Chapter 7 –Session 2 Overview: Model Evaluation: Establishing Model's Credibility

<u>Contents</u>

Session 2 Overview: Model Evaluation - Establishing Model's Credibility

7.0 Introduction

Session 1 described the substantial efforts required to develop a state-of-science air quality modeling system that supports the implementation of the National Ambient Air Quality Standards (NAAQS) under the Clean Air Act authority. This session presents the Division's model evaluation program that has been developed to support the Community Multiscale Air Quality (CMAQ) model. Photochemical air quality models like CMAQ are used to simulate ozone (O_3), particulate matter $\leq 2.5\ \mu gm^{-3}$ ($PM_{2.5}$) components, and other pollutants across regional domains. Performance evaluations play a critical role in both regulatory and research applications of the models. For example, State Implementation Plans (SIPs) include an evaluation of a "base" case simulation and air quality simulations demonstrating attainment of the NAAQS with proposed emission control strategies. In research applications, improvements to process-level model algorithms or inputs are judged, in part, based on whether these changes improved model performance. In model applications that have either or both regulatory and research purposes, models can further be used to infer relationships between atmospheric pollutant concentrations and the relevant processes, notably meteorology, chemistry, and emissions. Given the influence that model evaluation results can have on regulatory decisions and scientific conclusions about air pollution, it is critical that model evaluation studies are comprehensive and characterize model performance in insightful ways that not only reveal how well model-predicted pollutant levels compare to observed data, but also increase our confidence in the inputs (e.g., meteorology and emissions) and the modeled processes.

Program goals and strategic directions for research described in this session support Long Term Goal (LTG) 1 of the EPA Clean Air Research Multi-Year Plan (MYP) which aims to reduce uncertainty in standard setting and air quality management decisions through advances in air pollution science. The primary focus of this LTG is the development and implementation of NAAQS for particulate matter and ozone and other air quality regulations such as those for hazardous air pollutants.

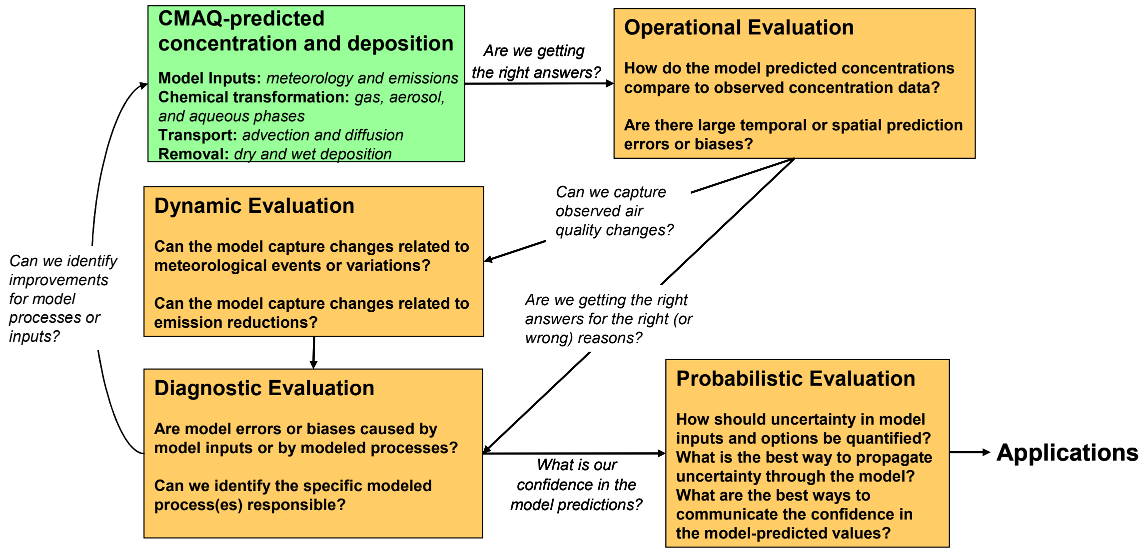
7.1 Research Summary

The evaluation program has been designed to assess CMAQ model performance for specific time periods and for specific uses of the model. Further, it has been a priority to identify improvements needed in model processes or inputs and better characterize and reduce model uncertainty. The Division has developed a framework (see Figure 7.1) to describe these different aspects of model evaluation under four categories, as outlined and illustrated below:

- Operational evaluation, as defined here, is a comparison of model-predicted and routinely measured concentrations of the end-point pollutant(s) of interest in an overall sense. This is the first phase of any model evaluation study.
- Diagnostic evaluation investigates the atmospheric processes and input drivers that affect model performance to guide CMAQ development and improvements needed in emissions and meteorological data.
- Dynamic evaluation assesses a model's air quality response to changes in meteorology or emissions, which is a principal use of an air quality model for air quality management.

- Probabilistic evaluation strives to characterize uncertainty in CMAQ model predictions for model applications such as predicted concentration changes in response to emission reductions.

Figure 7.1 AMAD Framework for Model Evaluation



Poster 2.1 provides further background and description of this model evaluation framework and the goals of the program. Research in this session is organized around four major areas of emphasis:

1. Evaluation efforts that support transfer of the CMAQ model to operational applications
2. Diagnostic research to identify the influence of modeled processes or emission inputs on CMAQ model performance
3. Dynamic evaluation efforts to assess the air quality model's response to emission changes
4. Probabilistic evaluation efforts to characterize the uncertainty in CMAQ predictions

7.1.1 Evaluation efforts that support transfer of the CMAQ model to operational applications

CMAQ is used by the EPA Office of Air Quality Planning and Standards (OAQPS) as well as many states and regional planning organizations for air quality management. These regulatory applications of the model require establishing the model's credibility through comprehensive evaluation results. CMAQ has been publicly available to the user community since 1998 with annual to biennial model releases, reflecting science improvements. While the user community has established strong capabilities to apply the model, the analysis and evaluation of model results have been limited by the lack of tools available to post-process the model output and develop both graphical and statistical evaluation results. Until recently many researchers and regulatory users were still using commercial software to create their own ad-hoc post-processing and evaluation tools for each new modeling application; a process that was inefficient and error-prone. The Atmospheric Model Evaluation Toolkit (AMET) was developed to provide standardized evaluation tools with the needed functionalities to advance the operational evaluation approaches. AMET was publicly released in 2008, with AMAD scientists collaborating with OAQPS colleagues on the features most beneficial to the regulatory users. AMET has extended the operational evaluation approaches to assess air quality model results at various spatial scales (e.g., individual sites, States, EPA Regions) and different temporal scales (e.g., hourly, daily, weekly, monthly, and seasonally). AMET can be used to evaluate both meteorological and air quality models, so that the impact of the meteorological model's performance on the air quality predictions can be more carefully considered. Poster 2.1 provides more information about AMET and example results.

