JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2018MS001457

Quantifying Uncertainties of Ground-Level Ozone Within WRF-Chem Simulations in the Mid-Atlantic Region of the United States as a Response to Variability

Andrew Thomas[1,2], **Amy K. Huff**[1], **Xiao-Ming Hu**[3], **and Fuqing Zhang**[1,2]

[1]Department of Meteorology and Atmospheric Science, Pennsylvania State University, University Park, PA, USA, [2]Center for Advanced Data Assimilation and Predictability Techniques, Pennsylvania State University, University Park, PA, USA, [3]Center for Analysis and Prediction of Storms, and School of Meteorology, University of Oklahoma, Norman, OK, USA

Key Points:
- Large uncertainties in simulations of daytime ground-level ozone derive from uncertainties in meteorological initial and boundary conditions
- The average ground-level ozone mixing ratio (mean bias) is nevertheless influenced more by uncertainties in the emissions inventory than in the meteorological initial and boundary conditions
- The error and bias of model-predicted ground-level ozone may be decreased as much as 0.59 ppbv per year due to progressive reduction over the past decade of emissions of nitrogen oxides (NO_x), a chemical precursor of ozone

Correspondence to:
A. K. Huff and F. Zhang,
akh157@psu.edu;
fzhang@psu.edu

Citation:
Thomas, A. M., Huff, A. K., Hu, X.-M., & Zhang, F. (2019). Quantifying uncertainties of ground-level ozone within WRF-Chem simulations in the mid-Atlantic region of the United States as a response to variability. *Journal of Advances in Modeling Earth Systems*, *11*, 1100–1116. https://doi.org/10.1029/2018MS001457

Received 26 JUL 2018
Accepted 1 APR 2019
Accepted article online 3 APR 2019
Published online 29 APR 2019

Abstract Understanding forecast uncertainties and error growth dynamics is a prerequisite for improving dynamical prediction of meteorology and air quality. While predictability of meteorology has been investigated over the past few decades, the uncertainties in air quality simulations are less well known. This study explores the uncertainties in predicting ground-level ozone (O_3) in the Mid-Atlantic region of the United States during June 2016 through a series of simulations using WRF-Chem, focusing on the sensitivity to the meteorological initial and boundary conditions (IC/BCs), emissions inventory (EI), and planetary boundary layer (PBL) scheme. The average uncertainty of ground-level maximum 8-hr average O_3 mixing ratio (MD8-O_3) was most sensitive to uncertainties in the IC/BCs, while uncertainty in the EI was of secondary importance, and was least sensitive was to the use of different PBL schemes. Updating the NO emissions in the EI had the greatest influence on the accuracy, with an estimated decrease of 0.59 ppbv/year in the root-mean-square error and an average decrease of 0.63 ppbv/year in the values of modeled MD8-O_3. Our study suggests using perturbations in IC/BCs may lead to a more dispersive ensemble of O_3 prediction than using different PBL schemes and/or different EI. However, considering the combined uncertainties from all three sources examined are still smaller than the averaged root-mean-square errors of predicted O_3 against observations, there are apparent other sources of uncertainties not studied that need to be considered in future ensemble predictions of O_3.

Plain Language Summary Ozone, the primary pollutant in photochemical smog, is harmful to human health, particularly for children, senior citizens, and people with existing heart or lung diseases, like asthma. To protect public health, air quality forecasts of ozone are issued across the United States, primarily for metropolitan areas, where ground-level ozone tends to be the highest. When ground-level ozone is predicted to be exceed the daily health standard, people are advised to take steps to limit their outdoor activities. Operational air quality forecasters use predictions of ground-level ozone from numerical air quality models as guidance for their public forecasts. To assess the uncertainty in these model predictions of ground-level ozone, a series of simulations using the air quality model WRF-Chem were conducted in this study through changing the inputs to the model, including starting weather conditions, pollutant emissions inventories, and other model settings. Results show that the largest uncertainty in ground-level ozone was predicted by the model simulations using different starting weather conditions. Using the results of this study, the year-to-year decrease in the ground-level ozone was simulated, based on reduction in the source emissions. This study provides guidance regarding how numerical air quality models should be configured for future operational applications, which may lead to more accurate predictions of ground-level ozone and, thus, more accurate air quality forecasts.

1. Introduction

Ground-level ozone (O_3), one of the U.S. Environmental Protection Agency (EPA)'s six criteria pollutants, has been linked to a litany of health problems, including cardiovascular disease (Azevedo et al., 2011), infant mortality (Bell, 2004), and asthma (Gent, 2003). Sensitive groups, including individuals with cardiovascular and pulmonary diseases, children, and senior citizens, are particularly susceptible to negative health effects associated with exposure to criteria pollutants. To protect the public from the harmful effects of ground-level O_3 and the other criteria pollutants, the U.S. EPA has established National Ambient Air Quality Standards (NAAQS). Primary NAAQS project public health, and secondary NAAQS project public welfare. For O_3, the

current (as of 1 October 2015) primary and secondary NAAQS are both 70 ppbv, averaged over an 8-hr time period. Compliance with the O_3 NAAQS is determined by the annual fourth-highest MD8-O_3, averaged over 3 years; if this value for a given location is ≥71 ppbv, the location is designated a nonattainment area for O_3.

To help protect the public from the adverse health effects of O_3, air quality forecasts of ambient O_3 are issued for major metropolitan regions and surrounding suburbs across the United States. These forecasts are communicated using EPA's the Air Quality Index (AQI), a color-coded, dimensionless scale (https://www3.epa.gov/airnow/aqi_brochure_02_14.pdf). An AQI of 101 or higher corresponds to an exceedance of the O_3 NAAQS (termed "O_3 exceedances" hereafter), therefore, it is most critical for air quality forecasters to issue accurate forecasts on days when the AQI is expected to reach 101 or higher.

Eulerian air quality modeling is a multifaceted problem that depends on economic forcing (Tong et al., 2016), meteorological conditions (Seaman, 2000), biogenic emissions (Bell & Ellis, 2004), and chemistry (Mar et al., 2016). Each of these processes is variable, which affects the practical predictability of O_3. Practical predictability is defined as the uncertainty of modeling with errors in either initial state or process that are considered acceptable for operational uses. Practical predictability is contrasted against intrinsic predictability—the limit of the uncertainty of a nearly perfectly modeled system (Lorenz, 1969; Melhauser & Zhang, 2012; Zhang et al., 2006). Practical predictability studies evaluate the current state of the ability to accurately model a phenomenon or variable, with the intention of identifying factors that may improve accuracy. The practical predictability of O_3 is essential for the proper implementation of models in air quality forecasting, since it allows model users to attribute error to each process. We are using the American Meteorological Society's definition of predictability (Predictability, 2012), which is "[t]he extent to which future states of a system may be predicted based on knowledge of current and past states of the system." Since it is common to quantify predictability using model uncertainty (Zhang et al., 2006; Melhauser & Zhang, 2012; Houtekamer & Zhang, 2016; Bei et al., 2010), we will often refer to uncertainty (2012), expressed through a standard deviation, as a metric to describe the practical predictability. We will also use variability to refer to the uncertainty of a process. For this study, we will only examine model uncertainty as it pertains to practical predictability, since we want to target research areas, such as the emissions inventory or initial and boundary conditions, to maximize the accurate prediction of ground-level O_3.

In this study, we used the Weather Research and Forecasting Model with Chemistry (WRF-Chem; Grell et al., 2005), rather than the Community Multiscale Air Quality Model (CMAQ; Byun et al., 1997), which is used by the National Air Quality Forecasting System (NAQFC; Otte et al., 2005), because WRF-Chem computes chemistry alongside the meteorology. In this way, WRF-Chem is not restrained to computing the chemical tendencies with meteorological information that is as old as the meteorological model output interval, as is the case with CMAQ (Grell et al., 2005). The downside is that the computational costs increase with the additional computations between output intervals (Grell et al., 2005). While another advantage of simulating chemistry alongside the meteorology is the possibility for chemistry to interact with model physics, we did not test such features in order to mitigate the role of errors in chemistry on physical and dynamical processes and to reduce the computational expenses attributed to modeling aerosols.

The model uncertainty of simulated ground-level O_3 is influenced by several settings. The most notable settings that contribute to the model uncertainty are the physical parameterizations, initial and boundary conditions, and anthropogenic emissions inventory (Cuchiara et al., 2014; Mallet & Sportisse, 2006; Mena-Carrasco et al., 2009; Misenis & Zhang, 2010). Physical parameterizations, equations that approximately emulate subgrid phenomena, are methods representative of the current understanding of the physical process. The variety of solutions of the parameterizations is representative of the diverse understanding of those processes and is one aspect of practical predictability of simulated ground-level O_3. Varying initial and boundary conditions are within the definition of practical predictability, for it is not reasonable to know the true state of the atmosphere, differentiating practical predictability from the intrinsic predictability (Melhauser & Zhang, 2012). The definition of practical predictability also contains the anthropogenic emissions inventory, as real-time point and area emissions are not available presently.

Several studies have been done on the model uncertainty of O_3 due to variability of the physical parameterizations (e.g., Hu et al., 2012; Hu et al., 2013; Hu et al., 2013; Žabkar et al., 2013). Mallet and Sportisse (2006) indicated that O_3 is most sensitive to the model uncertainty in the turbulence closure, which is embodied in the planetary boundary layer (PBL) scheme within WRF-Chem. Some studies indicate that the Yonsei

University (YSU) PBL scheme shows the closest agreement between predicted hourly O_3 and observed hourly O_3 (Yerramilli et al., 2012; Cuchiara et al., 2014; Cheng et al., 2012). Although Yerramilli et al. (2012) claim that O_3 is more sensitive to model uncertainty in the Land Surface Model (LSM), Pleim (2011) notes that there are deficiencies in that study, including the fact that the second version of the Asymmetric Convective Model (ACM2) at that time was not implemented for use with tracers within WRF-Chem. Since then, other studies, such as Cuchiara et al. (2014), have used ACM2 within WRF-Chem since diffusion was implemented in 2013. Hu, Klein, and Xue (2013) and Hu, Klein, Xue, Zhang et al. (2013) indicate that the mixing strength in the PBL scheme is critical to correctly simulate near-surface O_3 and its dry deposition during night time, which will affect the amount of O_3 in the convective boundary layer during the next day. Other studies, such as Misenis and Zhang (2010), corroborate the claim that O_3 may be more sensitive to model uncertainty in the LSM than the PBL scheme, though Hodnebrog et al. (2011) suggested that the model uncertainty of O_3 to variability in the LSM may be confined to certain regions. Additionally, the physical relationship between the LSM and O_3 is not as clear as the physical relationship between O_3 and the PBL. There are numerous settings specific to atmospheric chemistry that may impact the mixing ratios of O_3 and its precursors but are limited to "on" and "off" within WRF-Chem, such as dry deposition.

