A Multi-Model Assessment for the 2006 and 2010 Simulations under the Air Quality Model Evaluation International Initiative (AQMEII) Phase 2 over North America: Part II. Evaluation of Column Variable Predictions Using Satellite Data

Kai Wang*, Khairunnisa Yahya, and Yang Zhang*
Department of Marine, Earth, and Atmospheric Sciences, NCSU, Raleigh, NC 27695, USA
Christian Hogrefe and George Pouliot
Atmospheric Modeling and Analysis Division, Office of Research and Development, U.S. EPA, RTP, NC 27711, USA
Christoph Knote and Alma Hodzic
Atmospheric Chemistry Division, National Center for Atmospheric Research, Boulder, CO 80301, USA
Roberto San Jose and Juan L. Perez
Computer Science School, Technical University of Madrid, Campus de Montegancedo, Boadilla del Monte, 28660 Madrid, Spain
Pedro Jiménez Guerrero and Rocio Baro
Department of Physics, Ed. CIOyN, Campus de Espinardo, Regional Campus of International Excellence "Campus Mare Nostrum", University of Murcia, 30100 Murcia, Spain
Paul Makar
Air Quality Research Division, Environment Canada, Toronto, Ontario, M3H 5T4, Canada
Ralf Bennartz
Earth & Environmental Sciences, Vanderbilt University, Nashville, TN 37205, USA
Space Science and Engineering Center, University of Wisconsin-Madison, WI 53706, USA

Corresponding Authors: Tel.: +1 919 515 9688; Fax: +1 919 515 7802

Email Addresses: kwang@ncsu.edu (Kai Wang), yzhang9@ncsu.edu (Yang Zhang)

Abstract

Within the context of the Air Quality Model Evaluation International Initiative Phase 2 (AQMEII2) project, this part II paper performs a multi-model assessment of major column abundances of gases, radiation, aerosol, and cloud variables for 2006 and 2010 simulations with three online-coupled air quality models over the North America using available satellite data. It also provides the first comparative assessment of the capabilities of the current generation of online-coupled models in simulating column variables. Despite the use of different model configurations and meteorological initial and boundary conditions, most simulations show comparable model performance for many variables. The evaluation results show an excellent agreement between all simulations and satellite-derived radiation variables including downward surface solar radiation, longwave radiation, and top-of-atmospheric outgoing longwave radiation, as well as precipitable water vapor with domain-average normalized mean biases (NMBs) of typically less than 5% and correlation coefficient (R) typically more than 0.9. Most simulations perform well for column-integrated abundance of CO with domain-average NMBs of -9.4% to -2.2% in 2006 and -12.1% to 4.6% in 2010 and from reasonably well to fair for column NO_2, HCHO, and SO_2, with domain-average NMBs of -37.7% to 2.1%, -27.3% to 59.2%, and 16.1% to 114.2% in 2006, respectively, and, 12.9% to 102.1%, -25.0% to 87.6%, -65.2% to 7.4% in 2010, respectively. R values are high for CO and NO_2 typically between 0.85 and 0.9 (i.e., R^2 of 0.7-0.8). Tropospheric ozone residuals are overpredicted by all simulations due to overestimates of ozone profiles from boundary conditions. Model performance for cloud-related variables is mixed and generally worse compared to gases and radiation variables. Cloud fraction (CF) is well reproduced by most simulations. Other aerosol/cloud related variables such as aerosol optical depth (AOD), cloud optical thickness, cloud liquid water path, cloud condensation nuclei,

and cloud droplet number concentration (CDNC) are moderately to largely underpredicted by most simulations, due to underpredictions of aerosol loadings and also indicating high uncertainties associated with the current model treatments of aerosol-cloud interactions and the need for further model development. Negative correlations are found for AOD for most simulations due to large negative biases over the western part of the domain. Inter-model discrepancies also exist for a few variables such as column abundances of HCHO and SO_2 and CDNC due likely to different chemical mechanisms, biogenic emissions, and treatments of aerosol indirect effects. Most simulations can also capture the inter-annual trend observed by satellites between 2006 and 2010 for several variables such as column abundance of NO_2, AOD, CF, and CDNC. Results shown in this work provide the important benchmark for future online-couple air quality model development.

Keywords: Satellite data, online-coupled model, model evaluation, WRF/Chem, WRF-CMAQ, GEM-MACH, AQMEII

Highlights:

- Multi-model evaluation for column variables against satellite data performed for NA
- Radiation budgets and major column gases are either well or reasonably well predicted
- Large underpredictions for AOD, COT, LWP, CCN, and CDNC from most simulations
- High uncertainties associated with parameterizations of aerosol indirect effects

1. Introduction

Evaluation of air quality models (AQMs) is a key practice in advancing the scientific understanding of various physical/chemical processes treated in the models, since it can help to validate the formulations and parameterizations of major atmospheric processes introduced by model development and demonstrate their impact on capabilities of models in reproducing the

atmospheric observations. The evaluation of AQMs, especially on a regional scale, has conventionally focused on comparing model predictions with either ground-level measurements or to a lesser extent airborne in-situ data or ground-based remote sensing profiles. Not until recently, with the launches of many satellites by the U.S. National Aeronautics and Space Administration (NASA), the U.S. National Oceanic and Atmospheric Administration (NOAA), the European Space Agency (ESA), the Canadian Space Agency (CSA), and the Japan Aerospace Exploration Agency (JAXA) that can measure atmospheric constituents, radiation budgets, and cloud/aerosol properties, did the atmospheric science community start to realize the potential and feasibility of utilizing such data to evaluate regional-scale air quality models (Vijayaraghavan et al., 2008). With the development of more satellite instruments/sensors, more satellite data are now available with a global coverage and a large number of atmospheric constituents simulated by air quality models can be constrained.

Most current satellites commonly used for measuring atmospheric composition and aerosol/cloud properties are low polar-orbiting sun-synchronous satellites, which typically orbit at an altitude of 700-800 km and view the equator (or the low-mid latitudes) on the Earth at the same local time every day (Martin, 2008). Many sensors carried onboard those satellites passively detect the emitted or scattered radiation from atmospheric gases or aerosols. The detected radiances are then converted to geophysical quantities of interests through complex retrieval processes. Compared to other measurements, there are two major advantages for using satellite retrieval data for air quality applications: large synoptic spatial coverage and vertically integrated measures of atmospheric components aloft (Engel-Cox et al., 2004; Vijayaraghavan et al., 2008). Recently, an increasing number of air quality studies have utilized satellite data in many ways, e.g., identifying forest wildfires or dust storm events (Bian et al., 2007; Song et al.,

2008; Magi et al., 2009), tracing the long-range transport of air pollutants (Heald et al., 2003; Hodzic et al., 2007; Wang et al., 2009; Huang et al., 2013), deriving boundary/initial conditions (BCs/ICs) for regional air quality models (Tang et al., 2009), monitoring air quality in rural or remote regions where no ground-level network (Engel-Cox et al., 2004), conducting inverse modeling to estimate emission of precursors (Kopacz et al., 2009; Streets et al., 2013) or performing data assimilation to constrain/improve the model performance (Sandu and Chai, 2011; Miyazaki et al., 2012; Saide et al., 2013), and evaluating performance of regional and global AQMs (Kondragunta et al., 2008; Zhang et al., 2009, 2012a, b; Knote et al., 2011; Wang et al., 2012).

Significant progress has been achieved in the past decade in the development of online-coupled meteorology and chemistry/air quality modeling (Zhang, 2008; Baklanov et al., 2014). One of the key issues addressed by online-coupled models is to investigate the complex climate-chemistry-aerosol-cloud-radiation feedback processes, which are closely related with column abundance of atmospheric constituents such as ozone (O_3) and fine particular matter ($PM_{2.5}$) as well as aerosol/cloud properties such as aerosol optical depth (AOD) and cloud optical thickness (COT) in the troposphere. Accurately reproducing those column abundances and aerosol/cloud variables in the atmosphere is thus important in estimating the aerosol direct and indirect effects as well as interactions between meteorology/climate and air quality for online-couple models. The satellite retrieval products provide valuable and unique information for validation of the capabilities of models in representing column abundances and aerosol/cloud variables.

In Part I paper, a multi-model simulation intercomparison of O_3 and $PM_{2.5}$ formation indicators are conducted and a few key indicators are also evaluated using available surface and satellite observations (Campbell et al., 2014). In this Part II paper, a number of satellite

retrievals of column abundances of gases (e.g., carbon monoxide (CO) and nitrogen oxide (NO_2)), radiation budgets (e.g., downward surface solar radiation (SWDN) and outgoing top-of-atmosphere (TOA) longwave radiation (OLR)), and aerosol-cloud associated properties (e.g., AOD and COT) are used to evaluate results from three online-couple models from six research groups as part of the collaborative Air Quality Model Evaluation International Initiative Phase 2 (AQMEII2) project (Alapaty et al., 2012). AQMEII2 is targeted at evaluating the most advanced online-coupled AQMs with representation of climate-chemistry-aerosol-cloud-radiation interactions and examining their status in simulating these complex interactions. In the context of AQMEII2, the objectives of this Part II are twofold. First, to perform an operational evaluation of the column abundances of major gases and radiation/aerosol/cloud variables simulated by the participating models using satellite retrievals over the North America (NA) domain which covers the continental U.S., southern Canada, and northern Mexico for the years 2006 and 2010. Second, to examine the current status and capability of those state-of-the-science fully coupled AQMs in predicting those variables. This study provides the first comparative assessment of the capabilities of the current generation of online-coupled models in simulating column variables.

2. Model Description and Evaluation Protocols

2.1. Model Description

Six research teams apply three state-of-the-science online-coupled models over the NA domain which covers southern Canada, continental U.S., and northern Mexico during AQMEII 2. These models include the Weather Research Forecasting model (WRF) with chemistry (WRF/Chem) version 3.4.1 (Grell et al., 2005), the WRF coupled with the Community Multiscale Air Quality model system version 5.0.1 (WRF-CMAQ) (Wong et al., 2012), and the

Global Environmental Multi-scale-Modelling Air quality and Chemistry model version 1.5.1 (GEM-MACH) (Moran et al., 2010). Model/simulation configurations are summarized in Table 1 of Campbell et al. (2014). Four out of six research groups apply WRF/Chem with different model configurations. They are North Carolina State University, U.S., Technical University of Madrid, Spain, National Center for Atmospheric Research, U.S., and University of Murcia, Spain (the simulations from those groups are referred to as NCSU, UPM, NCAR, and UMU, respectively). The U.S. Environmental Protection Agency (EPA) uses WRF-CMAQ and Environment Canada (EC) uses GEM-MACH (their simulations are referred to as EPA and EC, respectively). In addition to slightly different domain sizes, large differences exist in the horizontal/vertical resolution, the physical and chemical modules, and natural emissions selected by each group. Among the six groups, NCSU, EPA, and EC conduct the full year simulations for both 2006 and 2010. UPM performs a simulation for 2006 only and UMU and NCAR perform a simulation for 2010 only. All WRF-based models use the Lambert Conformal projection while GEM-MACH uses a rotated polar projection. All groups simulate the secondary organic aerosol (SOA) formation except for UPM. All groups include aqueous-phase (AQ) chemistry and NCSU, EPA, and NCAR have included AQ chemistry for both convective and resolved clouds. Most groups treat online dust emissions except for UPM and EC. Aerosol indirect effects are considered by all the simulations except for EPA.

Despite different model configurations, all six simulations use the same set of anthropogenic emissions and chemical ICs/BCs, in order to minimize the differences caused by different chemical inputs. The anthropogenic emissions are comprised of data from the U.S., Canada, and Mexico. For the U.S. emissions, the 2008 National Emission Inventory (NEI) (version 2, released April 10, 2012) is used as the basis for both the 2006 and 2010 model ready

emission datasets (Pouliot et. al. 2014). The 2008-based modeling platform provides all the SMOKE inputs and datasets for processing with SMOKE (Pouliot et. al. 2014). These files contain the chemical speciation files, the temporal allocation, and spatial allocation used for emission processing with SMOKE. Year specific (2006 and 2010) updates for different sectors (i.e., on/off road transport, wildfires and prescribed fires, and Continuous Emission Monitoring (CEM)-equipped point sources) are used. Canadian emissions are derived from the Canadian National Pollutant Release Inventory and Air Pollutant Emissions Inventory for the year 2006. These included updated spatial allocations for Canadian mobile emissions for the emissions of NH_3, as well as other updates (Im et al., 2014a). Mexican emissions are based on a 1999 inventory and projected to year 2008 (Im et al., 2014a). Four groups use the Model of Emissions of Gases and Aerosols from Nature (MEGAN) version 2 that is embedded in WRF/Chem and two groups use different versions of the Biogenic Emissions Inventory System (BEIS), which may lead to large differences of isoprene emissions as indicated by Im et al. (2014a). The chemical ICs/BCs are provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast system (IFS)- Model for Ozone And Related Tracers (MOZART) model in the context of the Monitoring Atmospheric Composition and Climate (MACC) project for major gaseous and aerosol species with a 3-hr temporal resolution and 1.125° spatial resolution (Inness et al., 2013). These ICs/BCs are remapped based on different chemical speciation and aerosol size representations of the individual models.

2.2. Satellite Data Description

Table A1 in the supplementary material summarizes satellite data used in this study. These include tropospheric CO column abundances from the Measurements of Pollution in the Troposphere (MOPITT), tropospheric NO_2, formaldehyde (HCHO), and sulfur dioxide (SO_2)

abundances from the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIMACHY), the tropospheric O_3 residuals (TORs) derived from the Ozone Monitoring Instrument (OMI)/ Microwave Limb Sounder (MLS), SWDN and downward surface longwave radiation (LWDN) from the Cloud's and the Earth's Radiant Energy System (CERES), OLR from the Advanced Very High Resolution Radiometer (AVHRR), and AOD, COT, cloud fraction (CF), cloud condensation nuclei (CCN), and precipitable water vapor (PWV) from the Moderate Resolution Imaging Spectroradiometer (MODIS). Cloud droplet number concentration (CDNC) and cloud liquid water path (LWP) derived by Bennartz (2007) based on MODIS retrievals are also used. A brief description of those datasets is provided in the supplementary material.

In this study, all satellite data used are level-3 monthly average (except for CDNC, which is daily average) retrieval data from public resources with various resolutions (see Table A1) except for CDNC and LWP, which are derived based on MODIS data (Bennartz, 2007). All the level-3 data have been well validated and quality assured by the satellite data retrieval teams using independent aircraft and/or sonde data (Martin, 2008). The satellite data with different resolutions are mapped to the Lambert conformal projection used in all simulations using the bi-linear interpolation of the NCAR command language (http://www.ncl.ucar.edu/). The uncertainties associated with individual data/retrieval algorithms may help explain some differences between simulations and satellite-derived products and will be further discussed in Sections 3 and 4.

