Air Pollution Data for Model Evaluation and Application

Armistead Russell[1,*], James Boylan[2], and Sergey Napelenok[3]

1. Georgia Institute of Technology, Atlanta, GA
2. Georgia Department of Natural Resources, Atlanta, GA
3. US Environmental Protection Agency, RTP, NC

*Corresponding author

ABSTRACT

Confidence in air quality models is founded by continual model evaluation, which, in turn, relies on having the appropriate measurement data for evaluation. The type of data desired depends on the objectives of the intended model application. Data routinely available from state, local, tribal and EPA monitoring networks is sufficient for many common model application evaluations, including assessing the impact of emissions controls. However, advancing the underlying science in the model, and developing increased confidence for non-traditional uses, requires more specialized measurements, such as continuous methods for particulate matter speciation.

INTRODUCTION

One objective of designing an air pollution monitoring network is to obtain data for applying and evaluating air quality models that are used in the air quality management process (*1, 2*). As such, it is important to consider what data are desirable for evaluating air quality models. Unfortunately, since the atmosphere is a very complex system, that models cannot fully capture all relevant physical and chemical processes, and are inherently incomplete (*3*). Or, as it is often said, "all models are wrong, some are useful." A model evaluation that compares model results against measurements is one mechanism to provide confidence in their usefulness (*4, 5*).

Two issues that need to be addressed in considering the data requirements for a specific model evaluation: 1) what specific model is being evaluated and 2) what is the purpose of the evaluation. A model that has already undergone extensive evaluation of model performance and peer review, and has been open to public scrutiny, requires less evaluation than a new, less tested model. Still, the appropriate type of measurement (or observational) data needs to be identified to evaluate the type of model in consideration. Similarly, using a well tested model for a new, as opposed to routine, purpose requires more extensive evaluation to develop the desired level of confidence. But, again, the data used in the evaluation should be appropriate to address the purpose of the evaluation.

Air quality models serve a number of purposes. In the regulatory arena, they are used to evaluate the impacts of changing emissions, e.g., to assess the effectiveness of emissions controls, to estimate the impacts of specific sources on pollutant concentrations, and to provide information to assess pollutant exposures and potential effects. Scientific applications center around developing a better understanding of how chemical species evolve and interact with exposed ecosystems, and elucidating the processes that govern chemical evolution.

For the purpose of this article, we focus on chemical transport models (CTMs) because they are widely used in the two primary model application areas - public policy and science (26). Further, we focus on data desired for "field" applications, i.e., the use of CTMs to simulate actual ambient pollutant concentrations. Issues are addressed at a broader level to provide insight to other air quality models as well.

Recognizing that the type of evaluation being conducted is dependent on the model chosen and the objectives, Studies (*4, 6*) have identified four types of model evaluation:

- Operational – evaluates model predictions against observed concentrations using a variety of statistical measures and graphical plots
- Diagnostic – evaluates model processes (chemical and physical), model parameters, and model inputs
- Dynamic – evaluates model response to changes in emissions and/or meteorology
- Probabilistic – evaluates model uncertainty by using various statistical techniques

While those four categories were primarily developed for CTMs, they apply across all models. In general, dynamic evaluations are important for public policy applications, while all types of evaluations are important for scientific assessment applications. Increased confidence is built from applying all four types of evaluations.

CTMs can be used to predict pollutant concentrations in the atmosphere by simulating emissions, advection, diffusion, deposition, and chemistry. Typically, 3-D Eulerian CTMs are used to access how ozone (O_3) and particulate matter (PM) will respond to emission controls. These results can be used to develop control strategies in order to bring areas that are currently violating the National Ambient Air Quality Standards (NAAQS) into attainment. The two most widely used, and most thoroughly evaluated, CTMs in the U.S. are CMAQ (*7*) and CAMx (*8*). Their uses include the development of State Implementation Plans (SIPs), assessment of federal control policies (e.g., (*9*)), establishing source impact relationships (including health impacts), and elucidating atmospheric chemical dynamics. In addition to being widely used, those models also benefit from being descended from models that were also extensively tested. Over the years, CTMs have shown increasing skill in simulating O_3 and PM levels leading to their increased use in regulatory applications.

Critical inputs to CTMs include emission rates (allocated spatially and temporally by models and measurements), meteorology (from a model or measurements) and boundary and initial conditions (also from models and/or measurements). These inputs can have a major impact on the performance of the CTM and should be thoroughly evaluated (*4*), in addition to overall model performance. .

