Atmospheric Environment 116 (2015) 159–171

Contents lists available at ScienceDirect

Atmospheric Environment

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

Chemical transport model consistency in simulating regulatory outcomes and the relationship to model performance

Bonyoung Koo [a,*], Naresh Kumar [b], Eladio Knipping [b], Uarporn Nopmongcol [a], Tanarit Sakulyanontvittaya [a], Mehmet T. Odman [c], Armistead G. Russell [c], Greg Yarwood [a]

[a] ENVIRON International Corporation, 773 San Marin Drive, Suite 2115, Novato, CA 94998, USA
[b] Electric Power Research Institute, 3420 Hillview Avenue, Palo Alto, CA 94304, USA
[c] School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA 30332, USA

HIGHLIGHTS

- Two differently-configured chemical transport models are evaluated for O_3 and $PM_{2.5}$.
- Modeling results are applied to determine various policy-related modeling outcomes.
- Impact of using different modeling configurations on policy outcomes is investigated.

ARTICLE INFO

Article history:
Received 20 December 2014
Received in revised form
17 June 2015
Accepted 19 June 2015
Available online 21 June 2015

Keywords:
Chemical transport model
Regulatory outcomes
Model performance
CAMx
CMAQ

ABSTRACT

It is critical to evaluate an air quality model prior to regulatory applications to ensure model performance is adequate for regulatory decision making. However, no formal benchmarks currently exist in the United States (US) to judge whether a model's performance is acceptable for given purposes, and practitioners usually rely on criteria established in the past, when the extent of modeling domains and length of simulation periods were relatively limited. This study conducts modeling experiments to investigate the impact of using different modeling configurations on various policy-related modeling outcomes. Two widely used chemical transport models with different meteorological data, biogenic emissions, and aerosol modeling schemes are applied to an annual modeling period, and model performance for ozone and particulate matter (PM) constituents is evaluated for different timescales and geographical regions in the US. Results show that while both models can be considered acceptable based on criteria commonly used by modelers when evaluated on the annual basis, the model performance at finer levels can reveal differences between the two modeling configurations, which may lead to different policy outcomes. Model results for 2005 and 2014 are used to determine monitoring sites projected to violate current US standards for ozone and fine PM, sites significantly affected by emissions from two selected upwind states, and ratios of deposition to ambient concentration of oxides of sulfur and nitrogen, a.k.a. transference ratios, at each monitoring site in the eastern US. The two modeling configurations result in quite different lists of sites and transference ratios even though both show acceptable model performance in the conventional model performance evaluation. This calls for development of model performance evaluation criteria specific to various regulatory purposes.

1. Introduction

Three-dimensional air quality models, also called chemical transport models (CTMs), are the primary tools used by regulators, industry, and other agencies for a variety of applications, including development of regulations (EPA, 2011a, b), design of air quality control strategies and implementation plans (Odman et al., 2009), and short-term operational air quality forecasting (Hu et al., 2010; McHenry et al., 2004; Menut and Bessagnet, 2010; Otte et al., 2005). Recently, their use has been extended to assessing air quality standards when monitoring data are unavailable (EPA, 2011a) as well as providing exposure estimates for epidemiological studies

* Corresponding author.
E-mail address: bkoo@environcorp.com (B. Koo).

http://dx.doi.org/10.1016/j.atmosenv.2015.06.036
1352-2310/

evaluating relationships between air pollutants and potential adverse health effects (Baxter et al., 2013). The two most commonly used CTMs in the US are the Community Multiscale Air Quality (CMAQ) Model (Byun and Ching, 1999) and the Comprehensive Air Quality Model with Extensions (CAMx; ENVIRON, 2011).

Given the widespread use of air quality models, these models need to be rigorously evaluated to ensure that modeled results can be used reliably. Typically, models are evaluated by comparing simulated concentrations of various pollutants against measurements either during a high pollution episode or for a full year to determine whether the models can accurately simulate the formation and transport of pollutants. In the past, the US Environmental Protection Agency (EPA) had established model performance criteria that urban airshed models were required to satisfy before they could be used for regulatory applications for ozone (EPA, 1991). However, no formal guidelines currently exist for evaluation of regional scale air quality models for ozone or $PM_{2.5}$ (EPA, 2007, 2014). For ozone, most practitioners continue to use the guidelines established for urban scale models whereas evaluations for $PM_{2.5}$ are based on “criteria” or “goals” used by the community in general. Overall, there is no consistency in what statistical or graphical methods are used in such evaluations and what criteria need to be followed to judge a model's performance to be “adequate” or “acceptable”. Therefore, it is difficult to compare performance of one model application against another.