2.3. Evaluation Protocols

An operational performance evaluation is conducted in terms of the spatial distribution and domainwide performance statistics following the evaluation protocol from Zhang et al. (2006, 2009). The metrics used in this analysis include the normalized mean bias (NMB), the

normalized mean error (NME), the correlation coefficient (R) and the coefficient of determination (R^2), and the normalized standard deviation (NSD) (see Supplementary for associated formulas). Not all simulations predict or output all variables and only the available variables are used for the intercomparison (see Tables 1 and 2 for variable availability). The model outputs for all gas column abundances except for TORs and for all aerosol/cloud variables except for CDNC are vertically integrated up to the tropopause which is assumed to be 100 hPa (the exact choice has little influence on those variables) following Zhang et al. (2009) to generate the tropospheric amounts in order to match the satellite data. For TORs, since they are very sensitive to the choice of tropopause, the monthly average tropopause pressure provided by the NCEP reanalysis database (similar NCEP data was used for OMI/MLS retrievals) is used to calculate TORs from simulations. For CDNC, it is processed as within low level warm clouds (corresponding to pressure levels of 950-850 hPa) as suggested by Bennartz (2007). All the gas column abundances and AOD are further processed to include the values only under cloud-free conditions. As discussed in Section 5, no averaging kernels are applied for the processing of model data. All model outputs are also averaged at the same satellite crossing time in order to facilitate the comparison. Since the domain size of individual simulation is different, all simulation results have been re-gridded into the domain of NCSU as a common domain to ensure a fair intercomparison. All the results are analyzed as annual average for all variables for 2006 and 2010. In addition, the model performance from multiple models is examined using Taylor diagrams (Taylor, 2001) to provide a concise statistical summary with respect to the correlation, biases, and variances (as indicated by NSD).

3. Model Evaluation for 2006

3.1. Column Mass Abundance

Figure 1 compares the spatial distribution of tropospheric column abundances for CO, NO_2, HCHO, SO_2, and TOR between satellite observations and four simulations for 2006. The corresponding performance statistics are given in Table 1. For CO, both MOPITT observation and simulations show high CO abundances over the continental source regions (e.g., the eastern U.S., the Atlantic coast of the U.S., and California) and the trans-Pacific transport inflow regions (e.g., the Pacific Northwest Ocean) and low CO columns over elevated terrain (e.g., Rocky Mountains). All simulations underpredict CO columns with NMBs ranging from -9.4% (EPA) to -2.2% (EC) with systematic underpredictins despite the biases are typically small and within the retrieval uncertainties. As reported by Heald et al. (2003), regional emissions in particular biomass burning emissions, are expected to be the main contributor to elevated CO concentrations, thus determining the CO columns. Since all simulations use the same emission inventory, the systematic underpredictions by all simulations might therefore be caused by possible uncertainties (such as missing fire emissions) in CO emissions. Other possible contributing factors may include uncertainties associated with BCs from MACC and retrieval methods used for MOPITT data. For example, Heald et al. (2003) indicated that potential biases in the vertical profile of CO at higher altitudes from their global model (which are very sensitive to BCs) could be an important source for model biases against MOPITT observations. Emmons et al. (2009) also reported the possible positive biases for MOPITT CO retrievals over the continents as compared to oceans. The higher CO columns predicted by EC should be due to a much finer vertical resolution within the lower to free troposphere where column CO abundances are the highest (i.e., 24 layers vs. 16-17 layers for other models), which can better capture the elevated CO.

The spatial distribution of NO_2 columns is generally well reproduced by all four simulations and many hot spots of NO_2 columns observed by SCIAMACHY are captured in the Northeastern U.S., Midwest, Texas, and California, which correlate well with high NO_x emission source areas (e.g., industrialized and urban areas). The domainwide statistics show mixed performance for different simulations in terms of magnitude, with NMBs of -37.7% (EC), -14.7% (UMP), 2.1% (EPA), and 14.1% (NCSU), respectively. The discrepancies between simulations and satellite retrievals can be attributed to a few likely reasons. First, a previous study by Choi et al. (2011) suggested that NO_x emissions from the NEI 2005 have large uncertainties and may be overestimated in the southern U.S. Pouliot et al. (2014) showed that 2006 domainwide NO_x emissions are fairly similar between the NEI 2005 based AQMEII Phase 1 model inputs and the NEI 2008 based AQMEII Phase 2 model inputs, but also showed significant shifts in emission estimates for some source sectors such as mobile sources. The relative large biases for NO_2 columns by all simulations may be an indication that further work is needed to evaluate emission inputs. Second, since all the simulations use the same set of NO_x emissions, the mixed performance (i.e., overprediction vs. underprediction) also could be caused by different reaction rates used for NO_2 associated reactions simulated by different gas-phase mechanisms. Third, as reported by Martin (2008), tropospheric NO_2 and SO_2 concentrations are dominant in the planetary boundary layer (PBL) due to intensive surface sources and short lifetimes. As a result, both column NO_2 and SO_2 abundances in PBL can contribute to more than two-thirds of tropospheric NO_2 and SO_2 columns over polluted regions. Therefore, differences in PBL mixing processes simulated by the meteorological models may play an important role. An examination of PBL heights (PBLHs) (figures not shown) show that EC predicts the largest PBLH followed by EPA, UPM, and NCSU, although PBLHs between UPM and NCSU are very

close, due to the same Yonsei University PBL scheme. The pattern of PBLH can help to explain the predicted NO_2 columns from NCSU, EPA, and EC (i.e., the largest NO_2 columns from NCSU followed by EPA and EC). UPM predicting the smallest NO_2 despite with the lowest PBLH might be due to other reasons such as gas-phase mechanisms. Fourth, there might be missing processes such as the plume-in-grid in current model treatments, which was found to help improvements of column NO_2 performance by previous studies (Vijayaraghavan et al., 2009). Finally, there are uncertainties associated with satellite retrievals. Boersma et al. (2004) and some other studies (e.g., Martin et al., 2003; van Noije et al., 2006) showed that different NO_2 column retrieval approaches may lead to $\pm 5\times10^{14}$-1×10^{15} molecules cm^{-2} for additive error (±35% - 60% relative error) over polluted areas. The algorithms used to convert the measured irradiances to column values are in part dependent on air-quality models, which are used to calculate air mass factors used in the retrieval process. Recent work of McLinden et al. (2014) found air mass factors generated using higher resolution model and surface data allows significant local gradients to be resolved, increasing the retrieval estimated maximum vertical column densities of NO_2 by a factor of 2.

Both SCIAMACHY observations and four simulations show high HCHO abundances over the southeastern U.S., California, and coastal areas of Mexico (except EC which does not include the coastal areas of Mexico), where biogenic and biomass burning emissions are high. The correlation is moderate for all simulations with values of R ranging from 0.69 to 0.79 (i.e., R^2 of 0.48 to 0.62), suggesting that all simulations reproduce the spatial distribution relatively well. The discrepancies in magnitude between simulations and observations, however, are relatively large except for EPA, with NMBs of -27.3% (UPM), -24.5% (NCSU), -11.5% (EPA), and 59.2% (EC), respectively. The much larger HCHO columns predicted by EC could be due

306 to a few reasons. First, the photolysis rates of HCHO predicted by EC might be low due to much
307 higher predicted cloud water (Figure 3) leading to the lower destruction of HCHO. Second,
308 among all gas mechanisms in this study, ADOM-II simulates only isoprene without species
309 terpene/monoterpene from biogenic sources. All the terpene/monoterpene emissions are mapped
310 into a lumped species ethane. Instead of generating longer chain aldehydes and ketones, the
311 terpene/monoterpene masses from biogenic emissions in ADOM-II goes into HCHO upon
312 oxidation. This treatment leads to much higher HCHO formation compared to mechanisms that
313 explicitly represent terpene/monoterpene. Finally, as reported by Carlton and Baker (2011),
314 BEIS v3.14 tended to generate a factor of 1.5 higher HCHO emissions compared to MEGAN v2,
315 which may partially contribute the higher HCHO columns predicted by both EPA and EC
316 compared to NCSU and UPM. Due to the fact that the bulk of the NO_2 and HCHO columns are
317 within the lower PBL over polluted regions and are closely related to NO_x and VOC emission
318 sources, the ratio of column HCHO/NO_2 has been proposed as a robust indicator (Martin et al.,
319 2004) for surface photochemistry (especially NO_x- or VOC- limited O_3 chemistry) and has been
320 further examined by the Part I paper (Campbell et al., 2014).

321 All four simulations moderately or significantly overpredict SO_2 columns (NMBs ranging
322 from 16.1% to 114.2%) with moderate spatial correlation (values of R^2 ranging from 0.34-0.41).
323 Similar to NO_x, high SO_2 levels are predicted by all simulations over source regions and are
324 correlated with observations. The larger differences between simulations and observations for
325 SO_2 columns compared to other gases could be largely due to the larger uncertainties associated
326 with SO_2 retrievals. As reported by McLinden et al. (2014), SO_2 retrievals using higher
327 resolution profiles and surface data can increase maximum vertical columns of SO_2 by a factor of
328 1.4. The higher SO_2 predicted by EC is due to a lower oxidation rate of SO_2 by OH radicals (i.e.,

329 8.3×10^{-13} cm^3 $molecule^{-1}$ s^{-1} in ADOM-II vs. 8.8-9.5 $\times 10^{-13}$ cm^3 $molecule^{-1}$ s^{-1} in CBMZ and
330 CB05 under ambient temperature and pressure) (Lurmann et al., 1986) and by aqueous chemistry
331 (Makar et al., 2014). NCSU predicts the lowest SO_2 columns due to the inclusion of both
332 heterogeneous chemistry of SO_2 on aerosol particles and in convective clouds. Both treatments
333 convert a large amount of SO_2 from gas-phase into particulate sulfate. However, convective
334 cloud chemistry is also simulated by EPA which gives higher overpredictions of SO_2 columns
335 than NCSU, suggesting the important role of SO_2 heterogeneous chemistry.

336 Due to an erroneous mapping of O_3 profile from 50 hPa to 100 hPa in the simulation
337 conducted by EPA which leads to unrealistic high TORs, only TOR plots from the other three
338 simulations are shown in Figure 1 and Table 1 for 2006. All three simulations show systematic
339 overpredictions of TORs compared to OMI/MLS with NMBs of 19.9% (EC), 29.9% (UPM), and
340 38.0 (NCSU), respectively, which are mainly caused by the O_3 profiles provided by MACC. The
341 general better TOR performance from EC is associated with higher vertical resolution that can
342 represent the tropopause provided by NCEP reanalysis data better.

343 Figure 2a shows the Taylor diagram for the column abundances of the five gases from
344 simulations in 2006, which can help assess the general skill of the models. Due to the large
345 amplitude of SO_2 variations, all markers for SO_2 are displayed as outlier points outside the plot
346 area. Most simulations underpredict the amplitude of variability (with the NSD less than 1) of
347 column abundance of most gases. R values range between 0.8 and 0.9 (R^2 between 0.64 and
348 0.81) for most species, indicating the ability of all models in reproducing the spatial pattern of
349 column abundances. The large amplitude of SO_2 variability indicates potential issues associated
350 with aqueous chemistry of all models and high uncertainties with satellite retrievals. The large
351 variability for HCHO from EC may be associated several reasons as discussed earlier, in

particular its use of ethene as a surrogate for monoterpenes (Makar et al., 2014). Overall, the performance for CO columns is the best by all models, followed by NO_2, HCHO, TOR, and SO_2.

3.2. Aerosol and Cloud Variables

Figure 3 shows the spatial distribution of selected aerosol/cloud related variables (i.e., AOD, CF, LWP, and PWV) between satellite observations and predictions by different simulations for 2006. Figure A1 shows COT, CCN, and CDNC. LWP and CCN from observations are only available over the ocean. The domainwide statistics are summarized in Table 1. All simulations exhibit a systematic and large underprediction of AOD over the western U.S. and the spatial distributions are also quite different compared to MODIS AOD. The most noticeable differences are in western U.S. and northern Mexico, where simulations fail to capture the high level of MODIS AODs (up to 0.45) and are lower by factors of 3-4 than MODIS observations. The AOD underpredictions off the west coast are the least pronounced for UPM. Since this area is located close to the domain boundaries and AOD can be affected by the trans-Pacific transport of Asian air pollutants and dust storms, the differences in model performance suggest that different approaches used to map the aerosol boundary conditions from MACC to the regional models contribute to the model biases in AOD predictions, in particular when the representation of aerosol size bins used in MACC differs from that used in the regional model. Contrasting to western U.S., all simulations better estimate the MODIS AOD over eastern U.S., where anthropogenic aerosol loadings are high. The domainwide NMBs are -56.7% (EC), -35.8% (NCSU), -34.9% (EPA), and -3.8% (UPM), respectively. The model biases well exceed the uncertainties associated with MODIS retrievals (see Table A1) and several possible reasons may help explain the discrepancies between MODIS observations and simulations over the western part of domain. First, simulated AOD depends strongly on $PM_{2.5}$ mass concentration

predictions. The inaccurate prediction of $PM_{2.5}$ loadings, particularly from the dust emissions, may lead to the underprediction of AODs over the arid areas. Two out of four simulations (i.e., UPM and EC) lack of dust emissions and the other two (i.e., NCSU and EPA) may simply underpredict dust emissions. Second, higher uncertainties exist for MODIS AOD over the deserts of western U.S. and northern Mexico. A recent work by Drury et al. (2008) found that high positive biases of MODIS AOD exist over the above desert areas caused by some errors in the surface reflectance estimates from the MODIS retrieval algorithms. Using their improved AOD retrievals, they produced much lower AOD.

All simulations for which cloud fractions were submitted reproduce the spatial distribution of MODIS CF well with high values (>0.7) over the oceans, southeastern Canada, and northeastern U.S and low values (<0.4) over the mountainous areas of western U.S. and Mexico. All three simulations can also reproduce the magnitude of MODIS CF well with NMBs of -2.8% (NCSU), -2.4% (EPA), and 0.5% (UPM). All three simulations capture the high values (>75 g cm^{-2}) and general distribution of LWP off the Atlantic coasts and Pacific Northwest, but the magnitude is less than the satellite retrieval. The predicted pattern for LWP is correlated with CF. All simulations underpredict LWP with NMBs of -34.7% (EPA), -28% (NCSU) and -22.6% (UPM), which is mainly caused by the limitations in the cloud parameterizations of WRF/Chem for NCSU and UPM such as the inaccurate contribution of convective clouds to LWP (Zhang et al., 2012) and aerosol-cloud interaction treatments such as uncertainties associated with the Abdul-Razzak and Ghan (2002) scheme (AG) and the missing aerosol indirect effects in WRF-CMAQ for EPA. All three simulations show good agreement of PWV with MODIS retrievals in terms of both spatial distribution and magnitude. Consistent spatial gradients of PWV are shown between simulations and observations, with high values in low

latitude/altitude regions and low values in high latitude/altitude regions. The domainwide NMBs are -1.4% (NCSU), -0.2% (UPM), and 1.3% (EPA), respectively. The general pattern of PWV does not closely correlate with aerosol loadings and cloud covers (as demonstrated by AOD, CF, and LWP). This is due to the fact that on the regional scale PWV is largely a function of synoptic-scale meteorology rather than aerosols/cloud processes (Ten Hoeve et al., 2011).

As shown in Figure A1, COT is largely underpredicted by NCSU due to the missing COTs contributed by rain, snow, and graupel from WRF/Chem (Zhang et al., 2012). Another reason may be due to the underprdiction of LWP, which is ultimately determined by underprediction of aerosol loading and uncertainties in the cloud schemes and aerosol-cloud interaction parameterizations as mentioned earlier. Both NCSU and UPM underpredict CCN, in particular along the Atlantic coasts. Due to the fact that CCN is highly related to the amount of aerosols available for activation, the model underpredictions of CCN likely are caused by an underprediction of aerosol loadings and potential inaccurate representation of land-ocean interactions, which transport too little aerosols to marine areas. The result contrasts with the study by Zhang et al. (2012), in which too high CCN was predicted off the Atlantic coasts due to too strong transport of continental polluted air. Zhang et al. (2012) predicted much higher wind speeds compared to this study (Yhaya et al., 2014). Compared to CCN, the performance for CDNC is better for NCSU and UPM. All three simulations underpredict MODIS CDNC, with the lowest values (domainwide average of ~39 cm^{-3} for EC vs. ~93-121 cm^{-3} for NCSU and UPM) and a different spatial pattern by EC. MODIS, NCSU, and UPM all show high CDNC over the midwest, eastern U.S., and Atlantic Ocean. Since CDNC has substantial impacts on other predicted cloud properties such as COT and LWP, the results shown here are consistent with the underprediction of other variables. Besides the limitations associated with cloud

schemes, the uncertainties related to the aerosol activation scheme (i.e., AG scheme) for both WRF/Chem and GEM-MACH simulations may be another contributor to the underprediction of CDNC. Several studies (e.g., Ghan et al., 2011; Zhang et al., 2012; Gantt et al., 2014) showed that an aerosol activation parameterization based on Fountoukis and Nenes (2005) and its recent updates can give higher CDNC due to a higher activation fraction of aerosols, which should be considered in future model development to improve the model performance of CDNC, COT, and LWP.

