A YEAR-LONG MM5 EVALUATION USING A MODEL EVALUATION TOOLKIT

Robert C. Gilliam[1], Prakash V. Bhave[1], Jonathan E. Pleim[1] and Tanya L. Otte[1]
[1]Atmospheric Sciences Modeling Division, National Oceanic and Atmospheric Administration's Air Resources Laboratory, on assignment to the Environmental Protection Agency's National Exposure Research Lab.
Email: gilliam.robert@epa.gov
Web address: http://www.epa.gov/asmdnerl
Voice (919) 541-4593 Fax (919) 541-1379

1. INTRODUCTION

Air quality modeling has expanded in both sophistication and application over the past decade. Meteorological and air quality modeling tools are being used for research, forecasting, and regulatory related emission control strategies. Results from air quality simulations have far reaching implications and are closely linked to the meteorological model that drives chemical transport, diffusion, and reactions. Therefore, modeling systems should be evaluated by considering all components involved (Hogrefe et al. 2001). This connection can be achieved by linking the statistical analysis of the air quality model with that of the meteorological model in space and time, in order to distinguish how errors in the air chemistry model are attributed to errors in the meteorological modeling.

An evaluation tool is being developed that will 1) provide a better sense of meteorological model uncertainty; 2) standardize the evaluation process; 3) manage a large volume of evaluation results; 4) make the overall evaluation process more efficient and less labor intensive; and 5) directly link the meteorological model evaluation with the air quality model evaluation.

This study applies the model evaluation tool to a year-long simulation using the Pennsylvania State University (PSU) / National Center for Atmospheric Research (NCAR) fifth-generation mesoscale model (MM5). The results are reported not only to examine the MM5 model performance, but also to demonstrate the effectiveness of the evaluation system. Among the evaluations presented are surface-based 2 m temperature, 10 m wind, 2 m mixing ratio, precipitation and solar radiation. Wind profiler data are also used to examine the ability of MM5 to simulate the vertical distribution of wind over the diurnal cycle. Additionally, a direct linkage between the meteorological and the air quality model performance, specifically ozone and nitrate, is attempted. Only a brief summary of the results are presented here because of manuscript length requirements.

2. METHODOLOGY

2.1 Model Simulations

An important effort was recently undertaken to examine both air quality in general as well as the capability of simulating air quality over the Continental United States. A year-long Community Multiscale Air Quality (CMAQ) model (Byun and Ching, 1999) simulation was performed over the continental United States, with a horizontal grid spacing of 36 km, for the time period from January 01 – December 31, 2001. The required meteorological fields for the CMAQ simulation were supplied by the MM5 model (Grell et al., 1994) version 3.6.1. These model outputs are the focus of the evaluation. For specifics on the MM5 configuration refer to McNally (2003). For specifics on the CMAQ configuration refer to Eder (2004).

2.2 Model Evaluation Tool

The model evaluation tool is being designed to be flexible, extendable, and user-friendly. The current version is compatible with MM5, Weather Research and Forecast (WRF) model, and National Centers for Environmental Prediction (NCEP) Eta model output. Observations from National Oceanic and Atmospheric Administrations (NOAA) Forecast Systems Laboratory (FSL) Meteorological Assimilation Data Ingest System (MADIS) are matched in time and space with the model output. These matched model-observation pairs for each variable are immediately stored in a relational database, which is the core of the evaluation system. Various analysis programs connect to the database and extract user-specified data, then generate statistical and analysis plots. Currently, programs have been developed to produce spatial statistics over specified time periods, time series statistics, and time series plots for various observation sites, comparisons between modeled precipitation and the national gridded precipitation analysis, as well as a utility to extract and compute statistics for subsets of data (e.g., temperature statistics for the mountains of NC, during the winter nighttime hours when temperature < 0ºC and wind speed < 1 $m \cdot s^{-1}$). Also, a utility is available to examine the simulated vertical profiles of wind and temperature using observations from the National Profiler Network (NPN).

