Published in final edited form as:
Atmos Environ (1994). 2017 ; 164: 102–116. doi:10.1016/j.atmosenv.2017.05.020.

Dynamic Evaluation of Two Decades of WRF-CMAQ Ozone Simulations over the Contiguous United States

Marina Astitha[1,*], **Huiying Luo**[1], **S. Trivikrama Rao**[1,2], **Christian Hogrefe**[3], **Rohit Mathur**[3], and **Naresh Kumar**[4]

[1]Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT, USA

[2]North Carolina State University, Raleigh, NC, USA

[3]U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

[4]EPRI, Palo Alto, CA, USA

Abstract

Dynamic evaluation of the fully coupled Weather Research and Forecasting (WRF)- Community Multi-scale Air Quality (CMAQ) model ozone simulations over the contiguous United States (CONUS) using two decades of simulations covering the period from 1990 to 2010 is conducted to assess how well the changes in observed ozone air quality are simulated by the model. The changes induced by variations in meteorology and/or emissions are also evaluated during the same timeframe using spectral decomposition of observed and modeled ozone time series with the aim of identifying the underlying forcing mechanisms that control ozone exceedances and making informed recommendations for the optimal use of regional-scale air quality models. The evaluation is focused on the warm season's (i.e., May-September) daily maximum 8-hr (DM8HR) ozone concentrations, the 4th highest (4th) and average of top 10 DM8HR ozone values (top10), as well as the spectrally-decomposed components of the DM8HR ozone time series using the Kolmogorov-Zurbenko (KZ) filter. Results of the dynamic evaluation are presented for six regions in the U.S., consistent with the National Oceanic and Atmospheric Administration (NOAA) climatic regions. During the earlier 11-yr period (1990–2000), the simulated and observed trends are not statistically significant. During the more recent 2000–2010 period, all trends are statistically significant and WRF-CMAQ captures the observed trend in most regions. Given large number of sites for the 2000–2010 period, the model captures the observed trends in the Southwest (SW) and MW but has significantly different trend from that seen in observations for the other regions. Observational analysis reveals that it is the long-term forcing that dictates how high the ozone exceedances will be; there is a strong linear relationship between the long-term forcing and the 4th highest or the average of the top10 ozone concentrations in both observations and model output. This finding indicates that improving the model's ability to reproduce the long-term component will also enable better simulation of ozone extreme values that are of interest to regulatory agencies.

[*]Corresponding author: Marina Astitha, Civil and Environmental Engineering, University of Connecticut, 261 Glenbrook Road, Storrs, CT, 06269-3037, Phone: 860-486-3941, Fax: 860-486-2298, marina.astitha@uconn.edu.

Keywords

Model evaluation; WRF-CMAQ; spectral decomposition; ozone trends; ozone design value; decadal simulations

1. Introduction

Because regional-scale air quality models are being used to support policy decisions, it is important to critically assess these models' capability in reproducing the spatio-temporal features seen in observations. Confidence in the application of air quality models for forecasting and regulatory assessments is established by conducting four types of model performance evaluation: operational, dynamic, diagnostic, and probabilistic (Dennis et al. 2010). There have been many studies focusing on the operational evaluation of regional air quality models, but only few addressed the other three components of the model evaluation framework recommended by Dennis et al. (2010). Recent studies have summarized and recommended statistical analyses to assess air quality model performance (Simon et al. 2012; Emery et al. 2016). Hogrefe et al. (2008) stated that operational model evaluation provides little insight into the reliability of the estimation of relative changes which pertains to the actual model application in the regulatory setting and argued that more emphasis should be placed on the development and application of dynamic and retrospective model evaluation approaches. Dynamic evaluation entails the assessment of the model's ability to reproduce the changes in observed air quality conditions stemming from the changes in emissions and meteorology and is one of the key components for building confidence in a model's use for policy analysis.

Dynamic model evaluation has been conducted in the past with a focus on specific pairs of years and regional areas in the contiguous U.S. (Gilliland et al., 2008; Godowitch et al., 2010; Pierce et al., 2010; Napelenok et al., 2011; Kang et al., 2013; Foley et al., 2015a, b; Stoeckenius et al., 2015; Henneman et al., 2015) or on trend analysis using decadal model simulations for South Korea and Europe (Seo et al., 2014; Banzhaf et al., 2015, Xing et al., 2015). Dynamic model evaluation for a 21-yr timeframe over the contiguous U.S. (CONUS) has never been performed. The information embedded in observations and model results, namely, changes in emissions and meteorological conditions, over the 21-year time period provides a unique opportunity to evaluate the model's ability to reproduce the observed changes in ambient ozone concentrations.

Simulated and observed air pollution concentration time series have differences that are driven by the representativeness of spatial and temporal scales as well as the stochastic nature of the atmosphere vs. a deterministic air quality modeling approach. Measurements are used to assess the severity of air pollution problems across different states and nationally, and to evaluate the performance of air pollution modeling systems (Civerolo et al., 2003). Hogrefe et al. (2001) have demonstrated that regional-scale photochemical models are capable of better reproducing the long-term variations than the short-term variations embedded in time series of observed pollutant concentrations. Ozone can be thought of as the baseline of pollution (defined as the sum of seasonal and longer-term components) that

formed from precursor emissions and modulated by meteorological conditions (Rao et al., 1996; 2011). Previous work on this subject has indicated that most of the information in the observed data is included in the long-term component which, in turn, controls the exceedances (Rao et al. 1996 and 2011; Hogrefe et al. 2000 and 2001).

In this study, we analyze ozone concentrations from 21-yr simulations performed with the fully coupled Weather Research and Forecasting (WRF)–Community Multi-scale Air Quality (CMAQ) model over CONUS for the period 1990–2010 that were described in Gan et al. (2015). Emission reductions during this period (1990–2010) associated with the EPA's Nitrogen Oxides State Implementation Plan Call and significant reductions in mobile source emissions have improved ozone air quality over CONUS. The multi-decadal simulations and spectral separation of different components that are influenced by different forcings in time series of ozone concentrations provide a valuable framework to assess the model's ability to reproduce the observed changes as well as to identify which components might require further improvement in the model. Furthermore, the dynamic evaluation of ozone exceedances and temporal components (i.e., decomposed DM8HR ozone time series) is aimed at attributing changes to emissions or meteorology for both modeled and observed ozone values.

The analysis includes evaluation of ozone trends as well as absolute and relative changes in the 4th highest (referred to as *4th*) and average of the top10 (referred to as *top10*) ozone concentrations over six regions in CONUS. In addition, the Kolmogorov-Zurbenko (KZ) filter is applied to spectrally decompose ozone time series (Rao and Zurbenko, 1994; Rao et al., 1997; Eskridge et al., 1997; Hogrefe et al., 2000; Hogrefe et al., 2003; Porter et al., 2015; Solazzo and Galmarini, 2016). The KZ filter enables the assessment of variations in the strengths of the short-term forcing (i.e., fast changing emissions, boundary layer evolution, night and day differences, synoptic scale weather-induced variations) and baseline forcing (i.e., longer-term variations attributable to emissions, policy, large-scale background and trends) in the DM8HR time series.

Section 2 provides the description of WRF-CMAQ simulations and selection of observations from the EPA's Air Quality System (AQS). Section 3 presents the methods of analysis for the ozone spectral decomposition and the statistical metrics that accompany the dynamic model evaluation. The results from this analysis are discussed in section 4. In the results section, the model evaluation starts with the traditional operational evaluation approach and is followed by an extensive overview of the dynamic evaluation results for trends, absolute and relative changes in the 4th highest and top10 ozone concentrations as well as changes in the spectrally decomposed ozone time-series (i.e. temporal components). A summary of the findings and concluding remarks are presented in Section 5.

2. Description of the coupled WRF-CMAQ Simulations and Observations

Time series of simulated 1990 – 2010 summertime (May-September) DM8HR ozone concentrations are obtained from the work of Gan et al. (2015) using the coupled WRF-CMAQ model (Wong et al., 2012). The model is configured with a 36-km horizontal grid cell spacing (Fig. 1) over CONUS and 35 vertical layers of varying thickness extending from

the surface to 50 mb (Gan et al., 2015). Time varying chemical lateral boundary conditions (BC) are obtained from a 108km x 108km WRF-CMAQ hemispheric simulation (Xing et al., 2015). The simulations are driven with internally consistent historic emission inventories (Xing et al., 2013) that include gas-phase precursors for ozone and aerosols and primary particulate matter. Positive emission trends during 1990–2010 over Asia for NMVOC, SO2, NOx and NH3 are included in the hemispheric simulations and, result in increased ozone concentrations being simulated over Asia. The boundary conditions applied to the 21-year CMAQ simulations used in this study include the mentioned emission influences from the Asian continent on the western boundary condition but a quantitative analysis of that influence is beyond the scope of this study. The simulations analyzed in the current study include direct aerosol feedback effects. The reader is referred to Gan et al. (2015) for additional information on the WRF-CMAQ two-decadal simulations.

Observations are selected based on quality-assured surface ozone measurements from EPA's AQS for monitors that have at least 80% data coverage in each year. This threshold assures that inferences on trends and ozone changes drawn over the specific timeframe are robust and representative for each station and region. The station availability for the 21-yrs, after the application of the 80% threshold, is illustrated in Fig. 1a with a total of 259 sites. We performed dynamic evaluation also for the more recent 11-yrs (2000–2010) dataset to include additional stations in our analysis. The stations for the 2000–2010 time period are illustrated in Fig. 1b, with a total of 677 sites. Increasing the number of stations in the 11-yr analysis affects the statistical rigor of the results as the data sample becomes larger. Also, more regions are better represented by adding more sites in the analysis, compared to the 21-yr period. The regionalization of the results is common for both time periods and is based on NOAA's climatic regions with some minor alterations (sites from Northern Rockies and Plains are limited (only 1 for 1990–2010 and 4 for 2000–2010), thus are not considered). The six (6) regions are: Midwest (MW), Northeast (NE), Southcentral (SC), Southeast (SE), Southwest (SW) and West (W). For the analysis presented in this study, all available stations from the AQS database based on the 80% threshold were used, without discarding specific locations that might not be represented by the model's 36 km grid configuration (e.g. coastal or urban stations cannot be resolved with model's 36 km grid). Analysis of the same 21-yr simulations considering urban, rural and suburban stations has already been performed by Foley et al. (2015c).