3.3. Radiation Variables

Figure 4 shows the spatial distribution of radiation variables (i.e., SWDN, LWDN, and OLR) between satellite observations and 2006 simulations. All simulations reproduce the spatial distributions well for all three radiation variables with values decreasing with increasing latitude, which is driven by the strength of solar radiation. For SWDN, high values are also displayed at higher elevations due to less scattering of incoming solar radiation by atmospheric components. For LWDN, the high values at lower latitudes and low values over the Rocky Mountains correlate very well (with $R^2 > 0.96$, see Table 1) with high and low cloud coverage over those areas (see CF plots in Figure 3). The pattern of OLR is different from LWDN because of the larger impact of high level clouds on OLR. Overall, SWDN is slightly overpredicted by all simulations with NMBs of 0.4% (UPM), 2.6% (EC), 4.3% (NCSU), and 5.4% (EPA). LWDN and OLR are slightly underpredicted except for OLR of EPA with NMBs of -1.9% and -1.3% (NCSU), -0.3% and -2.2% (UPM), and -1.6% and 0.4% (EPA). It should be noted that the simulated aerosol/cloud properties play an important role in affecting the performance of radiation through aerosol direct and indirect effects. The overpredictions of SWDN and underpredictions of LWDN and OLR can be mainly attributed to underpredictions of AOD (due

to less scattering of solar radiation leading to higher SWDN), CF, COT, and LWP (due to less clouds that lead to less emissions of longwave radiation and less trapping of outgoing longwave radiation).

Figure 2b shows the Taylor diagram for selected radiation/aerosol/cloud related variables from four simulations in 2006. There are some outliers including the AOD from two simulations (i.e., NCSU and EPA) due to negative correlation and LWP from EC due to a large NSD. All simulations show a good agreement for SWDN, LWDN, ORL, and PWV. Simulations generally overestimate the amplitude of variability for SWDN (except NCSU), LWDN, and CF and underestimate it for most of other variables. Correlation is excellent for SWDN, LWDN, OLR, and PWV (typically > 0.9) and good for CF and LWP (typically between 0.6 and 0.9), which is consistent with Figures 1 and3-4. The negative correlation for AOD is mainly caused by the large overpredictions over western U.S. and slightly underprediction over eastern U.S. Overall, the results show the high uncertainties in simulating many cloud related variables and further model improvement for the related physical/chemical treatments (e.g., aerosol activation scheme and aqueous-phase chemistry scheme) is warranted.

4. Model Evaluation for 2010 and Its Comparison with 2006

4.1. Column Mass Abundance

Figure 5 shows the spatial distribution of tropospheric column abundances for CO, NO_2, HCHO, SO_2, and TOR between satellite observations and four 2010 simulations. The corresponding performance statistics are given in Table 2. Similar to 2006, all simulations can capture the spatial distribution of MOPITT CO columns well (e.g., they match the high and low abundances areas well). Most simulations underpredict CO columns with NMBs of -12.1% (EPA), -10.0% (NCAR), and -9.4% (NCSU) except for EC which has an NMB of 4.6%. The

potential reasons for the model biases have been discussed in Section 3.1. Compared to 2006, MOPITT CO columns are higher over the Pacific Northwest and southern Canada in 2010, indicating stronger trans-Pacific transport of Asian air pollutants in 2010, which is not well captured by most simulations except for EC. This finding suggests the importance of higher vertical resolution in free troposphere in simulating long lifetime species such as CO. Similar to 2006, the locations of hot spots associated with high NO_x emissions are well reproduced by all 2010 simulations. However, all simulations moderately or largely overpredict the NO_x abundances with NMBs of 12.9% (EPA), 31.8% (NCSU), 91.6% (EC), and 102.1% (NCAR). The domain-average reduction of SCIAMACHY NO_2 columns from 2006 to 2010 is ~18%, which agrees well with the reported NO_x emission reduction of 22% between 2006 and 2010 in EPA's NEI (Stoeckenius et al., 2014). Such a reduction is also reflected in the changes of simulated NO_2 columns for NCSU and EPA, by ~6% and ~10%, respectively. For HCHO, both SCIAMACHY and all four simulations show high column abundances over regions with high biogenic and biomass burning emissions in 2010, which is similar to2006. HCHO columns are underpredicted by NCSU and EPA with NMBs of -25.0% and -10.9%, while they are overpredicted by NCAR and EC with NMBs of 14.2% and 87.6%. The inter-model variability is likely caused by the differences in both biogenic emissions and gas-phase mechanisms. As discussed in Section 3.1, although BEIS used by EC and EPA predict higher HCHO emissions, NCAR predicts an order of magnitude higher isoprene emissions (i.e., 7.2 kton-C km^{-2} $year^{-1}$ vs. 0.02-0.58 kton-C km^{-2} $year^{-1}$) than other simulations (Im et al., 2014a), which lead to the overprediction of HCHO. For EC, both higher HCHO emissions and larger formation of HCHO through ADOM-II (see Section 3.1) result in the large overprediction of HCHO. Two major factors determine the annual changes of HCHO columns from SCIAMACHY observations

between 2010 and 2006. One factor is the change of meteorology. 2010 is considered as a general warmer year compared to 2006. Yahya et al. (2014b) found that the annual average surface temperature over the Clean Air Status and Trends Network (CASTNET) network increased from 11.7°C in 2006 to 15.9°C in 2010. The increase of temperature will increase the biogenic emissions thus leading to more HCHO. The other factor is the change of anthropogenic emissions. Stoeckenius et al. (2014) reported an overall reduction of anthropogenic VOC emissions from 2006 to 2010 that can lead to less HCHO. The two factors may compensate each other and thus create the interesting pattern for SCIAMACHY HCHO as shown in Figures 1 and 5, i.e., larger maximum HCHO over southeastern U.S. but lower domainwide mean values in 2010. Among the three simulations with both 2006 and 2010 results, only EPA reproduces this pattern. For SO_2, three simulations (i.e., NCSU, NCAR, and EPA) present very similar SO_2 columns in terms of both magnitude and spatial distribution while EC presents much higher SO_2 columns. All simulations miss some major hot spots over the western part of domain and oceans observed by SCIAMACHY, possibly due to missing source of SO_2 emissions (e.g., ship emissions) or uncertainties in retrievals. Most simulations underpredict SO_2 with NMBs of -65.2% (NCSU), -65.6% (NCAR), and -60.2% (EPA) except for EC that overpredicts it with an NMB of 7.4%. NCSU predicts the lowest SO_2 columns again in 2010 due to treatments of both SO_2 heterogeneous chemistry and convective cloud AQ chemistry as discussed in Section 3.1. The increasing trend shown in SCIAMACHY SO_2 columns between 2010 and 2006 contradicts with the reported SO_2 emissions reduction by ~40% from 2006 to 2010 by Stoeckenius et al. (2014) and suggests that further investigation of satellite retrievals is needed, considering a rigorous enforcement of SO_2 emission control programs in North America (Pouliot et al., 2014). All three simulations show overpredictions of TORs NMBs of 13.5% (EC), 19.3 (NCSU), and

43.7% (NCAR), respectively, which are due to uncertainties associated with O_3 profiles provided by MACC and emphasize the needs for carefully dealing with O_3 profiles in future studies. The spatial pattern from NCAR is different with both other simulations and OMI/MLS is due to the coarser vertical resolution especially between 350-200 hPa, where it cannot resolve the tropopause from NCEP data well.

Figure 6a shows the Taylor diagram for four gases from four simulations in 2010. Unlike 2006, the amplitude of SO_2 variability in 2010 is reduced due to much higher SO_2 columns from observations, despite lower correlations (~0.2-0.3 in 2010 vs. 0.5-0.6 in 2006) caused by much lower simulated SO_2 over western U.S. and oceans. Generally, the performance for CO is still the best in 2010 among all gases followed by HCHO, NO_2, TOR, and SO_2. A generally poorer performance is found for all species compared to 2006, particularly for NO_2 with two simulations becoming outliers and for HCHO with one simulation becoming an outlier.

4.2. Aerosol and Cloud Variables

Figure 7 shows the spatial distribution of selected aerosol/cloud related variables (i.e., AOD, CF, and PWV) between satellite observations and five 2010 simulations. Figure A2 shows the remaining variables (i.e., COT, CCN, CDNC). The domainwide statistics are summarized in Table 2. All simulations demonstrate a similar systematic underprediction of AOD over western U.S. shown in the 2006 simulations. NCSU and EPA slightly overpredict AOD and EC slightly underpredicts it over eastern U.S. NCAR shows a factor of two overprediction due to large overpredictions of dust contributions to $PM_{2.5}$ (Im et al., 2014b). The domainwide NMBs are -59.5% (EC), -36.1% (EPA), -29.5% (NCSU), and 42.3% (NCAR), respectively. MODIS AOD retrievals show a general decreasing trend from 2006 to 2010, especially over eastern U.S., likely associated with the reduction of anthropogenic emissions of aerosols and precursors. NCSU,

EPA, and EC reproduce this decreasing trend. The spatial distribution of MODIS CF is generally captured by all simulations. NCSU, EPA, and NCAR also reproduce the magnitude well with NMBs of 0.2%, -5.7%, and -9.1%, respectively, while UMU largely underpredict CF with NMBs of -33.2%. The trend for MODIS CF between 2006 and 2010 is not very apparent with slightly more domain average CF observed in 2010 and both NCSU and EPA reproduce the trend. Similar agreements with MODIS PWV retrievals in terms of both spatial distribution and magnitude are presented in 2010 compared to 2006. The domainwide NMBs are -3.2% (NCAR), -1.3% (UMU), -1.1% (NCSU), and 2.3% (EPA). As shown in Figure A2 NCSU shows similar large underpredictions for COT in 2010 due likely to the same reasons discussed in Section 3.2. MODIS COT shows a decreasing trend from 2006 to 2010 (i.e., domainwide average of 16.0 vs. 15.2; ~5% reduction), which is to a lesser extent captured by NCSU (i.e., 5.26 vs 5.15; ~2% reduction). Similar to 2006, NCSU largely underpredicts CCN in 2010 with an NMB of -68.6%. Despite the large underpredictions in both years, NCSU reproduces the decreasing trend of MODIS CCN from 2006 to 2010 (i.e., with domain averages of 0.34×10^9 cm^{-2} and 0.28×10^9 cm^{-2} in 2006 and 2010 for MODIS vs. 0.13×10^9 cm^{-2} and 0.09×10^9 cm^{-2} for NCSU). Both NCSU and EC also underpredict CDNC for 2010 with NMBs of -37.0% and -66.2%, respectively, and NCSU also reproduces the decrease of CDNC observed by MODIS in 2010 compared to 2006. In general, although relatively large biases still exist for most of the predicted aerosol/cloud variables in 2010, most simulations can capture the inter-annual changes of those variables between 2010 and 2006.

4.3. Radiation Variables

Figure 8 shows the spatial distribution of radiation variables (i.e., SWDN, LWDN, and OLR) between satellite observations and 2010 simulations. The model performance for all

radiation variables from all simulations is generally good in terms of both spatial distribution and magnitude. SWDN is overpredicted by all simulation with NMBs of 1.8% (EC), 2.7% (NCSU), 3.3% (EPA), 14.4% (NCAR), and 18.7% (UMU). LWDN and OLR are underpredicted except for OLR of UMU with NMBs of -0.9% and -0.8% (NCSU), -5.0% and -0.1% (NCAR), -4.1% and 3.9% (UMU), and -1.1% and -0.9% (EPA). Since NCSU, NCAR, and UMU all use the same WRF/Chem model and similar radiation schemes (i.e., either RRTMG or RRTM), the relatively larger overprediction of SWDN by NCAR and UMU should be due to the lower predicted CF comparing to other simulations. The satellite observations show a decrease for SWDN and LWDN and an increase for OLR between 2010 and 2006, which is consistent with the increase of CF. Both NCSU and EPA can reproduce the trend of SWDN but show the opposite trend for LWDN and OLR, possibly due to either uncertainties associated with aerosol indirect effect treatments in WRF/Chem or the missing indirect effects of aerosols in WRF-CMAQ.

Figure 6b shows the Taylor diagram for selected aerosol/radiation/cloud variables from five simulations in 2010. AOD are still outlier points due to their negative correlation. Compared to 2006, the overall performance for SWDN, LWDN, and OLR are slightly better. The performance for PWV is slightly worse. The performance for CF and CDNC is generally comparable. Overall, the performance for radiation variables is still the best in 2010, followed by PWV, CF, CDNC, and AOD.

5. Conclusions

In this study, a comparative evaluation is performed for simulations of 2006 and 2010 over the NA domain using three state-of-the-science online-coupled models (i.e., WRF/Chem, WRF-CMAQ, and GEM-MACH). A number of variables evaluated include column-integrated

gas abundances (i.e., tropospheric CO, NO_2, HCHO, SO_2, and TOR), aerosol and cloud properties (i.e., AOD, COT, CF, CCN, CDNC, LWP, and PWV), and radiation budgets (i.e., SWDN, LWDN, and OLR) against available satellite retrieval data (i.e., MOPITT, SCIAMACHY, MODIS, CERES, and AVHRR).

The comparison results show that all simulations can reproduce the MOPITT CO columns well with low biases and high correlations for both years. Larger discrepancies exist for NO_2, HCHO, SO_2 column abundances and TOR possibly due to several reasons including uncertainties in emissions for NO_2 and HCHO and simulated PBL mixing processes, missing model treatments such as plume-in-grid processes, uncertainties associated with BC/profiles for O_3, and uncertainties associated with satellite retrievals algorithms themselves. Inter-model variability is also more apparent for abundances of NO_2, HCHO, and SO_2 than CO due to several possible reasons such as different oxidation rates caused by different gas-phase mechanisms and different treatments of aerosol chemistry (e.g., AQ chemistry and heterogeneous chemistry). For example, the lowest SO_2 columns simulated by NCSU in both 2006 and 2010 are mainly due to the inclusion of both heterogeneous chemistry of SO_2 on aerosol particles and convective cloud chemistry in their model. NCSU, EPA, and EC simulations are performed for both years, which enable a comparison for the simulated inter-annual trend from 2006 to 2010 with that of satellite observations. Both NCSU and EPA are able to reproduce the reduction of SCIAMACHY NO_2 columns in 2010 caused by decreasing emissions compared to 2006. Among the three 2010 simulations, only EC captures the high MOPITT columns caused by the stronger trans-Pacific transport of Asian air pollutants in 2010 than 2006 and only EPA captures the trend of SCIAMACHY HCHO columns caused by the increase of biogenic emissions and decrease of anthropogenic emissions. SCIAMACHY shows an increasing trend of SO_2 column abundances

605 from 2006 to 2010, which is inconsistent with reported reductions of SO_2 emissions and the
606 resultant decreases in simulated SO_2 column abundances. Such an inconsistency is more likely
607 caused by uncertainties in the satellite data retrieval algorithms than uncertainties in the SO_2
608 emissions used in all model simulations, given a rigorous enforcement of SO_2 emission control
609 programs in North America.