To transition research models, such as CMAQ, to applications by the user community, the credibility of the model must be established. Recent EPA regulatory applications that used CMAQ were the Clean Air Interstate Rule (CAIR) and the Clean Air Mercury Rule (CAMR), after model comparisons demonstrated that CMAQ was ready for their use because of its superior evaluation results and computational efficiency. For the NOAA National Weather Service's (NWS) real-time air quality forecasting program, CMAQ again had to be vetted before it was operationally used. Poster 2.2 provides history on how the CMAQ model was adapted for the NWS meteorology forecasts, the emission forecasts were developed, and the CMAQ forecasts on eastern U.S. to continental U.S. domains were evaluated. All of these efforts led to the shift from experimental to operational application of CMAQ by the NWS for issuing national air quality forecast guidance.

7.1.2 Diagnostic evaluation research to identify the influence of modeled processes or model inputs on CMAQ predictions

Diagnostic evaluation informs and advances the development of CMAQ by identifying processes or input conditions most closely linked with the modeling errors. Session 1 discussed diagnostic evaluation efforts that were integral to the development of the CMAQ model. This included meteorological model development needs that are critical to air quality model predictions, as well as chemical and physical processes directly modeled in CMAQ. In addition to the diagnostic aspects of model development, this session focuses on research efforts to develop tools capable of diagnosing model errors and attributing those errors to specific processes or input datasets.

Instrumented versions of the CMAQ model have been created to examine specific atmospheric processes and contributions from emissions, boundary conditions, transport, etc. (Poster 2.3). A sulfur tracking version of CMAQ was developed to analyze the relative roles of gas and aqueous-phase production pathways in shaping the simulated three-dimensional sulfate distributions. Similarly, a carbon apportionment version of CMAQ was developed to track the size, composition and source of primary carbonaceous aerosols. The Decoupled Direct Method (DDM) version of CMAQ, which was originally developed by external research groups, has been included in the public release of CMAQv4.7 and is being used in our diagnostic research studies of emissions, boundary condition influences for $PM_{2.5}$, and for probabilistic modeling.

Another area of diagnostic evaluation described in Session 2 is inverse modeling studies to evaluate emissions (Poster 2.4). These studies are referred to as "top-down" because observed and predicted concentrations are used to infer whether the emissions are biased. Inverse modeling approaches have been used extensively in global scale modeling for long-lived, well mixed gases (e.g., chlorofluorocarbons, carbon dioxide), and the Division has adapted and extended an approach more suitable for regional-scale air quality modeling. Regional-scale air pollutants of concern have substantial variability in space and time, due to their shorter chemical lifetimes, heterogeneous source regions, and meteorological variability. Inverse modeling studies have been conducted to derive and evaluate seasonal NH_3 emissions and to assess NO_x emissions with NO_2 satellite data.

7.1.3 Dynamic evaluation efforts to assess the air quality model's response to emission changes

Traditionally, air quality model evaluation has expressly focused on model performance for a specific time period that has expanded from an ozone episode to annual and multi-year simulations. These studies have identified numerous areas needing improvement; however, traditional evaluation methods do not ensure that predicted concentration changes stemming from changes in emissions or meteorology are better captured. Given that regulatory application of air quality models relies on the model's response to emission changes, dynamic evaluation efforts were developed during the past four years. The NO_x SIP Call, which required NO_x controls on power plants by May 31, 2004, was identified as an excellent case study for dynamic evaluation because the emission reductions occurred over a short time period and because the emissions data are directly available from the Continuous Emission Monitor Systems (CEMS) on power plant stacks.

Gégo et al. (2007) and USEPA (2007) show examples of how observed O_3 levels have decreased notably after the NO_x SIP Call was implemented. Poster 2.5 describes how CMAQ-predicted changes in daily maximum 8-hour O_3 concentrations compared from summer 2002 to the summers of 2004 and 2005, or from before the emissions reductions to after. Gilliland et al. (2008) demonstrated that improvements in mobile emission estimates and a new CMAQ version that included an updated chemical mechanism led to the best model-predicted O_3 reductions. However, these results also revealed model underestimates of O_3 decreases in the northeastern states at extended downwind distances from the Ohio River Valley NO_x emission source region. Analysis methods, such as the e-folding distances and ozone production efficiencies (Gilliland et al., 2008; Godowitch et al., 2008), have been used to show that NO_x emissions in these simulations are not impacting O_3 levels as far downwind as observed, and oxidized nitrogen chemistry in the free troposphere has been identified as a potential cause for further investigation.

7.1.4 Probabilistic evaluation efforts to characterize the uncertainty in CMAQ predictions

When weighing the societal benefits of different air quality management strategies, policy-makers need quantitative information about the relative risks and likelihood of success of different options to guide their decisions. A key component in such a decision support system is an air quality model that can provide not only a single "best-estimate" but also a credible range of values to reflect uncertainty in the model predictions. Probabilistic evaluation of CMAQ seeks to answer these questions:

- How do we quantify the uncertainty in model inputs and parameterizations?
- How do we propagate model input uncertainty to the model outputs?
- How do we communicate the level of confidence in the model-predicted values in a way that is meaningful and useful to decision-makers?

