J10.1 THE EMERGENCE OF NUMERICAL AIR QUALITY FORECASTING MODELS AND THEIR APPLICATION

Kenneth Schere[1]*, Veronique Bouchet[2], Georg Grell[3], John McHenry[4], and Stuart McKeen[5]

[1]NOAA/ARL, Atmospheric Sciences Modeling Division, R.T.P., NC, on assignment to U.S. EPA, National Exposure Research Laboratory
[2]Meteorological Service of Canada, Dorval, Quebec, Canada
[3]NOAA/ESRL, Global Systems Division, Boulder, CO, on assignment from CIRES, University of Colorado
[4]Baron Advanced Meteorological Systems, Raleigh, NC
[5]NOAA/ESRL, Chemical Sciences Division, Boulder, CO, on assignment from CIRES, University of Colorado

1. INTRODUCTION

Over the past decade there has evolved an increasing interest by the public in the U.S. and other nations in the day-to-day air quality conditions to which they are exposed. Driven by the increasing awareness of the health aspects of air pollution exposure, especially by sub-populations most sensitive such as children and the elderly, short-term air pollution forecasts are being provided by more and more local authorities. Degradation of visibility in national parks and other pristine areas has also provided motivation for forecasts. Besides issuing alerts and warnings on air quality conditions, some local authorities are relying on air quality forecasts to put in place intermittent short-term management strategies such as free bus/rail fares, additional carpool strategies, burning bans, etc.

Efforts to produce these short-term (1-3 day) forecasts, usually of ozone (O_3), carbon monoxide (CO), fine particulate matter (PM2.5) and/or visibility rely on techniques ranging from persistence, to simple empirical local “rules-of-thumb”, to various statistical regression or neural network methods, to more complex models of the atmosphere. In a number of cases, nations or regions have launched efforts to use models that span large regional or national areas that can bridge the time and space scales between urban areas containing more dense air quality monitoring networks.

This paper discusses the emergence of national/regional numerical air quality forecast (NAQF) model systems based on three-dimensional grid (Eulerian) models, driven by mesoscale weather forecast models and source emissions models. We will focus on examples of such systems in the U.S. and Canada, although there are other notable forecast systems in place elsewhere. Recently a review of international air quality forecasting approaches (Cope and Hess, 2005) discussed example systems in Australia, Japan, Europe, and North America. They credit the availability of cost-effective high performance computers, advancements in high-resolution meteorological modeling, and the availability of real-time air quality monitoring data as having spurred the development of more sophisticated air quality model forecast systems. Most of the national efforts have been on regional-scale forecasting using models with horizontal grid sizes of 25-50 km, although model resolution is improving rapidly in many cases.

In the review done by Cope and Hess (2005) several Eulerian model systems are highlighted. In Europe such systems are used to produce regional forecasts by Meteo France (Europe/France), University of Cologne (Europe/Germany), Swedish Meteorological and Hydrological Institute (Europe), and Norwegian Institute for Air Research (Northern Hemisphere), among others. The Australian forecast system, one of the first to become operational, employs a nested grid system with 1-5 km grids over the target areas of Sydney and Melbourne, with twice-daily forecast cycles.

2. NORTH AMERICAN MODELING SYSTEMS

We present here several examples of NAQF modeling systems now in operation within the U.S. and Canada, including the system components, operational schedules, attributes, and outputs. While we use the terminology of “air quality forecast models” here, in practice these modeling systems provide numerical guidance for the operational forecaster to use in issuing the local forecast. The model guidance is combined with current monitoring data and the knowledge of local

* *Corresponding author address:* Kenneth Schere, U.S. EPA-E243-03, Research Triangle Park, NC 27711; e-mail: schere.kenneth@epa.gov

pollution patterns in creating the specific local air quality forecast.