3. Methods of analysis

3.1 Data analysis using the Kolmogorov-Zurbenko (KZ) filter

The distinguishable scales of variations in the hourly ozone concentration time series are: intra-day (associated with fast-changing emissions, rapid rise and fall of the planetary boundary layer height, and ventilation), diurnal (day and night differences in emissions loading, photolysis, and meteorological conditions), weekly (anthropogenic activity including weekday-weekend effects), synoptic scale (weather-induced variations, days to weeks), seasonal (annual solar cycle) and inter-annual (El-Nino, climate variability and the evolution of air pollution management policies and emissions both domestically and on a hemispheric scale). The KZ filter (Rao and Zurbenko, 1994; Rao et al., 1997) is applied to

separate different scales in May-September time series of observed and modeled ozone concentrations. Since the spectral decomposition is applied to the May-September DM8HR ozone time series rather than time series of year-round hourly values, the components that can be extracted are the short-term (synoptic-scale forcing, reflecting weather-induced variations) and long-term (baseline forcing, reflecting seasonality, emissions loading, policy, and slow-changing processes) components. The KZ is a low-pass filter that uses an iterated moving average (Zurbenko, 1986; Rao et al. 1997; Hogrefe et al. 2003). The [*KZ(m,k)*] filter application is characterized by two parameters, the length of the moving average window, m, and the number of iterations, k. The filtered time series contain low-frequency fluctuations (denoted as the Baseline (*BL)* forcing from now on) while the difference between the original time series ($O_3(t)$) and the filtered time series includes the high frequency variation of the signal (denoted as short-term (*SY)* forcing from now on).

In this study, we applied KZ(5,5) to daily time series, with a window size of 5 days and 5 iterations as used in Porter et al. (2015) which is equivalent to the characteristics of the KZ(103,5) filter applied to hourly time series in Hogrefe et al. (2000). This filtering separates the weather-induced variations (short-term component *SY*) in DM8HR time series from the longer-term baseline component (*BL*). In this study, since DM8HR ozone time series are used, the Nyquist interval is 2-days, indicating that features having time scales (e.g., intra-day forcing due to fast changing emissions, boundary layer evolution, and weather conditions, diurnal forcing due to night vs. day) less than 2 days cannot be resolved (see Fig. 2 in Dennis et al., 2010). The window length and the number of iterations determine the scale separation. The 50% cut-off frequency for the KZ(5,5) is about 11 days, and, hence, scales less than 22 days are embedded in the short-term or SY forcing.

$$BL(t) = KZ(5,5) \quad (1)$$

$$SY(t) = O_3(t) - KZ(5,5) \quad (2)$$

$$O_3(t) = SY(t) + BL(t) \quad (3)$$

Where $O_3(t)$ is the original time series of observed or modeled DM8HR ozone concentrations, *BL* is the baseline or long-term component and *SY* is the short-term component. It should be noted that scale separation is not perfect as there is no rigid boundary between short-term and longer-term forcings (i.e., there is some leakage of information between the two separated components). Nonetheless, the KZ filtering reveals important information on the two components in observed and modeled time series of air pollutant concentrations, helping us in better understanding the driving forces that control ozone exceedances.

The SY forcingis a zero-mean process (filter residual) in contrast to the BL component. It is worth noting that the BL Root Mean Squared Error (RMSE) that is analyzed in Section 4 includes both the mean and variance error components while RMSE for SY reflects the variance error only. In addition, the mean of the BL component (BLmean) and the standard deviation of the SY component (SYstdev) over the summertime period are included in the analysis. The BLmean indicates the magnitude of the long-term forcing in each year, whereas the SYstdev estimates the influence of the variations of the short-term forcing (synoptic scale weather) on the total ozone concentration in that year. More details on the relationship between the two spectral components and the ozone exceedances are discussed in Section 4.

3.2 Statistical metrics

The traditional statistical metrics used in this study are described in the Appendix. In addition to the Pearson correlation coefficient (R, Eq. A1), mean bias (BIAS, Eq. A2) and RMSE (Eq. A3), we follow Wilmott (1981) in separating the RMSE into the systematic and unsystematic components by regressing the modeled (P) and observed concentrations (O) to the best-fit line (Delle Monache et al. 2005; Kang et al. 2008; 2010).

The systematic (RMSEs) and unsystematic (RMSEu) components of RMSE are estimated by Eq. 4 and 5. The RMSEs estimates the model's systematic error and hence, the better the regression between simulations and observations, the smaller the systematic error. The RMSEu is a measure of how much of the discrepancy between estimates and observations is due to random processes.

$$RMSE_S = \sqrt{\frac{\sum_{i=1}^{N} (\hat{P}_i - O_i)^2}{N}} \quad (4)$$

$$RMSE_u = \sqrt{\frac{\sum_{i=1}^{N} (P_i - \hat{P}_i)^2}{N}} \quad (5)$$

A "good" model performance dictates that the RMSE will be small in magnitude, with the systematic error approaching zero and the unsystematic approaching the RMSE (McNally, 2010):

$$RMSE^2 = RMSE_s^2 + RMSE_u^2 \quad (6)$$

The percentage of systematic (PRMSEs) or unsystematic (PRMSEu) error is calculated as the ratio of the squared RMSEs or RMSEu divided by the squared total RMSE in accordance with Eq. 6. Finally, the assessment of changes in ozone concentration (modeled

and observed) for the time period of consideration is performed by the calculation of trends (ppb/yr) over a number of years as well as absolute changes (AC; ppb) and relative changes (RC; %) (Eq. 7, 8) between specific pairs of years.

$$AC = FutureYear - BaseYear \quad (7)$$

$$RC = 100 * (FutureYear - BazeYear)/BaseYear \quad (8)$$

4. Results and Discussion

4.1 Operational evaluation

Operational evaluation has been the commonly-used approach to examine how well the model-simulated concentrations compare to the observations in an overall sense (Dennis et al. 2010). The standard statistical metrics used in operational evaluation (R, BIAS, RMSE) provide a broad overview of the model performance. In the following, we present operational evaluation of modeled DM8HR, 4th highest, average of the top10 ozone concentrations as well as the BL and SY components of the ozone time series.

The density scatter plots for DM8HR ozone concentrations during the 21-yrs reveal a different behavior for each region in the CONUS (Fig. 2). The correlation ranges between 0.5–0.75 and the RMSE between 11–17 ppb. Most of the values are between 20 and 90 ppb for both observations and modeled concentrations, except for the Southwest region (where only 10 sites are available). In the Midwest (MW) and Northeast (NE), the model overestimates the observed DM8HR ozone values. Many of these high simulated values are found at stations near water bodies in the NE, SE, and MW regions, possibly due to low PBL heights (stable conditions) in the model near large water bodies (Athanassiadis et al. 2002; Rao et al. 2003). The modeled ozone concentrations for the West (W), with most of the sites primarily in California, reflect overestimation of the low observed DM8HR concentrations and underestimation of the high observed values. One potential reason for such discrepancies near complex terrain areas is the coarse resolution of the simulations (36 km) that hinders the accurate representation of observed ozone concentrations.

When the later 11-yrs of the same dataset (2000–2010) are considered, the RMSE for the DM8HR ozone values decreases within a range of 1–2.6ppb across all regions (results included in the Supplement). Since the RMSE is weighted towards large errors, when high ozone concentrations are reduced, a reduction in RMSE is also expected. For the 2000–2010 period with 677 stations available nation-wide (Fig. 1b), the correlation also increases and RMSE is reduced (Fig. S1 in the supplement) compared to the 21-yrs, with a notable change for the West where the underestimation of the high observed values is no longer evident. The comparison of model performance between 259 and 677 stations for the recent 11-yr period does not show substantial improvement (Fig. S1, S2), even though the stations are more than

doubled for some regions. This is an indication that the simulation period is more important than the number of available stations for the model evaluation.

Before continuing with the comparison of observed and modeled values, it is important to explain the rationale behind selecting the BLmean as one of the variables in this study. To illustrate how BL and SY components influence ozone exceedances (i.e., the 4th highest and top10), the 21-yr observed DM8HR ozone concentrations are spectrally decomposed at each station. The 21 SY components are superimposed on a given year's BL component and the resulting 4th highest and top10 are determined. Then, the highest and lowest 4th and top 10 from those 21 values associated with that BL forcing are determined and this is repeated for 21 BL components. The maximum 4th highest and top 10 values for each region are included in Table 1, indicating the ozone extreme values of concern for regulatory applications. As expected, the highest BLmean is recorded during the earlier years (1990–1998) where higher emissions loading was prevalent and the lowest BLmean during the more recent years (2009). The only exception is SW where both high and low BLmean values are recorded in the earlier decade. For example, in the Northeast region, with the BLmean level at 71 ppb in 1991, the 4th highest value reaches 124 ppb, but when the BL level is at 44 ppb in 2009, 4th highest value reaches only 89 ppb (Table 1, NE). These results illustrate that when the BL is low, even strong SY forcing did not lead to very high values for the 4th highest or top10 ozone concentrations. High ozone values are found for the West region (Table 1, W) for both high and low BLmean but the lowest BLmean is still rather high (71ppb) compared to other regions that have 44–50ppb (Table 1, Lowest BLmean column). The highest BLmean for the West region is 104ppb which gives a 4th highest of 163ppb. The results seen for the West are consistent with those in the other regions. This observational feature reveals that it is the BL that dictates how high the 4th highest or top10 value could reach under different short-term forcings.

The relationship between BLmean and 4th highest ozone concentration for the CONUS is highly linear as indicated by the observations (Fig. 3a) for all years in the 1990–2010 timeframe. The model captures very well this linear relationship seen in the observations, with the main difference seen at some stations (BLmean is lower and 4th highest is higher than the observed in the upper range of the concentration distribution) (Fig. 3b). Similar results have been shown by Porter et al. (2017) for the domain-wide monthly ozone. Linearity also exists between the standard deviation of the short-term component (SYstdev) and the 4th highest (Fig. 3c, d), as it is well-known that synoptic-scale meteorological conditions greatly influence ozone exceedances. These results lend further support to Hogrefe et al. (2000) that BL and SY forcings should be viewed as the necessary and sufficient conditions, respectively, for the observed extreme ozone values. In other words, even if the SY forcing is very strong, high ozone levels cannot be reached unless also accompanied by a reasonably high BL level. The general conclusion is that ozone exceedances are strongly controlled by the long-term component (BLmean) and if the model is not capable in accurately simulating the BLmean, it will most likely have difficulties in capturing the ozone exceedances.

The spatial distribution of correlation and RMSE (Fig. 4a, b) for seasonal BLmean for 1990 – 2010 indicates that the model is more skillful in reproducing the strength of the long-term

component for the Eastern U.S. compared to the West. Similar patterns apply to the variation of the SY component (Fig. 4c,d; SYstdev) as well as the 4th highest and top10 ozone concentrations (Fig. 4e-h). Large errors and low correlations are found in the Western U.S. as well as near coastal locations in the Eastern U.S. and near the Great Lakes. The BL component of the DM8HR ozone time series exhibits larger errors compared to the SY component (Fig. 5), keeping in mind that the values of BL are larger in magnitude than the SY component. In addition, as stated in Section 3.1, the SY component is a zero-mean process and, thus, the RMSE includes only the variance error, whereas the RMSE for BL includes both bias and variance. It is important to note here that the results shown in Fig. 4 are based on one value per station per year (i.e. 21 data points at each station) while the results in Fig. 5 are based on daily time-series (~153 × 21 values at each station). This means that the results in Fig. 4 (BLmean, SYstdev, 4th and top10) are affected by inter-seasonal variability while in Fig. 5 (BL and SY) both intra- and inter-seasonal variability is included.