To address these questions, a combination of deterministic air quality models and statistical methods are being used to derive probabilistic estimates of air quality. For example, an ensemble of deterministic simulations is frequently used to account for different sources of uncertainty in the modeling system (e.g. by varying emissions or meteorological inputs, boundary conditions, or parameterization of chemical or physical processes). One challenge in implementing ensemble approaches is that chemical transport models require significant input data and computational resources to complete a single simulation. The Decoupled Direct Method (CMAQ-DDM-3D, Poster 2.3) is used here to generate large member ensembles while avoiding the major computational costs of running the regional air quality model multiple times. Statistical methods are then used to post-process the ensemble of model runs based on observed pollutant levels. Maximum likelihood estimation is used to fit a finite mixture statistical model to simulated and observed pollutant concentrations. The final predictive distribution is a weighted-average of probability densities and the estimated weights can be used to judge the performance of individual ensemble members, relative to the observations.

These approaches provide an estimated probability distribution of pollutant concentration at any given location and time. The full probability distribution can be used in several ways, such as estimating a range of likely, or "highly probable" concentration values, or estimating the probability of exceeding a given threshold concentration of a particular pollutant. Poster 2.6 includes an example for the southeastern U.S., for current conditions as well as with a 50% reduction in NO_x emissions.

7.2 Future Research

The evaluation program plans continues to grow in all areas of CMAQ evaluation:

- Operational evaluation will consider more direct assessment of how meteorological model uncertainties impact air quality predictions. As these types of advancements are integrated into the AMET tool, the user community can readily use these approaches for a more thorough analysis of model results.
- Dynamic evaluation of the NO_x SIP Call will be used as an important test case for new model changes, and additional opportunities for evaluating air quality changes will be investigated. For example, we are investigating whether CMAQ is properly simulating the observed changes in air quality on daily, weekly, intra-seasonal, and longer-term time scales. Also, 2009 and 2010 emission reductions under the CAIR Rule were anticipated as good test cases for dynamic evaluation; however, this plan must be postponed since the CAIR was vacated by the courts.
- Evaluation approaches used to diagnose the role of model processes or inputs in model performance issues will continue to develop. CMAQ-DDM-3D, as well as the carbon apportionment and sulfate tracking models, have been released to the community as part of the CMAQv4.7 release. These instrumented models will continue to be used in the Division's model evaluation research as well as to inform improvements in the CMAQ model and/or model inputs. Inverse modeling, a specific example of instrumented modeling for emission evaluation, will continue to be used to address large emission uncertainties as data become available.
- Probabilistic evaluation approaches will continue to be developed for their use in regulatory applications. The goal will be to demonstrate practical, robust approach(es) to estimate confidence levels for CMAQ predictions that can aid the decision-making process.

Continued coordination of the evaluation efforts with the CMAQ model development will be a core priority for the program. Model issues that are identified through any of the above evaluation efforts will inform the next directions for ongoing and new developmental projects.

7.3 Impact

The CMAQ modeling system is used extensively by EPA and the states for air quality management (e.g., State Implementation Plans and rulemaking: Clean Air Interstate Rule, Clean Air Mercury Rule, Renewable Fuels Standard Act-2) and by the NOAA National Weather Service National Air Quality Forecasting system. The CMAQ evaluation program has supported CMAQ model releases by properly characterizing model performance and changes in model performance. By publishing evaluation results from the CMAQ model, the Division has provided benchmark model performance information to the user community. AMET has facilitated more detailed model evaluation in the user community by providing advanced tools for processing model output and making comparisons to observational data.

In addition to supporting the EPA program office and user community, the Division's evaluation program has provided scientific leadership to the user community. Our research publications demonstrate how instrumented models can be used to evaluate emissions via inverse modeling, quantify the contribution of various inputs and processes to concentrations, and to estimate model uncertainties. While there are too many specific contributions from these diagnostic efforts to list, highlights include the following:

- CMAQ-DDM-3D will contribute to Policy Relevant Background (PRB) $PM_{2.5}$ estimates in the upcoming ORD Integrated Science Assessment for the 5-year review of the $PM_{2.5}$ NAAQS.
- The carbon apportionment model identified emission biases in several primary organic carbonaceous aerosol sources that have since been improved in the National Emissions Inventory.
- CMAQ-based inverse modeling provided seasonal NH_3 emission estimates when none were available, and these results were integrated into the temporal factor estimates for the National Emission Inventory. These seasonality estimates have also been used by the GEOS-Chem global chemistry modeling community.
- CMAQ-DDM-3D inverse modeling demonstrated that NO_2 column data could be used to evaluate NO_x emissions, and also showed the importance of improving free tropospheric chemistry for oxidized nitrogen for emission budget studies.

Our dynamic evaluation studies were the first to directly evaluate model-predicted air quality changes against observed changes resulting from major emissions reductions. Results from these studies demonstrated that:

- Predicted O_3 concentration reductions after the NO_x SIP Call emissions changes will be consistent with the observed changes if the best emission information and newest CMAQ model was used.
- The model has difficulty in capturing reductions in O_3 farther downwind from the power plant emissions-rich region, which is likely related to tropospheric chemistry issues that need further investigation.

It is anticipated that the probabilistic evaluation work presented in this session will make a large impact on the user community because it will provide guidance on how uncertainty can be quantified for CMAQ applications. It has also provided diagnostic information about the sensitivity of O_3 predictions to variations in emissions and meteorological modeling options, which can be very helpful in informing and prioritizing further model development and testing.

7.4 Session Posters

The following six (6) posters will be presented in this session:

(1) Regional Air Quality Model Evaluation: Framework and Tools (2.1)
(2) Evaluation of O_3 Predictions from the National Air Quality Forecast System to Support Transition of Air Quality Modeling Research to Operations (2.2)
(3) Instrumented Models for Diagnostic Model Evaluation – Decoupled Direct Method in 3D, Carbon Apportionment, and Sulfur Tracking (2.3)
(4) Diagnostic Evaluation of Emissions via Top-down Inverse Modeling (2.4)
(5) Using Dynamic Evaluation to Assess CMAQ Model Response Induced by Emissions Changes and Meteorological Variability (2.5)
(6) New Directions in Air Quality Model Evaluation: Probabilistic Model Evaluation (2.6)

Abstracts for each of these posters follow.