2.1 *Eta-CMAQ Modeling System*

The Eta-CMAQ NAQF system was built as a partnership between NOAA's National Weather Service and Air Resources Laboratory, in collaboration with U.S. EPA, to provide model guidance initially for O_3 forecasts in the northeast U.S. The system components include NOAA's mesoscale Eta meteorological forecast model and EPA's Community Multiscale Air Quality (CMAQ) model (Otte et al., 2005; Byun and Schere, 2005). A CMAQ model preprocessor, PREMAQ, transforms the Eta meteorological fields to align with CMAQ's grid and coordinate system, and calculates pollutant deposition velocities and meteorologically-dependent pollutant emissions including those from mobile and biogenic sources. The mobile source emission procedure is based on a computationally-efficient grid-specific least-squares regression fit to the results of the comprehensive MOBILE6 emissions model results. The Sparse Matrix Operator Kernel Emissions (SMOKE) model (Houyoux et al., 2000) is used to produce all other emissions in an off-line manner using EPA's 2001 National Emissions Inventory (based upon NEI99) projected to the target forecast year.

Model domains and grid sizes: Operational model domain covers the eastern half of the U.S. with 12-km horizontal grid size. Pre-operational domain in testing covers the continental U.S. at same grid size.

Target air pollutants: O_3 in operational mode; O_3 and PM2.5 in pre-operational mode.

Initial/ boundary conditions: "Warm start" of CMAQ forecast runs obtain initial conditions from previous model cycle. Static climatological profiles of O_3 and other species are used along all boundaries, except O_3 data from NOAA's Global Forecast System are specified as lateral boundaries at CMAQ's top-most layer, representing stratospheric air influences.

Daily cycling: Model runs are performed by NOAA/NWS on IBM e-server mainframe computer using 65 processors. The current cycles include: 48-h CMAQ model run from 12 UTC initialization (for next-day forecast guidance); 48-h model run from 06 UTC initialization (for current-day update); and internal 6-h model runs at 18 and 00 UTC to provide continuity for warm start initial conditions.

Figure 1 provides an illustration of the Eta-CMAQ model forecast system results for estimated daily maximum of 8-hr average O_3 on August 4, 2005 from the 12 UTC model run on August 3.

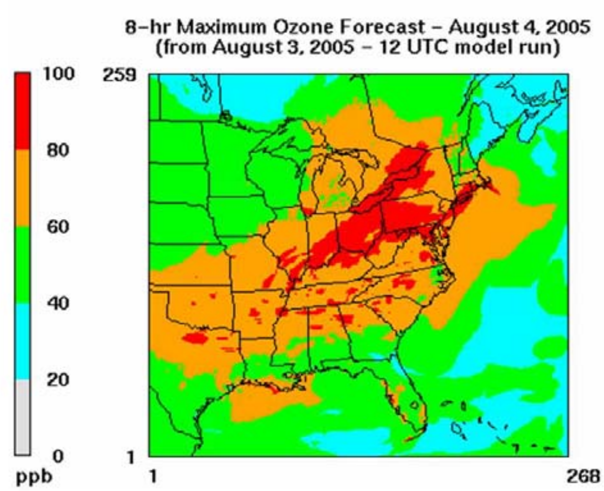

Figure 1. Lowest-layer model forecast estimates for 8-hr maximum O_3 concentrations for August 4, 2005 from Eta-CMAQ modeling system.

2.2 *MM5-MAQSIP-RT Modeling System*

The MM5-MAQSIP-RT air quality forecast system has evolved from initiatives from MCNC and Baron Advanced Meteorological Systems (BAMS) to provide customized numerical air quality forecast guidance to their clients. The system is composed of NCAR's MM5 mesoscale meteorological model and BAMS' Multiscale Air Quality Simulation Platform-Real Time (MAQSIP-RT; McHenry et al., 2004). The linkage between these models is provided by a custom-built MM5 coupling module (MCPL), within the MM5 model, that provides the required meteorological variables needed by the emissions model and the MAQSIP-RT model. The full SMOKE emissions model is part of the real-time modeling system, including the MOBILE5b mobile source emissions model and the BEIS3 biogenics model. Other emissions are projected to the target forecast year from EPA's NEI-99 base emissions inventory.

Model domains and grid sizes: Nested domains including the continental U.S. at 45-km resolution, 15-km resolution domain covering the eastern U.S., and several targeted metropolitan areas at 5-km resolution.

Target air pollutants: O_3 and PM2.5.

