PERGAMON

Atmospheric Environment 35 (2001) 4175–4188

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

Evaluating the performance of regional-scale photochemical modeling systems: Part II—ozone predictions

Christian Hogrefe[a], S. Trivikrama Rao[a,*], Prasad Kasibhatla[b], Winston Hao[c], Gopal Sistla[c], Rohit Mathur[d], John McHenry[d]

[a] *Department of Earth and Atmospheric Sciences, University at Albany, State University of New York, Albany, NY 12222, USA*
[b] *Nicholas School of the Environment, Duke University, Durham, NC 27708, USA*
[c] *New York State Department of Environmental Conservation, Albany, NY 12233, USA*
[d] *Environmental Programs, North Carolina Supercomputing Center, Research Triangle Park, NC 27709, USA*

Received 01 November 2000; accepted 12 March 2001

Abstract

In this paper, the concept of scale analysis is applied to evaluate ozone predictions from two regional-scale air quality models. To this end, seasonal time series of observations and predictions from the RAMS3b/UAM-V and MM5/MAQSIP (SMRAQ) modeling systems for ozone were spectrally decomposed into fluctuations operating on the intra-day, diurnal, synoptic and longer-term time scales. Traditional model evaluation statistics are also presented to illustrate how the scale analysis approach can help improve our understanding of the models' performance. The results indicate that UAM-V underestimates the total variance (energy) of the ozone time series when compared with observations, but shows a higher mean value than the observations. On the other hand, MAQSIP is able to better reproduce the average energy and mean concentration of the observations. However, both modeling systems do not capture the amount of variability present on the intra-day time scale primarily due to the grid resolution used in the models. For both modeling systems, the correlations between the predictions and observations are insignificant for the intra-day component, high for the diurnal component because of the inherent diurnal cycle but low for the amplitude of the diurnal component, and highest for the synoptic and baseline components. This better model performance on longer time scales suggests that current regional-scale models are most skillful in characterizing average patterns over extended periods, rather than in predicting concentrations at specific locations, during 1–2 day episodic events. In addition, we discuss the implications of these results to using the model-predicted daily maximum ozone concentrations in the regulatory framework in light of the uncertainties introduced by the models' poor performance on the intra-day and diurnal time scales.

Keywords: Air pollution modeling; Ozone models; Time series analysis; Model evaluation; Model uncertainty; Regulatory policies

1. Introduction

Many atmospheric processes contribute to the levels of ozone concentrations observed at the ground level; these processes themselves are influenced by meteorology, chemistry, emissions and land-use patterns, to name a few. Their complex nonlinear interactions are simulated by three-dimensional regional photochemical grid models such as URM (Kumar et al., 1994), UAM-V (Systems Applications International (SAI), 1995), CAMx (ENVIRON, 1997), SAQM (Chang et al., 1997), MAQSIP (Odman and Ingram, 1996; Kasibhatla and Chameides, 2000), MODELS-3 (United States Environmental Protection Agency (US EPA), 1998,

*Corresponding author. Present address: Department of Environmental Conservation, 625 Broadway, Albany, NY 12233–3259, USA. Tel.: +1-518-402-8402; fax: +1-518-402-9035.

E-mail address: strao@dec.state.ny.us (S.T. Rao).

1352-2310/01/$ - see front matter
PII: S1352-2310(01)00183-2

2000), etc. which are being used both in research and in regulatory applications for developing emission control strategies to reduce ozone concentrations to a level below the National Ambient Air Quality Standard (NAAQS) (United States Environmental Protection Agency (US EPA), 1991, 1999). Therefore, as discussed in Part I of this paper (Hogrefe et al., 2001), an evaluation of all aspects of photochemical modeling systems is critical to building confidence in the use of these models for regulatory decision making. In this paper, we focus on the evaluation of ozone concentrations simulated by two photochemical models being used for seasonal air quality simulations.

In the past, photochemical models were applied for the duration of one or a few historical ozone exceedance events, and model evaluation for a particular simulation was usually limited to the comparison of ozone predictions and measurements through a certain set of statistical performance measures (Tesche et al., 1990; United States Environmental Protection Agency (US EPA), 1991, 1994, 1999). This approach to model evaluation has been used extensively in the past (e.g. Tesche et al., 1996). In this study, we introduce the concept of scale analysis (Eskridge et al., 1997; Vukovich, 1997; Rao et al., 1997, 2000b; Sirois et al., 1999) to evaluate the regional-scale photochemical modeling systems, and apply it to the output from seasonal simulations of the time series of ozone concentration values with two different current-generation modeling systems. Preliminary results from the application of this technique to ozone predictions from a different seasonal modeling simulation were reported in Rao et al. (2000b) and Hogrefe et al. (1999). Analyses of ozone precursors and ozone-precursor relationships are presented in Part III of this series of papers (Biswas et al., 2001). By evaluating model performance on different time scales, we are able to illustrate that model performance is time scale specific. The implications of this time scale specific model performance to modeling applications such as exposure assessment/forecasting or research vs. regulatory policy-making are discussed.

2. Description of models and database

Part of the modeling simulations analyzed in this study have been performed as part of separate previous studies by different groups (Lagouvardos et al., 1997; Kasibhatla and Chameides, 2000) with no coordination of modeling options considered. For example, to isolate the effect of chemistry, it would be useful to perform additional seasonal simulations using the meteorological fields from either RAMS or MM5 to drive both photochemical models, but no such simulations have been performed by either group. However, it should be noted that Biswas and Rao (2001) have applied both RAMS and MM5 to the UAM-V photochemical model (Systems Applications International (SAI), 1995) for three episodes and found large differences in predicted ozone concentrations.

The first photochemical model used in this study is UAM-V which has extensively been used for regulatory applications in the past (e.g. Tesche et al., 1996). Simulations were performed for the 1 June–31 August 1995 period on a grid with horizontal grid dimensions of 36 km. The grid extends from 99°W to 67°W and from 26°N to 47°N. Fourteen vertical layers extend from the surface to about 4 km in the UAM-V model, with 10 layers being below 1500 m. The meteorological input fields were derived from a simulation with the RAMS3b model (Walko et al., 1995) and interpolated to the UAM-V grid system. More details on the setup of the meteorological model can be found in Part I of this series of papers (Hogrefe et al., 2001).

The second modeling system analyzed is the Seasonal Model for Regional Air Quality (SMRAQ) (SMRAQ, 1997; Kasibhatla and Chameides, 2000). The photochemical model used in the SMRAQ study is the Multiscale Air Quality Simulation Platform (MAQSIP) (Odman and Ingram, 1996). MAQSIP is a generalized coordinate, modular science process, multiscale air quality modeling system built as a prototype for EPA's Models-3 system, which borrows salient features from the SARMAP Air Quality Model (SAQM) (Chang et al., 1997). The modeling period covered from 15 May 1995 to 11 September 1995, encompassing the time period for which the RAMS3b/UAM-V simulation was carried out. The air quality simulation was carried out with a horizontal grid spacing of 36 km covering the eastern United States (Fig. 1). In the vertical direction, 22 sigma levels were used up to 100 mb, with 11 layers below 1500 m. The meteorological component of the SMRAQ system is the NCAR/PennState Fifth Generation Model

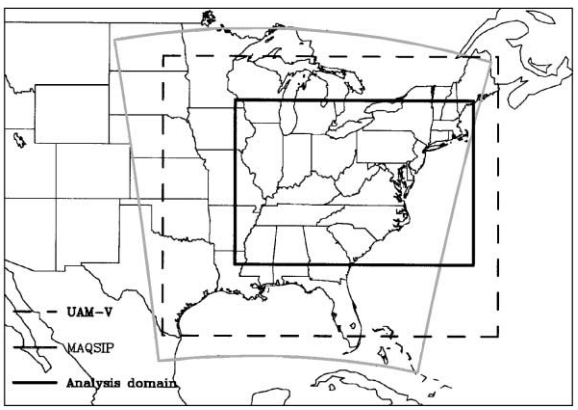

Fig. 1. Map showing the 36 km MAQSIP grid, the 36 km UAM-V grid, and the analysis domain used in this study.

