THE VALUE OF NUDGING IN THE METEOROLOGY MODEL FOR RETROSPECTIVE CMAQ SIMULATIONS

Tanya L. Otte*
NOAA Air Resources Laboratory, Atmospheric Sciences Modeling Division, RTP, NC, USA
(In partnership with the U.S. EPA National Exposure Research Laboratory)

1. INTRODUCTION

Using a nudging-based data assimilation approach throughout a meteorology simulation (i.e., as a "dynamic analysis") is considered valuable because it can provide a better overall representation of the meteorology than a pure forecast. Dynamic analysis is often used in the Pennsylvania State University/National Center for Atmospheric Research Mesoscale Model (MM5) to generate multi-day meteorology simulations that are used as background for the Community Multiscale Air Quality (CMAQ) Modeling System when CMAQ is used for retrospective research and for regulatory applications. The Weather Research and Forecasting (WRF) Model has been publicly available for several years as the next-generation model to succeed MM5. However, many CMAQ users, especially those who do not have forecasting-only applications, have not gravitated from MM5 to WRF, and the prominent explanation is that the nudging is not available yet in WRF. The initial release of a nudging capability for WRF is scheduled for fall 2006. Because the influence of the input meteorology fields (e.g., from MM5 or WRF) on the chemistry model simulation (e.g., using CMAQ) is significant, it may be considered intuitively obvious that using nudging to provide a dynamic analysis will lead to an improved air quality simulation. However, the penalty if nudging is not used in the meteorology model has not yet been quantified. This paper provides preliminary insights into the value to the chemistry model simulation of using nudging-based data assimilation for dynamic analysis in the meteorology fields that are input to CMAQ.

2. BACKGROUND AND MOTIVATION

For two decades, Eulerian (or gridded) chemistry models have been driven by meteorological fields that are generated by Eulerian meteorology models such as MM5, in part, because meteorological observations and archived forecast fields do not exist at a high enough temporal and spatial resolution to capture atmospheric features (e.g., mixing depth, column temperature and wind profiles) that have been considered important for near-surface regional-scale chemical transport modeling. Models such as MM5 have been able to bridge the gap by providing fields at the desired resolution. The accuracy of the meteorology fields from MM5 has also been improved in retrospective simulations by creating dynamic analyses by using Newtonian relaxation throughout the simulation period. The dynamic analysis has also extended the length over which MM5 fields could be created and used without reinitializing to a period of several days.

Newtonian relaxation or "nudging" (Stauffer and Seaman 1990, 1994) is one method of four-dimensional data assimilation that is implemented in MM5 and will be released in WRF. Nudging involves adding an artificial forcing term to the equations of motion that reflects the difference between the observed state and the model state at a given location and time. The nudging term is weighted by a coefficient that is selected so that its reciprocal value represents the *e*-folding time over which the model error will be reduced in the absence of any other model forcing, and it is at least one order of magnitude smaller than the dominant terms in the equations. Nudging can be applied to wind, temperature, and water vapor mixing ratio in any combination and with independent nudging coefficients in MM5 and in WRF. Nudging can be accomplished in MM5 by using either gridded analyses where there is a "true" observed state at each model grid point (i.e., "analysis nudging") or by using high-frequency and/or high-density observations as they occur in space and time (i.e., "observation nudging").

When nudging is used in MM5 to create input for CMAQ, it is assumed that the improvements that are seen in MM5 with nudging will translate into a corresponding improvement in the CMAQ simulation. The magnitude of the impact of nudging in MM5 on the CMAQ simulations, however, has not yet been quantified. Understanding the value of nudging in the

**Corresponding author:* Tanya L. Otte, EPA/NERL/AMD, 109 T.W. Alexander Dr., Mail Drop E243-03, Research Triangle Park, NC 27711; e-mail: otte.tanya@epa.gov

meteorology model to the CMAQ simulation can be important to identify sensitivities and to define how errors in meteorological fields impact pollutant fate and transport. In addition, there could be implications for defining the optimal simulation length for the meteorology model as well as optimal air-quality forecast periods using CMAQ. Furthermore, this analysis can help to improve the application of nudging in the meteorology model to create dynamic analyses, and it can help to focus areas of improvement in meteorological modeling to support air-quality applications.