While there is no bright line test to see if a model is "acceptable" or "not acceptable" for a particular application, there are a number of model to measurement comparisons that are performed to help make a "weight of evidence" conclusion. For $O_3$and PM modeling, it is critical to have measurements of hourly $O_3$and daily PM (total and speciated by component – sulfate, nitrate, ammonium, OC, EC, crustal materials). In addition, it is beneficial to have measurements of precursor emissions (NOx and VOCs for ozone; SO2, NOx, NH3, and VOCs for PM). Model performance statistics are calculated for ozone and PM (total and speciated by component). Various statistical metrics have been used such as mean bias, mean error, mean normalized bias, mean normalized error, mean fractional bias, and mean fractional error. Tesche et al (*10*) considered past model evaluations and concluded that ozone mean bias should be less than ±15% and the mean normalized error should be less than 30%.

Table 1: Measurements Typically Used for Operational and Diagnostic Model Evaluations and Corroborating Measurements Used for "Weight-of-Evidence" Demonstration

Model Type	Species Typically Used in Model Evaluation	Additional Species used for further "weight-of-evidence" or diagnostic evaluation
Chemical Transport Model-Ozone	O_3	NO, NO_2, speciated VOCs, CO
Chemical Transport Model-PM	PM2.5-mass, sulfate, nitrate, ammonium, EC, OC, soils	Speciated VOCs, NH_3, HNO_3, SO_2, PM elements (Ni, V, Fe, Si, Mn, As, Se, etc.)

Given that one of the major objectives of model applications is to assess how air pollutant concentrations will respond to emission controls, dynamic model evaluation is of direct importance. Unfortunately, it is nearly impossible to observe a fully controlled dynamic scenario in nature. Instead, previous studies have relied on comparing simulating air quality prior to and after periods separated by substantial and quantifiable changes in the parameters of interest (e.g., intervention studies, (*11*)) while assuming that other parameters have less significant impacts on the model output metrics of interests. Such intervention studies typically stem from either large emission reductions over time due to both regulatory action and technological improvement, or temporary emissions reductions due to plant outages and the displacement of sources (see examples in (*11*)). Data requirements to support such studies are essentially the same as the requirements for operational evaluation, except that the measurement data must exist for both the "before" and the "after" cases, and be for a long enough period to address changes in meteorology. Another aspect of dynamic evaluation is the desire to have confidence in the model's ability to predict long-term trends in pollutant concentrations. While traditional surface measurement networks are frequently sufficient for such an undertaking, satellite retrievals of air pollutants and meteorological parameters can supplement trend analysis (*12*). Some intensive field program also have implemented a number of aircraft to capture pollutant data aloft, although expensive and intermittent such studies provide information not available from surface-based measurements (*13, 14*).

Discussion

If it appears that the measurements needed to evaluate air quality models are similar to those that are available from routine monitoring networks, it is because these networks, like the models, have evolved in response to the changing needs of the air quality community. As a specific example, it was recognized that just having PM mass was not sufficient to apply RMs or to appropriately evaluate, and thus confidently use (or improve) CTMs to quantify the impact of sources on particulate matter (*15*). Given the importance, the Species Trends Network (STN; (*16*)) was developed and has become central to model application and evaluation. Similarly, operational model evaluation has evolved with the networks to make the best use of the routinely available data.

Is this to say that the only measurements desired are those available from the routine monitoring networks? Unfortunately, that is not the case. Models have evolved to the point that they are now being evaluated by more detailed diagnostic evaluations using a much broader range of observations of species not generally available, but generally viewed as being of importance to the formation and fate of ozone and PM. For example, the Southern California Air Quality Study (SCAQS: (*17*)) was designed and conducted to provide a much more comprehensive set of data for evaluating and improving CTMs. This has been followed by numerous additional "intensive" studies worldwide (*18-21*), including the PM Supersite program (*22*) providing opportunity to further assess model capabilities and needs. The Electrical Power Research Institute (EPRI) has funded a number of detailed measurement studies, such as the South Eastern Aerosol Research and Characterization (SEARCH) network (*23*), that include measurements of hourly PM by species (SO4, NO3, NH4, OC, EC), additional precursors (e.g., SO2, NO, NO2, NOy, HNO3, CO, NH4), meteorological parameters, and deposition. While the surface measurements continue to grow, the one area that still needs to be addressed is aloft measurements. Aloft measurements from aircraft and radiosondes (*24*) are useful to evaluate model transport and give us additional confidence in surface results as well as to evaluate model processes above the surface.