The coarse grid resolution of the WRF-CMAQ simulations is the main factor influencing the representation of complex terrain areas like coastal stations, urban areas or mountainous regions and is likely the main factor leading to poorer model performance in certain locations. Another factor is the long timeframe over which the comparison is done (21-years); more specifically, the model has shown better performance for the recent 11 years (2000–2010) of the 21-yr study period that is characterized by large reductions in emissions and associated ozone exceedances.

4.2 Dynamic evaluation

4.2.1 Trends in ozone concentrations—Trends in the high ozone concentrations for each U.S. region displayed a decreasing pattern from 1990 to 2010, which is expected given the significant emission reductions nationwide and has been noted in studies analyzing observed ozone trends (Simon et al., 2015). This section's discussion begins with the analysis of trends in the 10th, 50th and 90th percentiles of DM8HR ozone concentration from both observations and CMAQ results (Fig. 6). Looking at the trends over the 21-year period (1990–2010), the trends in the 10th percentile are mostly positive (except for SE), the trends in the 50th percentile have mixed features and the trends in the 90th percentile are mostly negative (Fig. 6a). The model generally follows the observed trend but in some cases it gives larger ozone reductions than the observations (SE, MW, NE). These results indicate that the ozone distributions have been getting narrower over these two decades. Trends for the earlier and more recent decades are analyze separately to further understand this behavior.

During the earlier decade (1990–2000, Fig. 6b), trends behave differently depending on the region. Trends for all percentiles are mostly negative for the West with the observations showing larger variability than the model trends. Southwest (SW) and Southcentral (SC) show positive trends, with SW observed trends also exhibiting larger variability. The model generally agrees with the observations for these three regions. One thing to mention here is that emission inventories from Asia, Mexico and Canada are included in the hemispheric CMAQ simulations that provided the chemical boundary conditions but the uncertainty of those inventories might be influential in the depiction of trends and their variability for the

western U.S. regions. For SE, MW and NE the modeled vs. observed trends have larger differences that sometimes lead to opposite signs. The analysis of the 4th highest trends has shown that the 1990–2000 time period has statistically insignificant mean trends for all regions (this will be discussed in detail later in this section). During the more recent decade (2000–2010; Fig. 6c), trends in the 10th percentile are mixed, are mostly negative in the 50th percentile and exclusively negative in the 90th percentile. The model agrees well with the observations in most cases for the recent decade. The improved model trends for each percentile is likely attributed to the accuracy of emission inventories for the more recent decade compared to the earlier period.

Regional averages of simulated and observed station-specific trends (in ppb/yr) calculated at each monitor are presented in Table 2. The 21-yr period is divided in two 11-yr periods, 1990–2000 and 2000–2010, due to different direction in the trend in the 4th highest ozone concentration at some locations. In addition, the 11-yr trend in the 4th highest during 2000–2010 for the 677 sites is included in Table 2. Statistically significant trends with $p<0.05$ (bold font in Table 2) are generally seen for the more recent 11-yr periods for both observations and simulations. Moreover, the 95% bootstrap confidence intervals (Efron, 1982; Rao et al. 1985; Porter et al. 1997) provide evidence on whether the observed and simulated trends are significantly different (Table 2 values in italics and Fig. 7).

Observed and simulated trends for 1990–2000 are not statistically significant and the model captures the observed trend with statistical confidence for SW, SC, NE and CONUS. During the more recent 11-yr period all trends are statistically significant (only exception is model trend for the West), regardless of the number of stations used in the analysis. The significance of ozone trends is most probably related to emission reduction policies over CONUS that led to steeper decreases of emissions during the more recent decade (2000–2010) (also shown in Fig. 12 of Xing et al. 2013). In contrast, the results deviate for the two sets of stations when looking at how significant the differences in the trends are (Fig. 7b, c). The model captures the trends seen in observations (i.e., overlapping confidence intervals) for SW, SE, MW and SC and under-predicts the trends for W, NE and CONUS when the 259 sites are used (Fig. 7b). The model captures the trend for MW and SW and under-predicts the trends for all other regions when 677 sites are considered in the analysis (Fig. 7c). In general, the model underestimates the magnitude of ozone reductions in all cases where the trends have been statistically significant.

Looking more carefully at each region, different features in the variability of modeled vs. observed trends are identified. For the Western U.S., observations exhibited larger variability than CMAQ-simulated values until the year 2000; during 2000–2010, this variability decreased. Data for SW U.S. was limited (Fig. S3b), but nevertheless the model showed larger spatial variability compared to observations for all years. For SC U.S., the simulated 4th highest is closer to the observed one with observations showing outliers mostly after the year 2000. The model has larger 4th highest values than the observations for all years for SE U.S. The same applies to MW and NE U.S. with many outliers associated with the model. From the previous spatial maps (Fig. 4e-f), it is evident that outliers were mainly concentrated in the coastal areas (N. Atlantic and the Great Lakes). Because of the strong relationship between the BL component and the 4th highest ozone concentrations (Fig. 2),

the observed and simulated BLmean trends are calculated for the 21-yr period (Fig. S3 lower panels). Except for the W and SW regions, simulated BLmean is larger than observed BLmean for all years in SC, SE, MW and NE. This explains the behavior of the 4th highest described previously. A closer look at the trends in 4th highest and BLmean for the more recent decade (2000–2010) is also provided in the supplement (Fig. S4), showing that the modeled 4th highest ozone trends are very close to the corresponding observed values for all regions, with the same outliers noted before for SW, NE and SE. The simulated BLmean is higher than the observed in some cases without an evident implication to the 4th highest. However, in the case of the MW and SE, the over-estimated BLmean (Fig. S4e-f, lower panels) likely contributed to the higher simulated 4th values (Fig. S4e-f, upper panels) given the linear relationship shown in Fig. 3.

The spatial distribution of the 2000–2010 trends for 4th, top10, BLmean, and SYstdev as well as observed vs. simulated trend differences (difference=model-observations) for every site are presented in Fig. 8. The simulated BLmean trend is generally smaller than the observed trend at most stations (Fig. 8a-b). The standard deviation of the short-term forcing (SYstdev) exhibits trends small in magnitude with the simulated being smaller than the observed ones in most regions (Fig. 8j-k). Exceptions are denoted for the BLmean and SYstdev trends in coastal areas. Trends in Top10 and 4th highest ozone are almost identical (Fig. 8d-i). During the 2000–2010 period, there is a stronger negative trend in the observed 4th and top10 compared to the simulated for the SC, SE, MW and NE, a feature also evident in Table 2. Typically, the sites with relatively larger magnitude of trend in BLmean also exhibit higher magnitudes in trends for top 10 and 4th highest, again suggesting that BLmean has an important impact on the model's ability to capture the trends in the upper tail of the ozone distribution. CMAQ captures the negative observed trend in all regions, but the magnitude is different. In the East, the model captures the observed 4th and top10 decreasing trends at most stations (Fig. 9e, h) although the modeled trends are less negative compared to those in observations, indicating a slower pace in the 4th and top10 ozone decreases estimated by the model. For comparison purposes, the spatial distribution of trends during the 21-yr period (1990–2010) is available in the supplement (Fig. S5).

4.2.2 Absolute and Relative Changes in Ozone Concentrations Between Specific Time Intervals—

The first part of dynamic model evaluation included analysis of temporal trends in the 4th highest and top10 values as discussed in the previous section (4.2.1). The second part is devoted to the assessment of how well the model simulated the changes in ozone concentration over specific time intervals when compared with observations. Having a 21-yr WRF-CMAQ simulation provides a unique opportunity to explore time intervals ranging from 2 to 15 years and identify strengths and weaknesses of simulated changes over those time intervals.

First, the change in top10 paired-in-time concentrations from 2000 to 2010 are analyzed (Table 3). The values in parenthesis denote unpaired-in-time values (e.g. when the simulated top10 values are calculated independent of the dates indicated by observations as was done for the analysis in the previous sections). The sign of the observed change between 2000 and 2010 is simulated accurately by the model with both paired- and unpaired-in-time combinations (except for the paired simulated change for the West). The ratio (CMAQ

change/observed change) ranges from 0.6 to 1.6 with CMAQ underestimating the change for all regions except for the SW. The unpaired comparison performed better than the paired comparison for the West U.S., where the model change has the same sign but the magnitude is almost half that of the observed one. In all subsequent analyses, unpaired-in-time is used rather than paired-in-time top10 values.

Next, the model's performance in simulating the changes in the 4th highest and top10 ozone concentrations across all possible 5-yr intervals over the 21-yr period is investigated. To this end, the BIAS, total RMSE and percentages of systematic and unsystematic components of RMSE are presented. These metrics are computed for absolute 5-yr changes (AC; Eq.7) and relative 5-yr changes (RC; Eq. 8) in the 4th highest ozone concentrations as well as the 4th highest concentrations across all 21 years (Fig. 9). The results clearly demonstrate the reduction of error and bias in changes compared to absolute concentrations (Fig. 9a-c, j-l) even for the Western U.S. The maximum RMSE is reduced from 85ppb to 30ppb and the range of bias from (−40ppb, 83ppb) to (−6ppb, 19ppb) when considering the modeled changes in 4th highest concentrations rather than the absolute 4th highest concentrations. Similar results are found for the average of the unpaired top10 ozone concentrations (not shown here).

An important feature of this analysis is the quantification of systematic vs. unsystematic components of the RMSE that delineate whether it is feasible to make model improvements or the model is as good as it can be, since model processes improvements target reduction of systematic biases (Eq. 3–7, Section 3.2). More specifically, systematic error was larger than unsystematic for the 4th highest ozone concentration (Fig. 10d, g) and especially for Western U.S. This indicates that improvements in the model are possible to reduce the systematic error. As expected, the absolute and relative changes (AC and RC) exhibited larger percentages of unsystematic errors compared to systematic (Fig. 9 e,h and f,i), which denote that using ozone concentrations in a relative sense than in their absolute concentration levels makes better use of the model's strengths (because random errors cannot be reduced). A few stations in S. California showed persistently large systematic errors even for AC and RC, which are worthy of further investigation to identify causes for these errors and to improve model simulations.

To expand on these findings, a comprehensive analysis of all possible time intervals is presented for changes in the BLmean, standard deviation of the SY (SYstdev), 4th highest, and unpaired top10 ozone concentrations for the 21-yr timeframe (Fig. 10 and more detailed boxplots in S6–S8). The statistical metrics discussed are bias, RMSE, RMSEu and RMSEs for every combination (Fig. 10). We focus on the BLmean, 4th highest and unpaired top10.

Starting with BLmean (Fig. 10a), the median bias of changes across all simulation year intervals (blue line) is smaller than the bias of the absolute BLmean value (blue dotted line). The same applied to RMSE (black) and RMSEs (green) which increase as the year-intervals increase from 1 to 15 years. The RMSEu remained almost constant and equal to the RMSEu of the actual BLmean. The median errors of changes in the 4th highest and unpaired top10 ozone concentrations showed a similar pattern, with a reduced bias compared to absolute values for all change intervals (Fig. 10 c, d). RMSEs increases with the time interval and

RMSEu is almost constant with a slight decrease after the 10-yr interval. Changes across time intervals between 2 and 15 years produce lower bias and systematic RMSE, higher RMSEu and similar RMSE compared with the absolute values, coinciding with the time span over which models are often used in a relative sense for regulatory applications.