610 Most simulations tend to underpredict most aerosol/cloud related variables due to
611 underpredictions of aerosol loadings, inaccurate treatments associated with aerosol-cloud
612 interactions, and uncertainties of satellite data. For example, all simulations significantly
613 underpredict AOD over the western part of the domain, but this could be the result of either
614 underestimation of dust aerosols or positive biases associated with MODIS retrievals. All
615 simulations also tend to significantly underpredict COT, CCN, and CDNC with NMBs generally
616 between -70% to -30% due to the underprediction of aerosol loadings, uncertainties associated
617 with cloud schemes, and potential underpredictions of aerosol activations. However, most
618 simulations perform better in reproducing PWV, due to the fact that it is more dependent on the
619 synoptic-scale meteorology and less dependent on aerosol loadings and cloud covers in the
620 current model treatments. The investigation of inter-annual trend for the above variables shows
621 that most simulations can reproduce the decreasing trend from 2010 to 2006 for variables AOD,
622 PWV, COT, CCN, and CDNC.

623 For radiation variables, all simulations show good agreement with satellite data with
624 NMBs of mostly less than 5%. This indicates good performance of aerosol radiation schemes
625 despite uncertainties still existing in the current model treatments of aerosol/cloud-radiation
626 feedbacks. The feedbacks of aerosols/clouds on radiation are reflected in the general
627 overprediction of SWDN and underprediction of LWDN and OLR, with the former due likely to

the underpredictions of aerosol loadings (e.g., AOD) and the latter due likely to the underpredictions of the magnitudes of cloud properties (e.g., COT, CF, LWP, and CDNC). NCSU, EPA, and EC can reproduce the inter-annual trend of SWDN observed by satellite. Trends in LWDN and OLR are not reproduced by EPA and NCSU possibly due to missing (i.e., EPA) or inaccurate (i.e., NCSU) aerosol indirect effect treatments in the model (LWDN and OLR were not stored in the EC simulations).

While the results in this study provide valuable information on model evaluation against satellite retrievals, this work is subject to several limitations in dealing with the simulation data processing that should be addressed in the future. First, all satellite data used in the work are level 3 data and are subject to higher uncertainties without applying the averaging kernels (AK, which is only available for level 2 data) for column abundances of gases. Therefore, the a priori profiles used by MOPITT, SCIAMACHY, and OMI/MLS retrievals may further contribute to the uncertainties for the comparison (applying AK in the processing of model data would have limited the impacts of the a priori profiles). However, a recent study by Zhang et al. (2010) found that applying AK from the MOPITT retrievals may introduce more noises from the a priori and thus this caveat should be noted for processing column CO. Another study of Schaub et al. (2006) compared ground-based measured NO_2 columns with and without applying the AK from the Global Ozone Monitoring Experiment (GOME) (which uses the similar retrieval methods as SCIMACHY) and found that both methods showed a good agreement with GOME retrievals under the clear sky conditions. Second, the processing of the model results used 100 hPa as a fixed cut-off for the tropopause which may further introduce uncertainties and a more accurate approach should be applied in the future. Third, AOD from different simulations are currently calculated by different methods assuming different preset complex refractive indexes

within individual models. A more consistent way such as using the same offline AOD calculation script but prognostic aerosol outputs from different models should be considered in the future study to allow for a more consistent comparison. Finally, some speculation analyses shown earlier in this study can only be validated through sensitivity simulations. Those simulations are out of the scope of this work and should be addressed in future studies.

Nevertheless, this study provides the first comparative assessment of the capabilities of the current generation of regional online-coupled models in simulating tropospheric columns of major atmospheric components and atmospheric radiation budgets, as well as cloud and aerosol properties. The analyses highlight the strength and deficiencies of current model treatments in simulating chemistry-aerosol-cloud-radiation interactions, in particular, aerosol indirect effects, in current generation of the online-coupled models. The study also identifies several key areas of further investigation and potential model improvements, such as using higher vertical resolution to better represent column abundances and using more advanced aerosol activation parameterization for aerosol-cloud interactions, thus providing the benchmark for future online-couple air quality model development and improvement, as well as re-assessment.

Acknowledgements

This work is supported by NSF Earth System Program (AGS-1049200). We gratefully acknowledge the contribution of various groups to the AQMEII activity: U.S. EPA, Environment Canada, 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) for providing and processing North American national emissions inventories and ECMWF/MACC project & Météo-France/CNRM-GAME for providing MACC model outputs used as chemical initial and boundary conditions in this work. We acknowledge

the free use of satellite data obtained from NASA's and ESA's web sites. The NCSU team would also like to acknowledge high-performance computing support from Yellowstone by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation and Stampede, provided as an Extreme Science and Engineering Discovery Environment (XSEDE) digital service by the Texas Advanced Computing Center (TACC). The UPM team thankfully acknowledges the computer resources, technical expertise and assistance provided by the Centro de Supercomputación y Visualización de Madrid (CESVIMA) and the Spanish Supercomputing Network (BSC). The views expressed here are those of the authors and do not necessarily reflect the views and policies of the U.S. EPA. This paper has been subjected to EPA review and approved for publication. Thanks are also due to Jian He from NCSU for providing the script for Taylor diagram. We also thank two reviewers for their comments to improve the manuscript.

References

Abdul-Razzak, H. and S. J. Ghan (2002), A parameterization of aerosol activation. 3. Sectional representation, J. Geophys. Res., 107(D3), 4026,doi:10.1029/2001JD000483.

Alapaty et al. (2012), New Directions: Understanding interactions of air quality and climate change at regional scales, Atmos. Environ. 49, 419-421.

Baklanov et al. (2014), Online Coupled Regional Meteorology-Chemistry Models in Europe: Current Status and Prospects, *Atmospheric Chemistry and Physics*, 14, 317-398, doi:10.5194/acp-14-317-2014.

Bennartz, R. (2007), Global assessment of marine boundary layer cloud droplet number concentration from satellite, J. Geophys. Res., 112, D02201, doi:10.1029/2006JD007547.

Bian, H., M. Chin, S. R. Kawa, B. Duncan, A. Arellano, and P. Kasibhatla (2007), Sensitivity of global CO simulations to uncertainties in biomass burning sources, J. Geophys. Res., 112, D23308, doi:10.1029/2006JD008376.

Boersma, K. F., H. J. Eskes and E. J. Brinksma (2004), Error analysis for tropospheric NO_2 retrieval from space, J. Geophys. Res. 109, D04311, doi:10.1029/2003JD003962.

Campbell et al. (2014), A multi-model assessment for the 2006 and 2010 simulations under the Air Quality Model Evaluation International Initiative (AQMEII) Phase 2, Indicators of the sensitivity of O_3 and $PM_{2.5}$ formation to precursor gases over North America, in review.

Carlton, A. and K. Baker (2011), Photochemical modeling of the Ozark isoprene Volcano: MEGAN, BEIS, and their impacts on air quality predictions, Environ. Sci. Technol., 45, 4438-4445.

Carlton, A., P. Bhave, S. L. Napelenok, E. O. Edney, G. Sarwar, R. W. Pinder, G. A. Pouliot, and M. Houyoux (2010), Model representation of secondary organic aerosol in CMAQv4.7, Environ. Sci. Technol., 44, 8553–8560.

Choi, Y. (2011), The uncertainty analysis of National Emission Inventory (NEI) 2005 NO_x emissions over the lower middle United States by utilizing top-down approach, International Workshop on Air Quality Forecasting Research, Potomac, MD, Nov 28-Dec 1, 2011.

De Smedt, I., J.-F. Müller, T. Stavrakou, R. J. van der, A, H. J. Eskes, and M. Van Roozendael (2008), Twelve years of global observations of formaldehyde in the troposphere using GOME and SCIAMACHY sensors. Atmos. Chem. Phys., 8(16), 4947-4963.

Drury, E., D. J. Jacob, J. Wang, R. J. D. Spurr, and K. Chance (2008), Improved algorithm for MODIS satellite retrievals of aerosol optical depths over western North America, *J. Geophys. Res., 113,* D16204, doi:10.1029/2007JD009573.

Emmons, L. K., D. P. Edwards, M. N. Deeter, J. C. Gille, T. Campos, P. Nédélec, P. Novelli, and G. Sachse (2009), Measurements of Pollution In The Troposphere (MOPITT) validation through 2006, Atmos. Chem. Phys., 9(5), 1795–1803, doi:10.5194/acp-9-1795-2009.

Engel-Cox J. A., M. H. Raymond and A. D. J. Haymet (2004), Recommendations on the use of satellite remote-sensing data for urban air quality, Journal of the Air & Waste Management Association, 54:11, 1360-1371, doi:10.1080/10473289.2004.10471005.

Fountoukis, C. and A. Nenes (2005), Continued development of a cloud droplet formation parameterization for global climate models, J. Geophys. Res., 110, D11212, doi:10.1029/2004jd005591.

Gantt, B., J. He, X. Zhang, Y. Zhang, and A. Nenes (2013), Incorporation of advanced aerosol activation treatments into CESM/CAM5: model evaluation and impacts on aerosol indirect effects, Atmos. Chem. Phys. Discuss., 13, 32291-32325.

Gao, B.-C., and Y. J. Kaufman (2003), Water vapor retrievals using Moderate Resolution Imaging Spectroradiometer (MODIS) near-infrared channels, J. Geophys. Res., 108(D13), 4389, doi:10.1029/2002JD003023.

Ghan, S. J., H. Abdul-Razzak, A. Nenes, Y. Ming, X. Liu, M. Ovchinnikov, B. Shipway, N. Mekhidze, J. Xu, and X. Shi (2011), Droplet nucleation: physically-based parameterizations and comparative evaluation, J. Adv. Model. Earth Syst., 3, M10001, doi:10.1029/2011MS000074.

Gong, S., L. A. Barrie, and J. P. Blanchet (1997), Modeling sea salt aerosols in the atmosphere: 1. Model development, J. Geophys. Res., 102, 3805–3818, doi:10.1029/96JD02953.

Gong, S. L., L. A. Barrie, and M. Lazare (2003), Canadian Aerosol Module (CAM): A size-egregated simulation of atmospheric aerosol processes for climate and air quality models 2. Global sea-salt aerosol and its budgets, J. Geophys. Res., 107, 4779, doi:10.1029/2001JD002004.

Grell G. A,, S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock, and B. Eder, (2005), Fully coupled 'online' chemistry in the WRF model. Atmos. Environ., 39:6957-6976.

Heald, C. L., et al. (2003), Asian outflow and trans-Pacific transport of carbon monoxide and ozone pollution: An integrated satellite, aircraft, and model perspective, J. Geophys. Res., 108(D24), 4804, doi:10.1029/2003JD003507.

Hodzic A, Madronich S, Bohn B, et al. (2007), Wildfire particulate matter in Europe during summer 2003: meso-scale modeling of smoke emissions, transport and radiative effects, Atmos. Chem. Phys., 7 (15), 4043-4064.

Hodzic, A. and Jimenez, J. L. (2011), Modeling anthropogenically controlled secondary organic aerosols in a megacity: a simplified framework for global and climate models, Geosci. Model Dev., 4, 901-917, doi:10.5194/gmd-4-901-2011.

Huang, M., K. W. Bowman, G. R. Carmichael, R. Bradley Pierce, H. M. Worden, M. Luo, O. R. Cooper, I. B. Pollack, T. B. Ryerson, and S. S. Brown (2013), Impact of Southern California anthropogenic emissions on ozone pollution in the mountain states: Model analysis and observational evidence from space, J. Geophys. Res. Atmos., 118, 12,784–12,803, doi:10.1002/2013JD020205.

Im et al. (2014a), Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone. Atmospheric Environment, in review.

Im et al. (2014b), Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part II: Particulate Matter, Atmospheric Environment, in review

Inness et al., 2013: The MACC reanalysis: an 8 yr data set of atmospheric composition, Atmos. Chem. Phys., 13, 4073-4109, doi:10.5194/acp-13-4073-2013.

Knote, C., Brunner, D., Vogel, H., Allan, J., Asmi, A., Äijälä, M., Carbone, S., van der Gon, H. D., Jimenez, J. L., Kiendler-Scharr, A., Mohr, C., Poulain, L., Prévôt, A. S. H., Swietlicki, E., and Vogel, B. (2011), Towards an online-coupled chemistry-climate model: evaluation of trace gases and aerosols in COSMO-ART, Geosci. Model Dev., 4, 1077-1102, doi:10.5194/gmd-4-1077-2011.

Kondragunta, S., et al. (2008), Air quality forecast verification using satellite data. J. Appl. Meteor. Climatol. 47, 425–442.

Kopacz, M., D. J. Jacob, D. K. Henze, C. L. Heald, D. G. Streets, and Q. Zhang (2009), A comparison of analytical and adjoint Bayesian inversion methods for constraining Asian sources of CO using satellite (MOPITT) measurements of CO columns, J. Geophys. Res., 114, D04305, doi:0.1029/2007JD009264.

Lee, C., R. V. Martin, A. van Donkelaar, G. O'Byrne, N. Krotkov, A. Richter, L. G. Huey, and J. S. Holloway (2009), Retrieval of vertical columns of sulfur dioxide from SCIAMACHY and OMI: Air mass factor algorithm development, validation, and error analysis, J. Geophys. Res., 114, D22303, doi:10.1029/2009JD012123.

Li, J. and H. W. Barker (2005), A radiation algorithm with correlated k-distribution, Part I: local thermal equilibrium, J. Atmos. Sci, 62, 286-309.

Liebmann, B. and C. A. Smith (1996), Description of a complete (interpolated) outgoing longwave radiation dataset, Bulletin of the American Meteorological Society, 77, 1275-1277.

Lurmann, F. W., A. C. Lloyd, and R. Atkinson (1986), A chemical mechanism for use in long-range transport/acid deposition computer modeling, J. Geophys. Res., 91, 10905-10936.

Magi, B. I., P. Ginoux, Y. Ming, and V. Ramaswamy (2009), Evaluation of tropical and extratropical Southern Hemisphere African aerosol properties simulated by a climate model, J. Geophys. Res., 114, D14204, doi:10.1029/2008JD011128.

Makar et al. (2014), Feedbacks between Air Pollution and Weather, Part 1: Effects on Chemistry, Atmospheric Environment, in review.

Martin et al., 2003 Martin, R. V., D. J. Jacob, K. Chance, T. P. Kurosu, P. I. Palmer, and M. J. Evans (2003), Global inventory of nitrogen oxide emissions constrained by space-based observations of NO_2 columns, J. Geophys. Res., 108(D17), 4537, doi:10.1029/2003JD003453.

Martin, R. V. (2008), Satellite remote sensing of surface air quality, Atmospheric Environment, 42(34), 7823-7843.

Martin, R. V., A. M. Fiore, and A. Van Donkelaar (2004), Space-based diagnosis of surface ozone sensitivity to anthropogenic emissions, Geophys. Res. Lett., 31, L06120, doi:10.1029/2004GL019416.

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

Miyazaki, K., H. J. Eskes, K. Sudo, M. Takigawa, M. van Weele, and K. F. Boersma (2012), Simultaneous assimilation of satellite NO_2, O_3, CO, and HNO_3 data for the analysis of tropospheric chemical composition and emissions, Atmos. Chem. Phys., 12, 9545–9579.

