and Evaluation Protocol

2.1 Model Configurations and Simulation Design

The model used is the modified WRF/Chem v3.6.1 with updates similar to those implemented into WRF/Chem v3.4.1 as documented in Wang et al. (2014). The main updates include the implementation of an extended version of Carbon Bond 2005 (CB05) (Yarwood et al., 2005) gas-phase mechanism with the chlorine chemistry (Sarwar et al., 2007) and its coupling with the Modal for Aerosol Dynamics in Europe/Volatility Basis Set (MADE/VBS) (Ahmadov et al., 2012). MADE/VBS incorporates a modal aerosol size distribution, and includes an advanced secondary organic aerosol (SOA) treatment based on gas-particle partitioning and gas-phase oxidation in volatility bins. The CB05-MADE/VBS option has also been coupled to existing model treatments of various feedback processes such as the aerosol semi-direct effect on photolysis rates of major gases, and the aerosol indirect effect on cloud droplet number concentration (CDNC) and resulting impacts on shortwave radiation. The main physics and chemistry options used in this study as well as their corresponding references can be found in Table 1. The simulations are performed at a horizontal resolution of 36-km with 148 × 112 horizontal grid cells over the CONUS domain and parts of Canada and Mexico, and a vertical resolution of 34 layers from the surface to 100-hPa. Considering the decadal applications of WRF/Chem in this work which is much longer than many past WRF/Chem applications, the simulations are reinitialized monthly (rather than 1-4 days used in most past WRF/Chem applications to short-term episodes that are on an order of months up to 1-year, e.g., Zhang et al., 2012a, b; Yahya et al., 2014, 2015b) to constrain

meteorological fields toward National Centers for Environmental Prediction (NCEP) reanalysis data while allowing chemistry-meteorology feedbacks within the system. As discussed in Sections 3.1 and 3.3, the reinitialization frequency of 1-month may be too large to constrain some of the meteorological fields such as moistures, which in turn affect other parameters, and a more frequent reinitialization may be needed to improve the model performance. The impact of the frequency of the reinitialization on simulated meteorological and cloud parameters will be further discussed in Sections 3.1 and 3.2. A list of acronyms used in this paper can be found in Table S1.

2.2 Processing of Emissions and Initial Conditions (ICs)/Boundary Conditions (BCs)

Global RCP emissions are available as monthly-average emissions for 2000, 2005, and for every 10 years between 2010 and 2100, at a grid resolution of 0.5°×0.5° (Moss et al., 2010; van Vuuren et al., 2011). The RCP emissions in 2000, 2005, and 2010 are used to cover the 10-year emissions needed for WRF/Chem simulations, i.e., the periods of 2001 – 2003, 2004 – 2006, and 2007 – 2010, respectively. Processing global RCP emissions in 2000, 2005, and 2010 into regional, hourly emissions needed for the 10-year WRF/Chem simulations requires essentially three main tasks. These include 1) mapping the RCP species to CB05 speciation used in WRF/Chem; 2) re-gridding the RCP emissions from 0.5 × 0.5° grid resolution to the 36 × 36 km grid resolution used for regional simulation over North America; and 3) applying species and location dependent temporal allocations (i.e., emissions variation over time) to the re-gridded RCP emissions. Table S2 shows the species mapping between RCP species and CB05 species. To map the RCP species to CB05 speciation, some assumptions are made due the relatively detailed speciation required by CB05. Some of the CB05 species are directly available in RCP; however, others are lumped into RCP groups, for example, the "other alkanals" and "hexanes and higher alkanes" in the RCP groups can be considered to approximately represent the acetaldehyde and higher aldehydes

emissions required by CB05, respectively (Table S2). For the CB05 species such as ethanol, methanol, internal and terminal olefin carbon bonds in the gas-phase, and elemental and organic carbon in the accumulation mode of the aerosol particles, other RCP groups are used to approximate these emissions (Table S2). For the remaining CB05 species that are not available in RCP (i.e. chlorine, HCl, HONO, NH_4^+, NO_3^-, PAR, unspeciated $PM_{2.5}$, H_2SO_4, and SO_4^{2-}), their 2000 emissions are based on the 2002 National Emission Inventory (NEI) (version 3, http://www.epa.gov/ttn/chief/emch/), while their 2005 and 2010 emissions are based on the 2008 NEI-derived emissions ~~2008 NEI~~ (version 2) from the Air Quality Modelling Evaluation International Initiative (AQMEII) project as described in Pouliot et al. (2015), ~~with~~ which include year-specific updates for on/off road transport, wildfires and prescribed fires, and Continuous Emission Monitoring-equipped point sources ~~(Pouliot et al., 2014)~~. To re-grid the RCP emissions, the RCP rectilinear grid is first interpolated to a WRF/Chem curvilinear grid using a simple inverse distance weighting (NCAR Command Language Function – rgrid2rcm), and a subset of the RCP grid that covers the WRF/Chem CONUS domain is then extracted. To derive a temporal allocation for monthly-averaged RCP emissions, hourly emission~~s~~ profiles are taken from those used in-house WRF/Chem simulations over CONUS during 2001 (Yahya et al., 2015a), and 2006 and 2010 as part of the AQMEII project (Yahya et al., 2014, 2015b). ~~For~~ The emissions for those existing in-house simulations were generated based on the 2002 NEI, the emissions were generated with the Sparse Matrix Operator Kernel Emissions (SMOKE) model version 2.3. The emissions for the existing in-house 2006 and 2010 simulations were generated based on the pre-merged emissions provided by the U.S. EPA, which were derived from the 2008 NEI with year-specific section emissions for 2006 and 2010 as part of the AQMEII. ~~for 2002 NEI and~~ SMOKE version 3.4 ~~for 2008 NEI with year-specific sector emissions for 2006 and 2010, which~~was used to prepare

the spatially, temporally, and chemically speciated "model-ready" emissions for the existing in-house 2006 and 2010 WRF/Chem simulations. Since NEI is updated and released every three years, the temporal profiles of emissions used in SMOKE for 2002, 2006 and 2010 are assumed to be valid for 3-4 years around the NEI years, i.e., 2001-2003, 2004-2006, and 2007-2010, respectively. The temporal allocations applied to the RCP emissions are therefore based on the SMOKE model's profiles for each species and source location, and include non-steady-state emissions rates (i.e., seasonal, weekday or weekend, and diurnal variability) that are valid for the entire simulation periods of 2001-2010. Specifically, the hourly re-gridded RCP emission rates for each species *E*, or E_{hr}^{RCP} are calculated by

$$E_{hr}^{RCP}(t,z,\mathrm{lat},\mathrm{lon}) = \mathrm{E}_{mon}^{RCP}(z,lat,lon) * \left[\frac{E_{hr}^{WRF}(t,z,\mathrm{lat},\mathrm{lon})}{E_{mon}^{WRF}(z,\mathrm{lat},\mathrm{lon})} \right] \quad (1)$$

where $E_{mon}^{RCP}, E_{mon}^{WRF}$, and E_{hr}^{WRF} represent the original monthly-averaged RCP emissions rates, the monthly-averaged WRF/Chem emissions rates, and the hourly WRF/Chem emission rates, respectively, which are valid at each model time *t*, layer *z*, and *lat* and *lon* grid points. The RCP elevated source emissions for sulfur dioxide (SO_2), sulfate (SO_4^{2-}), elemental carbon (EC) and organic carbon (OC) were also incorporated into the model-ready emissions for WRF/Chem using steps 1) – 3) and Eq. (1) above. Lastly, RCP aircraft source emissions for EC, nitric oxide (NO), and nitrogen dioxide (NO_2) are directly injected into the closest model layers. No temporal allocations are applied to the RCP aircraft source emissions.

Biogenic emissions are calculated online using the Model of Emissions of Gases and Aerosols from Nature version 2 (MEGAN2) (Guenther et al., 2006). Emissions from dust are based on the online Atmospheric and Environmental Research Inc. and Air Force Weather Agency

183 (AER/AFWA) scheme (Jones and Creighton, 2011). Emissions from sea salt are generated based
184 on the scheme of Gong et al. (1997).

185 The chemical and meteorological ICs/BCs come from the modified CESM/CAM5 version
186 1.2.2 with updates by He et al. (2014) and Glotfelty et al. (2015) developed at the North Carolina
187 State University (CESM_NCSU). WRF/Chem and CESM both use the CB05 gas-phase
188 mechanism (Yarwood et al., 2005), however, WRF/Chem includes additional chlorine chemistry
189 from Sarwar et al. (2007), whereas CESM_NCSU uses a modified version of CB05, the CB05
190 Global Extension (CB05GE) by Karamchandani et al. (2012). In addition to original reactions in
191 CB05 and chlorine chemistry of Sarwar et al. (2007), CB05GE includes chemistry on the lower
192 stratosphere, reactions involving mercury species, and additional heterogeneous reactions on
193 aerosol particles, cloud droplets and on polar stratospheric clouds (PSCs). Both WRF/Chem and
194 CESM_NCSU use a modal aerosol size representation, rather than a sectional size representation.
195 While WRF/Chem includes MADE/VBS with 3 prognostic modes (Ahmadov et al.,
196 2012),CESM_NCSU includes the Modal Aerosol Model with 7 prognostic modes (Liu et al., 2012)
197 is used in CESM_NCSU. In addition to similar gas-phase chemistry and aerosol treatments,
198 CESM_NCSU and WRF/Chem use the same shortwave and longwave radiation schemes (i.e., the
199 Rapid and accurate Radiative Transfer Model for GCM (RRTMG)), though they use different
200 cloud microphysics parameterizations, PBL, and convection schemes. As GCMs generally contain
201 systematic biases which can influence the downscaled simulation, the meteorological ICs/BCs
202 predicted by CESM_NCSU are bias corrected before they are used by WRF/Chem using the
203 simple bias correction technique based on Xu and Yang (2012). Temperature, water vapor,
204 geopotential height, wind, and soil moisture variables available every 6 hours from the NCEP Final
205 Reanalyses (NCEP FNL) dataset are used to correct the ICs and BCs derived based on results from

CESM_NCSU for WRF/Chem simulations. In this bias-correction approach, monthly climatological averages for ICs and BCs are first derived from both NCEP and CESM_NCSU cases. The differences between the ICs and BCs from the NCEP and CESM_NCSU climatological averages are then added onto the CESM_NCSU ICs and BCs to generate bias-corrected CESM_NCSU ICs/BCs. Assuming that the causes for the biases remain the same in future, ~~This~~ this bias correction technique can also be applied to future year simulations ~~where~~ for which NCEP FNL data is not available.