Application and evaluation of the models using intensive study data, along with decades of more routine applications, has led to some confidence for routine policy applications. However, their use in "new" applications does not enjoy the same level of confidence. For example, in developing a new, multi-pollutant, secondary standard for nitrogen and sulfur oxides, CMAQ was used to simulate deposition of the multiple oxides of nitrogen and sulfur (*25*). A recognized concern by the Clean Air Scientific Advisory Committee (CASAC) was that the model had not been thoroughly evaluated for such an application. In part, the desired data is not available. Another concern is driven by the new 1-hour NAAQS for NO_2, which includes a focus on near-road concentrations. Application of models to near-road dispersion and reaction of NOx, and potentially other pollutants including PM, will be a new challenge, and confidence can be developed by continuing evaluation using not only NO_2 observations, but ozone and NO as well, and will be enhanced by a more extensive network of monitors beyond those typically available to assess attainment with the NAAQS. Particularly useful measurements that might be added to the current network would include temporally more highly or finely resolved PM species (e.g., EC/OC or black carbon; and ionic compounds (*14*)) that could challenge models' abilities to correctly simulate the diurnal variation, additional trace gases (e.g., nitric acid, ammonium) that are important for assessing the inorganic gas-particle partitioning and nitrogen/acid deposition. Organic PM speciation added at PAMS monitors would provide a growing data base to help assess SOA formation approaches.

Summary

Evaluation, both historical and to the specific application at hand, is required to use an air quality model with confidence. Much of the confidence the various models now enjoy has been developed over years of evaluations using a range of air quality and meteorological measurement data obtained from both routine networks and intensive monitoring programs. At this point, model evaluation for routine applications typically gets by with routinely available measurements, with increased confidence coming from using additional observations when available. However, routine measurements will not provide the information to advance the science in the models. Significant model advancement, and likewise significantly increased confidence in their use in new applications, will rely on more comprehensive, and tailored, measurement studies, including measurement data aloft.

Disclaimer. The United States Environmental Protection Agency through its Office of Research and Development collaborated in the research described here. It has been subjected to Agency review and approved for publication.

1. National Research Council, *Air Quality Management in the United States*. (National Academy Press, Washington DC, 2004).
2. R. D. Scheffe *et al.*, *J. Air Waste Manage. Assoc.* **59**, 579 (May, 2009).
3. N. Oreskes, K. Shraderfrechette, K. Belitz, *Science* **263**, 641 (Feb 4, 1994).
4. C. Seigneur *et al.*, *J. Air Waste Manage. Assoc.* **50**, 588 (Apr, 2000).
5. A. G. Russell, *J. Air Waste Manage. Assoc.* **58**, 289 (2008).
6. R. Dennis *et al.*, *Environ. Fluid Mech.* **10**, 471 (Aug, 2010).
7. D. Byun, K. Schere, *Applied Mechanics Reviews* **59**, 51 (2006).
8. T. W. Tesche *et al.*, *Atmos. Environ.* **40**, 4906 (Aug, 2006).
9. US Environmental Protection Agency, "Air Quality Modeling Final Rule Technical Support Document" (U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC 27711, 2011).
10. T. W. Tesche, P. Georgopoulos, F. L. Lurmann, P. M. Roth, "Improvement of Procedures for Evaluating Photochemical Models, Draft Final Report" (California Air Resources Board, Sacramento, CA., 1990).
11. van Erp AM, Kelly FJ, Demerjian KL, Pope CA, Cohen AJ, *Air Quality and Atmospheric Health*, (2011).
12. G. J. Frost *et al.*, *J. Geophys. Res.-Atmos.* **111**, (Jun 20, 2006).
13. P. A. Solomon, P. K. Hopke, *J. Air Waste Manage. Assoc.* **58**, S1 (2008).
14. P. A. Solomon, C. Sioutas, *J. Air Waste Manage. Assoc.* **58**, 164 (Feb, 2008).
15. National Research Council Committee on Research Priorities for Airborne Particulate Matter, *Research Priorities for Airborne Particulate Matter: III Early Research Progress*. (National Academy Press, Washington, DC, 2001).
16. P. A. Solomon *et al.*, *J. Air Waste Manag. Assoc.* **In Preparation**, (2011).
17. *Japca-the International Journal of Air Pollution Control and Hazardous Waste Management* **37**, 842 (Jul, 1987).
18. X. S. Wang *et al.*, *Atmos. Environ.* **45**, 4941 (Sep, 2011).
19. G. Li *et al.*, *Atmos. Chem. Phys.* **11**, 3789 (2011).
20. B. K. Pun *et al.*, *J. Air Waste Manage. Assoc.* **50**, 961 (Jun, 2000).
21. E. B. Cowling, W. L. Chameides, C. S. Kiang, F. C. Fehsenfeld, J. F. Meagher, *J. Geophys. Res.-Atmos.* **103**, 22209 (1998).
22. P. A. Solomon, P. K. Hopke, J. Froines, R. Scheffe, *J. Air Waste Manage. Assoc.* **58**, S3 (2008).

