4.6 Has the Performance of Regional-Scale Photochemical Modelling Systems Changed over the Past Decade?

C. Hogrefe, J.-Y. Ku, G. Sistla, A. Gilliland, J.S. Irwin, P.S. Porter, E. Gégo, P. Kasibhatla, and S.T. Rao

Abstract This study analyzes summertime ozone concentrations that have been simulated by various regional-scale photochemical modelling systems over the Eastern U.S. as part of more than ten independent studies. Results indicate that there has been a reduction of root mean square errors (RMSE) and an improvement in the ability to capture ozone fluctuations stemming from synoptic-scale meteorological forcings between the earliest seasonal modelling simulations and more recent studies. However, even the more recent model simulations exhibit RMSE values of about 15 ppb and there is no evidence that differences in RMSE between these recent simulations are attributable to systematic improvements in modelling capability. Moreover, it was determined that certain aspects of model performance have not changed over the past decade. One such aspect is that the RMSE of simulated time series can be reduced by applying temporal averaging kernels of up to seven days while the benefit of longer averaging windows appears to vary from year to year. In addition, it is found that spatial patterns simulated by these modelling systems typically have lower correlations and higher centered RMSE than temporal patterns. Analogous to the errors in the simulated time series, these errors in the spatial patterns can be reduced through the application of spatial averaging kernels.

Keywords Regional-scale air quality modelling, model evaluation, model intercomparison

1. Introduction

In the United States, grid-based photochemical modelling systems consisting of separate modules for estimating emissions, meteorology, and air quality have been used for several decades to simulate ozone concentrations, most often in the context of assessing the effectiveness of emission control strategies. While early model applications were limited to a few ozone episodic days, there are increasingly more seasonal, annual and multi-year simulations over the past decade, and the scope of applications has broadened to include air quality forecasting and assessments of the impacts of climate change on air quality. This study analyzes summertime ozone concentrations that have been simulated as part of more than ten independent studies utilizing such modelling systems. While these studies were not coordinated to form harmonized modelling or a controlled ensemble, each of them represents best modelling practices, reflecting the state of science at the time the simulations were performed. The object of this analysis is to assess how our ability to simulate regional-scale ozone concentrations and their variability has changed over the past decade. To this end, we have attempted to quantify the ability of the various simulations to capture temporal and spatial patterns and to characterize model performance on different temporal and spatial scales. No attempt was made to perform diagnostic evaluations for determining the underlying reasons for differences in model behaviour. Moreover, the focus of the analysis is on the comparison between the observed and simulated spatial and temporal patterns rather than on differences in absolute values and biases. Section 2 contains a brief overview of the observations and modelling simulations analyzed in this study. Results are presented in Section 3, and Section 4 discusses the implications of our analysis for various modelling applications.

2. Description of Observations and Model Simulations

The observed daily maximum 8-hour ozone concentrations for the period 1993–2005 were determined from hourly ozone observations at surface monitors from the U.S. EPA's AQS database. In order to be included in the analysis, monitors had to be (a) located within the analysis domain spanning the land area common to all modelling simulations listed in the next section, and (b) have at least 50% non-missing days during June–August between 1993 and 2005. The application of these screening criteria resulted in the selection of 248 monitor locations.

Table 1 Overview of all modelling simulations analyzed in this study.

