Evaluating the performance of regional-scale photochemical modeling systems: Part I—meteorological predictions

Christian Hogrefe[a], S. Trivikrama Rao[a,*], Prasad Kasibhatla[b], George Kallos[c], Craig J. Tremback[d], Winston Hao[e], Don Olerud[f], Aijun Xiu[f], John McHenry[f], Kiran Alapaty[f]

[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] *Department of Physics, University of Athens, Greece*
[d] *Mission Research Corporation/ASTeR Division, Fort Collins, CO 80522, USA*
[e] *New York State Department of Environmental Conservation, Albany, NY 12233, USA*
[f] *Environmental Programs, North Carolina Supercomputing Center, Research Triangle Park, NC 27709, USA*

Abstract

In this study, the concept of scale analysis is applied to evaluate two state-of-science meteorological models, namely MM5 and RAMS3b, currently being used to drive regional-scale air quality models. To this end, seasonal time series of observations and predictions for temperature, water vapor, and wind speed 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 method of spectral decomposition can help provide additional insight into the models' performance. The results indicate that both meteorological models under-represent the variance of fluctuations on the intra-day time scale. Correlations between model predictions and observations for temperature and wind speed are insignificant on the intra-day time scale, 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 longer-term components. This better model performance on longer time scales suggests that current regional-scale models are most skillful for characterizing average patterns over extended periods. The implications of these results to using meteorological models to drive photochemical models are discussed. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Air pollution modeling; Meteorological models; Model evaluation; Time series analysis; Spectral analysis; Model uncertainty; Ozone modeling

1. Introduction

Three-dimensional photochemical modeling systems 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, 2000), etc. are the primary tools being applied by state and federal agencies for developing emission control strategies to reduce ambient ozone concentrations to a level below the National Ambient Air Quality Standard (NAAQS) (United States Environmental Protection Agency (US

*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).

EPA), 1991, 1999). Therefore, an evaluation of all components of the photochemical modeling systems is critical to building confidence in the use of these types of models for regulatory purposes. In a series of three papers, we introduce a new concept to model performance evaluation and apply it to analyze meteorological input parameters, ozone predictions, predictions of ozone precursors, and predictions of ozone-precursor relationships. In this first paper of the series, we focus on the evaluation of meteorological fields simulated by two mesoscale meteorological models being widely used for air quality simulations.

In the previous studies, meteorological models used in air pollution modeling (Hanna, 1994; Lyons et al., 1995; Olerud and Wheeler, 1997) have been evaluated using traditional statistical measures as well as qualitative assessments such as visual examinations of observed and simulated wind and temperature fields. As some of these studies noted, the evaluation of meteorological models being used in air pollution modeling is an 'inexact science'; the observed and predicted meteorological fields are not independent because of the use of four-dimensional data assimilation (4DDA), and there are no clear pass/fail criteria for assessing the performance of meteorological models being used in air quality simulations (Hanna, 1994; Olerud and Wheeler, 1997; Pielke and Uliasz, 1998). In addition, comparisons between observations and model predictions are complicated by the fact that observations are point measurements while model predictions are Reynold's average mean state variables.

In this study, we introduce the concept of scale analysis (Eskridge et al., 1997; Vukovich, 1997; Rao et al., 1997, 2000) to evaluate regional-scale photochemical modeling systems, and apply it to the output from seasonal simulations from two different current-generation modeling systems. The work presented in this study builds upon the preliminary comparisons between observations and model predictions for meteorological variables performed by Hogrefe and Rao (2000). By evaluating model performance on different time scales using a spectral filter, we are able to quantify the scale dependence of model performance. The results are discussed from the perspective of regulatory applications of meteorological/photochemical modeling systems. In the context of the other two papers in this series (Hogrefe et al., 2001; Biswas et al., 2001), an evaluation of the meteorological input fields applied in the photochemical model is intended to determine the time scales on which air quality predictions can—or cannot—be expected to capture the salient temporal and spatial features embedded in air quality observations. If, for example, wind speed (transport) predictions on a certain time scale were shown to be inconsistent with observations, one cannot expect ozone predictions from

the photochemical model to be accurate for the right reasons on that time scale.

2. Description of model and database

Both meteorological modeling simulations analyzed in this study have been performed as part of separate previous studies (Lagouvardos et al., 1997; SMRAQ, 1997a, b). As a result, modeling options such as cloud parameterization, soil moisture, 4DDA, etc., were treated differently in both simulations.

2.1. RAMS3B

The first meteorological model used in this study is RAMS3b (Walko et al., 1995), whose application to air quality studies was described by Lyons et al. (1995). The RAMS3b simulations were performed for the 1 June–31 August 1995 period on three nested grids with horizontal grid dimensions of 108, 36 and 12 km, respectively (Fig. 1). Four-dimensional data assimilation was employed using input fields blended from surface observations and the analysis fields available from the European Center for Medium Range Forecasting (ECMWF). The time interval for both the ECMWF fields and the surface observations was 6 h. Specifically, model predictions for wind, temperature and water vapor were nudged towards these blended input fields with a nudging factor of $4.62 \times 10^{-5} \, \text{s}^{-1}$ for all layers and grids. The outer 108 km RAMS3b grid in this simulation covers most of the continental US and has 28 vertical layers ranging from 69 m to about 16 km, and the 36 and 12 km grids have 34 layers, ranging from 17 m
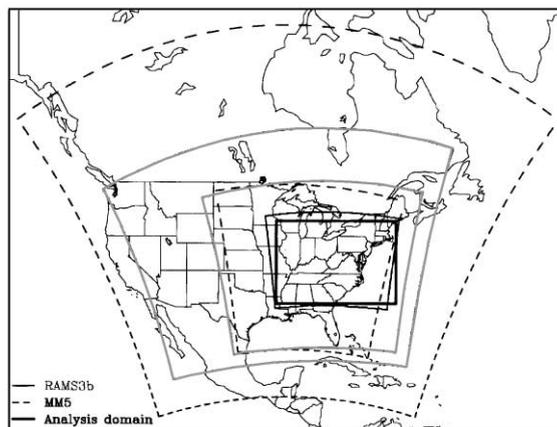

Fig. 1. Map depicting the three horizontal RAMS3b grids ($108 \times 36 \times 12$ km grid spacing), the two horizontal MM5 grids (108×36 km grid spacing), and the "analysis domain" used in this study.

