Interpreting the Information Embedded in Observed and Modeled Air Quality Time Series and Using Regional Air Quality Models for Regulatory Policies

Huiying Luo

University of Connecticut - Storrs, huiying.luo@uconn.edu

Follow this and additional works at: https://opencommons.uconn.edu/dissertations

Recommended Citation
https://opencommons.uconn.edu/dissertations/2371
Interpreting the Information Embedded in Observed and Modeled Air Quality Time Series and Using Regional Air Quality Models for Regulatory Policies

Huiying Luo, PhD
University of Connecticut, 2019

Abstract

Regional-scale air quality models are being used to perform research, forecasting, regulatory assessments and planning. Thus, it is important to critically assess the models’ capability in reproducing the spatio-temporal features seen in observations. The main scope of this work is to develop new methods for assessment and evaluation of EPA’s Community Multiscale Air Quality Modeling System (CMAQ) in the context of building confidence when regional-scale air quality models like CMAQ are used for policy support. The research is divided into three parts. In the first part, features embedded in two decades of ozone and PM$_{2.5}$ observations and coupled WRF-CMAQ model outputs are analyzed and dynamic evaluation of model simulations is conducted. In the second part, a probabilistic framework is developed based on 34 years of ozone observations to examine the efficacy of emission reduction policies on ozone concentrations using regional-scale air quality models. In the third part, the contributions of anthropogenic emissions to ozone concentrations are assessed towards defining the manageable portion of ozone concentration with the aid of specialized CMAQ model simulations.
Interpreting the Information Embedded in Observed and Modeled Air Quality Time Series and Using Regional Air Quality Models for Regulatory Policies

Huiying Luo

B.S., Central South University, 2012

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2019
Approval Page

Doctor of Philosophy Dissertation

Interpreting the Information Embedded in Observed and Modeled Air Quality Time Series and Using Regional Air Quality Models for Regulatory Policies

Presented by
Huiying Luo, B.S.

Major Advisor ______________________________________________________________________
Dr. Marina Astitha

Associate Advisor ___________________________________________________________________
Dr. George Kallos

Associate Advisor ___________________________________________________________________
Dr. S. Trivikrama Rao

Associate Advisor ___________________________________________________________________
Dr. Kristina Wagstrom

Associate Advisor ___________________________________________________________________
Dr. Guiling Wang

University of Connecticut
2019
Acknowledgments

I would like to express my special appreciation and thanks to my advisor Professor Marina Astitha for being such a tremendous mentor and a role model for me in numerous aspects. This Ph.D. journey would not have been so valuable without all her professional guidance, support, encouragement over the years. Many thanks to Professor S. Trivikrama Rao (North Carolina State University) who also guided me closely and inspired me with his genuine passion in scientific research. I sincerely thank my committee members Professor George Kallos (University of Athens), Professor Kristina Wagstrom and Professor Guiling Wang for all their brilliant comments and advices. I gratefully acknowledge our collaborators Dr. Christian Hogrefe and Dr. Rohit Mathur from the United States Environmental Protection Agency for their immense contributions to my research.

I appreciate all the professors, colleagues at the University of Connecticut. I’m thankful for all the splendid seminars, discussions and lectures; the warmest connections we share as members and former members of Atmospheric Modeling Group with Jaemo, Alex, Ghezae, Christina and Mike; the enjoyable friendship and memories with Yan, Junya, Chen, Di, Yi, Bowen and many more.

My special thanks to my family, especially my dear grandmother for always being such caring, and my friends back in China. Last but not least, I would like to thank my husband Hao Yuan for his warmest love and perpetual support.
Contents

Chapter 1  Introduction ................................................................. 1

Chapter 2  Dynamic evaluation of two decades of WRF-CMAQ ozone simulations over the Contiguous United States .......... 4
  2.1 Introduction ................................................................. 4
  2.2 Description of the coupled WRF-CMAQ simulations and observations ......................... 6
  2.3 Methods of analysis ................................................................. 8
    2.3.1 Data analysis using the Kolmogorov-Zurbenko (KZ) filter ............................... 8
    2.3.2 Statistical metrics ................................................................. 10
  2.4 Results and discussion ................................................................. 11
    2.4.1 Operational evaluation ................................................................. 11
    2.4.2 Dynamic evaluation ................................................................. 17
  2.5 Summary ................................................................. 28

Appendix ................................................................. 29

Additional Figures ................................................................. 31

Chapter 3  Operational evaluation of PM$_{2.5}$ simulations over the contiguous United States .... 39
  3.1 Introduction ................................................................. 39
  3.2 WRF-CMAQ simulations and Air Quality Observations ................................................. 40
  3.3 Results and discussion ................................................................. 41
  3.4 Conclusions ................................................................. 47

Chapter 4  Evaluating Trends and Seasonality in Modeled PM$_{2.5}$ Concentrations Using Empirical Mode Decomposition ......................................................... 49
  4.1 Introduction ................................................................. 49
  4.2 Coupled WRF-CMAQ PM$_{2.5}$ Simulations and Observations ................................................. 51
  4.3 Method ................................................................. 52
Chapter 6 On the manageable portion of the ozone non-attainment problem in the contiguous United States

6.1 Introduction

6.2 Data and methods

6.2.1 Model simulations and observations

6.2.2 Ozone decomposition and reconstruction

6.3 Results and Discussion

6.3.1 Modeled influence of anthropogenic emissions on ozone

6.3.2 Modeled manageable portion of ozone exceedances

6.4 Conclusions

Additional Figures

Concluding Remarks

References

Research Products
Chapter 1

Introduction

Air quality models are being used to perform research, forecasting, regulatory assessments, and planning. Thus, 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 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). 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 and is one of the key components for building confidence in a model’s use for policy analysis. Most past dynamic model evaluation focused 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 (Banzhaf et al., 2015; Seo et al., 2014; Xing et al., 2015). The 21-year model simulation by Gan et al. (2015) over the contiguous U.S. (CONUS), together with ozone and PM observations, provide a unique opportunity to evaluate the model’s ability to reproduce the observed changes in ozone and PM concentrations.

Year-to-year changes in air quality are attributable to variations in meteorological conditions, local-to-hemispheric scale emissions and transport and factors such as climate change and stratospheric-tropospheric exchanges. In current regulatory applications, a regional air quality model is applied for a base year and a future year with reduced emissions using the same meteorological conditions as those in the base
year (U.S. EPA, 2014). Because the meteorology used for the base year would never prevail for future emission control years, the impacts of emission reductions on extreme values of interest cannot be predicted perfectly even if the model and its input data were perfect. Hence, with the current methodology, we would not know how meteorological variability affects the efficacy of emission reduction policies in assuring compliance with the standards in the future year and there is no guarantee that the envisioned emission control strategy will in fact lead to ozone compliance in future years.

To help improve upon the current method for using regional air quality models in the regulatory context, 34 years of ozone observations (1981-2014) for the contiguous United States are analyzed with the objective of developing a method for estimating the confidence bounds for the projected ozone design value (DV), thereby advancing the scientific notion of future projections towards a probabilistic framework that explicitly accounts for prediction uncertainties due to the year-to-year variability in meteorology. The proposed new method can be utilized for building confidence when regional-scale air quality models are used for policy support.

Furthermore, the significant influence of anthropogenic emissions on modeled ozone concentrations and have been assessed by multiple studies in the past to determine a portion of the uncertainties that accompany regional air quality model simulations. Our interest in this study is to quantify the contribution from anthropogenic emissions and estimate the manageable portion of the air pollution burden in the continental United States and help make emission reduction policies more realistic and effective.

To facilitate all the above, we develop methodologies that utilize spectral and/or mode decomposition of the observed and modeled long-term air quality time series (Rao and Zurbenko, 1994; Rao et al., 1997; Eskridge et al., 1997; Hogrefe et al., 2000; Hogrefe et al., 2003; Colominas et al., 2014; Porter et al., 2015; Solazzo and Galmarini, 2016). The decomposition allows us to identify the underlying forcing mechanisms that control exceedances, and thus, provide informed recommendations for the optimal use of regional-scale air quality models.
This dissertation aims towards a more complete understanding of the information embedded in observations and simulations of ozone and fine particulate matter over continental US, advancing the implementation of regional air quality models in regulatory assessment with more confidence. The main objectives of this work are provided below:

Chapter 2-4. Analyze, interpret and compare the features embedded in decades of air quality observations and model outputs for ozone and PM$_{2.5}$.

Chapter 5. Develop new tools for examining the efficacy of emission reduction policies on ozone concentrations using regional-scale air quality models in a probabilistic framework.

Chapter 6. Quantify the influence of anthropogenic and wildfire emissions on ozone concentrations and estimate the modeled manageable portion of the ozone burden in continental United States.
Chapter 2
Dynamic evaluation of two decades of WRF-CMAQ ozone simulations over the Contiguous United States

2.1 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; Stoeckenuis 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, 2011; Hogrefe et al. 2000, 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, 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.

2.2 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. 2-1) 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. 2-1a 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. 2-1b, 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 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).
2.3 Methods of analysis

2.3.1 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.

\[
BL(t) = KZ(5,5) \quad (2 - 1)
\]
\[
SY(t) = O_3(t) - KZ(5,5) \quad (2 - 2)
\]
\[
O_3(t) = SY(t) + BL(t) \quad (2 - 3)
\]

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 \( SY \) forcing is 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 (SYstddev) over the summertime period are included in the analysis. The BLmean indicates the magnitude of the long-term forcing in each year, whereas the SYstddev 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.

2.3.2 Statistical metrics

The traditional statistical metrics used in this study are described in the Appendix. In addition to the Pearson correlation coefficient (R, Eq. 2-A1), mean bias (BIAS, Eq. 2-A2) and RMSE (Eq. 2-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. 2-4 and 2-5. 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.

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

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

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):

\[
RMSE^2 = RMSE_s^2 + RMSE_u^2 \quad (2-6)
\]

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. 2-6. 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. 2-7, 2-8) between specific pairs of years.

\[
AC = FutureYear - BaseYear
\]

\[
RC = 100 \times \frac{(FutureYear - BaseYear)}{BaseYear}
\]

2.4 Results and discussion

2.4.1 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, 4\textsuperscript{th} 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. 2-2). 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.

Figure 2-2. 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. 2-1a).

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 (see Additional Figures). 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. 2-1b), the correlation also increases and RMSE is reduced (Fig. 2-S1) 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. 2-S1, 2-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 2-1, 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 2-1, 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 2-1, 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 2-1, 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. 2-3a) 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. 2-3b). 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 (SYstddev) and the 4th highest (Fig. 2-3c, 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. 2-4a, 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. 2-4 c, d; SYstddev) as well as the 4th highest and top10 ozone concentrations (Fig. 2-4e-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. 2-5), 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. 2-4 are based on one value per station per year (i.e. 21 data points at each station) while the results in Fig. 2-5 are based on daily time-series (~153 x 21 values at each station). This means that the results in Fig. 2-4 (BLmean, SYstddev, 4th and top10) are affected by inter-seasonal variability while in Fig. 2-5 (BL and SY) both intra- and inter-seasonal variability is included.
Table 2-1. 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.

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

Figure 2-3. 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 2-4. 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.
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.

2.4.2 Dynamic evaluation

*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. 2-6). 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. 2-6a). 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. 2-6b), 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. 2-6c), 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.
Figure 2-6. Observed and simulated trends in the 10\textsuperscript{th}, 50\textsuperscript{th} and 90\textsuperscript{th} 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.

Regional averages of simulated and observed station-specific trends (in ppb/yr) calculated at each monitor are presented in Table 2-2. 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 4\textsuperscript{th} highest ozone concentration at some locations. In
addition, the 11-yr trend in the 4\textsuperscript{th} highest during 2000-2010 for the 677 sites is included in Table 2-2. Statistically significant trends with p<0.05 (bold font in Table 2-2) 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 2-2 values in italics and Fig. 2-7).

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. 2-7b, 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. 2-7b). 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. 2-7c). 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. 2-S3b), but nevertheless the model showed larger spatial variability compared to observations for all years. For SC U.S., the simulated 4\textsuperscript{th} highest is closer to the observed one with observations showing outliers mostly after the year 2000. The model has larger 4\textsuperscript{th} 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. 2-4 e-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. 2-2), the observed and simulated BLmean trends are calculated for the 21-yr period (Fig. 2-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 as Additional Figures (Fig. 2-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. 2-S4 e-f, lower panels) likely contributed to the higher simulated 4th values (Fig. 2-S4 e-f, upper panels) given the linear relationship shown in Fig. 2-3.

**Table 2-2.** 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. 2-1~2-7).

<table>
<thead>
<tr>
<th>Regions</th>
<th>Simulated Trend (ppb/yr)</th>
<th>Observed Trend (ppb/yr)</th>
<th>Simulated Trend (ppb/yr)</th>
<th>Observed Trend (ppb/yr)</th>
<th>Simulated Trend (ppb/yr)</th>
<th>Observed Trend (ppb/yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>-0.19</td>
<td>-1.52</td>
<td>-0.01</td>
<td>-0.53</td>
<td>-0.26</td>
<td>-0.67</td>
</tr>
<tr>
<td>SW</td>
<td>0.34</td>
<td>0.37</td>
<td>-0.70</td>
<td>-0.62</td>
<td>-0.80</td>
<td>-0.73</td>
</tr>
<tr>
<td>SC</td>
<td>0.79</td>
<td>0.65</td>
<td>-1.18</td>
<td>-1.40</td>
<td>-1.14</td>
<td>-1.53</td>
</tr>
<tr>
<td>SE</td>
<td>0.40</td>
<td>0.86</td>
<td>-1.36</td>
<td>-1.55</td>
<td>-1.31</td>
<td>-1.66</td>
</tr>
<tr>
<td>MW</td>
<td>-0.13</td>
<td>0.26</td>
<td>-1.49</td>
<td>-1.47</td>
<td>-1.46</td>
<td>-1.61</td>
</tr>
<tr>
<td>NE</td>
<td>-0.63</td>
<td>-0.60</td>
<td>-1.35</td>
<td>-1.78</td>
<td>-1.35</td>
<td>-1.79</td>
</tr>
<tr>
<td>CONUS</td>
<td>-0.09</td>
<td>-0.21</td>
<td>-1.07</td>
<td>-1.31</td>
<td>-1.11</td>
<td>-1.40</td>
</tr>
</tbody>
</table>
Figure 2-7. 95% bootstrap confidence intervals (ppb/yr) for observed (black) and modeled (gray) regional trends shown in Table 2-2. Non-overlapping bootstrapped intervals denoted with black squares show trends that are significantly different (also indicated by italics in Table 2-2).

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. 2-8. The simulated BLmean trend is generally smaller than the observed trend at most stations (Fig. 2-8a, 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. 2-8j, k). Exceptions are denoted for the BLmean and SYstdev trends in coastal areas. Trends in Top10 and 4th highest ozone are almost identical (Fig. 2-8d-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 2-2. 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. 2-9 e, 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 t (Fig. 2-S5).

**Figure 2-8.** 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. 2-1b).
The first part of dynamic model evaluation included analysis of temporal trends in the 4th highest and top10 values. 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 2-3). 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. 2-7) and relative 5-yr changes (RC; Eq. 2-8) in the 4th highest ozone concentrations as well as the 4th highest concentrations across all 21 years (Fig. 2-9). The results clearly demonstrate the reduction of error and bias in changes compared to absolute concentrations (Fig. 2-9 a-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).
Table 2-3. 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. 2-1a.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>56</td>
<td>67.8 (82.3)</td>
<td>70.8 (79.5)</td>
<td>75.8</td>
<td>71.1</td>
<td>2.9 (-2.8)</td>
<td>-4.7</td>
<td>-0.6 (0.6)</td>
</tr>
<tr>
<td>SW</td>
<td>10</td>
<td>67.3 (77.6)</td>
<td>59.9 (68.7)</td>
<td>75.5</td>
<td>70.0</td>
<td>-7.4 (-8.9)</td>
<td>-5.5</td>
<td>1.4 (1.6)</td>
</tr>
<tr>
<td>SC</td>
<td>21</td>
<td>82.1 (89.5)</td>
<td>67.5 (73.7)</td>
<td>88.8</td>
<td>71.3</td>
<td>-14.6 (-15.8)</td>
<td>-17.5</td>
<td>0.8 (0.9)</td>
</tr>
<tr>
<td>SE</td>
<td>33</td>
<td>78.6 (89.7)</td>
<td>68.6 (77.2)</td>
<td>84.3</td>
<td>68.8</td>
<td>-10.0 (-12.4)</td>
<td>-15.5</td>
<td>0.6 (0.8)</td>
</tr>
<tr>
<td>MW</td>
<td>81</td>
<td>73.4 (82.3)</td>
<td>68.5 (76.8)</td>
<td>76.3</td>
<td>68.6</td>
<td>-4.9 (-5.4)</td>
<td>-7.6</td>
<td>0.6 (0.7)</td>
</tr>
<tr>
<td>NE</td>
<td>58</td>
<td>76.0 (84.3)</td>
<td>75.6 (81.0)</td>
<td>81.6</td>
<td>75.0</td>
<td>-0.4 (-3.3)</td>
<td>-6.5</td>
<td>0.1 (0.5)</td>
</tr>
</tbody>
</table>

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. 2-7). More specifically, systematic error was larger than unsystematic for the 4th highest ozone concentration (Fig. 2-10d, 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. 2-9 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 (SYstddev), 4th highest, and unpaired top10 ozone
concentrations for the 21-yr timeframe (Fig. 2-10 and more detailed boxplots in 2-S6 – 2-S8). The statistical metrics discussed are bias, RMSE, RMSEu and RMSEs for every combination (Fig. 2-10). We focus on the BLmean, 4th highest and unpaired top10.

**Figure 2-9.** 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.

Starting with BLmean (Fig. 2-10a), 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. 2-10c, 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.

**Figure 2-10.** 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.
2.5 Summary

Two decades of WRF-CMAQ simulations provided a unique opportunity to assess changes in DM8HR ozone caused by emission reduction policies and to 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 the 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.