Regional Air Quality Model Evaluation: Framework and Tools (2.1)

Alice Gilliland, Wyat Appel, Kristen Foley, Robert Gilliam, Robert Pinder,
S.T. Rao, Kenneth Schere, and Jenise Swall

Collaborators: Christian Hogrefe, New York Department of Environmental Conservation; Steve Porter, University of Idaho; Sharon Phillips, USEPA Office of Air Quality Planning and Standards; Jerry Davis, NC State University; John Irwin, Irwin and Associates

To advance the comprehensive model evaluation effort that is a critical component of the Community Multiscale Air Quality (CMAQ) model program, we present an evaluation framework describing the roles of operational, diagnostic, dynamic, and probabilistic evaluation approaches. Comparison of criteria pollutant predictions to observations (e.g., ozone, fine particulate matter [$PM_{2.5}$] mass and species) is a fundamental part of evaluation protocols, and it is critical to assess the role of various processes and model inputs to those predictions. These operational and diagnostic evaluation approaches can provide important insights to issues that can inform and improve the air quality model or the meteorological or emission inputs. Further, evaluating an air quality model's response to emission changes is central to how the model is used for air quality management at the local, state, and federal levels. The NO_x State Implementation Plan (NO_x SIP) Call offered a unique opportunity to conduct a "dynamic evaluation" of CMAQ's predicted O_3 change as a result of a large, abrupt NO_x emission reduction in 2004. Uncertainties in model inputs and processes are very difficult to characterize in deterministic models such as CMAQ, but having reasonable uncertainty estimates could provide additional help and guidance to the air quality management community. Probabilistic evaluation approaches are under development to characterize the impact of uncertainties in emissions, meteorology, and chemistry on air quality predictions using CMAQ and the CMAQ-Decoupled Direct Method (CMAQ-DDM, Poster 2.3). Credible uncertainty estimates can provide valuable information to air quality management decisions about the confidence in the predicted air quality changes and likelihood for reaching attainment. Additionally, these uncertainty bounds can be useful to the research process by identifying model prediction errors outside the bounds of uncertainty that warrant further investigation. With this background on the structure and goals of the evaluation program, the individual posters in this session will describe specific research studies in more detail.

Evaluation of O_3 Predictions from the National Air Quality Forecast System to Support Transition of Air Quality Modeling Research to Operations (2.2)

Brian Eder, Rohit Mathur, Jonathan Pleim, Tanya Otte,
Kenneth Schere, Jeffrey Young and George Pouliot

Collaborators: Daiwen Kang, Shaocai Yu, Daniel Tong, Hsin-mu Lin, Tianfeng Chai (Science and Technology Corporation); Paula Davidson, Jeff McQueen (NOAA, NWS)

Although air quality has improved significantly in the decades following passage of the Clean Air Act and its Amendments, there are still many areas in the United States where the public is exposed to unhealthy levels of air pollutants, most notably ozone (O_3) and fine particulate matter ($PM_{2.5}$). The cost of poor air quality to the United States from pollution-related illnesses alone has been estimated at 150 billion dollars. For many citizens, especially those who suffer from respiratory problems, the availability of air quality forecasts (analogous to weather forecasts) could make a significant difference in how they plan their daily activities and in turn improve the quality of their lives. Accordingly, NOAA, which has environmental prediction as one of its core missions, has partnered with EPA, which has the protection of human health and welfare as a core mission, to develop, refine, and disseminate a real-time Nationwide Air Quality Forecast System (NAQFS).

The purpose of this research has been to provide a series of operational evaluations that characterize the performance of numerous iterations of the NAQFS (as it has expanded in both its capacity and coverage), supporting its transition into operational status. The evaluations have utilized a suite of metrics that examine the NAQFS performance of both *discrete forecasts* (observed versus modeled concentrations) for the maximum 8-hour O_3 concentrations (the focus of this poster), and *categorical forecasts* (observed versus modeled exceedances / non-exceedances) for the 8-hour (85 ppb) National Ambient Air Quality Standard. These evaluations typically covered five-month periods (1 May to 30 September) using O_3 concentration measurements obtained from EPA's AIRNow network.

Results of the numerous evaluations indicate that the NAQFS has performed well through its numerous refinements and expansions, and continues to do so. Mean, domain-wide correlations are typically ≥ 0.70. Values of normalized mean bias and error are generally within 10% and 20%, respectively. However, closer examination of these metrics over finer spatial and temporal scales has revealed some systematic deficiencies with the various NAQFS configurations. Recent examples include southern California, where the NAQFS tends to underpredict O_3 concentrations (especially on weekends), and the southeast Atlantic and Gulf coasts regions, where the model overpredicts. Subsequent analysis revealed that the incorrect temporal allocation of precursor emissions was likely the source of the underprediction in southern California, while inaccurate simulation of PBL heights likely contributed to the overprediction in the coastal regions.

When compared to the evolution of numerical weather prediction's incorporation into weather forecasting, which took decades, the success of this program, which involves numerical O_3 forecasts based on the WRF-CMAQ model, has been realized in a remarkably short time frame.

Instrumented Models for Diagnostic Model Evaluation:
Decoupled Direct Method in 3-D, Carbon Apportionment, and Sulfur Tracking (2.3)

Sergey Napelenok, Prakash Bhave, Shawn Roselle

Collaborators: Dan Cohan, Rice University; Yongtao Hu, Talat Odman, Armistead Russell, Georgia Institute of Technology

It is often useful to determine not only the state of an environmental system, but also its response to perturbations in various parameters that define it. Instrumented models offer a unique interpretation of standard air quality model output and are useful to better understand the physical and chemical processes occurring in the atmosphere. In terms of diagnostic model evaluation, instrumented models can (1) identify model processes that require further attention, (2) complement operational model evaluation (which is important to establish a general level of confidence in modeling results), and (3) identify model outputs that require further attention. Three instrumented models are presented: Decoupled Direct Method in three dimensions (CMAQ-DDM-3D), Carbon Apportionment (CMAQ-CA), and Sulfur Tracking (CMAQ-ST).