Initial/ boundary conditions: Warm start initial conditions are obtained from previous model run, typically 12-h old. Static climatological profiles of

O_3 and other species are used along all boundaries of the 45-km resolution domain. Boundaries of the inner nested domains are obtained from the outer nest.

Daily cycling: Model runs are performed by BAMS on various SGI servers using 4-24 processors. Continental U.S. 45-km grid, 120-h MAQSIP-RT model run from 18 UTC initialization (for next-day forecast guidance); 120-h model run from 06 UTC initialization (for current-day update). Eastern U.S. (15-km grid) domain is run for 30-42 h at 06 and 18 UTC, and the 5-km innermost nested domains are run for 24 h at 06 and 18 UTC.

Figure 2 provides an illustration of the MM5-MAQSIP-RT model forecast system results for maximum estimated 8-hr O_3 on August 8, 2005 for the 45-km grid continental U.S. domain.

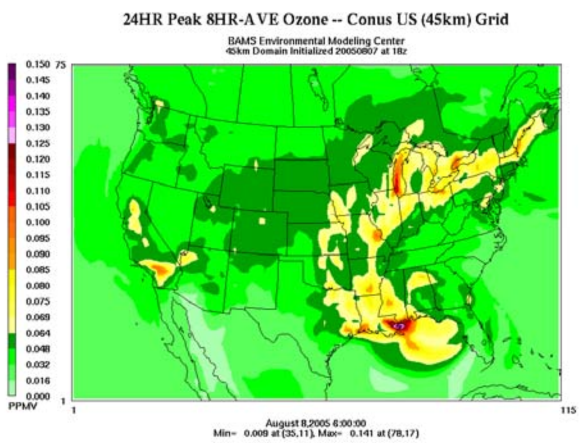

Figure 2. Lowest-layer model forecast estimates for daily maximum of 8-hr average O_3 on August 8, 2005 from the 12 UTC model run on August 7 from MM5-MAQSIP-RT modeling system.

2.3 *GEM-CHRONOS Modeling System*

The Canadian Hemispheric and Regional Ozone and NO_x System (CHRONOS) is the operational ozone air quality model of the Meteorological Service of Canada (MSC) for providing daily forecast guidance on a national and provincial scale. The CHRONOS model is based on the chemical-transport formulation of Pudykiewicz (1997), and has been in operation since 2001. Hourly meteorological data are provided to CHRONOS by the MSC Global Environmental Model (GEM), the principal mesoscale weather forecast model used in Canada, following an interpolation step to bring the meteorological fields onto the CHRONOS grid system. All anthropogenic emissions, including mobile emissions, are prepared by SMOKE off-line using climatological meteorology. Since June 2005, CHRONOS uses the 2000 Canadian and 2001 U.S. national emission inventories. Biogenic emissions are calculated on-line based on an early version of BEIS2.

Model domains and grid sizes: Operational domain covers North America with 21-km horizontal grid cell size.

Target air pollutants: O_3 and PM.

Initial/ boundary conditions: "Warm start" initial conditions are obtained from hour 24 of the previous day's forecast. Boundary conditions are treated via a zero-gradient scheme.

Daily cycling: Model runs are performed once each day with initialization at 00 UTC with a 48hr simulation duration. CHRONOS is run on MSC's Environment IBM supercomputer using 8 processors. Forecast guidance is available by 08 UTC.

Figure 3 provides an illustration of the GEM-CHRONOS model forecast system results for estimated O_3 on September 16, 2005 from the September 15, 2005, 00 UTC model run.

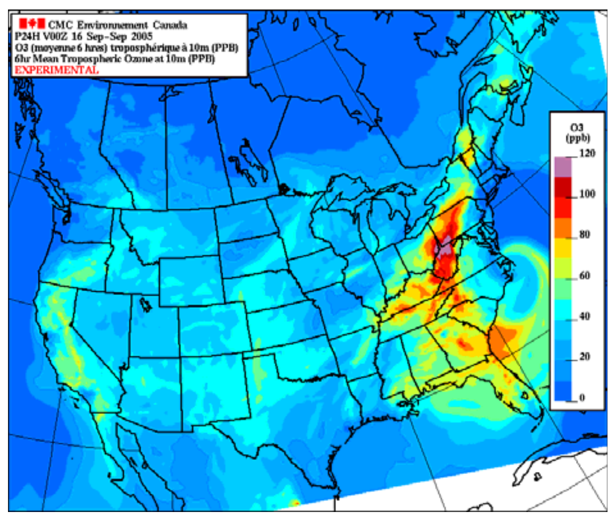

Figure 3. Lowest-layer model forecast estimates for 6-hr mean O_3 concentrations ending at 00 UTC on September 16, 2005 from GEM-CHRONOS modeling system.