Appendix

Statistical metrics employed in the evaluation of simulated ozone concentrations:

\[
R = \frac{\sum_{i=1}^{N} (P_i \cdot O_i) - \frac{\sum_{i=1}^{N} P_i \cdot \sum_{i=1}^{N} O_i}{N}}{\sqrt{\frac{\sum_{i=1}^{N} P_i^2}{N} - \left(\frac{\sum_{i=1}^{N} P_i}{N}\right)^2} \cdot \sqrt{\frac{\sum_{i=1}^{N} O_i^2}{N} - \left(\frac{\sum_{i=1}^{N} O_i}{N}\right)^2}}
\]  

\[
BIAS = \frac{\sum_{i=1}^{N} (P_i - O_i)}{N}
\]  

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{N} (P_i - O_i)^2}{N}}
\]
The linear regression provides an estimated (or regressed) prediction ($\hat{P}$) with $a$ and $b$ the least square regression coefficients (Eq. 2-A4).

$$\hat{P}_i = a + b \cdot O_i$$

(2 - A4)
Figure 2-S1. 11-yr density scatter plots (5 ppb bins) of daily maximum 8-hr (DM8HR) ozone concentrations across each region (2000-2010): 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 (#677 sites shown in Fig. 2-1b).
**Figure 2-S2.** 11-yr density scatter plot (5 ppb bins) of daily maximum 8-hr (DM8HR) ozone concentrations across each region (2000-2010): 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 (#259 sites shown in Fig. 2-1a).
Figure 2-S3. Observed (black) and modeled (gray) 4\textsuperscript{th} highest (ppb; a-f upper panels) and BLmean (ppb; a-f, lower panels) for each region (Fig. 2-1a) during 1990-2010 (#259 stations). The boxes indicate the 25\textsuperscript{th}, 50\textsuperscript{th} and 75\textsuperscript{th} percentiles, while the whiskers can extend to 1.5 times the interquartile range. Data beyond those are identified as outliers and denoted by crosses. Number of stations is shown in parenthesis next to the region.
Figure 2-S4. Observed (black) and modeled (blue) 4th highest (a-f upper panels) and BLmean (a-f, lower panels) for each region (Fig. 2-1b, #677 sites) during 2000-2010. The boxes indicate the 25th, 50th and 75th percentiles, while the whiskers extend to 1.5 times the interquartile range. Data beyond those are identified as outliers and denoted by crosses.
Figure 2-S5. Trends (ppb/yr) and differences between modeled and observed trends (ppb/yr) in the 4\textsuperscript{th} highest, top10, BLmean and SYstddev for the 21-yr period (1990-2010; #259 sites).
Figure 2-S6. Variation of RMSE for changes in the (a) BLmean, (b) SYstd, (c) 4th and (d) top10 ozone concentrations across CONUS and different year intervals (i=1, 2, 3, … 15). The 16th entry in the x-axis is the absolute concentration (not the change) boxplot.
Figure 2-S7. Same as in Fig. 2-S4 but for the unsystematic component of the RMSE (RMSEu).
Figure 2-S8. Same as in Fig. 2-S4 but for the systematic component of the RMSE (RMSEs).
Chapter 3
Operational evaluation of PM$_{2.5}$ simulations over the contiguous United States

3.1 Introduction

It is well recognized that inhalable fine particulate matter (PM$_{2.5}$) adversely impacts human health and the environment. Regional air quality models have been widely employed in health impact studies and policy-making related to particulate matter. Model simulations must be evaluated to ensure acceptable replication of observations so model users have confidence in the use of regional air quality modeling systems.

Operational evaluation provides a broad overview of the model performance by comparing modeled concentrations to in-situ observations through statistical metrics. Swall and Foley (2009) and Dennis et al. (2010) state that this type of evaluation includes an important inconsistency where volume-averaged model values are compared against point observations that characterize an individual event (referred to as incommensurability). Thus, operational evaluation is only one part of the model evaluation puzzle that complements the dynamic evaluation.

Some of the trend evaluations in the past have focused on specific pairs of years (Kang et al., 2013; Zhou et al., 2013; Foley et al., 2015b). The evaluation of trend based on one pair of specific years comes with the lack of accounting for the inter-annual variations and possible multiannual cycles between those periods. In this study, we adopt the commonly used linear trend analysis of annual indexes such as the annual mean and multiple annual percentiles based on long term simulations under the assumption of data linearity and stationarity and a record of multiannual mean (Hogrefe et al., 2011; Banzhaf et al., 2015; Yahya et al., 2016; Astitha et al., 2017; Karamchandani et al., 2017).
3.2 WRF-CMAQ simulations and Air Quality Observations

The two-way coupled WRF-CMAQ is configured with a 36 km horizontal grid spacing over CONUS with 35 vertical layers of varying thickness extending from the surface to 50 mb (Wong et al., 2012; Gan et al., 2015). Time varying chemical lateral boundary conditions are nested from the 108 km hemispheric WRF-CMAQ simulation from 1990 to 2010 (Xing et al., 2015). The simulations are driven by a comprehensive emission dataset which includes the aerosol precursors and primary particulate matter (Xing et al., 2013, 2015). The readers can refer to Gan at al. (2015) for additional model information and the trend evaluation against seven pairs of sites from CASTNET (Clean Air Status and Trend Network) and IMPROVE for 1995-2010. We obtained the 2000-2010 24-hour average $\text{PM}_{2.5}$ and its speciated time series from the above-mentioned set of simulations with direct aerosol feedback. The earlier years are not included in this evaluation because of the limited availability of $\text{PM}_{2.5}$ observations during the earlier decade.

We use observations of 24-hour average total $\text{PM}_{2.5}$ concentrations across CONUS from US EPA’s Air Quality System (AQS) based on data availability. Given that most sites have a sampling rate of one every three days, those 95 sites with data coverage above 30% each year (equivalent to 90% for 1-in-3-day sampling sites) are used for model evaluation (Fig. 3-1). The sites are grouped into six geographical regions adapted from NOAA’s U.S. Climate Regions by combining West and Northwest to West (W) and combining West North Central (only 1 site), East North Central and Central to Midwest (MW). The region list and number of sites in each region can be found in Fig. 3-1. Most of the sites are located in urban (42 sites) or suburban (48) regions, with only 5 in rural regions.
Figure 3-1. PM$_{2.5}$ monitoring stations from AQS with at least 30% annual data for 2000-2010 grouped into 6 geographical regions adapted from NOAA’s U.S. Climate Regions: West (W), Southwest (SW), South (S), Midwest (MW), Southeast (SE) and Northeast (NE). Number in parentheses following each region indicates the site count.

3.3 Results and discussion

Model performance varies with geographical locations and seasons (Fig. 3-2). Total PM$_{2.5}$ concentrations in SW and W regions are largely underestimated for all seasons with low correlations of 0.36-0.66 for W and 0.13-0.4 for SW (large scatter at low concentrations and underestimation of observed high concentrations, especially in the 1$^{\text{st}}$ and 4$^{\text{th}}$ quarters of the year (Q1 and Q4)). The poor performance in cooler seasons (RMSE around and above 10 μg/m$^3$) in W and SW calls for a closer evaluation of PM$_{2.5}$ components to identify the causes of these discrepancies. The other four regions (S, MW, SE and NE) have a moderate model performance and a visible seasonal feature. The model reproduces total PM$_{2.5}$ better in the MW, SE and NE regions (correlations range from 0.56 to 0.72) mostly during the warmer seasons (Q2 and Q3). All PM$_{2.5}$ concentrations in cooler seasons (Q1 and Q4) are generally overestimated with higher errors.
Annual mean PM$_{2.5}$ is better simulated by the model than the daily concentrations with a median correlation of 0.86 compared with 0.57 and median RMSE of 2.77µg/m$^3$ compared with 7.26µg/m$^3$ (Fig. 3-3 a and b). Yet, the highly linear correlated annual mean indicated by the high correlation and high systematic RMSE (RMSEs), with a median of 93.5%, has a bias that is almost identical to that of daily PM$_{2.5}$ concentration. Similar bias and higher percentage of RMSEs for the annual values indicate that the majority of the errors are systematic, and thus, improvements are feasible. On the contrary, the daily PM$_{2.5}$ for Eastern US shows low systematic errors which indicates that the errors are primarily random, and thus difficult to correct. The high RMSEs of both daily and annual mean PM$_{2.5}$ in Western US can being explained by the regularly underestimated W and SW PM$_{2.5}$ concentrations seen in scatter plots of Fig. 3-2 a-b.

Trends in the PM$_{2.5}$ annual indexes, including the quarter mean, annual mean, and percentiles varying from the lower 10th to the higher 98th, are commonly reported in most dynamic evaluation studies (Gan et al., 2015, 2016). In this section, we summarize the evaluation of trends in all these indexes for the period 2000-2010, which are close to that seen in Gan et al. (2015) using a smaller number of sites from CASTNET and IMPROVE spanning 1995-2010.

In terms of the regional trend, the model captures the decrease of both annual mean and 98$^{th}$ percentile, except for the notable decrease in the 98th percentile in the West region (Fig. 3-4). However, the magnitude of both annual mean and 98th percentiles are underestimated in W and SW and overestimated in S and NE. Slight overestimation also appears in MW and SE. The model reproduces the majority of the year-to-year changes in the annual mean, such as the drop between 2007 and 2008 in SE (Fig. 3-4a). Unlike that of annual mean, it is more challenging for the model to replicate the year-to-year variations in the 98th percentile (Fig. 3-4b). Another interesting feature seen from the ranges of the box whiskers is that the spatial variations are underestimated in the W and overestimated in the eastern regions (S, MW, SE and NE).
Figure 3-2. Density scatter plots (binned every 2 μg/m$^3$) of 24-hour average PM$_{2.5}$ concentrations across each region and quarter for 2000-2010: (a1-4) West (W), (b1-4) Southwest (SW), (c1-4), South (S), (d1-4) Midwest (MW), (e1-4) Southeast (SE), (f1-4) Northeast (NE); 1-4 represent Quarter 1 (Jan-Feb-Mar) to Quarter 4 (Oct-Nov-Dec).
Figure 3-3. Statistical metrics of simulated 24-hour average (left column) and annual mean (right column) PM$_{2.5}$ concentrations for 2000-2010. (a-b) RMSE, (c-d) Systematic RMSE (RMSEs) (e-f) correlation coefficient (R) and (g-h) Bias.
Figure 3-4. Regional observed and modeled PM$_{2.5}$ (a) annual mean and (b) 98th percentiles for 2000-2010. Each box indicates the median, 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using open circles.

By comparing the regional trend in multiple percentiles and all four quarters during 2000-2010, CMAQ agrees with the decreasing sign of the observed trends and captures the characteristic changes of different percentiles, meaning higher change rates in higher percentiles (Fig. 3-5). However, the decreasing trends are generally underestimated in W and SW and overestimated in S and NE. In terms of seasonality, cooler seasons (Q1 and Q4) are more challenging to simulate than warmer seasons, likely due to the different contributions from multiple species. However, because of the limited number of sites that provide
speciated PM$_{2.5}$ concentrations, we are unable to confirm this statement in general. We do not see a difference between urban and suburban sites in the trends of various percentiles and quarters (not shown here).

Figure 3-5. Regional boxplots of observed and modeled trends of PM$_{2.5}$ at multiple percentiles and all four quarters for 2000-2010: (a) 10$^{th}$, 50$^{th}$ and 90$^{th}$ percentiles; (b) Quarters 1-4. Each box follows the rule as Fig. 6 except that the outliers are noted in crosses.

Zooming in to the spatial distribution of the annual mean and 98th percentile trends, the general decreasing trends in CMAQ agree with the observations (what we also see in Fig. 3-4). The limited increasing trends (orange and red) and the smaller decreasing trends (yellow) are not statistically significant (Fig. 3-6). The median trend of the annual mean is -0.39 µg/m$^3$/yr which is being simulated as -0.37 µg/m$^3$/yr, a difference much smaller than the trend of the 98$^{th}$ percentile (-1.05 µg/m$^3$/yr simulated to be -0.83 µg/m$^3$/yr). Most underestimation of the magnitude of the decreasing trends, marked red in Fig. 3-6 (c,
f) lies in W, SW and Florida, while the overestimation, marked blue in Fig. 3-6 (c, f) occurs in MW and NE.

**Figure 3-6.** Trends and trend differences in observed and modeled PM$_{2.5}$ annual mean and 98$^{th}$ percentiles for 2000-2010. Trends in annual mean (a) observed, (b) CMAQ and (c) their difference denoted as CMAQ-obs; Trends in 98$^{th}$ percentiles (d) observed, (e) CMAQ and (f) their difference. Significant trends with p<0.05 in (a-b, d-e) are circled.

### 3.4 Conclusions

The decadal WRF-CMAQ simulations are generally overestimating the PM$_{2.5}$ concentrations in the eastern United States and underestimating in the west. Percentage of the 98$^{th}$ percentile’s systematic error (83.5%) is much larger than that of unsystematic error (16.5%), which suggests that CMAQ’s ability to simulate exceedances in total PM$_{2.5}$ could be improved. A universal decreasing trend in PM$_{2.5}$ 10$^{th}$, 50$^{th}$, 90$^{th}$ and 98$^{th}$ percentiles can be seen from both observations and CMAQ. However, the notable decreasing trend in the 98th percentile that was observed in the West is not captured by the model. Even though the operational evaluation provides information of how the model is reproducing total PM$_{2.5}$ concentrations, assumptions on linearity of trends are not necessarily correct as the underlying mechanisms are in fact
nonlinear and nonstationary. This problem is addressed in the next chapter, where the application of empirical mode decomposition (EMD) allows for non-linear trend evaluation as well as assessment of various temporal-scale trends of total PM$_{2.5}$ and its speciated components.
Chapter 4
Evaluating Trends and Seasonality in Modeled PM$_{2.5}$ Concentrations Using Empirical Mode Decomposition

4.1 Introduction

It is well recognized that inhalable fine particulate matter (PM$_{2.5}$) adversely impact human health and the environment. Regional-scale air quality models are being used in health impact studies and decision-making related to PM$_{2.5}$. Long-term model simulations of PM$_{2.5}$ concentrations using regional air quality models are essential to identify long-term trends and variations such as annual cycles in areas larger than what is covered by in-situ measurements. However, total PM$_{2.5}$ concentrations are challenging to predict because of the dependence on the contributions from individual PM$_{2.5}$ components, such as sulfates, nitrates, carbonaceous species, and other natural species. In this context, a detailed process-based evaluation of the simulated speciated PM$_{2.5}$ must be carried out to ensure acceptable replication of observations so model users can have confidence in using regional air quality models for policy-making. Furthermore, process-based information can be useful for making improvements to the model.

Some of the trend or step change evaluations of regional air quality models in the past have focused on specific pairs of years (Kang et al., 2013; Zhou et al., 2013; Foley et al., 2015). These studies do not properly account for the sub-seasonal and inter-annual variations between those specific periods. Trend evaluation is commonly done by linear regression of indexes such as the annual mean or specific percentiles, assuming linearity and stationarity of time series (Civerolo et al., 2010; Hogrefe et al., 2011; Banzhaf et al., 2015; Astitha et al., 2017). The problem with the linear trend evaluation is that there is no guarantee the trend is actually linear during the period of the study because the underlying processes are in fact nonlinear and nonstationary (Wu et al., 2007).

Seasonal variations are usually studied and evaluated by investigating the monthly or seasonal means (Civerolo et al., 2010; Banzhaf et al., 2015; Yahya et al., 2016; Henneman et al., 2017). Evaluation
of ten-year averaged monthly mean of PM$_{2.5}$ simulated with WRF/Chem against the Interagency Monitoring of Protected Visual Environments (IMPROVE) by Yahya et al. (2016) shows that the model captures the observed features of summer peaks in PM$_{2.5}$ with a phase shift of few months. However, according to the analysis (Fig. 10) in Henneman et al. (2017), the seasonality shown in monthly-averaged PM$_{2.5}$ time series is much less distinguishable compared with that of ozone and CMAQ (version 5.0.2) does not replicate the monthly PM$_{2.5}$ quite well with large underestimation in the summer months. In these studies, the seasonality might not be well represented by the preselected averaging window size of one or three months. In addition, averaging of those monthly or seasonal means across multiple years may conceal the long-term trends or interannual variations driven by climate change, emission control policies or other slow varying processes.

To address the above-mentioned problems, we propose a new method for conducting air quality model evaluation for PM$_{2.5}$ using improved CEEMDAN. Improved CEEMDAN is an Empirical Mode Decomposition (EMD)-based, data-driven intrinsic mode decomposition technique that can adaptively and recursively decompose a nonlinear and nonstationary signal into multiple modes called intrinsic mode functions (IMFs) and a residual (trend component) (Huang et al., 1998; Wu and Huang, 2009; Yeh et al., 2010; Torres et al., 2011; Colominas et al., 2014). It does not require any preselection of the temporal scales or assumptions of linearity and stationarity for the data, thereby providing some insights into time series of PM$_{2.5}$ concentrations and its components. Decomposed PM$_{2.5}$ long-term trend components and annual cycles from observed and simulated PM$_{2.5}$ serve as the intuitive carrier of the trend and seasonality evaluation. In the meantime, several other IMFs with characteristic time scales ranging from multiple days to years are also decomposed, enabling model evaluation of the less studied sub-seasonal and inter-annual variations.

Section 2 describes the coupled WRF-CMAQ model simulations and corresponding observations from multiple speciated PM$_{2.5}$ networks. Section 3 presents an overview of the EMD and improved CEEMDAN technique and the statistical metrics accompanying model evaluation, including the time-dependent intrinsic correlation (TDIC) on the decomposed IMFs (Chen et al., 2010; Huang and Schmitt, 2014; Derot et al., 2016). Section 4 describes the findings on the long-term trend and seasonality in total
PM$_{2.5}$ and its components, as resolved by the improved CEEMDAN technique and includes a discussion on the sub-seasonal, seasonal, and inter-annual variability. The conclusions from this work are presented in section 5.

4.2 Coupled WRF-CMAQ PM$_{2.5}$ Simulations and Observations

The two-way coupled WRF-CMAQ (version 5.0.2) is configured with a 36 km horizontal grid spacing over the contiguous United States (CONUS) with 35 vertical layers of varying thickness extending from the surface to 50 mb (Wong et al., 2012; Gan et al., 2015). Time-varying chemical lateral boundary conditions were derived from the 108 km resolution hemispheric WRF-CMAQ (Mathur et al., 2017) simulation for the 1990-2010 period (Xing et al., 2015). The simulations are driven by a comprehensive emission dataset which includes the aerosol precursors and primary particulate matter (Xing et al., 2013, 2015). The readers can refer to Gan et al. (2015) for additional model information and the trend evaluation against seven pairs of sites from the CASTNET (Clean Air Status and Trend Network) and IMPROVE networks for 1995-2010. We obtained the 2002-2010 daily average PM$_{2.5}$ and its speciated time series from the set of simulations with direct aerosol feedback. The earlier years of 1990-2001 are not included in this evaluation because of the limited availability of speciated PM$_{2.5}$ observations.

To avoid misinterpretation of data due to the presence of missing values, only sites with continuous complete long-term record for total PM$_{2.5}$ and its speciation including SO$_4$, NO$_3$, NH$_4$, OC, EC, and Cl are studied (Fig. 4-1). All of the selected sites have data coverage above 90% each year for at least six consecutive years between 2002 and 2010 (equivalent to 30% for 1-in-3 days sampling sites). This strict data selection led to the sparsity of this type of observations for the study period. QURE, a rural site carrying out 1-in-3 days sampling of total and speciated PM$_{2.5}$ of SO$_4$, NO$_3$, OC, EC, and Cl, is located in Quabbin Summit, MA. It is one of the three sites from the IMPROVE network that has at least six continuous years of speciated observations and was selected here to demonstrate the application of the proposed method in rural areas. It should be noted that the majority of the observed Cl in 2002 and 2003 is negative due to a filter issue problem which was not addressed until 2004 (White, 2008). Thus, simulations of Cl are only
evaluated during 2004-2007 at this site. Station RENO, located in urban Reno, NV, is also a 1-in-3 days sampling site of total and speciated PM$_{2.5}$ of SO$_4$, NO$_3$, NH$_4$, OC, and EC, and it is the only Chemical Speciation Network (CSN) site that fulfills this data coverage requirement. The third site ATL in the Southeastern Aerosol Research and Characterization Study (SEARCH) network is located 4.2 km northwest of downtown Atlanta, GA. It is the only long-term site available with daily sampling rate (Hansen et al., 2003; Edgerton et al., 2005) that meets the data coverage requirement. The best-estimate (BE), a calculated concentration intended to represent what is actually in the atmosphere (Edgerton et al., 2005), of the total PM$_{2.5}$ and SO$_4$, NO$_3$, NH$_4$, and EC components are retrieved for the evaluation. OC component is a direct measurement. These three sites have a continuous record covering at least 6 years (2002 – 2007 for QURE and ATL and 2002 – 2010 for RENO) that allows an evaluation of long-term trends.