2.3 Model Evaluation Protocol

The focus of the model evaluation is mainly to assess whether the model is able to adequately reproduce the spatial and temporal distributions of key meteorological and chemical variables as compared to observations on a climatological time scale. A scientific question to be addressed in this work is, is WRF/Chem sufficiently good for regional climate and air quality simulations on a decadal scale? A climatological month refers to the average of the month for all the 10 years. For example, January refers to the average for all the months of January from 2001 to 2010. Statistical evaluations such as mean bias (MB), Pearson's correlation coefficient (R), normalized mean bias (NMB), normalized mean error (NME) (The definition of those measures can be found in Yu et al. (2006) and Zhang et al. (2006)) and Index of Agreement (IOA) ranging from 0 to 1 (Willmott et al., 1981) for major chemical and meteorological variables are included. IOA can be calculated as,

$$IOA = 1 - \frac{\sum_{i}^{N} (O_i - S_i)^2}{\sum_{i}^{N} (|O_i - \overline{O}| + |S_i - \overline{S}|)^2} \quad (2)$$

where O_i and S_i denote time-dependent observations and predictions at time and location i, respectively, N is the number of samples (by time and/or location), $\overline{O}$ denotes mean observation and $\overline{S}$ denotes mean predictions over all time and locations, they can be calculated as:

$$\overline{O} = (1/N)\sum_{i=1}^{N} O_i \quad \overline{S} = (1/N)\sum_{i=1}^{N} S_i$$

IOA values range from 0-1, with a value of 1 indicating a perfect agreement.

For surface networks with hourly data, e.g., National Climatic Data Center (NCDC), the observational data are paired up with the simulated data on an hourly basis for each site. The observational data and simulated data are averaged out for each site. The statistics are then calculated based on the site-specific data pairs. The satellite-derived data are usually available on a monthly basis, and the simulated data are also averaged out on a monthly basis. The satellite-derived data are regridded to the same domain and number of grid cells similar to the simulated data. The time dimension is removed for the climatological evaluation, the statistics are based on a site-specific average or a grid cell average. The statistics are then calculated based on the paired satellite-derived vs. simulated grid cell values. The spatial and temporal analyses include spatial plots of MB over CONUS, spatial overlay plots of averaged simulated and observational data, monthly climatologically-averaged time series of major meteorological and chemical variables, annual average time series; probability distributions of major meteorological and chemical variables, and spatial plots of major aerosol and cloud variables compared with satellite data. A summary of the observational data from surface networks and satellite retrievals can be found in Table S3. The variables that are analyzed in this study include O_3, particulate matter with diameter less than and equal to 2.5 and 10 μm ($PM_{2.5}$ and PM_{10}, respectively), and $PM_{2.5}$ species including sulfate (SO_4^{2-}), ammonium (NH_4^+), nitrate (NO_3^-), EC, OC, and total carbon (TC = EC + OC), temperature at 2-m (T2), relative humidity at 2-m (RH2), and wind speed at 10-m (WS10), wind

direction at 10-m (WD10), precipitation, aerosol optical depth (AOD), cloud fraction (CLDFRA), cloud water path (CWP), cloud optical thickness (COT), CDNC, cloud condensation nuclei (CCN), downward shortwave radiation (SWDOWN), net shortwave radiation (GSW), downward longwave radiation (GLW), outgoing longwave radiation at the top of atmosphere (OLR), and shortwave and longwave cloud forcing (SWCF and LWCF). While uncertainties exit in all the observational data used, systematic uncertainty analysis/quantification is beyond the scope of this work. In this work, all observational data are considered to be the true values in calculating the performance statistics. The information on the accuracy of most data used in the model evaluation has been provided in Table 2 of Zhang et al. (2012a). Uncertainties associated with some of the observational data are discussed in Section 3.

3. Model Performance Evaluation

3.1 Meteorological Predictions

Table 2 summarizes the statistics for T2, RH2, WS10, WD10, and precipitation. The model performs very well for a 10-year average T2 with a slight underprediction (an MB of -0.3 °C). This is better or consistent with other studies which tend to report underpredictions in simulated T2. Brunner et al. (2014) reported a range of monthly MBs for T2 of -2 to 1 °C for simulations using a number of CTMs over individual years for 2006 and 2010 with reanalysis meteorological ICs/BCs. Seasonal temperature biases of -1.8 to -2.3 °C were reported from an ensemble of regional climate models (RCMs) for a simulation period of 1971 to 2000 over ~~the N~~northeastern U.S. (Rawlins et al., 2012). He et al. (2015) also showed biases of -3 to 0 °C over CONUS when compared against NCEP reanalysis data. Kim et al. (2013) compared the results of a number of RCMs over CONUS over a climatological period of 1980 to 2003 against Climatic Research Unit (CRU) surface analysis data at a 0.5^{o} resolution and reported T2 biases of -5 to 5 °C. Figure 9.2

from Flato et al. (2013) shows that the Coupled Model Intercomparison Project Phase 5 (CMIP5) models tend to underpredict T2 for the period of 1980 to 2005 over western U.S. by up to -3 °C. The slight bias in T2 can be attributed to errors in soil temperature and soil moisture (Pleim and Gilliam, 2009) or errors in the green vegetation fraction in the National Center for Environmental Prediction, Oregon State University, Air Force and Hydrologic Research Lab (NOAH) Land Surface Model (LSM) (Refslund et al., 2013). RH2 and WS10 are slightly overpredicted. Precipitation is largely overpredicted, consistent with overpredictions in precipitation from WRF and WRF/Chem simulations reported in literatures. For example, Caldwell et al. (2009) attributed the overprediction in precipitation to overprediction in precipitation intensity but underprediction in precipitation frequency. Otte et al. (2012) also reported that the precipitation predicted by WRF is too high compared to the North American Regional Reanalyses (NARR) data throughout the whole CONUS domain over a period of 1988 – 2007. Nudging and reinitialization have been most commonly used methods to control such errors. . Three sensitivity simulations are conducted for a summer month (July 2005) to pinpoint likely causes of the precipitation biases. The baseline simulation (**Base**) uses a monthly reinitialization frequency, CESM_NCSU ICs/BCs, and the Grell 3D cumulus parameterization. The sensitivity simulations include (1) **Sen1**, which is similar to the Base case except with a 5-day reinitialization period; (2) **Sen2**, which is similar to Base except using NCEP for the meteorological ICs/BCs; and (3) **Sen3**, which is similar to Base except using WRF/Chem v3.7 with the Multi-Scale Kain Fritsch (MSKF) cumulus parameterization, instead of Grell 3D. The differences in configuration setup in those sensitivity simulations are given in Table S4. The evaluation and comparison of the baseline and sensitivity results in July 2005 are summarized in Tables S5 and S6, and Figure S1 in the supplementary material. As shown in Tables S5-S6 and Figure S1, the precipitation bias can be attributed to several factors including the use of

Grell 3D cumulus parameterization scheme, the use of bias-corrected CESM_NCSU data (instead of NCEP reanalysis data), and the use of an reinitialization frequency of 1-month, among which the first factor dominates the biases in precipitation predictions. The simulated precipitation is very sensitivity to different cumulus parameterizations. Compared to scale-aware parameterizations such as the multi-scale Kain-Fritsch (MSKF) cumulus scheme, the Grell 3D parameterization has a tendency to overpredict precipitation, particularly over ocean.

Figure 1 shows the spatial distributions of MB for 10-year average predictions of T2, RH2, WS10, and precipitation. Figure 2 shows the time series of 10-year average monthly and annual average T2, WS10, RH2, precipitation, O_3, and $PM_{2.5}$ against observational data and IOA statistics. T2 (Figure 1a) tends to be underpredicted over eastern and western U.S. and overpredicted over the central U.S. The bias correction method itself may also contribute to the slight biases in T2. A single temporally averaged (2001 – 2010) NCEP reanalysis file is applied to the 6-hourly BCs for each individual year, which would in some cases contribute to the biases in the climatological 10-year evaluation. T2 also tends to be ~~underpredicted~~ overpredicted during the cooler months but ~~overpredicted~~ underpredicted during the warmer months (Figure 2a). While the bar charts in Figure 2 show domain- average mean observed and mean simulated T2, IOA performance takes into account the proportion of differences between mean observed and mean simulated values at different sites. ~~IOA can be calculated as,~~

$$IOA = 1 - \frac{\sum_{i}^{N}(O_i - S_i)^2}{\sum_{i}^{N}(|O_i - \overline{O}| + |S_i - \overline{S}|)^2} \quad \text{~~(2)~~}$$

where O_i and S_i denote time-dependent observations and predictions at time and location i, respectively, N is the number of samples (by time and/or location), $\overline{O}$ denotes mean observation and $\overline{S}$ denotes mean predictions over all time and locations, they can be calculated as:

$$\overline{O} = (1/N)\sum_{i=1}^{N} O_i \text{ , } \overline{S} = (1/N)\sum_{i=1}^{N} S_i \text{ ,}$$

IOA values range from 0-1, with a value of 1 indicating a perfect agreement.

The model performance in terms of IOA for T2 is slightly worse during the warmer months as compared to the cooler months; however, IOA values for all months are ≥ 0.9. The poorer IOA statistics for the warmer months are possibly influenced to a certain extent by the fact that the IOA tends to be more sensitive towards extreme values (when temperatures are maximum) due to the squared differences used in calculating IOA (Legates and McCabe, 1999). As shown in Figures 1b and 2b, the spatial distributions of MBs for RH2 follow closely the spatial distributions of MBs for T2, where T2 is underpredicted, RH2 is overpredicted and vice versa. Unlike T2, the IOA for RH2 is the highest during the warmer months and the lowest during the winter months, but IOA for RH2 is generally high (> 0.7) for all months. WS10 is also generally overpredicted along the coast, over eastern U.S. and some portions over the western U.S. (Figure 1c), consistent with overpredictions of T2 over the coast, and partially due to unresolved topographical features. In this case the topographic correction for surface winds used to represent extra drag from sub-grid topography (Jimenez and Dudhia, 2012) is used as an option in the 10-yr WRF/Chem simulations; however, WS10 is still overpredicted except for the areas of flat undulating land in the central U.S. Jimenez and Dudhia (2012) also suggested that the grid points nearest to the observational data might not be the most appropriate or most representative, and that the selection of nearby grid points can help to reduce errors in surface wind speed estimations. In this study, as the evaluation

is conducted over the whole CONUS, the nearest grid points are used for evaluation, which could also result in errors in wind speed evaluation. The positive T2 and WS10 bias along the coast could be due to the fact that the model grids for temperatures and wind speeds are located over the ocean, however, the observation points are located slightly inland. As shown in Figure 2, WS10 performs well on average for the months of April, May, and June, and is overpredicted for the other months. Nonetheless the climatological NMB for WS10 overall is low at 7.7% (Table 2). WS10 has higher IOA values during the spring months and the lowest IOA during the summer months and in November. The model performs relatively well in predicting WD10 variability with a Corr of 0.6, indicating overall a more southerly direction domain-wide predicted by the model compared to observations. Precipitation is overpredicted for all months except for June, especially during the summer months of July to August. Even with the inclusion of radiative feedback effects from the subgrid-scale clouds in the radiation calculations, precipitation is still overpredicted with the Grell 3D scheme, which is consistent with the results shown by Alapaty et al. (2012). Precipitation mainly has lower IOAs during the summer compared to other months, except in June which actually exhibits the largest IOA of all months. Even though June is considered a summer month, it does not show overprediction in precipitation compared to the other summer months. It is possible that in June, the overall atmospheric moisture content is low. This is consistent with simulated RH2 as June is the only month where RH2 is underpredicted compared to observations.