3. EXPERIMENT DESIGN

For this work, the meteorology, emissions, and chemistry modeling suite is run for two different configurations of input meteorology: one that uses analysis nudging (i.e., a dynamic analysis), and one that does not (i.e., effectively, a forecast). The simulations are performed on a horizontal domain with 36-km horizontal grid spacing that includes the continental United States and parts of Canada and Mexico (cf. Eder and Yu 2006). Thirty-four terrain-following layers are used for both the meteorology and chemistry simulations, and eighteen layers are in the lowest 2 km of the atmosphere.

MM5 (Grell et al. 1994) version 3.6 is used for the meteorology simulations. The background fields and lateral boundary conditions for MM5 originate from the National Centers for Environmental Prediction's North American Mesoscale Model (i.e., for this period, the Eta Model; Black 1994) 3-h analyses. The Eta Model analyses are interpolated to the MM5 domain and reanalyzed with surface and upper-air observations. The physics options used in MM5 in this study include the Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) for longwave radiation, the Dudhia shortwave radiation scheme (Grell et al. 1994), the Kain-Fritsch 2 convective model (Kain 2002), the Reisner 2 microphysics parameterization (Reisner et al. 1998), the Asymmetric Convective Model (ACM) for the planetary boundary layer (PBL) (Pleim and Chang 1992), and the Pleim-Xiu land-surface model (Xiu and Pleim 2001; Pleim and Xiu 2003).

In the MM5 simulation that includes nudging, 3-h 3D analyses of temperature, water vapor mixing ratio, and horizontal wind components are used with nudging coefficients of 3.0×10^{-4} s^{-1}, 1.0×10^{-5} s^{-1}, and 3.0×10^{-4} s^{-1}, respectively. Three-hourly surface analyses of horizontal wind components are also used with a nudging coefficient of 3.0×10^{-4} s^{-1}. There is no nudging of mass fields within the PBL following Stauffer et al. (1991).

The emissions are based on the U.S. Environmental Protection Agency 2001 National Emission Inventory. The emissions are processed using the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system (Houyoux et al. 2000) version 2.2. Mobile source emissions are processed with the MOBILE6 (EPA 2003) model within SMOKE. The biogenic emissions are processed using the Biogenic Emissions Inventory System, version 3 (BEIS3; Pierce et al. 1998). For this work, the biogenic and point-source emissions sectors are reprocessed for each MM5 simulation to capture the effects of the hourly meteorology on the emissions. All other emissions sectors are independent of the MM5 simulations.

The chemistry transport is modeled using CMAQ (Byun and Schere 2006) version 4.6. The 2005 update to the Carbon Bond chemical mechanism (CB05; Yarwood et al. 2005) is used. The PBL is modeled using the ACM version 2 (ACM2; Pleim 2006). The fourth version of the modal aerosol model ("AERO4"; Binkowski and Roselle 2003) is used for aerosol chemistry. Chemical dry-deposition velocities are computed using an electrical analog resistance model ("M3DRY"; Pleim et al. 2001). The chemistry lateral boundary conditions are prepared from a global simulation using the GEOS-CHEM model (Bey et al. 2001).

The MM5 simulations are run for the period 12 UTC 19 Jun – 00 UTC 4 Aug 2001. The period is broken into overlapping 5.5-day run segments. The first 12 h of each MM5 segment are a "spin-up" period for cloud processes, and they are not used for emissions or chemistry processing; the remaining five days are input for the chemistry model. All fields except soil moisture are reinitialized in each MM5 segment, as is typically done for regulatory modeling applications using CMAQ. The CMAQ simulations cover the period 00 UTC 20 Jun – 00 UTC 4 Aug 2001, but the first ten days are considered "spin up" to allow the chemistry to come into equilibrium, and they are not used in the analysis.

4. RESULTS

Two sets of MM5 and CMAQ simulations ("Nudge" and "NoNudge") for the 35-day period from 30 Jun – 3 Aug 2001 are analyzed to assess the impact of using nudging in MM5 on the CMAQ simulation. The 35-day period includes seven different MM5 run segments. Because the skill of the meteorology models degrades over time, particularly in the absence of nudging, the model performance (MM5 and CMAQ) is aggregated over time slices within each 5.5-day MM5 run segment (Table 1). As the first 12 h of each MM5 simulation is not used, "Day 1" refers to hours 13-36 of the MM5 simulation, "Day 2" refers to hours 37-60, and so on. The CMAQ performance is binned in time, as well, to determine the impact on the chemistry transport model as it corresponds to degraded MM5 model performance.

Table 1. Dates in 2001 used for analysis as given by time elapsed within each MM5 run segment.

