Atmospheric Environment 166 (2017) 551–565

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Air quality modeling for accountability research: Operational, dynamic, and diagnostic evaluation

Lucas R.F. Henneman [a, *], Cong Liu [b], Yongtao Hu [a], James A. Mulholland [a], Armistead G. Russell [a]

[a] *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA*
[b] *School of Energy and Environment, Southeast University, Nanjing, China*

HIGHLIGHTS

- CMAQ was applied and evaluated in an accountability framework.
- Most species simulated at accuracy consistent with previous applications.
- Biases in $PM_{2.5}$ species tend to cancel each other out, yielding low $PM_{2.5}$ bias.
- CMAQ and statistically-derived sensitivities agree in the Southeast.

ARTICLE INFO

Article history:
Received 3 May 2017
Received in revised form
21 July 2017
Accepted 26 July 2017
Available online 2 August 2017

Keywords:
Air pollution accountability
Model evaluation
Chemical transport modeling

ABSTRACT

Photochemical grid models play a central role in air quality regulatory frameworks, including in air pollution accountability research, which seeks to demonstrate the extent to which regulations causally impacted emissions, air quality, and public health. There is a need, however, to develop and demonstrate appropriate practices for model application and evaluation in an accountability framework. We employ a combination of traditional and novel evaluation techniques to assess four years (2001-02, 2011-12) of simulated pollutant concentrations across a decade of major emissions reductions using the Community Multiscale Air Quality (CMAQ) model. We have grouped our assessments in three categories: *Operational* evaluation investigates how well CMAQ captures absolute concentrations; *dynamic* evaluation investigates how well CMAQ captures changes in concentrations across the decade of changing emissions; *diagnostic* evaluation investigates how CMAQ attributes variability in concentrations and sensitivities to emissions between meteorology and emissions, and how well this attribution compares to empirical statistical models. In this application, CMAQ captures O_3 and $PM_{2.5}$ concentrations and change over the decade in the Eastern United States similarly to past CMAQ applications and in line with model evaluation guidance; however, some $PM_{2.5}$ species—EC, OC, and sulfate in particular—exhibit high biases in various months. CMAQ-simulated $PM_{2.5}$ has a high bias in winter months and low bias in the summer, mainly due to a high bias in OC during the cold months and low bias in OC and sulfate during the summer. Simulated O_3 and $PM_{2.5}$ changes across the decade have normalized mean bias of less than 2.5% and 17%, respectively. Detailed comparisons suggest biased EC emissions, negative wintertime SO_4^{2-} sensitivities to mobile source emissions, and incomplete capture of OC chemistry in the summer and winter. Photochemical grid model-simulated O_3 and $PM_{2.5}$ responses to emissions and meteorologically across the decade match results from receptor-based, statistical regression models. $PM_{2.5}$ sensitivities to mobile source emissions in the summertime have decreased substantially, but wintertime mobile sensitives remain largely unchanged because decreases in negative SO_4^{2-} sensitivities match decreases in positive sensitivities from other constituents. Similarly, NO_X emissions have led to decreased summertime O_3 and increased wintertime O_3 because of opposite sensitivities. Overall, results show that emissions reductions improved air quality across the domain and remain a viable option for improving future air quality.

---

* Corresponding author.
E-mail address: lhenneman@gmail.com (L.R.F. Henneman).

http://dx.doi.org/10.1016/j.atmosenv.2017.07.049
1352-2310/

1. Introduction

Rigorous assessment of current regulations plays an important role in shaping future policy decisions. To this end, air pollution accountability research seeks to causally attribute regulatory interventions to changes in emissions, air quality, exposure/dose, and public health—links in the so-called *Accountability Chain*—to specific regulations (Health Effects Institute, 2003). Causal linkages are difficult to establish, however, due to uncertainties in the links between each step. For example, the link between emissions and air quality is uncertain due to simultaneous variability in meteorology and other emissions sources. Photochemical grid models (PGMs) offer a mechanistic approach for quantifying these relationships for accountability research.

PGMs play a vital role in air quality regulatory frameworks in the United States and Europe (National Research Council, 2004), and are increasingly used elsewhere (Hu et al., 2016; Liu et al., 2010; Xing et al., 2015). These models, alternatively called chemical transport models (CTMs), simulate emissions, transport, formation, and fate of multiple air pollutants in the atmosphere. Results from PGMs are applied in a variety of capacities, including regulatory applications (Simon et al., 2013), air quality forecasting (Odman et al., 2007), atmospheric chemistry research (Brune et al., 2016; Park et al., 2004; Travis et al., 2016), and exposure studies (Bravo et al., 2012; Di et al., 2016; Fann et al., 2012; Muller et al., 2009; U.S. EPA, 2015b).