- CMAQ-DDM-3D provides an efficient and accurate approach for calculating the sensitivity of atmospheric pollutant concentrations to changes in photochemical model parameters (emissions, chemical reaction rates, initial/boundary conditions, etc.).
- CMAQ-CA allows quantification of absolute contributions from different emission sources to primary organic carbon (OC) and elemental carbon (EC). Molecular tracer and radiocarbon techniques have been developed to measure source-specific contributions. The combination of these measurements and carbon apportionment allows evaluation of both the air quality model and the emissions inventory.
- CMAQ-ST allows for analysis of the sulfate production pathways. It tracks sulfate production from gas-phase and aqueous-phase chemical reactions, as well as contributions from emissions and initial and boundary conditions.

All of these models have been made available to the regulatory and academic communities. Each one has been used in a wide range of applications, and has extracted additional utility from model predictions. During the 2008 model release cycle, the models are being made available for the first time, via the Community Modeling and Analysis System (CMAS) Center.

Aside from diagnostic model evaluation, instrumented models are useful in regulatory applications in the areas of source apportionment, estimation of uncertainty, selection and evaluation of control strategies, and future projections. In the future, these instrumented models will continue to be updated to incorporate the latest scientific advancements contained in the base CMAQ model, and distributed to the user community. Furthermore, the Division plans to build additional instrumented model capabilities, including developing adjoint versions of the CMAQ model.

Diagnostic Evaluation of Emissions via Top-down Inverse Modeling (2.4)

Robert Pinder, Sergey Napelenok, Alice Gilliland

Collaborator: Randall Martin, Dalhousie University

In model evaluation studies, deficiencies in emissions are often a likely culprit for model biases or errors. Unlike meteorological and air quality model predictions, regional-scale emissions estimates cannot be validated against direct measurements (with the one exception being measurements available from Continuous Emissions Monitoring Systems (CEMS) on electricity generating units). For mobile emissions, flux measurements can be made directly at the tailpipe to develop emission factors, but the aggregate estimate across national roads and highways is an extrapolation that includes numerous assumptions. Diffuse area sources, such as fertilizer applications, are much more difficult to characterize at individual sites, especially for gridded, regional-scale modeling.

To complement the immense effort required to develop “bottom-up” emission inventories, inverse modeling approaches have been introduced as a tool for evaluating and refining emission estimates. While some have referred to inverse modeling as running the model “backwards,” that is a misnomer. Instead, based on the estimated sensitivity of the pollutant concentration to emission changes (the Jacobian or K_t) and on estimated uncertainties, inverse methods can be used to infer how much the emissions would need to change to get optimal agreement with the observations. While originally used in global-scale modeling for long-lived species such as chloroflourocarbons, the Division’s work in this area has demonstrated that similar approaches can be effectively applied to regional-scale air quality modeling problems. In general, inverse modeling approaches estimate the emission levels that would minimize the differences between modeled and observed concentrations. The methods rely on the response of air quality model predictions to emission changes and uncertainty estimates of concentrations and emissions. There are two studies summarized here:

- The NH_3 study demonstrates that inverse methods can be used to characterize seasonal changes in NH_3 emissions, and it also shows that overestimating these emissions during cooler seasons causes large biases in total nitrate concentrations in the model.

- The NO_x study, which is focused on urban areas of high mobile NO_x emissions and surrounding rural areas, demonstrates that NO_2 satellite retrievals can serve as observational data for NO_x inverse modeling. The study also expands on the inverse modeling approach by characterizing individual source regions, which is possible only because of the denser satellite data and incorporating the CMAQ-DDM capabilities into the approach. The importance of considering modeling uncertainties is also highlighted in the NO_x study, where model uncertainties in the free tropospheric NO_x chemistry severely affected NO_2 comparisons in rural areas.

Using Dynamic Evaluation to Assess CMAQ Model Response Induced by Emissions Changes and Meteorological Variability (2.5)

James Godowitch, Alice Gilliland, Robert Pinder, Kristen Foley, S.T. Rao

Collaborators: Christian Hogrefe (SUNY-A/ASRC); Edith Gego, P. Steven Porter (U. of Idaho)

In efforts undertaken to demonstrate attainment of the ozone (O_3) standard, regional-scale air quality models are applied with potential emission control strategies in order to estimate pollutant response. A model performance evaluation is also typically performed using model-predicted concentrations with base case emissions to determine its ability to reproduce past observations. However, such an operational evaluation does not establish that a model will respond correctly to emission changes. A dynamic evaluation approach, which evaluates a model's ability to accurately simulate air quality changes related to actual changes in emissions, has been difficult until recently to undertake, yet it is critically relevant to regulatory applications.

A prototype dynamic evaluation effort has been undertaken to assess the Community Multiscale Air Quality (CMAQ) model's predicted O_3 response to substantial real-world NO_X point source emission reductions associated with the EPA's NO_X State Implementation Plan (SIP) Call program implemented from 2003 through May 2004 and to on-road mobile source emissions changes. The emission changes due to the NO_X SIP Call were well characterized from CEMS (Continuous Emissions Monitoring System) hourly measurements, and substantial changes were identified in observed O_3 levels. The selected modeling periods spanned three months (June, July, August), with summer 2002 representing the pre-NO_X SIP Call period. To consider the modeled response to emission changes in light of strong meteorological influences on O_3, two summer periods after the NO_X SIP Call (2004 and 2005) were included that exhibited very different meteorological conditions. Model configurations and inputs included the following: the CMAQ (v4.5) model with the CB4 and SAPRC99 chemical mechanisms, and CMAQ (v4.6) with CB05 chemistry; meteorological fields generated by the Penn State/NCAR Mesoscale Model (MM5 v3.6.3) with 4-D data assimilation (FDDA); and 3-D emissions from the SMOKE (v2.2) processing system. The National Emissions Inventory (NEI) 2001, adjusted to the particular modeling year, was used in simulations with the CB4 and SAPRC99 chemical mechanisms, while the more recent NEI 2002 and NEI 2005 were employed in the CB05 simulations. The MOBILE6 model estimated on-road vehicle emissions with projected VMT and control program information for a reference county only (CB4 and SAPRC99 runs), while the NEIs for 2002 and 2005 contained county-specific control program information. The hourly CEMS point source emissions data were available for each summer period, and the BEIS (v3.13) model generated biogenic emissions estimates.