5. Summary

Two decades of WRF-CMAQ simulations provided a unique opportunity to assess changes in DM8HR ozone caused by emission reduction policies and evaluate the model's capability to reproduce the observed changes. Dynamic evaluation of WRF-CMAQ simulations for the period 1990–2010 has been performed here by analyzing the behavior of temporal trends, absolute vs. relative changes in the 4th highest, top10 and spectrally-decomposed ozone temporal components (i.e., short-term and long-term forcings). The main objectives of this study are to assess the model's ability to reproduce the changes seen in the ozone observations and identify model's strengths and weaknesses in simulating the absolute vs. relative ozone concentrations. The main findings from this dynamic evaluation study are summarized below:

- Spectral decomposition of 21-yrs observed DM8HR ozone time series revealed that it is the magnitude of the long-term forcing (i.e. baseline component), not the strength of the short-term forcing (synoptic-scale weather-induced variations), that dictates how high the 4th highest or top10 value could reach.
- A strong linear relationship between the level of the baseline and the ozone exceedances (4th highest/top10) is evident in both observations and model simulations. This suggests that improving the model's ability to reproduce the long-term component would also benefit the simulation of extreme values that are of interest to regulatory agencies.
- The 21-year trends in the 10th percentile are mostly positive, mixed for the 50th and negative for the 90th percentile in most regions, indicating that ozone distributions are becoming narrower during these two decades. CMAQ successfully represented the trends in the more recent decade (2000–2010) and exhibited variable performance during the earlier decade.
- Trends in simulated and observed ozone are found to be statistically significant for the 2000–2010 period, but not during the earlier decade (1990–2000). For the more recent 2000–2010 period, WRF-CMAQ captured the observed trend in most regions and when utilizing a larger number of sites, the model had captured the observed trends in the Southwest and Midwest but underestimated the observed trends in the other regions.
- Model errors are smaller for temporal changes in the 4th and top10 values compared to those in their absolute concentration levels.
- Larger systematic model error compared to unsystematic for the 4th highest concentration indicates that improvements in the model are possible to help reduce systematic errors in simulating the extreme values.

- Changes across different time intervals ranging from 2 to 15 years produce similar RMSE, higher unsystematic and lower systematic errors and bias compared to the absolute 4th and top10 concentration levels, suggesting air quality models are more suitable for use in the relative than in the absolute sense for regulatory applications.

This dynamic evaluation study reveals that the CMAQ model is better at reproducing temporal changes in ozone exceedances than their absolute concentration levels. There is a dependency of the model performance on region and time period, which is mainly driven by emission inventory accuracy, uncertainties in boundary conditions and deposition. Additional research is necessary to understand these aforementioned influences and provide guidance for future improvements in model simulations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The multi-decadal WRF-CMAQ simulations analyzed in this work were conducted previously as part of a larger research effort on investigating trends in air pollution and radiation and we gratefully acknowledge the efforts of Chuen-Meei Gan, Jia Xing, David Wong, and Jon Pleim in performing these simulations and making model outputs available for this study. Although this work has been reviewed and approved for publication by the U.S. Environmental Protection Agency, it does not reflect Agency's views and policies. Two of the authors (MA and HL) gratefully acknowledge the support of the Electric Power Research Institute (EPRI) for this research under Contract #00–10005071, 2015–2017.

APPENDIX

Statistical metrics employed in the evaluation of simulated ozone concentrations:

$$R = \frac{\frac{\sum_{i=1}^{N}(P_i \cdot O_i)}{N} - \frac{\sum_{i=1}^{N} P_i}{N} \cdot \frac{\sum_{i=1}^{N} O_i}{N}}{\sqrt{\frac{\sum_{i=1}^{N} P_i^2}{N} - \left(\frac{\sum_{i=1}^{N} P_i}{N}\right)^2} \cdot \sqrt{\frac{\sum_{i=1}^{N} O_i^2}{N} - \left(\frac{\sum_{i=1}^{N} O_i}{N}\right)^2}} \quad \text{(A1)}$$

$$BIAS = \frac{\sum_{i=1}^{N}(P_i - O_i)}{N} \quad \text{(A2)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (P_i - O_i)^2}{N}} \quad \text{(A3)}$$

The linear regression provides an estimated (or regressed) prediction $(\widehat{P})$ with a and b the least squares regression coefficients (Eq. A4).

$$\widehat{P}_i = a + b * \ O_i \quad \text{(A4)}$$

References

Athanassiadis GA , Rao ST , Ku JY ., Clark R , 2002: Boundary Layer Evolution and its Influence on Ground-Level Ozone Concentrations. Environmental Fluid Mechanics (2002) 2: 339. doi: 10.1023/A:1020456018087.

Banzhaf S , Schaap M , Kranenburg R , Manders AMM , Segers AJ , Visschedijk AHJ , van der Gon HAC , Kuenen JJP , van Meijgaard E , van Ulft LH and Cofala J . Dynamic model evaluation for secondary inorganic aerosol and its precursors over Europe between 1990 and 2009. Geosci. Model Dev, 8, 1047–1070, 2015.

Civerolo K , Mao H , Rao ST , The Airshed for Ozone and Fine Particle Pollution in the Northeastern United States. Pure and Appl. Geophy., 160:81–105, 2003.

Delle Monache L , Nipen T , Deng X , Zhou Y , and Stull R (2006), Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction, J. Geophys. Res, 111, D05308, doi:10.1029/2005JD006311.

Dennis R , Fox T , Fuentes M , Gilliland A , Hanna S , Hogrefe C , Irwin J , Rao ST , Scheffe R , Schere K and Steyn D , 2010 A framework for evaluating regional-scale numerical photochemical modeling systems. Environmental Fluid Mechanics, 10(4), pp. 471–489.21461126

Efron B 1982: The Jackknife, the Bootstrap and other resampling plans Society for Industrial and Applied with the Regional Air Pollution Study Data Base. Mathematics, Philadelphia, PA.

Emery C , Liu Z , Russell AG , Odman MT , Yarwood G & Kumar N , 2016: Recommendations on Statistics and Benchmarks to Assess Photochemical Model Performance, Journal of the Air & Waste Management Association, 10.1080/10962247.2016.1265027.

Eskridge RE , Ku J-Y , Rao ST , Porter PS , Zurbenko IG , 1997 Separating different scales of motion in time series of meteorological variables. Bull. Am. Meteorol. Soc. 78, 1473–1483.

Foley KM , Hogrefe C , Pouliot G , Possiel N , Roselle SJ , Simon H and Timin B , 2015a Dynamic evaluation of CMAQ part I: Separating the effects of changing emissions and changing meteorology on ozone levels between 2002 and 2005 in the eastern US. Atmos. Environ, 103, pp. 247–255.

Foley KM , Dolwick P , Hogrefe C , Simon H , Timin B and Possiel N , 2015b Dynamic evaluation of CMAQ part II: Evaluation of relative response factor metrics for ozone attainment demonstrations. Atmospheric Environment, 103, pp. 188–195.

Foley KM , Hogrefe C , Xing J , Gan C , Wong D , Pleim J , Mathur R , and Roselle S , 2015c: A Comparison of Observed and Simulated 1990 – 2010 U.S. Ozone Trends, A&WMA 108th Annual Conference, Raleigh, NC, 6 25th, 2015.

Galmarini S , Kioutsioukis I , and Solazzo E , 2013: E pluribus unum*: ensemble air quality predictions. Atmos. Chem. Phys, 13, 7153–7182.

Gan C-M , Pleim J , Mathur R , Hogrefe C , Long CN , Xing J , Wong D , Gilliam R , and Wei C : Assessment of long-term WRF–CMAQ simulations for understanding direct aerosol effects on radiation “brightening” in the United States, Atmos. Chem. Phys, 15, 12193–12209, doi:10.5194/acp-15-12193-2015, 2015.

Gilliland AB , Hogrefe C , Pinder RW , Godowitch JM , Foley KL and Rao ST , 2008 Dynamic evaluation of regional air quality models: assessing changes in O 3 stemming from changes in emissions and meteorology. Atmospheric Environment, 42(20), pp. 5110–5123.

Godowitch JM , Pouliot G , Rao ST , 2010 Assessing multi-year changes in modeled and observed NOx concentrations from a dynamic evaluation perspective. Atmos. Environ. 44 (24), 2894–2901.

Godowitch JM , Gilliam RC and Rao ST , 2011: Diagnostic Evaluation of the Chemical and Transport Processes in a Regional Photochemical Air Quality Modeling System. Atmos. Environ. 45, 3977–3987.

Godowitch JM , Gilliam RC , Roselle SJ , 2015: Investigating the impact on modeled ozone concentrations using meteorological fields from WRF with an updated four–dimensional data assimilation approach. Atm. Poll. Res, 305–311, doi: 10.5094/APR.2015.034.

Henneman LRF , Holmes HA , Mulholland JA , Russell AG , 2015 Meteorological detrending of primary and secondary pollutant concentrations: Method application and evaluation using long-term (2000–2012) data in Atlanta. Atmospheric Environment 119 (2015) 201–210.

Herwehe JA , Otte TL , Mathur R , Trivikrama Rao S , 2011: Diagnostic analysis of ozone concentrations simulated by two regional-scale air quality models, Atmospheric Environment, Volume 45, Issue 33, October 2011, Pages 5957–5969, ISSN 1352–2310.

Hogrefe C , Trivikrama Rao S , Zurbenko Igor G. , and Porter P. Steven , 2000: Interpreting the Information in Ozone Observations and Model Predictions Relevant to Regulatory Policies in the Eastern United States. Bulletin of the American Meteorological Society, Vol. 81, No. 9, 9 2000.

Hogrefe C , Rao ST , Kasibhatla P , Hao W , Sistla G , Mathur R , McHenry J , 2001: Evaluating the performance of regional-scale photochemical modeling systems: Part II—ozone predictions, Atmospheric Environment, 35 (2001), pp. 4175–4188.

Hogrefe C , Somaraju Vempaty S Rao Trivikrama , Porter P.Steven , 2003: A comparison of four techniques for separating different time scales in atmospheric variables, Atmospheric Environment, Volume 37, Issue 3, January 2003, Pages 313–325, ISSN 1352–2310, 10.1016/S1352-2310(02)00897-X.

Hogrefe C , Civerolo KL , Hao W , Ku JY , Zalewsky EE and Sistla G , 2008 Rethinking the assessment of photochemical modeling systems in air quality planning applications. Journal of the Air & Waste Management Association, 58(8), pp. 1086–1099.

Hogrefe C , Barry Lynn , Richard Goldberg , Cynthia Rosenzweig , Eric Zalewsky , Winston Hao , Prakash Doraiswamy , Kevin Civerolo , Jia-Yeong Ku , Gopal Sistla , P.L. Kinney , 2009: A combined model–observation approach to estimate historic gridded fields of PM2.5 mass and species concentrations, Atmospheric Environment, 43, 16,Pages 2561–2570, ISSN 1352–2310, 10.1016/j.atmosenv.2009.02.031.