While the variety of choices of the PBL scheme represents a large model variability of a solitary meteorological process, emissions inventory variability, which represents the variability of anthropogenic emissions, provides a large component of model uncertainty of O_3. The emissions inventory includes, but is not limited to, emissions of carbon monoxide, volatile organic compounds (VOCs), and nitric oxide (NO) within the United States. O_3 is formed from the reaction of atomic oxygen with molecular oxygen following a series of complex reactions involving sunlight; NO and nitrogen dioxide (NO_2), collectively termed NO_x; and VOCs (Calvert et al., 2015). Within the United States, anthropogenic NO_x, which is emitted by high-temperature combustion processes, has been decreasing since at least 2005 due to regional emissions controls on large energy-generating units and mobile sources (Tong et al., 2015). Although NO_x emissions overall have been decreasing, they retain day to day temporal and spatial variability. Variability in emissions inventories is associated with changes in emissions control devices, activity, fuel sources, and changes in emission factors, among other reasons (Anderson et al., 2014; Castellanos et al., 2009; Frost et al., 2006; Kim et al., 2006; Tong et al., 2015; Vijayaraghavan et al., 2012). Of the emission inventories available, the gridded version of the National Emissions Inventory (NEI) series from the U.S. EPA has 4-km resolution, the highest spatial resolution among readily available emissions inventories. This aspect is critical, since higher-resolution emissions inventories are important for local air quality studies (Hodnebrog et al., 2011). Additionally, the program that adapts the NEI data to the WRF-Chem grid implements an algorithm for subgrid buoyant plume rise and elevated point source emissions, which may not otherwise be resolved by WRF-Chem. The NEI gridded emissions inventories are updated approximately every 6 years, with the latest emissions inventory being representative of a typical July weekday in 2011 (NEI-11). The point sources are compiled by the EPA, according to the Air Emissions Reporting Rule. The mobile emissions sources were processed using the Motor Vehicle Emissions Simulator, which has the most variability according to Anderson et al. (2014). For more information on the construction of the NEI-11, please see EPA (2015).

Accurate prediction of ground-level O_3 is contingent upon accurate prediction of NO_x. Errors in NO_x emissions are particularly important for the Mid-Atlantic region, which is NO_x limited (e.g., Butler et al., 2011; Duncan et al., 2010), meaning that during the summer O_3 season, increases in ambient NO_x emissions will increase production of O_3. Travis et al. (2016) showed that the NEI-11 may be overestimating NO_x emissions up to a factor of 2, demonstrating that the overestimation of NO_x emissions leads to overestimation of O_3 in NO_x-limited environments. While Travis et al. (2016) agree with Anderson et al. (2014), the recommended decrease in NO_x varies, indicating that there is ambiguity with the precise decrease needed. The NAQFC adjusted the NO_x emissions for 2012 and found that the NO_x bias decreased between 0.57 and 2.34 ppbv, while the decrease in O_3 bias was between 0.92 and 1.87 ppbv (Pan et al., 2014). Other studies in different areas yielded varying results. Zhong et al. (2016) provides one example of the model uncertainty of O_3 in China due to discrepancies between one regional and one global emissions inventory, with differences between 12 and 16 ppbv of O_3 in certain locations, suggesting that the model uncertainty of O_3 due to variability of the emissions inventory is location specific. Going further than simply comparing emissions

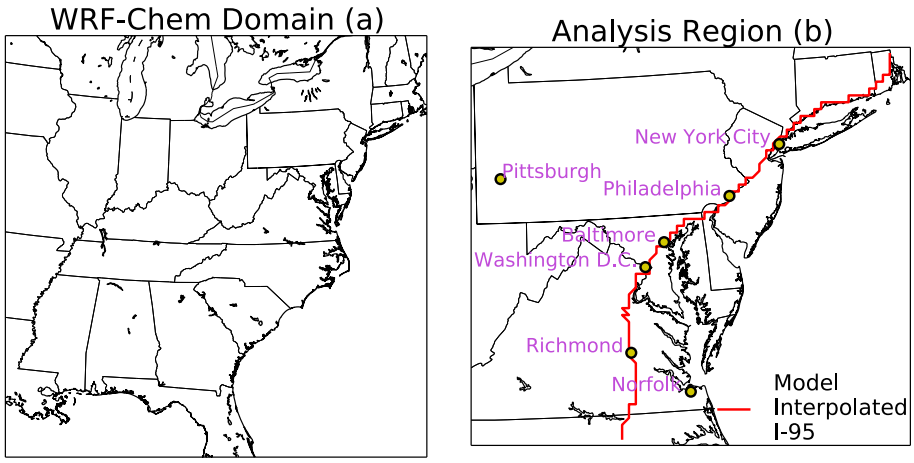

Figure 1. The simulated WRF-Chem domain (a), where the WRF-Chem simulations were conducted, and the analysis region (b), where our analysis occurs.

inventories, some studies have examined the response of O_3 due to certain sources (Vijayaraghavan et al., 2012) and temporal variability (Castellanos et al., 2009) of emissions.

While the emissions inventory is known to be a major source of model uncertainty for MD8-O_3, the initial and boundary conditions are critical for determining the evolution of the model. Lorenz (1963) demonstrated the influence of perturbations of the initial state on the evolution of nonperiodic flow. This dependency on the initial conditions also includes the initial composition. While chemical initial and boundary conditions are important (Berge et al., 2001), the influence of chemical initial conditions can be constrained with a spin-up time of 48 hr, to reduce the correlation on initial conditions by 10% (Jiménez et al., 2007). Bei et al. (2010) suggests that meteorological initial conditions contribute more uncertainty to O_3 mixing ratios in Mexico City, with the ensemble spread reaching 15 ppbv over Houston. Zhang et al. (2007) noted that, for a high O_3 event in Houston, Texas, the spread of ensemble members with different initial (and boundary) conditions peaked with 40 ppbv of hourly O_3 values over the Gulf of Mexico and 20 ppbv over one of the Texas stations. Gilliam et al. (2015) also observed a spread as high as 10–20 ppbv over the northeastern United States, while using the Short-Range Ensemble Forecasting system members as initial and boundary conditions. Hu et al. (2019) used a similar methodology as Gilliam et al. (2015) for the Dallas-Fort Worth area and found that the spread of the plume direction was most affected. Beekmann (2003) suggest that the O_3 sensitivities may range from 4–10 ppbv over Paris. Discrepancies in the ranges of sensitivities may be a result of a variety of reasons, including the methods of perturbations of initial and boundary conditions, the quantity of O_3 precursors emitted, the effect of local topography, and the modeled meteorological uncertainty.

We investigated and compared the relative impacts of the variability in the emissions inventory, the PBL scheme, and the meteorological initial and boundary conditions on the model uncertainty of MD8-O_3 by running different WRF-Chem simulation experiments over the eastern United States for June 2016; the model domain is shown in Figure 1a. While the simulations were conducted over the eastern United States, the focus was on the urbanized Interstate-95 (I-95) Corridor, running from New York City to Washington, DC, as shown in Figure 1b. This region routinely observes the highest MD8-O_3 in the Mid-Atlantic region. The goal was to quantify the model uncertainty of MD8-O_3 along the I-95 Corridor, with the aim of making recommendations to improve air quality model guidance in an area that is highly susceptible to O_3 exceedances. The motivation behind this research was to test potential settings needed to make near-continuous operational predictions of MD8-O_3 using WRF-Chem for the Mid-Atlantic Region. Those potential settings for the simulations included the YSU, Mellor-Yamada-Janjic scheme (MYJ), and ACM2 PBL schemes, the NEI-11, NEI-05, NEI-14 emissions inventories, and Global Forecasting System (GFS), European Center for Medium Range Weather Forecasting Interim Reanalysis (ERA-Interim), and the second version of the Modern Era Retrospective analysis for Research and Applications (MERRA-2) meteorological initial and boundary conditions. The three analysis data sets differ in the forecast model,

Table 1
CNTL Model Settings

Setting	Choice	Reference
Longwave radiation	RRTM	Mlawer et al. (1997)
Shortwave radiation	Dudhia	Dudhia (1989)
Cumulus	Grell-Devenyi	Grell and Dévényi (2002)
Microphysics	WSM6	Hong and Lim (2006)
Land surface model	NOAH	Tewari et al. (2004)
Chemistry	RACM	Stockwell et al. (1997)
Biogenic emissions	MEGAN	Guenther et al. (2006)
Photolysis	Madronich	Madronich (1987)

data assimilation methodology, and the amount of observations being assimilated, which can represent realistic uncertainties in the meteorological initial and boundary conditions.

This study complements much of the existing literature and provides an evaluation of the relative importance of the choice of different model settings. This work also supplements previous studies regarding the model uncertainty of MD8-O_3 as a response to variability in (a) the meteorological initial and boundary conditions, by testing different sources as a sampling method, and (b) the emissions inventory by testing qualitatively similar emissions inventories with the exactly the same meteorological scenarios. In this way, this research represents a nearly longitudinal modeling study of the trends of O_3 due to changes in the emissions inventory.

2. Methods

2.1. The Climatology of O_3 in the Mid-Atlantic Region

Historically, O_3 exceedance days within the Mid-Atlantic region have been characterized by a well-defined synoptic scale pattern associated with the western edge of the quasi-stationary Bermuda High extending into the Mid-Atlantic region. This pattern includes a ridge of high pressure aloft, with the ridge axis over or west of the Mid-Atlantic, and slowly eastward migrating surface high pressure. This synoptic pattern is conducive to sunny skies, above average temperatures, stagnating surface winds, and regional transport of O_3 precursor emissions from the historically NO_x-rich Ohio River Valley source region (Ryan et al., 1998). Under these conditions, local and regional O_3 formation was maximized, leading to multiday O_3 exceedance events.

During June 2016, the weather patterns observed during the most widespread MD8-O_3 exceedance days of 1 and 20 June were similar to those described by Ryan et al. (1998). On the days of 11, 15, 24, and 26 June, O_3 formation was enhanced by mesoscale features, including Atlantic Ocean sea breezes, Chesapeake Bay breezes, and an Appalachian lee trough, all of which increased O_3 production. An example of the enhancement of O_3 by a sea breeze circulation in the Mid-Atlantic region was given in Stauffer et al. (2015). Another example is shown in Seaman and Michelson (2000), which examined the influence of an Appalachian lee trough on O_3.