(MM5) (Grell et al., 1994). The MM5 simulation was carried out for the same time period as the air quality simulation and no interpolation of the meteorological fields was necessary since both MM5 and MAQSIP used the same grid projection. Additional details on the setup of MM5 can be found in Part I of this series of papers (Hogrefe et al., 2001).

Both air quality models used the Carbon Bond Mechanism (Gery et al., 1989), but the treatment of boundary conditions was different between the UAM-V and MAQSIP simulations. While time- and height-varying boundary conditions based on surface observations and available ozonesonde data were used in UAM-V, a time-invariant ozone boundary concentration at the background level of 35 ppb was prescribed for MAQSIP at all heights. Emissions for both modeling systems were based on the same available national emissions inventories for anthropogenic emissions, but there were differences in the treatment of sub-grid scale plumes. Biogenic emissions were calculated as a function of model meteorology using the Biogenic Emissions Inventory System 2 (BEIS2) (Geron et al., 1994). Because the temperatures and insolation differed between the two meteorological models, so did the modeled biogenic and mobile source emissions input to the two air quality models. The emission inventories for the UAM-V simulation analyzed in our study were prepared as described by Sistla et al. (2001) for their 12 km UAM-V simulation, and further details on the emission inventory for the MAQSIP simulation can be found in Houyoux et al. (2000).

In this study, we focus on the evaluation of Layer 1 model predictions using the hourly surface ozone observations from the US EPA's AIRS database. The height of the first layer was set at 50 m in RAMS3b/UAM-V and at 38 m in MM5/MAQSIP. The analysis domain for this study extends from 92°W to 69.5°W and 32°N to 44°N (see Fig. 1). Only monitoring data from stations within this analysis domain and corresponding model results are presented; the model results were bilinearly interpolated to the observational sites. To illustrate spatial patterns, a data interpolation algorithm was used (Civerolo and Rao, 2001).

3. Methods of analysis

3.1. Traditional statistics for model evaluation

For model evaluation in air quality management studies, the US EPA stipulated the application of the statistical measures listed in Table 1 to predicted ozone concentrations (United States Environmental Protection Agency (US EPA), 1991). Note that for normalization all of these measures are divided by the observed ozone concentrations at each hour and location rather than normalizing by the mean observed concentration after computation of the evaluation metric. Observation-prediction pairs were often excluded from the analysis if the observed concentration was below a certain cutoff; the cutoff levels varied from study to study but often a level of 60 ppb was used (Russell and Dennis, 2000). The sampling for these statistics can be done through space at each hour (i.e., time series of spatial statistics), through time at each location (i.e., spatial patterns of temporal statistics), or both through space and time (i.e., one number characterizes the entire simulation). Although there is no objective criterion set forth for a satisfactory model performance, US EPA suggested values of 5–15% for the mean normalized bias error (MNBE), 15–20% for the unpaired peak prediction accuracy (UPA), and 30–35% for the mean normalized gross error (MNGE) to be met by modeling simulations being used for regulatory applications. The values for the MNBE and UPA can be either positive or negative. As discussed by Russell and Dennis (2000), these statistics provide little insight into the physical behavior of the model. In the following section, we apply scale analysis to evaluate ozone predictions as an additional tool for performing model evaluations.

Table 1
Definition of the US EPA recommended statistical measures

Mean Normalized Bias Error (MNBE)	$\frac{1}{n}\sum_{i=1}^{n}\frac{C_{\mathrm{mod}}(x,t)-C_{\mathrm{obs}}(x,t)}{C_{\mathrm{obs}}(x,t)}$
Mean Normalized Gross Error (MNGE)	$\frac{1}{n}\sum_{i=1}^{n}\frac{\lvert C_{\mathrm{mod}}(x,t)-C_{\mathrm{obs}}(x,t)\rvert}{C_{\mathrm{obs}}(x,t)}$
Unpaired Peak Prediction Accuracy (UPA)	$\frac{C_{\mathrm{mod}}(x,t)_{\max}-C_{\mathrm{obs}}(x,t)_{\max}}{C_{\mathrm{obs}}(x,t)_{\max}}$

3.2. Definition of time scales and associated processes

It has been documented in the past that time series of ozone observations contain fluctuations occurring on many different time scales (Rao et al., 1997; Vukovich, 1997; Hogrefe et al., 2000; Sebald et al., 2000). Since we analyze hourly concentrations of both ozone observations and model predictions for a time period of three months in this study, the periods that can be resolved range from 2 h to 30–40 days. Spectral analysis of observed ozone time series at various locations for the summer of 1995 reveals that the single largest forcing in the hourly time series data is the diurnal forcing having a period of 24 h. Additional frequency bands of interest

are the intra-day range (periods less than 12 h), the synoptic range (periods of 2–21 days), and longer-term fluctuations (periods greater than 21 days). The choice of these periods is based on the analysis of power spectra as well as on a priori knowledge about the characteristic time scales of different processes affecting ozone concentrations (the intra-day component should include fast-acting, local processes, the diurnal component should be dominated by the 24 h periodicity, the synoptic component should contain fluctuations related to changing synoptic conditions, and the baseline should contain the low-frequency part of the signal). The actual choice of frequency ranges was made to minimize the covariance between the estimates of the different components. Details on the choice of filter parameters for the estimation of spectral components can be found in Rao et al. (1997) and Hogrefe et al. (2000). The definition of these time scales is identical to those for the evaluation of meteorological variables described in Part 1 of this series of papers (Hogrefe et al., 2001).

The atmospheric processes that contribute to intra-day fluctuations of ozone include the effects of turbulent horizontal and vertical mixing, local titration by fresh emissions of NO, and ozone response to fast-changing emission patterns during the rush traffic hours. Diurnal fluctuations in ground-level ozone are associated with the diurnal variation of the solar flux and the resulting differences between daytime photochemical production and nighttime removal of ozone as well as the diurnal cycle of boundary layer evolution and decay. The variations of ozone on the synoptic scale are caused by changing meteorological conditions such as the presence of a near-stagnant high pressure system or the passage of frontal systems. Fluctuations of the baseline are expected to be caused by such processes as the seasonal variation of the solar flux, changing large-scale flow patterns, and changes in vegetation coverage and biogenic emissions.

3.3. Spectral decomposition and log-transform of ozone time series

While any method that can cleanly decompose a time series into fluctuations of the desired time scales can be used, we used the Kolmogorov–Zurbenko (KZ) filter (Zurbenko, 1986) because of its powerful separation characteristics, simplicity, and ability to handle missing data. This technique is described in more detail in Eskridge et al. (1997) and Rao et al. (1997). A log-transform of time series data prior to analysis is a technique frequently used to stabilize the variance of a time series in cases where the local mean of the time series is proportional to the local standard deviation (Milionis and Davies, 1994). Since this is the case for time series of ozone concentrations (Rao and Zurbenko, 1994; Salcedo et al., 1999; Sebald et al., 2000), we analyzed the log-transformed observed and predicted hourly ozone concentrations.

