AN OPERATIONAL EVALUATION OF THE ETA - CMAQ AIR QUALITY FORECAST MODEL

Daiwen Kang, Brian Eder, Rohit Mathur, Shaocai Yu, and Ken Schere*

EPA/600/A-04/080

1. INTRODUCTION

The National Oceanic and Atmospheric Administration (NOAA), in collaboration with the Environmental Protection Agency (EPA), are developing an Air Quality Forecasting Program that will eventually result in an operational Nationwide Air Quality Forecasting System. The initial phase of this program, which couples NOAA's Eta meteorological model with EPA's Community Multi-scale Air Quality (CMAQ) model, began operation in May of this year and has been providing forecasts of hourly, maximum 1- and 8-hour ozone concentrations over the northeastern United States.

As part of this initial phase, an operational evaluation of the coupled modeling system is being performed in which both *discrete forecasts* (observed versus modeled concentrations) for hourly, maximum 1-hr, and maximum 8-hr O_3 concentrations and *categorical forecasts* (observed versus modeled exceedances / non-exceedances) for both the maximum 1-hr (125 ppb) and 8-hr (85 ppb) are evaluated. This paper examines one month (1- 30 June, 2004) of the evaluation, using hourly O_3 concentration measurements from the EPA's AIRNOW network.

2. THE MODELING SYSTEMS

The Eta-CMAQ Air Quality Forecasting (AQF) system is based on the National Centers for Environmental Prediction's (NCEP's) Eta model (Black 1994; Rogers et al., 1996) and EPA's CMAQ Modeling System (Byun and Ching 1999). A brief summary of

---

* Daiwen Kang#, Brian Eder@, Rohit Mathur@, Shaocai Yu#, and Kenneth Schere@, NERL, U.S. Environmental Protection Agency, RTP NC, 27711, USA. # On Assignment from Science and Technology Corp., VA 23666, USA. @On assignment from Air Resources Laboratory, National Oceanic and Atmospheric Administration, RTP, NC 27711, USA.

the linkage between the Eta and the CMAQ models, relevant to this study, is presented below. A more in depth description can be found in Otte et al. (2004).

The Eta model is used to prepare the meteorological fields for input to the CMAQ. The NCEP Product Generator software is used to perform bilinear interpolations and nearest-neighbor mappings of the Eta Post-processor output from Eta forecast domain to the CMAQ forecast domain. The processing of the emission data for various pollutant sources has been adapted from the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system (Houyoux et al., 2000) on the basis of the U.S. EPA national emission inventory. The Carbon Bond chemical mechanism (version 4.2) is used for representing the photochemical reactions.

Detailed information on transport and cloud processes in the CMAQ is given in Byun and Ching (1999). For this application, O_3 concentrations are forecast over the Northeast U.S. using a 12-km horizontal grid spacing on a Lambert Conformal map projection. The vertical resolution is 22 layers, which are set on a sigma coordinate, from the surface to ~100 hPa. Vertically varying lateral boundary conditions for O_3 are derived from daily forecasts of the Global Forecast System (GFS). 3D chemical fields are initiated from the previous forecast cycle. The Eta 12 UTC cycles are used for the forecast cycle (Otte et al., 2004). The primary Eta-CMAQ model forecast for next-day surface-layer O_3 is based on the current day's 12 UTC Eta cycle and products are issued daily no later than 1330 LDT. The target forecast period is local midnight through local midnight (04 UTC to 03 UTC for the Northeast U.S.). An additional 8-hr is required beyond midnight to calculate peak 8-h average O_3 concentrations. As a result, a 48-hr Eta-CMAQ forecast is needed (based on the 12 UTC initialization) to obtain the desired 24-hr forecast period. At the time of publication, one month of evaluation had been performed (1 June to 30 June). Additional time periods will be available at the time of the conference.