In general the model is able to reproduce the monthly trends in meteorological variables; for example, the predicted trend in T2 closely follows the observed trends by NCDC. The observed RH2 decreases from January to a minimum in April, and then increases from April to December. Although the model predicts a similar pattern in RH2, there is a lag in the RH2 minimum occurring two months later in June (Figure 2b). For WS10, the observation peaks in April, as compared to

359 the simulated peak in March. The model correctly predicts the observed WS10 minimum occurring
360 in August. The model trend in precipitation is similar to observations, except during the summer
361 months of July through September, where a large overprediction leads to a sharp increase in July,
362 followed by a gradual decrease through December.

363 Figures 2e – 2h show the annual time series trends for T2, RH2, WS10, and precipitation.
364 The model performs relatively well in predicting the annual mean T2 for most years (with MBs of
365 < 0.5 °C; Figure 2e). T2 also does not show an obvious decreasing or increasing T2 trend between
366 2001 and 2010. The IOA for annual T2 for all years are > 0.95. However for 2002, mean simulated
367 T2 is ~0.7 °C higher than the observational data. IOA is still high for 2002 which indicates
368 probably good performance of T2 at most sites, however with large overpredictions at a few sites
369 which could skew the mean observed and mean simulated value but not influence IOA
370 significantly. RH2 is consistently overpredicted by the model with the largest overprediction in
371 2009. With the exception of 2009, observed RH2 is rather steady (65 – 70 %) from 2001 to 2010.
372 IOA is also steady for RH2, except for 2009. As mentioned earlier, WRF tends to overpredict
373 WS10 in general. Figure 2g shows that observations indicate weaker wind speeds from 2001 to
374 2007. Model performance is better from 2007 to 2010 with higher IOAs compared to previous
375 years. WRF has worse performance especially at weaker wind speeds as is the case from 2001 to
376 2007. Model performance for precipitation is more variable year-to-year, with IOAs ranging from
377 0.4 to 0.7; however, there is a systematic positive bias during the 10 year period.

378 Figure 3 shows the probability distributions of T2, RH2, WS10, and precipitation against
379 NCDC and NADP for 10 years. The observed and simulated variables are averaged at each site
380 for the 10-year period, and the pairs are then distributed into a probability distribution over 30 bins
381 of observed and simulated values of T2. For T2, the simulated and observed probability

distributions are very similar (Figure 3a), consistent with the statistics for T2 which shows only a small cold bias. The model overpredicts T2 at sites where temperatures are very low. The probability distribution curve for simulated RH2 is also shifted to the right of the observed RH2 (Figure 3b), with an observed and modeled peak 74% and 78% respectively. The probability distribution of simulated WS10 is narrower (between 2 and 6 m s^{-1}) compared to that of observed WS10 (between 1 and 7 m s^{-1}). The model thus overpredicts when near-surface wind speeds are low, but underpredicts when wind speeds are very high. This suggests that the surface drag parameterization is still insufficient to help predict low wind speeds; however, it might have contributed to the reduction in the simulated high wind speeds (Mass, 2012). The probability distribution for simulated precipitation against NADP also shows a shift to the right, consistent with the statistics for overpredicted precipitation and also with the probability curve of RH2. Nasrollahi et al. (2012) examined 20 combinations of microphysics and cumulus parameterization schemes available in WRF and found that most parameterization schemes overestimate the amount of rainfall and the extent of high rainfall values. In this study, while Grell 3D Ensemble cumulus parameterization contributes in part to the overpredictions of precipitation, most overpredictions occur at high thresholds as shown in Figure 3 (d) and they are attributed to possible errors in the Morrison two moment scheme because the overpredictions of non-convective precipitation dominate the overpredictions of total precipitation.

3.2 Chemical Predictions

3.2.1 Ozone

Table 2 summarizes the statistics for major chemical species. The model overpredicts hourly O_3 mixing ratios on average against the Aerometric Information Retrieval System (AIRS) – Air Quality System (AQS) with an NMB of 9.7% and an NME of 22.4%, but underpredicts O_3

mixing ratios against the Clean Air Status and Trends Network (CASTNET) with an NMB of -8.8% and an NME of 19.8%. The O_3 mixing ratios are overpredicted at AIRS-AQS sites for all climatological months except for April and May (Figure 4a) but underpredicted at CASTNET sites for all months except for October with the largest underpredictions occurring in April and May where IOA statistics are the lowest (Figure 4b). IOA statistics for all climatological months range from 0.5 to 0.6 for AIRS-AQS and from 0.4 to 0.9 for CASTNET. In general, IOA values tend to be higher for CASTNET compared to AIRS-AQS during the fall and winter months of October to March. The IOA values for AIRS-AQS are rather steady on average over the 12 months compared to CASTNET. This can be attributed to the larger dataset of AIRS-AQS (> 1000 stations) compared to CASTNET (< 100 stations), the high and low undulations in O_3 averages at the CASTNET sites tend to be smoothed or averaged out in O_3 averages at the AIRS-AQS sites given larger AIRS-AQS dataset. The observed data from AIRS-AQS and CASTNET also show the highest monthly O_3 mixing ratios over April and May. This result is consistent with the findings of Cooper et al. (2014), who reported the highest mass of tropospheric O_3 for the northern hemisphere in April and May based on the Ozone Monitoring Instrument (OMI) measurements in 2004, which suggested that the column mass of O_3 is not necessarily proportional to nitrogen oxide (NO_x) emissions that peak during the summer. In addition, Cooper et al. (2014) attributed a shift in the seasonal O_3 cycle observed at many rural mid-latitude monitoring sites to emissions reductions in the U.S. The same study also reported that the summertime O_3 mixing ratios were lower in eastern U.S. between 2005 and 2010 when compared to previous years, while remaining relatively constant in spring. Thus the summer O_3 maximum during 2001- 2004 was replaced by a broad spring/summer peak in 2005 - 2010. Both the observed and simulated O_3 mixing ratios do not decrease for AIRS-AQS and CASTNET from 2001 to 2010 (Figures 4e and 4f). This is

somewhat consistent with Cooper et al. (2014) which showed that surface and lower tropospheric O_3 has a decreasing trend over eastern U.S. but an increasing trend over the western U.S. from 1990-1999 to 2010. The predicted annual average O_3 mixing ratios are consistent from 2001 to 2010, with overpredictions and IOAs of ~0.6 at the AIRS-AQS sites, and underpredictions and IOAs of ~0.6 to 0.8 at the CASTNET sites.

Figure 5 shows the probability distributions of maximum 1-hour and 8-hour O_3 mixing ratios against CASTNET and AIRS-AQS. The probability distributions of the observed and simulated O_3 mixing ratios are very similar. The model is able to simulate the range and probabilities of O_3 mixing ratios relatively well at both CASTNET and AIRS-AQS sites. At the CASTNET sites as shown in Figures 5a and b, the model accurately predicts the peak maximum 1-hour O_3 mixing ratio centered at ~60 ppb, however, slightly underpredicts the peak maximum 8-hour O_3 mixing ratio by a few ppb. At the AIRS-AQS sites as shown in Figures 5c and d, the predicted probability distribution curve is slightly shifted to the right of the observations for both maximum 1-hour and 8-hour O_3 mixing ratios. It is also interesting to note that the probability distributions for CASTNET and AIRS-AQS are quite different. O_3 at the AIRS-AQS sites has a unimodal normal distribution, while O_3 at the CASTNET sites has a bi-modal distribution, with a tail of the distribution extending toward lower O_3 mixing ratios (0 – 20 ppb). The peak distribution occurs at around 10 ppb, because the O_3 mixing rations are low at most CASTNET sites. The second peak at ~60 ppb for CASTNET occurs mainly around the summer months during which O_3 is produced through photochemistry involving its precursors. These distributions are attributed to the nature of the sites' locations, where the AIRS-AQS network includes a mixture of urban, suburban and rural sites, leading to a normal distribution of O_3 mixing ratios centered at relatively higher O_3 mixing ratios, while the CASTNET network includes mostly rural sites that exhibit a

low maximum 1-hour and 8-hour O_3 mixing ratios, thus leading to a distribution with a tail skewed towards the lower O_3 mixing ratios.

Figure 6 shows the diurnal variation of O_3 concentrations and IOA statistics for the four climatological seasons against CASTNET (Figures a to d) and AIRS-AQS (Figures e to h) (Winter - January, February and December (JFD); Spring - March, April, and May (MAM); Summer - June, July, and August (JJA); Fall - September, October, and November (SON). Figure 6a shows that in more rural sites (CASTNET) in winter O_3 tends to be underpredicted during the morning (01:00 – 09:00 local standard time (LST)) and evening hours (18:00 – 24:00 LST). However, Figure 6b shows that in general for all AIRS-AQS sites including urban sites, O_3 is systematically overpredicted for all hours of the day. The diurnal trends for CASTNET and AIRS-AQS are completely opposite for winter. As CASTNET sites are located in areas where urban influences are minimal, most of these sites are likely to be NO_x-limited sites (Campbell et al., 2014). Underpredicted NO_x emissions in rural areas can lead to underpredictions in O_3 concentrations in NO_x-limited areas. As shown in Figure 2a), T2 is generally overpredicted during the winter months, which explains the overpredictions in O_3 for most sites against AIRS-AQS. As shown in Figures 6a, b and c, for CASTNET, the diurnal variations of O_3 in MAM and JJA are similar to that in JFD. As shown in Figure 6d, slight overpredictions during the daylight hours of 10:00 to 17:00 LST occur in SON at the CASTNET sites, however the trends are similar for morning and evening hours as compared to the other seasons. Similar to SON at the CASTNET sites, for AIRS-AQS sites, overpredictions during daylight hours occur in JJA and SON (Figures 6 g and h), and also to a much lesser extent in MAM (Figure 6f). This is probably due to the overpredictions of T2, which are the smallest during MAM compared to other months as shown in Figure 2a.

