7.1

PREMAQ: A NEW PRE-PROCESSOR TO CMAQ FOR AIR-QUALITY FORECASTING

Tanya L. Otte[1], George Pouliot[1], and Jonathan E. Pleim[1]
NOAA Air Resources Laboratory, Research Triangle Park, North Carolina
e-mail: otte.tanya@epa.gov
Web address: http://www.epa.gov/asmdnerl
Voice: 919-541-7533 Fax: 919-541-1379

[1]On assignment to the National Exposure Research Laboratory, U.S. EPA, Research Triangle Park, NC

1. INTRODUCTION

On 6 May 2003, NOAA and the U.S. EPA signed a Memorandum of Understanding and a Memorandum of Agreement to expand their collaboration toward the development of a national air-quality forecasting (AQF) system. NOAA and the U.S. EPA are jointly developing various components of the AQF system, which will be run operationally at the NOAA's National Weather Service. Currently, the AQF system is based on the National Centers for Environmental Prediction's (NCEP's) Eta Model (Black 1994; Rogers *et al*. 1996) and the U.S. EPA's Community Multiscale Air Quality (CMAQ) Modeling System (Byun and Ching 1999). The Eta/CMAQ AQF system is designed to provide twice-daily numerical guidance for air-quality forecasts for the U.S. Initially, the capability includes surface-level O_3 forecasts of hourly average concentrations as well as 8-h averages. Future expansion of the system will include $PM_{2.5}$. The current forecast domain is the Northeast U.S. using a 12-km horizontal grid spacing on a Lambert Conformal map projection. Ultimately the forecast domain will be expanded to cover the continental U.S., and the horizontal grid spacing may decrease in conjunction with changes to the Eta Model and successor mesoscale models. The AQF system is targeted to become fully operational for O_3 predictions for the Northeast U.S. in September 2004.

A new pre-processor to CMAQ (PREMAQ) has been developed as part of the AQF system. PREMAQ combines the functionality of the Meteorology-Chemistry Interface Processor (MCIP) (Byun and Ching 1999) and parts of the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system (Houyoux *et al*. 2000) in a single real-time processor. PREMAQ was specifically designed to link NCEP's Eta Model with CMAQ, and it uses meteorological fields that are specific to the Eta Model output suite. The following is a description of the technical details of linking the Eta Model to CMAQ with PREMAQ in the national AQF system from Otte *et al*. (2004).

2. METEOROLOGY

The meteorology input for the national AQF system is based on output from NCEP's Eta Model, which is currently run four times daily for North America at a horizontal grid spacing of 12 km. The operational domain for the Eta Model covers the continent of North America. Ideally, CMAQ would use the same horizontal grid spacing and staggering, the same map projection, and the same vertical coordinate as the meteorological fields from the Eta Model to maintain mass consistency. However, the Arakawa E-grid (Arakawa and Lamb 1977) and the rotated latitude-longitude map projection that are used by the Eta Model are not currently supported by CMAQ, which uses the Arakawa C-grid and a number of other map projections. Also, although CMAQ uses a generalized vertical coordinate to facilitate close coupling with meteorology models (Byun 1999), the Eta Model's "step mountain" vertical coordinate would be especially difficult to replicate in the CMAQ model.

Another option for coupling the Eta Model with CMAQ would be to use data files that are already created by NCEP. Eta Model output is routinely disseminated to operational customers on various predefined GRIdded Binary (GRIB) domains with data written at three-hourly intervals and typically on pressure surfaces. CMAQ makes use of prognostic variables (e.g., mixing depth, land-use category, plant canopy water, canopy conductance) that are generally unavailable in the routinely available Eta Model output. Also, the pressure coordinate in the widely available Eta Model output files is not easily converted to the generalized coordinate system. Further, CMAQ requires meteorological fields at a minimum of hourly intervals. Thus, using available Eta Model output clearly would not provide close spatial and temporal coupling between the two modeling systems.

As a compromise, coupling between the Eta Model and CMAQ in the AQF system involves horizontal and vertical interpolation of the Eta Model output to a grid structure that is typically ingested by CMAQ. Using a series of post-processors, the Eta Model output is recast onto a hydrostatic sigma-P vertical coordinate structure and a Lambert conformal map projection of an Arakawa-C staggered horizontal grid with 12-km horizontal grid spacing. In addition, NCEP is creating hourly Eta Model forecast output to 48 h for the AQF system on the CMAQ forecast domain, and providing additional output variables post-processed from the Eta Model prognostic algorithms. This approach is advantageous because it did not entail any modifications to the Eta Model or to CMAQ, and it minimally impacted the operational suite at NCEP. To mitigate the effects of horizontal and vertical interpolation used in the Eta Model post-processing, a robust mass-correction algorithm (R. Yamartino, personal communication, 2002) was implemented in CMAQ to conserve air chemical species.