2.4 *WRF-Chem Modeling System*

An emerging air quality model forecasting system is based upon the Weather Research and Forecasting (WRF) meteorological forecast model, a new product from the National Center for Atmospheric Research (NCAR). In a joint effort between NCAR and NOAA, on-line chemistry has been added to the WRF model to create WRF-Chem (Grell et al., 2005), an integrated model of atmospheric meteorology, physics, and chemistry. The on-line nature of this model system distinguishes it from the other "off-line" model systems profiled here. WRF-Chem is undergoing rigorous daily testing by NOAA and other groups in advance of operational deployment. The WRF model is an evolutionary successor to the MM5 mesoscale model and contains many of the same dynamics and physics options as MM5, along with a new computational infrastructure. Biogenic source emissions are estimated dynamically during model simulations with meteorological inputs from the WRF model. All other emissions are based on static data sets derived from the EPA NEI99 national emissions inventory.

Model domains and grid sizes: Model domains include North America at 40-km horizontal grid resolution and eastern half of U.S. at 27-km resolution.

Target air pollutants: O_3 and PM2.5.

Initial/ boundary conditions: Warm start of WRF-Chem forecast runs obtain initial conditions from previous model cycle. Static boundary condition profiles of O_3 and other species are based on averages of mid-latitude aircraft profiles from several field studies over the eastern Pacific.

Daily cycling: Model runs are performed twice each day with initializations at 00 and 12 UTC. Simulations are run forward for 36-hr. All model runs are performed on the NOAA Global Systems Division Ijet supercomputer, using up to 120 processors, depending upon the application. Figure 4 provides an illustration of the WRF-Chem model forecast system results for estimated O_3 at 21 UTC on August 4, 2005 from the 00 UTC model run on the same date.

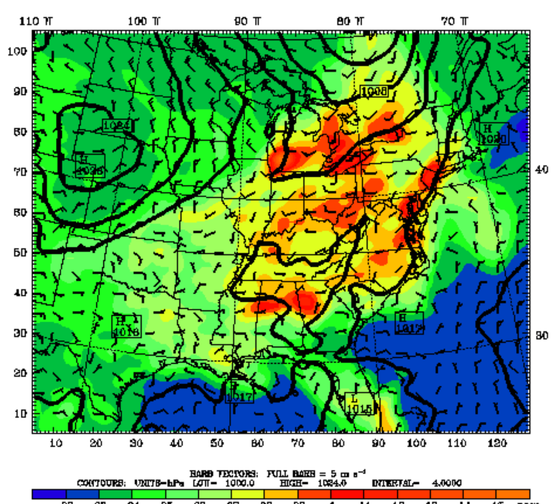

Figure 4. Lowest-layer model forecast estimates for 21 UTC on August 4, 2005 from WRF-Chem modeling system.

3. MODEL INTERCOMPARISONS

Real-time air quality forecasts were used as part of the field operational planning and later analysis in the International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) field study conducted over New England during the summer of 2004. McKeen et al. (2005) discuss the operational assessment of the air quality forecast results for O_3 across the eastern U.S. based on standard surface monitoring networks for 7 model forecast systems. The systems included the 4 discussed above (actually 5, as this study included both the 15- and 45-km versions of the MAQSIP-RT model), as well as two research models from MSC (AURAMS system) and the University of Iowa (STEM-2K3 system).

Each model was run on a daily basis by its home institution and results of the simulations were provided to NOAA. Each modeling system was run according to its home protocols in terms of meteorological and emissions inputs, initial/boundary conditions, and daily cycling. No attempts were made to normalize across modeling systems for data, model formulation, or operational procedure differences. Statistical evaluation was performed for the period 6 July through 30 September 2004 for daily values of maximum 8-hr average and maximum 1-hr average O_3.