3. O_3 DATA

Hourly, near real-time, O_3 (ppb) data obtained from EPA's AIRNOW program are used in the evaluation. Over 600 stations are available, mostly in urban areas, resulting in over 500,000 observations. In addition to the hourly data, both the maximum 1-hour and maximum 8-hour concentrations are calculated for each station and each day over the evaluation period. The calculation of 8-hour maximum is the same as model forecast using the forward calculation method (i.e. calculation of the last seven 8-hour maximum concentrations including data from next day). The maximum 1-hr and 8-hr concentrations are considered missing if half of the hourly observation data is missing for the day. If two or more monitoring stations are located within the same model grid cell, their average value is used as the representative measurement for that grid cell.

4. STATISTICS

A brief summary of the various statistics used in this evaluation are presented below. A more in depth discussion can be found in Kang et al. (2003, 2004).

4.1. Discrete Statistics

For the **discrete forecast** evaluation, basic summary statistics along with two standard and widely used measures of **bias**: the Mean Bias (MB) and the Normalized Mean Bias (NMB); and **error:** the Root Mean Square Error (RMSE) and Normalized Mean Error (NME) were selected and are defined below:

$$MB = \frac{1}{N}\sum_{1}^{N}(Cm - Co) \qquad (1)$$

$$NMB = \frac{\sum_{1}^{N}(Cm - Co)}{\sum_{1}^{N}Co} \cdot 100\% \qquad (2)$$

$$RMSE = \left(\frac{1}{N}\sum_{1}^{N}(Cm - Co)^2\right)^{0.5} \qquad (3)$$

$$NME = \frac{\sum_{1}^{N} Cm - Co}{\sum_{1}^{N}Co} \cdot 100\% \qquad (4)$$

Where *Cm* and *Co* are modeled and observed concentrations, respectively.

4.2. Categorical Statistics

For the **categorical forecast** evaluation, the models' Accuracy (A), Bias (B), Probability of Detection (POD), False Alarm Rates (FAR) and Critical Success Index (CSI) were calculated, based upon observed exceedances, non-exceedances *versus* forecast exceedance, non-exceedances for both the 1- and 8-hour O_3 standard. A graphical representation of the variables (a, b, c and d) used to formulate the categorical metrics is presented in Figure 1, where **a** would represent a forecast 1-hr exceedance (>125 ppb) that did not occur, **b**: a forecast 1-hr exceedance that did occur, **c**: a forecast 1-hr non-exceedance that did not occur and **d**: a non forecast 1-hr exceedance that did occur.

$$A = \left(\frac{b + c}{a + b + c + d}\right) \cdot 100\% \qquad (5)$$

$$CSI = \left(\frac{b}{a + b + d}\right) \cdot 100\% \qquad (6)$$

$$POD = \left(\frac{b}{b + d}\right) \cdot 100\% \qquad (7)$$

$$B = \left(\frac{a + b}{b + d}\right) \qquad (8)$$

$$FAR = \left(\frac{a}{a + b}\right) \cdot 100\% \qquad (9)$$

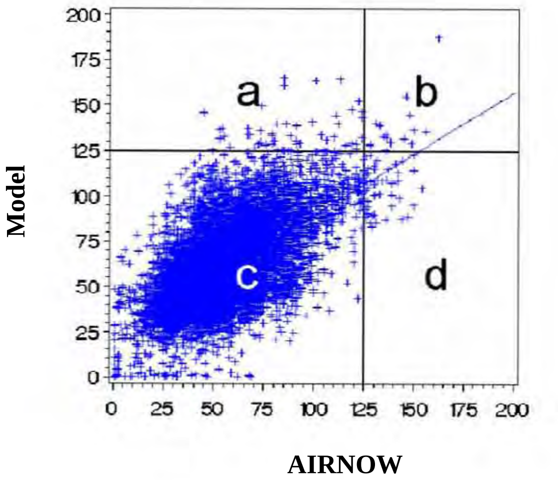


Figure 1. Example plot for the categorical evaluation.