2.2. WRF-Chem Simulations

A control simulation, designated CNTL, using the WRF-Chem version 3.6.1 (Grell et al., 2005) with the regional atmospheric chemistry mechanism (RACM; Stockwell et al., 1997), was performed with meteorological initializations starting every day at 12 UTC from 1 June to 29 June, with a spin-up period on 30 and 31 May. A 2-day spin-up period was chosen because there were precedents for it, or for even fewer hours, without extensive explanation of the accuracy of the chemical initial and boundary conditions (Jiménez et al., 2007; Tie et al., 2010; Zhang et al., 2006). Each initialization contained 48 hr of forecasts, such that each day in June 2016 was simulated. The chemical fields were reinitialized with the previous day's prediction of the chemical fields, such that the chemical fields for the first 24-hr period of each initialization was considered continuous. The reinitialization of the meteorological fields is assumed to have a small effect on the chemical tendencies, with only a hypothetical jump in the photolysis rates. The tendencies from the chemical mechanism are considered autonomous because the initialization of the meteorological fields was a dry start, so clouds were still forming, but since the initialization time was shortly after sunrise, the inconsistency in the photolysis rate is considered minimal. Inconsistencies in the horizontal advection and vertical mixing tendencies may have occurred due to changes in the momentum fields, but they are considered (a priori) to be small due to the timing of when reinitialization occurred. A preliminary test was conducted and showed that the inconsistencies between the soil data and meteorological data had a negligibly small effect on the result (not shown). The CNTL sensitivity experiment used the YSU (Hong et al., 2006) PBL scheme and the NEI-11 emissions inventory (EPA, 2015). In addition to the model settings listed in Table 1, all sensitivity experiments had a domain spanning from the Mississippi River to southern Maine, as is depicted in Figure 1a. The vertical model structure was the same as described by Hu et al. (2012) with 12-km grid spacing. In Figure 1b, major cities are marked and labeled in the analysis region which contains

Table 2
Sensitivity Experiment Names and Changed Model Settings

Sensitivity experiment	PBL	Meteorological initial and boundary conditions	Emissions
CNTL	YSU	GFS	2011
ACM2	ACM2	GFS	2011
MYJ	MYJ	GFS	2011
NEI-05	YSU	GFS	2005
NEI-14	YSU	GFS	2014
MERRA-2	YSU	MERRA-2	2011
ERA	YSU	ERA	2011

Note. ACM2 = Asymmetric Convective Model; CNTL = control simulation; ERA = European Center for Medium Range Weather Forecasting Interim Reanalysis; GFS = Global Forecasting System; MERRA = Modern Era Retrospective analysis for Research and Applications; MYJ = Mellor-Yamada-Janjic scheme; NEI = National Emissions Inventory; PBL = planetary boundary layer; YSU = Yonsei University.

the urbanized I-95 Corridor, marked in red. We made the WRF-Chem domain more expansive than the analysis region to account for anthropogenic sources of NO_x upwind.

Besides the CNTL sensitivity experiment, a total of six sensitivity experiments were performed, each of which varied either the PBL scheme, emissions inventory, or the meteorological initial and boundary conditions from that used in the CNTL sensitivity experiment. All sensitivity experiments, including the CNTL, are detailed in Table 2, along with the naming convention for each subgroup of sensitivity experiments.

The first subgroup of sensitivity experiments explored two different emissions inventories, the 2005 version of the National Emission Inventory (NEI-05) and a version of the NEI-11 updated to reflect emissions of NO in 2014 (NEI-14) in order to capture the annual variability of the emissions of NO and the corresponding model uncertainty of O_3. Utilizing the method of Tong et al. (2015), 2014 was the most recent year for which NO emissions could be updated. The weekday emissions of the NEI-05 were used to provide continuity with the NEI-11, since only the weekday emissions of the NEI-11 were available. The NO point emissions of the NEI-14 were updated for each hour by using the average of the Clean Air Market data (https://ampd.epa.gov/ampd/), which is provided by power plants to EPA (EPA: Air Markets Program Data, 2018). The area emissions were updated for each state by using state-wide ratios of NO_x emissions in 2014 to NO_x emissions within the NEI-11, which were contributed by the National Oceanic and Atmospheric Administration (Tong et al., 2015). The creation of the NEI-14 loosely follows the methodology used to update the NAQFC O_3 model (Pan et al., 2014). Figure 2 demonstrates the changes to each emissions inventory in the analysis region. Daily NO emissions decreased by 27.4% from the NEI-05 to the NEI-14 and daily emissions of VOCs decreased by 39.6% from the NEI-05 to the NEI-11; emissions of VOCs were unchanged by design from the NEI-11 to NEI-14.

The next subgroup of sensitivity experiments focused on two different PBL schemes, the ACM2 and MYJ schemes. As reviewed in Hu et al. (2010), the CNTL sensitivity experiment uses the YSU scheme, which is a nonlocal parabolic K-scheme that defines entrainment. The ACM2 is a hybrid local/nonlocal PBL scheme, which treats upward mixing nonlocally and downward mixing locally. In contrast to the YSU scheme used by the CNTL sensitivity experiment, the MYJ scheme is a local K-scheme. As noted by Skamarock et al. (2008), the WRF model constrains some PBL schemes to certain surface layer physics. Therefore, the sensitivity experiments that utilized the MYJ PBL scheme used the Eta similarity surface layer physics (Monin & Obukhov, 1954), while the ACM2 and YSU PBL schemes use the revised MM5 surface layer physics (Jiménez et al., 2012).

The final subgroup of sensitivity experiments used two different initial and boundary conditions: the second version of the MERRA-2 by the National Aeronautics and Space Administration and the ERA-Interim reanalysis from the European Centre for Medium-Range Weather Forecasts. MERRA-2 is a reanalysis data set that utilizes the Goddard Earth Observing System model with a 3D-Variational data assimilation system. The approximate resolution of MERRA-2 is 0.5° × 0.625° with 72 vertical levels (Gelaro et al., 2017). The ERA-Interim reanalysis uses the European Centre for Medium-Range Weather Forecasts model with T255 resolution (approximately 80 km) with a 4D-Variational data assimilation system (Dee et al., 2011). The soil temperature, moisture, and depth for all sensitivity experiments were taken from the GFS, such that only meteorological data was varied among this subgroup. The potential inconsistencies were found to be of little importance after comparing the ERA sensitivity experiment with GFS soil information and a simulation with the ERA initial and boundary meteorological and soil conditions (not shown). All inputted data sets were temporally interpolated, to the extent that the boundary conditions were updated hourly.

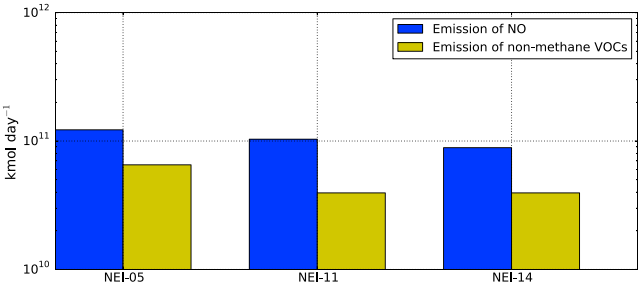

Figure 2. Average hourly emissions of NO and nonmethane volatile organic compounds from each emissions inventory used in this study.

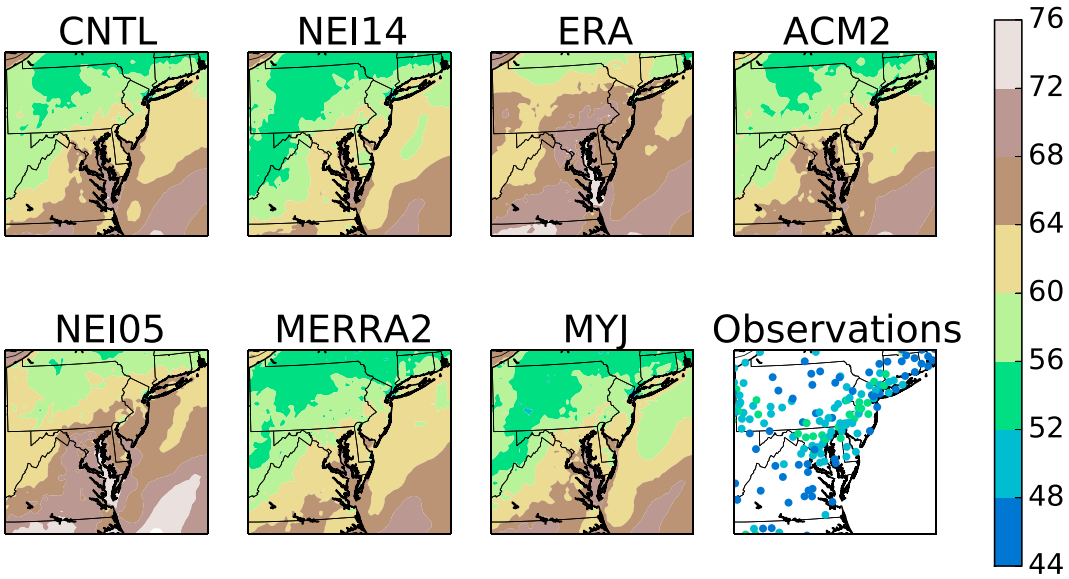

Figure 3. Temporal average of observed maximum daily 8-hr average ozone (MD8-O_3) during the modeling period, representing the average modeled and observed MD8-O_3.

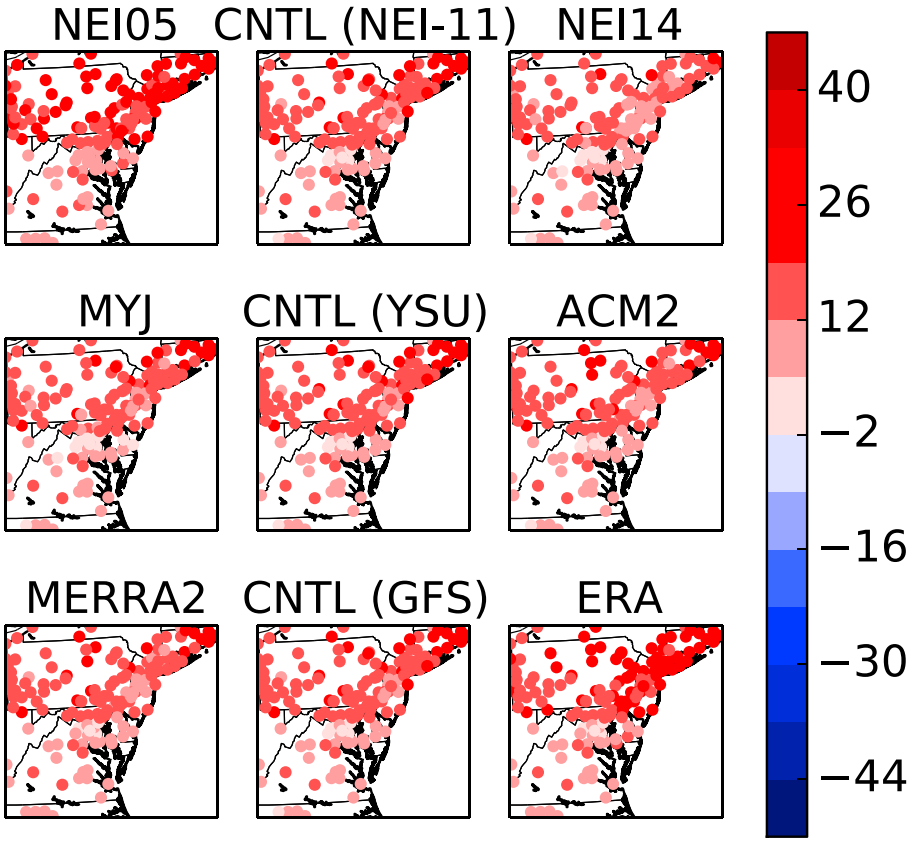

Figure 4. Mean bias of maximum daily 8-hr averaged ozone (MD8-O_3) for modeling period.