Figure 5 illustrates the bias and root mean square error (ppb) for each modeling system, without identifying particular systems. The last set of bars in the graph represents an "ensemble" simple average of all 7 models. Note that this set of model results is not a true ensemble as used in other meteorological applications. Rather, the ensemble concept is used more loosely here to represent a collection of air quality forecast models in this application. Evaluation statistics were compiled for a spatial region that was a common subset of all modeling system domains. The two-month set of statistics show model biases ranging from 4 ppb to 27 ppb, with the ensemble in the midrange at 8-10 ppb. Results for rms errors were similar, with the ensemble results falling in the mid-range of the individual models.

As an extension of this study, a simple bias-correction technique was applied to each modeling system as well as the ensemble result. After applying the bias-correction, the rms errors were lower for the ensemble result than for any of the individual system results, providing perhaps some early indication of the usefulness of an ensemble approach in air quality forecasting.

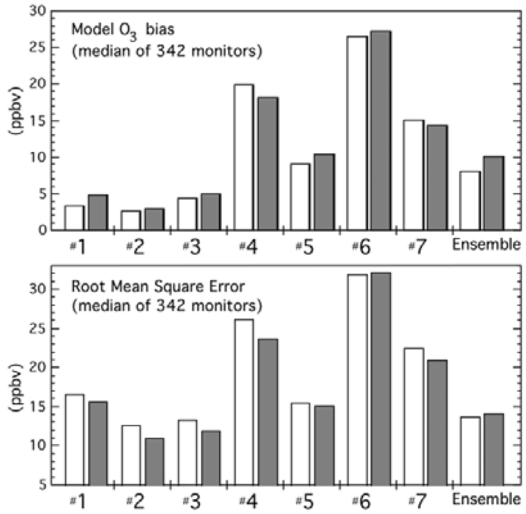

Figure 5. Model bias and root mean square errors for seven air quality models used in the ICARTT model intercomparison. The ensemble mean results from the seven models are also shown. White bars represent results for maximum 1-hr average O_3; shaded bars represent results for maximum 8-hr average O_3.

4. CURRENT AND FUTURE CHALLENGES

The NAQF systems profiled here have been in operations between 3 and 7 years for producing daily numerical guidance during the O_3 season. Over this period the model developers and evaluators have identified a number of issues they are currently working on.

The Eta-CMAQ system principal challenge is to link two models using quite different grid and coordinate systems together in a mass-consistent manner. The Eta uses a stepped-mountain vertical coordinate with a rotated latitude-longitude horizontal coordinate, while CMAQ typically uses a terrain-following sigma-p vertical coordinate and a Lambert-conformal horizontal coordinate. These different coordinate systems require a significant amount of spatial interpolation of the meteorological data before they are used by CMAQ. A mass continuity adjustment scheme has been included in the CMAQ model to adjust for any errors introduced by the spatial (and temporal) interpolations of the meteorological data.

Another current challenge has been to improve the specification of lateral inflow boundary concentrations for the CMAQ model. Generally climatological ozone profiles are used at the boundaries. However, O_3 concentrations from the NWS Global Forecast System (GFS) have recently been adapted to use in the upper portion of the boundary profile. Experience thus far has shown that the ozone data from the GFS model (no active photochemistry), which best represents the stratospheric burden, do not always improve the CMAQ ground-level forecasts. This issue is receiving additional study, including the possible increase of resolution near the tropopause.

One of the future directions for air quality forecasting is the use of on-line meteorological/chemical models, such as WRF-Chem. On-line models solve the issue of spatial/temporal interpolations of meteorological data for the chemical-transport model, since the chemistry is solved within the same executing model on the same coordinate system and grid as the meteorology. On-line models also allow for inclusion of radiative feedback effects associated with pollutant loading, on the modeled dynamics. Efficiency of the computations is one of the challenges for this type of model, although a single model run that produces a weather forecast and a “chemical weather” forecast at once is arguably an efficient system.