In the current implementation, forecast data from the 60 Eta Model layers are interpolated to 22 hydrostatic sigma-P layers for CMAQ using a modified version of the Eta Model Post-Processor. The full horizontal extent of the Eta Model forecast domain is included in the processing. The 22-layer configuration features most of the layers in the lower troposphere within the planetary boundary layer, where most of the photochemical activity takes place that is important for surface (and near-surface) O_3 generation. There are approximately 12 layers below 2 km AGL, and the lowest layer thickness is about 39 m. The 22 hydrostatic sigma layers are converted to the generalized vertical coordinate that is used in the CMAQ chemistry transport model. The geopotential height is interpolated to the layer interfaces (i.e., 23 levels) using virtual temperature, and that height is used to derive the temperature hydrostatically from thickness so that temperature and height are in hydrostatic balance. Where necessary, the Eta Model Post-Processor diagnoses additional forecast variables for use in the AQF system. Each hourly output file from this procedure contains about 400 MB of data.

The hydrostatic sigma data on the full horizontal extent of the Eta Model domain are then subset to the CMAQ forecast domain plus the lateral boundaries. NCEP's Product Generator software is used to perform the bilinear interpolations and nearest-neighbor mappings of the Eta Model Post-Processor output from Eta Model horizontal forecast domain to the CMAQ forecast domain. The Product Generator is used for the AQF system in largely the same form as it is used to generate all other widely available versions of the Eta Model output. The suite of input and output variables (and vertical surfaces) and the temporal frequency of the output are different from most other distributed configurations of the same Eta Model forecast. The Product Generator strictly performs grid-to-grid interpolations for a subset of the variables available in the Eta Model output files; no new variables are created in the Product Generator. The Product Generator produces a set of hourly output files that contain about 9 MB of data each, which represents a significant reduction in the size of the data files, largely due to extracting data from a small subset of the Eta Model forecast domain. In addition, the Product Generator places all variables on the unstaggered Arakawa A-grid and in GRIB format, which are standard procedures in the Eta Model post-processing.

The final step in preparing the meteorology for input is done through the new software preprocessor for CMAQ, PREMAQ. PREMAQ is based on algorithms from MCIP, but it also includes emissions processing in the AQF system. The purposes of PREMAQ are to put the Eta Model forecast fields onto the horizontal and vertical structure that CMAQ expects, to create all of the atmospheric state variables that are required in the chemistry transport model, and to perform time-dependent emissions calculations. PREMAQ does not perform any horizontal interpolation other than to translate from the Arakawa A-grid to the Arakawa C-grid. The vertical coordinate is converted from the hydrostatic sigma structure to the generalized coordinate, but no vertical interpolation is performed on the three-dimensional fields. PREMAQ calculates air density and Jacobian (i.e., a time-dependent function of surface pressure, air density, and gravity in the hydrostatic sigma coordinate system), which are used as state variables in the mass-conserving CMAQ general equations. PREMAQ also computes the dry deposition velocities for various photochemical species. Unlike MCIP, the meteorologically dependent emissions processing is contained in PREMAQ to streamline the operational pipeline and minimize input/output requirements. In the initial implementation of the AQF system, PREMAQ is run as a serial stand-alone pre-processor.

3. EMISSIONS

Ideally, near real-time emissions data would be used in the AQF system. However, real-time collection and transmission of emissions data from states to the U.S. EPA and then to NOAA does not currently exist and is therefore unavailable for AQF. As an alternative, the U.S. EPA's national emissions inventory (NEI; see http://www.epa.gov/ttn/chief/net/index.html) contains historical data for mobile sources (e.g., vehicular traffic), stationary area and point sources (e.g., power plants), and natural and agricultural sources (e.g., wild fires and animal operations). The NEI is generally updated tri-annually, but modifications can be made to project changes in emissions for any area and year. Some emissions data can be predefined based on the historical emissions patterns, while others must be set using criteria from the specific forecast day. Also, biogenic emissions depend strongly on meteorological factors such as temperature and solar insolation. These factors must be considered on a day-to-day basis to provide input for CMAQ.