Table 3
Error Metrics of Maximum Daily 8-hr Averaged O_3 (MD8-O_3) for Each Sensitivity Experiment

Sensitivity experiment	Mean bias		RMSE		Unbiased RMSE*	
	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
CNTL	14.23	14.3	20.1	19.55	14.2	13.33
NEI-05	17.59	17.56	23.41	22.7	15.45	14.39
NEI-14	12.14	12.38	18.08	17.78	13.4	12.75
ERA	17.49	16.29	23.63	21.52	15.89	14.06
MERRA2	13.84	13.65	19.43	19.34	13.65	13.7
ACM2	13.98	13.96	19.66	18.95	13.82	12.82
MYJ	13.32	13.08	19.68	18.72	14.48	13.39

Note. One asterisk (*) signifies that the day 1 error metric is statistically significantly greater than the day 2 error metric, using a 95% confidence interval for the Student's *T* test. ACM2 = Asymmetric Convective Model; CNTL = control simulation; ERA = European Center for Medium Range Weather Forecasting Interim Reanalysis; MERRA = Modern Era Retrospective analysis for Research and Applications; MYJ = Mellor-Yamada-Janjic scheme; NEI = National Emissions Inventory; RMSE = root-mean-square error.

As mentioned previously, the focus of the study was on the urbanized I-95 Corridor. This entire region, encompassing parts of eastern Pennsylvania, New Jersey, Delaware, Maryland, and northern Virginia, is in nonattainment for O_3, based on the 2008 O_3 NAAQS of 75 ppbv (since nonattainment is determined by a 3-year average of O_3 observations, nonattainment areas based on the 2008 NAAQS are the most recent data currently available). In the I-95 Corridor analysis region, we calculated the sensitivity of MD8-O_3, expressed by the standard deviation among the three sensitivity experiments of each subgroup. We also examined the vertical distribution of MD8-O_3, and the modeled uncertainty thereof, along I-95. Additionally, we evaluated the effect of the changing emissions inventory on predicted MD8-O_3.

3. Results and Discussion

3.1. Average Group Differences

Figure 3 shows the temporal average of ground-level MD8-O_3 for each sensitivity experiment in the analysis region, as compared to the similarly processed observations. The model-interpolated location of I-95 is marked in white. Each sensitivity experiment overproduced MD8-O_3 in the analysis region, but the ERA sensitivity experiment overproduced O_3 to the largest extent.

Among the sensitivity experiments, the ERA sensitivity experiment produced the most O_3, even more than the NEI-05 sensitivity experiment. The cause of this overprediction was not well understood, though one possible explanation is that the ERA sensitivity experiment inadequately modeled vertical transport of O_3 and O_3 precursors from the residual layer. The choice of PBL scheme led to less noticeable differences in the average MD8-O_3 compared to the subgroups with different initial and boundary conditions or different emissions inventories. This result is different from previous studies, which found that the PBL scheme selection was more critical under different circumstances, such as Yerramilli et al. (2012), Cuchiara et al. (2014), or Cheng et al. (2012).

More quantitatively, Figure 4 and Table 3 show the mean bias and error metrics of each sensitivity experiment during the modeling period. The NEI-05 and ERA sensitivity experiments had the most overprediction

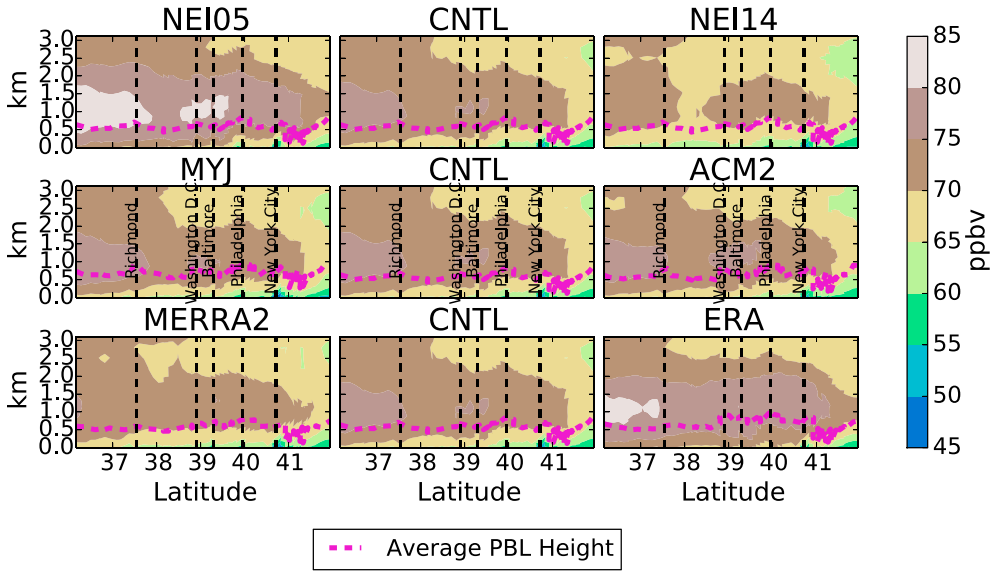

Figure 5. Average cross section of the maximum daily 8-hr average ozone (MD8-O_3) along I-95 during the modeling period.

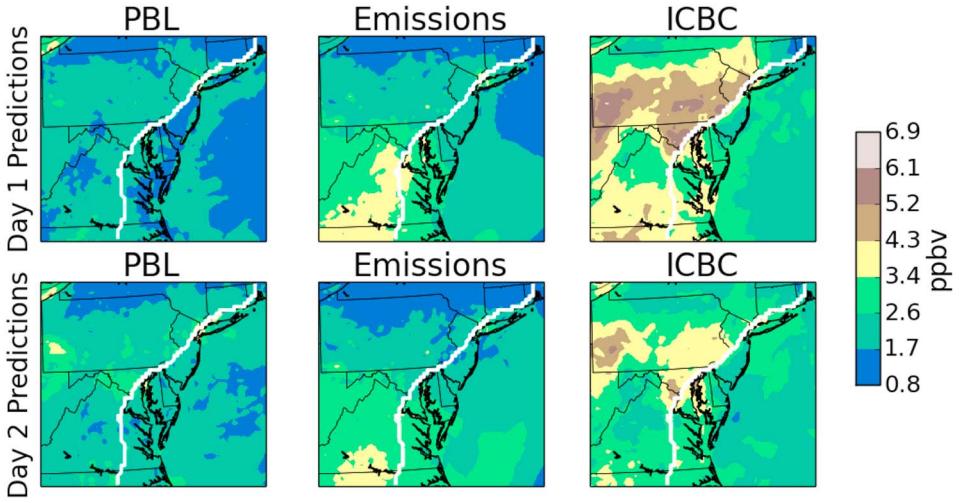

Figure 6. The average model uncertainty of the maximum daily 8-hr average ozone (MD8-O_3), expressed by the temporal average of the standard deviation of each subgroup, during the modeling period.

of MD8-O_3, while the NEI-14 experiment overpredicted less MD8-O_3 relative to the other sensitivity experiments in the emissions inventory subgroup (Table 3). The reduced overproduction of the NEI-14 sensitivity experiment suggests that an updated emissions inventory results in more accurate predictions of MD8-O_3. Also, the difference between the unbiased root-mean-square error (RMSE) of the day 1 and day 2 predictions of MD8-O_3 is statistically significant at the 95% confidence interval. While this is worth noting, the limited number of models used, as well as the fact that none of the other error tables within section 3.5 displays similar statistical significance, suggests that this is a Type 1 error.

The temporal average of the cross-sectional MD8-O_3 over I-95 is displayed in Figure 5. The latitudes of major cities are marked and labeled by dashed lines. The vertical coordinate was computed using the hypsometric equation of the averaged hydrostatic pressure and temperature during the same period that the ground-level MD8-O_3 was modeled. In the southern part of I-95, the NEI-05 sensitivity experiment showed the most

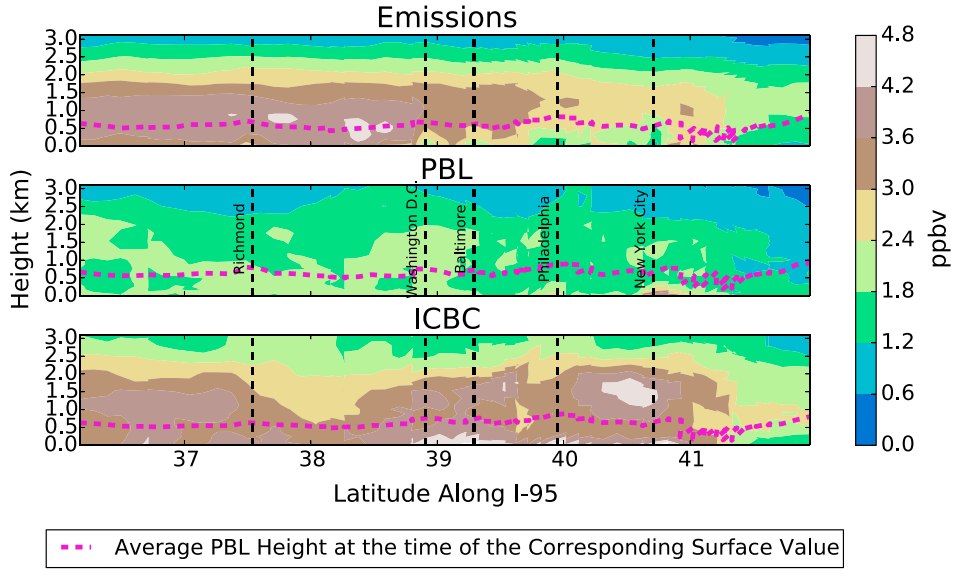

Figure 7. The average model uncertainty, expressed by the temporal average of the cross section of the standard deviation, of maximum daily 8-hr average ozone (MD8-O_3) during the 20% of days with National Ambient Air Quality Standards exceeding ozone observations. PBL = planetary boundary layer. ICBC= meteorological initial and boundary conditions.

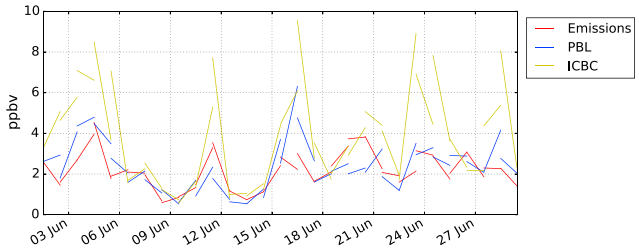

Figure 8. A time series of model uncertainty of the maximum daily 8-hr averaged ozone (MD8-O_3) for each O_3 observing monitor, expressed by the spatially averaged standard deviation, in June 2016. The beginning of each line segment indicates the uncertainty associated with the day 1 prediction of MD8-O_3, while the end represents the day 2 uncertainty. PBL = planetary boundary layer.ICBC = meteorological initial and boundary conditions

overproduction of MD8-O_3, with peak values of MD8-O_3 in rural and suburban areas. The overproduction of MD8-O_3 in the NEI-05 sensitivity experiment in the southern analysis region was most prominent 0.5–1 km aloft, near the height of the boundary layer, where the average mixing ratio reached over 80 ppbv. Additionally, the ERA sensitivity experiment showed that the overproduced O_3 was enhanced aloft by 5 ppbv over the cities of Washington D.C., Baltimore, and Philadelphia, as compared to both the MERRA-2 and CNTL sensitivity experiments. The overprediction of the ERA sensitivity experiment is likely due to its inadequate simulation of moist convection (not shown), leading to the enhancement of MD8-O_3. Thus, this cross section analysis demonstrates the impact that emission reductions can have on the vertical profile of MD8-O_3, as well as the uncertainty due to variability in the meteorological initial and boundary conditions.