to 16 km. In the RAMS3b simulation, the option of simulating clouds was not enabled because of computational constraints for performing the seasonal simulation. A more detailed description of the meteorological modeling used in this analysis can be found in Lagouvardos et al. (1997). Since the RAMS3b fields were used to drive the UAM-V photochemical model, they had to be mapped from the polar-stereographic coordinate systems used by RAMS3b on to the lat/lon coordinate system used by UAM-V. In particular, the RAMS3b fields were interpolated to a grid with horizontal grid dimensions of 36 km extending 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. These interpolated fields rather than the original fields are analyzed here since they were used to drive the photochemical simulations analyzed in the companion papers (Hogrefe et al., 2001; Biswas et al., 2001). Hereafter, we will refer to the meteorological input fields for the RAMS3b/UAM-V photochemical simulation as 'RAMS3b' with the understanding that the meteorological outputs from RAMS3b were mapped on to the UAM-V grids.

2.2. MM5

The second modeling system considered here is the Seasonal Model for Regional Air Quality (SMRAQ). The meteorological component of the SMRAQ system is the NCAR/PennState Fifth Generation Model (MM5) (Grell et al., 1994), and the photochemical model is the Multiscale Air Quality Simulation Platform (MAQSIP) (Odman and Ingram, 1996).

The modeling period for the SMRAQ study spanned from 15 May 1995–11 September 1995 (Kasibhatla and Chameides, 2000). This modeling period encompasses the same time period for which the RAMS3b/UAM-V simulation was carried out. The two horizontal grids used in the SMRAQ study had dimensions of 108 and 36 km. Fig. 1 depicts a map of the horizontal grids used in the two modeling systems. As with the RAMS3b simulation, 4DDA was used. The analysis fields were derived from a blend of ECMWF fields (6 h time intervals), surface observations (3–6 h time intervals), and rawinsonde observations (12 h time intervals). However, whereas nudging for RAMS3b was performed for wind speed and temperature for all model layers, for MM5 nudging was done for wind speed, temperature and water vapor mixing ratio above the boundary layer and for wind speed only within the boundary layer. The nudging factors used for wind speed and temperature in MM5 were $3.0 \times 10^{-4} \, \mathrm{s}^{-1}$ for the outer grid and $2.5 \times 10^{-4} \, \mathrm{s}^{-1}$ for the inner grid, and the nudging factor used for water vapor mixing ratio was $1.0 \times 10^{-5} \, \mathrm{s}^{-1}$. In other words, the nudging used for winds (and temperature above the boundary layer) in MM5 was stronger than the one used in RAMS3b, but temperature predictions within the boundary layer are purely prognostic for MM5. Since high-frequency fluctuations are smoothed out in the 4DDA fields used in both simulations, the stronger nudging in MM5 may suppress some high-frequency fluctuations. In contrast to the RAMS3b and UAM-V system, both MM5 and MAQSIP used the same grid systems so that no interpolation of meteorological fields into the photochemical model was needed. A major difference between the MM5 and RAMS3b simulations is the inclusion of a cloud parameterization in the MM5 simulation. For the outer grid, the Kuo cloud scheme (Kuo, 1974) was employed, and for the inner grid, the Kain–Fritsch cloud scheme (Kain and Fritsch, 1990, 1993) was employed. The simulated cloud cover in MM5 provides both nocturnal (long-wave) and daytime (short-wave) radiation modulation. Additional details about the SMRAQ project can be found in SMRAQ (1997a) and Kasibhatla and Chameides (2000). Hereafter, we will refer to the meteorological input fields for the SMRAQ study as 'MM5'.

2.3. Observations

For the evaluation of model predictions in Layer 1, hourly surface observations were retrieved from the Data Support Section at the National Center for Atmospheric Research (NCAR-DSS). Whereas the height of the first layer is 50 m for RAMS3b/UAM-V, it is 38 m for MM5/MAQSIP. As noted above, since 4DDA was performed for both the RAMS3b and MM5 simulations, the observed and predicted fields are not independent. 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 here; the model results were bilinearly interpolated to the observational sites. For the evaluation of upper air model predictions, we use a set of radiosonde and profiler measurements that were made as part of the NARSTO—Northeast research project during the summer of 1995. During the three high ozone events (19–20 June, 13–15 July, 31 July–2 August), additional radiosonde soundings were made to supplement the routine soundings at 00:00 GMT and 12:00 GMT at six sites in the northeastern United States. Continuous 1 h average wind measurements were obtained from 5 profilers for the 15 June–31 August 1995 time period. These profilers measured winds up to about 4 km; measurements are reported as layer-average for layer thicknesses of 57 and 105 m. The lowest layer ranges from 82 to 139 m. The profilers were located at Milestone Point, CT, Redhook, NY, Rutgers University, NJ, Gettysburg, PA, and Holbrook, PA. These additional soundings and profiler measurements provide

a good database for model evaluation since they were not used in 4DDA.

3. Methods of analysis

3.1. Traditional model evaluation statistics

In previous studies, several statistical measures have been used to evaluate meteorological models used for air pollution modeling (Willmott, 1982; Hanna, 1994; Olerud and Wheeler, 1997; McNally and Tesche, 1993). Some of these measures are listed in Table 1. 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). In addition, these statistics can be applied either to the entire data set (usually hourly averaged variables) or to extreme values such as the daily maximum or minimum to evaluate particular aspects of model performance.

The use of traditional statistical measures to evaluate model performance faces several shortcomings: the required statistical independence of the observed and predicted data sets is violated because of the use of 4DDA (Hanna, 1994; Olerud and Wheeler, 1997), and these statistics provide little insight into the physical behavior of the model. Therefore, additional model evaluation methods are needed to supplement these statistical methods, and some of such additional methods are introduced in the following section.