![Figure 4-1. Location and data coverage of the PM$_{2.5}$ monitoring sites QURE, RENO and ATL.](image)

**4.3 Method**

4.3.1 Empirical Mode Decomposition

The Empirical Mode Decomposition (EMD) technique, proposed in the late 1990s, is capable of adaptively and recursively decomposing a signal into multiple modes called intrinsic mode functions (IMFs), where each mode has a characteristic frequency, and a residual with at most one extremum (Huang et al., 1998). The decomposed signal then is expressed as the summation of all IMFs and the residual:
\[ x = \sum_{i=1}^{k} d_i + r \]  

where \( x \) is the original signal, \( d_i \) is the \( i \)th IMF, \( k \) is the number of the IMFs and \( r \) is the final residual. Each IMF has the following properties (Huang et al., 1998):

1) The number of extrema (maxima and minima) and the number of zero-crossings must be equal or differ at most by one;
2) The local mean at any point, the mean of the envelope defined by local maxima and the envelope defined by local minima, must be zero.

Nevertheless, “mode mixing” where oscillations with very disparate scales can be present in one mode or vice versa is commonly reported. To cope with this issue, multiple noise assisted EMD have been developed successively (Wu and Huang, 2009; Yeh et al., 2010; Torres et al., 2011; Colominas et al., 2014). It is evident that the latest improved Complete Ensemble EMD with Adaptive Noise (improved CEEMDAN) manages to alleviate the problem of mode mixing with the benefit of reducing the amount of noise presented and avoiding spurious modes (Colominas et al., 2014). Moreover, the end effects or boundary effects have been addressed by its predecessor EEMD (Ensemble Empirical Mode Decomposition) by extrapolating the maxima and minima, and behaved well in numerous time series with dramatically variant characteristics (Wu and Huang, 2009). The extrapolation of maxima and minima is proven to be more effective compared with the extrapolation of the signal itself such as repetition or reflection (Rato et al., 2008).

Given the EMD’s ability to deal with real-world nonstationary and nonlinear time series data, it is widely used in engineering, economics, earth and environmental sciences (e.g., Huang et al., 1998; Chang et al., 2003; Yu et al., 2008; Colominas et al., 2014; Derot et al., 2016). We use the most up-to-date noise-assisted improved CEEMDAN technique with at least hundreds of noise realizations to decompose observed and simulated PM\(_{2.5}\) time series. Readers can refer to Colominas et al. (2014) for detailed description of the technique and access to the corresponding MATLAB code. Trial and error attempts are made in setting the input of the improved CEEMDAN function to achieve best mode separation.
The impact of boundaries on the decomposed annual cycles and the residual is assessed by the variations (standard deviation) of hypothetical decomposed boundaries by cutting a continuous eighteen-year total PM$_{2.5}$ observation (North Little Rock, AR) 48 times at different years and times of the year (Fig. 4-S1). The standard deviation is found to largely diminish within half the annual cycles and could be negligible within one year for the annual cycle. This could very possibly expand to IMFs with other characteristic scales. Yet, trend components (residuals) show variability depending on the available time period after cutting. Most of the time, they follow the reference long-term trend reflected either by the residual or the summation of the residual and the IMF with longest temporal scale decomposed from the eighteen-year PM$_{2.5}$ (Fig. 4-S1c). This is in line with our expectations as a trend should exist within a given time span, following the definition in Wu et al. (2007): “The trend is an intrinsically fitted monotonic function or a function in which there can be at most one extremum within a given data span”. Although very strict data completeness requirement is employed for this study, it should not be conceived as a limitation of the method itself. A sensitivity test based on a period of nine years of total PM$_{2.5}$ observation at the same site with 99% data coverage shows that even though variability of annual cycles and long-term trends increases with decreased data availability (100%, 90%, ..., 10%), the structure of those components is consistent. The average of 40 realizations of annual cycles and long-term trend components in each data-completeness scenario is in perfect alignment with that of 100% data completeness (Fig. 4-S2 and 4-S3). Given the fact that those 40 realizations in each scenario are based on independent random samplings of the original observations, the increased variability could very possibly result from the difference in the sampled data itself rather than the method. Thus, the robustness of improved CEEMDAN decomposed annual cycles and long-term trend is justified. In fact, EMD has been proven to be an effective tool for data gap-filling (Moghtaderi et al., 2012).

The characteristic period of each IMF can be estimated by the peak period $t_p$ (days) where the power spectrum of the IMF peaks:

$$t_p = \frac{1}{f_p} \quad \text{(4 - 2)}$$
in which $f_p$ is the frequency that the power spectrum peaks in the unit of number of cycles per day. The peak estimates can be biased if more than one high-power frequency is located close to each other in one IMF. Thus, power spectrum is only used as a fast screening tool to determine if a desired decomposition is accomplished. As an alternative approach, the mean period $t_m$ can be estimated by:

$$t_m = \frac{\text{Time span}}{n_{\text{max}} + n_{\text{min}} + n_{\text{zero}}} \quad (4-3)$$

where $n_{\text{max}}$, $n_{\text{min}}$ and $n_{\text{zero}}$ are the number of maxima, minima and zero-crossings, respectively, during the Time span (days). As the frequency decreases, the mean period estimates become less accurate because of the limited time span compared with the length of the cycle and should be carefully interpreted.

An example of the total PM$_{2.5}$ decomposition with improved CEEMDAN at the QURE site shows modes ranging from very high frequency to very low frequency (IMF1 to IMF7) and a residual (Fig. 4-2). Mean ($t_m$) and peak ($t_p$) estimations of the characteristic periods of each IMF are presented on the right side of each mode. Annual cycles and long-term trend components are well represented by IMF6 and the residual, with the remaining IMFs carrying weekly, sub-seasonal, seasonal, and inter-annual variations, respectively, for both observed and simulated PM$_{2.5}$ (Fig. 4-2). We have noticed that in some rare cases, a spurious mode in the last IMF with synchronous signal and very close scales to its previous IMF exists. This is possibly due to the fact that the characteristic periods of those IMFs are in proximity to the span of the studied time span. In these cases, the last two modes are merged by adding those two modes together to conduct a detailed evaluation as discussed in Section 4.

Decomposition of observed (blue) and simulated (red) 24-hour average total PM$_{2.5}$ into 7 IMFs and a residual component (trend) at Quabbin Summit, MA using the improved CEEMDAN: (a) Time series of total PM$_{2.5}$, IMFs and the residual component (all with unit of $\mu g/m^3$); (b) Power spectrum of the corresponding time series. The colored numbers on the right side of time series are the mean period $t_m$ in days, while the ones on the right side of the power spectrum are the peak period $t_p$ in days, which are also indicated by the dashed vertical lines on the power spectrum. Note that all power spectra are in the log scale.
and those of the IMFs are zoomed in with a range of 100 to 104 on the y-scale for better visual clarity (total PM$_{2.5}$ and the residual component: 10^-2 to 10^7).

Figure 4-2. Decomposition of observed (blue) and simulated (red) 24-hour average total PM$_{2.5}$ into 7 IMFs and a residual component (trend) at Quabbin Summit, MA using the improved CEEMDAN: (a) Time series of total PM$_{2.5}$, IMFs and the residual component (all with unit of $\mu$g/m$^3$); (b) Power spectrum of the corresponding time series. The colored numbers on the right side of time series are the mean period $t_m$ in days, while the ones on the right side of the power spectrum are the peak period $t_p$ in days, which are also indicated by the dashed vertical lines on the power spectrum. Note that the scales for the time series are not all the same. Also, all power spectra are in the log scale and those of the IMFs are zoomed in with a range of 10^0 to 10^4 on the y-scale for better visual clarity (compared with 10^-2 to 10^7 for total PM$_{2.5}$ and the residual component).
4.3.2 Statistical metrics

EMD-decomposed IMFs and trend components allow for a detailed time-dependent evaluation of PM$_{2.5}$ and provide a novel opportunity to trace the performances of specific scales back to the corresponding speciated components. Note that the trend component is the decomposed residual component from the PM$_{2.5}$ in the unit of µg/m$^3$ and it is not the traditional concept of trend in concentration per time. In addition to a direct evaluation of its magnitude, we also calculated its derivative to identify the periods with higher or lower rate of change (concentration per time). Time-dependent intrinsic correlation (TDIC) is utilized to study the evolution of the model performance for cyclic variations throughout time (Chen et al., 2010; Huang and Schmitt, 2014; Derot et al., 2016). It is a set of correlations calculated for IMFs over a local period of time $I$ centered around time $t$:

$$ I(t) = \left[ t - \frac{t_w}{2}, t + \frac{t_w}{2} \right] $$

in which $t$ is the center time for the calculation of the correlation and $t_w$ is the moving window length. The minimum of $t_w$ is set to be the local instantaneous period of the IMF (larger of that in observation or simulation) using the general zero crossing method to ensure that at least one instantaneous period is included in calculating the local correlation coefficient (Chen et al., 2010). The maximum of $t_w$ is the entire data period with a traditional overall correlation being calculated. The empty spaces in the pyramids used to depict the TDIC are an indication that the correlation is not statistically significantly different from zero. With both decomposed observed and modeled concentrations in a narrow scale range, the correlation would no longer be contaminated by coexisting signals of different scales (Chen et al., 2010).

In order to summarize the performance of the decomposed trend component and IMFs, the ratio of the mean magnitudes of the trend components is defined as:

$$ r_{trend} = \frac{Mean_{CMAQ}}{Mean_{observation}} $$

(4 – 5)
where $Mean_{CMAQ}$ and $Mean_{observation}$ represent the mean of simulated and observed residual components respectively. The ratio of the mean amplitude of each IMF is defined by Equation 6, where an example for the annual cycles is provided:

$$
\tau_{\text{annual}} = \frac{RMS_{\text{CMAQ,annual}}}{RMS_{\text{observation,annual}}}
$$

(4 – 6)

where $RMS_{\text{observation,annual}}$ and $RMS_{\text{CMAQ,annual}}$ represent the root mean square of observed and simulated annual cycles respectively. Finally, the phase shift of an IMF $n$ is defined to be days an IMF decomposed from modeled time series has to shift in order to achieve the highest correlation ($R_{max}$) with the corresponding IMF with similar scale from observed PM$_{2.5}$ time series. In practice, $n$ could be as much as a few cycles of the mean period, $t_m$. Here, we limit the absolute number of shift days to not exceed a half cycle as a reference for the phase shift of an IMF. Thus, $n$ satisfies $-\left(\frac{t_m}{2}\right) \leq n \leq \left(\frac{t_m}{2}\right)$ with $t_m$ being the larger mean period in observation or simulation. It becomes $-0.5 \leq \frac{n}{t_m} \leq 0.5$ in terms of number of cycles.

4.4 Results and discussion

4.4.1 Temporal scales

Temporal scales in PM$_{2.5}$ resolved by EMD depend solely on the intrinsic properties of the data itself. These properties include underlying characteristics of specific PM$_{2.5}$ concentrations, the data sampling frequency, which determines the scales that can be resolved in the high frequency IMFs, and the time span for the data coverage, which could possibly play an important role in differentiating the low frequency IMFs from the trend component. Here, we first evaluate the scales represented by the mean period in the speciated PM$_{2.5}$ time series. Note that the mean period is only one indication of the model evaluation against observations, and it does not indicate any information on the magnitude or the phase of the time series, which will be further discussed in Sections 4.3 to 4.4.
Fig. 4-3a presents the characteristic scales of IMFs in observed and simulated total and speciated PM$_{2.5}$ of QURE. The CMAQ model compares well with the observations for IMFs 1 through 6 with cycles of 9, 19, 37, 78, 158 and 347 days (average of all observed and simulated total and speciated PM$_{2.5}$). Among all these IMFs, IMF6, which represents the annual cycles, shows the least variations in the characteristic scale (Fig. 4-3a) and highest peak energy from the power spectrum such as Fig. 4-2b for total PM$_{2.5}$, except for observed EC and OC where the power of half-year cycles is more dominant (Fig. 4-S4). These two features demonstrate a clear seasonality in both observed and simulated total and speciated PM$_{2.5}$, which would otherwise be concealed by practices such as monthly averaging. This can be further confirmed by the statistically significant annual cycles (except for observed EC and OC) (Fig. 4-S5) based on a Monte Carlo verified relationship between the energy density and mean period of IMFs (Wu and Huang, 2004; Wu et al., 2007). To explore the inter-annual cycles in more detail, mean periods of IMFs with scales longer than a year are being displayed in the top left panel of Fig. 4-3a. Some variability exists between the observation and model simulation to the extent that not all IMFs from observation are being simulated and vice versa. The estimated mean periods of the inter-annual cycles and the differences in the presence of slow varying cycles with the long characteristic scales are likely to be influenced by their proximity to the data time span of 6 years (4 years for Cl). This implies that the model evaluation shouldn’t go beyond 3 years (2 years for Cl) given the current data coverage. CMAQ captured the 3-year cycles in EC and total PM$_{2.5}$ and 2-year cycles in OC and Cl, despite an overestimation in the scales of 2-year cycles in observed SO$_4$ and NO$_3$.

Similar features in observed and simulated total and speciated PM$_{2.5}$ concentrations at RENO are presented in Fig. 4-3b. Likewise, the highest peaks in the power spectrum also sit in the annual cycles of IMF6 except for the observed OC and total PM$_{2.5}$ which have higher peak power at half-year cycles. All annual IMFs are statistically significant except for simulated NH$_4$ (Fig. 4-S5). The small variation in the estimated characteristic period of IMF6 is because this monitoring site is located in a wildfire prone region on the border of Nevada and California. Clear evidence can be seen from Fig. 4-4a that an extra annual cycle in the IMF6 of observations in the summer of 2008 is depicted, which is very possibly driven by the
2008 California Wildfires spanning from May until November. Unlike the diversified scales in IMF7 at QURE, IMF7 at RENO features universal 2-year cycles of all species as well as total PM$_{2.5}$ and all of them are well replicated by the model. However, variations in time scales are present in IMF8 possibly because of the limited data coverage. Thus, only species with time scales less than 4 years in both observations and model simulations are evaluated. It is evident that CMAQ has reproduced the 3-year cycles in SO$_4$ and NH$_4$.

**Figure 4-3.** The characteristic scales resolved in the IMFs of observed and simulated total and speciated PM$_{2.5}$ for (a) QURE, (b) RENO and (c) ATL. IMF1 to the last pair of IMFs with increasing characteristic periods are shown from bottom left to top right. Top left panel in each subplot shows characteristic scales in the unit of years (y-axis) of all IMFs with inter-annual cycles (the x-axis represents the IMF number). In the subplots, species decomposed from observations are connected by solid lines, while species decomposed from simulations are represented by smaller markers in darker shades connected by dashed lines.

ATL is the only speciated site with daily data coverage. Observed and simulated total and speciated PM$_{2.5}$ concentrations at the ATL site are decomposed into 9 or 10 IMFs (Fig. 4-3c). Because of the change in data frequency, high frequency scales such as weekly cycles can be evaluated and the significance tested (Fig. 4-S5) annual cycles with the highest peak power is represented by IMF8(IMF7 for SO$_4$ and NO$_3$). Annual cycles of SO$_4$ and NO$_3$ appeared in the earlier stage of decomposition in IMF7 because of their relatively weak half-year cycles, which largely led to the mixed signal of half year and annual cycles in IMF7 in total PM$_{2.5}$ as in Fig. 4-5b. This is more visible in the observed IMF7 where the energy of the one-year period surpasses that of the half year. Yet, clues can be seen from Fig. 4-5 that the amplitude and the
energy of annual cycles leaked into IMF7 is very limited compared to that remaining in IMF8, indicating that it is still safe to conduct model evaluation on the seasonality using IMF8 with an underestimation in the amplitude of observation. On the other hand, inferences should be made with caution for IMF7 because of the mixed modes. Scales up to 3 years are relatively well reproduced by the model.

**Figure 4-4.** Same as Fig. 4-2 but for RENO site with 8 IMFs.
4.4.2 Long-term trend

The EMD-decomposed long-term trend components for the observed and simulated total and speciated PM\textsubscript{2.5} concentrations are presented in Fig. 4-6. To better visualize the non-linearity of the trend component, the rates of change (temporal derivative of a trend component, which is the change in the consecutive concentration divided by the sampling rate of 1 or 3 days and converted to the unit of
µg/m³/year by multiplying 365 day/year) are added with a separate y-axis on the right side in each panel (gray colored scale). It is evident that PM$_{2.5}$ is changing at a varying rate, forming either a monotonic trend component or a trend component with one extremum, which cannot be fully represented by a single constant number using a traditional linear regression approach. Given that there are chemical species other than the ones studied in the total PM$_{2.5}$, not all performance issues can be fully explained by the five available species.

**Figure 4-6.** Trend components of observed and simulated total and speciated PM$_{2.5}$ for (a) QURE, (b) RENO and (c) ATL in µg/m$^3$ with dashed lines representing the rate of the change (temporal derivative of the trend component converted to µg/m$^3$/year) against the right-side y axis, with a reference line of no change in dark gray line in the center.

At the QURE site, CMAQ captures the general decreasing trend in observed total PM$_{2.5}$ which can mainly be traced back to NO$_3$ and OC, while both observed and simulated trend components in SO$_4$ and EC are relatively constant (Fig. 4-6a). Moreover, the periods with highest decreasing rate in observed total PM$_{2.5}$ during 2003-2004 with a decreasing rate of -0.44 µg/m$^3$/year is also well replicated by the model.
Nevertheless, the slightly increasing PM$_{2.5}$ level in the later years is simulated to be decreasing at a much higher rate, which is partly due to the overestimated decreasing rate in OC and species other than the five studied ones. The trend component of simulated Cl shows a cyclic-like feature because of proximity between the existence of a cycle of 4-5 years (by decomposing the simulation during the 6-year study period) and 4-year period limited by the available quality assured observations. The rate of change in the simulated trend component by decomposing the simulation during the 6-year study period would mimic that from the 4-year observation, both with a negligible negative value throughout 2004-2007. However, the magnitude of the trend component is almost doubled (1.8 times compared with observation) in the model with contribution from all species except for SO$_4$. A quantitative summary of the magnitude of the trend component can be found in Table 4-1.