An attraction of the relational database is the ability to link data sets (e.g., air quality with meteorological data). This is achieved by extracting a set of air quality data (model and observed values) and corresponding meteorology in time and space based on some criteria (e.g., 8-h ozone conc. > 60 ppb and region = eastern U.S.). Conversely, a set of meteorology and air quality data (model and observed values) can be extracted based using meteorological dependent criteria (e.g., 8-h ozone conc. when 2 m temperature > 300 K and 10-m wind speed < 3 m s^{-1} and 10-m wind direction between 180º and 330º). Ozone observations from the Aerometric Information Retrieval System (AIRS) network and particle phase nitrate observations from the Interagency Monitoring of Protected Visual Environments (IMPROVE) network are used in this study. Examples of linking these datasets are presented in Section 3.2

3. RESULTS

3.1 General Statistics

Domain-wide statistics provides a general measure of how well the model simulation represents the conditions that actually occurred. Table 1 is a compilation of several well-known model performance statistics for 2 m temperature from the 2001 MM5 simulation. These are presented for a variety of data subsets in order to show the flexibility of the model evaluation tool. For example, seasonal statistics were derived using date start and stop criteria. Data were also extracted from the database by station elevation criteria to examine model performance over and near the coastal waters (elev. < 25 m), inland (25 m < elev. < 500 m) and in higher elevations (> 500 m). Clustering days with similar weather patterns was also performed. Lastly, specific datasets can be examined, although not performed in this study, using land use, meteorological conditions, observations network, and/or geographical bounds as criteria.

Cases	Sample Size	MAE	BIAS	IOA
Winter	2135744	2.24	-1.25	0.95
Spring	2167129	1.85	-0.52	0.95
Summer	2222299	1.63	-0.18	0.94
Fall	2265935	1.72	-0.34	0.96
Inland*	1491581	1.65	-0.22	0.93
Mountains*	103664	1.85	0.22	0.92
Marine*	445854	1.69	-0.34	0.57
Cluster 1	1751747	1.87	-0.53	0.76
Cluster 4	794300	1.95	-0.48	0.84
Cluster 7	500047	1.90	-0.61	0.80
NC*	102700	1.63	-0.57	0.87
FL*	119404	1.73	-0.18	0.72
TX*	168723	1.76	-0.92	0.86
IA*	96080	1.57	-0.10	0.93
CO*	62021	2.72	-1.60	0.88
NY*	72920	1.69	-0.33	0.91
WA*	72231	2.00	-0.42	0.91
AZ*	46114	3.31	-2.44	0.83

Table 1. Mean Absolute Error (MAE), Mean Bias (BIAS) and Index of Agreement (IOA) calculated using hourly modeled and observed 2 m temperature (K) data from several data subsets. *Statistics were calculated for summer only (Jun 21-Sep 21).

Overall, the MM5 2001 simulation did well at representing the meteorological conditions near the surface. Table 1 indicates the model performed best in the summer when the temperature bias was -0.18ºC, mean absolute error (mae) 1.63ºC and an index of agreement (ioa) of 0.94. In winter, the model performed the poorest of all cases because of a cold bias at night. The overall bias was -1.25ºC, but the bias was worse in the early morning (~ -3.00ºC) and approached zero in the afternoon (diurnal bias not shown).

As expected, when examining the statistics as a function of elevation, regions that were not influenced by the coast and mountains were simulated best (bias: -0.22ºC, mae: 1.65ºC) followed closely by coastal areas. The statistics support that it is more difficult to simulate the meteorology over mountainous areas where mesoscale variations are not resolved, especially at a 36 km scale. Further support for this are the metrics that were calculated for Colorado over the same summer period, which indicate large temperature errors (2.72ºC) and bias (-1.60ºC) as do the statistics for Washington state (bias: -0.42ºC, mae: 2.00ºC).

Table 1 also presents the metrics of several cases for clusters of days that represent distinct synoptic patterns, determined by a map typing procedure (McKendry et al. 1995). Cluster 1 represents patterns where a Canadian high pressure was anchored over the northeast U.S. Cluster 4 represents a synoptic pattern where a cold front passes off the east coast of the U.S. and high pressure builds into the central U.S. Cluster 7 corresponds to weak large-scale flow over the entire eastern U.S., which typically occurs in the summer. The metrics are all similar for clusters 1, 4 and 7. This implies that the model performs consistently during various weather patterns. However, when the spatial distribution of these statistics were plotted (not shown), it was apparent that some areas (i.e., where fair weather exists) are better simulated than others.

Temperature statistics (Table 1) were also calculated for a number of states across the U.S. The 2 m temperature for the Midwest state of Iowa was best simulated, which is not surprising since it has a relatively homogenous surface. North Carolina and New York were also well simulated. As already noted, mountain states like Washington and Colorado were, comparatively, not as well simulated, as temperatures were simulated too cool. The warm season desert climate of Arizona was simulated worst among the states with an average error of 3.31ºC and bias of -2.44ºC.