The results of this first-of-its-kind dynamic evaluation revealed that the model tended to underestimate the maximum 8-h O_3 reductions observed after the NO_X SIP Call was implemented. Statistical results of relative (%) and absolute (ppb) observed and modeled changes for 2004 minus 2002 and 2005 minus 2002 in Gilliland et al. (2008) also indicated that the model predicted changes better at O_3 levels $\geq 95^{th}$ percentile than at the median of the distribution. The CMAQ model with CB05 using updated emission inventories showed incremental improvements in the modeled O_3 response to the NO_X emission reductions. While mobile emission uncertainties may be a factor, results of spatial correlation analysis (e-folding distances) and model comparisons with observations aloft suggest that the contribution of long-range transport of O_3 and precursors was underpredicted, especially when using the CB4 chemical mechanism.

New Directions in Air Quality Model Evaluation:
Probabilistic Model Evaluation (2.6)

Kristen Foley, Robert Pinder, Sergey Napelenok

Collaborators: Christopher Frey (NC State University), Halûk Özkaynak (NERL/HEASD)

When weighing the societal benefits of different air quality management strategies, policy-makers need quantitative information about the relative risks and likelihood of success of different options to guide their decisions. A key component in such a decision support system is an air quality model that can estimate not only a single "best estimate" but also a credible range of values to reflect uncertainty in the model predictions. Probabilistic evaluation is a relatively new component of AMAD's model evaluation framework. This aspect of the evaluation of the Community Multiscale Air Quality (CMAQ) model seeks to answer these questions: How do we quantify our uncertainty in model inputs and parameterizations? How do we propagate this uncertainty to the predicted model outputs? How do we communicate our level of confidence in the model-predicted values in a way that is valuable and useful to decision-makers?

To address these questions, we have deployed a combination of deterministic air quality models and statistical methods to derive probabilistic estimates of air quality. For example, an ensemble of deterministic simulations is frequently used to account for different sources of uncertainty in the modeling system (e.g., emissions or meteorological inputs, boundary conditions, parameterization of chemical or physical processes). A challenge with ensemble approaches is that chemistry-transport models require significant amounts of input data and computational resources to complete a single simulation. We have applied the Decoupled Direct Method (CMAQ-DDM-3D) to generate high-member ensembles while avoiding the major computational cost of running the regional air quality model multiple times. We have also used the Bayesian Model Averaging (BMA) statistical technique to postprocess the ensemble of model runs based on observed pollutant levels. Maximum likelihood estimation is used to fit a finite-mixture statistical model to simulated and observed pollutant concentrations. The final predictive distribution is a weighted average of probability densities, and the estimated weights can be used to judge the performance of individual ensemble members, relative to the observations. Daily ozone data from a set of Air Quality System (AQS) monitoring stations in the Southeast United States are used to select a set of weighted ensemble members that, when compared to observations, have minimum spread but still capture the observed variability.

The results from this research provide an estimated probability distribution of pollutant concentration at any given location and time in the study domain. The full probability distribution can be used in several ways, such as estimating a range of likely, or "highly probable," concentration values, or estimating the probability of exceeding a given threshold value of a particular pollutant. This information can be used to quantitatively compare the relative risks and rewards of air quality control options and to select the emissions control strategy that has the largest probability of success. The information contained in the ensemble of model simulations can also be used as a model evaluation tool by uncovering cases and locations that cannot be explained using a realistic range of model inputs.

7.5 Session Publications

This section presents the products (generally from 2004-2008) associated with each poster in this session. Some products are associated with multiple posters, so they are listed as products under more than one poster.

Regional Air Quality Model Evaluation: Framework and Tools (2.1)

Appel, K.W., **P.V. Bhave**, **A.B. Gilliland**, **G. Sarwar**, **S.J. Roselle**, Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model performance; Part II - particulate matter, Atmos. Environ, doi:10.1016/j.atmosenv.2008.03.036, 2008.

Appel, K.W., A.B. Gilliland, **G. Sarwar**, and **R.Gilliam**, Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model performance; Part I - ozone, Atmospheric Environment, 41, 9603-9615, 2007.

Dennis, R.L., T. Fox, M. Fuentes, **A.B. Gilliland**, S. Hanna, C. Hogrefe, J. Irwin, **S.T. Rao**, R. Scheffe, **K. Schere**, D. Steyn, A. Venkatram, On the Evaluation of Regional-Scale Photochemical Air Quality Modeling Systems, submitted to *Atmos. Environ*.

Dennis, R.L., **P.V. Bhave**, **R.W. Pinder**, Observable indicators of the sensitivity of PM2.5 nitrate to emission reductions—Part II: Sensitivity to errors in total ammonia and total nitrate of the CMAQ-predicted non-linear effect of SO_2 emission reductions. Atmospheric Environment, 42(6), 1287-1300 (2008)

Gilliland, **A.B., J.M. Godowitch**, C. Hogrefe, and **S.T. Rao**, Evaluating Regional-Scale Air Quality Models, Air Pollution Modeling and Its Application XIX, Chapter 4.8, 414-421, in press.

Gilliland, A.B., C. Hogrefe, **R.W. Pinder**, **J.M. Godowitch**, **K.L. Foley**, and **S.T. Rao**, Dynamic Evaluation of Regional Air Quality Models: Assessing Changes in O_3 Stemming from Changes in Emissions and Meteorology, Atmospheric Environment, doi:10.1016/j.atmosenv.2008.02.018, 2008.