4. Results and discussion

4.1. Traditional statistics for model evaluation

The results of model performance evaluation using the US EPA-recommended statistical measures described in Section 3.1 are presented in Table 2. In this analysis, sampling was performed over both space and time, i.e., these statistics were computed using values from all available stations for the entire simulation period. Table 2 also illustrates how the computed statistics vary for three different cutoff values of 60, 40 and 20 ppb (i.e., observation-prediction pairs are excluded from the analysis if the observed concentration is below the cutoff concentration). It can be seen that both UAM-V and MAQSIP meet the US EPA-recommended criteria of 5–15% for the mean normalized bias, and 30–35% for the normalized gross error for acceptable model performance when the statistics are computed for cutoff levels of 60 and 40 ppb for both hourly and daily maximum concentrations. For the unpaired peak prediction accuracy, only UAM-V satisfies the criterion of having an absolute value not exceeding 15–20%. While there is obviously no impact of the choice of the cutoff value on the UPA (the observed and predicted peaks over the entire domain and season are, of course, larger than 60 ppb), the MNBE and the MNGE show larger values for smaller cutoff concentrations. This increase is larger for hourly concentrations than for the daily maximum concentrations. The increase of both the MNBE and MNGE with decreasing cutoff values is larger for UAM-V than MAQSIP, indicating that MAQSIP has a greater ability to predict lower observed concentrations correctly. More generally, the use of these cutoff-value-dependent performance statistics in the regulatory process might lead to model-tuning geared to accurately predict the peak ozone values without evaluating the capability of the model to properly simulate the ozone accumulation process, i.e., the difference between daily minimum and maximum concentrations. In addition, the regulatory practice of excluding observation-prediction pairs from the analysis if the *observed* concentration is below the cutoff concentration regardless of the predicted concentration creates a bias estimate that is too low—some model underpredictions (observed concentration above and modeled concentration below the cutoff) are included in the estimate, but some overpredictions (observed concentration below and modeled concentration above the cutoff) are excluded. Historically, this "negative bias of the bias estimate" was deemed desirable from a regulatory perspective, since it is presumed that such a

Table 2
US EPA recommended evaluation statistics for UAM-V and MAQSIP ozone predictions

	Cutoff 60 ppb				Cutoff 40 ppb				Cutoff 20 ppb			
	Daily Max.		Hourly		Daily Max.		Hourly		Daily Max.		Hourly	
	UAM-V (%)	MAQSIP (%)	UAM-V (%)	MAQSIP (%)	UAM-V (%)	MAQSIP (%)	UAM-V (%)	MAQSIP (%)	UAM-V (%)	MAQSIP (%)	UAM-V (%)	MAQSIP (%)
UPA	−9	25	−9	25	−9	25	−9	25	−9	25	−9	25
MNBE	5	−5	5	−11	12	1	14	−5	17	6	33	5
MNGE	19	19	20	23	23	22	27	27	27	25	43	36

model would lead to more stringent emission reduction estimates. However, if these statistics are to be used in a scientific model evaluation, only observation-prediction pairs where both the observed *and* modeled concentration are above the cutoff should be included in the analysis.

4.2. Distribution of variance

Since the object of this study is to evaluate the accuracy of model predictions for ozone on different time scales, a first step of this evaluation is to compare the relative importance of the individual components to the overall ozone process for both observations and model predictions. To this end, the variance of each component was computed and divided by the sum of all component variances for both observations and model predictions. The results (Fig. 2) illustrate that the diurnal component is the largest contributor to the overall variance in observations and predictions from both models, followed by the synoptic, baseline and intra-day component. However, there are clear differences between the modeling systems: while UAM-V underestimates the relative contribution of the intra-day and diurnal fluctuations to the overall process variance and overestimates the strength of fluctuations on the synoptic and baseline time scale, MAQSIP predicts a variance distribution that is very close to the observations. Both models underestimate the total variance of the time series significantly, but the underestimation is more severe for UAM-V than for MAQSIP. The numbers in the pie charts in Fig. 2 were computed as spatial averages for all observational sites and the corresponding model predictions and, therefore, provide an average estimate of process energies in observations and model predictions. Clearly, specific sites could show distinctly different patterns depending on which processes are dominant in driving ozone fluctuations at this site (e.g. differences between urban and rural sites). Also, there is a significant amount of covariance between the separated components as indicated by the difference between the sum of the component variances and the variance of the undecomposed time series.

While Fig. 2 provides information about the relative importance of the individual processes in both model predictions and observations, it does not allow us to easily compare the absolute amount of energy on different time scales between observations and model predictions. For this purpose, Table 3 lists the ratios of the variances of the modeled to observed time series for different time scales for both modeling systems. It can be seen that, as discussed above, UAM-V strongly underestimates the variance of the raw (unfiltered) time series, while the underestimation is less severe for MAQSIP. For UAM-V, the underestimation of variance is pronounced on all time scales except for the baseline.

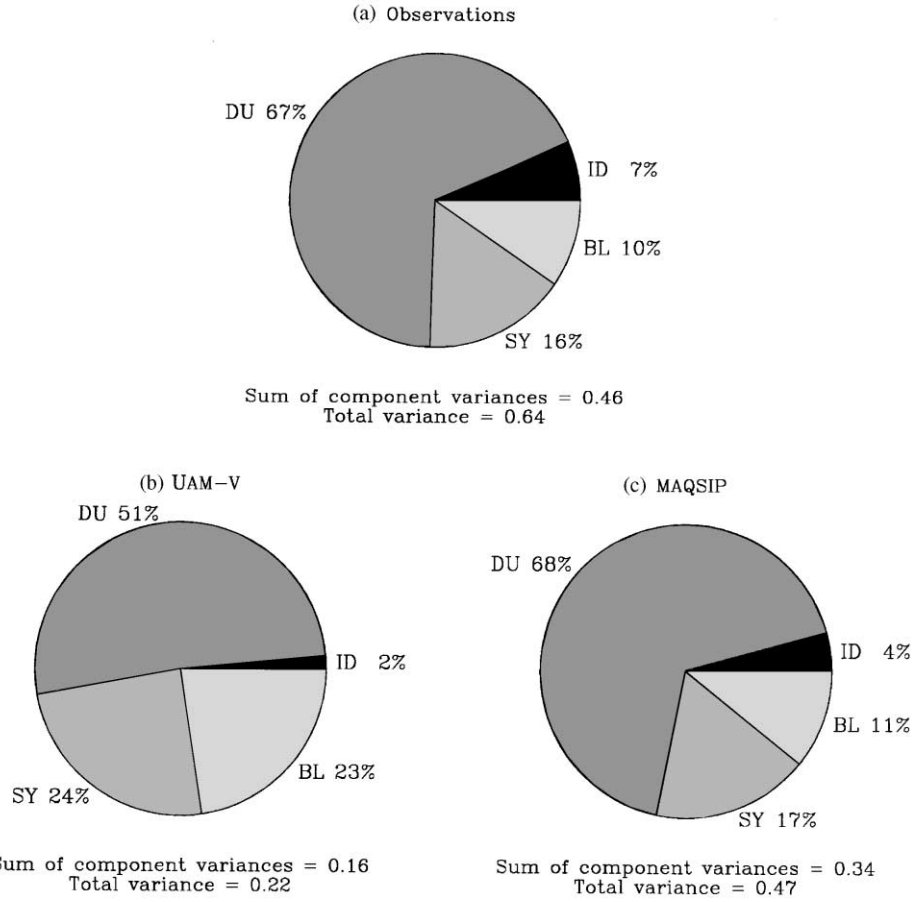

Fig. 2. Pie charts of the relative contribution of the variances of the component time series to the sum of the component variances for observations and model predictions. (a) Observations, and (b) UAM-V, and (c) MAQSIP.

For MAQSIP, on the other hand, only the underestimation of the variance on the intra-day and, to a lesser extent, on the synoptic time scale is significant, while the amount of energy is captured for both the diurnal and baseline time scale. *The only time scale for which the amount of energy is captured by both modeling systems is the baseline time scale*.