In general, these statistics are comparable and in most instances better than previous model evaluations. Hogrefe et al. (2001) presented performance statistics from a similar model setup over the Eastern U.S. In that study the mean absolute error for the entire domain was 2.22ºC, the bias was -0.93ºC. Other more localized

MM5 studies (Baker 2004; Stauffer and Deng 2003) report mean absolute errors on the order of 1.50ºC at best to 3.00ºC at worst, with temperature bias on the order of -2.00ºC to 0.50ºC.

Table 2. Descriptions of C_3 and temperature datasets from which Table 3 statistics were calculated. All cases included only Eastern U.S. observations from May 15 to Sep. 15, 2001

3.2 Linked Air Quality-Meteorology Evaluation

Case	Description
Case1	All available samples
Case 2	All available where mean daytime temperature (1000-1800 LST) > 31ºC (88º F)
Case 3	Same as Case 2 but only instances where 8-h O_3 > 60 ppb
Case 4	Same as Case 3 but only instances where there was no model or observed precip.
Case 5	Same as Case 3, but where there was model precip, but no observed precip.
Case 6	Same as Case 3, but where there was observed precip, but no model precip.

Several chemical species modeled in CMAQ are especially sensitive to meteorological conditions, including ozone (O_3) and nitrate (NO_3^-), which are controlled in part by the amount of incoming solar radiation at the surface and relative humidity, respectively. In theory, biases in the modeled meteorology will lead to biases in the predicted air quality. To explore this relationship, the meteorology and air quality concentrations were extracted for the same location and time using the evaluation database.

Ozone and associated 2 m temperature were obtained for a variety of cases/criteria for the Eastern U.S., and from these data subsets the model biases were calculated. The ozone values compared are the daily 8-h maximum for each site over the May 15, 2001 to September 15, 2001 period. These O_3 data were linked with the mean observed and modeled 2 m temperature between 1000-1800 LST from the nearest (within 10 km, on average) weather station. Table 2 provides a description of the various cases. Table 3 lists the ozone and 2 m temperature bias for each case. In theory, one would expect that a cold model bias during the daytime is in many instances, related to over-simulated cloudiness and lower incoming solar radiation at the surface. This situation would lead to lower ozone concentrations in the air quality model. This logic is weakly supported by the bias comparison between temperature and ozone. Case 2 has a temperature bias of -4.75ºC and ozone bias of -2.17 ppb, conversely Case 1 has biases of -0.36ºC and 6.16 ppb. In fact, the correlation coefficient of the ozone and temperature bias for these six cases is 0.38, so it appears that there is at least a weak relationship between the accuracy of the simulated temperature and ozone.

Table 3. Number of data points (Count), Mean Bias (BIAS) and Normalized Mean Bias (NMB) calculated using the linked ozone (O_3) in units of ppb, and temperature (T_{2m}) datasets described in Table 2. Correlation (O_3 bias, T_{2m} bias) = 0.38.

	Var	Sample Size	BIAS	NMB(%)
Case1	O_3	35854	6.16	12
	T_{2m}	35854	-0.36	
Case 2	O_3	1303	4.37	7
	T_{2m}	1303	-2.70	
Case 3	O_3	874	-0.30	0
	T_{2m}	874	-2.64	
Case 4	O_3	3406	1.90	2.
	T_{2m}	3406	-1.90	
Case 5	O_3	3060	-2.17	-3
	T_{2m}	3060	-4.09	
Case 6	O_3	598	0.84	1
	T_{2m}	598	1.56	

Aerosol nitrate was examined in a similar manner. Daily modeled and observed nitrate were matched with the mean observed and modeled relative humidity (2 m) between 0200-0800 LST. The meteorology was evaluated for this period because aerosol nitrate generally peaks in the early morning and is most influenced by high relative humidity (Nenes et al. 1998) which typically occurs overnight into the early morning.

Table 4. Descriptions of NO_3^- and relative humidity datasets from which Table 5 statistics were calculated.

Case	Description
Case1	All instances where mean modeled RH between 0200-0800 LST > 90% and mean mod-obs RH >0
Case 2	Same as Case1, but only instances where mod-obs RH < 0
Case 3	Same as Case 1, but for the period from Jan 01-May 31
Case 4	Same as Case 2, but for the period from Jan 01-May 31

As in the ozone analysis, several cases were chosen to examine this relationship (Table 4). Table 5 provides the nitrate (NO_3^-) and 2 m relative humidity (RH_{2m}) biases for each case. Higher RH_{2m} results in larger NO_3^- concentrations (Nenes et al. 1998). Therefore, if the meteorological model overpredicts the RH_{2m}, the air quality model will likely overpredict NO_3^-. The bias table illustrates this relationship, especially for cases 3 and 4 which include cooler months (Jan-May) when particle phase NO_3^- concentrations are overall higher. Case 3, which includes all data when RH_{2m} >