The input emission data for the AQF system are based on area, non-road, and point source inventories from the U.S. EPA 2001 NEI with some enhancements to better estimate the emissions for the current year. In particular, point source emission estimates include regional adjustments to the nitrogen oxides (NO_x) emissions based on projected energy usage (U.S. DOE 2004), and the 2002 commercial cooking inventory was also included. The 1995 Canadian and 1999 Mexican Big Bend Regional Aerosol and Visibility Observational Study (BRAVO) emission inventories were combined with the 2001 NEI to form the continental data set. The Vehicle Miles Traveled (VMT) data from the 1999 NEI is used in the AQF system as background for the mobile source emission calculations.

The processing of the emissions data for the AQF system has been adapted from SMOKE. SMOKE uses sparse-matrix algorithms to efficiently generate emission files required by air quality models such as CMAQ. This approach permits rapid and flexible processing of emissions data, both of which are critical in an operational environment. The processing steps of chemical speciation, temporal allocation, and spatial allocation in SMOKE are separated into independent operations. The results from these steps are merged

together at a final stage of processing using vector-matrix mathematics.

For the AQF system, the emissions processing is divided into two components: calculation of emissions fields that are independent of meteorological fields (and can be made available *a priori*), and calculation of emissions fields that are dependent on meteorology. The emissions processing that is independent of the meteorological fields is pre-computed outside of the AQF system using the SMOKE model and stored in static files that can be ingested on the appropriate day. The emissions processing that is dependent on the meteorological fields is integrated into PREMAQ so that the emissions can be calculated on an hourly basis in real time.

Emissions from area sources (e.g., agricultural fields, large open mining operations, forests, or aggregates of closely spaced point sources such as residential housing) are assumed to vary in predetermined spatial and temporal patterns that can be calculated in advance for any day of the year. Therefore, area source emissions are pre-calculated from the EPA inventory and stored as static files for the AQF system.

Emissions from point sources (e.g., industrial stacks) are assumed to have a predetermined temporal variability. However, since point sources are released at different heights in the atmosphere (depending on the height of the stack) usually as a heated gas, a plume rise algorithm is required to calculate the top and bottom heights of the plume. Plume rise is calculated from Eta Model forecasts of vertical and horizontal wind components, air temperature, heat flux, mixing height, and water vapor mixing ratio. Given these, the plume is partitioned into each of the model layers intersecting the plume based on the pressure in each layer. Only the plume rise that occurs in the vertical column of cells at the horizontal location of the stack is calculated in PREMAQ; the calculation of plume transport to neighboring model grid cells is done in CMAQ.

Biogenic emissions (e.g., hydrocarbon emissions from vegetation and nitric oxide emissions from soils) are highly dependent on meteorological fields. Therefore, the Biogenic Emissions Inventory System version 3 (BEIS3) (Pierce *et al*. 1998; Pierce *et al*. 2002) is directly integrated into PREMAQ. The biogenic emissions are calculated using the Eta Model forecasts of solar radiation, surface air temperature, surface pressure, and 24-h accumulated rainfall.

Mobile source emissions processing depends on temperature, vehicle activity, and vehicle fleet information. A highly detailed emissions model of on-road sources (MOBILE6) has been developed by the U.S. EPA (2003), and it has also been integrated into SMOKE. However, the SMOKE implementation of MOBILE6 is computationally expensive and inefficient for use in the AQF system. Therefore, an efficient method for estimating mobile source emissions based on SMOKE and MOBILE6 has been developed for the AQF system. First, mobile source emissions are computed using SMOKE and MOBILE6 with temperature fields from a previous time period. Using these data, a relationship between the mobile source emissions and the temperature was obtained using a nonlinear least squares fit of the emission data for each grid cell in the forecast domain, each chemical species, each hour of the day, and each day of the week. This level of detail is necessary because of the complex assumptions built into the MOBILE6 model, especially with respect to vehicle activity, time of day, and day of week. In addition, separate coefficients are calculated for the evaporative and exhaust components of the emission estimates. A quadratic function was used to fit the emission data to the temperature data. The coefficients calculated using the nonlinear least squares fit are saved for each grid cell, each chemical species, each hour of the day, and each day of the week. These coefficients are used with the Eta Model forecast temperature field to calculate the mobile source emissions in the AQF system. Only this final calculation for mobile source emissions is included in PREMAQ since the coefficients can be determined *a priori*. This calculation is very efficient, and it generates mobile source emission estimates that are highly correlated with using the complete SMOKE and MOBILE6 method. However, this method is not a replacement for MOBILE6 since it depends entirely on the detailed information contained in MOBILE6.

At the conclusion of the PREMAQ processing, emissions from area, point, biogenic, and mobile sources are combined to form a single 3-D gridded representation of the emissions for various photochemical species. This emissions file reflects hourly meteorological variations, as well as climatological and seasonal effects.