A calculation of the bias in the predicted seasonal arithmetic mean ozone concentration (predicted minus observed) indicates that over large regions of the modeling domain UAM-V overpredicts the mean (Fig. 3) while the bias is generally smaller and even negative for larger areas for MAQSIP. In general, both models show a larger bias for the southern part of the modeling domain than for the northern portion. These results combined with the previous analysis of the predicted and observed variability on different time scales suggest that the underprediction of variability by UAM-V (especially for the intra-day and diurnal time scales) is compensated by a higher mean ozone concentration. In other words, the net effect of processes that cause fluctuations (e.g. vertical and horizontal transport, photochemical production, surface removal for the diurnal component) on ozone concentrations is underestimated by UAM-V, and any agreement between observed and predicted daily maximum ozone concentrations is the result of a compensating positive bias. The standard practice of excluding ozone concentrations below 60 ppb from the model evaluation process (United States Environmental Protection Agency (US EPA), 1991) could conceivably exacerbate the situation where the desired agreement between observed and predicted daily maximum ozone concentrations can be reached by simply increasing the mean concentration rather than improving the model's ability to correctly simulate the atmospheric processes leading to the observed variability on these time scales. In contrast to the UAM-V simulation, the MAQSIP simulation was able to better capture the mean and variability present in ozone time series.

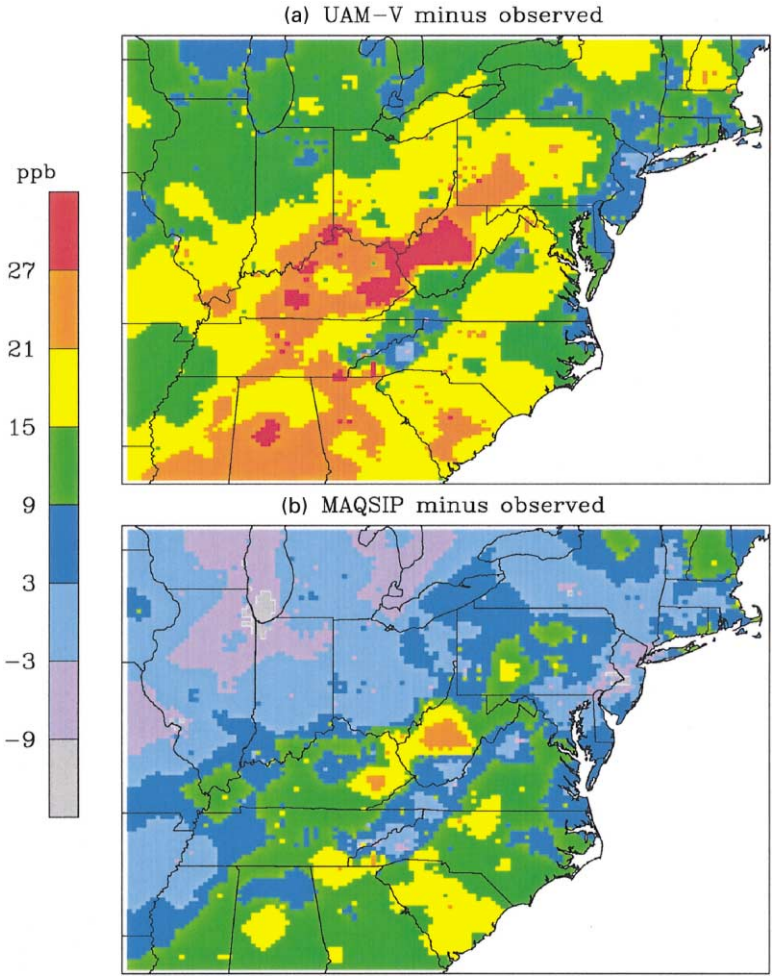

Fig. 3. Bias of the seasonal mean ozone concentration derived from hourly concentrations. The bias is calculated as (modeled) − (observed). (a) UAM-V minus observed, and (b) MAQSIP minus observed.

4.3. Diurnal component

As mentioned before, the diurnal component distinguishes itself from the other components by its well-defined periodicity, namely, 24 h. Furthermore, it is the largest contributor to the overall variance as discussed in the previous section. The observed and predicted diurnal cycles averaged over all stations and the entire summer season and the average diurnal curves for the rates of change are presented in Figs. 4a and b, respectively. It is evident that the two parameters characterizing the diurnal oscillation, namely the amplitude and phase, are different between observations and UAM-V predictions. For these average cycles, UAM-V underestimates the amplitude (compare with Table 3, ratio of variance of predicted to observed diurnal component) and displays a phase lag of about one hour. It can also be seen in the corresponding diurnal curves for the rates of change that UAM-V underestimates the rates of change (i.e. ozone tendancy) during the morning and evening hours. On the other hand, the observed and MAQSIP-predicted diurnal cycles of the diurnal component and the rate of change are very similar.

Morning-hour rates of change (ozone tendencies) are determined by the amount of ozone depletion during nighttime, chemical removal and subsequent production caused by fresh emissions, and the downward mixing of ozone trapped above the nocturnal inversion (e.g. Kleinman et al., 1994; Zhang and Rao, 1999). The underestimation of the ozone tendency by UAM-V suggests that the model does not capture the contributions of these processes to the overall diurnal cycle

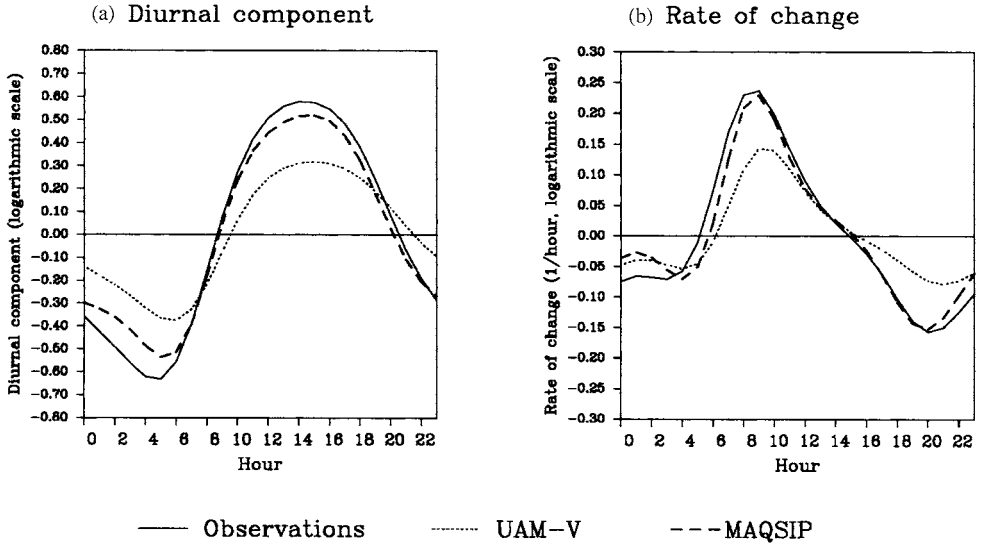

Fig. 4. Average diurnal cycles of ozone observations and model predictions. (a) Diurnal component, and (b) Rate of change of the diurnal component.

Table 3
Ratio of variances of modeled to observed ozone time series on different time scales

	Original	Intra-day	Diurnal	Synoptic	Baseline
UAM-V/obs.	0.39	0.10	0.37	0.57	0.94
MAQSIP/obs.	0.82	0.54	0.97	0.87	0.99

properly and, therefore, future model developments would have to focus on these processes to improve the representation of the diurnal cycle. MAQSIP captures the amplitude and rate of change of the diurnal cycle better than UAM-V. Since peak ozone predictions are very sensitive to the rate of growth of the mixed layer in the morning hours (Berman et al., 1997), an adequate treatment of ozone tendencies in the morning as the boundary layer grows is essential in photochemical modeling (Rao et al., 2001).