Estimation of real-time source emissions is another forecasting challenge. Efforts to date have focused on meteorological modulation of biogenic and mobile source emissions. However, other emission sources, such as electric utilities (variable power demand), surface coating (painting, asphalting, etc.), agricultural and construction activities, among others also are influenced by meteorology. Real-time representation of episodic air quality impacts of wildfires and dust storms also pose significant challenges. Building real-time components to handle these extensions to current emissions systems will be needed.

NAQF forecast guidance for ground-level O_3 is now operational, and the capability for PM2.5 is emerging. Other pollutant issues of interest that may have short-term forecast capabilities developed in the future include visibility/haze, CO, and possibly some air toxics species of interest such as benzene. Forecast durations should increase over time from next-day, to 1-3 day, to 7-day forecasts. Modeling domains now cover most of continental North America and we can expect grid resolution to increase significantly, to less than 10 km over the next decade.

As indicated by several modeling groups, issues of boundary condition specification are a significant issue. A logical extension from current capabilities would include the modeling of air quality on hemispheric to global scales to provide the large scale input to continental and regional models. There are several global air chemistry models being used now in research studies (e.g., GEOS-Chem, MOZART) and they have been linked to regional-scale models for assessments. Operational forecast centers may include future extensions of air chemistry in their global systems. Fully nested global to regional air quality systems are another potential future development. There are already plans to develop a global WRF model, making global to regional nested meteorological forecast guidance available.

Lastly, perhaps the greatest challenge to the advancement of air quality forecast systems is the development of chemical data assimilation systems for model system initialization, and the establishment of an international chemical profiling capability to feed the data assimilation system. The history of numerical weather prediction shows that the greatest advancements were made after the establishment of such monitoring and assimilation infrastructures. An international program know as the Global Earth Observation System of Systems (GEOSS; www.noaa.gov/eos.html) has begun the discussions to establish coordinated earth monitoring systems, including those for air quality, based on satellite, aircraft, profilers, and surface-based observations. Research studies have shown early promise in developing 3-D and 4-D variational adjoints of air quality modeling systems. These are components of data assimilation systems. The inclusion of chemical data assimilation for model initialization at global and regional scales combined with the development of bias correction techniques and model output statistics as NAQF systems mature, should make for increasingly accurate and useful forecast guidance for air pollutants. These advancements are dependent upon the evolution of data collection, archival, and dissemination activities such as those proposed through the GEOSS program.

ACKNOWLEDGMENTS

Special thanks go to contributions from Don Olerud, Jeff Vukovich, Ted Smith, Carlie Coats, and Bob Imhoff (Baron Advanced Meteorological Systems). The research presented here was performed, in part, under the Memorandum of Understanding between the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) and under agreement number DW13921548. This work constitutes a contribution to the NOAA Air Quality Program. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.

REFERENCES

Byun, D.W. and K.L. Schere, 2005: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Reviews* (in press).

Cope, M. and D. Hess, 2005: Air quality forecasting: A review and comparison of the approaches used internationally and in Australia. *Clean Air and Environmental Quality*, **39**, 52-60.

Grell, G.A. and Co-authors, 2005: Fully coupled "online" chemistry within the WRF model. *Atmos. Environ.*, **39**, 6957-6975.

Houyoux, M.R., and Co-authors, 2000: Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SMRAQ) project. *J. Geophys. Res.*, **105**, 9079-9090.

McHenry, J.N. and Co-authors, 2004: A real-time Eulerian photochemical model forecast system: Overview and initial ozone forecast performance in the northeast U.S. corridor. *Bull. Amer. Meteor. Soc.*, **85**, 525-548.

McKeen, S. and Co-authors, 2005: Assessment of an ensemble of seven real-time ozone forecasts over Eastern North America during the summer of 2004. *J. Geophys. Res.* (submitted)

Otte, T. and Co-authors, 2005: Linking the Eta model with the Community Multiscale Air Quality (CMAQ) modeling system to build a national air quality forecasting system. *Wea. Forecasting*., **20**, 367-384.

Pudykiewicz, J., A. Kalleur, and P.K. Smolarkiewicz, 1997: Semi-Lagrangian modeling of tropospheric ozone. *Tellus*, **49B**, 231-258.