Figure 7 compares the spatial distributions of 10-year average of the predicted and observed hourly O_3 mixing ratios. The O_3 mixing ratios tend to be underpredicted in eastern and northeastern U.S., where most of the CASTNET sites are located (Figure 7a). This is consistent with the diurnal trends from Figures 6a to d which also show underpredictions for CASTNET sites. From Figure 1a, T2 is underpredicted on average over northeastern U.S., which results in underpredictions in biogenic emissions in the rural areas from MEGAN2. This would in turn reduce O_3 mixing ratios in VOC-limited areas. O_3 photochemical reactivities would also be reduced due to reduced T2. O_3 mixing ratios are, however, overpredicted over northwestern U.S., and also near the coastline of western U.S. The overprediction of O_3 mixing ratios in northwestern U.S. can be attributed to an overprediction in the chemical BCs from CESM, as indicated by the high O_3 mixing ratios near the northwestern region of the domain boundary.

3.2.2 Particulate Matter

The 10-year average $PM_{2.5}$ concentrations are overpredicted with an NMB of 23.3 % against IMPROVE, and underpredicted with an NMB of -10.8 % against the Speciated Trends Network (STN) (Table 2). In addition, the IOA trend in Figure 4c shows very good performance for $PM_{2.5}$ against the Interagency Monitoring of Protected Visual Environments (IMPROVE) with IOA values > 0.8. IOA values for $PM_{2.5}$ against STN are high (~ 0.6 – 0.8) during the spring and summer months, but lower (~ 0.4) during the winter months (Figure 4d). The IMPROVE surface network covers generally rural areas and national parks while the STN surface network covers urban sites. The horizontal resolution of 36×36 km^2 used in this study may be too coarse to resolve the locally high $PM_{2.5}$ concentrations at urban sites in STN which are in proximity of significant point sources, especially during the fall and winter. During these colder seasons, $PM_{2.5}$ concentrations over the U.S. in general tend to be higher due to an extensive use of woodstove and

cold temperature inversions, which trap particulates near the ground (EPA, 2011). As shown in Table 2, the concentrations of $PM_{2.5}$ species such as SO_4^{2-}, OC, and TC are overpredicted at the IMPROVE sites, while the concentrations of the other main $PM_{2.5}$ species NO_3^-, NH_4^+, and EC are underpredicted at both IMPROVE and STN sites. TC concentrations, which are the sum of OC and EC, are overpredicted due to larger overpredictions of OC compared to the underpredictions of EC. The model also simulates both primary organic aerosol (POA) and secondary organic aerosol (SOA). OC is calculated as the sum of POA and SOA divided by the ratio of OA/OC, which is assumed to be a constant of 1.4 (Aitken et al., 2008). This calculation of OC using a constant of 1.4 is an approximation, which is subject to uncertainties when comparing simulated OC against observational data, as the ratio of OA/OC can be different in different environments (Aitken et al., 2008).

As shown in Table 2, at the STN sites, the model slightly overpredicts the concentrations of SO_4^{2-}, while underpredicting those of NO_3^-, NH_4^+, and EC. The overpredictions of SO_4^{2-} are likely due to the uncertainties that arise from processing of the RCP SO_2 emissions. The RCP SO_2 emissions are only available as a total emission flux, and they are not vertically distributed to the important point sources such as furnaces and stacks. In this work, two steps are taken to resolve the RCP elevated SO_2 emissions in each emission layer. First, a set of factors are derived from the fraction of the elevated emissions in each layer to the vertical sum of emissions for NEI used by default in the SMOKE model with the NEI data. Second, these factors are applied to the total RCP emissions to obtain SO_2 emissions in each emission layer. The total RCP SO_2 emissions were higher than the total NEI emissions, resulting in higher surface and elevated SO_2 emissions. Figures 4g and 4h compare the modeled annual average time series for $PM_{2.5}$ against IMPROVE and STN observations, respectively. In general, the model performs well for $PM_{2.5}$ at the

519 IMPROVE (IOA > 0.8) and STN (IOA ~ 0.5 – 0.7) sites. A declining trend in $PM_{2.5}$ observed and
520 simulated concentrations are also observed over the years. For the later years (2007 to 2010), the
521 model performs significantly better against IMPROVE compared to STN. As 2010 NEI emissions
522 are used for the years 2007 to 2010, there are not many variations in the simulated $PM_{2.5}$
523 concentrations over these 4 years.

524 Figures 7 and 8 show the spatial plots of 10-yr average of simulated 24-hour average ,
525 PM_{10}, $PM_{2.5,}$ and $PM_{2.5}$ species concentrations, overlaid with observations from both STN and
526 IMPROVE. The underpredictions of PM_{10} are dominated by an underprediction in the wind-blown
527 dust emissions, especially in western U.S. (Figure 7b). This is confirmed in Table 2, which shows
528 an MB of -11.5 $\mu g\ m^{-3}$ and an NMB of -51.2% against PM_{10} observations at AIRS-AQS sites. The
529 observational data indicate the elevated concentrations of dust over portions of Arizona and
530 California (> 50 $\mu g\ m^{-3}$), which are not reproduced by the simulations (the simulated
531 concentrations are much lower, < 20 $\mu g\ m^{-3}$). The AER/AFWA dust module (Table 1) does not
532 produce sufficient dust in this case, even though WS10 is overpredicted and is proportional to the
533 dust emissions. The sea-salt emission module by Gong et al. (1997), however, seems to produce a
534 reasonable amount of sea-salt as shown by the similar concentrations between simulated and
535 observational data for PM_{10} near the coastlines. In addition, the MADE/VBS module in
536 WRF/Chem does not explicitly simulate the formation/volatilization of coarse inorganic species.
537 The coarse inorganic species are available, however, in the emissions and are transported and
538 deposited in a manner that is similar to non-reactive tracers.

539 The model performs well for $PM_{2.5}$ over eastern U.S. (Figure 7c), where modeled
540 concentrations are close to the observations; however, over the western U.S. there are
541 underpredictions in $PM_{2.5}$, especially in central to southern California. Even though Table 2 shows

542 in general an overprediction of SO_4^{2-} against STN sites, the model underpredicts SO_4^{2-} in regions
543 of elevated SO_4^{2-} concentrations, in particular, where concentrations are above 10 μg m^{-3} in the
544 vicinity of significant point sources of SO_2 and SO_4^{2-} over eastern U.S. (Figure 7d). This is likely
545 due to the coarse resolution ($0.5^o \times 0.5^o$) of RCP emissions, which probably results in a general
546 overprediction of SO_2 emissions over a grid but cannot resolve point sources smaller than the grid
547 resolution. A similar pattern is found for NH_4^+ over eastern U.S. due to underpredictions of high
548 concentrations of SO_4^{2-} (Figure 8a). There are also large underpredictions in NH_4^+ over the western
549 U.S. The underpredictions in NH_4^+ are likely due to underpredictions of NH_3 emissions from RCP.
550 The NH_3 emissions from RCP are much lower than those of NEI emissions over western U.S., by
551 more than a factor of 5, especially over portions of California. Large underpredictions occur over
552 both eastern and western U.S. for NO_3^-, EC, and TC (Figures 8b, c, and d). The underpredictions
553 in NO_3^- are more likely influenced by the underpredictions of NH_4^+ rather than NO_x emissions.
554 NO_x emissions for NEI are higher than those of RCP for a number of point sources, however, in
555 general RCP has higher NO_x emissions. Other possible reasons for the underpredictions of NO_3^-
556 concentrations include both prediction and measurement errors associated with SO_4^{2-} and TNH_4
557 that can greatly affect the performance of NO_3^-, inaccuracies in the assumptions used in the
558 thermodynamic model (e.g., the assumption that inorganic ions are internally mixed and the
559 equilibrium assumption might not be representative, especially for particles with larger diameters),
560 as well as inaccuracies in T2 and RH predictions (Yu et al., 2005). The statistics for IMPROVE
561 TC indicate overpredictions; however the statistics for STN TC indicate larger underpredictions
562 with an MB of -2.0 μg m^{-3}, which would explain the large underpredictions in $PM_{2.5}$ concentrations
563 over western U.S. The large underpredictions are in part impacted by uncertainties in emissions as
564 well as due to uncertainties in the precursor gas emissions for these species, especially for TC. The

RCP emissions of EC and POA are lower when compared to those of NEI. NEI emissions have a higher spatial resolution, and thus more adequately represent the emissions from point sources compared to RCP. The underpredictions of TC are also more likely due to underpredictions in EC as compared to OC, as shown in underpredictions of EC by Figure 8c. As T2 is slightly underpredicted, these could have resulted in underpredictions in isoprene and terpene, which are major gas precursors of biogenic SOA, resulting in lower SOA and OC concentrations. In addition, the emissions of anthropogenic VOC species from RCP which are also of a lower spatial resolution compared to their emissions in the NEI tend to also be lower than NEI levels especially at point sources. The underpredictions for these particulate species, especially for water-soluble species including NH_4^+ and NO_3^- are also likely impacted by overpredictions in precipitation (Figure 2d), which leads to an overprediction in their wet deposition rates and thus a reduction of their ambient concentrations. The overpredictions in WS10 also help contribute to the deposition of $PM_{2.5}$ and $PM_{2.5}$ species onto the ground (Sievering et al., 1987).

3.3 Aerosol, Cloud, and Radiation Predictions

There are uncertainties in the satellite retrievals of various aerosol-cloud-radiation variables from the Clouds and the Earth's Radiant Energy System (CERES) and the Moderate Resolution Imaging Spectroradiometer (MODIS). Loeb et al. (2009) reported that the major uncertainties of the top of atmosphere radiative fluxes from CERES are derived from instrument calibration (with a net error of 4.2 W m^{-2}), and the assumed value of 1 W m^{-2} for total solar irradiance. However, there is good correlation ($R > 0.8$) between the model and CERES for the radiation variables SWDOWN, GSW, and GLW, which are all measured at the surface (Table 2). Modeled OLR at the top of the atmosphere also has relatively good correlation ($R \sim 0.6$).

SWDOWN and GLW are both slightly overpredicted due to influences from biases in PM concentrations and clouds, but GSW and OLR are slightly underpredicted.