3.2. Definition of time scales and associated processes

It is well known that time series of meteorological variables contain fluctuations occurring on different time scales. However, the highest and lowest resolvable frequencies are determined by the sampling interval and the length of the data record, respectively. Since we analyze hourly time series of both observations and model predictions for a time period of three months, the periods that can be resolved range from 2 h to 30–40 days. The choice of the different frequency bands in our analysis is based on the power spectrum as well as a priori knowledge about different physical processes of interest to the simulation of atmospheric dynamics for air quality simulations. For many atmospheric variables, the single largest forcing in the hourly time series data is the diurnal (DU) forcing with a period of 24 h. Additional frequency bands of interest are the intra-day (ID) range (periods less than 12 h), the synoptic (SY) range (periods of 2–21 days), and longer-term (BL) fluctuations (baseline containing periods greater than 21 days). While an approximate choice of these periods was based on physical considerations (the intra-day component reflects turbulent and local-scale processes such as local convective cloud and precipitation events; the diurnal component will be dominated by the day-and-night differences with 24 h periodicity; the synoptic component will contain fluctuations related to 500 and 300 mb Rossby and short-wave evolution/propagation; and the baseline will contain low-frequency oscillations forced by subseasonal solar-seasonal to inter-annual-processes including, for example, ENSO), the actual choice of frequency bands was made to minimize the

Table 1
Definition of the traditional evaluation statistical measures

Mean bias error[a] (MBE)	$\overline{a_{mod}(x,t) - a_{obs}(x,t)}$
Fractional mean bias[b] (FB)	$\dfrac{\overline{a_{mod}(x,t)} - \overline{a_{obs}(x,t)}}{0.5(\overline{a_{mod}(x,t)} + \overline{a_{obs}(x,t)})}$
Mean absolute gross error[a] (MAGE)	$\overline{|a_{mod}(x,t) - a_{obs}(x,t)|}$
Root mean square error[a] (RMSE)	$[\overline{|a_{mod}(x,t) - a_{obs}(x,t)|^2}]^{0.5}$
Fractional mean square error[b] (NMSE)	$\dfrac{\overline{|a_{mod}(x,t) - a_{obs}(x,t)|^2}}{\overline{a_{mod}(x,t)} \times \overline{a_{obs}(x,t)}}$
Standard deviation of residuals[a] (SD$_R$)	$[\overline{|(a_{mod}(x,t) - a_{obs}(x,t)) - MBE|^2}]^{0.5}$
Correlation coefficient[b] R	$\dfrac{\overline{(a_{mod}(x,t) - \overline{a_{mod}}) \times (a_{obs}(x,t) - \overline{a_{obs}})}}{[(a_{mod}(x_i,t) - \overline{a_{mod}})^2]^{0.5} \times [(a_{obs}(x_i,t) - \overline{a_{obs}})^2]^{0.5}}$

[a] From McNally and Tesche (1993).
[b] From Hanna (1994).

covariance between the estimates of the different temporal components. It should be noted that not all meteorological variables will show significant variations on all of these time scales, but we still chose to use the same spectral decomposition for all variables for the sake of consistency and inter-comparison.

3.3. Spectral decomposition of 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. A detailed discussion of the KZ filter as well as comparisons to other separation techniques can be found in Eskridge et al. (1997), Rao et al. (1997), and Hogrefe et al. (2000). In contrast to time series of ozone concentrations, the time series of meteorological variables are not log-transformed prior to analysis.

4. Results

4.1. Traditional statistics

The results of the traditional evaluation statistical metrics discussed in Section 3.1 are presented in Tables 2a and b for the surface temperature and surface wind speed for both RAMS3b and MM5. These values reflect averages over space (all monitoring stations) and time (all hours). In most previous studies, only values for the hourly data were reported (e.g. Olerud and Wheeler,

1997). Additional information about the models' ability to reproduce the amplitude of the diurnal cycle can be obtained by applying these statistics to daily minimum and maximum values as well. The results for surface temperature, presented in Table 2a, show that—for hourly data—RAMS3b predictions have a positive bias of about 1°C while MM5 underpredicts temperature by about 1°C. All other statistics show similar values for the two modeling systems for the hourly values, with rather small absolute and mean square errors and high correlation coefficients. When the statistics for the daily minimum and maximum temperature are compared, it is evident that RAMS3b overpredicts both the mean of the hourly values as well as the mean of the daily maxima and minima by roughly the same amount, while MM5 underpredicts the mean of the daily maximum values much stronger than the mean of the hourly values and overpredicts the mean of the daily minimum values. This indicates that MM5's layer-1 diurnal temperature range was smaller and less-well "represented" than the RAMS3b modeling system. In general, however, this was expected because cloud cover—fully modeled in MM5 while entirely omitted from RAMS3b—provides both nocturnal (long-wave) and daytime (short-wave) radiation modulation, simulating clouds' first order control on surface temperature. In addition, it should be kept in mind, that when comparing the surface observations with model predictions in layer 1, strong gradients near the surface could distort the comparison. In other words, the diurnal variation of the temperature at 19 m or 25 m, representative for the model's first layer, might be considerably less than at the surface in the presence of strong gradients. Taking this into account,

Table 2

	Hourly data		Daily maximum		Daily minimum	
	RAMS3b	MM5	RAMS3b	MM5	RAMS3b	MM5
(a) *Traditional evaluation statistics for RAMS3b and MM5 predicted temperature.*						
Mean bias error (°C)	1.38	− 0.93	1.19	− 3.45	0.55	1.31
Fractional mean bias (%)	5.57	− 3.96	3.95	− 12.4	2.85	6.67
Mean absolute gross error (°C)	2.29	2.22	2.09	3.57	1.70	1.85
Root mean square error (°C)	3.03	2.89	2.68	4.03	2.22	2.42
Fractional mean square error (%)	1.50	1.51	0.8	2.1	1.34	1.53
Standard deviation of residual distribution SD_R (°C)	2.69	2.74	2.41	2.08	2.15	2.04
Correlation coefficient R	0.88	0.88	0.83	0.87	0.85	0.87
(b) *Traditional evaluation statistics for RAMS3b and MM5 predicted wind speed.*						
Mean bias error (m s^{-1})	0.61	0.28	− 0.68	− 1.01	1.31	1.02
Fractional mean bias (%)	18.8	8.5	− 12.6	− 19.4	99.0	87.0
Mean absolute gross error (m s^{-1})	1.41	1.34	1.46	1.41	1.45	1.20
Root mean square error (m s^{-1})	1.80	1.71	1.93	1.88	1.71	1.55
Fractional mean square error (%)	31.0	30.8	12.9	13.2	223	215
Standard deviation of residual distribution SD_R (m s^{-1})	1.70	1.69	1.81	1.59	1.10	1.16
Correlation coefficient R	0.58	0.58	0.48	0.57	0.48	0.56

the underestimation of the diurnal temperature amplitude by MM5 might not be as severe, and the RAMS3b amplitude might actually be an overestimate. As discussed in SMRAQ (1997b), a further reason for the negative bias of the MM5 surface temperatures could be caused by the Kain–Fritsch cloud parameterization used in MM5 that tends to create unrealistically large and long-lasting "cold pools" under weakly forced situations. In addition, problems with the treatment of radiation in cloudy conditions may also contribute to this bias.