3.2. Model Uncertainty of MD8-O_3

Figure 6 displays the average standard deviation of near-surface MD8-O_3 for each sensitivity experiment subgroup. The model-interpolated location of I-95 is denoted by the white line. This figure illustrates that the meteorological initial and boundary conditions subgroup had a larger modeled uncertainty of MD8-O_3 than the uncertainties of modeled MD8-O_3 due to variabilities in the emissions inventory and PBL parameterizations for the study period. Variations in using different initial and boundary conditions led to peak modeled uncertainty (mean standard deviation among the subgroup) for O_3 of 5.2–6.1 ppbv, which is considerably higher that the peak sensitivities of about 3.4–4.3 ppbv due to changes in either the emissions inventory or the PBL parameterization scheme. This implies that the meteorological conditions are a primary source of O_3 modeling uncertainties, which is consistent with findings in Zhang et al. (2007). Despite having a similar peak model uncertainty of MD8-O_3, the variations in the emissions inventory experiments have a greater impact over a larger area on variations of MD8-O_3 than do variations due to the PBL schemes. Although the model uncertainty of MD8-O_3 may be largest for the initial and boundary conditions, it does not necessarily indicate that the model uncertainty was largest at every point. For example, MD8-O_3 in southern New Jersey was more sensitive to variability in the emissions inventory than variability induced by using different meteorological initial and boundary conditions (by approximately 1 ppbv). Additionally, the largest uncertainties due to different meteorological initial and boundary conditions may not be collocated with the highest values of MD8-O_3. For example, I-95, which connects major cities in the analysis region, runs along the outside edge of the peak uncertainties. This highlights the model uncertainty of the transport of near-surface O_3 resulting from NO_x plume sources, such as I-95.

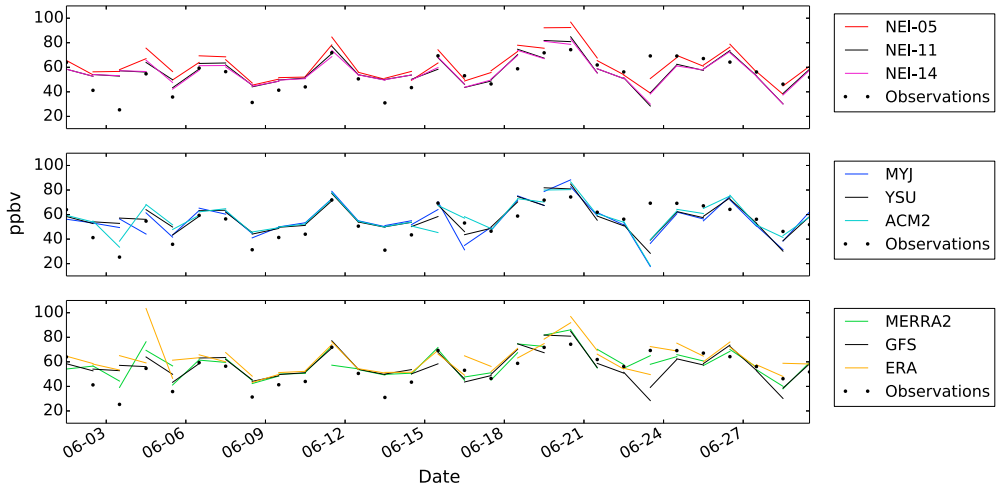

Figure 9. Time series of the maximum daily 8-hr averaged ozone (MD8-O_3) at Rutgers University.

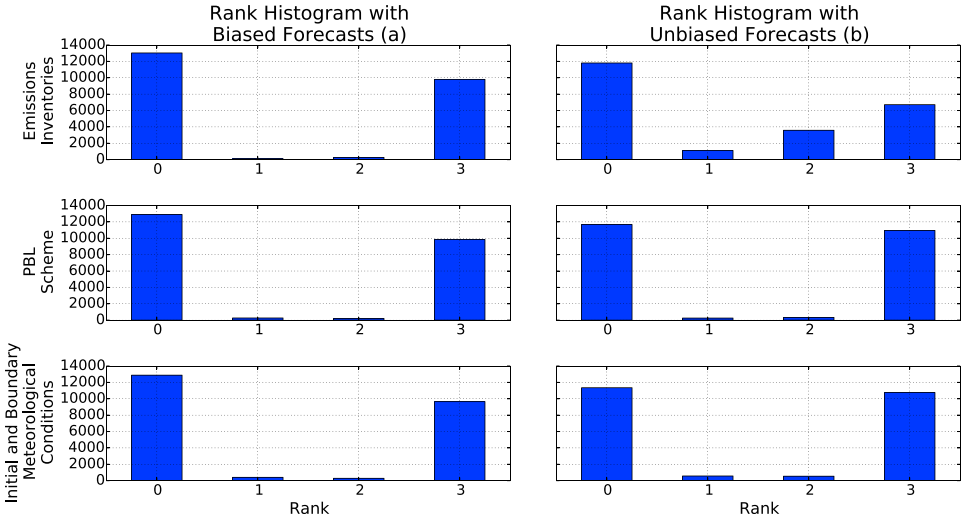

Figure 10. Rank histogram for biased (a) and unbiased (b) predictions of the maximum daily 8-hr averaged ozone (MD8-O_3) within the entire analysis region.

Figure 7 illustrates the vertical profile in the lowest 3-km altitude above I-95 for the temporally averaged model uncertainty of MD8-O_3. While the MD8-O_3 aloft was most sensitive to uncertainties in the meteorological initial and boundary conditions, the model uncertainty due to variability in the emissions inventory was larger aloft over the southern part of the analysis region. This result agrees with Travis et al. (2016) that the emissions inventory represents a large source of O_3 modeling error for the southeastern United States. Also, the most model uncertainty of MD8-O_3 associated with the emissions and initial/boundary conditions subgroups was located near the top of the boundary layer. This increase in the sensitivity model uncertainty of O_3 with height is consistent with the hypothesis that the ERA sensitivity experiment inadequately modeled convective vertical transport of O_3 and O_3-precursors.

Figure 8 shows which days have the most model uncertainty of MD8-O_3, by displaying the average standard deviation of each subgroup at each observing location. Consistent with Figure 6 and Figure 7, the initial and boundary conditions subgroup frequently had the largest uncertainties among the three sensitivity subgroups. However, the average standard deviation of each subgroup is considerably smaller than the RMSE, which is defined as the square root of the average squared distance between the observations and predictions, among all sensitivity experiments. This result also shows a smaller standard deviation than the findings of Gilliam et al. (2015), but this may be due to the use of only three sensitivity experiments in each subgroup, as well as the use of days with low O_3, which is inherent in the utilization of the

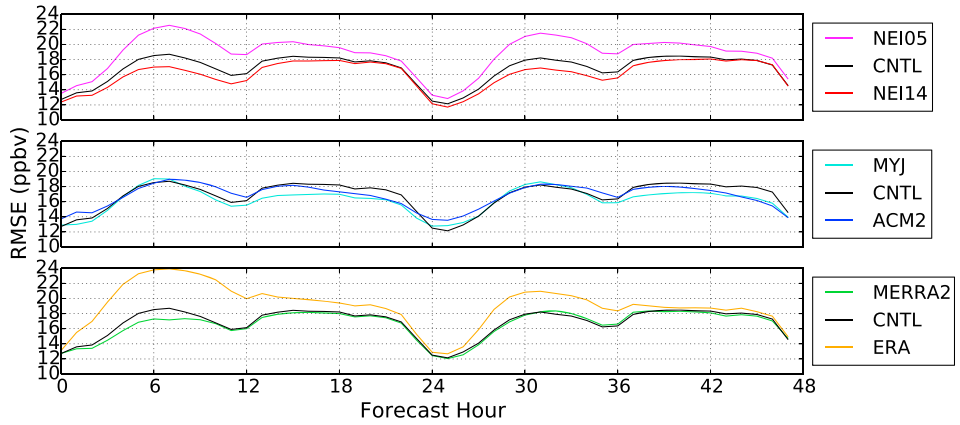

Figure 11. Root-mean-square error (RMSE) of hourly O_3 as a function of forecast hour.

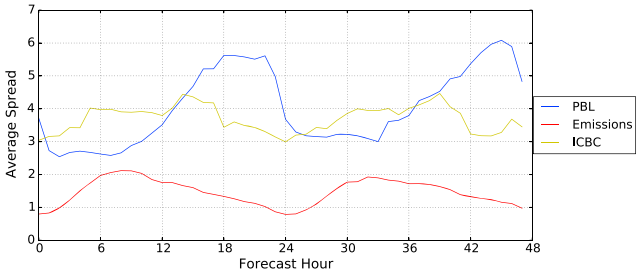

Figure 12. Station-averaged uncertainty of hourly O_3 as a function of forecast hour. PBL = planetary boundary layer. ICBC= meteorological initial and boundary conditions.

entire month of June. By comparison, Gilliam et al. (2015) used 10 Short-Range Ensemble Forecasting-initialized WRF-CMAQ simulations, as compared to our three multimodel initialized WRF-Chem simulations. Our results for the model uncertainty in MD8-O_3 due to variability in the initial and boundary conditions over land were of similar order to those presented in the "half-run" by Zhang et al. (2007). The similarity of these results implies that the initial differences between the CNTL, ERA, and MERRA-2 sensitivity experiments were representative of about half the variability (i.e., model uncertainty) of the initial ensemble employed by Zhang et al. (2007).

Although the initial and boundary conditions have the most model uncertainty among the subgroups, the initial and boundary condition subgroup is still underdispersive, suggesting other sources of uncertainties beyond what are examined in this study also need to be considered simultaneously in probabilistic O_3 modeling. This is evidenced by Figure 9, which shows a time series of MD8-O_3 of each subgroup at the Rutgers University O_3 monitor, which is representative of the New York City metropolitan area. The Rutgers University monitor was chosen because it observed 4 days with NAAQS exceedances of MD8-O_3 during the study period. Since many of the observations were outside the ranges of each subgroup, all of the subgroups are underdispersive, which is an important consideration when data assimilation of air quality models is conducted (Houtekamer & Zhang, 2016). Additionally, Figure 9 supports the idea that older emissions inventories produce more MD8-O_3, enhancing the positive bias. Figure 10, which shows a rank histogram of MD8-O_3 within the entire analysis region, broadens this idea of underdispersion beyond a single monitor. Even with the bias removed, Figure 10 shows signs of underdispersion.