There have been some attempts at developing model performance criteria for regional models. EPA conducted a workshop on “PM Model Evaluation” in February 2004 (Timin, 2004). In August 2007, a follow-on workshop was conducted by EPA in collaboration with the American Meteorological Society on “Evaluation methodologies for regional-scale air quality modeling systems” (Dennis et al., 2010). To date, however, there is no consensus on developing model performance criteria or how to assess the performance of the models for different applications. Boylan and Russell (2006) made recommendations for $PM_{2.5}$ performance metrics that could be used to assess whether a model has met model performance goals or criteria. Some of those recommendations have been informally used by other researchers (Tesche et al., 2006; Zhang et al., 2009); however, there exists no systematic methodology or acceptable criteria for conducting model performance that is accepted by the research or regulatory community. The Air Quality Model Evaluation International Initiative (AQMEII), which involved research groups from multiple countries and explored various approaches to model evaluation, did not develop any specific guidelines for model performance in various types of applications (e.g., Rao et al., 2011 and references therein).

While regulatory modeling typically applies operational evaluations where model predictions are evaluated against ambient observations, a model's ability to accurately predict air quality response to given emission changes also needs to be assessed. This need is addressed by dynamic evaluation which requires retrospective analysis of case(s) where significant air quality differences resulting from well-defined emissions changes (Dennis et al., 2010). EPA's NOx State Implementation Plan (SIP) Call (EPA, 2005) provided such a case and has been the subject of dynamic evaluation studies (Gilliland et al., 2008; Napelenok et al., 2011). The impact of shorter term emission changes has also been investigated using the weekend ozone effect (Pierce et al., 2010) and the 2003 North American blackout (Hu et al., 2006).

In the absence of model performance criteria, it is difficult to assess if the models are being thoroughly tested, and whether their performance is such that they can be confidently used for different applications. In fact, one could potentially obtain inaccurate and misleading results to guide policy questions such as “what level or types of emissions reductions are needed to achieve a desired level of air quality in the future?” Further, it is not apparent that another, more effective, strategy would be identified if a different model configuration were chosen that more accurately simulated the transport and transformation of pollutants. Pun et al. (2008) showed that relative response factors (fractional changes in pollutant concentrations due to future emissions changes) can be quite different when using different air quality models that meet a certain level of performance. Koo et al. (2012a) showed that the choice of modeling systems and simulation scenario, each with different model performance results, affected the ratio of deposition to ambient concentration, the so-called “transference ratio” that EPA has proposed to use in linking deposition to ambient concentrations as part of a consideration for a future Secondary National Ambient Air Quality Standard (NAAQS) for SOx and NOy (EPA, 2011a).

In this study, we conducted modeling experiments to see how the use of different models, inputs and configurations, consistent with the state of the science prevalent in the community, can affect various policy outcomes, and how these relate to model performance as measured by various metrics. In addition to the different models, we intentionally used different modeling configurations (e.g., meteorological data, biogenic emissions, secondary organic aerosol (SOA) schemes) to create distinct modeling frameworks that may be used in regulatory applications. The goal is to better understand how model performance and evaluation practices relate to assessing air quality model results and the responses to emission controls.

2. Methods

2.1. Overall approach

We ran two regional air quality models predominantly used in the US, CAMx and CMAQ to simulate the year 2005, which was used in prior regulatory modeling analyses (e.g., EPA, 2011a, b), using available emissions and meteorological inputs. Model predictions were evaluated against measurement data from multiple monitoring networks using common statistical measures (Table S1). Absolute measures such as mean bias/error are less useful if used alone because it is difficult to interpret their significance without knowing typical magnitudes of observation. Past EPA guidance for ozone (EPA, 1991) established model performance criteria using mean normalized bias (MNB) and error (MNE) but recently it was noted that these metrics can skew statistics (if used without a cutoff) because they are easily dominated by very low observed concentrations (Simon et al., 2012). In contrast, mean fractional bias (MFB) and error (MFE) are symmetric around zero bias and bounded by finite numbers (Boylan and Russell, 2006). Here, model evaluation was performed using normalized mean bias (NMB) and error (NME), which weigh all observations equally. As will be shown in subsequent sections, both models were found to perform adequately based on “acceptance criteria” used by the air quality community. We then ran the two models with emissions estimated for 2014 reflecting proposed air quality control policies in the US and determined the response (i.e., change in ozone and PM concentrations) of the models. We also examined the transference ratios of SOx and NOy as determined from the two models both for 2005 and 2014. In addition, we conducted source attribution studies to determine the contributions of emissions from two states (Georgia and Pennsylvania) to downwind air quality by zeroing out anthropogenic emissions from each state. While such a “zero-out” approach is not realistic and leads to the model being put in a state that is not consistent with likely conditions, at least locally, it has been used in the past providing a first estimate of maximum possible reductions (Marmur et al., 2006; Koo et al., 2009). We