MM5 Segment	Day 1	Day 2	Day 3	Day 4	Day 5
3	30 Jun	1 Jul	2 Jul	3 Jul	4 Jul
4	5 Jul	6 Jul	7 Jul	8 Jul	9 Jul
5	10 Jul	11 Jul	12 Jul	13 Jul	14 Jul
6	15 Jul	16 Jul	17 Jul	18 Jul	19 Jul
7	20 Jul	21 Jul	22 Jul	23 Jul	24 Jul
8	25 Jul	26 Jul	27 Jul	28 Jul	29 Jul
9	30 Jul	31 Jul	1 Aug	2 Aug	3 Aug

4.1 Meteorology

The MM5 performance is assessed for the two simulations using a standard suite of statistical measures by comparing against surface meteorological observations. This analysis is performed to gauge the relative improvement in the surface meteorological fields, which greatly impact near-surface chemistry modeling, when nudging is used in MM5. No upper-air meteorological verification is performed in this study because there are no comparisons against upper-air chemical observations, and it is well-known that a dynamic analysis with MM5 is generally statistically superior to a forecast. The statistics for the "Nudge" simulation compare favorably for the 35-day period with those computed for the full summer over the same domain and year but using a different MM5 configuration (Gilliam et al. 2006), so it is assumed that the seven MM5 run segments used here are representative of the same summer period. Figure 1 shows the root-mean-square error (RMSE) for 2-m temperature and 10-m wind speed calculated against all surface stations in the MM5 domain and binned by day following Table 1. The MM5 simulation with nudging ("Nudge") performs with reasonable consistency through the MM5 segment (i.e., little change in statistical skill with increased run length). However, the MM5 simulation without the nudging shows a marked decrease in skill with increased run length, as expected. The 2-m temperature RMSE in "Nudge" ranges from 2.55–2.60 K for each of the five days, while it rises from 2.82 to 3.60 K over the same time period in "NoNudge". The RMSE for 10-m wind speed is ~1.8 m s^{-1} on each of the five days in "Nudge", but it grows from 2.1 to 2.4 m s^{-1} in "NoNudge". A similar pattern holds for other discrete statistical measures including index of agreement, mean absolute error, and correlation (not shown), and for 2-m water vapor mixing ratio and 10-m wind direction (not shown).

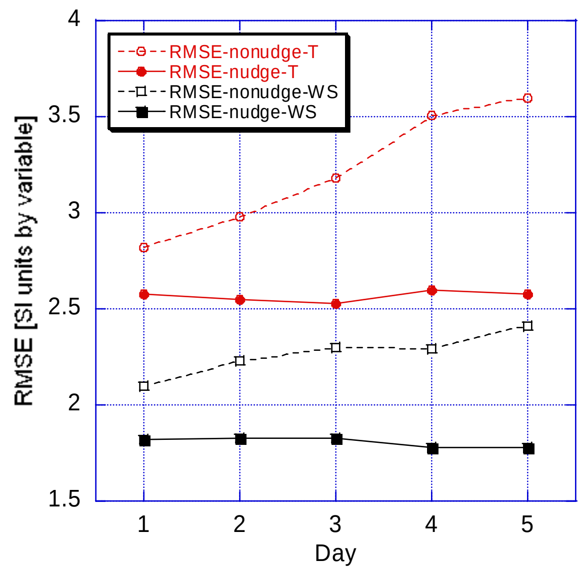

Figure 1. RMSE calculated at all observation sites for 2-m temperature (in red) and 10-m wind speed (in black) for 35 days of MM5 simulations. Days are binned following Table 1. The "Nudge" simulation is denoted by solid lines, and the "NoNudge" simulation is denoted by dashed lines.

4.2 Air Quality

Because of the time binning used in this analysis, it is necessary to compare against chemical observations that are available with a high temporal frequency (i.e., no coarser than daily) and a high spatial coverage. The CMAQ simulations are compared against surface hourly ozone and daily maximum 1-h ozone data from the EPA's Air Quality System (AQS) database. More than 1000 ozone monitors are used, with the highest observation densities in the eastern United States (east of the Mississippi River) and in California. Initially, discrete statistics are used.