Abbreviation	Summer Season	Grid Spacing (km)	Met. Model	Chem. Model	Chem. Mech.	Emissions	Reference
M1	1995	12	RAMS	UAM-V	CB4	OTAG 1995	Hogrefe et al. 2001
M2	1995	36	RAMS	UAM-V	CB4	OTAG 1995	Hogrefe et al. 2001
M3	1995	36	MM5	MAQSIP	CB4	OTAG 1995	Kasibhatla and Chameides, 2000
M4	2001	36	MM5	CMAQ4.4	CB4	NEI 2001	Eder and Yu, 2006
M5	2005	12	MM5	CMAQ4.4	CB4	NEI 2001 projected to 2005	Hogrefe et al., 2007a
M6	2001-2003	36	MM5	CMAQ4.5	SAPRC	NEI2001	Nolte et al., 2007
M7	2001	12	MM5	CMAQ4.5	CB4	NEI 2001	Appel et al., 2007
M8	1993-1999	36	MM5	CMAQ4.5	CB4	NEI 1993 - 1999	Hogrefe et al., 2007b
M9	2002	12	MM5	CMAQ4.5	CB4	NEI 2001 2002 Adjustments	Gilliland et al. 2007
M10	2004	12	MM5	CMAQ4.5	CB4	NEI 2001 2004 Adjustments	Gilliland et al. 2007
M11	2002	12	MM5	CMAQ4.5	SAPRC	NEI 2001 2002 Adjustments	Gilliland et al. 2007
M12	2004	12	MM5	CMAQ4.5	SAPRC	NEI 2001 2004 Adjustments	Gilliland et al. 2007
M13	2002	12	MM5	CMAQ4.5	CB4	OTC BaseB1 2002	Ozone Transport Commission, 2007

An overview of all modelling simulations analyzed in this study is provided in Table 1. In the following sections, the various modelling simulations will be referred to by the abbreviations listed in the first column. All simulations are listed roughly in chronological order based on when they were performed as part of various studies. In particular, the simulations for the summer of 1995 by M1–M3 were performed significantly earlier (in the late 1990s) compared to all other simulations analyzed here. Also included in Table 1 is a reference to the publications which provide more details about the individual simulations. For the comparison with observations, model values were extracted for the grid cells in which the 248 monitors described above were located.

3. Results and Discussion

3.1. Model evaluation of temporal and spatial patterns

Model performance for all simulations was summarized through the use of root mean square errors (RMSE) and the match between observed and simulated temporal and spatial patterns. Pattern matching, in turn, was quantified through the use of correlation coefficients, the ratio of simulated to observed standard deviations, and the centered pattern RMSE. The centered pattern RMSE is calculated after the means of observed and simulated ozone concentrations are subtracted from each observed and simulated data point (Taylor, 2001). As shown in Taylor (2001), these correlation coefficients, ratio of standard deviations, and centered pattern RMSE d between observations and model predictions is an array of data points sampled through time and/or space. In this analysis, we focus on the daily maximum 8-hour ozone concentrations and determine the models' ability to capture both the temporal patterns (time series, results are presented in Figure 1a, b) and spatial patterns (maps of concentrations, results are presented in Figure 2a, b). Figure 1a displays the so-called Taylor diagram (Taylor, 2001) in which the match between the observed and simulated spatially-averaged time series of June–August daily maximum 8-hour ozone at 248 monitors for each modelling simulation is indicated by the position of a single letter on a polar diagram. In this polar coordinate system, the position of the letter for a given modelling simulation on the diagram indicates (a) the correlation between observed and simulated time series as angle counter-clockwise from 90°, (b) the ratio of simulated to observed standard deviation of the time series as radial distance from the origin with a radius of 1 corresponding to am exact match between observed and simulated standard deviations, and (c) the centered pattern RMSE as distance from the reference point indicated by a black dot (Taylor, 2001). In terms of correlation coefficients, most simulations with the exception of the simulation of the summer of 1995 by models M1–M3 exhibit values greater than 0.8, ranging as high as 0.95 for the simulation of the summer of 2002 by both models M6 and M10. In terms of standard deviations, almost all simulations underestimate the observed standard deviations, typically by about 20%, but by as much as 40% for the simulation of the summer of 2002 by model M9. The centered pattern RMSE, i.e. the distance from the reference point indicated by a black dot, is largest for the simulations of the summer of 1995 by models M1–M3 and for the summers of 1993, 1994 and 1996 by model M8. Most other modelling simulations show similar performance to each other, with correlations between 0.85 and 0.95 and an underestimation of observed standard deviations by about 20%. While Figure 1a compared the behaviour of observed and simulated time series of daily maximum 8-hour ozone concentrations that were spatially averaged on each day, Figure 1b shows box plots of the total RMSE of the simulated time series of daily maximum 8-hour ozone concentrations, grouped by simulation. The distributions depicted by each box/whisker represent the RMSE of time series calculated separately at each of the 248 monitor locations. It is evident that the simulations of the summer of 1995 by models M1–M3 have a larger RMSE compared to all other simulations. In particular, they show a larger RMSE compared to the simulation of the same summer by model M8 which was performed almost a decade after these earlier simulation, indicating an improvement in model performance as measured by RMSE over this time span. Furthermore, the variations in model performance for more recent simulation periods (2001 or later) do not appear to be attributable to any systematic improvement in modelling capabilities. This is evidenced by the