The results for the surface water vapor mixing ratio (not shown) suggest good model performance for this variable. Table 2b presents the evaluation statistics for the surface wind speed. Both modeling systems predict the mean hourly observed wind speed to within about $0.5 \, \mathrm{m \, s^{-1}}$ (the rather high values for the fractional errors reflect the low values of the mean observed wind speed). A feature common to both modeling systems is the underprediction of the daily maximum wind speed and the overprediction of the daily minimum wind speed. This can be expected because we are comparing model predictions of Reynold's-average mean state variables with point measurements through these statistics, and sub-grid scale wind speed fluctuations can be substantial (e.g. Hanna, 1994; Hanna and Chang, 1992). Therefore, one would not expect a coarse-resolution discrete model to capture the full-amplitude range of observed surface winds. In addition, the use of 4DDA can also damp-out features with higher frequency temporal–spatial characteristics, because significant smoothing is applied to the analyzed fields prior to their use in 4DDA. This dependence of model performance on the amount of energy present within different time scales for a particular variable will be discussed in more detail in the following section.

4.2. Scale analysis

To examine the relative contribution of the different temporal components to the overall process energy, we calculated the variances of the seasonal time series of the components for both observations and model predictions for temperature, water vapor mixing ratio, and wind speed, and expressed them as fractions of the sum of the variances of all components. This allows us to compare the relative importance of different dynamic processes to the overall process for both observations and model predictions. The results, presented in Figs. 2a–c, represent spatial averages over all observation stations and the corresponding model grid cells. It can be seen that 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 (raw) time series. For the observed and RAMS3b

predicted temperature components (Fig. 2a), it is evident that temperature fluctuations caused by the day-and-night differences account for most of the variability in hourly temperature time series, followed by longer-term and synoptic-scale (weather-induced) fluctuations, while intra-day (local-scale) variations contribute little to the overall fluctuation. The RAMS3b model overestimates the relative contribution of the diurnal component to the overall variance and, as a consequence, understates the relative importance of the remaining components. However, the absolute amount of variance is captured closely by RAMS3b. For the MM5 predicted temperature time series, the total variance is underestimated by a factor of 2, and the importance of the diurnal fluctuation is strongly underestimated. This is consistent with the finding discussed in Section 4.1 which showed that MM5 underpredicts the diurnal temperature variation, and—as discussed in that section—is likely attributable to the inclusion of clouds in MM5 as well as the comparison of layer 1 model predictions with the surface observations.

For the observed and predicted mixing ratio time series from the two models, it can be seen that longer-term fluctuations are the largest contributors to the total variance, followed by synoptic-scale and diurnal fluctuations (Fig. 2b). The calculations for the wind speed time series (Fig. 2c) show significant differences between observations and model predictions. While in the observations, the intra-day component accounts for more than 20% of the variance, there is little contribution from the intra-day component to the overall variance in the predictions from either model. While Figs. 2a–c are useful to study the differences in the observed and predicted relative contributions of individual components to the overall variance, they do not allow us to compare the absolute magnitude of the fluctuations on different time scales between observations and model predictions directly. To this end, Table 3 displays the ratio of the variance of the predicted time series over the variance of the observed time series on different time scales for temperature, water vapor, and wind speed.

For the raw (unfiltered) temperature time series, the ratio of RAMS3b-predicted to observed variance is close to 1, while MM5 underpredicts the total variance by a factor of 2 (cf. Fig. 2a). For the intra-day temperature component, the variance is underestimated by both models, while the variance of the diurnal temperature component is overestimated by RAMS3b and underestimated by MM5. The variance of the synoptic and baseline components for temperature is underestimated by both models by approximately equal amounts, with the underestimation being stronger for the synoptic than the baseline component. As noted above, the underestimation of the diurnal temperature amplitude by MM5 is partially expected from a modeling simulation
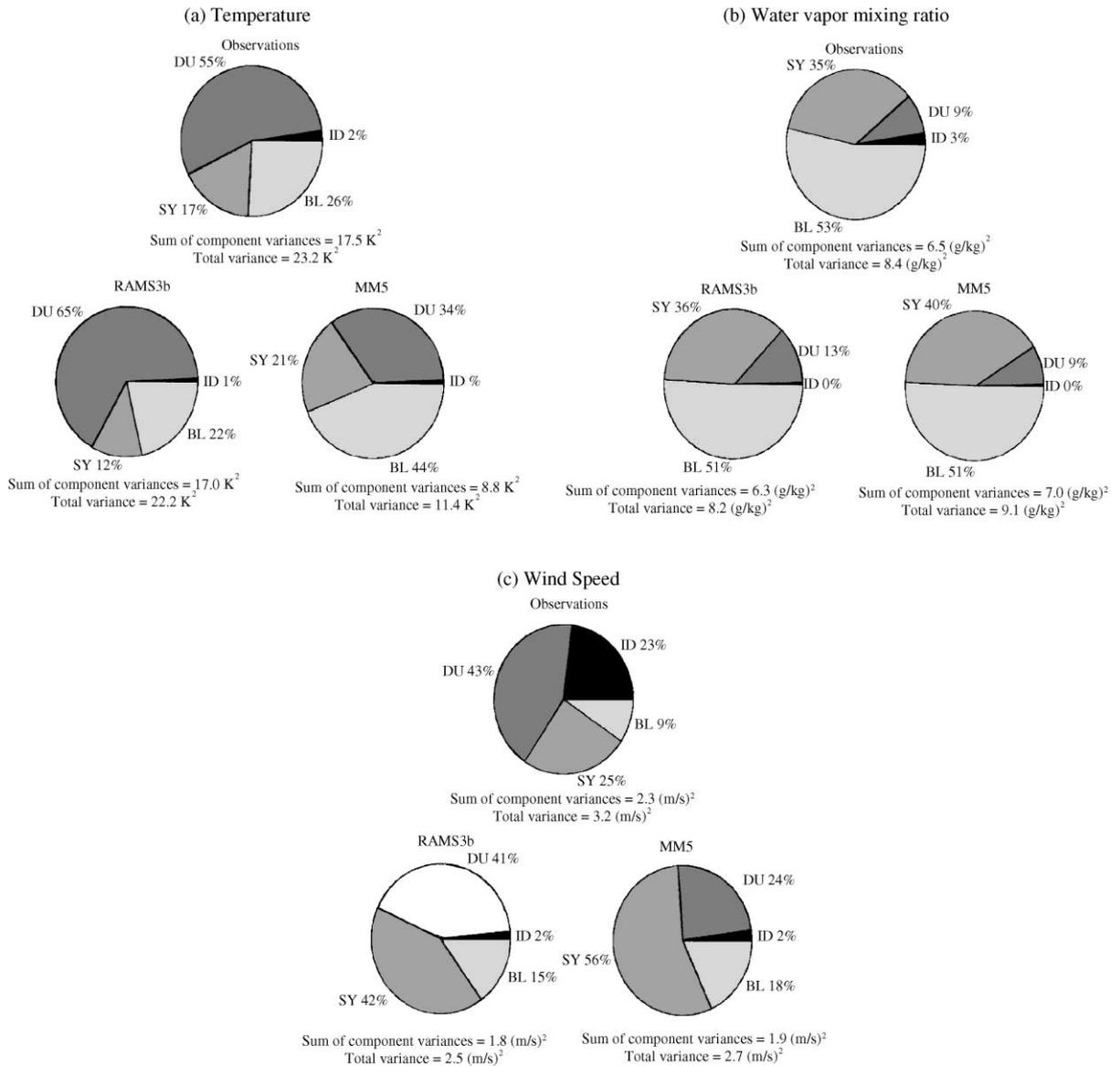
Fig. 2. Pie charts depicting the relative contribution of the variances of the component time series to the sum of the component variances for observations and model predictions. (a) Temperature. (b) Water vapor mixing ratio. (c) Wind speed.