Moran M.D., S. Ménard, D. Talbot, P. Huang, P.A. Makar, W. Gong, H. Landry, S. Gravel, S. Gong, L-P. Crevier, A. Kallaur, M. Sassi, 2010. Particulate-matter forecasting with GEM-MACH15, a new Canadian air-quality forecast model. In: Steyn DG, Rao ST (eds) Air Pollution Modelling and Its Application XX, Springer, Dordrecht, pp. 289-292.

Odum, J. R., Jungkamp, T. P. W., Griffin, R. J., Flagan, R. C., and Seinfeld, J. H. (1996), The atmospheric aerosol-forming potential of whole gasoline vapour, Science, 276, 96–99.

Pouliot, G., H. D. van der Gon,, J. Kuenen, P. Makar, J. Zhang, and M. Moran (2014), Analysis of the Emission Inventories and Model-Ready Emission Datasets of Europe and North America for Phase 2 of the AQMEII Project, Atmospheric Environment, in review.

Remer, L. A., et al. (2005), The MODIS aerosol algorithm, products, and validation, J. Atmos. Sci., 62, 947–973, doi:10.1175/JAS3385.1.

Saide, P. E., G. R. Carmichael, Z. Liu, C. S. Schwartz, H. C. Lin, A. M. da Silva, and E. Hyer (2013), Aerosol optical depth assimilation for a size-resolved sectional model: impacts of observationally constrained, multi-wavelength and fine mode retrievals on regional scale analyses and forecasts, Atmos. Chem. Phys., 13, 10425–10444, doi:10.5194/acp-13-10425-2013.

Sandu, A. and T. Chai (2011), Chemical data assimilation – an overview, Atmosphere, 3, 426–463, 2011.

Schaub, D., K. F. Boersma, J. W. Kaiser, A. K.Weiss, D. Folini, H. J. Eskes, and B. Buchmann (2006), Comparison of GOME tropospheric NO2 columns with NO_2 profiles deduced from ground-based in situ measurements, Atmos. Chem. Phys., 6, 3211–3229.

Shaw, P. (2008), Application of aerosol speciation data as an in situ dust proxy for validation of the Dust Regional Atmospheric Model (DREAM), Atmos. Environ., 42, 7304–7309, doi:10.1016/j.atmosenv.2008.06.018.

Song, C. H., M. E. Park, K. H. Lee, H. J. Ahn, Y. Lee, J. Y. Kim, K. M. Han, J. Kim, Y. S. Ghim, and Y. J. Kim (2008), An investigation into seasonal and regional aerosol characteristics in East Asia using model-predicted and remotely-sensed aerosol properties, Atmos. Chem. Phys., 8, 6627–6654.

Stoeckenius, T., C. Chemel, J. Zagunis, T. Sturtz, and T. Sakulyanontvittaya (2014), A Comparison between 2010 and 2006 Air Quality and Meteorological Conditions, and Emissions and Boundary Conditions used in Simulations of the AQMEII-2 North American Domain, this issue

Streets, D. G., et al. (2013), Emissions estimation from satellite retrievals: A review of current capability, Atmospheric Environment, 77, 1011-1042.

Tang, Y., et al. (2009), The impact of chemical lateral boundary conditions on CMAQ predictions of tropospheric ozone over the continental United States, Environ Fluid Mech, 9:43–58 doi:10.1007/s10652-008-9092-5.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106(D7), 7183–7192, doi:10.1029/2000JD900719.

Ten Hoeve, J. E., L. A. Remer, and M. Z. Jacobson (2011), Microphysical and radiative effects of aerosols on warm clouds during the Amazon biomass burning season as observed by MODIS: impacts of water vapor and land cover, Atmos. Chem. Phys., 11, 3021-3036.

van Noije, T. P. C., et al. (2006), Multi-model ensemble simulations of tropospheric NO_2 compared with GOME retrievals for the year 2000, Atmos. Chem. Phys., 6, 2943-2979.

Vijayaraghavan, K., H. E. Snell, and C. Seigneur (2008), Practical aspects of using satellite data in air quality modeling, Environ. Sci. Technol., 2008, 42 (22), 8187-8192, doi: 10.1021/es7031339.

Vijayaraghavan, K., Y. Zhang, C. Seigneur, P. Karamchandani, and H. E.Snell (2009), Export of reactive nitrogen from coal-fired power plants in the USA: Estimates from a plume-in-grid modeling study, J. Geophys. Res., 114, D04308, doi:10.1029/2008JD010432.

Wang, K. and Y. Zhang (2012), Application, evaluation and process analysis of U.S. EPA's 2002 multiple-pollutant air quality modeling platform, Atmos. Clim. Sci., 2, 254-289.

Wang, K., Y. Zhang, C. Jang, S. Phillips, and B. Wang (2009), Modeling intercontinental air pollution transport over the trans-Pacific region in 2001 using the Community Multiscale Air Quality modeling system, J. Geophys. Res., 114, D04307, doi:10.1029/2008JD010807.

Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee, G. L. Smith, and J. E. Cooper (1996), Clouds and the Earth's Radiant Energy System (CERES): An earth observing system experiment, Bull. Amer. Meteor. Soc., 77, 853–868.

Wong, D. C., Pleim, J., Mathur, R., Binkowski, F., Otte, T., Gilliam, R., Pouliot, G., Xiu, A., Young, J. O., and Kang, D. (2012), WRF-CMAQ two-way coupled system with aerosol feedback: software development and preliminary results, Geosci. Model Dev., 5, 299-312, doi:10.5194/gmd-5-299-2012.

Yahya, K., K. Wang, M. Gudoshava, T. Glotfelty, and Y. Zhang (2014a), Application of WRF/Chem over the continental U.S. under the AQMEII Phase II: Comprehensive evaluation of 2006 Simulation, this issue.

Yahya, K., K. Wang, Y. Zhang, C. Hogrefe, G. Pouliot, T. Kleindienst (2014b), Application of WRF/Chem over the continental U.S. under the AQMEII phase II: Responses of air quality and meteorology-chemistry interactions to changes in emissions and meteorology from 2006 to 2010, this issue.

Zhang L., D. J. Jacob, X. Liu, J. A. Logan, K. Chance, A. Eldering, and B. R. Bojkov (2010), Intercomparison methods for satellite measurements of atmospheric composition: application to tropospheric ozone from TES and OMI, Atmos. Chem. Phys., 10, 4725–4739, doi:10.5194/acp-10-4725-2010.

Zhang, Y. (2008), Online coupled meteorology and chemistry models: History, current status, and outlook, Atmos. Chem. Phys., 8, 2895–2932, doi:10.5194/acp-8-2895-2008.

Zhang, Y., K. Vijayaraghavan, X.-Y. Wen, H. E. Snell, and M. Z. Jacobson (2009), Probing into regional ozone and particulate matter pollution in the United States: 1. A 1 year CMAQ simulation and evaluation using surface and satellite data, J. Geophys. Res., 114, D22304, doi:10.1029/2009JD011898.

Zhang, Y., P. Liu, B. Pun, and C. Seigneur (2006), A comprehensive performance evaluation of MM5-CMAQ for the Summer 1999 Southern Oxidants Study episode-Part I: Evaluation protocols, databases, and meteorological predictions, Atmos. Environ., 40, 4825-4838.

Zhang, Y., Y. Chen, G. Sarwar, and K. Schere (2012a), Impact of gas-phase mechanisms on Weather Research Forecasting Model with Chemistry (WRF/Chem) predictions: Mechanism implementation and comparative evaluation, J. Geophys. Res., 117, D01301, doi:10.1029/2011JD015775.

Zhang, Y., P. Karamchandani, T. Glotfelty, D. G. Streets, G. Grell, A. Nenes, F.-Q. Yu, and R. Bennartz (2012b), Development and Initial Application of the Global-Through-Urban Weather Research and Forecasting Model with Chemistry (GU-WRF/Chem), *Journal of Geophysical Research*, 117, D20206, doi:10.1029/2012JD017966.

Ziemke, J. R., S. Chandra, B. N. Duncan, L. Froidevaux, P. K. Bhartia, P. F. Levelt, and J. W. Waters (2006), Tropospheric ozone determined from Aura OMI and MLS: Evaluation of measurements and comparison with the Global Modeling Initiative's Chemical Transport Model, J. Geophys. Res., 111, D19303, doi:10.1029/2006JD007089.

Table 1. Statistics summary for all models in 2006

Species*	Satellite	NCSU			UPM			EPA			EC		
		NMB (%)	NME (%)	R	NMB (%)	NME (%)	R	NMB (%)	NME (%)	R	NMB (%)	NME (%)	R
CO	MOPITT	-9.3	9.7	0.90	-7.7	8.4	0.88	-9.4	9.9	0.83	-2.2	5.3	0.85
NO2	SCIAMACHY	14.1	33.6	0.90	-14.7	33.2	0.93	2.1	34.9	0.86	-37.7	45.2	0.89
HCHO	SCIAMACHY	-24.5	29.0	0.77	-27.3	30.7	0.78	-11.5	30.6	0.51	59.2	59.8	0.71
SO2	SCIAMACHY	16.1	76.5	0.59	26.2	82.3	0.62	42.1	91.0	0.59	114.2	144.7	0.64
TOR	OMI/MLS	38.0	38.0	0.47	29.9	29.9	0.56	--	--	--	19.9	19.9	0.87
OLR	NOAA/CDC	-1.3	2.4	0.93	-2.2	3.3	0.86	0.4	1.5	0.97	--	--	--
LWDN	CERES	-1.9	2.5	0.99	-0.3	2.1	0.98	-1.6	2.0	0.99	--	--	--
SWDN	CERES	4.3	7.2	0.93	0.4	7.0	0.86	5.4	6.1	0.97	2.6	6.0	0.96
AOD	MODIS	-35.8	46.1	-0.02	-3.8	31.5	0.08	-34.9	39.4	-0.04	-56.7	56.7	0.08
COT	MODIS	-64.1	64.1	0.68	--	--	--	--	--	--	195.5	197.0	0.70
CF	MODIS	-2.8	10.4	0.81	0.5	11.6	0.76	-2.4	8.7	0.90	--	--	--
CCN	MODIS	-64.0	64.0	0.52	-48.5	48.9	0.56	--	--	--	--	--	--
CDNC	MODIS	-33.6	47.7	0.18	-16.1	44.7	0.12	--	--	--	-76.3	76.5	0.39
LWP	MODIS	-28.0	29.9	0.67	-22.6	29.8	0.57	-34.7	44.7	0.41	222.2	230.4	0.88
PWV	MODIS	-1.4	8.7	0.97	-0.2	8.5	0.97	1.3	8.7	0.98	--	--	--

*CO, NO_2, HCHO, SO_2, and TOR are all tropospheric abundance with units of 10^{18} molecules cm^{-2}, 10^{15} molecules cm^{-2}, 10^{15} molecules cm^{-2}, DU, and DU, respectively; OLR, LWDN, and SWDN with units of W m^{-2}; AOD, COT, and CF are unitless; CCN with unit of 10^9 cm^{-2}; CDNC with unit of cm^{-3}; LWP with unit of g cm^{-3}; PWV with unit of cm.

-- Simulation results either not available or have issues.

Table 2. Statistics summary for all models in 2010

Species*	Satellite	NCSU			NCAR			UMU			EPA			EC		
		NMB (%)	NME (%)	R	NMB (%)	NME (%)	R	NMB (%)	NME (%)	R	NMB (%)	NME (%)	R	NMB (%)	NME (%)	R
CO	MOPITT	-9.4	9.6	0.93	-10.0	10.2	0.93	--	--	--	-12.1	12.4	0.82	4.6	5.6	0.89
NO2	SCIAMACHY	31.8	42.4	0.89	102.1	105.0	0.89	--	--	--	12.9	38.7	0.81	91.6	101.9	0.76
HCHO	SCIAMACHY	-25.0	33.8	0.69	14.2	32.2	0.69	--	--	--	-10.9	34.0	0.53	87.6	88.3	0.67
SO2	SCIAMACHY	-65.2	71.4	0.31	-65.6	71.5	0.30	--	--	--	-60.2	68.7	0.32	7.4	85.6	0.20
TOR	OMI/MLS	19.3	19.4	0.64	43.7	43.7	-0.20	--	--	--	--	--	--	13.5	14.4	0.78
OLR	NOAA/CDC	-0.8	1.9	0.95	-0.1	2.4	0.91	3.9	4.0	0.97	-0.9	1.6	0.97	--	--	--
LWDN	CERES	-0.9	2.0	0.98	-5.0	5.0	0.99	-4.1	4.2	0.99	-1.1	2.0	0.98	--	--	--
SWDN	CERES	2.7	6.8	0.91	14.4	15.0	0.89	18.7	18.7	0.93	3.3	5.6	0.95	1.8	5.8	0.96
AOD	MODIS	-29.5	42.7	-0.09	42.3	67.5	-0.17	--	--	--	-36.1	43.1	-0.18	-59.5	59.7	-0.08
COT	MODIS	-63.2	63.2	0.60	--	--	--	--	--	--	--	--	--	213.4	214.7	0.54
CF	MODIS	0.2	9.0	0.87	-9.1	13.1	0.74	-33.2	33.2	0.78	-5.7	10.7	0.90	--	--	--
CCN	MODIS	-68.6	68.7	0.49	--	--	--	--	--	--	--	--	--	--	--	--
CDNC	MODIS	-37.0	47.5	0.26	--	--	--	--	--	--	--	--	--	-66.2	67.3	0.36
PWV	MODIS	-1.1	10.1	0.96	-3.2	11.7	0.96	-1.3	11.1	0.96	2.3	10.8	0.96	--	--	--

*CO, NO_2, HCHO, SO_2, and TOR are all tropospheric abundance with units of 10^{18} molecules cm^{-2}, 10^{15} molecules cm^{-2}, 10^{15} molecules cm^{-2}, DU, and DU, respectively; OLR, GLW, and SWDN with units of W m^{-2}; AOD, COT, and CF are unitless; CCN with unit of 10^9 cm^{-2}; CDNC with unit of cm^{-3}; PWV with unit of cm; LWP from MODIS is not available for 2010.
-- Simulation results either not available or have issues.

List of Figures

Figure 5. Spatial distribution of tropospheric column gas abundances (from top to bottom: column CO, column NO_2, column HCHO, column SO_2, and TOR) between satellite observation and different models for year 2010 (blank color denotes to missing values; due the erroneous mapping of O_3 profile, TOR from EPA is not shown).

Figure 6. Taylor diagram with NSD (desire value is 1), R (desired value is 1), and NMB (desire value is 0) for (a) selected column gas species and (b) radiation/aerosol/cloud variables among 5 simulations for year 2010. See Figure 2 caption for the meanings of coordinates and markers in the Taylor diagram.

Figure 7. Spatial distribution of aerosol/cloud related variables (from top to bottom: AOD, CF, and PWV) between satellite observation and different models for year 2010 (blank color denotes to missing values; CF/PWV from model EC and CF from model EPA are not available).

Figure 8. Spatial distribution of radiation (from top to bottom: SWDN, LWDN, and OLR) between satellite observation and different models for year 2010 (blank color denotes to missing values; LWDN and OLR from model EC are not available).