90% and the RH_{2m} bias is > 0 (model larger than observation) has a significant NO_3^- bias of 0.44 μg·m^{-3}, which is about 50% of the mean observed value. In case 4, the bias is calculated for data where the modeled RH_{2m} > 90%, but the RH_{2m} bias is < 0 (model less than observation). The resulting bias in NO_3^- is reduced to 0.03 μg·m^{-3} or 4% of the mean observed value. To a lesser degree, Case 1, when the RH_{2m} is > 90% and bias is > 0 in the warmer months, the NO_3^- is mean normalized bias lessens from -50%, in the opposite case (Case2) to -15%.The relationship between RH and NO_3^- is stronger than between O_3 and 2 m temperature because the linkage is more direct.

Table 5. Number of data points (Count), Mean Bias (BIAS) and Normalized Mean Bias (NMB) calculated using the linked nitrate (NO_3^-) in units of μg·m^{-3} and percent relative humidity (RH_{2m}) datasets described in Table 4.

	Var	Sample Size	BIAS	NMB(%)
Case 1	NO_3^-	315	-0.06	-15
	RH_{2m}	347	6.8	
Case 2	NO_3^-	373	-0.18	-51
	RH_{2m}	373	-2.8	
Case 3	NO_3^-	225	0.44	51
	RH_{2m}	225	16.62	
Case 4	NO_3^-	97	0.03	4
	RH_{2m}	97	-2.88	

4. DISCUSSION

Results obtained from a recently developed model evaluation tool using year-long MM5 and CMAQ model simulations were presented. The capability of subsetting evaluation results by a wide range of criteria was shown to provide important insights into how the meteorological model performed in a variety of instances. For example, it was shown that arid and mountainous climates were simulated worst in term of temperature. This may imply that the soil model is not suited for strict use in desert regions, and finer model resolution may be required in mountainous areas.

Additionally, it was shown that making the linkage between the air quality and meteorological errors is useful to isolate air quality prediction errors that result from meteorological errors. A weak relationship between ozone errors and 2 m temperature errors was found as well as a strong correlation between aerosol nitrate concentration errors and relative humidity errors. Future efforts will focus on examining the relationship of other chemical species with meteorology, as well as developing more specialized analysis tools for the evaluation toolkit.

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 DW 13921548. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.

5. REFERENCES

Baker, K., 2004: Meteorological modeling protocol for application to $PM_{2.5}$/haze/ozone modeling projects. Lake Michigan Air Directors Consortium, 9pp.

Byun, D.W., and J.K.S. Ching, 1999: Science algorithms of the EPA Models-3 Community Multiscale Air Quality Model (CMAQ) modeling system. EPA/600/R-99/030, U. S. Environmental Protection Agency, Office of Research and Development, Washington, DC 20460.

Eder, B.K., Y. Shaocai, 2004: A performance evaluation of the 2004 release of MODELS-3 CMAQ. 27th NATO/CCMS, Banff, Alb. Canada, Oct. 25-29, 2004.

Grell, G. A., J. Dudhia, and D.R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note, 138 pp., TN-398 + STR, National Center for Atmospheric Research, Boulder, CO.

Hogrefe, C., S. T. Rao, P. Kasibhatla, G. Kallos, C. J. Tremback, W. Hao, D. Olerud, A. Xiu, J. McHenry, and K. Alapaty, 2001: Evaluating the performance of regional-scale photochemical modeling systems: Part 1-meteorological predictions. *Atmospheric Environment*, **34**, 4159-4174.

McKendry, I.G., D.G. Steyn, and G. McBean, 1995: Validation of synoptic circulation patterns simulated by the Canadian Climate Center General Circulation Model for Western North America. *Atmosphere-Ocean*. **33**, 809-825.

McNally, D., 2003: Annual application of MM5 for calendar year 2001, Alpine Geophysics, Arvada, CO, March 2003, 179 pp.

Nenes, A., S. N. Pandis, and C. Pilinis, 1998: ISORROPIA: A New Thermodynamic Equilibrium Model for Multiphase Multicomponent Inorganic Aerosols. *Aquatic Geochemistry*, **4**,123-138.

Stauffer, D. R., and A. Deng, 2003: Improving fine grid meteorological simulations for air quality applications. Second Quarterly Report for CRN project A-46, 18 pp., Dept. of Meteorology, Pennsylvania State University, University Park, Pennsylvania.