Table 3
Ratio of variances of modeled to observed temporal components of time series for different variables

		Original	Intra-day	Diurnal	Synoptic	Baseline
Temperature	RAMS3b/obs	0.99	0.39	1.24	0.70	0.84
	MM5/obs	0.50	0.17	0.34	0.67	0.88
Water vapor	RAMS3b/obs	1.00	0.12	1.48	1.02	0.95
	MM5/obs	1.12	0.18	1.09	1.32	1.08
Wind speed	RAMS3b/obs	0.81	0.06	0.78	1.39	1.52
	MM5/obs	0.87	0.09	0.48	1.98	1.84

at a relatively coarse 36 km grid spacing and 38 m layer 1 height. The fact that RAMS3b does not show this underestimation might suggest that it provides a better agreement for the wrong reasons, namely, the lack of cloud effects. In short, the simulation of cloud processes is essential to create the right spectral response on the diurnal time scale for the right reason; however, it has to be kept in mind that there is considerable uncertainty in accurately simulating clouds with the current generation meteorological models.

For water vapor, both models approximately capture the variance of the original time series. As for temperature, the variance of the intra-day component is underestimated by both models. While the variance of the diurnal component is overestimated by RAMS3b and approximately captured by MM5, the situation is reversed for the variance of the synoptic component. The corresponding results for the wind speed components show that total variance is slightly underestimated by both models. Again, the models strongly underestimate the variability present on the intra-day time scale, whereas the amount of synoptic-scale and baseline variability for wind speed is over-estimated.

In summary, both the comparison of relative contributions of individual components to the total variance and the computation of ratios of predicted to observed variance for individual components provide useful insights into the ability of the models to correctly predict the amount of energy present on different time scales. Since fluctuations of these time scales are associated with different physical processes, this provides information about the ability of the model to capture these processes, and this information could not be as well-quantified without spectral decomposition of the time series.

The preceding analysis shows that model predictions could not properly capture the temporal variability present in the observed intra-day component for any variable. An additional way of evaluating model predictions against observations is to determine the average spatial correlation structure for each component. This is achieved by determining the correlation between the seasonal time series of a given temporal component at two different locations, determining the distance separating the two stations, repeating this procedure for all possible station pairs, and then determining the average correlation between two stations being separated by a certain distance. Figs. 3a–d show the results from this analysis (results are shown for wind speed only). While the spatial correlation for the observed intra-day component for wind speed drops to values below 0.3 in less than 15 km, the predictions from both modeling systems retain a strong spatial correlation on this time scale even beyond 100 km. In other words, the spatial variability of high-frequency, local-

scale wind fluctuations is not simulated well by the models (i.e., highly smoothed wind fields are reflected in the models). For the diurnal component, the spatial extent of the correlation structure is underestimated by MM5. For the other components, there is a good agreement between the observed and predicted average spatial correlation structures.

In addition to the above analysis, we also developed spatial maps of the correlation coefficient between observed and predicted time series of meteorological variables on different time scales. These maps permit us to identify the time scales on which the predicted time series are closest to the observed time series in different geographical areas. These results are presented in Fig. 4 for temperature and in Fig. 5 for wind speed. The panels in Fig. 4 illustrate that the time series of raw data, diurnal, synoptic and longer-term components are captured well by both models, with the highest correlations for the longer-term components. However, the time series of the intra-day component and amplitude of the diurnal cycle (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) show a weaker correlation.

For the wind speed (Fig. 5), correlations between observations and predictions are generally lower than that for the temperature. Most striking is the lack of correlation between the observed and predicted intra-day components, again illustrating the models' inability to capture fluctuations on this time scale. It is evident from Fig. 5 that correlations are highest for fluctuations having time scales longer than 1 day.