The overpredictions of the surface radiation variables are also impacted by the underpredictions in AOD and COT. AOD is underpredicted with an NMB of -24.0%, and COT is underpredicted with an NMB of -44.3%. These underpredictions indicate that less radiation is attenuated (i.e., absorbed or scattered) or reflected while traversing through the atmospheric column and clouds, thus allowing more radiation to reach the ground. Using the CESM model, He et al. (2015) also showed underpredictions in AOD and COT over CONUS against MODIS satellite retrievals. Figure 9 compares the spatial distributions of the 10-year average predictions of AOD (a and b) against the satellite retrieval data from MODIS. The simulated AODs show relatively large values over eastern U.S., due to the relatively higher PM concentrations in this region of the U.S. The MODIS AOD, however, shows slightly elevated values over eastern U.S., but the magnitudes are not as high as the simulated AOD over eastern U.S. MODIS-derived AOD is also higher over western U.S. compared to eastern U.S., and this trend is not found in the simulated AOD. The differences between the MODIS AOD and the simulated AOD are likely due to the differences in the algorithms used to retrieve AOD based on MODIS measurements and calculate AOD in WRF/Chem. For MODIS, AOD is calculated by matching the spectral reflectance observations with a lookup table based on a set of aerosol parameters including the aerosol size distributions from a variety of aerosol models, which differ based on seasons and locations (Levy et al., 2007). There are also different algorithms for dark land, bright land, and over oceans (Levy et al., 2013). The MODIS data are aggregated into a global 1^{o} gridded (Level-3) dataset with monthly (MOD08_M3) temporal resolution (https://www.earthsystemcog.org/site_media/projects/obs4mips/TechNote_MODIS_L3_C5_Aer

610 osols.pdf). The inaccuracies for the calculation of AOD in WRF/Chem include biases in aerosol
611 size distribution, aerosol composition, aerosol water content, and reflectances. They can also arise
612 from parameterizations in the calculations including the assumption of an internally-mixed aerosol
613 composition. Therefore, caution should also be taken when comparing simulated AOD with the
614 satellite-derived AOD products. Toth et al. (2013) compared Aqua MODIS AOD products over
615 the mid to high latitude Southern Ocean where a band of enhanced AOD is observed, to cloud and
616 aerosol products produced by the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP)
617 project; and AOD data from the Aerosol Robotic Network (AERONET) and the Maritime Aerosol
618 Network (MAN). They concluded that the band of enhanced AOD is not detected in the CALIOP,
619 AERONET, or MAN products. The enhanced AOD band is attributed to stratocumulus and low
620 broken cumulus cloud contamination, as well as the misidentification of relatively warm cloud
621 tops compared with surrounding open seas.

622 Figure 9 also shows spatial distributions of the 10-year average predictions of CDNC (c
623 and d), CWP (e and f), and COT (g and h), compared against the satellite retrieval data from
624 MODIS. The cloud variables CDNC, CWP, and COT tend to be underpredicted for most of the
625 regions over the U.S. However, CWP is largely overpredicted over the Atlantic ocean. This is also
626 likely due to the build-up of moisture over the Atlantic ocean, also influencing precipitation as
627 mentioned previously. CDNC is overpredicted over some regions in eastern U.S., but there are
628 also relatively large areas of underpredictions over both the land and ocean. This leads to an
629 average domain-wide underprediction for CDNC (Table 2). This is likely due to the differences in
630 deriving CDNC in the model and in the satellite retrievals. CDNC in the model is calculated based
631 on the activation parameterization by Abdul Razzak and Ghan (2000) based on the aerosol size
632 distribution, aerosol composition, and the updraft velocity. The MODIS-derived CDNC from

633 Bennartz (2007) is calculated based on cloud effective radius and COT, which would explain the
634 differences in spatial patterns between model and observed data. As indicated by Bennartz (2007),
635 the errors in CDNC can be up to 260%, especially for regions with low CF (< 0.1). The model and
636 MODIS spatial patterns are similar for CWP and COT over land, although the model values are
637 underpredicted. King et al. (2013) reported that the MODIS retrieval of cloud effective radius
638 when compared to in-situ observations is overestimated by 13% on average. Combined with
639 overestimations in COT, this leads to overestimation of liquid water path. In addition, there can
640 also be differences in satellite-derived cloud products from different satellites. For example, Shan
641 et al. (2011) showed that the derived CLDFRA from MODIS and another satellite, the Polarization
642 and Directionality of Earth Reflectances (POLDER) can differ with a global average of 10%.
643 Figure 10 shows similar spatial plots for modeled versus CERES derived SWDOWN,
644 OLR, SWCF, and LWCF. We note that modeled SWCF is calculated based on the differences
645 between the net cloudy sky and net clear sky shortwave radiation at the top of atmosphere, which
646 in turn are dependent on cloud properties including the CLDFRA, COT, cloud asymmetry
647 parameter, and cloud albedo. It is possible that due to the overprediction of CLDFRA, the
648 magnitudes of the simulated SWCF are greater than those from CERES (Figures 10c and 10g),
649 even though the other cloud variables are underpredicted. LWCF is calculated based on the
650 differences in clear-sky OLR and cloudy-sky OLR, which in turn are dependent on CLDFRA,
651 COT, and absorbance and radiance due to atmospheric gases. The underprediction of total-sky
652 OLR (Table 2 and Figures 10b and 10f) leads to an overprediction in LWCF. SWCF is largely
653 overpredicted over eastern U.S. and especially over the Atlantic ocean (Figures 10c and 10g).
654 LWCF is also overpredicted by the model in similar locations as SWCF, such as in southeastern
655 U.S., and over the ocean in the eastern portion of the domain (Figures 10d and 10h). This is further

confirmed by the underpredictions in SWDOWN over the Atlantic ocean and in general over the eastern portion of the domain, as increased clouds (as a consequence of overpredicted AOD, CWP and COT) and SWCF lead to less SWDOWN reaching the ground (Figures 10a and 10e) which also eventually leads to a reduction in the OLR also over the eastern portion of the domain. The larger negative SWCF and positive LWCF in the model compared to CERES, however, lead to an overall good agreement with CERES for the net cloud forcing (SWCF + LWCF; not shown). The mean bias for SWCF against CERES of 7.8 W m^{-2} and that for LWCF against CERES of 6.9 W m^{-2} are comparable to the results from the CMIP5 models of -10 to 10 W m^{-2} over CONUS region (Figure 9.5 in Flato et al., 2013). The evaluation of 10-year averaged predictions of aerosol-cloud-radiation variables is similar to the results from the WRF/Chem simulations in 2006 and 2010 by Yahya et al. (2014 and 2015). For example WRF/Chem generally performs well for cloud fraction but AOD, CDNC, CWP and COT are underpredicted in both studies, which possibly indicate consistent biases for every year contributing to climatological biases.

4. Summary and Conclusions

Overall, the model slightly underpredicts T2 with a mean bias of ~-0.3 °C, which is consistent or better than other studies based on chemical transport models and regional climate models. The underpredictions in T2 correlate to the overpredictions in RH2. WS10 biases are likely due to issues with unresolved topography or due to inaccuracies in the selection of representative grid points. There are seasonal biases in precipitation, where overpredictions tend to occur largely over the summer months; however, precipitation is overpredicted every year between 2001 and 2010 likely due mainly to uncertainties in WRF cumulus and microphysics parameterizations. in particular, the use of a different cumulus parameterization scheme, e.g., based on the MSKF available in WRF/Chem version 3.7 or newer has been shown in the sensitivity

679 study to significantly reduce precipitation biases. Other factors contributing to the precipitation
680 bias include the use of bias-corrected CESM_NCSU data (instead of NCEP reanalysis data), and
681 the use of an reinitialization frequency of 1-month. A satisfactory model performance for
682 meteorological variables is important and necessary when simulating future years, as data
683 evaluation is not possible. Meteorological variables such as temperature, humidity, wind speed
684 and direction, PBL height, and radiation have a strong impact on chemical predictions, and thus
685 are critical to the satisfactory model performance when predicting chemical variables such as O_3
686 and $PM_{2.5}$. Biases in O_3 and $PM_{2.5}$ concentrations can be attributed to biases in any of the
687 meteorological and chemical variables. The model performs generally well for radiation variables,
688 as well as for the main chemical species such as O_3 and $PM_{2.5}$, which indicates that the processed
689 RCP 8.5 emissions are reasonably accurate to produce acceptable results for the concentrations of
690 chemical species.

691 Modeled O_3 mixing ratios at the CASTNET sites are slightly underpredicted, but are
692 slightly overpredicted at AIRS-AQS sites, in part due to the fact that the CASTNET sites are
693 classified as rural, while the AIRS-AQS sites are classified as both urban and rural. O_3 mixing
694 ratios at the AIRS-AQS sites tend to be overpredicted during the colder fall and winter seasons,
695 and annually, O_3 mixing ratios are overpredicted every year from 2001 to 2010. O_3 mixing ratios
696 at the CASTNET sites are underpredicted for all climatological months, while the largest
697 underpredictions are observed from January to May. However, on a decadal time scale,
698 WRF/Chem adequately represents the different O_3 probability distributions at the AIRS-AQS and
699 CASTNET sites. This study also showed that peak O_3 mixing ratios are observed over April and
700 May rather than June to August, which is consistent with Cooper et al. (2014) who attributed this
701 to emission reductions and opposite trends in O_3 mixing ratios over eastern and western U.S. over

the last 20 years. Modeled $PM_{2.5}$ concentrations tend to be overpredicted at the IMPROVE sites but underpredicted at the STN sites. $PM_{2.5}$ at the IMPROVE sites tend to be underpredicted in spring and summer but overpredicted in fall and winter, while $PM_{2.5}$ concentrations against STN are persistently underpredicted for all climatological months. The IMPROVE and STN sites are classified as rural and urban, respectively. Due to the relatively coarse horizontal resolution of the model (36 × 36 km), the model is unable to capture the locally higher $PM_{2.5}$ concentrations at the STN sites. In general, however, the model performs relatively well for total $PM_{2.5}$ concentrations at the IMPROVE and STN sites with NMBs of within ±25%, although larger biases exist for $PM_{2.5}$ species. Model performance for PM_{10} should be improved, as PM_{10} also has important impacts on climate through influencing the radiative budget both directly and indirectly due to its larger size and higher concentrations. The choice of observational networks for model evaluation are therefore important as both networks can show positive and negative biases depending on the type and location of the sites (e.g., O_3 against AIRS-AQS and CASTNET, and $PM_{2.5}$ against STN and IMPROVE). The major uncertainties lie in the predictions of cloud-aerosol variables. As demonstrated in this study, large biases and error in simulating cloud variables even in the most advanced models such as WRF/Chem, indicating a need for future improvement in relevant model treatments such as cloud dynamics and thermodynamics, as well as aerosol-cloud interactions. In addition, there are large uncertainties in satellite retrievals of cloud variables for evaluation. In this study, most of the cloud-aerosol variables including AOD, COT, CWP, and CDNC are on average underpredicted across the domain; however, the overpredictions of cloud variables including COT and CWP over the Atlantic ocean and eastern U.S. lead to underpredictions in radiation and overpredictions in cloud forcing, which are important parameters when simulating future climate change.