3.3. Uncertainty and Error for Each Forecast Hour

Figure 11 shows the RMSE for O_3 predicted each hour within the analysis region for each subgroup. For the first few hours, the model error increases greatly. The ERA sensitivity experiment shows the largest error. The error also decreases to a relative minimum near hour 24. This minimum is similar to the minimum found at hour 0. This result may be caused by the daily reinitialization, but that does not sufficiently explain the drastic decrease in RMSE preceding reinitialization. However, one possibility is that diurnal pattern of the errors may be associated with the diurnal pattern of the sensitivity, related to the treatment of the PBL scheme. Figure 11 also shows that the errors in day 2 predictions of hourly O_3 may be due to the a priori assumption of the small impact of reinitialization error. This explanation is particularly relevant to the ERA and NEI-05, which have the most amount of error in all of the sensitivity experiments. Moreover, this pattern

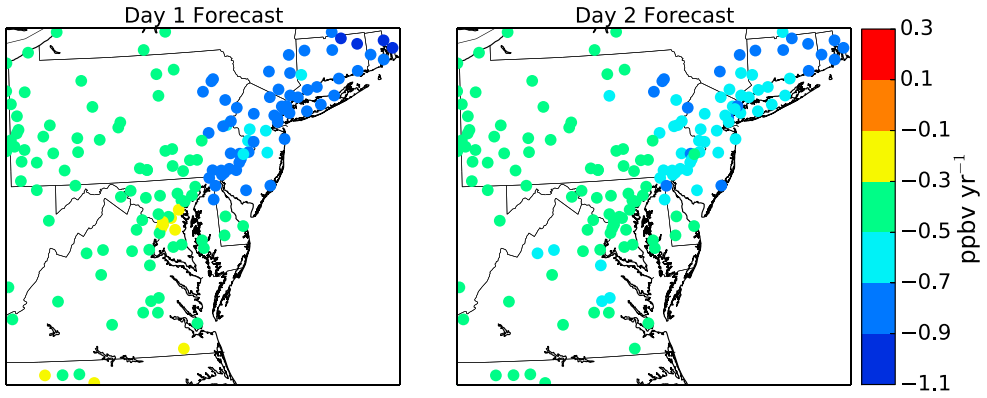

Figure 13. Slope of the least squares line regression of the root-mean-square error (comparing model predictions against observations) of the maximum daily 8-hr average ozone (MD8-O_3) per emissions inventory year during the 20% of days with the most National Ambient Air Quality Standards exceeding observations of ozone as a response to the different emissions inventory years.

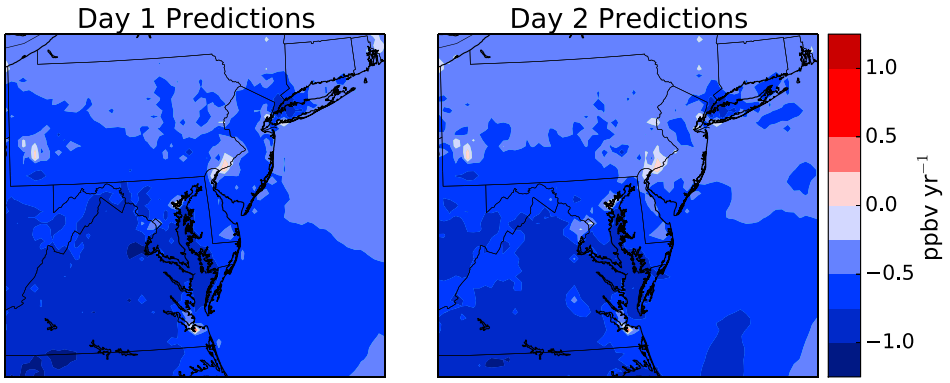

Figure 14. Slope of the least squares line regression for the modeled maximum daily 8-hr average ozone (MD8-O_3) as a response to the different emissions inventory years.

may be broadened to the idea that variability in the meteorological initial and boundary conditions are as important as variability in the emissions inventory.

Figure 12 shows the station-averaged uncertainty (root-mean-square difference or spread among members) of O_3 at each hour for each of the three subgroups as a response to variability. All three subgroups display a diurnal pattern. The uncertainty of the PBL subgroup shows negatively skewed behavior with peaks during the nighttime. The uncertainty of the emissions inventory subgroup is the smallest but shows a positive skewness and peak during the morning hours. The uncertainty of the emissions inventory subgroup also shows a small decrease in the afternoon and evening, implying a trend toward increased confidence. The meteorological initial and boundary conditions uncertainty is slightly larger in the daytime hours, but the diurnal pattern is not as pronounced compared to the emissions inventory or PBL subgroups.

3.4. Emissions Inventory Analysis

Since predicted MD8-O_3 has a clear positive bias among all sensitivity experiments, as seen in Figure 4, and the newer emissions inventories decrease error, as described previously in section 3.1, we estimated the average decrease of the RMSE due to emissions inventory updates. Figure 13 illustrates the slope of the least squares line of the RMSE of MD8-O_3 as a response to the emissions inventory year. There was one caveat: since only the emissions of NO were updated in the NEI-14, the changes in emissions of VOCs may alter the finding, but because O_3 in the Mid-Atlantic region is NO_x-sensitive, the result of changing emissions of VOCs will likely be small. Based on this analysis, an average decrease of approximately 0.59 ppbv/year of the RMSE and mean bias of MD8-O_3 was estimated.

Using a methodology similar to that used to estimate the average decrease of the RMSE, the average decrease of MD8-O_3 was also found, as shown in Figure 14. Predictions of MD8-O_3 decreased across the analysis region with each progressive emissions inventory. Over land, the average decrease of MD8-O_3 was 0.63 ppbv/year.

Table 4
Error Metrics of Daily Maximum O_3 for Each Sensitivity Experiment

Sensitivity experiment	Mean bias		RMSE		Unbiased RMSE	
	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
CNTL	14.12	14.27	21.94	21.83	16.8	16.51
NEI-05	18.32	18.67	26	25.98	18.45	18.07
NEI-14	11.62	11.94	19.63	19.71	15.82	15.68
ERA	17.38	16.4	25.42	23.91	18.55	17.4
MERRA2	14.18	14.25	21.81	22.13	16.57	16.93
ACM2	13.71	14.06	21.25	21.12	16.23	15.76
MYJ	14.35	14.31	22.75	21.65	17.65	16.24

Note. ACM2 = Asymmetric Convective Model; CNTL = control simulation; ERA = European Center for Medium Range Weather Forecasting Interim Reanalysis; MERRA = Modern Era Retrospective analysis for Research and Applications; MYJ = Mellor-Yamada-Janjic scheme; NEI = National Emissions Inventory; RMSE = root-mean-square error.

3.5. Analysis of Errors for Other O_3 Metrics

The results described to this point have focused on analysis of model-predicted MD8-O_3, due to its practical relevance for ambient air quality. However, analysis of other model-predicted metrics, such as daily average O_3, daily maximum O_3, and daily minimum O_3, can also provide important information and thus are also examined. For brevity, here only the errors of a few of such additional O_3 metrics are discussed. The model prediction period is 12 to 12 UTC daily, so there will be a difference in each metric when compared to the corresponding metric based on local standard time. This is not usually an issue for daily maximum O_3, but it can lead to differences for daily minimum O_3 and daily average O_3.

Table 5
Error Metrics of Daily Average O_3 for Each Sensitivity Experiment

Sensitivity experiment	Mean bias		RMSE		Unbiased RMSE	
	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
CNTL	17.02	16.81	21.02	20.51	12.34	11.75
NEI-05	19.24	18.92	23.28	22.59	13.1	12.33
NEI-14	15.86	15.73	19.82	19.46	11.89	11.45
ERA	19.42	18.48	23.46	22.07	13.16	12.07
MERRA2	16.84	16.84	20.65	20.71	11.96	12.06
ACM2	13.61	13.7	19.02	19.09	13.29	13.3
MYJ	15.03	14.8	19.56	19.21	12.52	12.24

Note. ACM2 = Asymmetric Convective Model; CNTL = control simulation; ERA = European Center for Medium Range Weather Forecasting Interim Reanalysis; MERRA = Modern Era Retrospective analysis for Research and Applications; MYJ = Mellor-Yamada-Janjic scheme; NEI = National Emissions Inventory; RMSE = root-mean-square error.

Table 4 lists the errors for daily maximum O_3, which are similar to results to Table 3, with a few exceptions. One difference is that the unbiased RMSE for day 1 predictions is not (statistically significantly) greater than the RMSE for day 2 predictions. Another difference is that the magnitude of the errors of MD8-O_3 (Table 3) are greater than the magnitude of the errors of daily maximum O_3 (Table 4), except for the mean bias of both days of the CNTL, NEI-14, and day 1 predictions of the ERA and ACM2.

Table 5 lists the errors in daily average O_3. In general, the mean bias of daily average O_3 was greater than that of MD8-O_3 (Table 3), with the exception of the ACM2 experiment. Both the RMSE and unbiased RMSE are greater for daily average O_3 compared to MD8-O_3. This increase in RMSE and mean bias with a widening averaging gap supports the idea that WRF-Chem predicts maximum O_3 most accurately, with errors in daily average O_3 inflated due to large errors in prediction of nocturnal O_3 (Hu, Klein, & Xue, 2013). It is also worth noting that the unbiased RMSE for daily average O_3 is smaller than that of the other O_3 metrics.

Table 6 lists the errors in daily minimum O_3. Both the mean bias and RMSE of daily minimum O_3 are greater than the corresponding values for MD8-O_3, maximum O_3, and daily average O_3, with the exception of the mean bias of the ACM2 experiment. The difference between the mean bias of the ACM2 experiment and the other sensitivity experiments suggests that the treatment of the nocturnal boundary layer in the ACM2 experiment produces less of a biased prediction, but the unbiased RMSE value indicates that this does not necessarily correspond to a more accurate prediction. Hu, Klein, & Xue (2013) provides a more in depth discussion on the treatment of the PBL scheme in relation to nocturnal O_3.

3.6. Limitations

One important limitation of this study was the number of sensitivity experiments conducted. As with most research involving groups of simulations, the number of simulations used was limited by computational resources. Since WRF-Chem was computationally more expensive than the stand-alone WRF, and we ran WRF-Chem for 1 month, we limited the number of sensitivity experiments in consideration of finite computing resources. Additionally, we wanted to compare the same number of sensitivity experiments, and we had three emissions inventories available, so we were limited by our choice to test the model uncertainty of ground-level O_3 to the variability of the emissions inventory. This limitation may be one reason the model uncertainty of all subgroups was less than the error of the observations, which further suggests there are other sources of uncertainties, such as biogenic emissions, dry deposition, chemical mechanisms, photolysis parameterizations, and convective parameterizations, in O_3 modeling over the study region. This limitation on the number of sources of sampled uncertainties are not trivial. For example, Mar et al. (2016) showed that difference in chemical mechanisms can lead to differences in predicted O_3 of 5-10 ppbv. These additional uncertainties will be subject of future research.