It should be noted that the AQS observations are recorded from midnight to midnight, local standard time (LST), and the MM5 and CMAQ simulation days are defined using Universal Time Constant (UTC). The software program that is used to calculate the statistics for CMAQ stores the data in terms of LST, so all model-observation pairings are made in LST. Therefore, in the analysis of the CMAQ simulations shown below, the day bins cannot be exactly compared with the data shown for MM5. This issue can be important for AQS sites in California, where MM5/CMAQ "days" are 1600 LST to 1600 LST, which may not include the daily 1-h maximum surface ozone that corresponds to the calendar day. In the eastern United States, the daily 1-h maximum surface ozone is generally recorded within the same day using either UTC or LST. Further inspection of the data is required to determine the impact of this idiosyncrasy on the interpretation of the results, particularly for "Day 1". However, the trends for days 2–5 are robust.

Figure 2 shows the RMSE for the surface daily 1-h maximum ozone broken into systematic and unsystematic vector components (RMSEs and RMSEu, respectively; cf. Willmott 1982). The RMSEs should account for processes that the model does not simulate well, whereas the RMSEu could be attributed to subgrid-scale processes or random errors. The CMAQ simulation that uses input meteorology from "Nudge" tends to be a better overall simulation than the simulation that uses meteorology from "NoNudge" as reflected in both the RMSEs and the RMSEu. The RMSEu does not exhibit a significant change through the MM5 run segment for either "Nudge" or "NoNudge". However, the RMSEs shows a marked decrease in skill and increases by nearly 4 ppb between days 2 and 5 for "NoNudge". Somewhat surprisingly, there is also a gradual decrease in skill in RMSE from day 2 to day 5 when "Nudge" is used, as seen in both RMSEs and RMSEu.

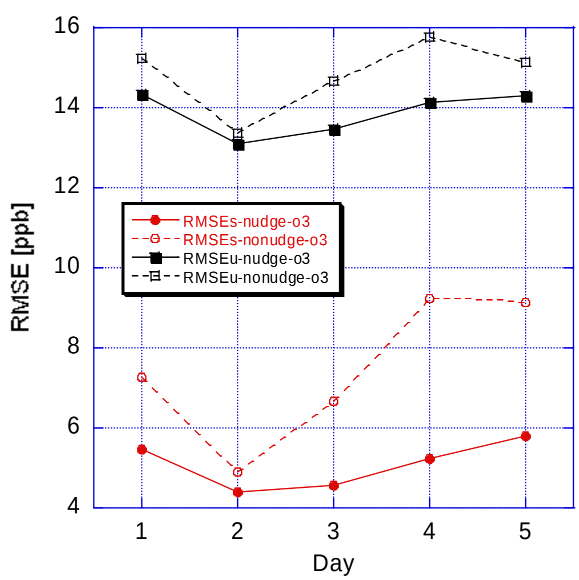

Figure 2. RMSE calculated at all AQS sites for surface daily 1-h maximum ozone by systematic (in red) and unsystematic (in black) vector components. Days are binned following Table 1. The CMAQ simulation that uses meteorology from "Nudge" is denoted by solid lines, and the CMAQ simulation that uses meteorology from "NoNudge" is denoted by dashed lines.

Figure 3 shows spatial comparisons of RMSE for daily 1-h maximum ozone for days 2 and 5 for the CMAQ simulations that used "NoNudge" and "Nudge". Figure 3a shows that the Day 2 CMAQ simulations with meteorology from "NoNudge" typically have widespread RMSE of 5–15 ppb. The Day 2 CMAQ simulations with meteorology from "Nudge" (Fig. 3b) indicate slightly smaller RMSE, often 5–10 ppb. This illustrates a fairly consistent spatial improvement in CMAQ by using nudging in MM5. By Day 5, the CMAQ simulation that used meteorology from "NoNudge" (Fig. 3c) experiences a widespread decrease in statistical skill compared to both the Day 5 "Nudge" (Fig. 3d) and to the Day 2 "NoNudge". The RMSE in Day 5 "NoNudge" are generally 10–20 ppb, compared to RMSE of 5–15 ppb in Day 5 "Nudge". Figures 3b and 3d show that the gradual decrease in statistical skill over time with "Nudge" (which is also seen in Fig. 2) is observed throughout the simulation domain. Further analysis of the data is required to test for statistical significance and to determine whether or not the statistics are driven by a single event that was poorly simulated or if this is a general conclusion.