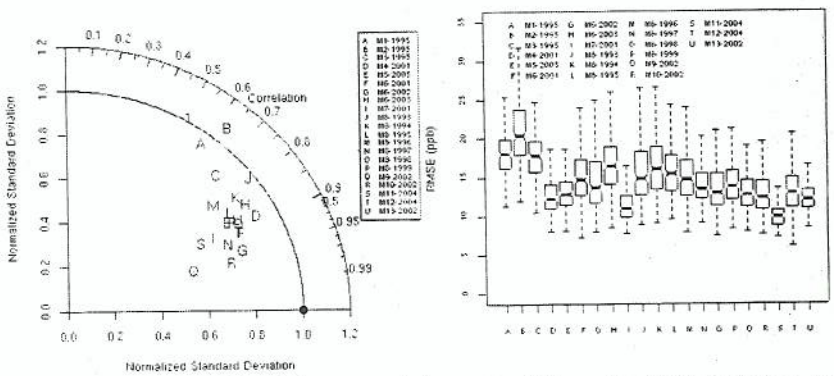

Fig. 1 (a) Taylor diagram of spatially averaged time series, (b) boxplot of RMSE of June–August time series at 248 monitors

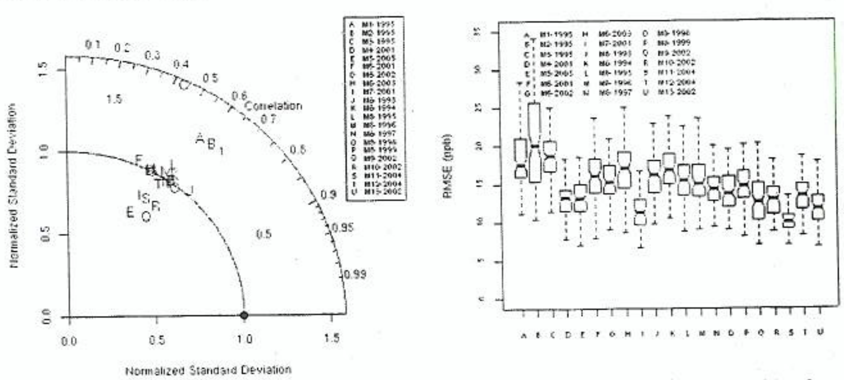

Fig. 2 (a) Taylor diagram of summertime averaged spatial patterns, (b) boxplot of RMSE of spatial patterns across 248 monitor locations on 92 summer days

substantial within-group variations of "F" – "H" and "Q" – "T", two sets of simulations that were each performed as part of a single study. These within-group variations as large or larger than the variations between all of the simulations for more recent time periods.