Table 4-1. The ratio of mean magnitude of the trend component $r_{trend}$ (CMAQ/observation). Boldface values indicate a relatively good estimate of the magnitude (0.7 - 1.3). “-” indicates the data is not available (same applies for Tables 4-2 and 4-3).

<table>
<thead>
<tr>
<th></th>
<th>TOT</th>
<th>SO$_4$</th>
<th>NO$_3$</th>
<th>NH$_4$</th>
<th>OC</th>
<th>EC</th>
<th>Cl</th>
</tr>
</thead>
<tbody>
<tr>
<td>QURE</td>
<td>1.8</td>
<td>0.9</td>
<td>3.5</td>
<td>-</td>
<td>1.4</td>
<td>1.7</td>
<td>1.3</td>
</tr>
<tr>
<td>RENO</td>
<td>0.8</td>
<td>1.3</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>-</td>
</tr>
<tr>
<td>ATL</td>
<td>1.2</td>
<td>1.0</td>
<td>2.1</td>
<td>1.0</td>
<td>0.9</td>
<td>1.4</td>
<td>-</td>
</tr>
</tbody>
</table>

RENO is located close to the border with California and is affected by large wildfire breakouts in the western U.S. as can been seen in the spikes of the observed total PM$_{2.5}$ (Fig. 4-4a). The model simulates large increasing rate up to 1.03 $\mu$g/m$^3$/year and decreasing rate up to -0.80 $\mu$g/m$^3$/year before and after the 2006-2007 winter season and fails to reproduce the relative stable condition seen in the observations with only -0.09 $\mu$g/m$^3$/year decreasing in 2004-2005 and 0.04 $\mu$g/m$^3$/year increasing in 2008-2009 (Fig. 4-6b). Similar feature is found for combustion related OC and EC species. The observed slightly decreasing trends in SO$_4$ and NH$_4$ during 2005-2009 are not being captured in the model simulations. The magnitude of the
trend component is slightly underestimated with $r_{trend}$ of 0.8 with contribution from all species except for SO$_4$ as well (Table 4-1).

During the period of 2002-2007, observations at ATL reveal a slightly increasing PM$_{2.5}$ trend that cannot be explained by the five listed PM$_{2.5}$ components trend (Fig. 4-6c), possibly indicating a contribution of the remaining species such as the non-carbonaceous portion of organic matter. Non-carbonaceous organic matter can account for more than half of total organic matter, which, in turn, can account for a large portion of the total PM$_{2.5}$ mass (Edgerton et al., 2005). In contrast, the model shows a slight decreasing trend with a peak decreasing rate in 2003 and misses the peak increasing rate of 0.23 $\mu$g/m$^3$/year in the winter season of 2005. Similarly, reversed trends are also simulated for SO$_4$, OC and EC, while the change rate in NO$_3$ is well captured. Unlike the previous sites, magnitude of trend components in total and speciated PM$_{2.5}$ are well simulated except for EC (1.4 times the observation) and NO$_3$ (2.1 times).

To sum up, the long-term trend at QURE is well simulated by the model. The occurrence of large wildfires lasting for several months have significantly impacted the long-term trend component at RENO and the model failed to capture those combustion-related species and total PM$_{2.5}$ primarily due to limitations in the historical data used to specify day-specific wildfire emissions (Xing et al., 2013). Slightly increasing levels of PM$_{2.5}$ and its species observed at ATL are simulated to be slightly decreasing, except for NO$_3$ which is well simulated. The magnitude of the long-term trend components of total PM$_{2.5}$ and SO$_4$ are well represented by CMAQ (Table 4-1). The model performs differently across the sites in terms of the magnitudes of the trend component in NO$_3$, NH$_4$, Cl, OC and EC. Species other than those in the available dataset may also play a considerable role in driving the agreements or disagreements between model simulations and observations of total PM$_{2.5}$.

4.4.3 Seasonality

The EMD-assisted seasonality evaluations utilize the decomposed IMF with characteristic period of one year to evaluate the amplitude and phase of the model simulation, both of which are time-dependent.
We first demonstrate the evaluation for total PM$_{2.5}$ at QURE (Fig. 4-7a). The top panel shows the annual cycle components and the bottom panel shows its TDIC pyramid. The decreasing amplitude of the annual cycles throughout 2002-2007 is almost perfectly represented with an overall ratio $r_{\text{annual}}$ being 1.0 (Table 4-2). Each pixel in the TDIC pyramid is the correlation (color-coded) calculated during a period of time $I(t)$ with width of $t_w$ days (y-axis) centered at a specific day (x-axis) as introduced in Section 3.2. The annual cycle mean periods are identical between CMAQ and observations (350 days, Fig. 4-2a IMF6), but there is a phase shift for all years with the entire TDIC pyramid being close to -1. By shifting the CMAQ annual cycles backward 159 days (almost half year), the overall correlation of the annual component can reach up to a peak of 0.9 (Table 4-3).

**Figure 4-7.** Decomposed annual cycles (IMF6) from observed (blue) and simulated (red) concentrations ($\mu g/m^3$) of (a) total PM$_{2.5}$, (b) SO$_4$, (c) NO$_3$, (d) Cl, (e) OC and (f) EC and their corresponding TDIC at Quabbin Summit, MA. The window size $t_w$ indicates the width of the window used to calculate a specific correlation centered at the day represented in x-axis.

What are the driving factors for the above phase shift in modeled total PM$_{2.5}$ at Quabbin Summit, MA? The illustrations in Fig. 4-7a for total PM$_{2.5}$ alone cannot provide useful information that will allow
the modeler to improve the model’s performance. This is accomplished by applying the EMD method to the PM$_{2.5}$ speciated components (Fig. 4-7b-f). Traces of the semi-annual phase shift (-159 days) of annual cycles or large overestimation in the winter and underestimation in the summer is because of the largely overestimated amplitude of NO$_3$ (4.3 times that of observation) which peaks in the winter and the almost semi-annual shifted OC (-147 days), as well as contributions from EC and Cl. NO$_3$ has a mean amplitude reaching almost half of that of the total PM$_{2.5}$. OC directly drives both the observed and simulated annual components to be negatively correlated. EC follows the feature of OC in the first four years or so and the feature of NO$_3$ in 2006 and 2007 and contributes to the half year shifted total PM$_{2.5}$. The magnitude of winter-peaking Cl cycles are overestimated with a phase shift of one month. However, the contribution of Cl is very limited because of the tiny amplitude in both observed and simulated annual cycles. In addition, annual cycles in SO$_4$ are well reproduced for the entire time span with an amplitude ratio of 0.7. A quantitative summary of the evaluation of the annual cycles at this site can be found in Tables 4-2 and 4-3.

**Table 4-2.** The ratio of mean amplitude of the annual component $r_{\text{annual}}$ (CMAQ/observation). Bold indicates a magnitude with a ratio close to 1 ($0.7 – 1.3$).

<table>
<thead>
<tr>
<th></th>
<th>TOT</th>
<th>SO$_4$</th>
<th>NO$_3$</th>
<th>NH$_4$</th>
<th>OC</th>
<th>EC</th>
<th>Cl</th>
</tr>
</thead>
<tbody>
<tr>
<td>QURE</td>
<td>1.0</td>
<td>0.7</td>
<td>4.3</td>
<td>-</td>
<td>1.6</td>
<td>3.1</td>
<td>1.6</td>
</tr>
<tr>
<td>RENO</td>
<td>1.2</td>
<td>0.5</td>
<td>0.1</td>
<td>0.2</td>
<td>1.5</td>
<td><strong>0.9</strong></td>
<td>-</td>
</tr>
<tr>
<td>ATL</td>
<td>0.5</td>
<td><strong>0.7</strong></td>
<td>2.4</td>
<td>0.4</td>
<td><strong>1.2</strong></td>
<td><strong>1.0</strong></td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 4-3.** Phase shift ($n$) of CMAQ simulated annual cycle components in days. The background color indicates the maximum correlation ($R_{\text{max}}$) that can be reached by shifting the CMAQ time series $n$ days with respect to observations: white = [0.8, 1], light grey = [0.6, 0.8], grey = [0.4, 0.6), dark grey = (0.2, 0.4). The bold shows number of shifts less than a month while the italic shows shifts longer than three months.

<table>
<thead>
<tr>
<th></th>
<th>TOT</th>
<th>SO$_4$</th>
<th>NO$_3$</th>
<th>NH$_4$</th>
<th>OC</th>
<th>EC</th>
<th>Cl</th>
</tr>
</thead>
<tbody>
<tr>
<td>QURE</td>
<td>-159</td>
<td>-6</td>
<td><strong>3</strong></td>
<td>-</td>
<td>-147</td>
<td>-105</td>
<td>-30</td>
</tr>
<tr>
<td>RENO</td>
<td>78</td>
<td>36</td>
<td><strong>12</strong></td>
<td>-21</td>
<td>33</td>
<td>96</td>
<td>-</td>
</tr>
<tr>
<td>ATL</td>
<td>-132</td>
<td>0</td>
<td><strong>8</strong></td>
<td>-17</td>
<td>-24</td>
<td>-54</td>
<td>-</td>
</tr>
</tbody>
</table>
Both observed and simulated annual cycles at the RENO site are largely contaminated by the extreme events lasting for several months that are not properly simulated, indicating the need for more appropriate emissions allocation. Overall, annual variations for total and speciated PM$_{2.5}$ are largely underestimated except for the total PM$_{2.5}$ and combustion-driven EC and OC from 2005 to 2007 (Fig. 4-8). The modeled phase of SO$_4$, NO$_3$, NH$_4$ and OC agrees with that of observation with exception for a length of about two years in each that missed the phasing: 2009-2010 for SO$_4$, summer 2005-summer 2007 for NO$_3$, 2006-2007 for NH$_4$ and 2004-2005 for OC. It is also notable that the TDIC pyramid of EC mimics that of total PM$_{2.5}$, implying the existence of errors in modeled EC in processes such as emissions, transport, and deposition that affected the model performance for total PM$_{2.5}$. In comparison, SO$_4$ and OC are relatively well simulated with a mean amplitude ratio of 0.5 and 1.5 and a phase shift of 36 and 33 days, respectively.

![Figure 4-8](image)

Figure 4-8. Same as in Fig. 4-7 for Reno, NV, except that (d) represents NH$_4$ rather than Cl.

Observed annual cycles of total PM$_{2.5}$ at the ATL site features a slightly increasing amplitude of annual variations from 2002 to 2006 which then decreased to the original state in 2007 (Fig. 4-9a).
Conversely, model-simulated annual cycles became weaker throughout the period, with an overall $r_{\text{annual}}$ of 0.5. As at the QURE site, the simulated annual components at the ATL site also show a shift of several months (-132 days). Specifically, traces of these phase shifts or large overestimation in the winter and underestimation in the summer can be seen from the more than doubled amplitude of NO$_3$ which peaks in winter and underestimated SO$_4$ and NH$_4$ in the warm seasons as well as the -54 days shifted EC. The anti-correlated remaining species other than those in the available dataset clearly played a role in driving the discrepancies seen in the total PM$_{2.5}$ annual cycles (Fig. 4-10). Specifically, the anti-correlation likely points to an inaccurate representation of the seasonal variation of the non-carbonaceous portion of organic matter due to an improper representation of organic aerosols in the model version analyzed here; this problem has since been corrected in more recent releases of the CMAQ model. The underestimated annual variations in the remaining components closely resemble that of the annual variation in total PM$_{2.5}$. The phase of simulated SO$_4$, NO$_3$, NH$_4$, and OC species is in good agreement with those in observations and the amplitude of simulated annual cycles in SO$_4$, OC and EC agree well with that in the observations (Tables 4-2 and 4-3).

In sum, annual cycles of PM$_{2.5}$ are also time-dependent and the phase in the annual cycles for total PM$_{2.5}$, OC and EC reveal a general shift of up to half a year (Table 4-3); this indicates a potential problem in the allocation of emissions during this study period and/or the treatment of organic aerosols in this version of the model. CMAQ generally simulated the phase in SO$_4$, NO$_3$, Cl and NH$_4$ quite well but did not always capture the magnitude of their variations (Table 4-2).
Figure 4-9. Same as in Fig. 4-7 for Atlanta, GA, except that the annual component is resolved in IMF8 (IMF7 for SO$_4$ and NO$_3$) because of the difference in sampling rate and characteristic embedded in the time series at ATL site and (d) represents NH$_4$ rather than Cl.

Figure 4-10. Decomposed annual cycles in Atlanta, GA for the remaining components presented in total PM$_{2.5}$ other than the five species in Fig. 4-9.
4.4.4 Sub-seasonal and inter-annual variability

In this section, model performance at multiple sub-seasonal and inter-annual scales with cycles less than 3 years, presented in the total and speciated PM$_{2.5}$, is evaluated following an approach similar to that for the annual cycles in Section 4.3 (Fig. 4-11). First, IMFs from observations and model simulations are paired based on their characteristic periods following the discussion in Section 4.1. Then, the magnitude of specific scales is evaluated using $r_{IMFn}$ following Equation 6 of the $r_{annual}$ for annual cycles. The phase shifts of the time series are assessed by the proportion of shifted days relative to the mean characteristic scales of the corresponding observed and simulated IMFs ($\frac{n}{T_m}$). For example, a phase shift of 0.1 cycles in the 2-year cycles is approximately 73 days while it would be 18 days for the half-year cycles.

The performance of the simulated amplitude of the sub-seasonal and inter-annual cycles is relatively stable from a few days to semi-annual scales and $r_{IMFn}$ is close to 1 in most cases (Fig. 4-11a-c). CMAQ captures the features seen in the observations at QURE, except for the large overestimation of NO$_3$ ($r_{IMFn}$ ranges from 2.6 to 3.7 at the sub-seasonal scale and reaches up to 13.8 for the 3-year cycles). Similar overestimation of NO$_3$ is also found at ATL ($r_{IMFn}$ ranges from 2.0 to 3.4, except for the 2-year cycles). In contrast, NO$_3$ at RENO is strongly underestimated with $r_{IMFn}$ ranging from 0.1 to 0.3 and reaching its minimum at the 2-year cycles. Likewise, all time scales of NH$_4$ at RENO are also being underestimated with $r_{IMFn}$ decreasing from 0.4 to only 0.1 at the 3-year cycles. The coexistence of underestimation of NO$_3$ and NH$_4$ variability, as well as their trend component, likely points to the insufficient grid resolution in representing ammonium nitrate episodes associated with stagnant meteorology in the mountainous regions as illustrated by Kelly et al. (2019). To sum up, model has simulated the magnitude of features across all scales in most of the studied cases. However, fluctuations in NO$_3$ are constantly being largely over- or under-estimated and improvements to the model are required to better replicate its variability (Fig. 4-11a-c).

A high $R_{max}$ of corresponding IMFs can only be achieved when the characteristic scales of those from observations and model simulations are close, there is minimal mode mixing, and negligible irregular
change of amplitude exists during the study period. Thus, $R_{max}$ tends to be small for all oscillations at RENO because of the irregular impact from events such as wildfires. Thus, the interpretation of phase shift is focused on the components and time scales having correlations above 0.4 only.

Results show that the sub-seasonal cycles at QURE all have a negligible phase shift of less than 0.1 cycles (Fig. 4-11d). The semi-annual cycles at RENO have around 0.2 cycle phase shifts in total PM$_{2.5}$ (-0.2), NH$_4$(0.2), OC (-0.2), and EC (-0.2) while negligible phase shifts of less than 0.1 cycles are simulated in SO$_4$ ranging from 9 days to semi-annual in scale. As at QURE, multiple sub-seasonal cycles at ATL all have a negligible phase shift of less than 0.1 cycles, with the exception of semi-annual OC which has a phase shift of nearly -0.4 cycles with a marginal correlation of around 0.4. Unlike the relatively stable $R_{max}$ throughout the time scales within each of the species for QURE and RENO, the $R_{max}$ at ATL tends to be much higher (roughly 0.6-0.8) in the scales of 6 to 25 days, except for NO$_3$, indicating the model’s success in simulating those weather-induced air quality fluctuations at this site as reflected by their negligible phase shifts.

However, the physical meaning of each sub-seasonal IMF is not yet fully understood and requires further study. Synoptic scale IMFs (IMFs with scale less than/around a month) usually have large variance and are not statistically significant different from white noise except for observed SO$_4$ and NH$_4$ (Fig. 4-S5). Yet, observed and simulated total and some speciated PM$_{2.5}$ at QURE and ATL (except IMF1) can achieve moderate to high $R_{max}$ at these time scales (Fig. 4-11 g-i), indicating a potential physical explanation of those time scales using meteorological variables. IMFs with scales longer than a month but less than half year possess much less variance and are usually not statistically significant different from noise. Exceptions are also found at the Atlanta site where observed IMFs are mostly significant different from noise. Whereas semi-annual cycles are mostly statistically significant (note that semi-annual SO$_4$ and NO$_3$ at ATL are too weak to be decomposed into a separate IMF). In a previous study, He et al. (2014) found semi-annual oscillations in the corrected AErosol RObotic NETwork (AERONET) Aerosol Optical Depth (AOD) and PM$_{10}$ mass concentrations are primarily caused by the change of wind directions in Hong Kong.
**Figure 4-11.** Model performance at all temporal scales for sites QURE, RENO and ATL. (a-c) ratio of mean amplitude of corresponding IMFs with similar characteristic mean periods (ideal ratio=1.0); (d-f) the phase shift $n$ in the number of mean periods (average mean period of corresponding IMFs decomposed from observation and model simulation); (g-i) maximum correlation $R_{max}$ can be achieved by shifting the modeled time series. The average mean period of corresponding IMFs decomposed from observations and CMAQ of total and speciated PM$_{2.5}$ are represented on the x-axis; all metrics on the y-axis are unitless. Horizontal reference lines are drawn at 0.7 and 1.3 in (a-c). Weekly, annual and inter-annual (2- to 3-year) scales are marked with vertical dashed lines.

The evaluation and interpretation of inter-annual cycles are constrained by the limited available speciated observations for a period of 6 to 9 years (4 years for Cl at QURE). Thus, only 2- to 3-year cycles are presented (Fig. 4-11) and evaluated. Among the 2- to 3-year inter-annual cycles at QURE, there is minimal phase shift for total PM$_{2.5}$, SO$_4$, Cl, and EC with moderate to high $R_{max}$. At RENO, the model
presents negligible shifts in 2-year cycles of OC and NH₄ while phase shifts of 0.3 and -0.5 cycles are simulated in the 3-year cycles for SO₄ and NH₄. At ATL, the phase shift of -0.2 to -0.4 cycles are simulated for PM₂.₅, NH₄, OC, and EC with periods of 2- to 3-year cycles; while 2- to 3-year SO₄ cycles have a half-year cycle shift.