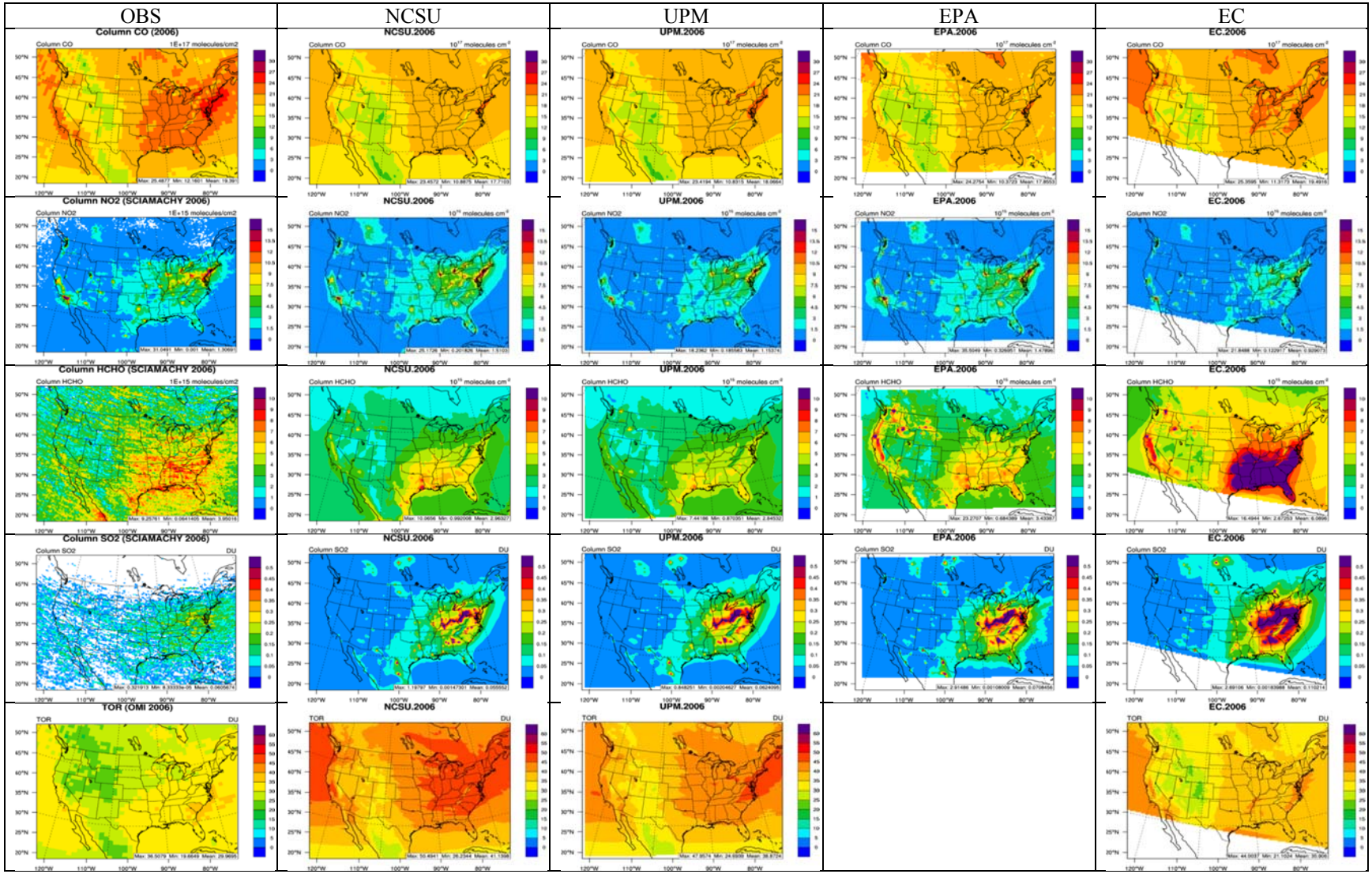

Figure 1. Spatial distribution of tropospheric column gas abundances (from top to bottom: column CO, column NO_2, column HCHO, column SO_2 and TOR) between satellite observation and different models for year 2006 (blank color denotes to missing values; due the erroneous mapping of O_3 profile, TOR from EPA is not shown).

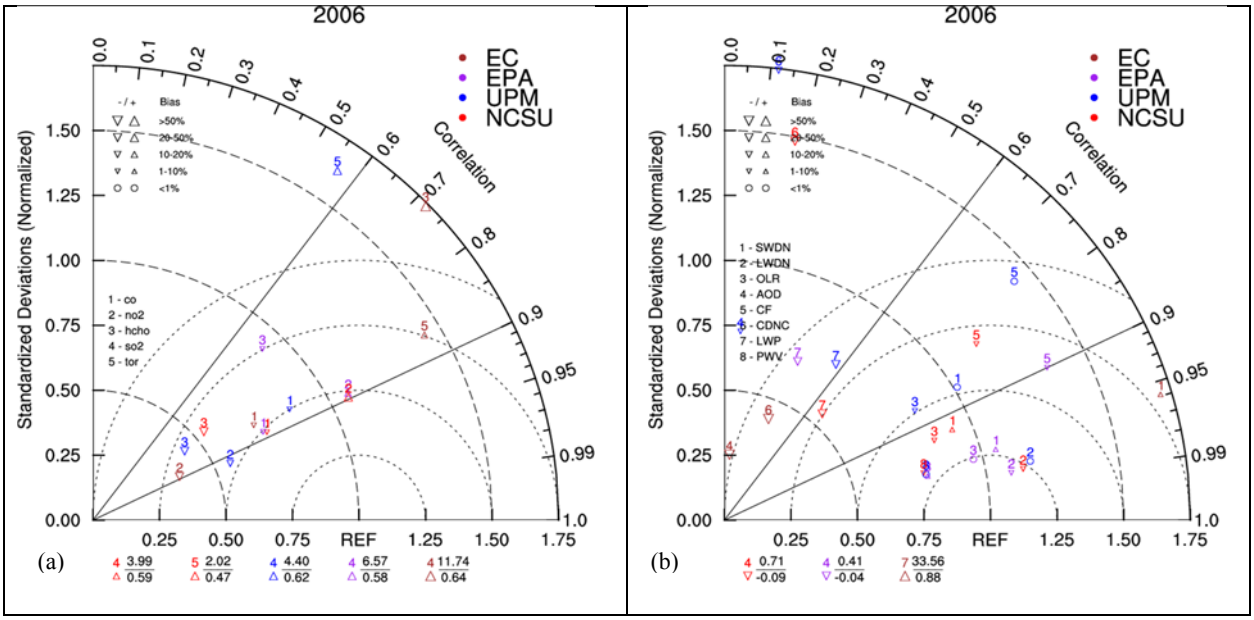

Figure 2. Taylor diagram with NSD (desire value is 1), R (desired value is 1), and NMB (desire value is 0) for selected (a) column gas species and (b) radiation/aerosol/cloud variables among 4 simulations for year 2006. Note that the point marked REF on the X-axis represents the observed field and all markers on the plot area represent the simulation results. The distance between the markers and the REF point is a measure of model performance, with smaller distances indicating better model performance. The closer the markers are to the X-axis, the better the model is able to reproduce the observed spatial pattern. The closer the marker is to the isoline crossing REF (i.e., the NSD is equal to 1), the better the model is able to reproduce the amplitude of variations in the satellite data. The hemispherical lines centered over "REF" on the horizontal axis represent the combined desired level of NSD and correlation values (the closer the markers to the inner hemispherical lines, the better overall model performance in terms of both magnitude and correlation). The size of the markers is proportional to the magnitude of an NMB, with smaller markers indicating smaller NMBs (regular triangle representing positive bias and inverse triangle representing negative biases).

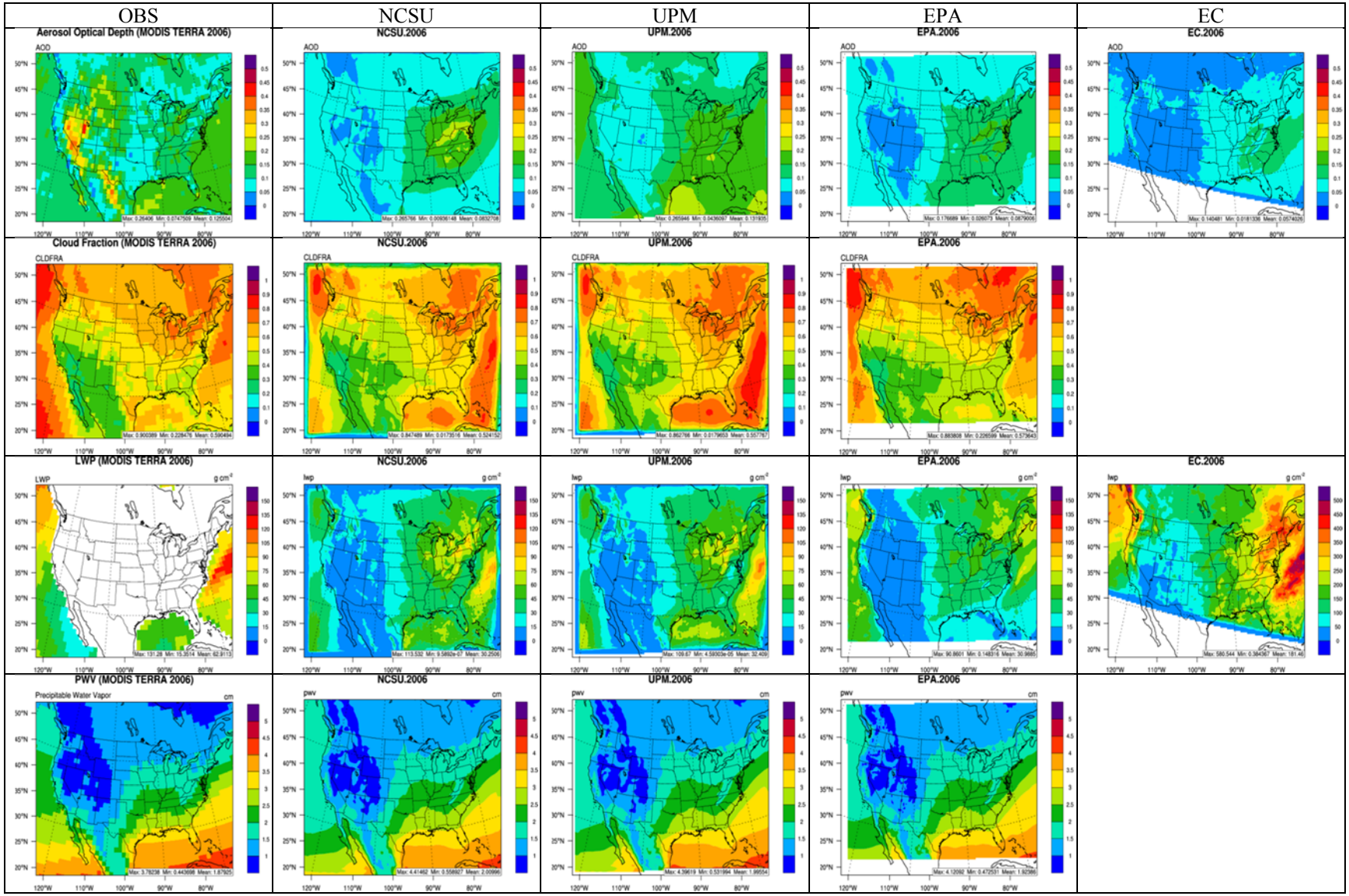

Figure 3. Spatial distribution of aerosol/cloud related variables (from top to bottom: AOD, CF, LWP, and PWV) between satellite observation and different models for year 2006 (blank color denotes to missing values; CF/PWV from model EC and CF from model EPA are not available; scale for LWP of EC is different).

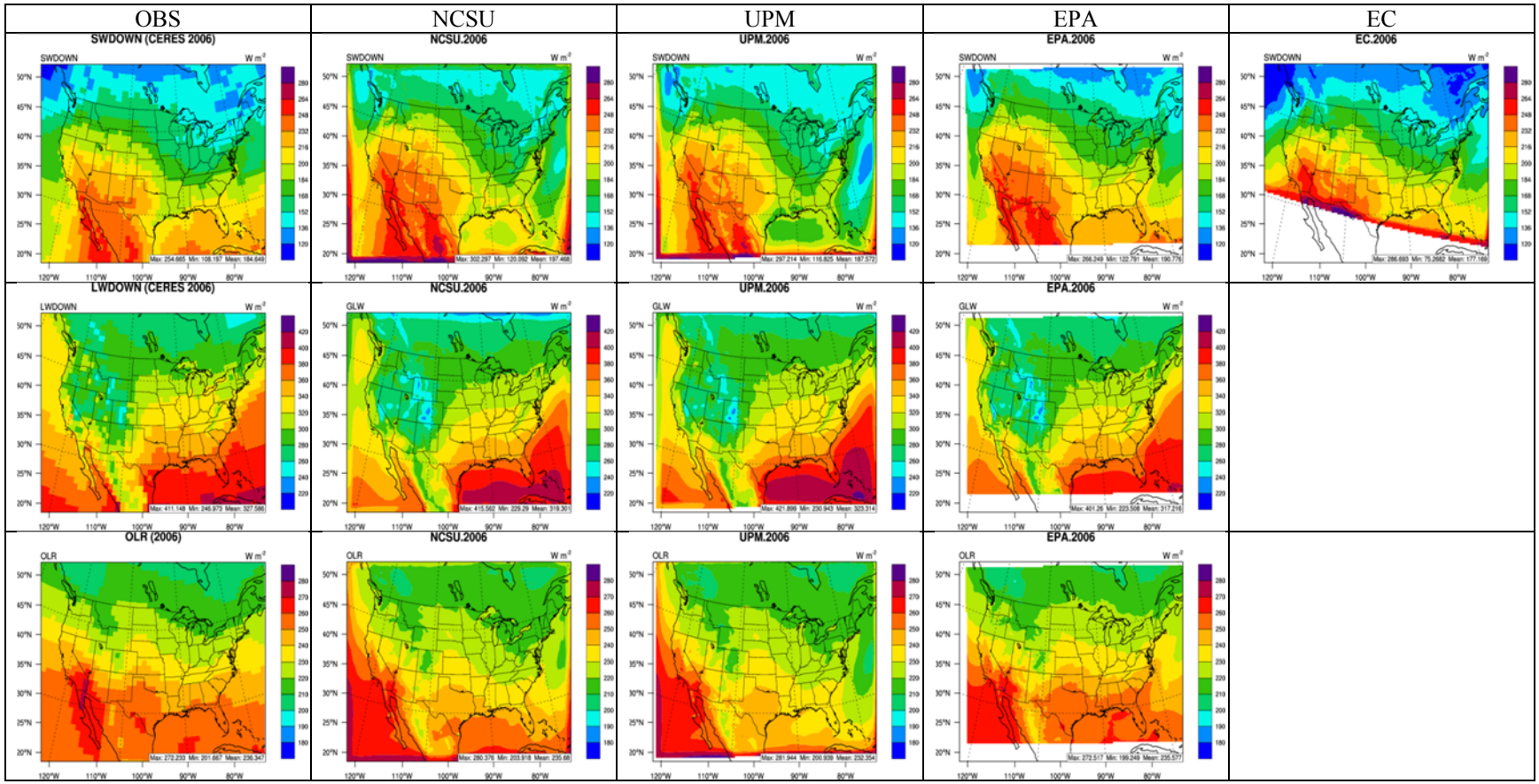

Figure 4. Spatial distribution of radiation (from top to bottom: SWDN, LWDN, and OLR) between satellite observation and different models for year 2006 (blank color denotes to missing values; LWDN and OLR from model EC are not available).