23. E. S. Edgerton *et al.*, *J. Air Waste Manage. Assoc.* **55**, 1527 (Oct, 2005).
24. P. Solomon, E. Cowling, G. Hidy, C. Furiness, *Atmos. Environ.* **34**, 1885 (2000, 2000).
25. R. Scheffe *et al.*, *EM* 16 (September, 2011).

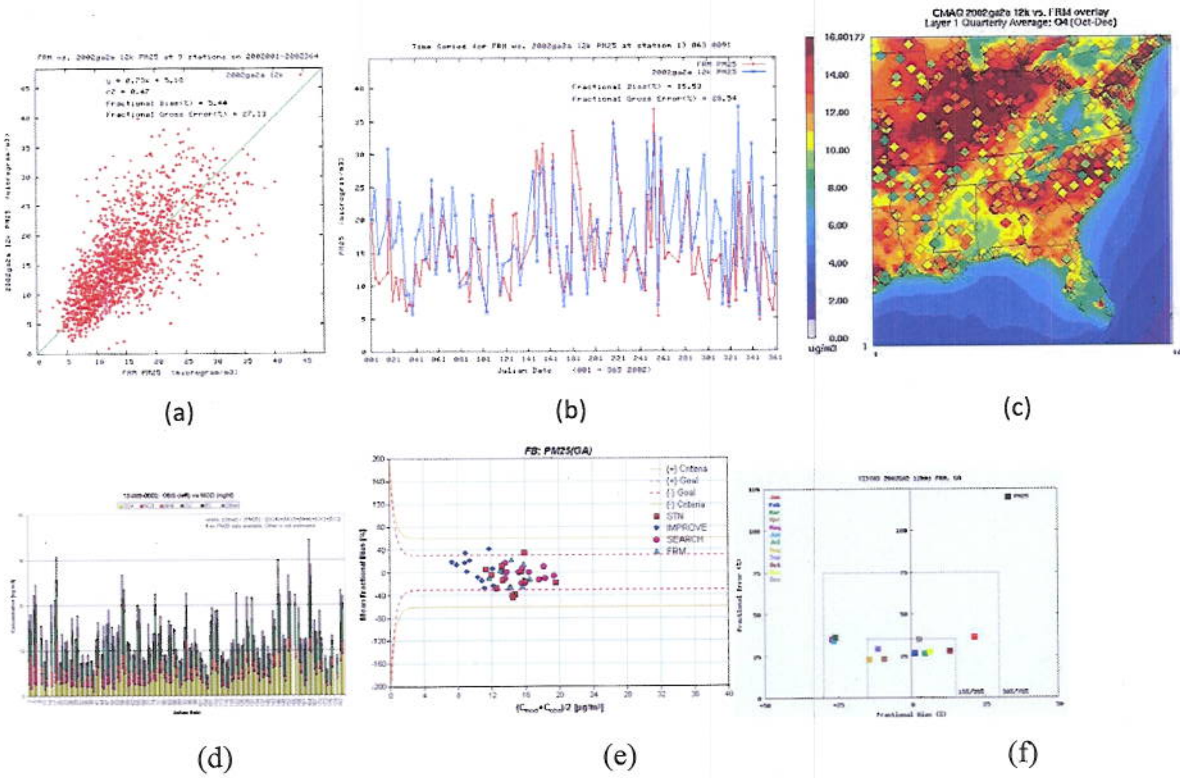

Figure 1. Example model performance plots: (a) PM scatter plot, (b) PM time series plot, (c) PM spatial plot (with observations overlaid), (d) PM stacked bar chart, (e) PM bugle plot showing concentration dependent goals (dotted line) and criteria (solid line), (f) PM soccer goal plot.