4.5 Conclusions

The main advantage for using EMD to evaluate PM₂.₅ and its speciated components is that it decomposes nonlinear and nonstationary signals into multiple modes and a residual trend component. It does not require any preselection of the temporal scales and assumptions of linearity and stationarity for the data, thereby providing insights into time series of PM₂.₅ concentrations and its components. Using improved CEEMDAN, we are able to assess how well regional-scale air quality models like CMAQ can simulate the intrinsic time-dependent long-term trend and cyclic variations in daily average PM₂.₅ and its species. This type of coordinated decomposition and evaluation of total and speciated PM₂.₅ provides a unique opportunity for modelers to assess influences of each PM₂.₅ species to the total PM₂.₅ concentration in terms of time shifts for various temporal cycles and the magnitude of each component including the trend.

A demonstration of how improved CEEMDAN could be applied to time series data at three sites over CONUS that provide speciated PM₂.₅ data reveals the presence of the annual cycles in PM₂.₅ concentrations and time-dependent features in all decomposed components. At these three sites, the model generally is more capable of simulating the change rate in the trend component than the absolute magnitude of the long-term trend component. However, the magnitude of SO₄ trend components is well represented across all three sites. Also, the model reproduced the amplitude of the annual cycles for total PM₂.₅, SO₄ and OC. The phase difference in the annual cycles for total PM₂.₅, OC and EC reveal a shift of up to half-year, indicating the need for proper allocation of emissions and an updated treatment of organic aerosols compared to the earlier model version used in this set of model simulations. The consistent large under/over-prediction of NO₃ variability at all temporal scales and magnitude in the trend component, as
well as the abnormally low correlations of synoptic scale NO$_3$ at ATL, calls for better representation of nitrate partitioning and chemistry. Wildfires have the potential to elevate PM$_{2.5}$ for months and can alter its variability at scales from few days to the entire year. Thus, more accurate fire emission data should be incorporated to improve model simulation, especially in those fire-prone regions.
Additional Figures

Figure 4-S1. Evaluation of boundary effect on annual cycles and long-term trend with eighteen years of total PM$_{2.5}$ observation at North Little Rock, AR. (a) time series of total PM$_{2.5}$ (average data coverage: 97%; gaps are filled with most recent available observation); (b) annual cycles; (c) long-term trend; (d) boundary effect on annual cycles. Vertical colored lines in (a) indicate the 48 hypothetical boundaries by cutting the data before the left boundaries (light red) and after the right boundaries (light blue). In (b) and (c), annual cycles and long-term trend decomposed from the entire studied period are used as reference (blue) and those from the cropped time series are plotted in red. Specifically for the long-term trend, the summation of the residual and the last IMF ($IMF_l$) with longest temporal scale are shown in dashed blue for better comparison with the trend of the shorter cropped time series. The boundary effect in (d) is assessed by the statistics of the difference ($Diff = Ann\ cut - Reference$) in annual cycles as a function of number of days to the hypothetical boundaries: mean of all difference in black, standard deviation of all difference in red, standard deviation of those from the left/right boundaries in light red/blue as in (a). Difference in the level of standard deviation of left/right boundaries is very likely driven by the amplitude of the local annual cycles.
Figure. 4-S2. Time series of (a) annual cycles and (b) long-term trend decomposed from the same total PM$_{2.5}$ observation as in Fig. 4-S1, but for the period of 2009-2017 when average data coverage is almost 99%. Light gray lines indicate the components decomposed from PM$_{2.5}$ time series with data coverage of 100%, 90%…,10% (40 realizations in each data completeness scenario; in each realization, certain percentage of observations are randomly drawn from the original time series and filled with linear interpretation before the decomposition) from top to bottom. Red thick line represents the average of those 40 realizations in each scenario and thin dark blue lines indicates the average ± two times the standard deviation.
**Figure 4-S3.** The impact of data completeness on decomposed annual cycles and long-term trend. Time series of (a) averaged annual cycles and (b) averaged long-term trend of 40 realizations in each data completeness scenario as shown with red thick lines in Fig. 4-S2; standard deviation distribution (9*365 data points in each distribution) of (c) annual cycles and (d) long-term trend for different data completeness scenarios. Same color scheme applies to all subplots as shown in (c). Vertical dashed lines in (c) and (d) are their corresponding median value of the distribution. Note the perfect alignment of the components shown in (a) and (b).

**Figure 4-S4.** The peak of power spectrum of each IMF of observed and simulated total and speciated PM$_{2.5}$ for (a) QURE, (b) RENO and (c) ATL. IMF1 to the last pair of IMFs with increasing characteristic periods are shown from left to right. Species decomposed from observations are connected by solid lines, while species decomposed from simulations are represented by smaller markers in darker shades connected by dashed lines.
**Figure 4-S5.** Statistical significance test results of IMFs decomposed from observed (blue) and simulated (red) total and speciated PM$_{2.5}$ for (a-f) QURE, (g-l) RENO and (m-r) ATL following the method used in Wu et al. (2007). IMF1 to the last IMF with increasing characteristic periods are shown from left to right with filled circles representing statistically significant IMFs and plus sign indicating those not distinguishable from the corresponding IMFs of a pure white noise series. Vertical dashed black lines indicate the period of a month (left) or year (right). The colored solid lines indicate the variance of IMFs of white noise with its first IMF containing the same variance as that decomposed from observation (light blue) or model simulation (light red). The dashed lines above/below the solid line in the corresponding color are the upper/lower bound of a 95% confidence interval estimated by taking three times/one third of the variance of the white noise decomposed IMFs. Any IMFs above the upper bound is deemed statistically significant from those decomposed from white noise.
Chapter 5
A new method for assessing the efficacy of emission control strategies

5.1 Introduction

In the United States, regional-scale air quality models are used to design emission controls needed to comply with the ozone standard (i.e., the design value, defined as the 4th highest daily maximum 8-hour ozone concentration in each year averaged over a consecutive 3-year period, to not exceed a specified level). Year-to-year changes in ozone air quality over the contiguous United States (CONUS) are attributable to variations in meteorological conditions, local-to-hemispheric scale emissions loadings and factors such as stratospheric-tropospheric exchanges and climate change. In current regulatory applications, a regional air quality model is applied for a base year and a future year with reduced emissions using the same meteorological conditions as those in the base year. According to the current recommended modeling guidance, the observed base year design value is multiplied by the ratio of the average of the top 10 modeled ozone concentrations for the base and future years to assess whether the estimated future year design value meets the ozone standard (U.S. EPA, 2014). Hence, with the current methodology, we would not know how the variability in meteorological conditions affects the efficacy of emission reduction policies in assuring compliance with the ozone standard in the future year. Also, because the same meteorological conditions as those in the base year would not prevail in future years and observations for the future year are not available, the current attainment demonstration methodology can never be evaluated relative to observations.

Past studies have revealed that the absolute levels of ozone concentrations simulated by regional-scale air quality models can be subject to significant bias and errors (Schere et al., 2012; Hogrefe et al., 2014; Gilliam et al., 2015; Porter et al., 2015, 2017; Astitha et al., 2017; Emery et al., 2017; Henneman et al., 2017; Solazzo et al., 2017a, b). This is particularly true at the upper tail of the concentration distribution.
Astitha et al. (2017) evaluated the year-to-year changes in ozone air quality induced by variations in meteorology and emissions during 1990 to 2010 using spectral decomposition of both observed and WRF-CMAQ simulated daily maximum 8-hr ozone concentration time series with the objective of identifying the underlying forcing mechanisms that control ozone exceedances. Analysis of the information embedded in the daily maximum 8-hr ozone time series from the past few decades reveals that the long-term forcing, which is attributable to the emissions loading on local-to-hemispheric scales, climate change, variations in the stratospheric-tropospheric exchange processes, and the short-term forcing, which is attributable to synoptic-scale weather fluctuations, are to be viewed as the necessary and sufficient conditions, respectively, for controlling the ozone exceedances (Rao et al., 1996, 2011; Hogrefe et al., 2000; Porter et al., 2017; Astitha et al., 2017).

To help improve upon the current method for using regional air quality models in the regulatory context, we analyzed 34 years of ozone observations (1981-2014) having at least 80% data coverage over the contiguous United States. The main objectives of this study are to 1) develop a method for estimating the confidence bounds for the projected ozone design value (DV), and 2) demonstrate how future projections could be communicated in a probabilistic framework that explicitly accounts for the prediction uncertainties stemming from the year-to-year variability in meteorology. We examine the variability in observed extreme values stemming from changes in emissions and meteorology during the 1981-2014 time period and present a new method for building confidence when regional-scale air quality models are used for policy support. This new probabilistic method for assessing the efficacy of emission control strategies and ozone data employed are described in Section 2, and evaluated in Section 3. Attainment demonstration in a real-world application using the suggested method and the CMAQ model is discussed in Section 4. Summary and conclusions are presented in Section 5.
5.2 Data and methods

5.2.1 Ozone measurements and CMAQ simulations

Ground-level daily maximum 8-hr ozone concentration (DM8HR) observations during May to September covering the contiguous United States (CONUS) were obtained from the U.S. Environmental Protection Agency’s (EPA) Air Quality System (AQS). A valid ozone season consists of at least 80% data coverage during May to September at each station. A total of 156 stations with at least 30 valid years (to provide enough synoptic conditions, noted as 30+ in this paper) from 1981 to 2014 were analyzed.

Two sets of WRF/CMAQ model simulations covering the CONUS for the year 2010 were employed to demonstrate this new approach from regulatory perspective. These simulations were performed as part of the third phase of the Air Quality Model Evaluation International Initiative (AQMEII3). The first simulation, hereafter referred to as BASE, uses meteorology, emissions, and lateral boundary conditions representative of 2010 conditions, whereas the second simulation, hereafter referred to GLO, reflects to a 20% reduction of anthropogenic emissions loading both in the global model simulations providing lateral boundary conditions and within the WRF/CMAQ modeling domain while using the same 2010 meteorology as in BASE. A detailed description of the AQMEII3 modeling scenarios can be found in Galmarini et al. (2017). Further details on the configuration of the WRF/CMAQ system can be found in Hogrefe et al. (2018). Model evaluation results for the BASE simulation can be found in Solazzo et al. (2017a, b) and Hogrefe et al. (2018). These two simulations are representative of how an air quality model is used by the regulatory agencies for attainment demonstration.

5.2.2 A Kolmogorov-Zurbenko (KZ) filter

In this study, we examine the year-to-year variability in the relative strengths of the short-term component induced by synoptic-scale weather fluctuations (SY) and the magnitude of the long-term component (BL) induced by changes in anthropogenic emissions and non-anthropogenic background ozone, and other longer-term factors embedded in observed ozone time series during 1981-2014. This is
accomplished by decomposing the May-September DM8HR with the Kolmogorov-Zurbenko (KZ) low-pass filter, a commonly-used filter to separate various temporal signals (Rao and Zurbenko, 1994; Rao et al., 1995, 1997; Eskridge et al., 1997; Milanchus et al., 1998; Gardner and Dorling, 2000; Ibarra-Berastegi et al., 2001; Hogrefe et al., 2003; Vukovich and Sherwell, 2003; Maxwell-Meier and Chang, 2005; Wise and Comrie, 2005a, b; Yang and Zurbenko, 2010; Galmarini et al., 2013; Kang et al., 2013; Sá et al., 2015; Phalitnonkiat et al., 2016; Zhang et al., 2018; Zurbenko and Smith, 2018).

The KZ\(m,k\) filter is defined as \(k\) applications of a simple moving average of \(m\) points. The moving average can be defined as follows:

\[
Y(t) = \frac{1}{m} \sum_{s=-\frac{m-1}{2}}^{\frac{m-1}{2}} X(t + s)
\]  

(5 – 1)

where time series \(Y(t)\) is the long-term component of time series \(X(t)\), \(m\) is the window length of the filter. The process is repeated \(k\) times by using \(Y(t)\) in the previous step as \(X(t)\) in the next iteration. The window length \((m)\) and the number of iterations \((k)\) together determine the scale separation between the frequencies present in the long-term component and those present in the residual short-term component (original time series – final longer-term component). An analytical representation of the filter transfer function that can be used to calculate the separation frequency for a given combination of \(m\) and \(k\) can be found in Rao et al. (1997) and a comparison against other spectral decomposition approaches is provided in Hogrefe et al. (2003).

The 50% cutoff frequency of the KZ filter, meaning the energy at this frequency is separated half-half by the KZ\((m,k)\), can be expressed as (see Rao et al., 1997 for details on filter characteristics):

\[
\omega_0 \approx \frac{\sqrt{6}}{\pi} \sqrt{\frac{1 - 0.5\frac{1}{k}}{m^2 - 0.5\frac{1}{k}}}
\]  

(5 – 2)

The parameters of KZ filter are selected based on the scale of the synoptic events, which is 2-21 days (Eskridge et al., 1997; Hogrefe et al., 2001). Slow variations with period of longer than about 3 weeks
in the signal should go into BL component. In this paper, we use KZ(5, 5) with window length of 5 days and 5 iterations to decompose May-September time series of observed and simulated daily maximum 8-hr ozone concentrations into the short-term synoptic (SY) and long-term baseline (BL) components for each year (Porter et al., 2015; Hogrefe et al., 2000; Astitha et al., 2017). The 50% cutoff frequency for KZ(5,5) is 0.0411, which gives a cutoff period for daily sampling of 1/0.0411≈24 days. The decomposition can be expressed as:

\[ BL(t) = KZ(5,5) \]

\[ SY(t) = O_3(t) - KZ(5,5) \]

where \( t = 1, 2, \ldots, 153 \) stands for the days from May 1st to September 30th, \( O_3(t) \) is the original time series of observed DM8HR ozone concentrations (the orange \( O_3 \) in Fig. 5-2a for instance), \( BL(t) \) is the baseline or long-term component (the red BL in Fig. 5-2a) and \( SY(t) \) is the residual short-term component.

An example of the spectral decomposition of 2014 DM8HR at Stratford, CT (Fig. 5-1a) and the power spectrum (Fig. 5-1b) of corresponding time series illustrated below reveals that there is good separation between the two forcings at around 24 days with some leakage of information presented. Since we are dealing with non-linear and non-stationary ozone time series, some leakage of information between the two separated components is expected.

**Figure 5-1.** KZ filter decomposition of observed ozone concentrations for Stratford, CT. (a) Ozone time series (\( O_3 \)), baseline component (BL) and synoptic component (SY) in 2014. (b) Power spectrum of each component and the ozone concentration that designates the time period (in days) when the two components separate.
5.2.3 Ozone time series reconstruction and the importance of the baseline component

Ozone time series reconstruction is accomplished by adding the synoptic and baseline components (Eq. 5-5), a reverse process of the KZ decomposition described in the previous section. The ozone baseline component represents the ozone pollution that is created primarily by emission sources whereas the prevailing synoptic-scale meteorology controlling the regional-scale transport of ozone and its precursors is reflected in the synoptic component.

\[ O_3(t) = SY(t) + BL(t) \] (5 – 5)

A linear relationship between the strength of the synoptic forcing, defined as the standard deviation of the weather-induced variations in ozone time series data (SYstd) and the mean of the baseline concentration (BLmean) has been found at each monitoring site in CONUS; an example of this linear relationship between BLmean and SYstd is illustrated in Fig. 5-S1.

\[ SYstd = f(BLmean) = p \times BLmean + q \] (5 – 6)

Given the high correlation between the strength of the synoptic forcing and the magnitude of the baseline concentration (Porter et al., 2017; Astitha et al., 2017), each observed synoptic component SY(t) is adjusted according to the mean of a target BL using the relationship below before being superimposed (for example, if we want to reconstruct O₃ concentration for 2010 (target) with the synoptic component of year 1998 (y), we need to adjust the synoptic component with the BL of 2010):

\[ SY_{y,adj}(t) = SY_y(t) \times \frac{f(BLmean_{target})}{f(BLmean_y)} \] (5 – 7)

in which y can be 1981,1982…2014, indicating the year of which the synoptic forcing is decomposed from (e.g. 1998 in the example given above). The target BL can be any baseline of interest (e.g. BL for the year 2010 in the example given above). The daily ozone concentration is then reconstructed by:

\[ O_3^{Y,target}(t) = BL_{target}(t) + SY_{y,adj}(t) \] (5 – 8)

An example of the decomposition and reconstruction is shown in Fig. 5-2, where the observed DM8HR time series from May through September in 2002 and 2014, along with the embedded baseline
concentration in Stratford, CT are illustrated. There has been a substantial improvement in ozone air quality between 2002 and 2014, likely due to the implementation of emission control programs on the electric generation units and motor vehicles that helped significantly reduce emissions of NOx and VOCs, given that the meteorological conditions including daily precipitation, average wind speed, mean and maximum temperature between 2002 and 2014 are not significantly different. This substantial improvement is indicated by the reduction in the 4th highest ozone concentration from 103 ppb in 2002 to 74 ppb in 2014 and a reduction in the number of days exceeding the 2008 NAAQS standard (75 ppb) from 27 in 2002 to only 3 in 2014 (Fig. 5-2ab). The large reduction in the baseline concentration as well as the temporal variability of the baseline concentration within the same year is evident.

To illustrate the relative importance of changes in the baseline vs. variability in the strength of the synoptic forcing, we reconstructed the ozone time series by adding the adjusted 2014 synoptic forcing to the 2002 baseline forcing (Fig. 5-2c) and vice versa (Fig. 5-2d). Combining the higher baseline in 2002 with the synoptic forcing that prevailed in 2014, we see that the 4\textsuperscript{th} highest ozone changed from 103 to 95 ppb and the number of exceedance days changed from 27 to 23, whereas combining the lower baseline in 2014 with the 2002 synoptic forcing led to 80 ppb (instead of 74 ppb) for the 4\textsuperscript{th} highest ozone and changed the number of exceedances from 3 to 4. Thus, the increase seen in the number of ozone exceedances is primarily influenced by the baseline level and not by changes in the synoptic forcing. Similar results were found when different years and stations were used without the adjustment for the synoptic component (see Rao et al., 2011).
Figure 5-2. Observed ozone time series (O$_3$) and long-term forcing (BL) along with the 4t highest ozone concentration and number of ozone exceedances at Stratford, CT during May-September: (a) 2002; (b) 2014; (c) reconstruction of O$_3$ using 2002 BL and adjusted 2014 synoptic forcing (SY); (d) reconstruction of O$_3$ using 2014 BL and adjusted 2002 SY. Reconstructed ozone O$_3$ by superimposing different synoptic forcings from 34 years on the same baseline concentration in: (e) 2002 (95% confidence bounds of the 4$^{th}$ highest: 94-114 ppb, actual=103 ppb); (f) 2014 (95% confidence bounds of the 4$^{th}$ highest: 73-84 ppb, actual=74 ppb). Crosses in (e-f) denote the 4$^{th}$ highest ozone concentrations with the 34 different synoptic conditions and the solid straight black line the observed 4$^{th}$ highest ozone concentration with respect to the baseline year.

The above conclusions are further supported if we expand the experiment to superimpose 34 years of synoptic forcings on the baselines that prevailed in 2002 and 2014 for the same Stratford, CT station.
(Fig. 5-2ef). It is evident that the 4th highest ozone concentrations associated with a given baseline (BL) can vary depending upon the prevailing synoptic forcing (see crosses in Fig. 5-2e and f denoting the 4th highest for each “reconstructed” year), but these variations are much smaller than the differences seen between the magnitudes of baselines in 2002 and 2014. Moreover, the position of the crosses in these time series (Fig. 5-2ef) also indicates that the 4th highest values associated with 30+ years of synoptic forcing tend to cluster together in time and coincide with times when the baseline level reaches a local maximum, further emphasizing the notion that elevated baseline levels are necessary for observing ozone exceedances.