Hogrefe C , Hao W , Zalewsky EE , Ku J-Y , Lynn B , Rosenzweig C , Schultz MG , Rast S , Newchurch MJ , Wang L , Kinney PL , and Sistla G , 2011: An analysis of long-term regional-scale ozone simulations over the Northeastern United States: variability and trends, Atmos. Chem. Phys, 11, 567–582, www.atmos-chem-phys.net/11/567/2011/. doi:10.5194/acp-11-567-20112010.

Kang D , Mathur R , Rao ST , and Yu S (2008), Bias adjustment techniques for improving ozone air quality forecasts, J. Geophys. Res, 113, D23308, doi:10.1029/2008JD010151.

Kang D , Mathur R , and Rao Trivikrama S. : Assessment of bias-adjusted PM2.5 air quality forecasts over the continental United States during 2007, Geosci. Model Dev, 3, 309–320, doi:10.5194/gmd-3-309-2010, 2010.

Kang D , Hogrefe C , Foley KL , Napelenok SL , Mathur R and Rao ST , 2013 Application of the Kolmogorov–Zurbenko filter and the decoupled direct 3D method for the dynamic evaluation of a regional air quality model. Atmospheric Environment, 80, pp. 58–69.

McNally D , 2010: The MMIFstat Statistical Analysis Package, Version 1. Prepared for Tim Allen (Fish and Wildlife Service) and John Vimont (National Park Service). Available at: https://www3.epa.gov/scram001/models/relat/mmifstat/MMIFStat.Users.Guide.2010.02.22.pdf.

Napelenok SL , Foley KM , Kang D , Mathur R , Pierce T and Rao ST , 2011 Dynamic evaluation of regional air quality model's response to emission reductions in the presence of uncertain emission inventories. Atmospheric Environment, 45(24), pp. 4091–4098.

Pierce T , Rao ST , Porter PS , and Ku J-Y , 2010: Dynamic Evaluation of a Regional Air Quality Model: Assessing the Weekly Cycle in the Observations and Model Outputs, Atmos. Env., 44, 3583–3596, doi:10.1016/j.atmosenv.2010.05.046

Porter PS , Rao ST , Hogrefe C , Mathur R , 2017: A reduced form model for ozone based on two decades of CMAQ simulations for the continental United States, J. Atmos. Poll. Res., 10.1016/j.apr.2016.09.00, In Press.

Porter PS , Rao ST , Hogrefe C , Gego E and Mathur R , 2015 Methods for reducing biases and errors in regional photochemical model outputs for use in emission reduction and exposure assessments. Atmospheric Environment, 112, pp. 178–188.

Porter PS , Rao ST , Ku JY , Peroit R , Dakins M , 1997: Small Sample Properties of Nonparametric Bootstrap T Confidence Intervals, J. Air & Waste Manage. Assoc, 47: 1197–1203, 11, 1997.

Rao ST , Sistla G , Pagnotti V , 1985: Resampling and extreme value statistics in air quality model performance evaluation, Atmospheric Environment 1985 , 19, 1503.

Rao ST , Zurbenko IG , 1994 Detecting and tracking changes in ozone air quality. Journal of Air and Waste Management Association 44, 1089–1092.

Rao ST , Zurbenko I , Porter PS , Ku M and Henry R , 1996: Dealing with the ozone non-attainment problem in the northeastern United States, Environmental Manager, p17–31, January 1996.

Rao ST , Zurbenko IG , Neagu R , Porter PS , Ku JY , and Henry RF , 1997: Space and time scales in ambient ozone data. Bull. Amer. Meteor. Soc, 78, 2153–2166.

Rao S , Ku J , Berman S , Zhang K , Mao H , 2003: Summertime Characteristics of the Atmospheric Boundary Layer and Relationships to Ozone Levels over the Eastern United States. Pure Appl. Geophys. (2003) 160: 21. doi:10.1007/s00024-003-8764-9.

Rao ST , Porter P , Mobley D , Hurley F , 2011: Understanding the spatio-temporal variability in air pollution concentrations, Environmental Manager, p42–48, November, 2011.

Seaman NL , 2000: Meteorological modeling for air-quality assessments. Atmos. Environ, 34, 2231–2259.

Seo J , Youn D , Kim JY and Lee H , 2014 Extensive spatiotemporal analyses of surface ozone and related meteorological variables in South Korea for the period 1999–2010. Atmospheric Chemistry and Physics, 14(12), pp. 6395–6415.

Simon H , Reff A , Wells B , Xing J , and Frank N , 2015: Ozone Trends Across the United States over a Period of Decreasing NOx and VOC Emissions, Environmental Science & Technology 2015 49 (1), 186–195, DOI: 10.1021/es504514z

Simon H , Baker KR , and Phillips S . 2012 Compilation and interpretation of photochemical model performance statistics published between 2006 and 2012. Atmos. Environ. 61, 124–139. doi: 10.1016/j.atmosenv.2012.07.012.

Sistla G , Zhou N , Hao W , Ku J-K , Rao ST , Bornstein R , Freedman F , Thunis P , 1996: Effects of uncertainties in meteorological inputs on Urban Airshed Model predictions and ozone control strategies. Atmos. Environ, 30, 2011–2025.

Solazzo E and Galmarini S : Error apportionment for atmospheric chemistry-transport models – a new approach to model evaluation, Atmos. Chem. Phys, 16, 6263–6283, doi:10.5194/acp-16-6263-2016, 2016.

Stoeckenius TE , Hogrefe C , Zagunis J , Sturtz TM , Wells B and Sakulyanontvittaya T , 2015 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. Atmospheric Environment, 115, pp. 389–403.

Willmott CJ (1981), On the validation of models, Phys. Geogr, 2, 184–194.

Wong DC , Pleim J , Mathur R , Binkowski F , Otte T , Gilliam R , Pouliot G , Xiu A , Young JO , and Kang D : WRFCMAQ 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, 2012.

Xing J , Mathur R , Pleim J , Hogrefe C , Gan C-M , Wong DC , Wei C , Gilliam R , and Pouliot G : Observations and modeling of air quality trends over 1990–2010 across the Northern Hemisphere: China, the United States and Europe, Atmos. Chem. Phys, 15, 2723–2747, doi:10.5194/acp-15-2723-2015, 2015.

Xing J , Pleim J , Mathur R , Pouliot G , Hogrefe C , Gan C-M , and Wei C : Historical gaseous and primary aerosol emissions in the United States from 1990 to 2010, Atmos. Chem. Phys, 13, 7531–7549, doi:10.5194/acp-13-7531-2013, 2013.

Zurbenko IG , 1986: The Spectral Analysis of Time Series.North Holland, 241 pp.

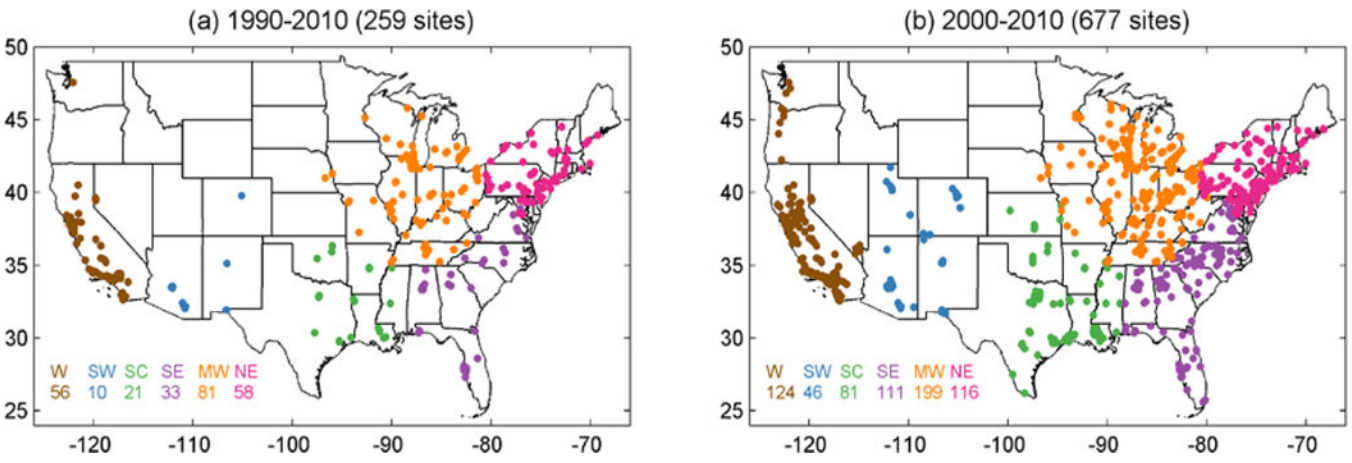

Figure 1.

AQS sites having at least 80% data coverage for each year for **(a)** 1990–2010 and **(b)** 2000–2010, which are grouped into 6 regions: Midwest (MW, orange), Northeast (NE, pink), Southcentral (SC, green), Southeast (SE, purple), Southwest (SW, blue) and West (W, brown).

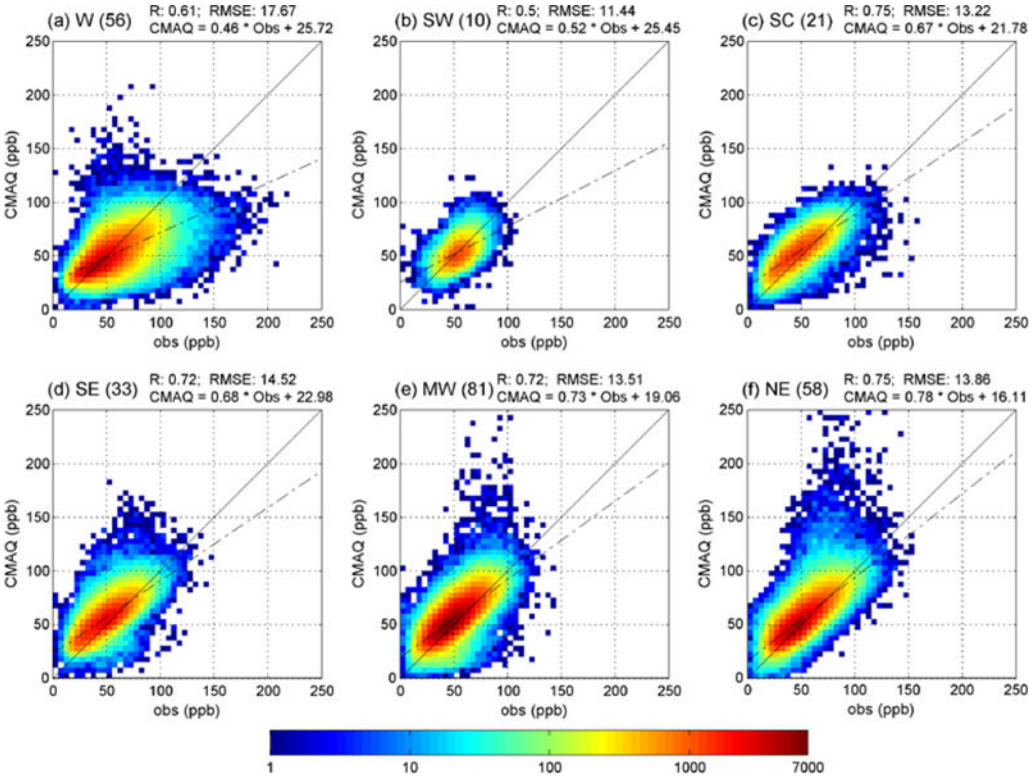

Figure 2.