In summary, the model is able to predict O_3 mixing ratios and $PM_{2.5}$ concentrations relatively well with regards to decadal scale air quality and climate applications. The model is able to predict meteorological variables satisfactorily and with results comparable to RCM and GCM applications from literatures. Possible reasons behind the chemical and meteorological biases identified through this work should be taken into account when simulating longer climatological periods and/or future years. Aerosol-cloud-radiation variables are important for climate simulations, the performance of these variables are not as good as that of the chemical and meteorological variables. They contain consistent biases in single-year evaluations of WRF/Chem. However, magnitudes of biases for SWCF and LWCF are comparable to those from literature, which suggests that model improvements should be made in terms of bias correction of downscaled ICs/BCs as well as aerosol-cloud-radiation parameterizations in the model. In addition, having consistent physical and chemical mechanisms between the GCM and RCMs could help to reduce uncertainties in the results (Ma et al., 2014). Although CESM and WRF/Chem use similar chemistry and aerosol treatments in this work, they use somewhat different physics schemes which may contribute to such uncertainties. The development of scale-aware parameterizations that can be applied at both global and regional scales would help reduce uncertainties associated with the use of different schemes for global simulations and downscaled regional simulations.

Acknowledgments

This study is funded by the National Science Foundation EaSM program (AGS-1049200) at NCSU. For WRF/Chem simulations, we would like to acknowledge high-performance computing support from Yellowstone (ark:/85065/d7wd3xhc) provided by NCAR's

~~Computational and Information Systems Laboratory, sponsored by the National Science Foundation.~~

Code and Data Availability

The WRF/Chem v3.6.1 code used in this paper will be available upon request. However, we highly encourage users to download the latest available version of the WRF/Chem code from NOAA's web site at http://www2.mmm.ucar.edu/wrf/users/download/get_source.html. The updates in our in-house version of WRF/Chem v3.6.1 has been implemented into WRF/Chem v3.7 and WRF/Chem v3.7.1 for scientific community release. The WRF/Chem v3.7 and WRF/Chem v3.7.1 codes are now publicly available at http://www2.mmm.ucar.edu/wrf/users/download/get_source.html. These latest versions of the source codes contain all major changes in the standard version of WRF/Chem v3.6.1 used in for this study. In addition, they have been rigorously tested for compatibility and compiling issues on various platforms. The inputs including the meteorological files, meteorological initial and boundary conditions, chemical initial and boundary conditions, model set-up and configuration, and the namelist set-up, and instructions on how to run the simulations for a 1-day test case, as well as a sample output for 1-day test can be provided upon request.

Acknowledgments

This study is funded by the National Science Foundation EaSM program (AGS-1049200) at NCSU. The emissions for chemical species that are not available from the RCP emissions are taken from the 2008 NEI-derived emissions for 2006 and 2010 provided by the U.S. EPA, Environment Canada, and Mexican Secretariat of the Environment and Natural Resources (Secretaría de Medio Ambiente y Recursos Naturales-SEMARNAT) and National Institute of

Ecology (Instituto Nacional de Ecología-INE) as part of the Air Quality Model Evaluation International Initiative (AQMEII). The hourly temporal profiles of 2001, 2005, and 2010 RCP emissions are based on the 2002NEI and the AQMEII 2006 and 2010 emissions derived based on the 2008 NEI. The authors acknowledge use of the WRF-Chem preprocessor tool mozbc provided by the Atmospheric Chemistry Observations and Modeling Lab (ACOM) of NCAR and the script to generate initial and boundary conditions for WRF based on CESM results provided by Ruby Leung, PNNL. For WRF/Chem simulations, we would like to acknowledge high-performance computing support from Yellowstone (ark:/85065/d7wd3xhc) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation.

References

Abdul-Razzak, H. and Ghan, S. J.: A parameterization of aerosol activation, 2. Multiple aerosol types, J. Geophys. Res., 105(5), 6837-6844, 2000.

Aitken, A.C., DeCarlo, P.F., Kroll, J.H., Worsnop, D.R., Huffman, J.A., Docherty, K.S., Ulbrich, I.M., Mohr, C., Kimmel, J.R., Sueper, D., Sun, Y., Zhang, Q., Trimborn, A., Northway, M., Ziemann, P.J., Canagaratna, M.R., Onasch, T.B., Alfarra, M.R., Prevot, A.S.H., Dommen, J., Duplissy, J., Metzger, A., Baltensperger, U. and Jimenez, J.L.: O/C and OM/OC ratios of primary, secondary and ambient organic aerosols with high-resolution time of flight aerosol mass spectrometry, Environ. Sci. Technol., 42, 4478 – 4485, 2008.

Alapaty, K., Herwehe, J., Nolte, C.G., Bullock, R.O., Otte, T.L., Mallard, M.S., Dudhia, J. and Kain, J.S.: Introducing subgrid-scale cloud feedbacks to radiation in WRF, the 13th WRF Users Workshop, Boulder, CO, June 26 to 29, 2012.

793 Ahmadov, R., McKeen, S.A., Robinson, A.L., Bareini, R., Middlebrook, A.M., De Gouw, J.A.,
794 Meagher, J., Hsie, E.-Y., Edgerton, E., Shaw, S. and Trainer, M.: A volatility basis set model
795 for summertime secondary organic aerosols over the eastern United States in 2006, J.
796 Geophys. Res. 117, D06301, doi:10.1029/2011JD016831, 2012.
797 Beniston, M., Stephenson, D.B., Christensen, O.B., Ferro, C.A.T., Frei, C., Goyette, S.,
798 Halsnaes, K., Holt, T., Jylha, K., Koffi, B., Palutikof, J., Scholl, R., Semmler, T. and Woth,
799 K.: Future extreme events in European climate: an exploration of regional climate model
800 projections, Clim. Change, 81, 71 – 95, doi: 10.1007/s10584-006-9226-z, 2007.
801 Bennartz, R.: Global assessment of marine boundary layer cloud droplet number concentration
802 from satellite, J. Geophys. Res-Atmos, 112(D2), D02201, doi:10.1029/2006JD007547, 2007.
803 Brunner, D., Savage, N., Jorba, O., Eder, B., Giordano, L., Badia, A., Balzarini, A., Baro, R.,
804 Bianconi, R., Chemel, C., Curci, G., Forkel, R., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A.,
805 Hozak, L., Im, U., Knote, C., Makar, P., Manders-Groot, A., van Meijgaard, E., Neal, L.,
806 Perez, J.L., Pirovano, G., San Jose, R., Schroder, W., Sokhi, R.S., Syrakov, D., Torian, A.,
807 Tuccella, P., Werhahn, J., Wolke, R., Yahya, K., Zabkar, R., Zhang, Y., Hogrefe, C. and
808 Galmarini, S.: Comparative analysis of meteorological performance of coupled chemistry-
809 meteorology models in the context of AQMEII phase 2, Atmos. Environ., in press,
810 doi:10.1016/j.atmosenv.2014.12.032, 2014.
811 Caldwell, P., H.-N.S. Chin, D.C. Bader, and G. Bala (2009), Evaluation of a WRF dynamical
812 downscaling simulation over California, Clim. Change., 95, 499-521.
813 Campbell, P. C**.,** Zhang, Y., Yahya, K., Wang, K., Hogrefe, C., Pouliot, G., Knote, C., Hodzic,
814 A., San Jose, R., Perez, J., Jimenez-Guerrero, P., Baro, R. and Makar, P.: A Multi-Model
815 Assessment for the 2006 and 2010 Simulations under the Air Quality Model Evaluation

International Initiative (AQMEII) Phase 2 over North America: Part I, Indicators of the Sensitivity of O_3 and $PM_{2.5}$ Formation Regimes, Atmos. Environ., in press, doi:10.1016/j.atmosenv.2014.12.026, 2014.

Chen, F. and Dudhia, J.: Coupling an advanced land-surface/hydrology model with the Penn State/NCAR MM5 modeling system. Part I: Model implementation and sensitivity. Mon. Wea. Rev., 129, 569-585, 2001.

Clough, S.A., Shephard, M.W., Mlawer, J.E., Delamere, J.S., Iacono, M.J., Cady-Pereira, K., Boukabara, S. and Brown, P.D.: Atmospheric radiative transfer modeling: a summary of the AER codes, J. Quant. Spectrosc. Radiat. Transfer, 91(2), 233 – 244, doi: 10.1016/j.qsrt.2004.05.058, 2005.

Cooper, O.R., Parrish, D.D., Ziemke, J., Balashov, N.V., Cupeiro, M., Galbally, I.E., Gilge, S., Horowitz, L., Jensen, N.R., Lamarque, J.-F., Naik, V., Oltmans, S.J., Schwab, J., Shindell, D.T., Thompson, A.M., Thouret, V., Wang, Y. and Zbinden, R.M.: Global distribution and trends of tropospheric ozone: An observation-based review, Elem. Sci. Anth., 2, 000029, doi:10.12952/journal.elementa.000029, 2014.

Dasari, H.P., Salgado, R., Perdigao, J. and Challa, V.S.: A regional climate simulation study using WRF-ARW model over Europe and evaluation for extreme temperature weather events, Intl., J. of Atmos. Sci., ID 704079, doi:10.1155/2014/704079, 2014.

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G. and Tarpley, J.D.: Implementation of NOAH land surface model advances in the National Centers for Environmental Prediction operational mesoscale model, J. Geophys. Res., 108, D22, 8851, doi:10.1029/2002JD003296, 2003.