While Figure 1a, b illustrate the ability of the various simulations to capture temporal patterns (i.e. time series), Figure 2a, b present the corresponding results when spatial patterns are evaluated. For Figure 2a, the Taylor diagram was constructed for the spatial maps of observed and simulated summer-average daily maximum 8-hour concentrations for each simulation. Several features are noteworthy. First, the correlations between observed and simulated spatial patterns of summertime-average ozone concentrations are markedly lower than those between the observed and simulated spatially-averaged time series (Figure 1a). Second, the centered pattern RMSE (as indicated by the distance from the reference point) is similar for all simulations with the exception of the simulation of the summer of 1995 by model M3. Figure 2b shows box plots of the total RMSE of the simulated spatial patterns of daily maximum 8-hour ozone concentrations, grouped by simulation. Similar to the box plot for the time series in Figure 1b, the simulations of the summer of 1995 by models M1–M3 show a larger RMSE compared to other simulations. In particular, they show a larger RMSE compared to the simulation of the same summer by model M8, which was performed almost a decade after these earlier simulation, indicating an improvement in the performance of these modelling systems as measured by RMSE over this time span.

The results presented in this section present evidence that model performance can vary based on the meteorological and emission conditions simulated and can exhibit trends. It is beyond the scope of this study to unequivocally ascribe the changes in model performance to improvements in any particular component of the modelling system (emissions, meteorological modelling, or photochemical modelling). However, variability in meteorology certainly affects model performance; therefore, the following section focuses on investigating the modelling systems' ability to capture the effect of synoptic-scale variations.

3.2. Model performance for synoptic regimes

Because synoptic-scale meteorological conditions exert a significant influence on the ground-level ozone concentrations, it is of interest to evaluate the modelling systems' response to these forcings. To this end, we characterized meteorological conditions through a map-typing procedure applied to gridded fields of mean sea level pressure (MSLP). The Kirchhofer correlation-based map typing procedure was used to determine the 10 most frequent MSLP patterns from all summer days between 1995–2005 (the gridded MSLP fields used in this analysis were not available prior to 1995), and each day was then assigned to the MSLP pattern best representing it. Details of this procedure are described in Hegarty et al. (2007). Next, both observed and simulated average ozone concentrations were computed for each pattern and at each station, and the all-pattern observed or simulated average was subtracted to determine the observed and simulated anomaly for each pattern and each station. Finally, at each station, the correlation between the observed and simulated anomalies across the 10 MSLP patterns was computed, and the results for the medians, 10th and 90th percentiles across all sites are shown in Table 2. In general, median correlations exceed 0.6, indicating that the modelling systems typically catch a substantial portion of the meteorologically-induced ozone variability on the synoptic-scale. In addition, the simulations of the summer of 1995 by models M1–M3 stand out as having the lowest correlations. In particular, the correlations are lower than for the simulation of the same summer by model M8. This suggests that the ability of photochemical modelling systems to capture the phase of synoptic-scale ozone build-up and removal events has improved over the past decade. On the other hand, simulations for more recent time periods exhibit large interannual variability in model performance but no systematic change.

Table 2 Correlation coefficients between the observed and simulated anomalies across the ten MSLP patterns for all model simulations. No results are shown for M8-1993 and M8-1994 because the gridded MSLP fields used in the synoptic typing analysis were not available for these time periods.

	M1	M2	M3	M4	M5	M6-2001	M6-2002	M6-2003	M7	M8-1995	M8-1996	M8-1997	M8-1998	M8-1999	M9	M10	M11	M12	M13
10th %	0.24	0.22	0.22	0.67	0.49	0.66	0.76	0.6	0.64	0.43	0.44	0.67	0.41	0.51	0.77	0.77	0.55	0.59	0.72
Median	0.67	0.68	0.64	0.86	0.81	0.87	0.89	0.79	0.85	0.78	0.73	0.86	0.75	0.85	0.89	0.89	0.8	0.84	0.87
90th %	0.88	0.87	0.88	0.94	0.94	0.94	0.95	0.94	0.94	0.92	0.88	0.95	0.92	0.95	0.95	0.95	0.94	0.94	0.95

3.3. Evaluation of model performance of different temporal and spatial scales

Hogrefe et al. (2001) showed that regional-scale modelling systems typically perform better in capturing signals on time scales longer than one day. To investigate this issue further, we constructed time series of running average one-day, three-day, ..., 31-day time series for both observations and model predictions and computed the standardized centered RMSE of these time series for each averaging period and model simulation. Results are presented as box plots in Figure 3a. The standardized centered RMSE in this plot was normalized by the observed standard deviation to account for reduced variability when averaging kernels are applied. The box/whiskers represent results across the 21 model simulations, for each model simulation, the median time series across all 248 monitors was chosen. The median standardized centered pattern RMSE generally decreases for averaging lengths up to 15 days but then shows an increase for greater averaging lengths. For individual models, the centered pattern RMSE begins to increase even beyond averaging lengths of seven days.