Density scatter plots (5 ppb bins) of daily maximum 8-hr (DM8HR) ozone concentrations across each region for 1990–2010 (21-yrs): a) West (W), b) Southwest (SW), c) Southcentral (SC), d) Southeast (SE), e) Midwest (MW) and f) Northeast (NE). The values in parenthesis denote the number of sites at each region (sites as seen in Fig. 1a).

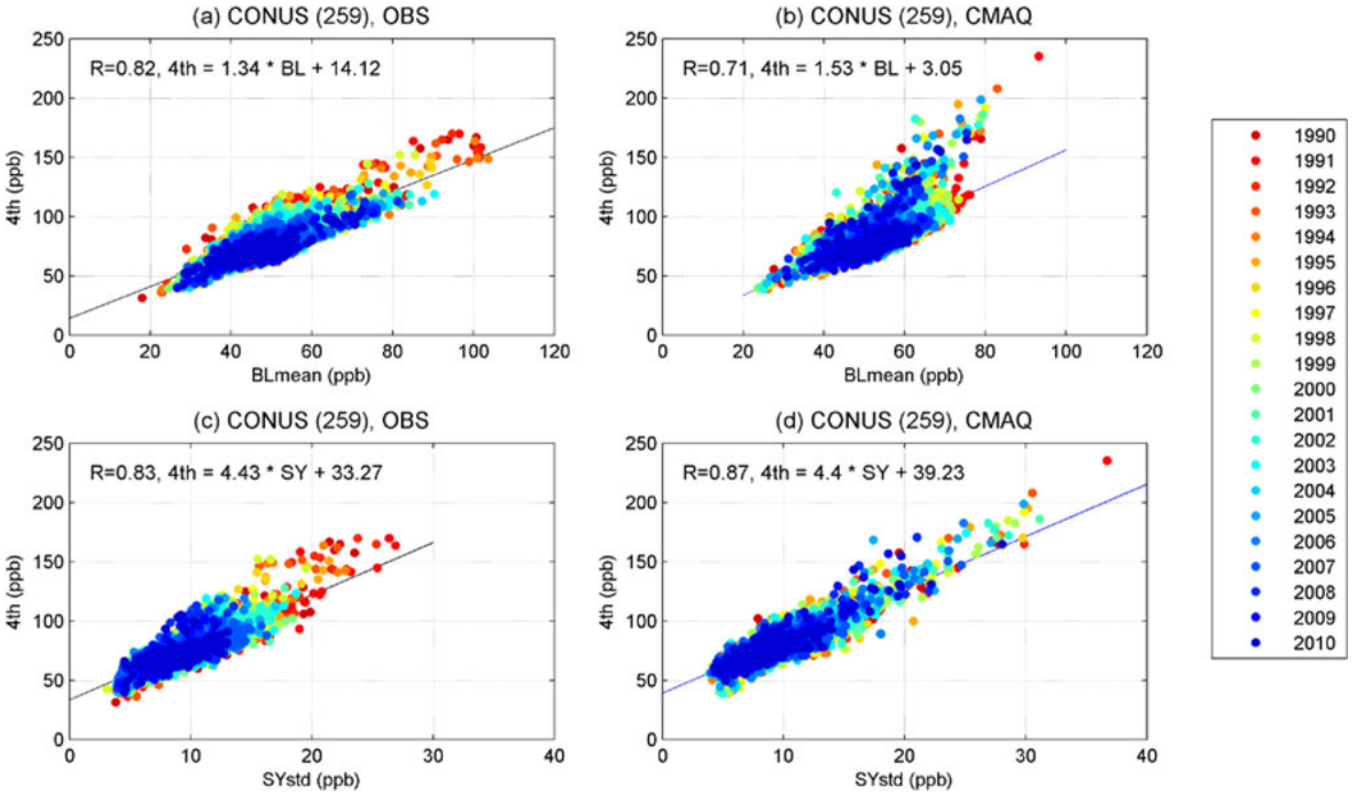

Figure 3.

Correlation between BLmean and SYstdev with 4th highest ozone concentration, color-coded by year. (a, c) Observations across the continental US (CONUS) and (b, d) corresponding CMAQ values.

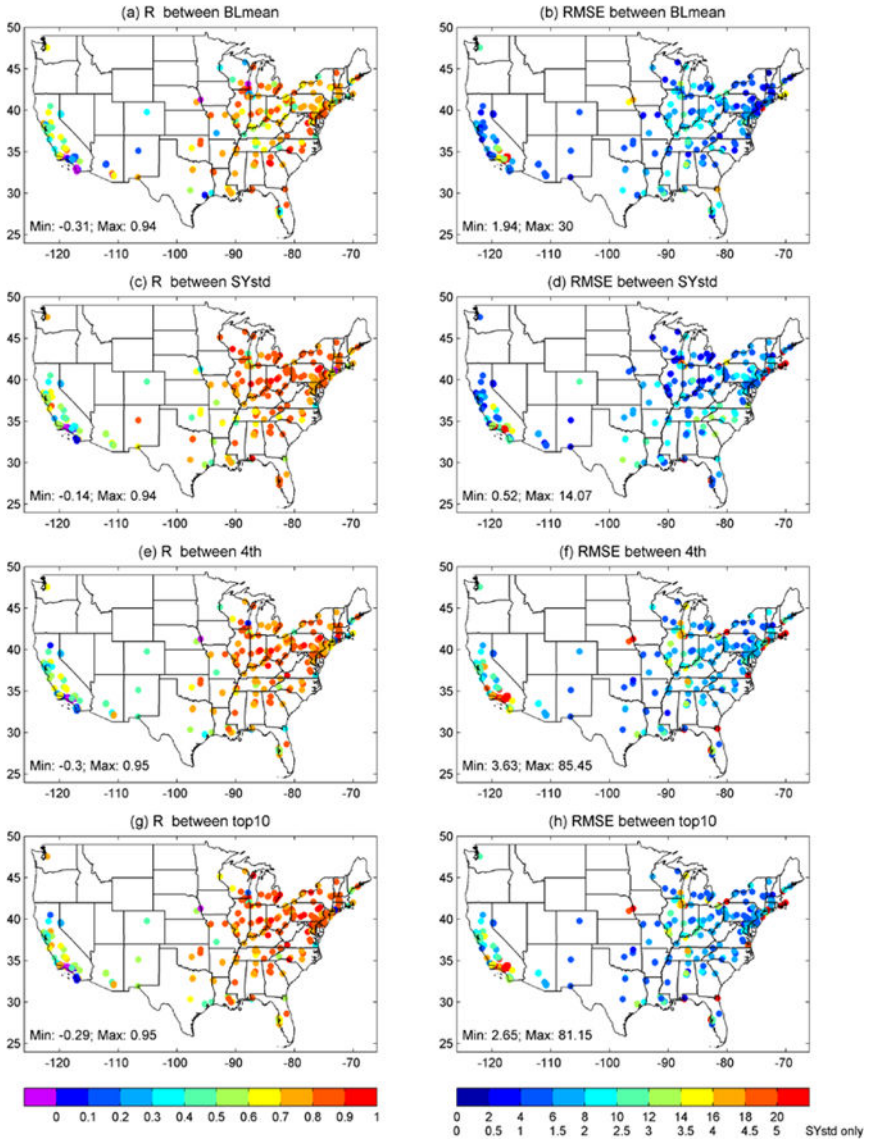

Figure 4.

R and RMSE between observed and modeled (a-b) Baseline mean (BLmean), (c-d) standard deviation of the synoptic component (SYstd), (e-f) unpaired 4th highest ozone concentration (4th), and (g-h) unpaired average of top 10 ozone concentrations (top10) of the DM8HR ozone, for 1990–2010. All RMSE units are in ppb.

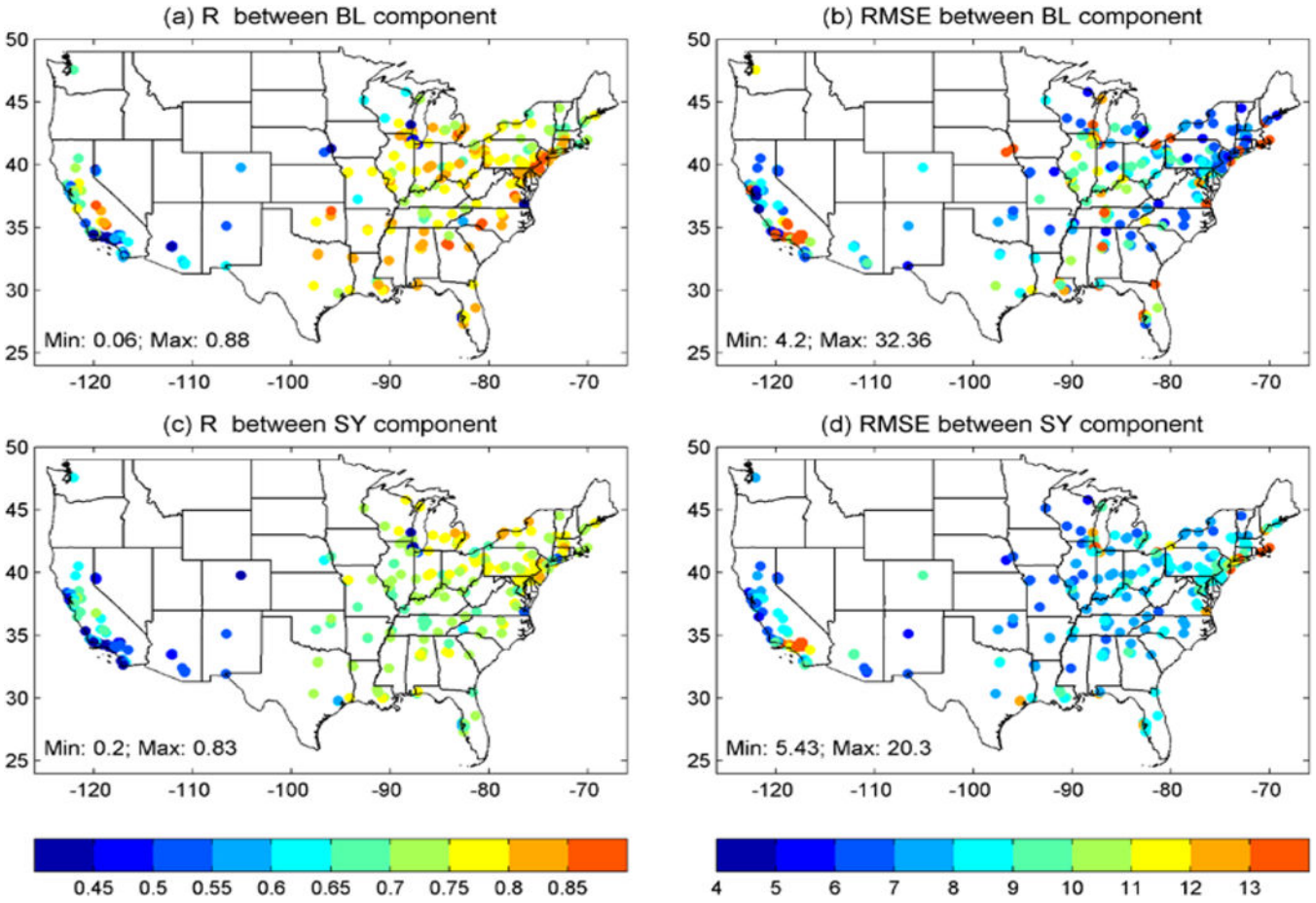

Figure 5.