Hogrefe, C., P.S. Porter, E. Gego, **A. Gilliland**, **R. Gilliam**, **J. Swall**, J. Irwin, and **S.T. Rao**, Temporal features in observed and predicted $PM_{2.5}$ concentrations over the Eastern U.S., Atmospheric Environment, 5041-5055, 2006.

Pinder, R.W., **R.C. Gilliam**, **K.W. Appel**, **S.L. Napelenok**, **A.B. Gilliland**, Efficient Probabilistic Estimates of Ozone Concentration Using an Ensemble of Model Configurations and Direct Sensitivity Calculations, submitted to *Environ. Sci. & Technol.*.

Pinder, R.W., R.L. Dennis, P.V. Bhave, Observable indicators of the sensitivity of $PM_{2.5}$ nitrate to emission reductions—Part I: Derivation of the adjusted gas ratio and applicability at regulatory-relevant time scales. Atmospheric Environment, 42(6), 1275-1286. 2008

Swall, J. L., **K.M. Foley,** The impact of incommensurability on model evaluation strategies: Moving beyond the comparison of matched observations and model grid cells, *Atmos. Environ.*, in press.

Public release of the Atmospheric Model Evaluation Tool (AMET): February 2008 (www.cmascenter.org).

Evaluation of O_3 Predictions from the National Air Quality Forecast System to Support Transition of Air Quality Modeling Research to Operations (2.2)

Eder, B. , D. Kang , **S. T. Rao**, **R. Mathur**, S. Yu, **T. Otte**, **K. Schere**, R. Wayland, S. Jackson, P. Davidson and J. McQueen, 2008: A demonstration of the use of National air quality forecasts for developing local air quality index forecasts. Submitted to *Bull. AMS.*

Eder, B., D. Kang, **R. Mathur**, **J. Pleim** and S. Yu, 2008: A performance evaluation of the National Air Quality Forecast Capability fo the summer of 2007. Submitted to *Atmos. Environ.*

Eder, B., D. Kang, **R. Mathur**, S. Yu and **K. Schere**, 2006: An Operational evaluation of the Eta-CMAQ air quality forecast model. *Atmos. Environ*. **40**: 4894 - 4905.

Eder, B. and S. Yu, 2006: A performance evaluation of the 2004 release of Models-3 CMAQ**.** *Atmos. Environ*. **40**: 4811-4824.

Kang, D., **R. Mathur**, **S. T. Rao** and S. Yu: 2008 Bias-adjustment techniques for improving ozone air quality forecasts, In press, *J. of Geophys. Res.*
Kang, D., **R. Mathur, K. Schere**, S. Yu and **B. Eder:** 2007, New categorical metrics for air quality model evaluation. *J. Appl. Met. and Climate*, **46**, pp. 549–555.
Kang, D., **B. Eder**, A. Stein, G. Grell, S. Peckham and J. McHenry, 2005: The New England Air Quality Forecasting Pilot Program: Development of an evaluation protocol and performance benchmark. *J. Air & Waste Manage. Assoc***. 55:** 1782-1796.
Mathur, R.: 2008, Estimating the Impact of the 2004 Alaskan Forest Fires on Episodic Particulate Matter Pollution over the Eastern United States through Assimilation of Satellite Derived Aerosol Optical Depth in a Regional Air Quality Model, *J. of Geophys. Res.,* 113, D17302, doi:10.1029/2007JD009767, 2008.
Mathur, R., S. Yu, D. Kang, and **K. Schere**: 2008, Assessment of the Winter-time Performance of Developmental Particulate Matter Forecasts with the Eta-CMAQ Modeling System, *J. Geophys. Res*., 113, D02303, doi:10.1029/2007JD008580.
Otte, T. and coauthors, 2005: Linking the Eta model with the Community Multi-scale Air Quality (CMAQ) modeling system to build a national air quality forecasting system. *Weather and Forecasting,* **20**, 367-384.
Yu, S., **R. Mathur, K. Schere**, D. Kang, **J. Pleim, J. Young**, D. Tong, **G. Pouliot**, S. McKeen, and **S.T. Rao**: 2008, Evaluation of real-time PM2.5 forecasts and process analysis of PM2.5 formation over the eastern U.S. using the Eta-CMAQ forecast model during the 2004 ICARTT study, *J. Geophys. Res*., 113, D06204, doi:10.1029/2007JD009226.
Yu, S., **R. Mathur**, **K. Schere**, D, Kang, **J. Pleim**, and **T. Otte**: 2007, A detailed evaluation of the Eta-CMAQ forecast model performance for O_3, its related precursors, and meteorological parameters during the 2004 ICARTT study, *Journal of Geophysical Research-Atmospheres*, 112, D12S14, doi:10.1029/2006JD007715.
Yu, S., **B. Eder**, **R. Dennis**, S. Chu and S. Schwartz, 2006: New unbiased metrics for the evaluation of air quality models. *Atmos. Sci. Let.* **7**: 26-34.
Yu, S., **R. Mathur**, D. Kang, **K. Schere**, **B. Eder** and **J. Pleim**: 2006, Performance and Diagnostic Evaluation of Ozone Predictions by the Eta-Community Multiscale Air Quality Forecast System during the 2002 New England Air Quality Study. *J. Air & Waste Manage. Assoc.* **56**: 1459–1471.