To illustrate the poor correlation of the amplitudes of the observed and predicted diurnal components for temperature, we present time series of observed and predicted daily maximum temperatures obtained from the raw data in Fig. 6a, time series of the amplitudes of the observed and predicted diurnal components in Fig. 6b, and time series of the normalized amplitudes of the observed and predicted diurnal components in Fig. 6c for Pittsburgh, PA. Fig. 6a shows that the observed and predicted daily maximum temperature time series follow each other closely. This is due to the influence of the synoptic-scale and longer-term forcings present in the time series, since these components are captured well by the models (cf. Fig. 4). On the other hand, Figs. 6b and c illustrate that the time series of the observed diurnal amplitude display a large variability that is not captured in the time series of the RAMS3b-predicted diurnal amplitude. In other words, the variability of the day-to-night temperature fluctuations due to the diurnal heating/cooling cycle (after removal of synoptic-scale and longer-term temperature variations caused by processes such as advection of air masses
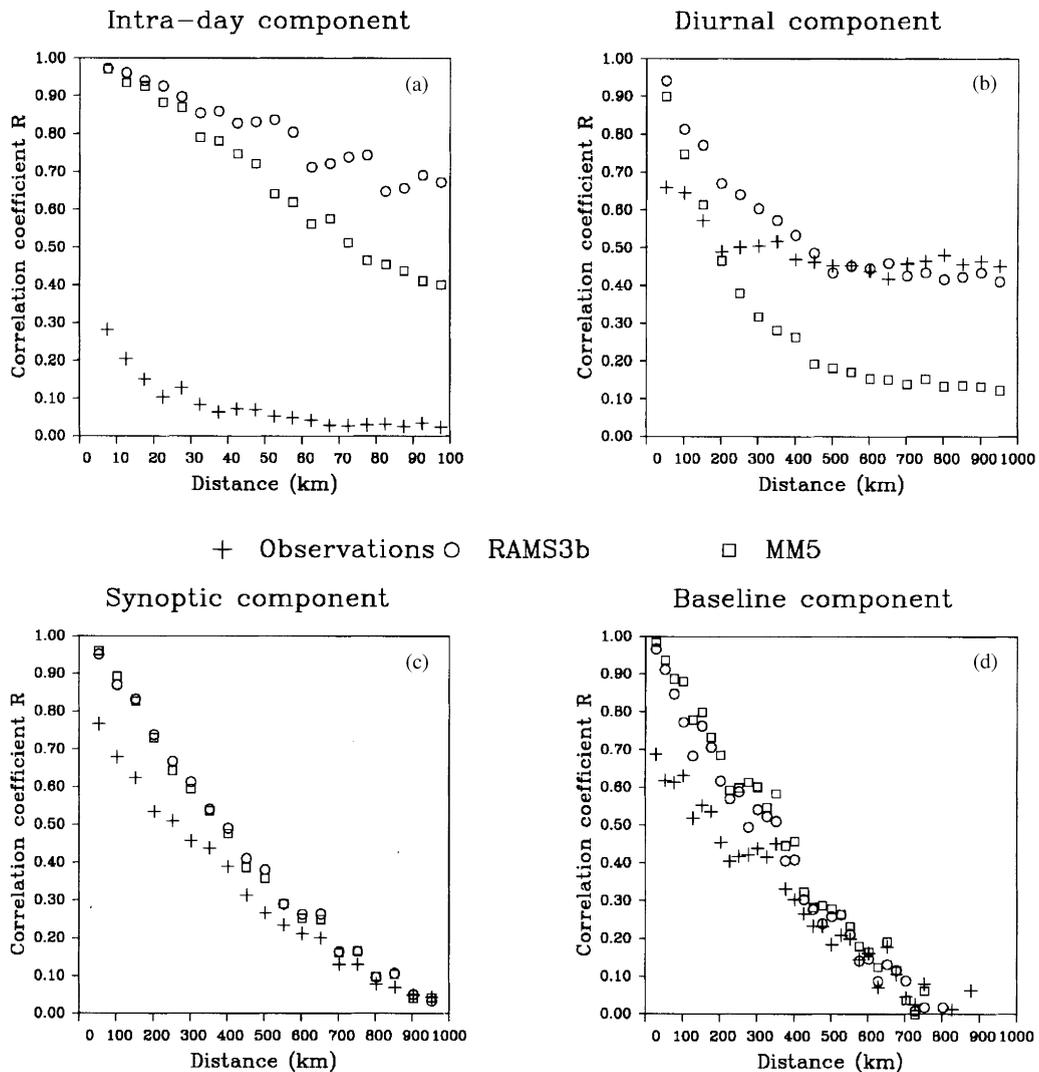
Fig. 3. Decay of correlations between wind speed time series at station pairs as function of distance between station pairs for observations and model predictions. (a) Intra-day component. (b) Diurnal component. (c) Synoptic component. (d) Baseline component.

and frontal passages) in the observations is not simulated well by the RAMS3b model. A likely reason for this lack of variability in the RAMS3b predictions is that the meteorological model was run without the simulation of cloud processes. Therefore, clear sky conditions are always present in RAMS3b, leading to much less day-to-day variability in the predicted diurnal heating/cooling cycle. To investigate this hypothesis further, we calculated the correlation between the time series of the amplitude of the observed diurnal temperature component and the observed daily mean opaque cloud cover at each observation station. The spatially averaged correlation coefficient is −0.74. This high negative correlation indicates that days with high