Reconstruction of ozone time series using the observed BL in 2002 and 34 years of SY (Fig. 5-2e) yields 34 values of the 4th highest ozone concentrations, one from each reconstructed ozone time series (Eq. 5-8). This allows the determination of the 95% confidence bound for the 4th highest (CBs, defined by the 2.5th and 97.5th percentile values of the distribution; more details are provided in Section 2.4) as 94-114 ppb with the actual observed being 103 ppb; similarly, the 95% CBs for the number of exceedance days are 19-31 with 27 being observed. Likewise, 95% CBs for the 4th highest ozone concentrations with the 2014 BL (Fig. 5-2f) is 73-86 ppb with the observed being 74 ppb and the number of exceedances is 2-10 with the actual being 3. Comparing the influence of the two baseline forcings, it is evident that the baseline with the higher magnitude in 2002 (Fig. 5-2e) yields higher 4th highest ozone concentrations than those in 2014 (Fig. 5-2ef) with 34 years of possible synoptic forcings. Analysis like that presented in Fig. 5-2 is given for four additional sites in the supplemental material (Fig. 5-S2–5-S5) and the conclusions are identical to those discussed for Stratford, CT. This analysis demonstrates that change in the baseline level induced by emission reduction rather than differences in the meteorology is the main driver for the improvements seen in ozone air quality between 2002 and 2014. The above results indicate that the magnitude of the baseline concentration dictates the peak ozone levels that can be reached for a given synoptic forcing, further validating the conclusions reached in previous studies (Rao et al., 2011; Astitha et al., 2017; Porter et al., 2017).
5.2.4 Estimating confidence bounds

It should be noted that the 4\textsuperscript{th} highest concentration (or any other extreme value) in any given year is a single event out of a population; that is, even if emissions were to remain the same, stochastic variations and meteorological conditions in a different year could lead to a different 4\textsuperscript{th} highest ozone concentration. From a statistical perspective, an observation on any given day represents a single realization out of an unknown underlying distribution. Therefore, it is important to characterize and quantify the distribution of the possible future 4\textsuperscript{th} highest ozone concentrations and design values associated with each emissions loading scenario, instead of relying on a single value. Once the distribution of the 4\textsuperscript{th} order statistic is determined, it is possible to estimate the 95\% confidence bounds for the 4\textsuperscript{th} highest ozone concentration. This can be done using various methods: 1) extreme value theory (EVT) based on its parent DM8HR ozone time series, 2) \( \mu \) (mean) \( \pm 2\sigma \) (standard deviation) provided that those 30+ 4\textsuperscript{th} highest values are normally distributed or 3) the 2.5\textsuperscript{th} percentile and 97.5\textsuperscript{th} percentile values of the variable distribution based on bootstrapping.

The first approach, the extreme values theory (EVT), provides the theoretical framework to estimate the distribution of any extreme values of interest, say, the 4\textsuperscript{th} highest value for ground level ozone for this study, from the underlying DM8HR ozone time series (Gumbel, 1958; Roberts, 1979; David, 1981; Chock, 1984; Rao et al., 1985; Hogrefe and Rao, 2001). However, the application of the exact theory of extreme values relies on the knowledge of the analytical form of the underlying DM8HR ozone distribution, which is generally unclear (Gumbel, 1958; Roberts, 1979; David, 1981; Chock, 1984; Rao et al., 1985; Hogrefe and Rao, 2001). It has been shown that the analytical cumulative distribution function (CDF) of the upper tail of most parent distributions can be described by an exponential distribution:

\[
F(x) = 1 - e^{-\lambda(x-B)}
\]  \hspace{1cm} (5 – 9)

where \( \lambda \) and \( B \) are the shape and location parameters, respectively,

\[
\lambda = \left[ \frac{1}{n-1} \times \sum_{i=1}^{n} (x_{(i)} - x_{(n)}) \right]^{-1}
\]  \hspace{1cm} (5 – 10)
\[ B = x_{(n)} \]  

(5 – 11)

in which, \( x \) is the upper 20\% of the ozone concentration distribution at any given site in one specific year and \( n \) is the number of observations in \( x \) (\( n \) ranges from 24 to 31 years in this study, which is 20\% of valid days during 153 days of May - September); \( x(1), x(2), \ldots, x(n) \) are the highest, second-highest, \ldots, smallest of the \( n \) tail values, respectively.

The exact theory of extreme values can then be applied to estimate the CDF of the 4\textsuperscript{th} order statistic (i.e., the 4\textsuperscript{th} highest ozone concentration) from the following expression:

\[
G(x) = \sum_{j=0}^{3} \binom{n}{j} \times [1 - F(x)]^j \times [F(x)]^{n-j}
\]

(5 – 12)

The 95\% confidence bounds for the 4\textsuperscript{th} highest ozone is determined by the 2.5\textsuperscript{th} and 97.5\textsuperscript{th} percentiles from the CDF \( G(x) \).

The first method of calculating the confidence bounds using the EVT, is not applicable to the design value since it requires averaging the 4\textsuperscript{th} highest ozone value over a consecutive 3-year period. However, it can serve to validate the confidence bounds derived from historical ozone observations described in Section 3.1. The second method (mean ±2 times the standard deviation) requires the 4\textsuperscript{th} highest ozone concentrations to be normally distributed, which may not be valid in all cases. Therefore, we utilize the bootstrap method (i.e., resampling with replacement) to develop the distribution of design values by randomly drawing the 30+ synoptic conditions 2000 times to represent the stochastic variation of synoptic conditions among the 3 consecutive years and utilize the same manner for 4\textsuperscript{th} highest (the 2000 random samples are to be considered as bootstrap estimates).

5.2.5 Probabilistic assessment of emission control strategies using the Baseline Projection Method based only on observations

In the following, we discuss the development and details of the proposed baseline projection methodology, described step by step in Fig. 5-3, to assess ozone extreme values for a given emissions
loading scenario under varying meteorological conditions based on 30+ years of ozone observations in CONUS. As noted before, time series of the DM8HR ozone concentrations during May to September at each site are spectrally decomposed into the baseline and synoptic components with the KZ filter for all years as in Fig. 5-1a.

**Figure 5-3.** Representation of the proposed probabilistic assessment of emission control strategies with the Baseline Projection Method. BL=baseline forcing; base=base year; future=future year; SY=synoptic forcing. Example of a 12-yr projection is given (2002-2014). Detailed description of each step is given in the text (section 5.2.5).

Given the importance of the baseline level discussed in Section 2.3, we propose to use the Baseline Projection Method to construct the baseline component for the future by projecting the base year observed daily BL onto a future year by performing linear regression of rank-ordered base and future year baseline concentrations. In Step 1 of Fig. 5-3, \( a \) and \( b \) are slope and intercept of the linear regression between the rank-ordered base year’s and future year’s BL. In real-world applications (i.e., ozone attainment demonstration) where future observations are not available, \( a \) and \( b \) are determined from the CMAQ model simulations of base year’s and emissions control scenario’s BL. BL\(_{\text{future,proj}}\) is the projected future BL from...
the observed base year’s BL (BL_{base,obs}). As an example, baseline linear regression for Stratford is shown (observed baselines are used here to demonstrate the method) in Fig. 5-4a and the 2014 projected baseline component is the green solid line in Fig. 5-4b, which is projected from the observed 2002 baseline as in Fig. 5-2a. This process is repeated for two additional pairs of base and future years to obtain a total of three years of projected BL time series given the need of 3-year averaging for the calculation of the design value.

Regression on rank-ordered ozone baseline concentration is necessary when we use only observations in Step 1 since it helps to eliminate the influence of changing meteorological conditions imbedded in observations in base and future years as well as the weekday-weekend differences and holidays between the two years; high or low ozone concentrations tend to occur under similar meteorological conditions for any given year, thereby permitting examination of the change in the ozone baseline induced predominantly by changing emissions. Since model projections use the same meteorology as in the base year, adding the difference between the base and future year baselines to the base year BL for each day works equally well.

In Step 2, ozone time series are reconstructed by adding the projected BL to the 30+ historically observed synoptic components adjusted for the specific baseline level, providing sufficient information to develop the possible distribution of the 4th highest and the design value (DV). That way, the distribution of the 4th highest and the DVs are generated for the future year by 2000 bootstrap replications to represent the stochastic variability in synoptic forcing. In Step 2 of Fig. 5-3, ¥_{yi,adj} represents the BL_{future,proj} adjusted SY from year $y$ ($y^{th}$ year in 1981-2014) and $O_3^{yi,future}$ is the corresponding reconstructed DM8HR time series; similar notations are used for DV and 4th highest ozone. The possible 4th highest values from the multiple reconstructed time series in 2014 for Stratford are marked as crosses in Fig. 5-4b.

Once the distribution of the 4th highest ozone or the design value is estimated, the confidence bounds (CBs) and probability of the 4th highest ozone or the design value exceeding a given threshold (Pex), say, the 2008 NAAQS standard pf 75 ppb, can be estimated for a given year or emissions loading scenario. In Step 3 of Fig. 5-3, CBs stand for 95% confidence bounds and Pex is the probability of DV exceeding the
NAAQS ozone standard. Consequently, CBs for ozone exceedances for 2014 at Stratford using the Baseline Projection Methodology (observations-based analysis only) is 73-88 ppb for the 4th highest and 3-10 for the number of days with exceedances (Fig. 5-4b), well-resembling the results when the observed baseline is used (Fig. 5-2f). Again, the actual ozone measurements fall within the estimated confidence bounds (actual 4th highest 74 ppb, actual exceedance days 3).

This exercise illustrates the methodology for the BL projection and probabilistic approach for predicting the 4th highest ozone concentration and number of ozone exceedances. A more comprehensive evaluation of the proposed method will be presented in the following section. It should be noted that since the time series of DM8HR for a future year is reconstructed, parameters such as median DM8HR can also be derived in a similar manner.

**Figure 5-4.** Baseline projection from 2002 to 2014 at Stratford, CT. (a) Rank-ordered observed baseline regression between base year 2002 and future year 2014 ($a=0.47$, $b=22.51$); (b) Reconstructed ozone for 2014 by using the projected 2014 BL from the 2002 BL based on their regression relationship in (a) and 34 years of SY forcings. Crosses denote the 4th highest ozone concentrations during the 34 synoptic years and the straight black line the observed 4th highest ozone concentration in 2014.

### 5.3 Results and discussion

The use of multi-year ozone observations rather than model simulations in this section demonstrates the appropriateness of the proposed method without having to delve into the impacts of model performance in reproducing the BL concentration and changes in the BL stemming from emissions reductions.
Furthermore, it lends credence to the proposed method since observation-based analysis can be validated. This enables comparison of the estimated confidence bounds for 2007 to 2014 projections against actual observed ozone design values. We selected projections that span at least 5 years to allow sufficient time for changes in emissions from emission control policies to become apparent. The main reason that the projections are focused from 2000 onward, is that the 4th highest observed ozone trends in the earlier decade (namely, 1990-2000) have been shown to be not statistically significant (Astitha et al., 2017), whereas the ozone trends during 2000 to 2010 are statistically significant and we wanted to demonstrate the BL projection method in a period covering significant changes in ozone exceedances driven primarily by emission changes.

5.3.1 Width of the 95% confidence bound

The exact theory of extreme values (EVT) provides the theoretical framework to examine the appropriateness of the proposed method to estimate confidence bounds of 4th highest values. The EVT method estimates the confidence bounds of the extreme value (4th highest in this application) directly from the future observed DM8HR time series, accounting for its statistical property. Thus, it is treated here as the ground truth. These two CBs (from EVT and the proposed BL method) are not expected to be identical since they are based on different mathematical approaches and statistical assumptions; nevertheless, they should be of similar magnitude.

The width of CBs for the 4th highest ozone calculated from the proposed method over all stations in CONUS is very similar to that estimated by the EVT (Fig. 5-5; box plots show the spatial distribution of the CBs width). The range of 10-15 ppb derived from the proposed method is consistent with the magnitude of the width of CBs derived from the EVT method. In addition, there is a capability to project into the future year using a variety of base years since the estimated CBs for a specific future year from multiple different base years, with different projection intervals, yield very similar distributions (Fig. 5-5). When more years
of historical observations than 34 considered here are included, we expect to see even better agreement between the widths of the CBs derived from the proposed method and EVT.

**Figure 5-5.** Box plots of 95% Confidence Bounds (CBs) width for the 4th highest (in ppb) for multiple future projections (≥5yrs) and all stations (colored boxes denote different projection intervals in years). CBs width from the exact theory of extreme values (EVT) is shown with the grayscale background boxplots. Boxes are marked at 25th, 50th, 75th percentile. Maximum whisker length equals 1.5 times the inter quantile range. Topmost outliers of CBs width in all projections come from one station at Pomona, CA.

5.3.2 Are the observed Design Values within the estimated Confidence Bounds?

It is crucial for the “future year” observed design value to fall within the predicted 95% confidence bounds for a valid projection method. As shown in Fig. 5-6, a clear majority (92–99% of the sites) of the observed DVs fall within the estimated confidence bounds. There are few stations where the observed DV does not fall within the confidence bounds and those vary depending on the projected year. More than half of the unsuccessful cases are the result of rounding up of the DV and CB values, with an actual decimal bias of less than one ppb. Other than that, these cases generally occur near large water bodies for specific future years when few local extreme events elevated the observed DV remarkably, leading to an
underestimation of the CBs. Although we accounted for data completeness by selecting only stations with at least 80% date coverage per ozone season, missing data still play a noticeable role and is one of the reasons to miss the observed DV at some stations. The success rates for all possible projections except the ones shown here are also in the 92-99% range. This high success rate, together with the fact that the widths of the confidence bounds for the 4th highest values are similar to those derived from the exact theory of extreme values, provides confidence regarding the reliability and appropriateness of the proposed methodology for the ozone attainment demonstration that would be based upon air quality modeling simulations in real-world future applications.

**Figure 5-6.** Maps indicating whether the observed DV falls within the 95% confidence bounds calculated using the 5-yr projection for (a-h) 2007-2014. The colors in each subplot denote the observed DV relative to the estimated CBs at each station: Green circle=within the CBs; red triangle= outside the CBs (observed DV is higher than estimated CBs); red inverted triangle=outside the CBs (observed DV is lower than estimated CBs).
5.3.3 Probabilistic assessment of ozone exceedances

An important feature of the proposed method is its ability to provide probabilistic assessments of future exceedances of given ozone threshold concentration (National Ambient Air Quality Standards, NAAQS). This is possible through the construction of a potential DV (or 4\textsuperscript{th} highest) distribution dictated by 30+ years of observed synoptic forcings influencing a given baseline ozone concentration (long-term forcing). The probability of non-attainment of the ozone standard is estimated by dividing the total projected DVs that are greater than the threshold by the size of sample pool used to formulate the DV distribution (2000 in the proposed approach, Fig. 5-3). The result is illustrated in Fig. 5-7, where the actual ozone attainment (DV>75 ppb) in the binary form for 2014 (Fig. 5-7a) is contrasted against the probability of the DV exceeding the NAAQS, \( P_{ex}(DV>75 \text{ ppb}) \) constructed with the actual observed 2014 baseline and the baseline projected from 2002 into 2014 (Fig. 5-7bc). The probability of the DV exceeding the NAAQS is a good indicator of the attainment status and conveys additional information for stations when DV is close to the level of the NAAQS (Fig. 5-7ab). The two maps with either the actual observed 2014 baseline (Fig. 5-7b) or the projected 2014 baseline (Fig. 5-7c) convey almost identical situations, validating the proposed BL projection methodology.

![Figure 5-7. Attainment determination. (a) Binary representation of ozone attainment status of observed 2014 DV (red=non-attainment, blue=attainment). (b) Probability of 2014 DV exceeding the ozone standard using the actual observed baseline. (c) Probability of projected 2014 DV exceeding the ozone standard using the proposed method over the 12-yr projection interval.](image-url)
Information on the probability of exceedance for each station would allow decision makers to explore the effects of selecting different cut-off probabilities for reaching attainment across the domain. Having the option of a probabilistic assessment allows the policy-makers to decide upon emissions control strategies informed by the range of possible attainment/non-attainment scenarios to choose from. This is discussed in more detail in the following section where the attainment demonstration procedure is illustrated for a real-world future year projection scenario.

5.4 Attainment demonstration using regional air quality modeling simulations

The probabilistic assessment using the baseline projection method has been illustrated and evaluated in the previous sections. The proposed method is adopted for regional air quality models, specifically the CMAQ model, in this section to help develop robust ozone control strategies. The current ozone NAAQS attainment demonstration method uses the ratio of the mean of the top 10 modeled ozone concentrations in the emission control case to those in the base case simulated by a regional air quality model (U.S. EPA, 2014). The observed base year design value (defined as three consecutive DVs, by considering weighted-average of five years 4th highest) is then multiplied by this ratio to determine whether that emission control case would reduce the design value to the level of the ozone standard. In those cases when the projected future design value is close to the NAAQS, more rigorous Weight of Evidence Analysis should be completed, such as supplemental trend analysis, diagnostic analysis and implementation of additional model simulations (Galmarini et al., 2013; Solazzo and Galmarini, 2016; U.S. EPA, 2014). If non-attainment is determined based on the current year design value, increasing emissions reduction cases are modeled until the design value is at or below the level of the standard. Because the meteorology used for the base year would never prevail for future emission control years, the impacts of emission reductions on extreme values of interest cannot be predicted perfectly even if the model and its input data were perfect. Hence, there is no guarantee that the envisioned emission control strategy will in fact lead to ozone compliance in future years with the current method. Therefore, we present an alternative approach with
emission reduction scenarios are accompanied by confidence bounds and probabilities of exceedance that account for the ever-present uncertainties in the numerical prediction of atmospheric processes and the inter-annual variability in meteorology when regional-scale air quality models are used.