838 EPA.: Our Nation's Air – Status and Trends through 2010, Particle Pollution, Report by the U.S.
839 EPA, 4pp, http://www.epa.gov/airtrends/2011, 2011, last accessed July 6th, 2015.
840 Fan, F., Bradley, R.S. and Rawlins, M.A.: Climate change in the northeastern U.S.: regional
841 climate validation and climate change projections, Clim. Dyn., 43, 145 – 161,
842 doi:10.1007/s00382-014-2198-1, 2014.
843 Feser, F., Rockel, B., Von Storch, H., Winterfeldt, J. and Zahn, M.: Regional climate models add
844 value to global model data, Bull. Amer. Meteor. Soc., 92, 1181 – 1192, 2011.
845 Flato et al.: Evaluation of Climate Models, In: Climate Change 2013: The Physical Science
846 Basis. Contribution of Working Group I to the Fifth Assessment Report of the
847 Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M.
848 Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)],
849 Cambridge University Press, Cambridge, United Kingdom and New York, NY, U.S.A.,
850 2013.
851 Gao, Y., Fu, J.S., Drake, J.B., Liu, Y. and Lamarque, J.F.: Projected changes of extreme weather
852 events in the eastern United States based on a high resolution climate modeling system,
853 Environ. Res. Lett., 7, 044025, 2012.
854 Gao, Y., Fu, J.S., Drake, J.B., Lamarque, J.-F. and Liu, Y.: The impact of emission and climate
855 change on ozone in the United States under representative concentration pathways (RCPs),
856 Atmos. Chem. Phys., 2013, 9607 – 9621, 2013.
857 Glotfelty, T., He, J. and Zhang, Y.: Updated organic aerosol treatments in CESM/CAM5:
858 development and initial application, in preparation, 2015.
859 Gong, S., Barrie, L.A. and Blanchet, J.P.: Modeling sea salt aerosols in the atmosphere: 1. Model
860 development, J. Geophys. Res., 102, 3805-3818, doi:10.1029/96JD02953, 1997.

Grell, G.A., Knoche, R., Peckham, S.E. and McKeen, S.A.: Online versus offline air quality modeling on cloud-resolving time scales, Geophys. Res. Lett., 31 (16), doi:10.1029/2004GL020175, 2004.

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, 2005.

Grell, G.A. and Freitas, S.R.: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, Atmos. Chem. Phys., 14, 5233-5250, doi:10.5914/acp-14-5233-2014, 2014.

Guenther, A., Kart, T., Harley, P., Wiedinmyer, C., Palmer, P.I. and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181-3210, 2006.

He, J., Zhang, Y., Glotfelty, T., He, R., Bennartz, R., Rausch, J. and Sartelet, K.: Decadal simulation and comprehensive evaluation of CESM/CAM5.1 with advanced chemistry, aerosol microphysics and aerosol-cloud interactions, J. Adv. Model. Earth Syst., 7, 110 – 141, doi:10.1002/2014MS000360, 2015.

Hong, S.-Y., Noh, Y. and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Wea. Rev., 134, 2318-2341, 2006.

Hong, S.-Y.: A new stable boundary-layer mixing scheme and its impact on the simulated East Asian summer monsoon, Q.J.R. Meteorol. Soc., 136, 1481 – 1496, doi:0.1002/qj.665, 2010.

Hurrell, J.W., Holland, M.M., Gent, P.R., Ghan, S., Kay, J.E., Kushner, P.J., Lamarque, J.-F., Large, W.G., Lawrence, D., Lindsay, K., Lipscomb, W.H., Long, M.C., Mahowald, N., Marsh, D.R., Neale, R.B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W.D.,

884 Hack, J.J., Kiehl, J. and Marshall, S.: The Community Earth System Model: A framework for
885 collaborative research, Bull. Am. Meteorol. Soc., 94, 1339 – 1360, doi:10.1175/BAMS-D-
886 12-00121.1, 2013.
887 Iacono, M.J., Delamere, J.S., Mlawer, E.J., Shepard, M.W., Clough, S.A. and Collins, W.D.:
888 Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative
889 transfer models, J. Geophys. Res., 113, D13103, doi:10.1029/2008JD009944, 2008.
890 IPCC : Climate change 2013: the physical science basis. In: Stocker, T.F., Qin, D., Plattner, G.-
891 K., Tignor, M.M.B., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M.
892 (Eds.), Contribution of Working Group I to the Fifth Assessment Report of the
893 Intergovernmental Panel on Climate Change, Summary for Policymakers, 2013.
894 Jacob, D., Barring, L., Christensen, O.B., Christensen, J.H., de Castro, M., Deque, M., Giorgi, F.,
895 Hagemann, S., Hirschi, M., Jones, R., Kjellstrom, E., Lenderink, G., Rockel, B., Sanchez, E.,
896 Schar, C., Seneviratne, S.I., Somot, S., van Ulden, A. and van den Hurk, B.: An inter-
897 comparison of regional climate models for Europe: model performance in present-day
898 climate, Clim. Change, 81, 31 – 52, 2007.
899 Jimenez, P.A. and Dudhia, J.: Improving the representation of resolved and unresolved
900 topographic effects on surface wind in the WRF model, J. Appl. Meteor. Climatol., 51, 300 –
901 316, 2012.
902 Jones, R.G., Noguer, M., Hassell, D.C., Hudson, D., Wilson, S.S., Jenkins G.J. and Mitchell,
903 J.F.B.: Generating high resolution climate change scenarios using PRECIS, Met Office
904 Hadley Centre, Exeter, UK, 40 pp., April 2004, 2004.
905 Jones, S. and Creighton, G.: AFWA dust emission scheme for WRF/Chem-GOCART, 2011
906 WRF workshop, June 20-24, Boulder, CO, USA, 2011.

907 Karamchandani, P., Zhang, Y., Chen, S.-Y., and Balmori-Bronson, R.: Development of an
908 extended chemical mechanism for global-through-urban applications, Atmos. Poll. Res., 3, 1
909 – 24, doi:10.5094/apr.2011.047.
910 Kim, J., Waliser, D.E., Mattmann, C.A., Mearns, L.O., Goodale, C.E., Hart, A.F., Crichton,
911 D.J., McGinnis, S., Lee, H., Loikith, P.C. and Boustani, M.: Evaluation of the surface
912 climatology over the conterminous United States in the North American Regional Climate
913 Change Assessment Program Hindcast Experiment using a regional climate model evaluation
914 system, J. Climate, 26, 5698 – 5715, 2013.
915 King, N.J., Bower, K.N., Crosier, J. and Crawford, I.: Evaluating MODIS cloud retrievals with in
916 situ observations from VOCALS-REx, Atmos. Chem. Phys., 13, 191 – 209, 2013.
917 Legates, D.R. and McCabe Jr., G.J.: Evaluating the use of "goodness-of-fit" measures in
918 hydrologic and hydroclimatic model validation, Water Resour. Res., 35(1), 233 – 241,
919 doi:10.1029/1998WR900018, 1999.
920 Levy, R.C., Remer, L.A. and Dubovik, O.: Global aerosol optical properties and application to
921 Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land, J. Geophys.
922 Res., 112(D13), doi:10.1029/2006JD007815, 2007.
923 Levy, R.C., Mattoo, S., Muchak, L.A., Remer, L.A., Sayer, A.M., Patadia, F., Hsu, N.C.: The
924 Collection 6 MODIS aerosol products over land and ocean, Atmos. Meas. Tech., 6, 2989 –
925 3034, 2013.
926 Leung, R.L., Qian, Y. and Bian, X.: Hydroclimate of the Western United States based on
927 Observations and Regional Climate Simulation of 1981 – 2000, Part I: Seasonal Statistics, J.
928 Clim., 16(12), 1892 – 1911, 2003.

929 Liu, X., Easter, R.C., Ghan, S.J., Zaveri, R., Rasch, P., Shi, X., Lamarque, J.-F., Gettelman, A.,
930 Morrison, H., Vitt, F., Conley, A., Park, S., Neale, R., Hannay, C., Ekman, A.M.L., Hess, P.,
931 Mahowald, N., Collins, W., Iacono, M.J., Bretherton, C.S., Flanner, M.G., and Mitchell, D.:
932 Toward a minimal representation of aerosols in climate models: description and evaluation in
933 the Community Atmosphere Model CAM5, Geosci. Mod. Dev., 5, 709 – 739,
934 doi:10.5194/gmd-5-709-2012, 2012.
935 Loeb, N.G., Wielicki, B.A., Doelling, D.R., Smith, L., Keyes, D.F., Kato, S., Manalo-Smith, N.
936 and Wong, T.: Toward Optimal Closure of the earth's top-of-atmosphere radiation budget, J.
937 Climate, 22, 748 – 766, 2009.
938 Ma, P.-L., Rasch, P.J., Fast, J.D., Easter, R.C., Gustafson Jr., W.I., Liu, X., Ghan, S.J. and Singh,
939 B.: Assessing the CAM5 physics suite in the WRF-Chem model: implementation, resolution
940 sensitivity, and a first evaluation for a regional case study, Geosci. Model Dec., 7, 755 – 778,
941 2014.
942 Mass, C.: Improved subgrid drag or hyper PBL/vertical resolution? Dealing with the stable PBL
943 problems in WRF, presented at the 13th WRF Users' Workshop, June 26 – 29, Boulder, CO,
944 2012.
945 Molders, N., Bruyere, C.L., Gende, S. and Pirhala, M.A.: Assessment of the 2006-2012
946 Climatological Fields and Mesoscale Features from Regional Downscaling of CESM Data by
947 WRF/Chem over Southeast Alaska, Atmos. Clim. Sci., 4, 589 – 613, 2014.
948 Morrison, H., Thompson, G. and Tatarskii, V.: Impact of cloud microphysics on the development
949 of trailing stratiform precipitation in a simulated squall line: Comparison of One- and Two-
950 Moment Schemes, Mon. Wea. Rev., 137, 991-1007, 2009.

951 Moss, R. H., Edmonds, J.A., Hibbard, K.A., Manning, M.R., Rose, S.K., van Vuuren, D.P.,
952 Carter, T.R., Emori, S., Kainuma, M., Kram, T., Meehl, G.A., Mitchell, J.F.B., Nakicenovic,
953 N., Riahi, K., Smith, S.J., Stouffer, R.J., Thomson, A.M., Weyant, J.P. and Wilbanks, T.J.:
954 The next generation of scenarios for climate change research and assessment, Nature, 463,
955 747 – 756, doi: 10.1038/nature0882, 2010.
956 Nasrollahi, N., AghaKouchak, A., Li, J., Gao, X., Hsu, K. and Sorooshian, S.: Assessing the
957 Impacts of Different WRF Precipitation Physics in Hurricane Simulations, Wea. Forecasting,
958 27, 1003 – 1016, 2012.
959 Neale R.B., Jadwiga, H.R., Conley, A.J., Park, S., Lauritzen, P.H., Gettelman, A., Williamson,
960 D.L., Rasch, P., Vavrus, S.J., Taylor, M.A., Collins, W.D., Zhang, M. and Lin, S.-J.:
961 Description of the NCAR Community Atmosphere Model (CAM 5.0), NCAR Tech. Note
962 NCAR/TN-486+STR, Natl. Cent. for. Atmos. Res., Boulder, CO, available at
963 http://www.cesm.ucar.edu/models/cesm1.0/cam/docs/description/cam5_desc.pdf, 2010, last
964 accessed July 6th, 2015.
965 Otte, T.L., Nolte, C.G., Otte, M.J. and Bowden, J.H.: Does Nudging squelch the extremes in
966 regional climate modeling? J. Clim., 25, 7046 – 7066, doi:10.1175/JCLI-D-12-00048.1,
967 2012.
968 Penrod, A., Zhang, Y., Wang, K., Wu, S-Y. and Leung, R.L.: Impacts of future climate and
969 emission changes on U.S. air quality, Atmos. Environ., 89, 533 – 547, 2014.
970 Petikainen, J.-P., O'Donnell, D., Teichmann, C., Karstens, U., Pfeifer, S., Kazil, J., Podzun, R.,
971 Fiedler, S., Kokkola, H., Birmili, W., O'Dowd, C., Baltensperger, U., Weingartner, E.,
972 Gehrig, R., Spindler, G., Kulmala, M., Feichter, J., Jacob, D. and Laaksonen, A.: The
973 regional aerosol-climate model REMO-HAM, Geosci. Mod. Dev., 5, 1323 – 1339, 2012.