4.4. Correlations between the modeled and observed ozone concentrations

In addition to comparing the variances and spatial correlation structures for both observations and model predictions on different time scales, it is of interest to examine the models' capability to reproduce spectral components themselves via spatial images of the correlation coefficient between the observed and modeled time series of each of the components at each observation station for the entire length of simulation (Fig. 5). The correlations for the raw time series range between 0.3 and 0.8 for both modeling systems, with the highest correlations along the eastern seaboard. The correlations for the intra-day component, however, are less than 0.3 almost everywhere, suggesting that the processes contributing to the high-frequency fluctuations in the intra-day component are not captured by either model. The correlations for the diurnal component are high over most of the domain, as is expected due to the inherent diurnal cycle caused by the day and night differences. These correlations are somewhat higher for UAM-V than for MAQSIP. However, when the amplitude of the diurnal component (which is created by taking the difference between the maximum and minimum values of the diurnal component on each day and, thus, does not contain the quasi-sinusoidal pattern of the hourly diurnal component stemming from night/day differences) is considered, the correlations between observations and predictions from both models are poor for most of the modeling domain. The synoptic component displays correlations between 0.4 and 0.7 for the northern part of

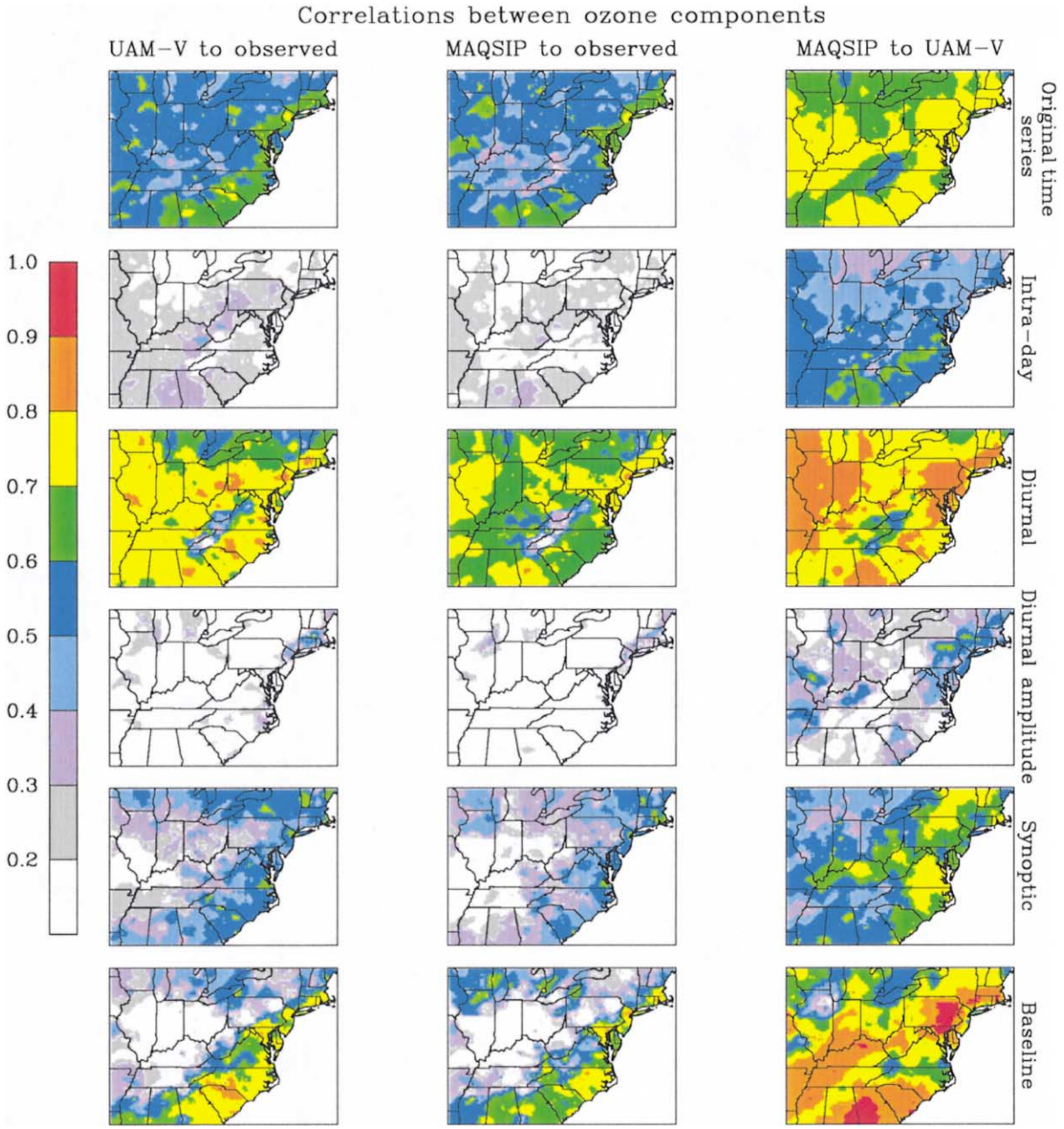

Fig. 5. Correlation between observed and predicted ozone time series on different time scales. Correlations between observations and UAM-V predictions on different time scales are shown in the left column, and correlations between observations and MAQSIP predictions on different time scales are shown in the right column.

the modeling domain and the Atlantic Coast, while the correlations are poor for the central Midwest. This feature is even stronger for the baseline component, where the correlations are higher than 0.7 for most of the Atlantic Coast, but insignificant over large portions of Indiana, Illinois, Ohio and Kentucky. In general, the correlation patterns are remarkably similar between the two modeling systems.

Two important inferences can be drawn from the above analysis. First, model performance as measured by correlation is time scale-specific: the intra-day component and the amplitude of the diurnal component

are poorly captured by both modeling systems, and the longer-term components (synoptic and baseline) on the other hand, show higher correlations. Clearly, the time and space scales needed for ozone forecasting and exposure assessment purposes (namely, the intra-day component and the magnitude of the diurnal component which are needed in addition to an accurate description of the larger-scale background) are not captured by either modeling system. On the intra-day time scale, an increased spatial resolution, both of the meteorological and emission fields, along with a dense observational network, might improve model performance, and, therefore, forecasting and exposure assessment capabilities. An analysis of UAM-V predictions from the simulation with 12 km grid spacing used by Hogrefe et al. (2000) (not presented here) shows that there is no improvement in the simulation of the intra-day component when compared to the 36 km simulation; therefore, presumably a grid spacing even smaller than 12 km would be needed to resolve the intra-day component. This result is consistent with the finding reported in Part I of this paper that also for wind speed fluctuations there is no significant improvement in the correlations for the 12 km simulation compared to the 36 km simulation (Hogrefe et al., 2001). Therefore, the model must have proper representation of dynamical processes operating on the inter-day scale. On the other hand, the longer time scales (synoptic and baseline) which are more relevant to investigations of the effects of emission control strategies on ozone concentrations (Porter et al., 2001) are represented quite well by the models. For regulatory modeling analysis, extended simulation periods (as done in this study) compared to episodic modeling covering smaller spatial domains appear to make the best use of the models' skills rather than increased resolution for shorter simulation periods. Second, the separation of the original time series into different spectral components allows us to identify the time scales that are responsible for the correlation structure in the raw data. In this case, it was shown that poor correlations in the Midwest are due to a poor representation of the synoptic forcing and especially the baseline component. In other words, to improve model performance in the central Midwest, a more accurate description of processes acting on time scales greater than 2 days is necessary. While our procedures cannot exactly determine which processes are poorly represented, these processes might include the characterization of changing deposition patterns due to growing and harvesting in the agricultural areas in the Midwest.