Taking the baseline component of the ozone time series that is representative of the long-term variation dictated by emissions, policy changes, background, and trends, we apply the Baseline Projection Methodology in the context of attainment demonstration. The regression coefficients $a$ and $b$ of the BASE-GLO projection are estimated using a rank-ordered regression of the baseline between the BASE CMAQ simulation for 2010 as the base year and the GLO scenario simulation with 20% global reduction of anthropogenic emissions as the future year. Then, the same set of regression coefficients is employed to estimate the projected BL from observed BL in 2008, 2009 and 2010 (Step 1, Fig. 5-3; BL). Following Steps 2 and 3 in the methodology (see Fig. 5-3), we determine the distribution of the DV. The model simulations are used only to project the baseline making use of the linear relationship between the baseline in the base and the projected year anchored to the observed baseline. The observed DV in 2010 (Fig. 5-8a) shows widespread non-attainment of the 2015 ozone NAAQS of 70 ppb with some extreme high values at sites of southern California. The median DV for the BASE-GLO projection (Fig. 5-8c) shows a decrease of 3 ppb across all stations. This indicates the impact in the future DV across CONUS stemming from the 20% reduction in anthropogenic emissions. Whether this reduction is sufficient or not for attainment is determined from the probabilistic viewpoint of the 2015 NAAQS exceedances. With the median DV across all stations (71 ppb) above the 2015 NAAQS in the base year 2010, the chance of the median DV exceeding 70 ppb is reduced from 68% to only 20% (Fig. 5-8 b, d). It is apparent that this reduction is still not adequate for some areas, especially some California and Northeastern stations, to reach attainment (Fig. 5-8cd). This BASE-GLO emission control scenario has a BL component reduction of around 5% nationwide with more prominent impacts in the Eastern US and central coastal area of California (Fig. 5-8e shows the reduction of the BL component from BASE to GLO).
Figure 5-8. Attainment demonstration with two emission control scenarios compared with base year conditions. **Base case:** (a) observed DV in base year 2010; (b) Pex (Probability of DV exceeding the 2015 ozone NAAQS of 70 ppb) with observed base year baselines. **BASE-GLO emission control scenario:** (c) median DV of the projected DV distribution; (d) Pex; (e) Percentage reduction of baseline mean from BASE to GLO. **5% BL reduction control scenario:** (f) median DV of the projected DV distribution; (g) Pex.

When we apply a hypothetical 5% BL reduction scenario across the board (instead of using the model simulation reflecting the GLO scenario), the results are very similar to the BASE-GLO projection (Fig. 5-8fg). This result indicates that a 5% BL reduction is approximately equivalent to a 20% anthropogenic emissions reduction GLO scenario. Given that the 5% BL reduction stemming from 20% emission reduction (i.e., GLO projection) is not adequate for attainment at many sites in California and NE
US, we apply a hypothetical emission reduction scenario yielding in 15% baseline reduction in 2010 from some potential VOC and NOx emission control strategies. This reflects a stricter emission control scenario compared to the BASE-GLO case and the decrease in DV is more pronounced nationwide (Fig. 5-9). The probabilities of exceeding the NAAQS (Pex) are diminished for this stricter emission control scenario, except for some areas in California which still face non-attainment (see Fig. 5-9b).

**Figure 5-9.** Attainment demonstration with the 15% baseline reduction scenario: (a) median DV of the projected DV distribution and (b) Probability of DV exceeding the 2015 ozone NAAQS of 70 ppb.

Experimenting with various emission control scenarios that can be different for each region or state using regional air quality models by controlling sectors such as on-road NOx and industrial VOC provides a probabilistic assessment of non-attainment that is much more informative and robust than providing a simple “yes/no” answer on the efficacy of a control strategy to policy-makers. In other words, the proposed methodology enables the policy-makers to assess the probability of success in achieving the intended target with each envisioned emission control strategy. Ultimately, decisions pertaining to the selection of emission control strategies and the probability threshold for exceeding the ozone standard are the prerogatives of policy-makers.
5.5 Summary and Conclusions

Prediction of the absolute pollutant concentration levels and changes in the peak ozone concentration values is challenging for regional air quality models because of uncertainties in input data as well as model physics and chemistry. Given the strong linkage between the magnitude of the baseline concentration level and peak ozone values, we developed a new method, namely the BL Projection, for estimating the associated 4th highest ozone, the design value and number of ozone exceedances and determine their confidence bounds by superimposing multi-decadal historical synoptic forcings acting on the prevailing BL concentration. This is demonstrated using 34-years of observations to establish the validity and robustness of the proposed methodology. As additional data become available in the future, the database for the synoptic forcings will expand beyond the 34 years used in this study, thereby facilitating more reliable and robust policy decisions on emissions management.

The use of the CMAQ model with the BL regression methodology is demonstrated using two simulations that share the same meteorological conditions (base case simulation for 2010 and 20% global anthropogenic emission reduction). Results indicate that the global 20% anthropogenic emission reduction yield a baseline change of only 5% across CONUS and is not adequate in some areas, especially stations in California and northeastern US. Thus, regulatory modeling assessments that focus on the impact of emission changes on the baseline ozone concentration level instead of on few high values of the ozone distribution would provide greater confidence in policy-making. Therefore, it is important that regional air quality models accurately simulate the baseline concentration and its changes stemming from emission reductions because it is the baseline component that strongly influences peak ozone levels or ozone exceedances.

The key results from this investigation are summarized below:

- The efficacy of emissions reductions in achieving ozone compliance is better assessed by focusing on the longer-term forcing, namely, BL concentration levels, rather than on the peak ozone concentration
levels. The BL Projection Method is easy to implement in the regulatory framework, thereby helping build confidence in the use of air quality models in the attainment demonstration process.

- The change in the long-term forcing from the base year stemming from a given emission reduction scenario when combined with the synoptic-scale weather fluctuations embedded in time series of historical ozone observations enables us to estimate the probability of exceeding the ozone standard associated with future emission reduction scenarios.

- Modeling results analyzed and presented in this probabilistic manner enable more explicit consideration of the ever-present uncertainty in projected changes in air quality needed to comply with the ozone standard.
Figure 5-S1. Relationship of the strength of synoptic forcing (standard deviation of SY component) and baseline mean (BLmean) at Stratford, CT. Crosses denote pairs of observed SYstd and BLmean for a specific year during 1981-2014.
**Figure 5-S2.** Same as Fig. 5-2 but for station in Bakersfield, CA.
Figure 5-S3. Same as Fig. 5-2 but for station in Atlanta, GA.
Figure 5-S4. Same as Fig. 5-2 but for station in Cary, IL.
Figure 5-S5. Same as Fig. 5-2 but for station in Fort Worth, TX.
Chapter 6
On the manageable portion of the ozone non-attainment problem in the contiguous United States

6.1 Introduction

Ground-level ozone is a widespread secondary air pollutant that has been linked to harmful impacts on human health, ecosystem and a variety of materials. It is formed in the presence of sunlight involving chemical reactions of Nitrogen Oxides (NOx, including NO and NO₂) and Volatile Organic Compounds (VOC). According to the 2014 National Emissions Inventory (NEI), biogenic VOC emissions account for 70% of the total VOC burden in the U.S. and the remainder originate from mobile and stationary sources; NOx is mostly anthropogenically emitted (> 93%) by mobile and stationary sources such as electricity generation, industrial processes, and boilers (U.S. EPA, 2017). Moreover, processes such as stratospheric-tropospheric intrusion and intercontinental transport are also identified as important ozone sources (Jaffe et al., 2018).

Regional-scale air quality models, such as the U.S. Environmental Protection Agency (EPA)'s Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006), are widely used to design emission control strategies needed to comply with the ozone standard by perturbing emissions, but keeping the future meteorological conditions the same as those occurred in the base year. An extreme hypothetical emissions reduction scenario where all domestic anthropogenic emissions are set to zero, i.e. the “zero-out” approach, is the most common modeling approach to quantify the background ozone concentration (Jaffe et al., 2018). In this study, we explore the manageable portion of the ozone non-attainment problem as it pertains to processes that can be managed via regulatory policies (i.e. controls that can be placed on anthropogenic emissions). We employ two sets of CMAQ (version 5.0.2) simulations using meteorology from the Weather Research and Forecasting (WRF) model: the base case model simulation reflecting the “current” conditions for the year 2010 (hereafter referred to as BASE) and the “zero-out” scenario where
all anthropogenic and wildfire emissions are turned-off (hereafter referred to as EM_ZERO). We assess the impacts of the “zero-out” emission scenario on ozone levels for different times of the year and at various percentiles of the ozone distribution, and quantify the contribution of anthropogenic emissions loading to ozone extreme values using the probabilistic method described by Luo et al. (2019). Further, we estimate the lower limit to the achievable ozone standard using 25 years of historical ozone time series data.

The description of the data and method of analysis is presented in Section 2; Section 3 has the discussion of the results, and conclusions are provided in Section 4.

6.2 Data and methods

6.2.1 Model simulations and observations

Time series of two sets of WRF/CMAQ (version 5.0.2) model simulations for the year 2010 were employed to study the maximum influence of the anthropogenic emissions on the ground-level ozone concentrations over the contiguous United States (CONUS). These simulations, covering CONUS, southern part of Canada and northern part of Mexico, were performed with horizontal grid spacing of 12km×12km in the offline mode (uncoupled chemical transport and meteorological models) by EPA as part of the third phase of the Air Quality Model Evaluation International Initiative (AQMEII3) (Rao et al., 2011; Galmarini et al., 2017; Hogrefe et al., 2018). The first simulation, BASE, uses the meteorology, emissions and lateral boundary conditions representative of 2010 conditions, while the second simulation, EM ZERO, reflects the condition where all anthropogenic and wildfire emissions were eliminated within the WRF/CMAQ modeling domain (but all other settings are kept the same). Further details on the configuration of the WRF/CMAQ system and evaluation of meteorological fields and air quality variables simulated in the BASE simulation can be found in Solazzo et al. (2017a, b) and Hogrefe et al. (2018).

Observations of the daily maximum 8-hr ozone concentration time series (DM8HR) were obtained from U.S. EPA’s Air Quality System (AQS). A total of 86 stations with data coverage above 80% each year during the 1990-2014 period (to provide a variety of synoptic weather conditions) were analyzed. They
are grouped into seven geographical regions following the U.S. Climate Regions (https://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php): Southwest (SW), South (S), West (W), Southeast (SE), Central (C), Northeast (NE) and West North Central (WNC) with the exception of one site in Virginia, which is being grouped to the adjacent NE instead of SE since it is too far from the other SE sites concentrated in Florida. Number of sites within each region can be found in the legend of Fig. 6-1a. Only one remote site in the Glacier National Park is available in the WNC region. The influence of emissions on ground-level ozone are also studied across the rural, suburban and urban sites (Fig. 6-1b). The location setting is determined based on the percentage of land use and land cover where the site falls within the grid cell of National Land Cover Database 2011 (https://www.mrlc.gov/data/legends/national-land-cover-database-2011-nlcd2011-legend): [0, 20) for rural (No development and Developed, Open Space), [20, 50) for suburban (Developed, Low Intensity), and [50 100] for urban (Developed, Medium and High Intensity). Given our interest on the manageable portion of the ozone distribution, which is reflected by zeroing out anthropogenic emissions, we gear our discussion in Section 3.2 around sites that were not heavily affected by the wildfires in 2010 (Fig. 6-S3).

![Figure 6-1. AQS sites grouped by (a) geographical regions: Southwest (SW), South (S), West (W), Southeast (SE), Central (C), Northeast (NE) and West North Central (WNC) and (b) location settings: rural, suburban and urban. Numbers in the legends indicate the number of sites in the corresponding group.](image-url)
6.2.2 Ozone decomposition and reconstruction

It is well recognized that various atmospheric processes operating on different timescales are embedded in ambient ozone time series data (see Fig. 2 in Dennis et al., 2010). Different filtering techniques such as the Empirical Mode Decomposition (Huang et al., 1998), Elliptic filter (Poularika, 1998), Kolmogorov-Zurbenko (KZ) filter (Rao and Zurbenko, 1994), Adaptive Filtering Technique (Zurbenko, et al., 1996), and Wavelet (Lau and Weng, 1995) can be used to achieve scale separation in time series of meteorological and air quality variables. Rao et al. (2019) applied a modified version of the Empirical Mode Decomposition (EMD), known as the Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (Colominas et al., 2014) and the KZ filter to the observed ozone time series in 2010 and found that both methods yielded similar results for the DM8HR time series data (see Fig. 2 in Rao et al., 2019). Therefore, only the results from the KZ filter are presented in this paper to examine the impacts of the emissions forcing and stochastic nature of the atmosphere on the observed ozone time series data. Here, the low-pass KZ(5,5) filter has been employed to decompose 25 years of observed and the above mentioned two sets of simulated DM8HR time series into short-term synoptic component (SY) and long-term baseline components (BL) with filter parameters of 5-day moving average window and 5 iterations chosen to extract the synoptic-scale forcing (time scales less than 21 days) (Rao and Zurbenko, 1994; Eskridge et al., 1997; Rao et al., 1997; Hogrefe et al., 2000; Hogrefe and Rao, 2001; Porter et al., 2015; Astitha et al., 2017; Luo et al., 2019; Rao et al., 2019). The 50% cutoff frequency of the KZ filter, meaning the energy at this frequency is separated half-half by the KZ (5, 5) is 0.0411 cycles per day which gives a cutoff period of $1/0.0411 \approx 24$ days (Rao et al., 1997). The zero-mean SY component, which mainly contains the synoptic-scale weather-induced fluctuations in ozone time series, is derived by subtracting the BL from the original or raw ozone time series. The magnitude of the BL component is largely governed by the anthropogenic emissions, background ozone, and other slow-changing atmospheric processes. The spectral decomposition can be expressed as:

$$BL(t) = KZ(5,5)$$

(6 − 1)
\[ SY(t) = O_3(t) - KZ(5,5) \]  \hspace{1cm} (6 - 2)

where \( t = 1, 2, \ldots, 365 \) (or 366 for a leap year) denotes the days from Jan. 1 to Dec. 31 and \( O_3 \) is the original DM8HR time series. During each iteration, the average of each window is placed to its center day (third day) for the next iteration or as BL component in the 5th iteration. Note that the KZ filter works well in the presence of missing data unless there are too many continuous days with missing values, which is one of the main advantages of the KZ filter as compared to other filtering methods. If more than half of the days (3, 4 or 5 days) are missing in a window such as the boundary days in early January and late December, the average value would be marked as missing. For example, Fig. 6-2a presents the observed ozone DM8HR time series (thin red; \( O_3 \) in equation 2) and the decomposed BL component (thick red; BL in equation 1) in Dona Ana, NM. Although there are \( \sim 10 \) days missing at the beginning and at the end because of the moving average effect (e.g., for the first iteration, Jan. 1-2 and Dec. 30-31 would be marked as missing value), this is not pose a problem for regulatory analysis even at the extreme scenario when all anthropogenic emissions are removed because of the unfavorable meteorological conditions for the formation of high ozone levels during those missing days of data (Fig. 6-2b). Only sites with at least 80% data coverage each year are analyzed in this study.

Ozone reconstruction can be accomplished by adding the SY components back onto the BL components following Luo et al. (2019). Because the strength of the synoptic forcing (defined as the standard deviation of the data in the SY component, SYstd) is highly correlated with the magnitude of the baseline level (mean of BL component, BLmean) (Astitha et al., 2017; Porter et al., 2017; Luo et al., 2019; Rao et al., 2019), SY components are also adjusted to the target BL based on a linear relationship between BLmean and SYstd before the reconstruction following Eq. 6 and 7 of Luo et al. (2019). In other words, the greater the pollution (i.e., high BL) the greater the influence of prevailing meteorology on pollutant transport (i.e., high SY). A high BL level along with strong SY forcing is required to observe ozone exceedances; however, if the magnitude of the BL level is low, ever higher SY forcing cannot lead to ozone exceedances. Thus, BL is to be considered as the deterministic part in ozone time series.
The entire year of 2010 is investigated in this study as opposed to the typical ozone season of May-September that was used in Luo et al. (2019) since large perturbation of emission would very possibly result in high ozone concentrations during times other than the typical ozone season. The reasoning behind adjusting the SY component is that if we want to reconstruct ozone concentrations for the 2010 BASE case (target) with the synoptic component of year 1990 (y), we need to adjust the synoptic component to be consistent with the BL of 2010. The reconstruction follows:

\[ O_{3}^{\text{target}}(t) = BL_{\text{target}}(t) + SY_{y,\text{adj}}(t) \]  

(6 - 3)

where \( y \) can be 1990, 1991…2014, indicating the year which the synoptic forcing is decomposed from. The target BL can either be a set of baseline time series decomposed from 2010 observations (OBS) or a projected baseline (PROJ) based on the observations (OBS) and the change simulated in the baseline components of the modeled BASE and EM ZERO as shown below:

\[ BL_{\text{proj}} = BL_{\text{obs}} + (BL_{\text{EM_ZERO}} - BL_{\text{BASE}}) \]  

(6 - 4)

For instance, \( O_{3}^{1993,\text{OBS}} \) represents the ozone time series reconstructed with observed BL in 2010 \( (BL_{\text{obs}}) \) under the 1993 meteorological conditions \( SY_{1993,\text{adj}} \). Below, we illustrate the details of our approach on the ozone decomposition and reconstruction at one monitoring site.

Observed ozone in 2010 is reconstructed to build the potential distribution of ozone concentrations given the same BL as that in 2010, but with different synoptic components embedded in ozone time series data during the 1990-2014 period (25 years). Fig. 6-2a depicts the actual ozone (thin red line), observed 4th highest ozone concentration (red cross) and observed BL (bold red line) in 2010 (Eq. 6-1 and 6-2). The reconstructed or pseudo ozone time series (thin grey lines, \( O_{3}^{\text{OBS}} \)) based on observed baseline in 2010 and their associated 4th highest values (grey crosses) are also presented. It should be noted that 4th highest values tend to be clustered at times when the magnitude of the BL is high. The number of days exceeding the current NAAQS ozone standard of 70 ppb (referred to as “ex”) are also derived from each reconstructed time series and shown in brackets along with the actual exceedances in Fig. 6-2. We refer to this ozone reconstruction and distribution based on the observed baseline in 2010 as “OBS”.

114
In the same manner, to build the ozone concentration distribution for the EM_ZERO scenario, we use the ozone changes in the two CMAQ model simulations (BASE and EM_ZERO) in a relative sense, meaning we use the difference between the BL components in the BASE and EM_ZERO scenarios, anchored to the actual observed BL for 2010 (Eq. 6-4). This is the BL that is depicted in Fig. 6-2b with the bold blue line. The reconstructed ozone concentrations for 25 years of adjusted synoptic components (gray lines in Fig. 6-2b) related to the EM_ZERO scenario are derived using Eq. 6-3 with BL$_{\text{proj}}$ in the place of BL$_{\text{target}}$ and 25 synoptic conditions adjusted to the BL$_{\text{proj}}$ based on Eq. 6 and 7 of Luo et al. (2019). We refer to this set of ozone reconstruction and distribution as “PROJ”.

**Figure 6-2.** Ozone decomposition and reconstruction for (a) 2010 observed (red), (b) projected scenario with anthropogenic and wildfire emissions removed (blue), at an AQS site in Dona Ana, NM. The actual observed 4th highest ozone concentration is marked as red cross in (a). The 4th highest ozone concentrations for reconstructed ozone concentrations are marked with grey crosses. The range (minimum to maximum) of the reconstructed 4th highest concentrations and number of days exceeding the NAAQS ozone standard is shown in brackets. In (a), the number before each range is the one derived from the 2010 observations. The dashed reference lines at 70 ppb representing the current level of the ozone standard.
6.3 Results and Discussion

6.3.1 Modeled influence of anthropogenic emissions on ozone

To study the influence of the modeled anthropogenic and wildfire emissions at different times and for various percentiles, the long-term BL components of OBS and PROJ are being compared. The reasons for choosing the BL components rather than the DM8HR ozone time series are: 1) emissions-related contributions are mostly contained in the BL component as shown in previous studies (Astitha et al., 2017; Porter et al., 2015 and 2017; Luo et al., 2019; Rao et al., 2019; 2) BL is the deterministic part in ambient ozone data; and 3) filtering out the high frequency zero-mean process (i.e., SY forcing) denoises the ozone signals throughout the year, which would otherwise hinder proper interpretation of the modeling results (Fig. 6-S1).