5. RESULTS

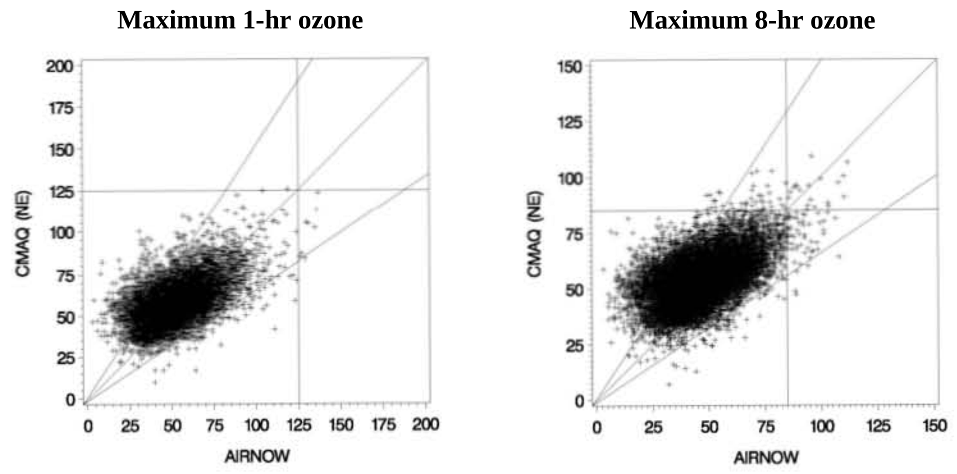
5.1. Discrete Evaluations

As seen in Table 1, discrete evaluations were performed for hourly, maximum 1-hr and maximum 8-hr O_3 forecast. To differentiate model performance at levels above typical background concentrations, metrics were also computed using a 40 ppb observation threshold. For the most part, observed O_3 concentrations were very low over most of the domain during the month of June. In fact, the mean observed O_3 concentration of 29.2 ppb was actually lower than boundary conditions used by the model. Therefore it is not surprising that the model tended to over predict O_3 concentrations for all the categories in Table 1, resulting in positive biases. Much of this overprediction is eliminated however, when the 40 ppb observed threshold is considered. As an example, the NMB for the hourly forecast falls from 42.4% (all obs. considered) to 1.9 % (obs > 40 ppb). Similarly, the NMB of the maximum 1- and 8-hour forecast falls from 11.1 to 2.0% and from 16.9 to 6.5%, respectively. With the exception of correlation coefficients, which actually decline slightly (due to the removal of the model's innate ability to simulate the diurnal variability), similar improvement is seen in the other metrics.

Table 1. Summary of discrete statistics for forecasts of hourly, maximum 1-hr, and 8-hr O_3 for all the concentration range and the range for all the observed concentrations greater than 40 ppb

Metrics	Hourly		Max 1-hr		Max 8-hr	
	All	>40 ppb	All	>40 ppb	All	>40 ppb
N	536,623	142,913	18,389	12,218	18,389	12,218
Obs. Mean	29.2	51.3	51.7	59.1	46.1	53.1
Mod. Mean	41.6	52.3	57.5	60.3	53.9	56.5
MB (ppb)	12.4	1.0	5.7	1.2	7.8	3.4
NMB (%)	42.4	1.9	11.1	2.0	16.9	6.5
NME (%)	53.7	17.0	20.8	14.3	24.0	15.5
RMSE (ppb)	19.7	11.4	13.8	10.8	14.1	10.4
r	0.54	0.45	0.54	0.52	0.51	0.48

Scatter plots of the model forecasts versus AQS observations (for both the maximum 1- and 8-hr ozone concentrations) are presented in Figure 2. In addition to illustrating the exceedance threshold areas (which were used in calculation of the categorical statistics), the plots also provide the 1:1.5 factor lines. As discussed above and evidenced in the scatter plot, most of the over prediction occurs at the lower concentrations; when observed concentrations >40 ppb, majority of the model forecast are within a factor of 1.5.


Maximum 1-hr ozone

Maximum 8-hr ozone

Figure 2. Scatter plots of the model versus AIRNOW for both 1- and 8-hour maximum ozone concentrations (ppb) with exceedance thresholds indicated.