cloud cover tend to have lower day/night temperature differences, and vice versa. Thus, the absence of the day-to-day variability of the diurnal temperature amplitude in RAMS3b is mostly attributable to the absence of clouds in the RAMS3b simulation. The inclusion of both explicit, grid-scale clouds and implicit, parameterized deep convection in MM5 compared to RAMS3b is likely to increase the day-to-day variability of the amplitude of the diurnal temperature component. This is indeed the case, as can be seen in Figs. 6b and c. However, although MM5's day-to-day variability of the diurnal temperature amplitude is larger than for RAMS3b, the actual amplitude is smaller than in the observations, consistent with the discussion of MM5
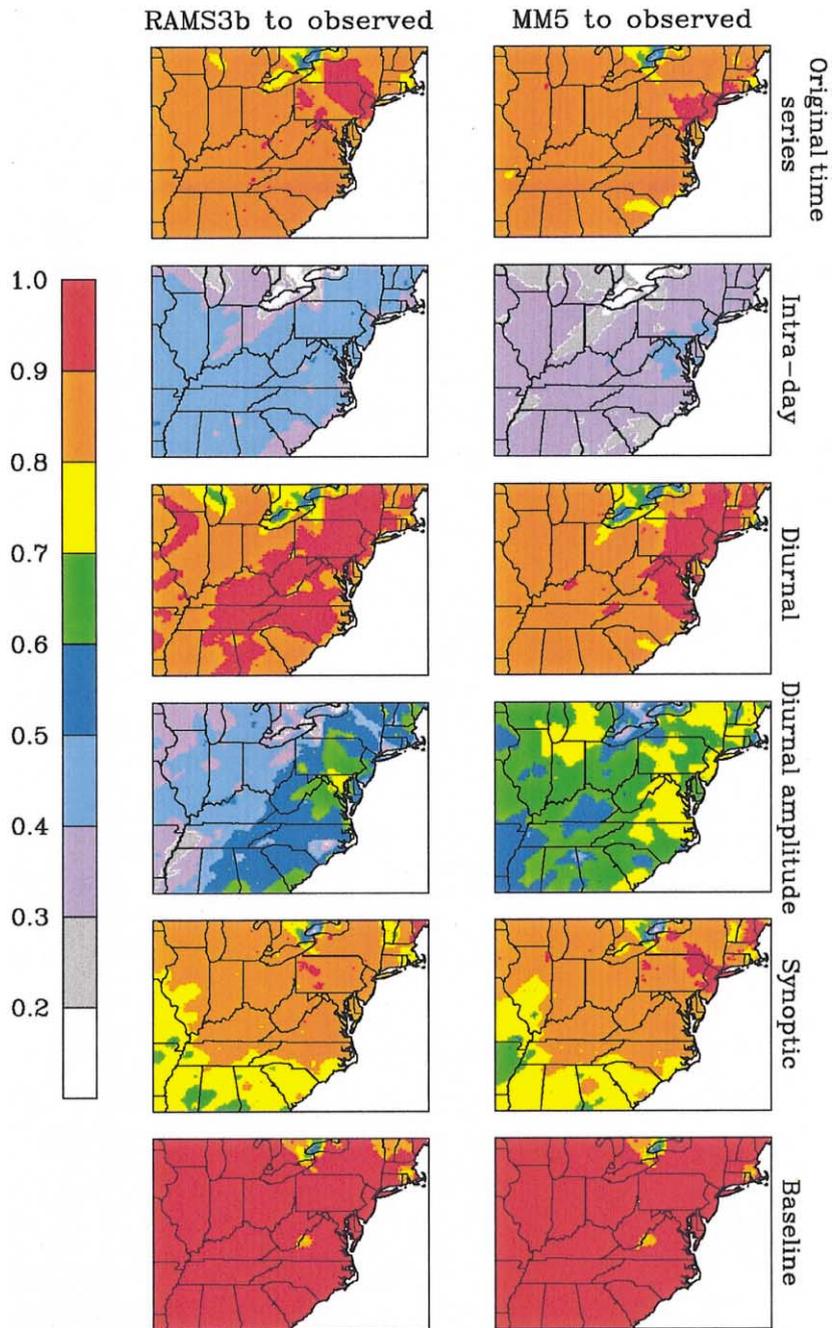
Fig. 4. Spatial images of correlation coefficients between observed and predicted time series of temperature for different spectral components. Correlations between RAMS3b predictions and observations are shown on the left, correlations between MM5 predictions and observations are shown on the right.

temperature predictions above. The correlations between the observed and MM5-predicted amplitude of the diurnal temperature cycle are improved compared to the RAMS3b simulation, but still lower than those for the synoptic and baseline components. Therefore, the inclusion of cloud processes has improved the simula-

tion of the diurnal time scale for temperature, but still the model performance is best on the synoptic and baseline time scales. This points to problems of consistently simulating the clouds at the right location at the right time, which is expected as clouds are highly scale-dependent: while synoptically driven major
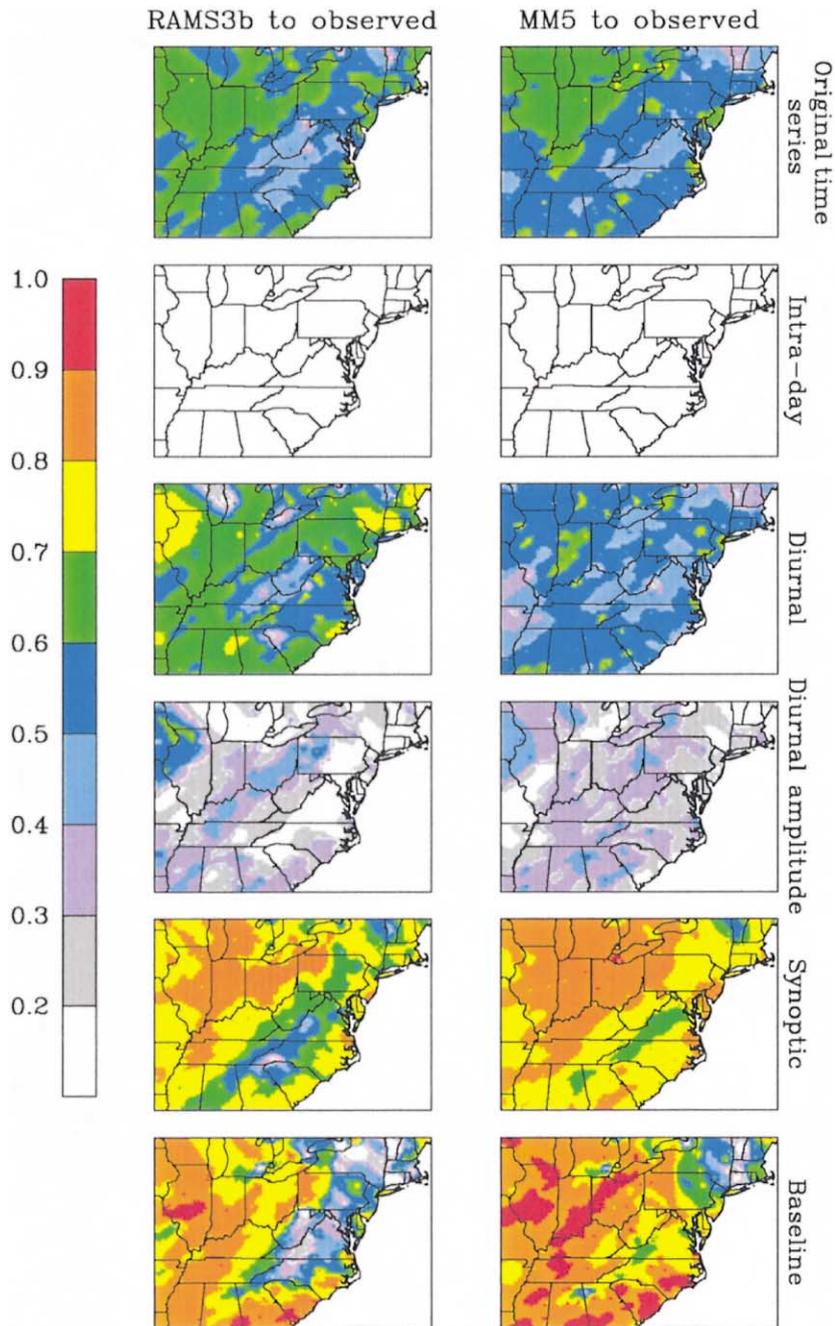
Fig. 5. Spatial images of correlation coefficients between observed and predicted time series of wind speed for different spectral components. Correlations between RAMS3b predictions and observations are shown on the left, correlations between MM5 predictions and observations are shown on the right.