We also investigated the effects of spatial averaging on model performance. To this end, a fine 1 × 1 km grid was overlayed on the analysis domain, and each observation and corresponding model value was assigned to the closest grid cell. Moving-average kernels of 1 × 1, 41 × 41, ..., 601 × 601 grid cells were then applied to these gridded fields, and the standardized centered RMSE was computed for the spatial patterns obtained by each averaging kernel. Only the averaged values at the original monitor locations were considered. Results are shown in Figure 3b. The box/whiskers represent results across the 21 model simulations. It can be seen that spatial averaging decreases the standardized centered pattern RMSE for all model simulations analyzed here for all averaging distances, i.e. all modelling systems are better able to capture the large-scale concentration patterns than localized features in the observed maps.

3.4. Implications

The results presented in the previous sections have considerable implications for applications of regional-scale photochemical modelling systems. Despite a reduction of RMSE and an improvement in the ability to capture ozone fluctuations stemming from synoptic-scale meteorological variability between the earliest seasonal modelling simulations and more recent studies, RMSE of modelled ozone time series still show values of 15 ppb. While this error can be reduced by applying temporal averaging kernels of up to seven days, the benefit of longer averaging windows appears to vary from year to year. For forecasting applications in which temporal averaging is not feasible and where the focus is on predicting single-day peak concentrations, this implies that bias-correction approaches such as the Kalman filter are needed to improve the accuracy of the model-based forecast. Second, spatial patterns simulated by these modelling systems typically have lower correlations and larger centered RMSE than temporal patterns. For studies seeking to utilize model-predicted concentration maps for applications such as health impact assessments, these points to the need for developing and applying statistical techniques aimed at combining information from both observations and model simulations to best represent spatial variability.

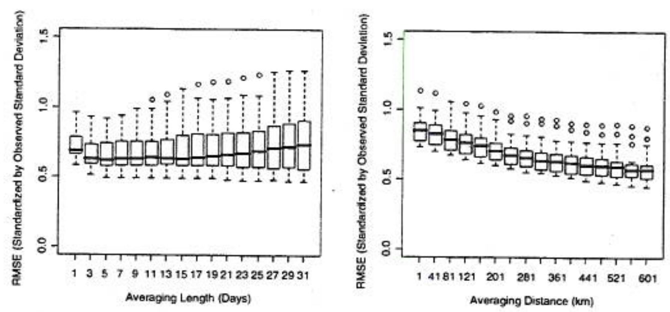

Fig. 3 (a) Boxplots of standardized centered RMSE for simulated time series as function of temporal averaging window length across 21 model simulations. (b) As in (a) but for spatial patterns and spatial averaging windows

Acknowledgments and disclaimer Christian Hogrefe gratefully acknowledges partial support for this work through a research fellowship from the Oak Ridge Institute for Science and Education (ORISE). Part of the work also was performed with support from the New York State Department of Environmental Conservation and under the Memorandum of Understanding between the U.S. Environmental Protection Agency and the U.S. Department of Commerce's National Oceanic and Atmospheric Administration and under agreement number DW13921548. The results presented in this paper do not necessarily reflect the policies or views of the supporting agencies.

References

Appel KW, Gilliland AB, Sarwar G, Gilliam RC (2007) Evaluation of the Community Multiscale Air Quality model version 4.5: Uncertainties and sensitivities impacting model performance; Part I – ozone, Atmos. Environ., Vol. 41, doi:10.1016/j.atmosenv.2007.08.044, pp. 9603–9615.