4.5. Implications to model applications in the regulatory setting

In this section, we translate the result of time scale–dependent model performance—as measured by correlations between observed and model predicted component time series—to an estimate of model uncertainty and discuss the implications to model applications in the regulatory setting. To this end, we introduce the concept of 'inherent' and 'reducible' uncertainty. We define the 'inherent' uncertainty to be the inability of the grid-based models to capture the observed fluctuations that are caused by processes acting on scales that are not resolvable with the grid cell size used in the model simulation.'Reducible' uncertainty arises from imperfect scientific understanding on how to best describe certain atmospheric processes that can be resolved by the models (e.g. the inability of SMRAQ to capture the day-to-day variations of the diurnal amplitude despite the simulation of clouds, the uncertainty about the proper parameterization for the PBL evolution in mesoscale models, inadequacies in the model input data). It manifests itself in the model-to-model and model-to-observation differences of ozone concentrations predicted by state-of-science modeling systems using different scientifically-sound process formulations, and model users are confronted with this 'reducible' uncertainty when applying a model to a particular situation. While the 'inherent' uncertainty is the theoretical lower bound of model uncertainty even for a 'perfect' model, the sum of 'inherent' and 'reducible' uncertainties is still a lower bound for the total modeling uncertainty in practical applications.

In light of the above discussions, the failure of the models to predict the intra-day component (as indicated by the poor correlations in Fig. 5) should be viewed as the 'inherent' uncertainty when the application of the models is restricted to spatial scales coarser than that needed to resolve the largest eddy motions. It is also noteworthy that the overall model uncertainty is not necessarily reduced as the grid resolution is increased to reduce 'inherent' uncertainty. For example, even if the grid resolution were fine enough to explicitly model eddies and, thus, capture the variance at the intra-day time scale, we would expect problems of predicting the eddy motions at the right time and place due to uncertainty in large eddy simulations. The standard deviation of the observed intra-day component during afternoon hours (1200–1700 EST) was calculated at each station for the summer of 1995 (June–August); the exponent of the resulting standard deviation, when multiplied by 100, then gives the percentage uncertainty in ozone concentration caused by the intra-day component. Fig. 6 presents a map of the 'inherent' uncertainty in predicting the daily maxima due to the models' inability to simulate intra-day fluctuations. The map illustrates that the 'inherent' uncertainty ranges from 4% to 14%, with higher values in the urban areas and lower values in the rural areas. The 'inherent' uncertainty averaged over all rural stations is 7%, while it is 11% for urban stations. As noted above, this 'inherent'

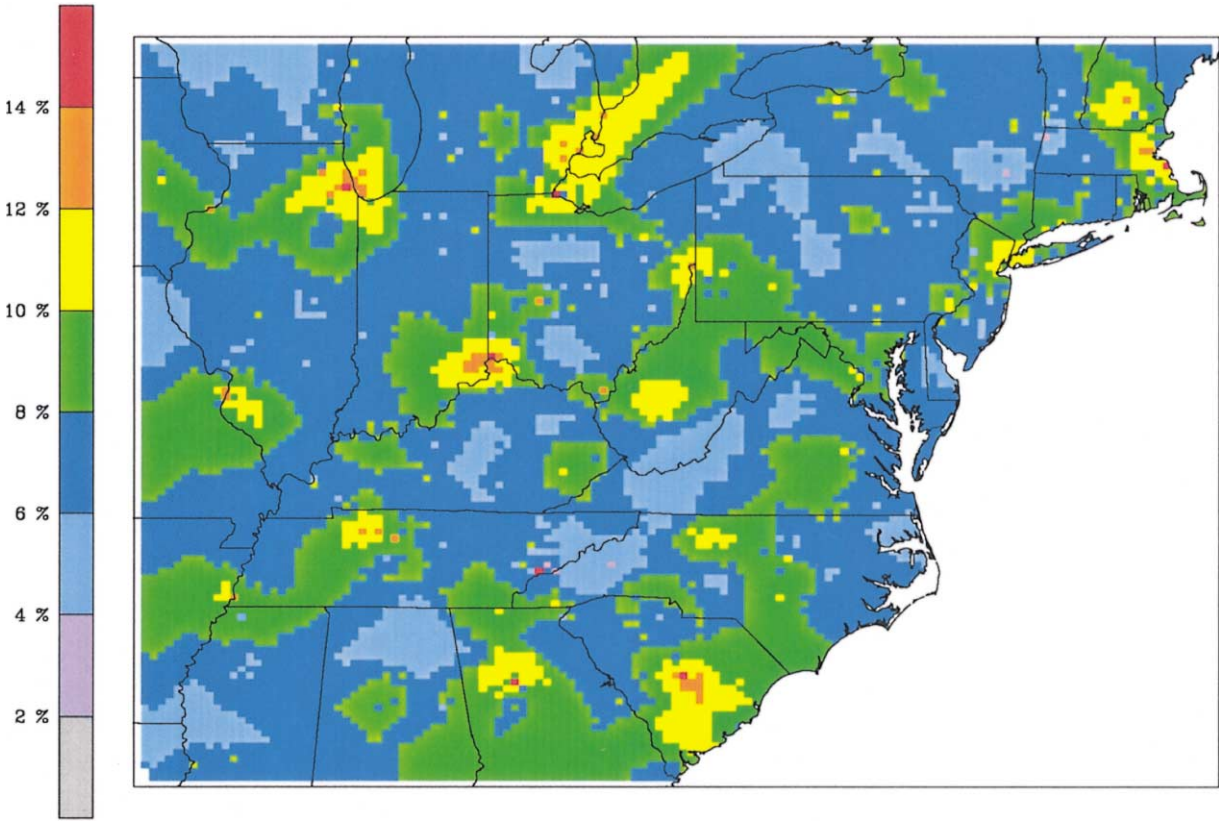

Fig. 6. Spatial map depicting the 'inherent' uncertainty in predicting daily maximum ozone concentrations.

uncertainty has to be viewed as the theoretical lower bound of model uncertainty even for a 'perfect' model.

As stated above, 'reducible' uncertainty arises from our imperfect scientific understanding on how to best describe certain processes (PBL evolution, soil moisture, land use/land cover, emissions, etc.) on scales that are resolvable with the grid spacing used in the simulation. The results presented in this study suggest that the largest part of the 'reducible' uncertainty in the current generation models stems from the models' inability to properly predict the day-to-day variations of the diurnal amplitude. The effect of meteorology, especially vertical mixing and ventilation, on model-predicted ozone concentrations was also illustrated by Biswas and Rao (2001) who found that the uncertainty of daily maximum ozone predictions due to the use of two different state-of-science mesoscale meteorological models in the air quality simulation was on the order of 20%. Since the simulations by Biswas and Rao (2001) were carried out only for the duration of three episodes and, therefore, little synoptic or baseline scale variability was modeled, most of this uncertainty has to be attributed to the uncertainty of predicting the diurnal component of ozone. The less-than-perfect model performance on the synoptic and baseline scales also contributes to the 'reducible', but to a much lesser extent than the processes on the diurnal scale since model-to-observation and model-to-model correlations are higher (Fig. 5).

The use of 8 h daily maximum ozone concentrations for regulatory purposes, as promulgated recently by the US EPA, reduces the influence of the intra-day fluctuation on the daily maxima (i.e., the 'inherent' uncertainty associated with individual 8 h daily maximum values) to about 1%. The failure to properly capture the processes on the diurnal time scale, however, still introduces uncertainty to the 8 h daily maximum ozone predictions, which is part of the 'reducible' uncertainty confronting model users as discussed above.

As an important implication to the use of models for regulatory purposes, our result of better model performance on longer time scales supports the approach of averaging the predicted daily maximum ozone concentrations when using these models for regulatory purposes as suggested by Sistla et al. (2001). This view is also supported by a recent study (Rao et al., 2000a) that investigated the differences in the daily maximum ozone predictions from a variety of episodic simulations with

various combinations of state-of-the-art photochemical modeling systems.