5. DISCUSSION

This paper describes the preliminary analysis of the value of using nudging in the meteorology model on the retrospective CMAQ simulation. A 35-day period is examined by binning the CMAQ simulation days according to time elapsed in each 5.5-day MM5 simulation segment. As expected, initial results confirm that the CMAQ simulations that use MM5 with nudging compare more favorably with the daily maximum 1-h surface ozone observations than the CMAQ simulations that used MM5 "forecast" fields. For 1-h maximum ozone, widespread increases in RMSE of 5 ppb from Day 2 to Day 5 are seen in the simulations where nudging is not used. There is no appreciable decline in the statistical skill of the MM5 simulation (i.e., for near-surface wind, temperature, and moisture) with nudging as run length increases. However, there are more subtle but widespread decreases in statistical skill for 1-h maximum ozone in CMAQ as simulation length increases when the MM5 dynamic analysis is used.

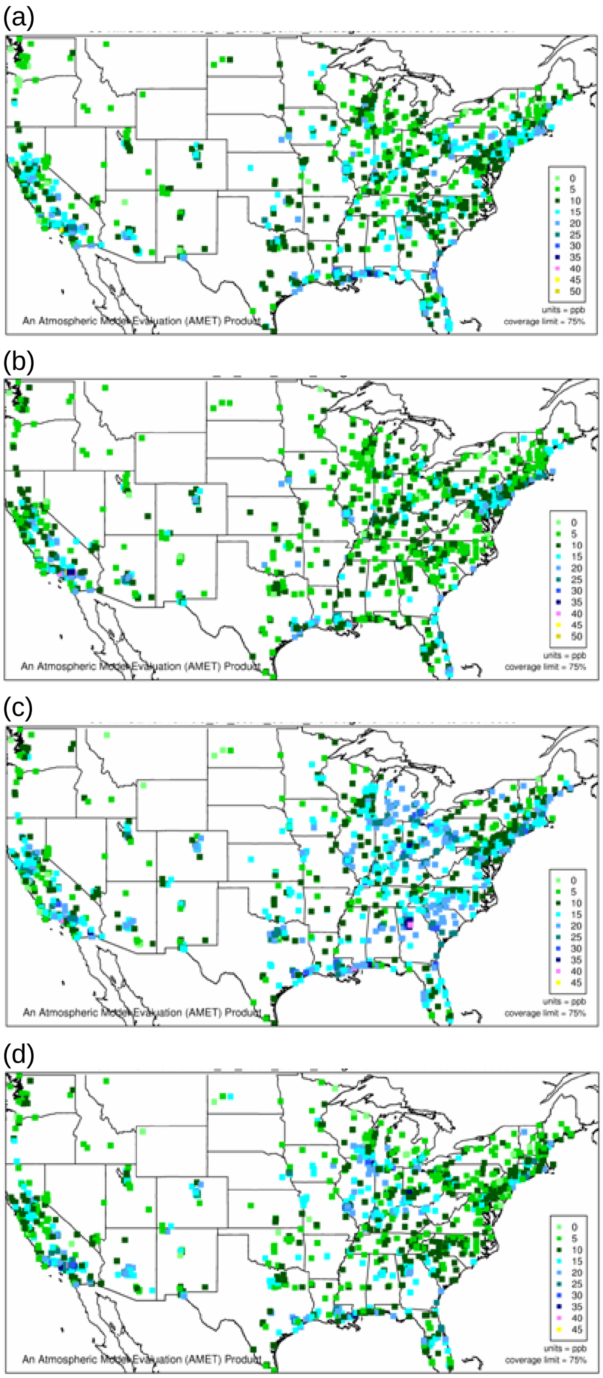

Figure 3. Spatial distribution of RMSE for daily 1-h maximum ozone at AQS sites. (a) Day 2 for “NoNudge”; (b) Day 2 for “Nudge”; (c) Day 5 for “NoNudge; and (d) Day 5 for “Nudge”.

Because the analysis is just beginning, it may be premature to draw firm conclusions. Further study of additional statistical measures (both discrete and categorical) is planned, as is a more thorough examination of the individual MM5 run segments to evaluate whether the bulk trends can be generalized to each MM5 simulation. In addition, hourly ozone and 8-h maximum ozone will be examined. Additional investigation of the feedbacks of the meteorological fields (e.g., PBL height, temperature, moisture, wind speed and direction, and precipitation) on the chemistry output is also warranted. This analysis technique may also be applied for a field-study period where other species are collected on a daily basis. This work may also be extended to WRF when the analysis nudging is fully implemented.