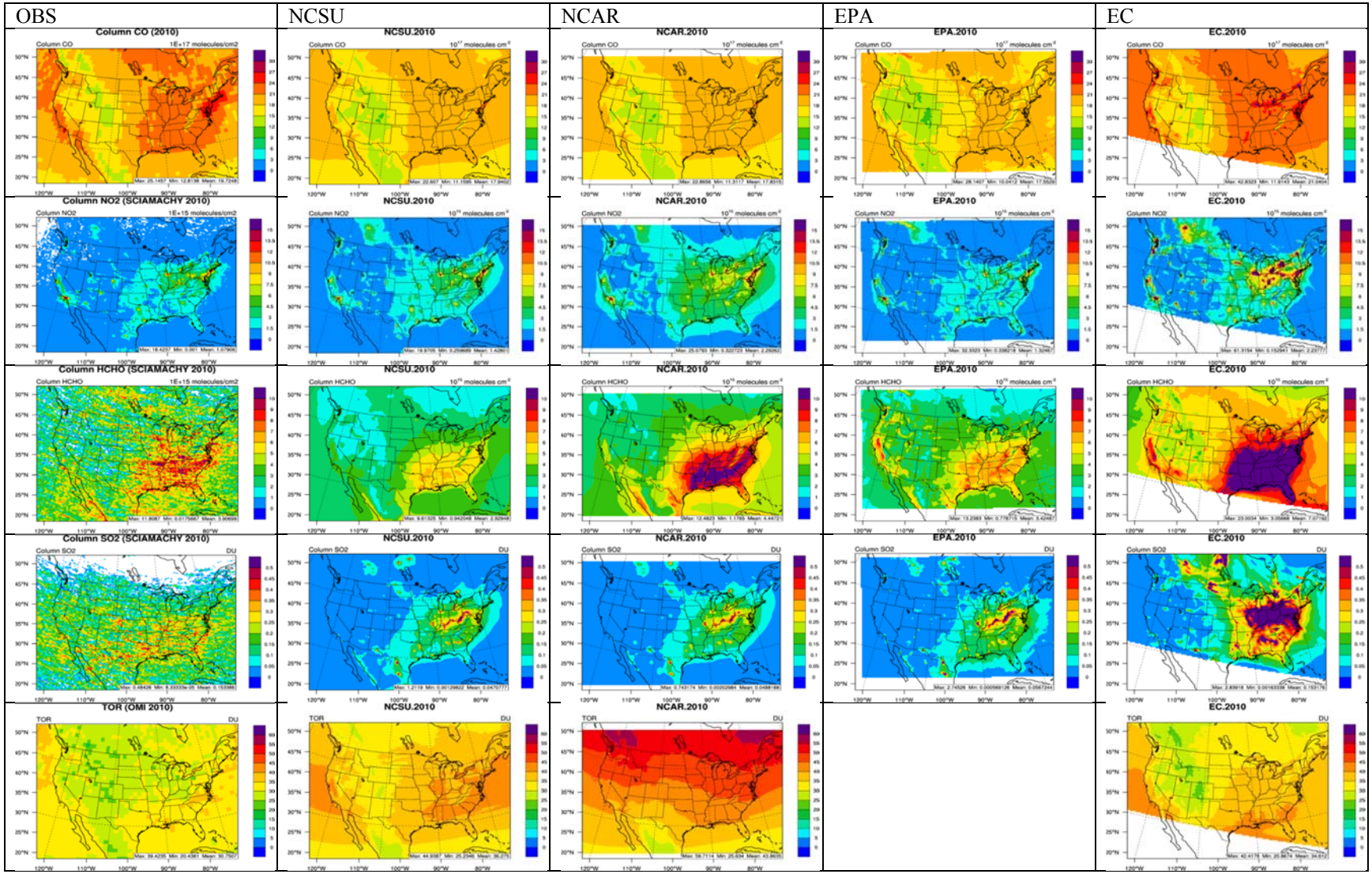

Figure 5. Spatial distribution of tropospheric column gas abundances (from top to bottom: column CO, column NO_2, column HCHO, column SO_2, and TOR) between satellite observation and different models for year 2010 (blank color denotes to missing values; due the erroneous mapping of O_3 profile, TOR from EPA is not shown).

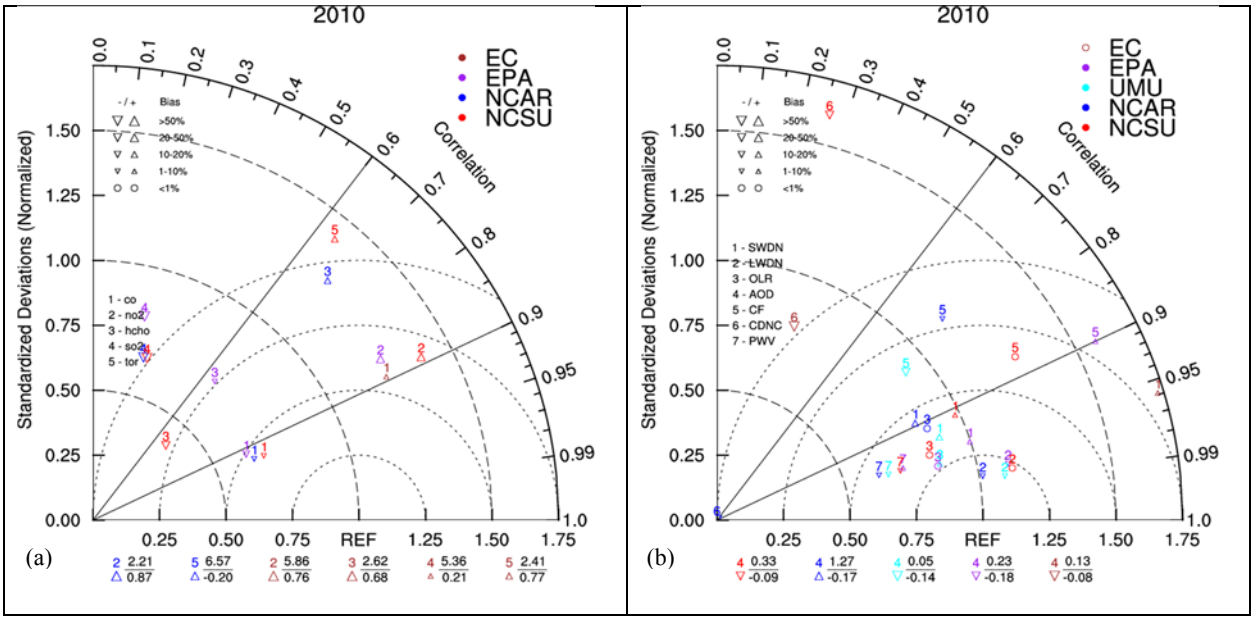

Figure 6. Taylor diagram with NSD (desire value is 1), R (desired value is 1), and NMB (desire value is 0) for (a) selected column gas species and (b) radiation/aerosol/cloud variables among 5 simulations for year 2010. See Figure 2 caption for the meanings of coordinates and markers in the Taylor diagram.

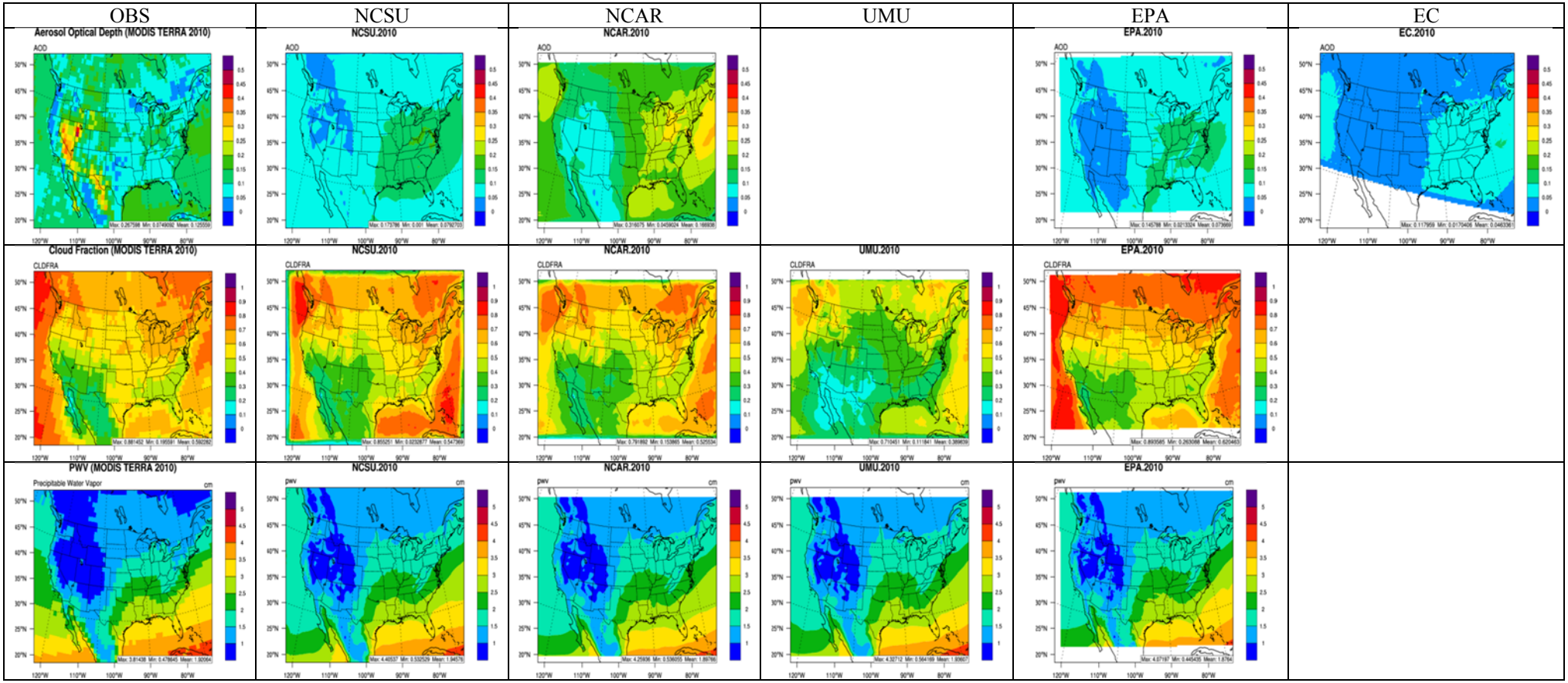

Figure 7. Spatial distribution of aerosol/cloud related variables (from top to bottom: AOD, CF, and PWV) between satellite observation and different models for year 2010 (blank color denotes to missing values; CF/PWV from model EC and CF from model EPA are not available).

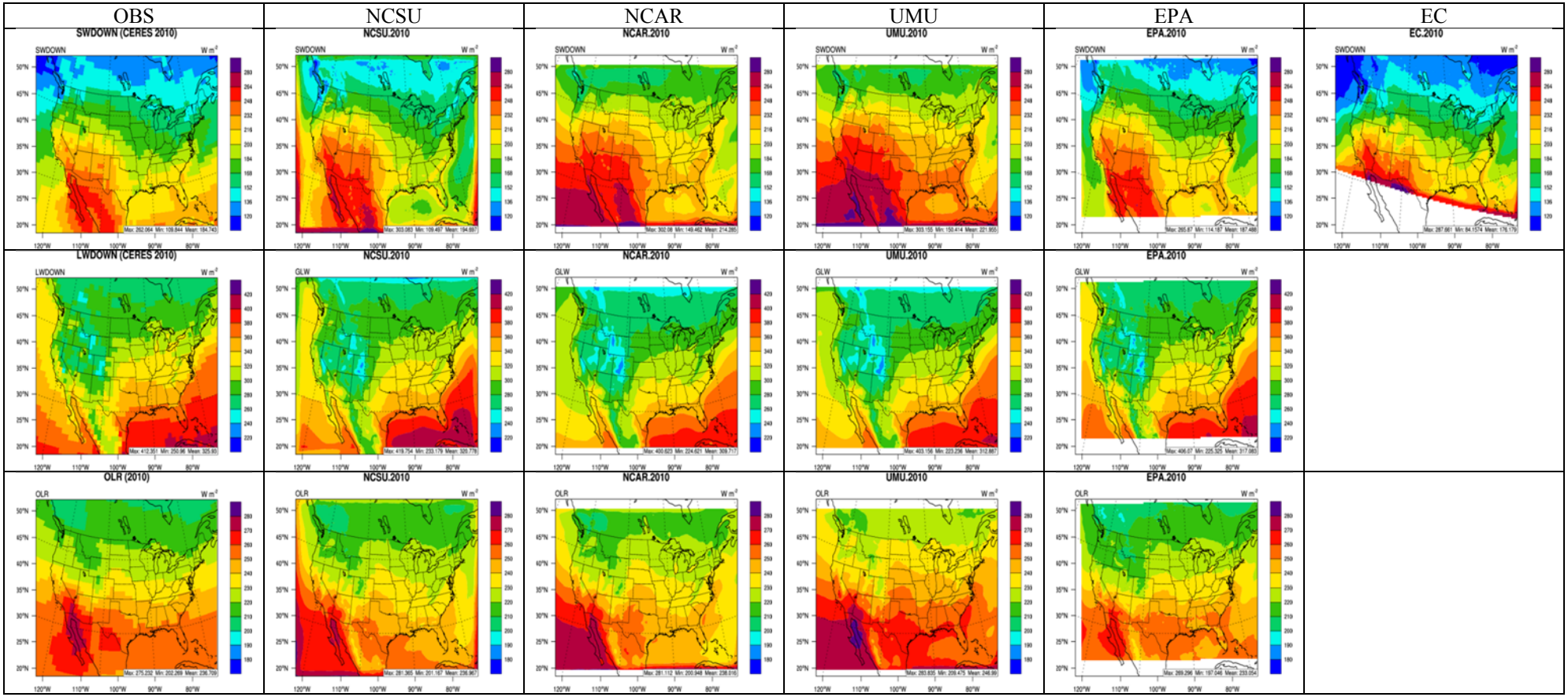

Figure 8. Spatial distribution of radiation (from top to bottom: SWDN, LWDN, and OLR) between satellite observation and different models for year 2010 (blank color denotes to missing values; LWDN and OLR from model EC are not available).