Performance evaluation provides information relating to the reliability of model results, the magnitude of biases and uncertainties, and the applicability of model results to various problems. Existing literature contains suggested model performance metrics and benchmarks. U.S. EPA (1991), for example, recommended a set of performance criteria for 1-hour ozone. Boylan and Russell (2006) compiled results from a variety of studies using PGM modeling and considered how such models are used in the regulatory process to develop concentration-dependent performance guidance for particulate matter (PM) with diameter less than 2.5 μm ($PM_{2.5}$). Simon et al. (2012) later summarized 69 studies containing model evaluation information to provide a contextual background for ozone, PM and PM species modeling performance expectations. Emery et al. (2017) used results from Simon et al. (2014) and more recent studies to recommend numerical criteria and goals for ozone, $PM_{2.5}$, and its species, and further suggested temporal and spatial scales in which these recommended numbers apply, i.e., about 1000 km, less than one month for ozone, and one month or one season for $PM_{2.5}$ and $PM_{2.5}$ species.

Dennis et al. (2010) proposed a four-tier framework—which is reflected in regulatory language currently under development by the US EPA (U.S. EPA 2014)—for model performance evaluation: operational, dynamic, diagnostic, and probabilistic. Among the four tiers, operational evaluation, in which modeled results are compared against corresponding measured data, is applied most frequently. Dynamic evaluation results, which quantify how well models capture the impact of emissions changes, have important regulatory implications because PGMs are often applied to estimate expected changes in concentration due to proposed controls (e.g., U.S. EPA, 1998, 2005). Diagnostic evaluation assesses processes and components of the modeling system, and probabilistic evaluation compares modeled and observed distributions of specific variables, instead of focusing on matched spatio-temporal data. A detailed knowledge of model limitations in this domain is critical to regulatory decision-making as manifested in control strategy development.

Unlike operational and dynamic evaluations, diagnostic evaluations often rely on comparisons with more empirical methods and quantities inferred or derived from observations (e.g., Digar et al., 2013; Xie et al., 2011). Godowitch et al. (2011), for example, compared ozone production efficiency and results from a metric to assess wind profiles to evaluate CMAQ's ability to simulate these variables. Marmur et al. (2006, 2009) compared various observation and CMAQ-based source apportionment techniques to assess their agreement. While neither the empirical or PGM-based methods are infallible, these comparisons provide important points of context for unobservable quantities, such as sensitivities to emissions and meteorology. Consistency between the various approaches provides increased confidence in the accuracy of each of the methods; alternatively, differences provide evidence of uncertainties in both methods to be further explained.

Notably, the four-tier framework has been applied in the Air Quality Model Evaluation International Initiative (AQMEII—(Hogrefe et al., 2015; Rao et al., 2011; Xing et al., 2015)). Im et al. (2015a, 2015b), for example, evaluated multiple models' abilities to simulate O_3 and $PM_{2.5}$ over United States and Europe. Work under this initiative has applied the multi-tiered framework for model evaluation, with an emphasis on diagnostic and probabilistic evaluations (Solazzo et al., 2017b, 2017a) (Solazzo and Galmarini, 2016). separated model error into components and related these to model processes at various time scales, and determined that long-term processes and input fields (e.g., emissions and boundary conditions) contributed most to the model error.

Modeling requirements for accountability studies align with needs for detailed model assessments over periods of changing emissions and meteorology. Several studies have applied dynamic evaluation to simultaneously test impacts of emissions reductions on air quality and assess the model's ability to capture the reduction (Banzhaf et al., 2015; Cohan and Chen, 2014; Daskalakis et al., 2016; Foley et al., 2015a; Gégo et al., 2008; Godowitch et al., 2010; Simon et al., 2014), and a few have applied diagnostic evaluation techniques to attribute variability in concentrations between meteorology and emissions differences (Foley et al., 2015b; Gilliland et al., 2008; Godowitch et al., 2008; Napelenok et al., 2011). Most studies to date in this domain have focused on ozone concentrations; few studies have assessed $PM_{2.5}$ and $PM_{2.5}$ species.

As part of an accountability study on the effectiveness of regulations implemented under the 1990 Clean Air Act Amendments, we applied the Community Multiscale Air Quality model (CMAQ) version 5.0.2 over a domain covering the eastern United States to 2 two-year periods spanning a period of large emissions reductions: 2001–2002 and 2011–2012. With the Decoupled Direct Method (DDM) extension, we calculated air pollution concentration sensitivities to electricity generating unit (EGU) and on-road mobile sources. Results from CMAQ were evaluated using multiple evaluation tiers. First, we apply an operational evaluation to answer the question: *How well does the PGM, as applied, capture observed concentrations?* Next, we apply a dynamic evaluation to answer the question: *Does the PGM capture observed air quality changes?* In a diagnostic analysis, we ask two novel questions: *Can the PGM reproduce empirically-derived pollutant variability attributed to emissions and meteorology?* and *How well does the PGM reproduce the air pollutant sensitivities to emissions as simulated via empirical statistical modeling?* Together, the analysis finds that observed air quality changes during the period are attributable primarily to emissions reductions, and supports the use of the PGM in an accountability applications.

2. Methods

2.1. CMAQ modeling and inputs

CMAQ (version 5.0.2, using CB05 (Byun and Schere, 2006,