974 Pleim, J.E. and Gilliam, R.: An indirect data assimilation scheme for deep soil temperature in the
975 Pleim-Xiu Land Surface Model, J. Appl. Meteor. Climatol., 48, 1362 – 1376, 2009.
976 Pouliot, G., van der Gon, H.A.C.D., Kuenen, J., Zhang, J., Moran, M. and Makar, P.: Analysis of
977 the Emission Inventories and Model-Ready Emission Datasets of Europe and North America
978 for Phase 2 of the AQMEII Project, Atmos. Environ., in press,
979 doi:10.1016/j.atmosenv.2014.10.061, 2014.
980 Rawlins, M.A., Bradley, R.S. and Diaz, H.F.: Assessment of regional climate model simulation
981 estimates over the northeast United States, J. Geophys. Res., 117, D23112,
982 doi:10.1029/2012JD018137, 2012.
983 Refslund, J., Dellwik, E., Hahmann, A.N., Barlage, M.J. and Boegh, E.: Development of satellite
984 green vegetation fraction time series for use in mesoscale modeling: application to the
985 European heat wave 2006, Theor. Appl. Climatol., 117, 377-392, doi:10.1007/s00704-013-
986 1004-z, 2014.
987 Sarwar, G., Luecken, D.J. and Yarwood, G.: Developing and implementing an updated chlorine
988 chemistry into the Community Multiscale Air Quality Model, presented at the 28th
989 NATO/CCMS International Technical Meeting, Lepzig, Germany, May 15 – 19, 2006.
990 Sarwar, G., Luecken, D. and Yarwood, G.: Chapter 2.9: Developing and implementing an
991 updated chlorine chemistry into the community multiscale air quality model, Developments
992 in Environmental Science, Volume 6, C. Borrego and E. Renner (Eds.), Elsevier Ltd,
993 DOI:10.1016/S1474-8177(07)06029-9, 168 pp., 2007.
994 Sarwar, G., Fahey, K., Napelenok, S., Roselle, S. and Mathur, R.: Examining the impact of
995 CMAQ model updates on aerosol sulfate predictions, the 10th Annual CMAS Models-3
996 User's Conference, October, Chapel Hill, NC, 2011.

997 Shan, Z., Parol, F., Riedi, J., Cornet, C. and Thieuleux, F.: Examination of POLDER/PARASOL
998 and MODIS/Aqua cloud fractions and properties representativeness, J. Climate, 24, 4435 –
999 4450, 2011.
1000 Sievering, H.: Small-particle dry deposition under high wind speed conditions: Eddy flux
1001 measurements at the boulder atmospheric observatory, Atmos. Environ., 21 (10), 2179 –
1002 2185, 1987.
1003 Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M.A., Mitchell, K., Ek, M., Gayno, G.,
1004 Wegiel, J. and Cuenca, R.H.: Implementation and verification of the unified NOAH land
1005 surface model in the WRF model. 20th conference on weather analysis and forecasting/16th
1006 conference on numerical weather prediction, pp. 11 – 15, 2004.
1007 Toth, T.D., Zhang, J., Campbell, J.R., Reid, J.S., Shi, Y., Johnson, R.S., Smirnov, A., Vaughan,
1008 M.A. and Winker, D.M.: Investigating enhanced Aqua MODIS aerosol optical depth
1009 retrievals over the mid-to-high latitude Southern Oceans through intercomparison with co-
1010 located CALIOP, MAN and AERONET data sets, J. Geophys. Res: Atmos, 18, 1- 15, 2013.
1011 van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C.,
1012 Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith,
1013 S.J. and Rose, S.K.: The representative concentration pathways: an overview, Climate
1014 Change, 109, 5 – 31, doi: 10.1007/s10584-011-0148-z, 2011.
1015 Wang, K., Zhang, Y., Yahya, K., Wu, S.-Y. and Grell, G.:, Implementation and initial
1016 application of new chemistry-aerosol options in WRF/Chem for simulating secondary
1017 organic aerosols and aerosol indirect effects for regional air quality, Atmos. Environ., in
1018 press, doi: 10.1016/j.atmosenv.2014.12.007, 2014a.

1019 Wang, K., Yahya, K., Zhang, Y., Hogrefe, C., Pouliot, G., Knote, C., Hodzic, A., San Jose, R.,
1020 Perez, J.L., Guerrero, P.J., Baro, R. and Makar, P.: Evaluation of Column Variable
1021 Predictions Using Satellite Data over the Continental United States: A Multi-Model
1022 Assessment for the 2006 and 2010 Simulations under the Air Quality Model Evaluation
1023 International Initiative (AQMEII) Phase 2, Atmos. Environ., in press,
1024 doi:10.1016/j.atmosenv.2014.07.044, 2014b.
1025 Warrach-Sagi, K., Schwitalla, T., Wulfmeyer, V. and Bauer, H.-S.: Evaluation of a climate
1026 simulation in Europe based on the WRF-NOAH model system: precipitation in Germany,
1027 Clim. Dyn., 41, 755 – 774, doi:10.1007/s00382-013-1727-7, 2013.
1028 Willmott, C. J.: On the validation of models, Phys. Geog., 2, 184 – 194, 1981.
1029 Xing, J., Mathur, R., Pleim, J., Hogrefe, C., Gan, C.-M., Wong, D.C., Wei, C., Gilliam, R. and
1030 Pouliot, G.: Observations and modeling of air quality trends over 1990-2010 across the
1031 Northern Hemisphere: China, the United States and Europe, Atmos. Chem. Phys., 15, 2723 –
1032 2747, doi:10.5194/acp-15-2723-2015.
1033 Xu, Z. and Yang, Z.-L.: An improved dynamical downscaling method with GCM Bias
1034 Corrections and Its Validation with 30 years of climate simulations, J. Clim., 25, 6271 –
1035 6286, 2012.
1036 Yahya, K., Wang, K., Gudoshava, M., Glotfelty, T. and Zhang, Y.: Application of WRF/Chem
1037 over North America under the AQMEII Phase 2. Part I. Comprehensive Evaluation of 2006
1038 Simulation, Atmospheric Environment, in press, doi:10.1016/j.atmosenv.2014.08.063, 2014.
1039 Yahya, K., He, J., and Zhang, Y.: Multi-Year Applications of WRF/Chem over Continental U.S.:
1040 Model Evaluation, Variation Trend, and Impacts of Boundary Conditions over CONUS, J.
1041 Geophy. Res., in review, 2015a.

1042 Yahya, K., Wang, K., Zhang, Y. and Kleindienst, T.E.: Application of WRF/Chem over North
1043 America under the AQMEII Phase 2. Part II. Comprehensive Evaluation of 2010 Simulation
1044 and Responses of Air Quality and Meteorology-Chemistry Interactions to Changes in
1045 Emissions and Meteorology from 2006 to 2010, Geosci. Model Dev., in press, 2015b.
1046 Yarwood, G., Rao, S., Yocke, M. and Whitten, G.Z.: Final Report – Updates to the Carbon Bond
1047 Chemical Mechanism: CB05, Rep. RT-04-00675, 246 pp., Yocke and Co., Novato, Calif.,
1048 2005.
1049 Yu, S., Dennis, R., Roselle, S., Nenes, A., Walker, J., Eder, B., Schere, K., Swall, J., and
1050 Robarge, W.: An assessment of the ability of 3-D air quality models with current
1051 thermodynamic equilibrium models to predict aerosol NO3-, J. Geophys. Res., 110, D07S13,
1052 doi:10.1029/2004JD004718, 2005.
1053 Yu, S., Eder, B., Dennis, R., Chu, S.-H., and Schwartz, S.: New unbiased symmetric metrics for
1054 evaluation of air quality models, Atmos. Sci. Lett., 7, 26 – 34, 2006.
1055 Yu, S., Mathur, R., Pleim, J., Wong, D., Gilliam, R., Alapaty, K., Zhao, C., and Liu, X.: Aerosol
1056 indirect effect on the grid-scale clouds in the two-way coupled WRF-CMAQ: model
1057 description, development, evaluation and regional analysis, Atmos. Chem. Phys., 14, 11247 –
1058 11285, doi:10.5194/acp-14-1-2014, 2014.
1059 Zhang, Y., Liu, P., Pun, B., and Seigneur, C.: A comprehensive performance evaluation of
1060 MM5-CMAQ for summer 1999 Southern Oxidants Study Episode, Part-I. Evaluation
1061 Protocols, Databases and Meteorological Predictions, Atmos. Environ., 40, 4825 – 4838,
1062 2006.

1063 Zhang, Y., Wen, X.-Y. and Jang, C.J.: Simulating chemistry-aerosol-cloud-radiation-climate
1064 feedbacks over the CONUS using the online-coupled Weather Research Forecasting Model
1065 with chemistry (WRF/Chem), Atmos. Environ., 44, 3568 – 3582, 2010.
1066 Zhang, Y., Y.-C. Chen, G. Sarwar, and K. Schere, 2012a, Impact of Gas-Phase Mechanisms on
1067 Weather Research Forecasting Model with Chemistry (WRF/Chem) Predictions: Mechanism
1068 Implementation and Comparative Evaluation, J. Geophys. Res., 117, D1,
1069 doi:10.1029/2011JD015775.
1070 Zhang, Y., P. Karamchandani, T. Glotfelty, D. G. Streets, G. Grell, A. Nenes, F.-Q. Yu, and R.
1071 Bennartz, 2012b, Development and Initial Application of the Global-Through-Urban
1072 Weather Research and Forecasting Model with Chemistry (GU-WRF/Chem), J. Geophys.
1073 Res., 117, D20206, doi:10.1029/2012JD017966.
1074
1075