The role of both observational errors and unresolved emissions is also of note. Evaluating each sensitivity experiment based on observations of O_3 is subject to factors such as representativeness and measurement uncertainty, and in most cases the measurement uncertainty is considered of negligible influence. Moreover, we are comparing errors between models, and in this way, our study can also be interpreted as comparing the representativeness of the different model settings.

Table 6
Error Metrics of Daily Minimum O_3 for Each Sensitivity Experiment

Sensitivity experiment	Mean bias		RMSE		Unbiased RMSE	
	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
CNTL	22.28	21.91	27.37	27.07	15.89	15.9
NEI-05	22.45	22.15	28.11	27.79	16.91	16.78
NEI-14	22.44	22.04	27.22	26.89	15.42	15.41
ERA	23.29	23.14	28.42	28.23	16.29	16.18
MERRA2	22.25	22.14	27.66	27.6	16.45	16.48
ACM2	13.26	13.41	26.07	25.45	22.45	21.63
MYJ	19.38	19.01	25.31	25.03	16.29	16.28

Note. ACM2 = Asymmetric Convective Model; CNTL = control simulation; ERA = European Center for Medium Range Weather Forecasting Interim Reanalysis; MERRA = Modern Era Retrospective analysis for Research and Applications; MYJ = Mellor-Yamada-Janjic scheme; NEI = National Emissions Inventory; RMSE = root-mean-square error.

4. Summary and Conclusion

Predictability within the field of meteorology has been comprehensively investigated in the past few decades and understanding of modeled uncertainties and error growth dynamics has proved critical for improving

weather forecasting. In contrast, this topic in air quality modeling has been less explored. Uncertainties of simulated near-surface O_3 in the eastern United States and its predictability were investigated in this study. We assessed the predictability of O_3 by running seven WRF-Chem experiments with different meteorological initial and boundary conditions, emissions inventories, and PBL schemes over the Mid-Atlantic region for June 2016. Our results demonstrate that these particular sources of model uncertainty or possibly the methods in sampling the variability, even without air quality data assimilation, are underdispersive. Simulated ground-level O_3 was most sensitive to variability in the meteorological initial and boundary conditions. This heightened model uncertainty emphasizes the need for accurate data assimilation of meteorological data, to ensure accurate modeling of ground-level O_3. In addition, our results demonstrate how the choice of initial and boundary conditions may result in systematic biases for prediction of ground-level O_3. Therefore, the choice of meteorological initial and boundary conditions, such as the choice of GFS or ERA, may influence the findings of other research. Of the three initial and boundary conditions tested, the MERRA-2 sensitivity experiment had the smallest bias and RMSE, suggesting it is most appropriate for operational use, though in the future we will examine other operational analysis and forecast datasets as initial and boundary conditions. The enhanced model uncertainty due to variability in the initial and boundary conditions also suggests that, pending other findings, air quality ensembles for modeling O_3 should consist of perturbations in initial and boundary conditions.

Furthermore, our results show a correspondence of emissions inventory age to RMSE in predicted O_3. While newer emissions inventories lead to more accurate results than older emissions inventories, with an improvement of approximately 0.59 ppbv/year, emissions inventories are updated infrequently. Therefore, updates in emissions modeling, or inverse modeling algorithms to remedy outdated emissions—especially in NO and VOCs—may improve O_3 model accuracy.

To further generalize the findings from this study, future research is needed to extend the experiments to different months and years, and with combinations of simultaneously changing emissions inventories, PBL schemes, and meteorological initial and boundary conditions. Further research will also focus on determining the underlying reason why ground-level O_3 is more sensitive to variability in the initial and boundary conditions rather than variability in the emissions inventory or the PBL scheme.

It is also worth noting that, given the total uncertainties from the three selected sources (emissions inventories, PBL schemes, and meteorological initial and boundary conditions) are still on average smaller than the RMSEs of predicted O_3 verifying against ground observations, the current study does not fully capture all sources of O_3 forecast uncertainties. Future work will investigate differences in O_3 tendencies, including, but not limited to, differences in the photolysis rate, chemical tendency, horizontal transport, and photostationary O_3. Additional sources of uncertainty will also be investigated, such as the chemical mechanism and photolysis parameterization. Similar methods may be applied to trace chemicals beyond O_3, such as O_3 precursors or the various VOCs within the chemical mechanism.

Acknowledgments
The data used for providing initial and boundary conditions is obtained from publicly accessible data archives at NCAR and NASA. The authors thank Daniel Tong for his help with developing the NEI-14 emissions inventory. The authors thank the two anonymous reviewers for their comments that improved the manuscript, and the editors for their support. Partial financial support for this study was provided by the Pennsylvania Department of Environmental Protection. Computing was performed at the Texas Advanced Computing Center, an NSF HPC Center, where the modeling output used in this study is archived and accessible.

References

Anderson, D. C., Loughner, C. P., Diskin, G., Weinheimer, A., Canty, T. P., Salawitch, R. J., & Dickerson, R. R. (2014). Measured and modeled CO and NO_y in DISCOVER-AQ: An evaluation of emissions and chemistry over the eastern US. *Atmospheric Environment*, *96*, 78–87. https://doi.org/10.1016/j.atmosenv.2014.07.004

Azevedo, J. M., Gonçalves, F. L. T., & de Fátima Andrade, M. (2011). Long-range ozone transport and its impact on respiratory and cardiovascular health in the north of Portugal. *International Journal of Biometeorology*, *55*(2), 187–202. https://doi.org/10.1007/s00484-010-0324-2

Beekmann, M. (2003). Monte Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the atmospheric pollution over the Paris area (ESQUIF) campaign. *Journal of Geophysical Research*, *108*(D17), 8559. https://doi.org/10.1029/2003JD003391

Bei, N., Lei, W., Zavala, M., & Molina, L. T. (2010). Ozone predictabilities due to meteorological uncertainties in the Mexico City basin using ensemble forecasts. *Atmospheric Chemistry and Physics*, *10*(13), 6295–6309. https://doi.org/10.5194/acp-10-6295-2010

Bell, M. L. (2004). Ozone and short-term mortality in 95 US urban communities, 1987-2000. *JAMA*, *292*(19), 2372. https://doi.org/10.1001/jama.292.19.2372

Bell, M., & Ellis, H. (2004).Sensitivity analysis of tropospheric ozone to modified biogenic emissions for the mid-Atlantic region. *Atmospheric Environment*, *38*(13), 1879–1889. https://doi.org/10.1016/j.atmosenv.2004.01.012

Berge, E., Huang, H.-C., Chang, J., & Liu, T.-H. (2001). A study of the importance of initial conditions for photochemical oxidant modeling. *Journal of Geophysical Research*, *106*(D1), 1347–1363. https://doi.org/10.1029/2000JD900227

Butler, T. J., Vermeylen, F. M., Rury, M., Likens, G. E., Lee, B., Bowker, G. E., & McCluney, L. (2011). Response of ozone and nitrate to stationary source NO_x emission reductions in the eastern USA. *Atmospheric Environment*, *45*(5), 1084–1094. https://doi.org/10.1016/j.atmosenv.2010.11.040

Byun, D., Young, J., Gipson, G., Godowitch, J., & Binkowski, F. (1997). Description of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system.

Calvert, J. G., Orlando, J. J., Stockwell, W. R., & Wallington, T. J. (2015). *The mechanisms of reactions influencing atmospheric ozone*. New York: Oxford University Press.

Castellanos, P., Stehr, J. W., Dickerson, R. R., & Ehrman, S. H. (2009). The sensitivity of modeled ozone to the temporal distribution of point, area, and mobile source emissions in the eastern United States. *Atmospheric Environment*, *43*(30), 4603–4611. https://doi.org/10.1016/j.atmosenv.2009.05.045

Cheng, F. Y., Chin, S. C., & Liu, T. H. (2012). The role of boundary layer schemes in meteorological and air quality simulations of the Taiwan area. *Atmospheric Environment*, *54*, 714–727. https://doi.org/10.1016/j.atmosenv.2012.01.029

Cuchiara, G. C., Li, X., Carvalho, J., & Rappenglück, B. (2014). Intercomparison of planetary boundary layer parameterization and its impacts on surface ozone concentration in the WRF/Chem model for a case study in Houston/Texas. *Atmospheric Environment*, *96*, 175–185. https://doi.org/10.1016/j.atmosenv.2014.07.013

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Dudhia, J. (1989). Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, *46*(20), 3077–3107. https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2

Duncan, B. N., Yoshida, Y., Olson, J. R., Sillman, S., Martin, R. V., Lamsal, L., et al. (2010). Application of OMI observations to a space-based indicator of NO_x and VOC controls on surface ozone formation. *Atmospheric Environment*, *44*(18), 2213–2223. https://doi.org/10.1016/j.atmosenv.2010.03.010

EPA (2015). 2011 National Emissions Inventory, version 2 technical support document, (August), 37–38, 273–274. Retrieved from https://www.epa.gov/sites/production/files/2015-10/documents/nei2011v2_tsd_14aug2015.pdf

EPA: Air Markets Program Data (2018) May 01, https://ampd.epa.gov/ampd/

Frost, G. J., McKeen, S. A., Trainer, M., Ryerson, T. B., Neuman, J. A., Roberts, J. M., et al. (2006). Effects of changing power plant NO_x emissions on ozone in the eastern United States: Proof of concept. *Journal of Geophysical Research*, *111*, D12306. https://doi.org/10.1029/2005JD006354

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Gent, J. F. (2003). Association of low-level ozone and fine particles with respiratory symptoms in children with asthma. *JAMA*, *290*(14), 1859. https://doi.org/10.1001/jama.290.14.1859

Gilliam, R. C., Hogrefe, C., Godowitch, J. M., Napelenok, S., Mathur, R., & Rao, S. T. (2015). Impact of inherent meteorology uncertainty on air quality model predictions. *Journal of Geophysical Research: Atmospheres*, *120*, 12,259–12,280. https://doi.org/10.1002/2015JD023674

Grell, G. A., & Dévényi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, *29*(14), 12609. https://doi.org/10.1029/2002GL015311

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment*, *39*(37), 6957–6975. https://doi.org/10.1016/j.atmosenv.2005.04.027

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., & C., G. (2006). Estimates of global terrestrial isoprene emissions using MEGAN. *Atmospheric Chemistry and Physics Discussions*, *6*(1), 107–173. https://doi.org/10.5194/acpd-6-107-2006

Hodnebrog, Ø., Stordal, F., & Berntsen, T. K. (2011). Does the resolution of megacity emissions impact large scale ozone? *Atmospheric Environment*, *45*(38), 6852–6862. https://doi.org/10.1016/j.atmosenv.2011.01.012

Hong, S., & Lim, J. (2006). The WRF single-moment 6-class microphysics scheme (WSM6). *Journal of the Korean Meteorological Society*. Retrieved from http://www2.mmm.ucar.edu/wrf/users/phys_refs/MICRO_PHYS/WSM6.pdf

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*(9), 2318–2341. https://doi.org/10.1175/MWR3199.1

Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*(12), 4489–4532. https://doi.org/10.1175/MWR-D-15-0440.1

Hu, X. M., Doughty, D. C., Sanchez, K. J., Joseph, E., & Fuentes, J. D. (2012). Ozone variability in the atmospheric boundary layer in Maryland and its implications for vertical transport model. *Atmospheric Environment*, *46*, 354–364. https://doi.org/10.1016/j.atmosenv.2011.09.054

Hu, X. M., Klein, P. M., & Xue, M. (2013). Evaluation of the updated YSU planetary boundary layer scheme within WRF for wind resource and air quality assessments. *Journal of Geophysical Research: Atmospheres*, *118*, 10,490–10,505. https://doi.org/10.1002/jgrd.50823

Hu, X. M., Klein, P. M., Xue, M., Zhang, F., Doughty, D. C., Forkel, R., et al. (2013). Impact of the vertical mixing induced by low-level jets on boundary layer ozone concentration. *Atmospheric Environment*, *70*, 123–130. https://doi.org/10.1016/j.atmosenv.2012.12.046

Hu, X.-M., Nielsen-Gammon, J. W., & Zhang, F. (2010). Evaluation of three planetary boundary layer schemes in the WRF model. *Journal of Applied Meteorology and Climatology*, *49*(9), 1831–1844. https://doi.org/10.1175/2010jamc2432.1

Hu, X.-M., Xue, M., Kong, F., & Zhang, H. (2019). Meteorological conditions during an ozone episode in Dallas-Fort Worth, Texas and impact of their modeling uncertainties on air quality prediction. *Journal of Geophysical Research: Atmospheres*, *124*, 1941–1961. https://doi.org/10.1029/2018JD029791

Jiménez, P., Parra, R., & Baldasano, J. M. (2007). Influence of initial and boundary conditions for ozone modeling in very complex terrains: A case study in the northeastern Iberian Peninsula. *Environmental Modeling and Software*, *22*(9), 1294–1306. https://doi.org/10.1016/j.envsoft.2006.08.004

Jiménez, P. A., Dudhia, J., González-Rouco, J. F., Navarro, J., Montávez, J. P., & García-Bustamante, E. (2012). A revised scheme for the WRF surface layer formulation. *Monthly Weather Review*, *140*(3), 898–918. https://doi.org/10.1175/MWR-D-11-00056.1

Kim, S.-W., Heckel, A., McKeen, S. A., Frost, G. J., Hsie, E.-Y., Trainer, M. K., et al. (2006). Satellite-observed U.S. power plant NO_x emission reductions and their impact on air quality. *Geophysical Research Letters*, *33*, L22812. https://doi.org/10.1029/2006GL027749

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*(2), 130–141. https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Lorenz, E. N. (1969). Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, *26*(4), 636–646. https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2

Madronich, S. (1987). Photodissociation in the atmosphere: 1. Actinic flux and the effects of ground reflections and clouds. *Journal of Geophysical Research*, *92*(D8). https://doi.org/10.1029/JD092iD08p09740

Mallet, V., & Sportisse, B. (2006). Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling. *Journal of Geophysical Research*, *111*, D02311. https://doi.org/10.1029/2005JD006149

Mar, K. A., Ojha, N., Pozzer, A., & Butler, T. M. (2016). Ozone air quality simulations with WRF-Chem (v3.5.1) over Europe: Model evaluation and chemical mechanism comparison. *Geoscientific Model Development*, *9*(10), 3699–3728. https://doi.org/10.5194/gmd-9-3699-2016, https://doi.org/10.5194/gmd-9-3699-2016

Melhauser, C., & Zhang, F. (2012). Practical and intrinsic predictability of severe and convective weather at the mesoscales. *Journal of the Atmospheric Sciences*, *69*(11), 3350–3371. https://doi.org/10.1175/JAS-D-11-0315.1

Mena-Carrasco, M., Carmichael, G. R., Campbell, J. E., Zimmerman, D., Tang, Y., Adhikary, B., et al. (2009). Assessing the regional impacts of Mexico City emissions on air quality and chemistry. *Atmospheric Chemistry and Physics*, *9*(11), 3731–3743. https://doi.org/10.5194/acp-9-3731-2009

Misenis, C., & Zhang, Y. (2010). An examination of sensitivity of WRF/Chem predictions to physical parameterizations, horizontal grid spacing, and nesting options. *Atmospheric Research*, *97*(3), 315–334. https://doi.org/10.1016/j.atmosres.2010.04.005

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*(D14), 16663–16682. https://doi.org/10.1029/97JD00237

Monin, A. S., & Obukhov, A. M. (1954). Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib Geophys Inst. Acad Science USSR*, *24*(151), 163–187.

Otte, T. L., Pouliot, G., Pleim, J. E., Young, J. O., Schere, K. L., Wong, D. C., et al. (2005). Linking the Eta model with the Community Multiscale Air Quality (CMAQ) modeling system to build a national air quality forecasting system. *Weather and Forecasting*, *20*(3), 367–384. https://doi.org/10.1175/WAF855.1

Pan, L., Tong, D., Lee, P., Kim, H. C., & Chai, T. (2014). Assessment of NO_x and O_3 forecasting performances in the U.S. National Air Quality Forecasting Capability before and after the 2012 major emissions updates. *Atmospheric Environment*, *95*(x), 610–619. https://doi.org/10.1016/j.atmosenv.2014.06.020

Pleim, J. E. (2011). Comment on Simulation of surface ozone pollution in the central gulf coast region using WRF/Chem model: Sensitivity to PBL and land surface physics. *Advances in Meteorology*, *2011*, 1–3. https://doi.org/10.1155/2011/464753

Predictability (2012). In American Meteorological Society glossary of meteorology. Retrieved from http://glossary.ametsoc.org/wiki/Predictability

Ryan, W. F., Doddridge, B. G., Dickerson, R. R., Morales, R. M., Hallock, K. A., Roberts, P. T., et al. (1998). Pollutant transport during a regional O_3 episode in the mid-Atlantic states. *Journal of the Air and Waste Management Association*, *48*(9), 786–797. https://doi.org/10.1080/10473289.1998.10463737

Seaman, N. L. (2000). Meteorological modeling for air-quality assessments. *Atmospheric Environment*, *34*(12–14), 2231–2259. https://doi.org/10.1016/S1352-2310(99)00466-5

Seaman, N. L., & Michelson, S. A. (2000). Mesoscale meteorological structure of a high-ozone episode during the 1995 NARSTO-northeast study. *Journal of Applied Meteorology*, *39*(3), 384–398. https://doi.org/10.1175/1520-0450(2000)039<0384:MMSOAH>2.0.CO;2

Skamarock, W. C., Klemp, J. B., Dudhi, J., Gill, D. O., Barker, D. M., Duda, M. G., et al. (2008). A description of the advanced research WRF version 3. *Technical Report*, (June), 113. https://doi.org/10.5065/D6DZ069T

Stauffer, R. M., Thompson, A. M., Martins, D. K., Clark, R. D., Goldberg, D. L., Loughner, C. P., et al. (2015). Bay breeze influence on surface ozone at Edgewood, MD during July 2011. *Journal of Atmospheric Chemistry*, *72*(3–4), 335–353. https://doi.org/10.1007/s10874-012-9241-6

Stockwell, W. R., Kirchner, F., Kuhn, M., & Seefeld, S. (1997). A new mechanism for regional atmospheric chemistry modeling. *Journal of Geophysical Research*, *102*(D22), 25847–25879. https://doi.org/10.1029/97JD00849

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M. A., Mitchell, K., ... & Cuenca, R. H. (2004). Implementation and verification of the unified NOAH Land Surface Model in the WRF model. In *20th conference on weather analysis and forecasting/16th conference on numerical weather prediction* (Vol. 1115).

Tie, X., Brasseur, G., & Ying, Z. (2010). Impact of model resolution on chemical ozone formation in Mexico City: Application of the WRF-Chem model. *Atmospheric Chemistry and Physics*, *10*(18), 8983–8995. https://doi.org/10.5194/acp-10-8983-2010

Tong, D., Pan, L., Chen, W., Lamsal, L., Lee, P., Tang, Y., et al. (2016). Impact of the 2008 global recession on air quality over the United States: Implications for surface ozone levels from changes in NO_x emissions. *Geophysical Research Letters*, *43*, 9280–9288. https://doi.org/10.1002/2016GL069885

Tong, D. Q., Lamsal, L., Pan, L., Ding, C., Kim, H., Lee, P., et al. (2015). Long-term NO_x trends over large cities in the United States during the great recession: Comparison of satellite retrievals, ground observations, and emission inventories. *Atmospheric Environment*, *107*(2), 70–84. https://doi.org/10.1016/j.atmosenv.2015.01.035

Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., et al. (2016). Why do models overestimate surface ozone in the southeast United States? *Atmospheric Chemistry and Physics*, *16*(21), 13561–13577. https://doi.org/10.5194/acp-16-13561-2016

Uncertainty (2012). In American Meteorological Society glossary of meteorology. Retrieved from http://glossary.ametsoc.org/wiki/Predictability.

Vijayaraghavan, K., Lindhjem, C., DenBleyker, A., Nopmongcol, U., Grant, J., Tai, E., & Yarwood, G. (2012). Effects of light duty gasoline vehicle emission standards in the United States on ozone and particulate matter. *Atmospheric Environment*, *60*, 109–120. https://doi.org/10.1016/j.atmosenv.2012.05.049

Yerramilli, A., Challa, V. S., Dodla, V. B. R., Myles, L., Pendergrass, W. R., Vogel, C. A., et al. (2012). Simulation of surface ozone pollution in the central gulf coast region during summer synoptic condition using WRF/Chem air quality model. *Atmospheric Pollution Research*, *3*(1), 55–71. https://doi.org/10.5094/APR.2012.005

Žabkar, R., Koračin, D., & Rakovec, J. (2013). A WRF/Chem sensitivity study using ensemble modeling for a high ozone episode in Slovenia and the northern Adriatic area. *Atmospheric Environment*, *77*, 990–1004. https://doi.org/10.1016/j.atmosenv.2013.05.065

Zhang, F., Bei, N., Nielsen-Gammon, J. W., Li, G., Zhang, R., Stuart, A., & Aksoy, A. (2007). Impacts of meteorological uncertainties on ozone pollution predictability estimated through meteorological and photochemical ensemble forecasts. *Journal of Geophysical Research*, *112*, D04304. https://doi.org/10.1029/2006JD007429

Zhang, F., Odins, A. M., & Nielsen-Gammon, J. W. (2006). Mesoscale predictability of an extreme warm-season precipitation event. *Weather and Forecasting*, *21*(2), 149–166. https://doi.org/10.1175/WAF909.1

Zhong, M., Saikawa, E., Liu, Y., Naik, V., Horowitz, L. W., Takigawa, M., et al. (2016). Air quality modeling with WRF-Chem v3.5 in East Asia: Sensitivity to emissions and evaluation of simulated air quality. *Geoscientific Model Development*, *9*(3), 1201–1218. https://doi.org/10.5194/gmd-9-1201-2016