In summary, the use of grid-based photochemical models for predictions of maximum ozone concentrations on individual days at individual grid cells has substantial uncertainties that—at this point—warrant against using them for this purpose in the regulatory setting. These uncertainties are caused not only by the inadequate spatial resolution of the models themselves, imperfect scientific understanding of the underlying processes, and random and systematic errors in the input data used for photochemical model simulations, but also the lack of sufficient and adequate spatio-temporal resolution of required input fields (e.g. emissions, soil moisture). It is important to realize that a modeling system is only as strong as its weakest link. However, since the coarser scale models can capture the critically important longer-temporal signals quite well, their use in policy-making applications is appropriate—particularly in light of the growing understanding that longer time scales need to be considered when assessing emission control strategies (Porter et al., 2001; Hogrefe et al., 2000). This latter view—which supports the importance of long-range spatio-temporal processes (the BL, SY, and to some extent DU signals) in the ozone problem—is also reflected in the current US EPA guidance on the use of photochemical models that calls for the use of ozone concentrations averaged over all episode days modeled rather than individual daily maximum concentrations for policy purposes (United States Environmental Protection Agency (US EPA), 1999). In related studies, Kasibhatla and Chameides (2000) also argue that current-generation modeling systems are better able to predict the ensemble of events that shape the spatial distribution of ozone on a seasonal scale rather than predict individual episodes, and Bouchet et al. (1999) report that the climatology of ozone in July during a five-year period is better characterized by an air quality model than the ozone distribution during a specific time period. The results of our study lend further support to these viewpoints.

5. Summary

In this study, ozone predictions from two photochemical modeling systems were evaluated using the scale analysis concept. Seasonal time series of observations and predictions for ozone from the RAMS3b/UAM-V and MM5/MAQSIP modeling systems were spectrally decomposed into intra-day, diurnal, synoptic and baseline time scales. While traditional statistics for model evaluation reveal the presence of a positive bias in the RAMS3b/UAM-V simulation, they do not provide any further insight into the strengths and shortcomings of the models. Scale analysis results reveal that UAM-V underestimates the relative strength of the intra-day and diurnal fluctuations and overestimate the relative strength of the longer-term fluctuations while MAQSIP approximately captures the relative contributions of all components. The absolute variance is underestimated by both modeling systems; however, the underestimation is more severe for UAM-V. Whereas UAM-V overestimates mean concentrations, the bias is much smaller for MAQSIP. The standard practice of excluding ozone concentrations below 60 ppb from the model evaluation process (United States Environmental Protection Agency (US EPA), 1991) can mask this behavior of UAM-V.

A detailed analysis of the average diurnal cycle reveals that UAM-V does not capture its amplitude and its rate of change in the morning hours. A plausible explanation is the inability of the model to properly simulate the effects of clouds, vertical mixing, and nighttime removal processes. The average diurnal cycle is properly captured by MAQSIP. The poor simulation of the diurnal amplitude in UAM-V points to shortcomings in the model's treatment of processes that dominate the diurnal cycle of ozone that need to be addressed in future model improvements. Correlations between model predictions and observations are insignificant for the intra-day component, high for the diurnal component because of the inherent diurnal cycle but low for the amplitude of the diurnal component, and highest for the synoptic and baseline components for both modeling systems. While the poor performance on the intra-day time scale is at least partially related to the horizontal grid dimension used in the model and the inability of the input fields (meteorology, emissions, soil moisture, vegetation cover, etc.) to capture this scale in time and space, the poor correlation for the amplitude of the diurnal cycle for both models points to uncertainties associated with modeling processes on this time scale that need to be addressed in future model improvements. A decrease in grid spacing from 36 to 12 km had not led to a better model performance for the intra-day component; therefore, a grid spacing even smaller than 12 km would be needed presumably to improve correlations on the intra-day component. Future studies using higher resolution modeling could employ the spectral decomposition technique introduced in this paper to demonstrate the potential benefits of high resolution modeling.

In sum, the better performance of the model on longer time scales suggests that modeling periods should be longer than the duration of a single episode to increase confidence in the regulatory modeling process. In addition, the results illustrate that predictions of the daily maximum ozone concentrations on individual days at individual grid cells are subject to considerable uncertainty. Therefore, it is important to average the daily maximum ozone concentrations over longer time

periods since it is in better accord with the strengths of the model to best characterize the synoptic and baseline time scales. This longer-term averaging is also most relevant to regulatory policies aimed at meeting and maintaining the standards.

Acknowledgements

This work is supported by the U.S. Environmental Protection Agency under grant R825260-01-0, EPRI under contract WO3189-12, and the New York State Energy Research & Development Authority under contract No. 6085. Additional support is provided by grants EA8009 from NC-DENR to Duke University and subgrant G-35-W62-G2 from Georgia Tech to Duke University (prime grant R826372-01-0 from US EPA). The authors would like to thank Mr. Patrick Dolwick of US EPA for assistance with the model simulations. Thanks are extended to Professors Igor Zurbenko and Steven Porter for many helpful discussions relating to the scale analysis. The authors would also like to thank the anonymous reviewers for helping us in improving the clarity of the paper.

References

Berman, S., Ku, J.Y., Zhang, J., Rao, S.T., 1997. Uncertainties in estimating the mixing depth-comparing three mixing-depth models with profiler measurements. Atmospheric Environment 31, 3023–3039.

Biswas, J., Rao, S.T., 2001. Uncertainties in episodic ozone modeling stemming from uncertainties in meteorological fields. Journal of Applied Meteorology 40, 117–136.

Biswas, J., Rao, S.T., Hogrefe, C., Hao, W., Sistla, G., 2001. Evaluating the performance of regional-scale photochemical modeling systems: Part III—precursor predictions and ozone-precursor relationships. Atmospheric Environment 35, (in press).

Bouchet, V.S., Laprise, R., Torlaschi, E., McConnell, J.C., Plummer, D.A., 1999. Studying ozone climatology with a regional climate model 2. Climatology. Journal of Geophysical Research 104, 30373–30385.

Chang, J.S., Jin, S., Li, Y., Beauharnois, M., Lu, C.H., Huang, H.C., Tanrikulu, S., DaMassa, J., 1997. The SARMAP Air Quality Model, Final Report. Air Resource Board, California Environmental Protection Agency, Sacramento, CA.

Civerolo, K., Rao, S.T., 2001. Space–time analysis of precipitation-weighted sulfate concentrations over the eastern United States. Atmospheric Environment (in press).

ENVIRON, 1997. User's guide to the Comprehensive Air Quality Model with Extensions (CAMx). ENVIRON International Corporation, 101 Rowland Way, Novato, CA 94945.

Eskridge, R.E., Ku, J.Y., Rao, S.T., Porter, P.S., Zurbenko, I.G., 1997. Separating different scales of motion in time series of meteorological variables. Bulletin of the American Meteorological Society 78, 1473–1483.

Geron, C.D., Guenther, A.B., Pierce, T.E., 1994. An improved model for estimating emissions of volatile organic compounds from forests in the eastern United States. Journal of Geophysical Research 99, 12773–12791.

Gery, M.W., Whitten, G.Z., Killus, J.P., Dodge, M.C., 1989. A photochemical kinetics mechanism for urban and regional-scale computer modeling. Journal of Geophysical Research 94, 12925–12956.

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

Hogrefe, C., Rao, S.T., Kasibhatla, P., 2001. Evaluating the performance of regional-scale photochemical modeling systems: Part I—meteorological input variables. Atmospheric Environment 35, (in press).

Hogrefe, C., Rao, S.T., Zurbenko, I.G., Porter, P.S., 1999. Seasonal photochemical modeling over the eastern United States: model performance on different space and time scales and implications to the use of models in a regulatory setting. Preprints, Symposium on Interdisciplinary Issues in Atmospheric Chemistry, Dallas, TX, American Meteorological Society, pp. 42–49.