R and RMSE (ppb) between observed and modeled (a-b) Baseline component (BL) and (c-d) Short-term component (SY) of decomposed DM8HR ozone, for 1990–2010.

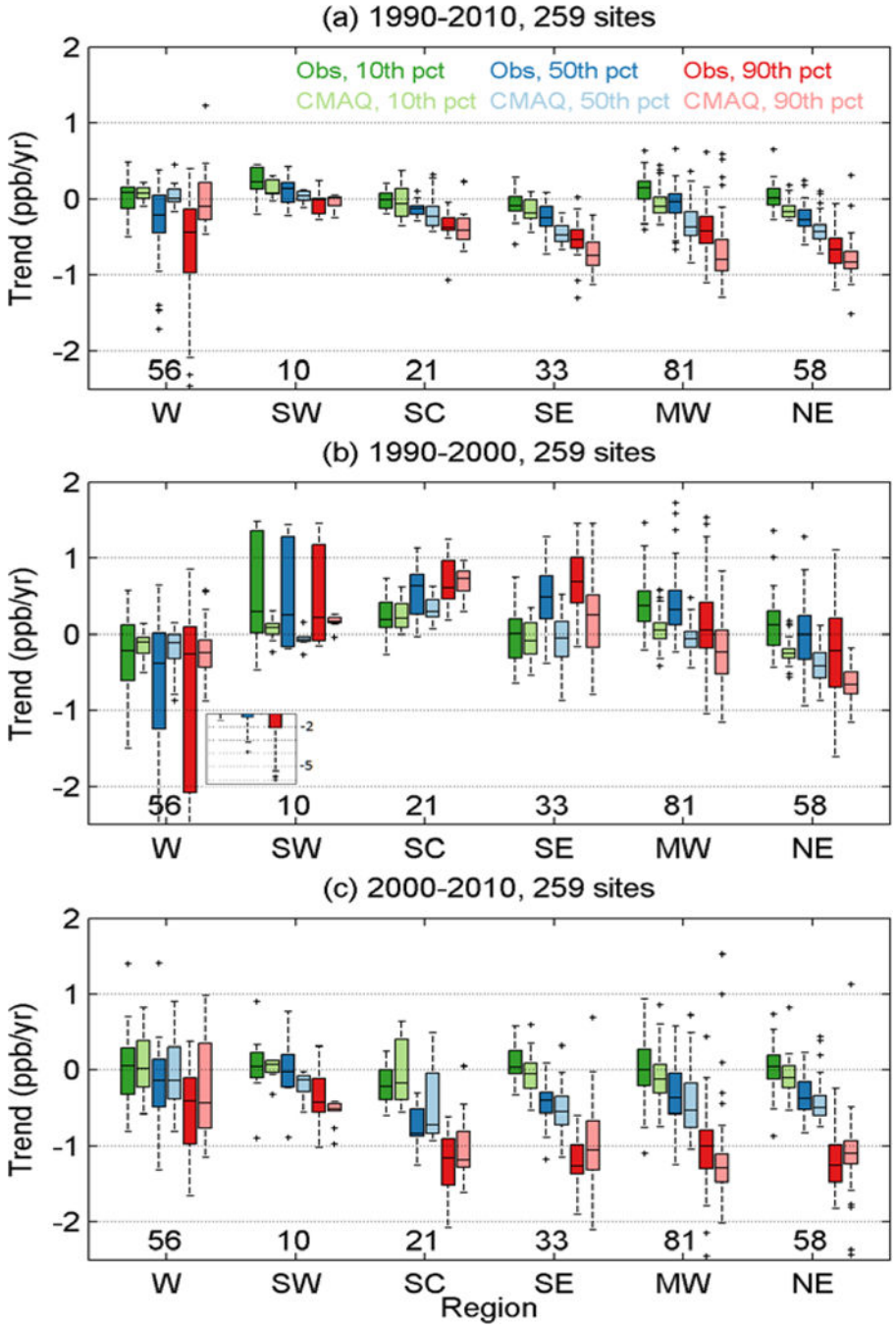

Figure 6.

Observed and simulated trends in the 10th, 50th and 90th percentile of DM8HR ozone concentration per region using 259 stations. a) 21-years (1990–2010), b) earlier 11-years (1990–2000), c) recent 11-years (2000–2010). The number of stations in each region is reported in the x-axis legend.

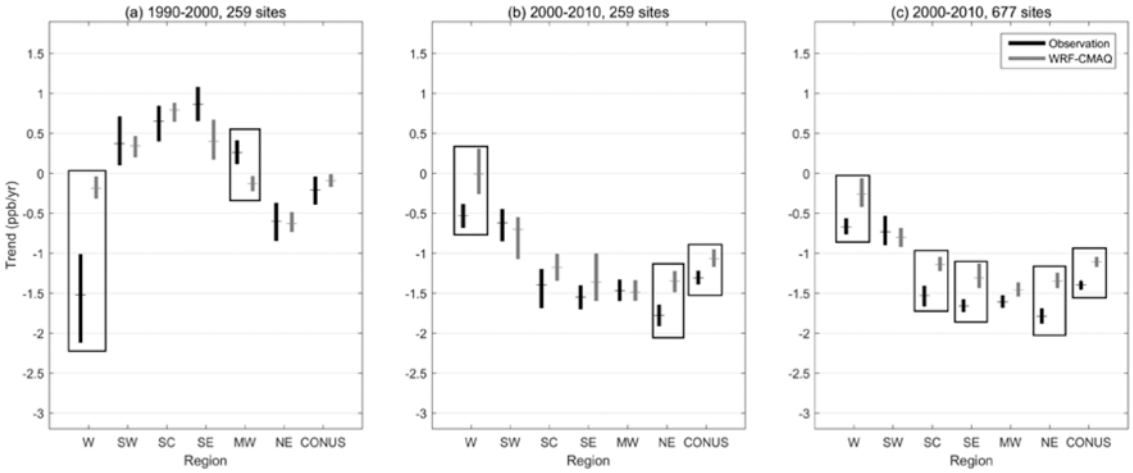

Figure 7.

95% bootstrap confidence intervals (ppb/yr) for observed (black) and modeled (gray) regional trends shown in Table 2. Non-overlapping bootstrapped intervals denoted with black squares show trends that are significantly different (also indicated by italics in Table 2).

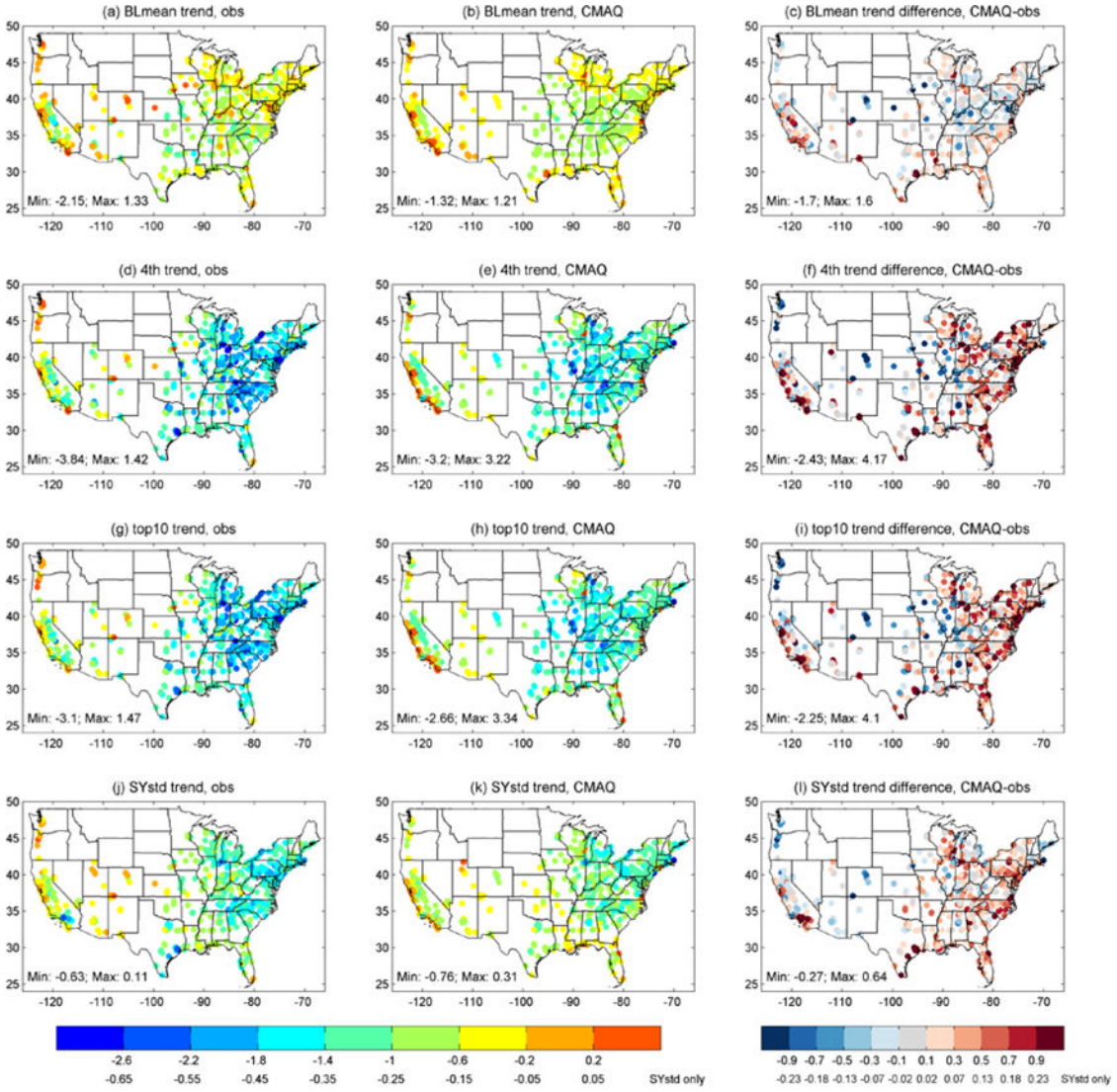

Figure 8.