Evaluation of model diurnal performance was also performed as seen in Figure 3, where boxplot (denoting 75th, 50th, 25th percentiles, max., min. and mean) of simple bias (CMAQ-observations) over the entire analysis period of June is provided. Although the model over predicts throughout the diurnal period, the positive bias is most prevalent during night, due in large part to the model system's difficulty in simulating the evolution of the nocturnal boundary layer and its impact on surface O_3 concentrations.

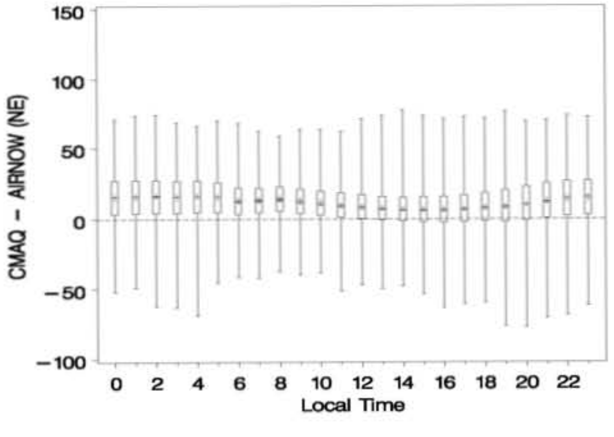


Figure 3. Boxplots of the diurnal variation (model – AIRNOW) for hourly ozone concentrations (ppb)

5.2. Time series

Figure 4 displays a time series of correlation coefficient (top panel), NMB (middle panel), and NME (bottom panel) for the hourly, maximum 1-hr, and maximum 8-hr concentrations throughout the month of June. The correlation coefficients typically fluctuate between 0.4 and 0.8, with one major exception, the 13th of June. On this date, the correlation coefficients collapsed (especially for the max. 1-hr and max. 8-hr concentrations). Detailed examination of the model's simulation on that date indicated its performance suffered due mainly to extensive cloud cover and precipitation across the modeling domain that had not accurately forecast. The NMB typically ranges from 0 to

Figure 4. Time series of correlation coefficient (top panel), NMB (middle panel), and NME (bottom panel) for hourly, maximum 1-hr, and maximum 8-hour ozone concentrations (ppb).

30% across the month for the maximum 1- and 8-hour forecasts, while the NME ranges from 20 to 40%. Note that the NMB and NME are systematically larger for the hourly forecasts due mainly to the models difficulty in simulating nighttime concentrations.

Closer examination of the time series figure reveals a systematic pattern of varied modeled performance (i.e. several days of good performance, followed by several days of poor performance). This systematic pattern could be traced back to the “synoptic-scale” meteorology that was impacting the domain. During days when high pressure, relative clear skies and little precipitation occurred within the domain (all conditions conducive to O_3 formation), the model performed well (i.e. 2-7, 19-21 June). Other days within the domain that were characterized by extensive cloud cover and precipitation (conditions not conducive to O_3 formation) associated with either fronts or areas of low pressure resulted in poor model performance. (i.e. 10, 15, 16, 17, 25 June)

5.3. Categorical Evaluations

Because of the prevalence of low O_3 concentrations during June, an insufficient number of maximum 1-hour exceedances occurred to provide meaningful categorical statistics (see Figure 2). Accordingly, discussion in this session will focus on the maximum 8-hour exceedances, of which, there were a sufficient, though not ideal number. As shown in Table 2, the **Accuracy** (A) for model prediction, which indicates the percent of forecasts that correctly predict an exceedance or non-exceedance, is close to 100%. However, care must be taken in interpretation of this metric, as it is greatly influenced by the overwhelming number of correctly forecast non-exceedances. To circumvent this inflation (which is common when evaluating the prediction of rare events like O_3 exceedances), the **Critical Success Index** (CSI) is often a better (though stringent) metric of model performance. The CSI provides a measure of how well the exceedances were predicted, without regard to the large occurrence of correctly predicted non-exceedances. For our evaluation, the CSI for the 8-hr exceedance is about 16%. The **Probability Of Detection** (POD) metric is similar to the CSI , (though less stringent) in that it measures the number of times a model predicted an exceedance when one actually occurred. In our evaluation, the POD for maximum 8-hr forecast is 25%. Measures of **Bias** (B), which for a categorical forecast indicates if forecast exceedances are under predicted ($B < 1$) or over predicted ($B > 1$), indicate that the model somewhat under predicts the number of maximum 8-hour exceedances. And finally, the fifth categorical metric, the **False Alarm Rate** (FAR), which indicates the number of times that the model predicted an exceedance when none occurred, was 67.8%. Though high, the FAR, as well as the other categorical metrics are comparable with other similar regional scale forecast models (Kang et al., 2003).