Hogrefe, C., Rao, S.T., Zurbenko, I.G., Porter, P.S., 2000. Interpreting the information in ozone observations and model predictions relevant to regulatory policies in the eastern United States. Bulletin of the American Meteorological Society 81, 2083–2106.

Houyoux, M.R., Vukovich, J.M., Coats Jr., C.J., Wheeler, N.J.M., Kasibhatla, P., 2000. Emission inventory development and processing for the seasonal model for regional air quality. Journal of Geophysical Research 105, 9079–9090.

Kasibhatla, P., Chameides, W.L., 2000. Seasonal modeling of regional ozone pollution in the eastern United States. Geophysical Research Letters 27, 1415–1418.

Kleinman, L., Lee, Y.-N., Springston, S.R., Nunneracker, L., Zhou, X., Brown, R., Hallock, K., Leahy, D., Lee, J.H., Newman, L., 1994. Ozone formation at a rural site in the southeastern United States. Journal of Geophysical Research 99, 3469–3482.

Kumar, N., Odman, M.T., Russell, A.G., 1994. Multiscale air quality modeling: application to Southern California. Journal of Geophysical Research 99, 5385–5397.

Lagouvardos, K., Kallos, G., Kotroni, V., 1997. Modeling and analysis of ozone and its precursors in the northeast U.S.A. (Atmospheric Model Simulations). Final Report to Electric Power Research Institute, Palo Alto, CA, 46 pp. (Available from Office of Science and Technology, NYSDEC, 50 Wolf Rd., Albany, NY 12233–3259).

Milionis, A.E., Davies, T.D., 1994. Regression and stochastic models for air pollution—II: application of stochastic models to examine the links between ground-level smoke concentrations and temperature inversions. Atmospheric Environment 28, 2811–2822.

Odman, M.T., Ingram, C.L., 1996. Multiscale Air Quality Simulation Platform (MAQSIP): source code documentation and validation. Technical Report, ENV-96TR002-v1.0. MCNC, Research Triangle Park, NC 27709, 83 pp.

Porter, P.S., Rao, S.T., Zurbenko, I.G., Dunker, A.M., Wolff, G.T., 2001. Ozone air quality over North America: part II—an analysis of trend detection and attribution techniques. Journal of the Air and Waste Management Association 51, 283–306.

Rao, S.T., Hogrefe, C., Mao, H., Hao, W., Ku, J.-Y., Sistla, G., Zhang, K., Biswas, J., Kasibhatla, P., Hansen, D.A., 2000a. Using photochemical modelling systems in designing emissions control strategies. Proceedings of the 15th International Clean Air & Environment Conference, Sydney, Australia, 26–30 November, 2000.

Rao, S.T., Hogrefe, C., Sistla, G., Wu, S.-Y., Hao, W., Zalewsky, E.E., Zurbenko, I.G., Porter, Kallos, G., Lagouvardos, K., Kotroni, V., Hansen, D.A., Smith, M., 2000b. An integrated modeling and observational approach for designing ozone control strategies for the eastern United States. Air Pollution Modeling and its Applications XIII. Kluwer Academic Plenum Publishers, New York, NY, pp. 3–16.

Rao, S.T., Ku, J.-Y., Berman, S., Zhang, K., Mao, H., 2001. Summertime characteristics of the atmospheric boundary-layer and relationships to ozone levels over the Eastern United States. Pure and Applied Geophysics (in press).

Rao, S.T., Zurbenko, I.G., 1994. Detecting and tracking changes in ozone air quality. Journal of Air and Waste Management Association 44, 1089–1092.

Rao, S.T., Zurbenko, I.G., Neagu, R., Porter, P.S., Ku, J.Y., Henry, R.F., 1997. Space and time scales in ambient ozone data. Bulletin of the American Meteorological Society 78, 2153–2166.

Russell, A., Dennis, R., 2000. NARSTO critical review of photochemical models and modeling. Atmospheric Environment 34, 2283–2324.

Salcedo, R.L.R., Alvim Ferraz, M.C.M., Alves, C.A., Martins, F.G., 1999. Time series analysis of air pollution data. Atmospheric Environment 33, 2361–2372.

Sebald, L., Treffeisen, R., Reimer, E., Hies, T., 2000. Spectral Analysis of Air Pollutants. Part 2: ozone time series. Atmospheric Environment 34, 3503–3509.

Sirois, A., Pudykiewicz, J.A., Kallaur, A., 1999. A comparison between simulated and observed ozone mixing ratios in eastern North America. Journal of Geophysical Research 104, 21397–21423.

Sistla, G., Hao, W., Ku, J.-Y., Kallos, G., Lagouvardos, K., Papadopoulos, A., Zhang, K., Mao, H., Rao, S.T., 2001. An operational evaluation of two regional-scale ozone air quality modeling systems over the eastern United States. Bulletin of the American Meteorological Society 82, 945–964.

SMRAQ, 1997. The SMRAQ project: development and application of a seasonal model for regional air quality. Available online at http://envpro.ncsc.org/SMRAQ.

Systems Applications International (SAI), 1995. Users Guide to the Variable Grid Urban Airshed Model (UAM-V). Systems Applications International, San Rafael, CA, 131 pp. (Available from Systems Applications International, 101 Lucas Valley Rd., San Rafael, CA 94903).

Tesche, T.W., Georgopoulos, P., Seinfeld, J.H., Lurmann, F., Roth, P.M., 1990. Improvements in procedures for evaluating photochemical models. Report A832-103, California Air Resources Board, Sacramento, CA.

Tesche, T.W., McNally, D.E., Wilkinson, J.G., Loomis, C.F., Emigh, R.A., 1996. Initial estimates of emissions reductions needed for ozone attainment in the Pittsburgh-Beaver Valley ozone nonattainment area. Report prepared for: Southwestern Pennsylvania Clean Air Stakeholder Group, Alpine Geophysics, LLC, Covington, KY.

United States Environmental Protection Agency (US EPA), 1991. Guideline for regulatory application of the Urban Airshed Model. EPA-450/4-91-013, July 1991, United States Environmental Protection Agency, Research Triangle Park, NC 27711.

United States Environmental Protection Agency (US EPA), 1994. Guidance on Urban Airshed Model (UAM) reporting requirements for attainment demonstrations. EPA-454/R-93-056, United States Environmental Protection Agency, Research Triangle Park, NC 27711.

United States Environmental Protection Agency (US EPA), 1998. EPA third-generation air quality modeling system. Models-3 Volume 9B: User manual. EPA-600/R-98/069(a), June 1998, United States Environmental Protection Agency, Research Triangle Park, NC 27711.

United States Environmental Protection Agency (US EPA), 1999. Draft report on the use of models and other analyses in attainment demonstrations for the 8hour ozone NAAQS. EPA-44/R-99-0004, May 1999, United States Environmental Protection Agency, Research Triangle Park, NC 27711.

United States Environmental Protection Agency (US EPA), 2000. Models-3 Home Page. Available online at http://www.epa.gov/asmdnerl/models3/.

Vukovich, F.M., 1997. Time scales of surface ozone variations in the regional, non-urban environment. Atmospheric Environment 31, 1513–1530.

Walko, R.L., Tremback, C.J., Hertenstein, R.F.A., 1995. RAMS-The Regional Atmospheric Modeling System, Version 3b, User's guide. ASTER Division, Mission Research Corporation, Fort Collins, CO.

Zhang, J., Rao, S.T., 1999. On the role of vertical mixing in the temporal evolution of the ground-level ozone concentrations. Journal of Applied Meteorology 38, 1674–1691.

Zurbenko, I.G., 1986. The Spectral Analysis of Time Series. North-Holland, Amsterdam, 236 pp.