Instrumented Models for Diagnostic Model Evaluation: Decoupled Direct Method in 3-D, Carbon Apportionment, and Sulfur Tracking (2.3)

Bhave, P.V., Pouliot, G.A., Zheng, M., Diagnostic model evaluation for carbonaceous $PM_{2.5}$ using organic markers measured in the southeastern U.S., *Environmental Science & Technology*, 41, 1577-1583, 2007.
Liao, K.-J., Tagaris, E., Manomaiphiboon, K., **Napelenok, S.L.**, Woo, J.-H., He, S., Amar, P., Russell, A.G., Sensitivities of ozone and fine particulate matter formation to emissions under the impact of potential future climate change, *Environmental Science & Technology*, 41, 8355-8361, 2007.
Napelenok, S.L., Cohan, D.S., Odman, M.T., Tonse, S., Extension and evaluation of sensitivity analysis capabilities in a photochemical model, *Environmental Modelling & Software*, 23(8), 994-999, 2008.
Napelenok, S.L, Pinder, R.W., Gilliland, A.B., Martin, R.V., A method for evaluating spatially-resolved NOx emissions using Kalman filter inversion, direct sensitivities, and space-based NO_2 observations, *Atmospheric Chemistry & Physics*, 8, 5603-5614, 2008.
Pinder, R.W., Dennis, R.L., Bhave, P.V., Observable indicators of the sensitivity of $PM_{2.5}$ nitrate to emission reductions – part I: derivation of the adjusted gas ratio and applicability at regulatory-relevant time scales, *Atmospheric Environment*, 42(6), 1275-1286, 2008a.
Pinder, R.W., Gilliam, R.C., Appel, K.W., Napelenok, S.L., Foley, K.M., Gilliland, A.B. Efficient probabilistic estimates of surface ozone concentration using an ensemble of model configurations and direct sensitivity calculations, *Environmental Science & Technology*, 2008b (in review).
Reff, A., Bhave, P.V., Pouliot, G.A., Pace, T.G., **Mobley, J.D.**, Houyoux, M., Emissions Inventory of $PM_{2.5}$ Trace Elements across the United States, *Environmental Science & Technology*, 2008 (in review).

Diagnostic Evaluation of Emissions via Top-down Inverse Modeling (2.4)

Gilliland, A.B., **K.W. Appel**, **R. Pinder**, **S.J. Roselle**, and **R.L. Dennis**, Seasonal NH_3 emissions for an annual 2001 CMAQ simulation: inverse model estimation and evaluation, Atmospheric Environment, 4986-4998. 2006.

Gilliland, A.B., **R.L. Dennis**, **S.J. Roselle**, and **T.E. Pierce**, Seasonal NH_3 emission estimates for the Eastern Unites States using ammonium wet concentrations and an inverse modeling method, Journal of Geophysical Research-Atmospheres, 108, NO. D15, 4477, 10.1029/2002JD003063, 2003.

Napelenok, S.L., **R.W. Pinder**, **A.B. Gilliland**, R.V. Martin, A method for evaluating spatially-resolved NOx emissions using Kalman filter inversion, direct sensitivities, and space-based NO2 observations, Atmospheric Chemistry and Physics (ACP), 8, 5603-5614, 2008.

Napelenok, S.L., **R.W. Pinder**, **A.B. Gilliland**, R.V. Martin, Developing a Method for Resolving NOx Emission Inventory Biases Using Discrete Kalman Filter Inversion, Direct Sensitivities, and Satellite-Based NO2 Column, Air Pollution Modeling and Its Application XIX, Chapter 6.1, 322-330, 2008.

Pinder, R.W., P.J. Adams, S.N. Pandis, **A.B. Gilliland**, Temporally resolved ammonia emission inventories: Current estimates, evaluation tools, and measurement needs, Journal of Geophysical Research-Atmospheres, 111, doi:10.1029/2005JD006603, 2006.

Using Dynamic Evaluation to Assess CMAQ Model Response Induced by Emissions Changes and Meteorological Variability (2.5)

Gego, E.**, A. Gilliland, J. Godowitch, S.T. Rao,** P.S. Porter, C. Hogrefe**.** Modeling analyses of the effects of changes in nitrogen oxides emissions from the electric power sector on ozone levels in the eastern United States. J. Air & Waste Manage. 58:580-588, DOI:10.3155/1047- 3289.58.4.580, (2008).

Gego, E., P.S. Porter**, A. Gilliland, S.T. Rao.** Observation-based assessment of the impact of nitrogen-oxide emissions reductions on ozone air quality over the eastern United States. J. Appl. Meteorol. and Clim., 46, 994-1008, (2007).

Gilliland, A.B., C. Hogrefe**, R.W. Pinder, J.M. Godowitch, K.L. Foley, and S.T. Rao.** Dynamic Evaluation of Regional Air Quality Models: Assessing Changes in O_3 Stemming from Changes in Emissions and Meteorology, Atmospheric Environment, 42, 5110-5123, doi:10.1016/ j.atmosenv.2008.02.018, (2008).

Gilliland, A.B., J.M. Godowitch, C. Hogrefe**, and S.T. Rao.** Evaluating Regional-Scale Air Quality Models, Air Pollution Modeling and Its Application XIX, Chapter 4.8, 414-421, in press

Godowitch, J.M., C. Hogrefe**, S.T. Rao.** Diagnostic analyses of a regional air quality model: Changes in modeled processes affecting ozone and chemical-transport indicators from NO_X point source emission reductions. J. Geophysical Research-Atmospheres,113, D19303,doi:10.1029/2007JD009537, (2008).

Godowitch, J.M., A.B. Gilliland, R.R. Draxler**, S.T. Rao.** Modeling assessment of point source NO_X emission reductions on ozone air quality in the eastern United States. Atmospheric Environment, 42, 87-100, (2008).

New Directions in Air Quality Model Evaluation: Probabilistic Model Evaluation (2.6)

Foley, K.M., R.W. Pinder, S.L. Napelenok, and H.C. Frey. Probabilistic Estimates of Ozone Concentrations from an Ensemble of CMAQ Simulations, poster for Models-3 Users' Conference, Chapel Hill, October 2008.

Özkaynak, H., H.C. Frey, J. Burke, and **R.W. Pinder.** Analysis of coupled model uncertainties in source to dose modeling of human exposures to ambient air pollution: a PM2.5 case- study, submitted to *Atmospheric Environment*

Pinder, R.W., R.C. Gilliam, K.W. Appel, S.L. Napelenok, K.M. Foley, and A.B. Gilliland. Efficient Probabilistic Estimates of Ozone Concentration Using an Ensemble of Model Configurations and Direct Sensitivity Calculations, submitted to *Environmental Science & Technology.*