Trends (ppb/yr) and differences in the trends (ppb/yr) between modeled and observed BLmean, 4th highest, top10 and SYstdev for 2000–2010 (#677 stations shown in Fig. 1b).

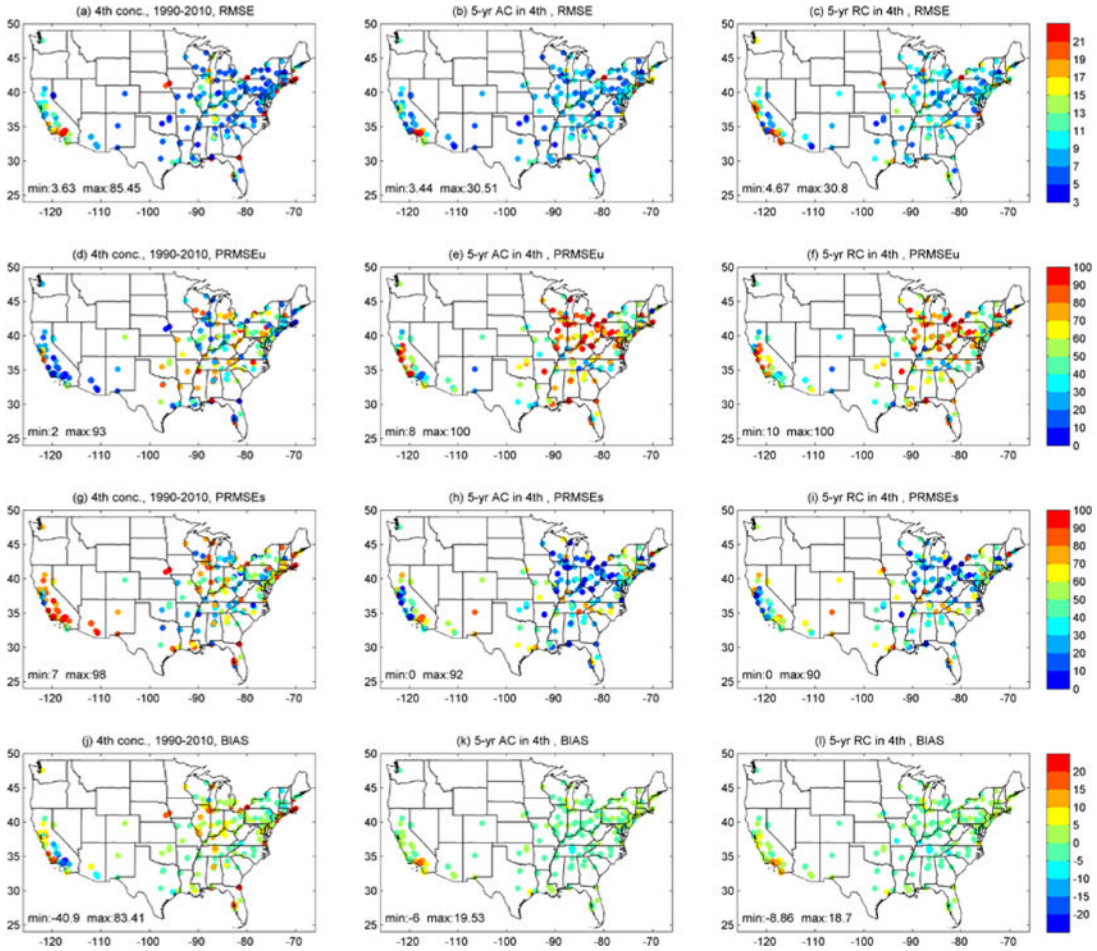

Figure 9.

Error of the 4th highest (left, ppb), absolute change in 4th highest for all 5-year intervals (AC; middle, 16 pairs, ppb) and relative change in 4th highest for all 5-year intervals (RC; right, 16 pairs, %): (a-c) RMSE; (d-f) Percentage of the unsystematic RMSE (PRMSEu); (g-i) Percentage of the systematic RMSE (PRMSEs); and (j-l) BIAS.

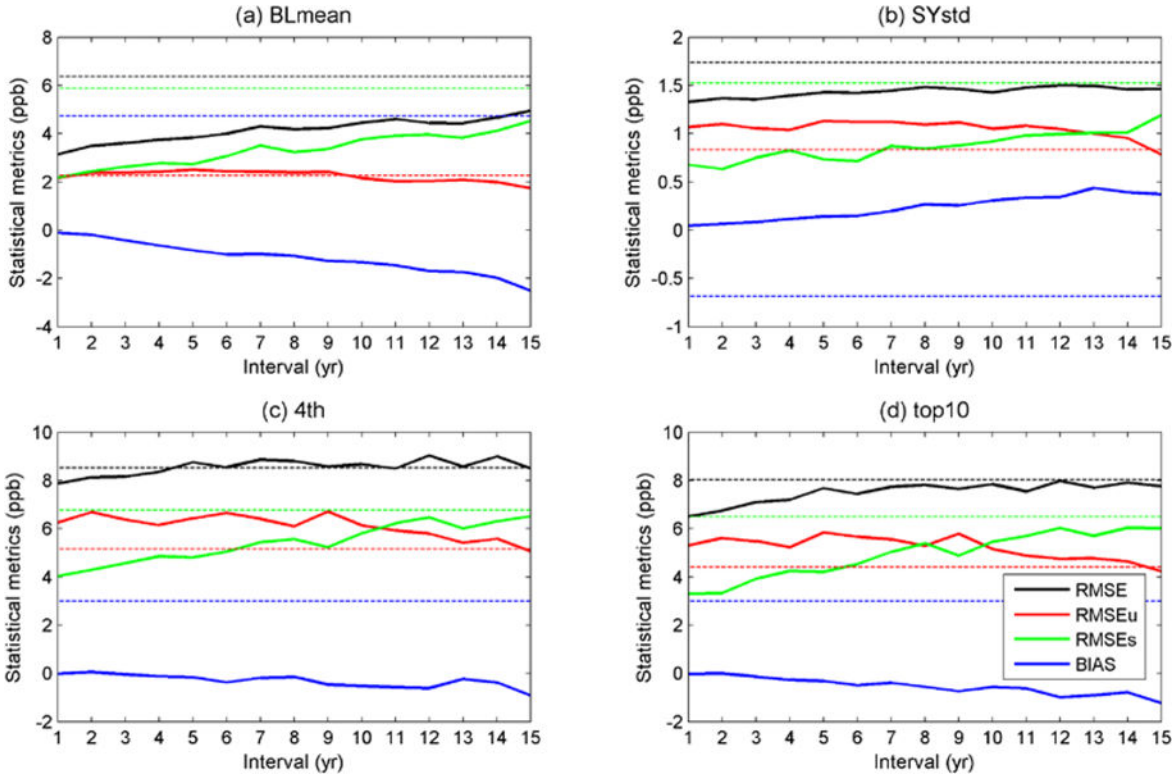

Figure 10.

Median error of changes in (a) BLmean, (b) SYstd, (c) 4th and (d) top10 ozone concentrations across CONUS for various change intervals (i=1, 2, 3, … 15). For example, number 5 on the x-axis shows the median change calculated from all possible 5 year change interval combinations (16 pairs for 1990–2010). For comparison, dashed lines show the corresponding metrics of absolute values (instead of the changes) computed across the entire 1990 – 2010 time period.

Table 1.

Maximum 4th highest and top10 observed ozone concentrations (ppb) calculated by superimposing 21-yrs (1990–2010) of short-term components (SY) to the baseline (BL) that exhibits the highest and lowest BLmean at each station (O3=BL+SY). The station that exhibits the highest BLmean per region is included in the table to demonstrate the role of BL in controlling ozone exceedances.

	Highest BLmean				Lowest BLmean			
Region	**Year**	**BLmean**	**4th max**	**top10 max**	**Year**	**BLmean**	**4th max**	**top10 max**
W	1994	104	163	162	2009	71	126	122
SW	1998	65	91	89	1991	50	71	70
SC	1998	61	95	93	2009	44	88	82
SE	1990	71	121	118	2009	44	84	82
MW	1998	74	114	113	2009	45	81	77
NE	1991	71	124	122	2009	44	89	88

Table 2.

Regional trends of the 4th highest (ppb/yr) from WRF-CMAQ and observations for 1990–2000 and 2000–2010. Values in **bold** denote statistically significant trends with p<0.05. Colored cells show percentage of stations within each region with statistically significant trends: white=0–20%, light grey=20–40%, grey=40–60%, dark grey=60–80%. Values in italics indicate trends that are significantly different (non-overlapping confidence intervals shown in Fig. 7).

	#259 sites				#677 sites	
Regions	**Simulated Trend (ppb/yr) 1990–2000**	**Observed Trend (ppb/yr) 1990–2000**	**Simulated Trend (ppb/yr) 2000–2010**	**Observed Trend (ppb/yr) 2000–2010**	**Simulated Trend (ppb/yr) 2000–2010**	**Observed Trend (ppb/yr) 2000–2010**
W	*−0.19*	***−1.52***	*−0.01*	***−0.53***	*−0.26*	***−0.67***
SW	0.34	0.37	**−0.70**	**−0.62**	**−0.80**	**−0.73**
SC	0.79	0.65	**−1.18**	**−1.40**	***−1.14***	***−1.53***
SE	*0.40*	*0.86*	**−1.36**	**−1.55**	***−1.31***	***−1.66***
MW	*−0.13*	*0.26*	**−1.49**	**−1.47**	**−1.46**	**−1.61**
NE	−0.63	−0.60	***−1.35***	***−1.78***	***−1.35***	***−1.79***
CONUS	−0.09	−0.21	***−1.07***	***−1.31***	***−1.11***	***−1.40***

Table 3.

Regional change in average of top 10 paired-in-time concentrations from 2000 to 2010 (ppb). The values in parenthesis denote changes calculated using the unpaired top10 concentrations. The ratio is marked with bold font for the best correspondence between model and observed changes. Sites correspond to Fig.1a.

Region	# of Sites	CMAQ 2000	CMAQ 2010	Observed 2000	Observed 2010	CMAQ Change	Observed Change	Ratio CMAQ/Obs. Change
W	56	67.8 (82.3)	70.8 (79.5)	75.8	71.1	2.9(−2.8)	−4.7	−0.6 (**0.6**)
SW	10	67.3 (77.6)	59.9 (68.7)	75.5	70.0	−7.4 (−8.9)	−5.5	**1.4** (1.6)
SC	21	82.1 (89.5)	67.5 (73.7)	88.8	71.3	−14.6 (−15.8)	−17.5	0.8 (**0.9**)
SE	33	78.6 (89.7)	68.6 (77.2)	84.3	68.8	−10.0 (−12.4)	−15.5	0.6 (**0.8**)
MW	81	73.4 (82.3)	68.5 (76.8)	76.3	68.6	−4.9 (−5.4)	−7.6	0.6 (**0.7**)
NE	58	76.0 (84.3)	75.6 (81.0)	81.6	75.0	−0.4 (−3.3)	−6.5	0.1 (**0.5**)