4. DRY DEPOSITION

Chemical dry deposition velocities are computed in PREMAQ using an electrical resistance analog model (Pleim *et al*. 2001). Atmospheric and boundary layer resistances are based on atmospheric surface layer parameters from the Eta Model (e.g., friction velocity and surface heat exchange coefficient). Canopy resistance is a parallel combination of surface resistances (leaf cuticle and ground) and stomatal resistance. The bulk stomatal resistance is derived from the moisture canopy conductance from the Eta Model. Surface resistances are scaled by solubility and chemical reactivity of each chemical species. Several surface parameters from the Eta Model (e.g., leaf area index, fractional vegetation coverage, canopy water content, and roughness length) are also used in the surface resistance calculations. These surface parameters often have strong ties to the land-use (or vegetation type) database that is used in the Eta Model. The forecast dry deposition velocities are projected to improve when a higher-resolution land-use database is incorporated into the Eta Model.

5. ANTICIPATED FUTURE MODIFICATIONS

As the modeling domain becomes larger and horizontal grid spacing decreases, the processing time for PREMAQ becomes more of a computational burden. Meeting the operational timelines at NCEP is important,

and PREMAQ has begun to take proportionally too much of the AQF system's operational time window. Therefore, PREMAQ will undergo software parallelization with the hope of using approximately four processors to perform the meteorology and emissions tasks. In addition, since the Weather Research and Forecast (WRF) Model is scheduled to *replace* the Eta model at NCEP in October 2005, a linkage between WRF (NCEP's "NMM" core) and CMAQ will need to be in place for the 2005 forecast season to prepare for the 2006 forecast season. To that end, a tighter off-line coupling of CMAQ to the WRF-NMM grid staggering and projection is also envisioned.

Disclaimer: *The research presented here was performed 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*. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.*

6. REFERENCES

Arakawa, A., and V. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Comput. Phys*., **17**, 173-265.

Black, T., 1994: The new NMC mesoscale Eta Model: description and forecast examples. *Wea. Forecasting*, **9**, 265-278.

Byun, D. W., 1999: Dynamically consistent formulations in meteorological and air quality models for multiscale atmospheric studies. Part I: Governing equations in a generalized coordinate system. *J. Atmos. Sci*., **56**, 3789-3807.

Byun, D. W., and J. K. S. Ching, Eds., 1999: Science algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. EPA-600/R-99/030, Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C. [Available from U.S. EPA, ORD, Washington, D.C. 20460.]

Houyoux, M. R., J. M. Vukovich, C. J. Coats, Jr., N. M. Wheeler, and P. S. Kasibhatla, 2000: Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SMRAQ) project. *J. Geophys. Res*., **105**, 9079-9090.

Otte, T. L., and Coauthors, 2004: Linking the Eta Model with the Community Multiscale Air Quality (CMAQ) Modeling System to build a national air quality forecasting system. Submitted, *Wea. Forecasting*.

Pierce, T., C. Geron, L. Bender, R. Dennis, G. Tonnesen, and A. Guenther, 1998: Influence of increased isoprene emissions on regional ozone modeling. *J. Geophys. Res*., **103**, 25611-25629.

Pierce, T., C. Geron, G. Pouliot, E. Kinnee, and J. Vukovich**,** 2002: Integration of the Biogenic Emissions Inventory System (BEIS3) into the Community Multiscale Air Quality Modeling System. Preprints, *12th Joint Conf. on the Appls. of Air Pollut. Meteor. with the A&WMA*, Norfolk VA, Amer. Meteor. Soc., J85-J86.

Pleim, J. E., A. Xiu, P. L. Finkelstein, and T. L. Otte, 2001: A coupled land-surface and dry deposition model and comparison to field measurements of surface heat, moisture, and ozone fluxes. *Water Air Soil Pollut*., **1**, 243-252.

Rogers, E., T. Black, D. Deaven, G. DiMego, Q. Zhao, M. Baldwin, N. Junker, and Y. Lin, 1996: Changes to the operational "early" Eta Analysis/Forecast System at the National Centers for Environmental Prediction. *Wea. Forecasting*, **11**, 391-413.

U.S. DOE, 2004: Annual Energy Outlook 2004 with Projections to 2025. DOE/EIA-0383, Energy Information Administration, Washington DC. [Available online at http://www.eia.doe.gov/oiaf/aeo/index.html.]

U.S. EPA, 2003: User's guide to MOBILE6.1 and MOBILE6.2 (Mobile Source Emission Factor Model). EPA420-R-03-010. Office of Transportation and Air Quality, Ann Arbor, MI, 262 pp. [Available online at http://www.epa.gov/otaq/m6.htm.]