Table 2. Summary of categorical statistics for forecasts of maximum 8-hr O_3

	A (%)	CSI (%)	POD (%)	B	FAR (%)
Max 8-hr	99.48	16.25	25.33	0.79	67.8

6. SUMMARY

The purpose of this research has been to provide an operational evaluation of the Eta-CMAQ air quality forecast system using O_3 observations obtained from EPA's AIRNOW program and a suite of statistical metrics for both discrete and categorical forecasts. Results from this evaluation revealed that modeling system performed reasonably well, in this, its first major attempt at forecasting ozone concentrations over the Northeastern United States. The quality of the discrete forecasts of the maximum 1-hour concentrations (r = 0.54, NMB = 11.1%, NME = 20.8%) and maximum 8-hour concentrations (r = 0.51, NMB = 16.9%, NME = 24.0.0%) were comparable, if not better than similar model applications during the summer of 2002 (Kang, 2004). Because of the prevalence of low O_3 concentrations during June, an insufficient number of maximum 1-hour exceedances occurred to provide meaningful categorical statistics. The categorical evaluation therefore focused on the maximum 8-hr forecast and also revealed results comparable to those found with similar model applications during the summer of 2002.

Time series of the metrics associated with the discrete forecasts revealed a systematic pattern of varied modeled performance that could be traced back to the "synoptic-scale" meteorology impacting the domain. During days when high pressure, relative clear skies and little precipitation occurred within the domain (all conditions conducive to O_3 formation), the model performed well. Conversely, on those days characterized by extensive cloud cover and precipitation (conditions not conducive to O_3 formation) associated with either fronts or areas of low pressure, the model performed poorly. This performance characteristic is likely attributable to the fact that CMAQ was designed and developed to perform well in O_3 conducive conditions. As the summer of 2004 progresses, and the likelihood of such conditions increases, the performance of the forecast system is expected to improve.

6. REFERENCES

Black, T. The new NMC meso-scale Eta Model: description and examples. Wea. Fore, 9, 265-278, 1994.

Byun, D.W. and J.K.S. Ching, Eds.,: Science algorithms of the EPA Models-3 Community Multi-scale Air Quality (CMAQ) modeling system, EPA/600/R-99/030, Office of Research and Development, U.S. Environmental Protection Agency, 1999.

Houyoux, M.R., J.M. Vukovich, C.J. Coats Jr., N.M. Wheeler, and P.S. Kasibhatla, Emission inventory development and processing for the seasonal model for regional air quality (SMRAQ) project. J. Geophys. Res. 105, 9079-9090, 2000.

Kang, D., B.K. Eder, and K.L. Schere, The evaluation of regional-scale air quality models as part of NOAA's air quality forecasting pilot program, Preprints, 26th NATO/CCMS International Technical Meeting on Air Pollution Modeling and its Application, 26-30 May 2003, Istanbul, Turkey, 404-411, 2003.

Kang, K., B. Eder, A. Stein, G. Grell, S.Peckham, The New England air quality forecasting pilot program:development of an evaluation protocol and performance benchmark. JAWMA, in press, 2004.

Otte, T.L., et al., Link the Eta model with the community multi-scale air quality (CMAQ) modeling system to build a real-time national air quality forecasting system. Weather and Forecasting, 2004 (in review).

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

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.