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

Gilliland AB, Hogrefe C, Godowitch JL, Rao ST (2007) Dynamic evaluation of regional air quality models: assessing changes in O_3 stemming from emissions and meteorology, Atmos. Environ., doi:10.1016/j.atmosenv.2008.02.018.

Hegarty J, Mao H, Talbot R (2007) Synoptic Controls on Summertime Surface Ozone in the Northeastern U.S., J. Geophys. Res., 112, D14306, doi:10.1029/2006JD008170

Hogrefe C, Rao ST, Kasibhatla P, Hao W, Sistla G, Mathur R, McHenry J (2001) Evaluating the performance of regional-scale photochemical modeling systems: Part II – ozone predictions. Atmos. Environ, 35, 4175–4188.

Hogrefe CW, Hao K Civerolo, J-Y Ku, Gaza RS, Sedefian L, Sistla G, Schere K, Gilliland A, Mathur R (2007a) Daily photochemical simulations of ozone and fine particulates over New York State: findings and challenges, J. Appl. Meteor, 46, 961–979.

Hogrefe C, Lynn B, Knowlton K, Goldberg R, Rosenzweig C, Kinney PL (2007b) Long-term regional air quality simulations in support of health impact analyses, preprints, NATO 28th ITM, Aveiro, Portugal, September. 25–29, 2007.

Kasibhatla P, Chameides WL (2000) Seasonal modeling of regional ozone pollution in the eastern United States. Geophys. Res. Lett., 27, 1415–1418.

Nolte CG, Gilliland AB, Hogrefe C (2007) Linking global to regional models to assess future climate impacts on air quality in the United States: 1. surface ozone concentrations. J. Geophys. Res., in press.

Ozone Transport Commission (2007) Draft Modeling Technical Support Document. http://www.otcair.org/projects_details.asp?FID=101&fview=modeling#

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192.

Discussion

S.T. Rao:

Do you expect that there is a limit to model improvement, and do you think that we have reached that now for ozone?

C. Hogrefe:

Yes, I do expect that there is a limit to model improvement. At least in an empirical sense, the results presented in this paper suggest that this limit has been reached for ozone. While there may be the potential for better model predictions of ozone through the use of higher-resolution modelling or updated chemical mechanisms, my expectation is that the resulting improvement in model performance would be incremental at best.

P. Builtjes:

With regard to Peter Builtjes comment about model performance would have been better if high resolution modelling (say 4 km) were performed, you should refer to the study by C. Mass in the Bulletin of American Meteorological Society, which analysed the performance of a meteorological model with two different grid cell size, 12 and 4 km, for a summer season. These results revealed the lack of superiority of the 4 km modelling over the 12 km modelling for the meteorological variables he had analyzed. Reviews of these results, there is no assurance that air quality models would perform better with higher resolution. Of course, there may be case studies where people showed better performance, but these are not long-term simulations. With episodic type modelling (two to three days simulations), there is not enough data to properly evaluate model predictions and errors.

C. Hogrefe:

I agree with this comment. Conceptually, higher-resolution modelling may lead to improved performance in some areas of strong gradients in terrain or emission densities, but this hypothesis can only be tested with longer-term simulations and monitoring networks that are denser than the routine meteorological and air quality networks. Moreover, the analysis you are referring to suggests that there is no systematic improvement in the performance of meteorological models run at 4 km vs 12 km resolution, implying that the potential benefits of higher resolution air quality modelling may also be limited and sporadic

Y.P. Kim:

Since UAM does not contain aerosol module, the ozone levels might be different from the models (e.g. CMAQ) with the aerosol module.

C. Hogrefe:

The effect of ozone-aerosol interactions on simulated ozone concentrations typically is a few ppb or less during summertime conditions. Therefore, I do not expect this effect to a major contributor to the differences in model performance between UAM-V and CMAQ seen in this study.