SUPPLEMENTARY MATERIALS

A Multi-Model Assessment for the 2006 and 2010 Simulations under the Air Quality Model Evaluation International Initiative (AQMEII) Phase 2 over North America: Part II. Evaluation of Column Variable Predictions Using Satellite Data

Kai Wang*, Khairunnisa Yahya, and Yang Zhang*

Department of Marine, Earth, and Atmospheric Sciences, NCSU, Raleigh, NC 27695, USA

Christian Hogrefe and George Pouliot

Atmospheric Modeling and Analysis Division, Office of Research and Development, U.S. EPA, RTP, NC 27711, USA

Christoph Knote and Alma Hodzic

Atmospheric Chemistry Division, National Center for Atmospheric Research, Boulder, CO 80301, USA

Roberto San Jose and Juan L. Perez

Computer Science School, Technical University of Madrid, Campus de Montegancedo, Boadilla del Monte, 28660 Madrid, Spain

Pedro Jiménez Guerrero and Rocio Baro

Department of Physics, Ed. CIOyN, Campus de Espinardo, Regional Campus of International Excellence "Campus Mare Nostrum", University of Murcia, 30100 Murcia, Spain

Paul Makar

Air Quality Research Division, Environment Canada, Toronto, Ontario, M3H 5T4, Canada

Ralf Bennartz

Earth & Environmental Sciences, Vanderbilt University, Nashville, TN 37205, USA

Space Science and Engineering Center, University of Wisconsin-Madison, WI 53706, USA

*Corresponding Authors: Tel.: +1 919 515 9688; Fax: +1 919 515 7802

Email Addresses: kwang@ncsu.edu (Kai Wang), yzhang9@ncsu.edu (Yang Zhang)

Sensors/Satellite Description

The MOPITT instrument aboard the NASA's Terra satellite allows retrievals of tropospheric column CO with a horizontal resolution of 22 × 22 km^2 at nadir by detecting the infrared radiation emissions in the 4.7 μm band of CO. The retrieval of MOPITT CO has a target accuracy and precision of 10% (Emmons et al., 2009). The SCIAMACHY instrument aboard the ESA's Environmental satellite (ENVISAT) allows retrievals of a few species including tropospheric column NO_2, HCHO, and SO_2 with a typical horizontal resolution of 30 × 60 km^2 and as low as 30 × 30 km^2 by measuring the backscattered solar radiation with the spectrum bands over 214-1750 nm, 2.0 μm, and 2.3 μm. Compared with an older instrument Global Ozone Monitoring Experiment (GOME) which also detects column NO_2 and HCHO, SCIAMACHY has much higher resolution and thus provide better resolved variability of retrievals. The retrievals from SCIAMACHY have varying uncertainties for different species and are typically higher for SO_2 and HCHO as compared to NO_2 as shown in Table A1. The OMI instrument aboard the NASA's Aura satellite measures the backscattering of solar radiation over the spectrum 270-500 nm with a spatial resolution of 13 × 24 km^2. The TORs are retrieved with a residual technique that uses both total column O_3 from OMI and stratospheric column O_3 from the Microwave Limb Sounder (MLS). The general uncertainties associated with TOR from OMI/MLS are 4-5 Dobson Units (DUs). The OMI instrument aboard the

NASA's Aura satellite measures the backscattering of solar radiation over the spectrum 270-500 nm with a spatial resolution of 13 × 24 km^2. The CERES instrument aboard the NASA's Terra satellite is designed to measure the shortwave and longwave radiation fluxes from the TOA to the Earth's surface. Each CERES instrument is a radiometer which has three channels: a shortwave channel to measure scattered sunlight in the 0.3-5 μm region, a channel in the 8-12 μm regions to measure thermal radiation emissions from the Earth, and a total channel to measure the entire spectrum of Earth's outgoing radiation (Wielicki et al., 1996). The CERES sensor has a high resolution of 20 × 20 km^2 at nadir and relatively low uncertainties for both SWDN (1%) and LWDN (0.5%). The AVHRR sensor on NOAA-18 satellite has five channels sensing in the visible, near-infrared, and thermal infrared spectrum with high resolution up to 1.1 × 4.4 km^2 and has been used to retrieve the TOA OLR. The MODIS instrument also aboard the NASA's Terra satellite provides unprecedented information about aerosol and cloud properties at a very high spatial resolution of 10 × 10 km^2 at nadir (which may reach 1-5 km for some cloud products; Martin, 2008). MODIS has 36 spectral channels (compared to 4-8 for most other sensors), ranging from 0.41 to 15 μm, which bestow MODIS the unique ability to retrieve AOD with much higher accuracy. The MODIS sensor is designed to systematically retrieve aerosol/cloud properties over both land and ocean on a daily basis. The typical AOD retrieval from MODIS is at 550 nm interpolated from two other independent retrievals at 470 and 660 nm. The uncertainties for various products associated with MODIS are also summarized in Table A1.

Statistical Equations

$$NMB = \frac{\sum_{i=1}^{N}(M_i - O_i)}{\sum_{i=1}^{N} O_i} \times 100$$

$$NME = \frac{\sum_{i=1}^{N}|M_i - O_i|}{\sum_{i=1}^{N} O_i} \times 100$$

$$R = \frac{\sum_{i=1}^{N}(M_i - \overline{M})(O_i - \overline{O})}{\left\{\sum_{i=1}^{N}(M_i - \overline{M})^2 \sum_{i=1}^{N}(O_i - \overline{O})^2\right\}^{1/2}}$$

$$NSD = \frac{\sum_{i=1}^{N}(M_i - \overline{M})^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2}$$

Table A1. Summary of satellite data used in the model evaluation.

Variables[a]	Sensors/Satellites[b]	Spatial/Temporal Resolutions/Equator Crossing Time for Raw Measurements	Level-3 Data Spatial Resolution	Time Resolution for Evaluation	Data Uncertainties	References
CO	MOPITT/Terra	22 × 22 km^2/once per day/10:30 am	1° × 1°	Monthly	±10%	Emmons et al. (2009)
NO_2	SCIAMACHY/ENVISAT	As low as 30 × 30 km^2/once per day/10:00 am	0.25° × 0.25°	Monthly	5×10^{14}-1×10^{15} molecules cm^{-2} (or 35%-60%) over highly-polluted areas	Boersma et al. (2004)
HCHO	SCIAMACHY/ENVISAT	As low as 30 × 30 km^2/once per day/10:00 am	0.25° × 0.25°	Monthly	0.5-2.0 × 10^{15} molecules cm^{-2}	De Smedt et al. (2008)
SO_2	SCIAMACHY/ENVISAT	As low as 30 × 30 km^2/once per day/10:00 am	0.25° × 0.25°	Monthly	3-7 × 10^{15} molecules cm^{-2} (40-80%) over the U.S.	Lee et al. (2009)
TOR	OMI-MLS/Aura	13 × 24 km^2/once per day/1:45 pm	1° × 1.25°	Monthly	±4-5 DUs	Ziemke et al. (2006)
SWDN	CERES/Terra	20 × 20 km^2/once per day/10:30 am	1° × 1°	Monthly	1%	Wielicki et al. (1996)
LWDN	CERES/Terra	20 × 20 km^2/once per day/10:30 am	1° × 1°	Monthly	0.5%	Wielicki et al. (1996)
OLR	AVHRR/NOAA-18	1.1 × 4.4 km^2/once per day/2:00 pm	2.5° × 2.5°	Monthly	N/A	Liebmann and Smith (1996)
AOD	MODIS/Terra	22 × 22 km^2/once per day/10:30 am	1° × 1°	Monthly	±0.05±0.15r over land and ±0.03±0.05r over ocean	Remer et al. (2005)

COT	MODIS/Terra	22 × 22 km^2/once per day/10:30 am	1° × 1°	Monthly	8% (random error); 13% (mean error)	Remer et al. (2005)
CF	MODIS/Terra	1-5 km/once per day/10:30 am	1° × 1°	Monthly	10% (random error)	Remer et al. (2005)
CCN	MODIS/Terra	1-5 km/once per day/10:30 am	1° × 1°	Monthly	N/A	Remer et al. (2005)
CDNC	MODIS/Terra	1-5 km/once per day/10:30 am	1° × 1°	Daily	< 10% when CF > 0.8 and LWP > 25 gm^{-2} over NA	Bennartz (2007)
LWP	MODIS/Terra	1-5 km/once per day/10:30 am	1° × 1°	Monthly	15-25 gm^{-2} (random error)	Bennartz (2007)
PWV	MODIS/Terra	1-5 km/once per day/10:30 am	1° × 1°	Monthly	5-10%	Gao and Kaufman (2003); Remer et al. (2005)

[a]TOR: tropospheric ozone residuals; SWDN: downward surface solar radiation; LWDN: downward surface longwave radiation; OLR: TOA outgoing longwave radiation; AOD: aerosol optical depth; COT: cloud optical thickness; CF: cloud fraction; CCN: cloud condensation nuclei; CDNC: cloud droplet number concentration; LWP: cloud liquid water path; PWV: precipitable water vapor.
[b]MOPITT: Measurements of Pollution in the Troposphere; SCIAMACHY: Scanning Imaging Absorption Spectrometer for Atmospheric Chartography; ENVISAT: Environmental Satellite; OMI: Ozone Monitoring Instrument; MLS: Microwave Limb Sounder; CERES: Cloud's and the Earth's Radiant Energy System; AVHRR: Advanced Very High Resolution Radiometer; MODIS: Moderate Resolution Imaging Spectroradiometer.

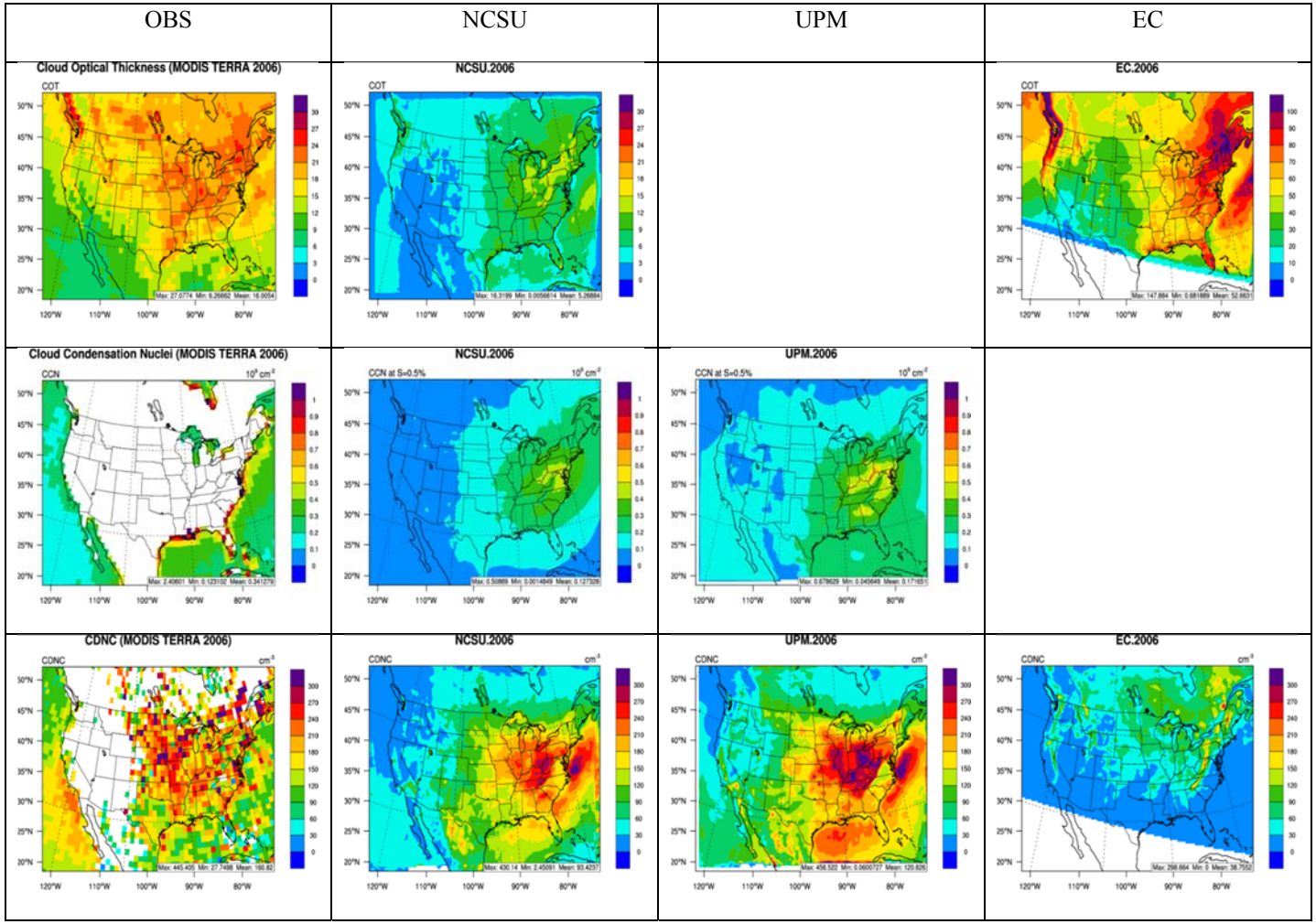

Figure A1. Spatial distribution of aerosol/cloud related variables (from top to bottom: COT, CCN, and CDNC) between satellite observation and different models for year 2006 (blank color denotes to missing values; only the available variables from limited simulations are displayed; scale for COT of EC is different).

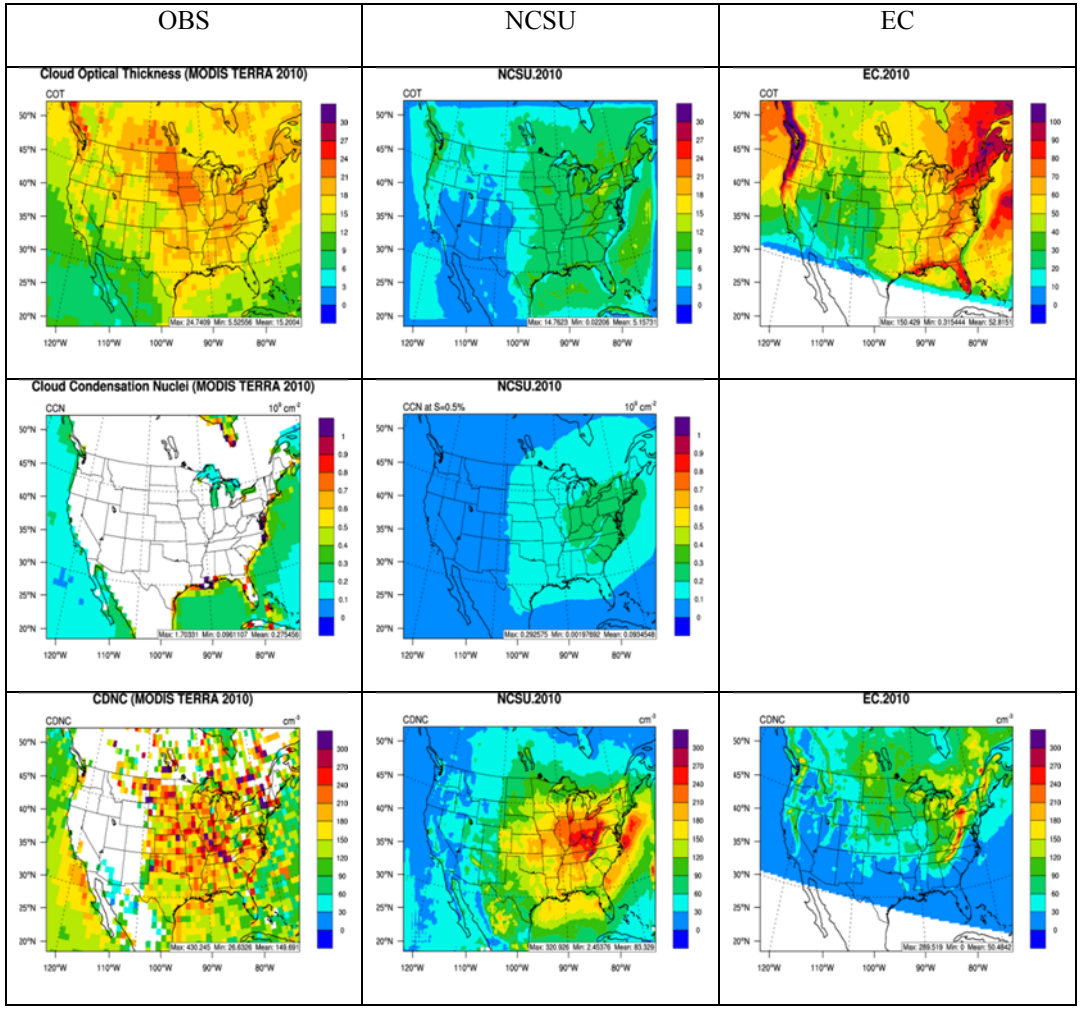

Figure A2. Spatial distribution of aerosol/cloud related variables (from top to bottom: COT, CCN, and CDNC) between satellite observations and different models for year 2010 (blank color denotes to missing values; only the available variables from limited simulations are displayed; scale for COT of EC is different).