6. ACKNOWLEDGMENTS

This research could not have been performed without leveraging the past and on-going research of my colleagues on the Model Development and Evaluation Teams. The MM5 preprocessing and the MM5 simulations that included the nudging were created by Lara Reynolds (CSC). The MM5-independent emissions files were created by Allan Beidler (CSC), Charles Chang (CSC), and Ryan Cleary (CSC). Guidance on meteorology-dependent emissions processing was generously provided by George Pouliot. Golam Sarwar (EPA/NERL/AMD) created the photolysis files. Harvard University provided the GEOS-CHEM simulations, and Steven Howard created the chemistry lateral boundary conditions files for June and July 2001. Shawn Roselle created the chemistry initial condition file and the CMAQ executable. Observation processing (meteorology and chemistry) and additional support with the Atmospheric Model Evaluation Tool was generously provided by Robert Gilliam and K. Wyat Appel. Technical reviews of this manuscript were performed by Brian Eder, Robert Gilliam, and S.T. Rao.

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. 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.

7. REFERENCES

Bey, I., D. J. Jacob, R. M. Yantosca, J. A. Logan, B. D. Field, A. M. Fiore, Q. Li, H. Y. Liu, L. J. Mickley, and M. G. Schultz, 2001: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *J. Geophys. Res*., **106**, 23 073–23 096.

Binkowski, F. S., and S. J. Roselle, 2003: Models–3 Community Multiscale Air Quality (CMAQ) model aerosol component. 1. Model description. *J. Geophys. Res*., **108**, 4183, doi:10.1029/2001JD001409.

Black, T., 1994: The new NMC mesoscale Eta model: Description and forecast examples. *Wea. Forecasting*, **9**, 265–278.

Byun, D. W., and K. L. Schere, 2006: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. *Appl. Mech. Rev*., **59**, 51–77.

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

EPA, 2003: User's guide to MOBILE6.1 and MOBILE6.2 (Mobile Source Emission Factor Model). U.S. Environmental Protection Agency Rep. EPA420-R-03-010, 262 pp.

Gilliam, R. C., C. Hogrefe, and S. T. Rao, 2006: New methods for evaluating meteorological models used in air quality applications. *Atmos. Environ*., **40**, 5073–5086.

Grell, G. A., J. Dudhia, and D. R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). NCAR Tech. Note NCAR/TN-398+STR, 138 pp.

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.

Kain, J. S., 2002: The Kain-Fritsch convective parameterization: An update. *J. Appl. Meteor*., **43**, 170–181.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res*., **102**, 16 663–16 682.

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**, 25 611–25 629.

Pleim, J. E., 2006: A combined local and non-local closure model for the atmospheric boundary layer. Part 1: Model description and testing. *J. Appl. Meteor. and Clim.*, in press.

Pleim, J. E., and J.S. Chang, 1992: A non-local closure model for vertical mixing in the convective boundary layer. *Atmos. Environ*., **26A**, 965–981.

Pleim, J. E., and A. Xiu, 2003: Development of a land surface model. Part II: Data assimilation. *J. Appl. Meteor*., **42**, 1811–1822.

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. Focus*, **1**, 243–252.

Reisner, J., R. J. Rasmussen, and R. T. Bruintjes, 1998: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Quart. J. Roy. Meteor. Soc*., **124B**, 1071–1107.

Stauffer, D. R., and N. L. Seaman, 1990: Use of four-dimensional data assimilation in a limited-area forecast model. Part I: Experiments with synoptic-scale data. *Mon. Wea. Rev*., **118**, 1250–1277.

Stauffer, D. R., and N. L. Seaman, 1994: Multiscale four-dimensional data assimilation. *J. Appl. Meteor*., **33**, 416–434.

Stauffer, D. R., N. L. Seaman, and F. S. Binkowski, 1991: Use of four-dimensional data assimilation in a limited-area mesoscale model. Part II: Effects of data assimilation within the planetary boundary layer. *Mon. Wea. Rev*., **119**, 734–754.

Willmott, C. J., 1982: Some comments on the evaluation of model performance. *Bull. Amer. Meteor. Soc*., **63**, 1309–1313.

Xiu, A., and J. E. Pleim, 2001: Development of a land surface model. Part I: Application in a mesoscale meteorology model. *J. Appl. Meteor*., **40**, 192–209.

Yarwood, G., S. Rao, M. Yocke, and G. Whitten, 2005: Updates to the Carbon Bond chemical mechanism: CB05. Final Report to the U.S. EPA, RT-0400675. [Available on-line at www.camx.com.]