The contribution of anthropogenic and wildfire emissions defined as the portion of ozone reduced from OBS to PROJ (i.e., the BASE to the EM ZERO scenario) has a seasonal feature for most regions. It is clear that as emissions within CONUS are removed, the period with the highest level of ozone indicated by its baseline level shifts to springtime in all seven geographical regions (Fig. 6-3) due to the active biogenic emissions and stratospheric-tropospheric ozone exchanges during springtime. The dominant peak in spring and secondary peak seen in W is consistent with that of Trinidad Head, CA, which is widely considered to reflect the background ozone level for the Western U.S. (Fig. 6-S2). Note that the regional median BL in Fig. 6-3 is not at any single site and, hence, its magnitude should not be compared with the denoised monthly mean time series at Trinidad Head, CA. The longer life-time of ozone in the free troposphere also contributes to a longer period of time with high ozone levels (spanning spring and early summer) in the SW (regional median BL level of 40-50 ppb) because of their high elevations (≥ 1 km). The reduction of the regional median BL, signifying the influence of anthropogenic and wildfire emissions remain at 20 ppb for a large portion of the typical ozone season of May to September except for WNC, where the anthropogenic emissions contribute around 7 ppb (Fig. 6-3). The one site in WNC is not heavily influenced by wildfire emissions (see map in Fig. 6-S3), thus we can attribute the changes to the
anthropogenic emissions only. Even with zero-out anthropogenic and wildfire emissions within the CMAQ modeling domain, there is a smaller decrease in the median BL during February to May for W, SW, FL, MW and NE, which cover the majority of the sites (blue and red lines in Fig. 6-3). On the other hand, removing the anthropogenic emissions could even increase the winter time BL (W, S, NE, C in Fig. 6-3) and thus the ozone level when SY components are added (Fig. 6-S1). Fig. 6-S4 shows that the increased winter ozone BL level is more pronounced in the urban and suburban regions. The highest increase occurs in W and SW with almost doubled ozone BL level in December. This is possibly due to the dramatic changes in the relative concentrations of VOC and NOx that shifts the chemistry of the tropospheric ozone production (note, biogenic emissions are included in the EM ZERO simulations). The increased wintertime ozone agrees with the increasing trend seen in Simon et al. (2015) during a period of decreasing NOx and VOC emissions (1998-2013) based on observations; they also concluded that the increasing ozone with emission reduction occurred more often in the urbanized regions during winter when the ozone is usually at lower mixing ratios.

To further examine the contribution of anthropogenic and wildfire emissions at each monitored location, we calculate the BL reduction at various percentiles (5th - 95th percentiles representing levels of daily BL component from lower tail to higher tail at each site) (Fig. 6-4). The reductions in BL are relatively consistent with a median value of around 9 ppb at all percentiles and locations, with higher reductions at the 95th percentiles and lower reductions or even increases at the 5th percentile. At the 75th and 95th percentiles, the regions with higher baseline levels experience the most reduction, especially for Southeastern California where the baseline can be reduced by up to 29 ppb. Similar features are also seen in the relative reduction of the BL level at all percentiles. At the 5th percentile, the impact of the anthropogenic emission removal has a divergent feature mostly because of the urban/rural difference (vast majority of the increases in BL e is seen at urban sites).
Figure 6-3. Change of regional median BL component from OBS to PROJ: (a) West (W), (b) Southwest (SW), (c) South (S), (d) Central (C), (e) Southeast (SE), (f) Northeast (NE), and (g) West North Central (WNC). The median BL component across the sites in the region each day is plotted against the y-axis on the right. The change in time series (grey shades) is based on the daily regional median BL component in OBS and PROJ scenario and is calculated by $BL_{PROJ} - BL_{OBS}$.

Figure 6-5 summarizes the continental median reduction rate at all percentiles induced by zeroing out all anthropogenic and wildfire emissions. The median BL and ozone reduction (in ppb) are relatively constant around 8 ppb from 1st to 60th percentiles and then rise with an increasing rate to 18 ppb for the 100th percentile. When divided by the local OBS level, the reduction rate forms a “U” shape from lower to higher percentiles for both baseline and DM8HR. Their reduction rates are almost identical from the 30th to 90th percentiles. The DM8HR has a more pronounced reduction rate at the lower and higher percentiles, because of the fluctuations caused by the varying synoptic components. This signifies that with an extreme scenario of zero-out of anthropogenic and wildfire emissions, we see a decrease ranging from 20% to 36% for the BL level and from 19% to 45% for the DM8HR ozone.
Figure 6-4. Observed BL in 2010, BL reduction (OBS - PROJ) and BL reduction rate (1 - PROJ / OBS) *100 % at various percentiles: (a-c) 5th, (d-f) 25th, (g-i) 50th, (j-l) 75th, and (m-o) 95th percentiles. The negative reductions (increase of BL with zero anthropogenic and wildfire emissions) from the 50th percentiles and up is seen only at two non-rural sites located in the San Francisco Peninsula.
Figure 6-5. Continental median concentration and reduction of (a) BL (b) DM8HR ozone. Orange line in (a): median observed BL concentration across all site at 1-100th percentiles. The reduction (OBS - PROJ) and reduction rate (1 - PROJ / OBS) *100 % are first calculated at each site and then the median of them across all sites is shown as solid and dashed blue lines. Same applies in (b) except that the DM8HR ozone at a specific site is the median of 25 reconstructed ozone time series based on the observed BL (OBS) or projected BL (PROJ) rather than directly observed or projected.

6.3.2 Modeled manageable portion of ozone exceedances

The 4th highest ozone concentration is used to determine ozone exceedances and denotes the upper tail of the annual ozone distribution. It is highly variable and susceptible to synoptic weather fluctuations and exceptional events such as stratospheric intrusion and wildfires. Moreover, it is usually more difficult to accurately simulate the 4th highest than the baseline component (Porter et al., 2015 and 2017; Rao et al., 2019). In this section, we utilize 25 years of historical ozone observations (1990-2014) to examine the impact of anthropogenic and wildfire emissions on the 4th highest ozone concentration towards defining the modeled manageable portion of ozone exceedances and the lowest achievable ozone standard. The former is represented by the reduction rate from OBS to PROJ, i.e. 2010 base case to zero-out emissions scenario, while the latter is reflected in the minimum probability of non-attainment (Pex = number of 4th highest values larger than the ozone standard in PROJ realizations / 25 × 100 %) at various hypothetical lower
levels for the ozone standard based on the PROJ scenario. It should be noted that the exclusion of wildfire emissions in the EM ZERO simulation hinders our ability to robustly quantify the “manageable” portion of the ozone burden at every site. However, given the significance of anthropogenic emissions for the ozone formation on a daily basis vs. the episodic nature of wildfire emissions, we believe we come very close to the upper threshold for the decrease in the ozone burden.

![Figure 6-6. Median of 4th highest ozone distribution based on ozone reconstructions with 25 years of SY components: (a) with BL decomposed from observations in 2010, (b) with projected baseline PROJ, (c) reduction and (d) reduction rate (i.e. the manageable portion) by eliminating anthropogenic and wildfire emissions. The median in a) and b) is of the 25 4th highest from the 25 reconstructed time series of each scenario. The two negative reduction sites in c) and d) are located in the San Francisco Peninsula.](image)

For the 2010 base year, the median values for the 4th highest ozone concentration based on ozone reconstructions with 25 years of SY components ranges from 48 to 97 ppb (Fig. 6-6a); most of the sites with median 4th highest ozone >70 ppb are in the S, NE and inland W regions, as well as some sites in the SW region. By eliminating all anthropogenic and wildfire emissions, all median 4th highest ozone values fall below or at the current national ozone standard of 70 ppb as expected and levels for the 4th highest >50
ppb are mostly found in W, SW, S and NE regions (Fig. 6-6b). The higher 4th highest ozone values in the projected scenario are found at sites with elevations close to or >1 km in SW, reflecting the high contribution of background ozone from the global ozone burden and stratospheric-tropospheric ozone exchanges (Jaffe et al., 2018).

The modeled manageable portion of the 4th highest ozone (Fig. 6-6d) shares almost identical geographical features with the reduction in the 4th highest between OBS and PROJ (Fig. 6-6c). The two exceptionally negative reductions are seen at sites located in the San Francisco Peninsula, with an increase of 4.2 and 13.6 ppb for the median 4th highest ozone value corresponding to an increase of 8% and 28%, respectively. Figure 6-7 summarizes the manageable portion of modeled ozone for different regions. Larger manageable portion of the 4th highest ozone is seen in the four eastern regions of S (17-35%), C (18-33%), SE (23-38%) and NE (18-42%). This implies that the S and NE regions where 4th highest ozone is >70 ppb is widespread in the year 2010 (Fig. 6-7a) are very likely to improve their air quality by implementing additional emission reductions. Likewise, the W sites with 4th highest ozone >70 ppb (Fig. 6-6a) generally feature a large manageable portion of the 4th highest ozone (above 30%). However, it could be difficult for some of the SW sites with 4th highest ozone >70 ppb to see much improvements by controlling emissions since more than 80% of the ozone problem originates from transport into CONUS. This is especially important given the impact of wildfires in Southern California and the SW sites, which are zeroed-out in the CMAQ EM_ZERO scenario.

The main message from this analysis is that sites in the W and SW regions, will experience decreased probabilities of meeting a lower standard even with the extreme scenario of zero anthropogenic emissions within the modeling domain (Fig. 6-7 d-f). Moreover, the actual manageable portion of the ozone burden is even lower than what we presented here because of the wildfire emission and emission outside CONUS within the modeling domain. For a hypothetical ozone standard of 55 ppb, the majority of the sites would bear some possibility of ozone non-attainment even without any domestic anthropogenic emissions (Fig. 6-7f). This modeling exercise provides insights on potential ozone standards that can be considered as more manageable since it is possible to reduce anthropogenic emissions by a small percentage,
compared to the extreme and unrealistic zero-out scenario. For example, if the ozone standard is set at 60 ppb, several sites would remain in non-attainment with probabilities of exceeding the 60-ppb ozone threshold by $> 60\%$ (Fig. 6-7e).

Figure 6-7. Regional summary of manageable portion of 4th highest and the lowest achievable ozone standards: box plot of median of 4th highest ozone distribution reconstructed with BL of (a) OBS and (b) PROJ; (c) theoretical manageable portion of ozone; probability of 4th highest ozone exceeding the theoretical ozone standards of (d) 65 ppb, (e) 60 ppb and (f) 55 ppb.

6.4 Conclusions

In this study, we examined the maximum impacts of anthropogenic and wildfire emissions on the ozone long-term component, the ozone concentration distribution function, and the 4th highest ozone concentration with the aid of CMAQ model simulations for the 2010 base year and a future year projection with zero-out anthropogenic and wildfire emissions. The modeled manageable portion of ozone exceedances is also estimated based on the probabilistic method described in Luo et al. (2019) under different historical synoptic forcings observed during 1990 to 2014 to reveal the theoretical upper limit for the manageable portion of ozone exceedances. The results indicate that the typical ozone season (May to September) would witness an ozone level reduction of 20 ppb in most regions if the anthropogenic and wildfire emissions were eliminated. While higher levels of ozone are indicated by higher baseline levels, occurrences of the 4th highest ozone tend to appear more frequently in the springtime due to shifts in the
chemical regime of the ozone cycle when anthropogenic emissions are non-existent. This study provides insights into the maximum ozone improvement that can be achieved under the current global ozone burden. In general, levels of ozone above the annual 90\textsuperscript{th} percentile could be reduced by no more than 12-18 ppb across CONUS. More substantial manageable portions of the 4th highest ozone are seen in the four eastern regions of S (17-35\%), C (18-33\%), SE (23-38\%), and NE (18-42\%). However, some W and SW sites exhibit much less influence from the domain-wide anthropogenic and wildfire emissions because of the influence of intercontinental ozone transport or stratospheric intrusions. It would be challenging to comply with the hypothetical 60 ppb ozone standard in the presence of the current global ozone burden.
Additional Figures

**Figure 6-S1.** Same as in Figure 6-3 but for DM8HR ozone time series.

**Figure 6-S2.** Denoised observed monthly mean ground level ozone at Trinidad Head, CA (blue) and its multiannual trend. Ozone has decreased until 2009 but it then started to increase due to the growth in global economy. The values for the year 2010 are zoomed in on the top right panel. The original ozonesonde ground level observation (roughly per week (irregularly)) is averaged per month to form a monthly mean ozone time series and then decomposed using improved Complete Ensemble Empirical Mode
Decomposition with Adaptive Noise (improved CEEMDAN). See Chapter of 4 and Rao et al. (2019) for the application and interpretation of the decomposed components. The first high frequency intrinsic mode function (IMF1) contain mostly noise and, hence, is removed to form the “denoised monthly mean”. The “long term trend” is the addition of the residual and last two IMFs.

**Figure 6-S3.** U.S. wildfires for 2010 obtained from EcoWest ([https://vis.ecowest.org](https://vis.ecowest.org)).

**Figure 6-S4.** Change in regional median BL component for locations at (a) urban, (b) suburban and (c) rural areas. The change is calculated in the same manner as in Fig. 6-3 and the number of sites available in each region are shown in the corresponding parentheses.
Concluding Remarks

Prediction of the absolute pollutant concentration levels is challenging for regional air quality models because of uncertainties in input data as well as model physics and chemistry. The main scope of this research is to advance the understanding of regional-scale air quality model simulations and provide helpful guidance on model improvement and the use of such models for regulatory purposes. The operational and dynamic evaluation of a set of decadal WRF-CMAQ simulations for ozone and PM$_{2.5}$ is presented in Chapters 2-4, with concentrations decomposed to multiple process-based temporal scales using the KZ filter and improved CEEMDAN techniques. The 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 $4^{th}$ highest or top10 ozone concentration could reach.

Similar to the findings from the dynamic evaluation of ozone, the WRF-CMAQ model is generally more capable of simulating the change rate in the trend component than the absolute magnitude of the long-term trend component for the total PM$_{2.5}$ concentrations. Our proposed technique reveals the influence of time shifts in PM$_{2.5}$ speciated components to the overall model performance for total PM$_{2.5}$, indicating the need for proper allocation of emissions and an updated treatment of organic aerosols compared to the earlier model version used in this set of model simulations.

Given the strong linkage between the magnitude of the baseline ozone concentration level and peak ozone values from the dynamic model evaluation in Chapter 2, a new probabilistic method for estimating the efficacy of emission reductions in achieving ozone compliance associated with the $4^{th}$ highest ozone concentration, the ozone design value and the number of ozone exceedances was developed. The new method allows determining the ozone confidence bounds by superimposing multi-decadal historical weather-induced synoptic forcings on the prevailing baseline concentration. The validity and robustness of the proposed methodology were proven based on retrospective analysis, and it suggests that the efficacy of emissions reductions in achieving ozone compliance is better assessed by focusing on the longer-term
forcing. Furthermore, modeling results analyzed and presented in this probabilistic manner enable more explicit consideration of the ever-present uncertainty in projected changes in air quality needed to comply with the ozone standard.

With the aid of idealized CMAQ model simulations, we assessed the maximum impacts of anthropogenic and wildfire emissions on the baseline component of ozone and the 4th highest concentration. The theoretically manageable portion of ozone exceedances is also estimated based on the probabilistic method described in Chapter 5. The results indicate that the typical ozone season (May to September) would witness an ozone level reduction of 20 ppb in most regions if the anthropogenic and wildfire emissions were eliminated. While higher levels of ozone are indicated by higher baseline levels, occurrences of the 4th highest ozone tend to appear more frequently in the springtime due to shifts in the chemical regime of the ozone cycle when anthropogenic emissions are non-existent. This study provides insights into the maximum ozone improvement that can be achieved under the current global ozone burden. In general, levels of ozone above the annual 90th percentile could be reduced by no more than 12-18 ppb across CONUS. More substantial manageable portions of the 4th highest ozone are seen in the four eastern regions of S (17-35%), C (18-33%), SE (23-38%), and NE (18-42%). However, some W and SW sites exhibit much less influence from the domain-wide anthropogenic and wildfire emissions because of the influence of intercontinental ozone transport or stratospheric intrusions. It would be challenging to comply with the hypothetical 60 ppb ozone standard in the presence of the current global ozone burden.

Future model development and regulatory applications should give priority to the long-term ozone components rather than the extreme values since air quality models are generally more capable of simulating the long-term trend and changes in the ozone concentrations. Investing in devising techniques that make use outputs of air quality models in a probabilistic manner would benefit the confidence we put in those model simulations that help determine air quality regulations. However, further investigation is needed to trace the high systematic error in the absolute magnitude of long-term components, as well as the systematic shifted seasonality seen in the simulated PM$_{2.5}$ and its constituents back to sources such as errors in emission inventories, physical and chemical treatment of the pollutants, or the relationship between point
measurement and the spatial mean at various spatial resolutions used in the gridded models. The methodologies developed for this dissertation could be used in the future to dynamically evaluate updated versions of the air quality modeling systems as well as inform on necessary changes to improve prediction of surface ozone and PM$_{2.5}$ concentrations and its speciated components.
References


meteorology on ozone levels between 2002 and 2005 in the eastern US. *Atmospheric Environment* 103, 247–255.


He, J., Zhang, M., Chen, X., & Wang, M., 2014. Inter-comparison of seasonal variability and nonlinear trend between AERONET aerosol optical depth and PM10 mass concentrations in Hong Kong. *Science China Earth Sciences*, 57(11), 2606-2615.


Lau, K.-M., and Weng, H.-Y., 1995 Climate signal detection using wavelet transform: how to make a time


Mathur, R., Xing, J., Gilliam, R., Sarwar, G., Hogrefe, C., Pleim, J., Pouliot, G., Roselle, S., Spero, T.L.,
Wong, D.C., Young, J., 2017. Extending the Community Multiscale Air Quality (CMAQ)
Modeling System to Hemispheric Scales: Overview of Process Considerations and Initial


and Wildlife Service) and John Vimont (National Park Service). Available at:

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented
at the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),
pp. 3821–3824.

regional air quality model’s response to emission reductions in the presence of uncertain emission

Phalitnonkiat, P., Sun, W., Grigoriu, M.D., Hess, P., Samorodnitsky, G., 2016. Extreme ozone events: Tail
behavior of the surface ozone distribution over the U.S. Atmos. Environ. 128, 134–146.


Research Products

**Peer-reviewed publications**


**Conference presentations and proceedings**


M. Astitha, J. Yang, H. Luo, S.T. Rao, 2015: Inherent uncertainties in atmospheric modeling: weather and air pollution. 34th International Technical Meeting on Air Pollution Modelling and its Application, Montpellier, France. 4-8 May, 2015. (Conference proceedings)