precipitation systems and their clouds are likely to be well-simulated by a mesoscale model at 36 km grid spacing, diurnally-driven precipitation and clouds under weaker forcing are likely to be less well-represented, and locally-driven boundary layer clouds that act on the intra-day time scale are likely to be represented poorly.

4.3. Upper air data

While the previous sections focused on the evaluation of Layer 1 model predictions, we present a brief evaluation of model predictions for all model layers in this section. As mentioned before, additional radiosonde
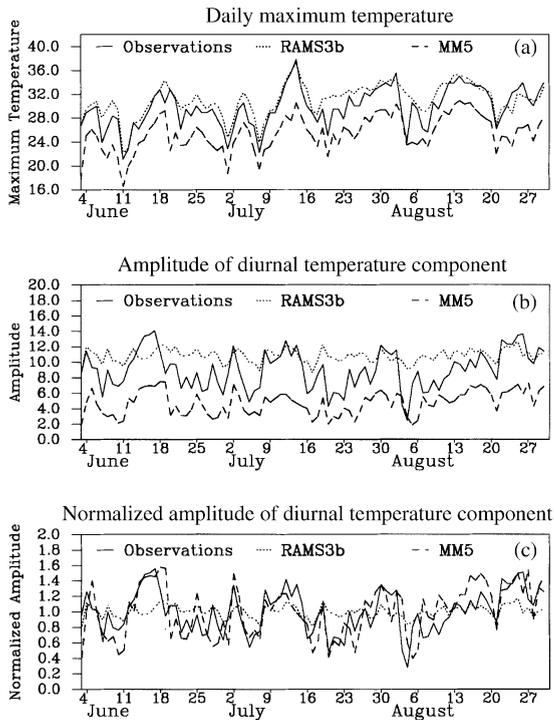
Fig. 6. (a) Time series of observed and predicted daily maximum temperature at Pittsburgh, PA. (b) Time series of the observed and predicted amplitude of the diurnal temperature component at Pittsburgh, PA. (c) Time series of the observed and predicted normalized amplitude of the diurnal temperature component at Pittsburgh, PA.

soundings were made in 1995 in addition to the routine soundings at 00:00 GMT and 12:00 GMT at a number of sites in the northeastern United States during the three ozone episodes, and profilers measured wind speed and direction for layers between 100 m and 4 km at five additional sites. These soundings and profiler measurements provide a good database for model evaluation since they were not used for 4DDA. For each model layer, the available observations within this layer were averaged corresponding to each sounding or profiler measurement, and then these pairs of observed and modeled values were compared. Because the additional soundings were performed only on a limited number of high ozone days, the spectral decomposition technique could not be applied to this set of observations. Therefore, we computed the mean and standard deviation of the differences of all pairs of observed and predicted data (summed over all non-routine soundings and all six sites) for each layer for temperature, water vapor mixing ratio and wind speed. The results indicate that neither modeling system exhibits a bias at any level for any variable that is significant at the 1-σ level with the exception of temperature, which is underpredicted

by both models above 3 km. The NARSTO-NE profiler measurements provide us with a much more dense temporal coverage and, therefore, allow us to perform spectral decomposition. Figs. 7a–c show profiles of the relative contribution of individual components to the total sum of the component variances for wind speed observations and predictions from the two models averaged over all five profiler sites. It can be seen that in the observations, the diurnal component is the largest forcing for the lowest layers only, while at higher levels, the synoptic component becomes the dominant forcing and the baseline component becomes the second largest forcing above 2 km (Fig. 7a). The relative contribution of the intra-day component decreases sharply between about 100 and 400 m, but remains almost constant at about 8% at higher levels. MM5 captures the shape of the observed profiles for the DU and BL components better than RAMS3b, but both models overestimate the importance of the synoptic component and underestimate the importance of the diurnal component above 500 m and the intra-day component for all layers (Figs. 7b and c). Figs. 8a–b show profiles of the correlation coefficients between observed and predicted time series of different components. The correlations profiles are very similar for both model simulations; the correlations are highest for the synoptic component below 1200 m and for the baseline component above 1200 m. On the other hand, correlations between the observed and predicted intra-day components are poor for all layers in both models. The better model performance on longer time scales as illustrated above for surface data holds true for upper air data also.

4.4. Grid resolution issue: 36 vs. 12 km grid spacing for RAMS3B

As mentioned in Section 2, the results from the RAMS3b output interpolated to the 36 km UAM-V grid were analyzed in the previous Sections 4.1–4.3 so that these results could be compared to the MM5 model simulations which were carried out with a grid spacing of 36 km. However, since our findings indicate that both models are unable to capture the energy and the temporal variations of the intra-day component for any variable, it is of interest to address the question whether these results would improve if the grid resolution is increased. For this purpose, we repeated the analysis after the RAMS3b 12 km predictions were interpolated to the UAM-V grid with a horizontal spacing of 12 km extending from 92°W to 69.5°W and 32°N to 44°N. These meteorological fields were used in the seasonal UAM-V simulation described in Hogrefe et al. (2000). Note that the 12 km UAM-V grid is identical to the analysis domain shown in Fig. 1. To illustrate the effects of the different grid spacings, we present Table 4 illustrating the ratios of modeled to
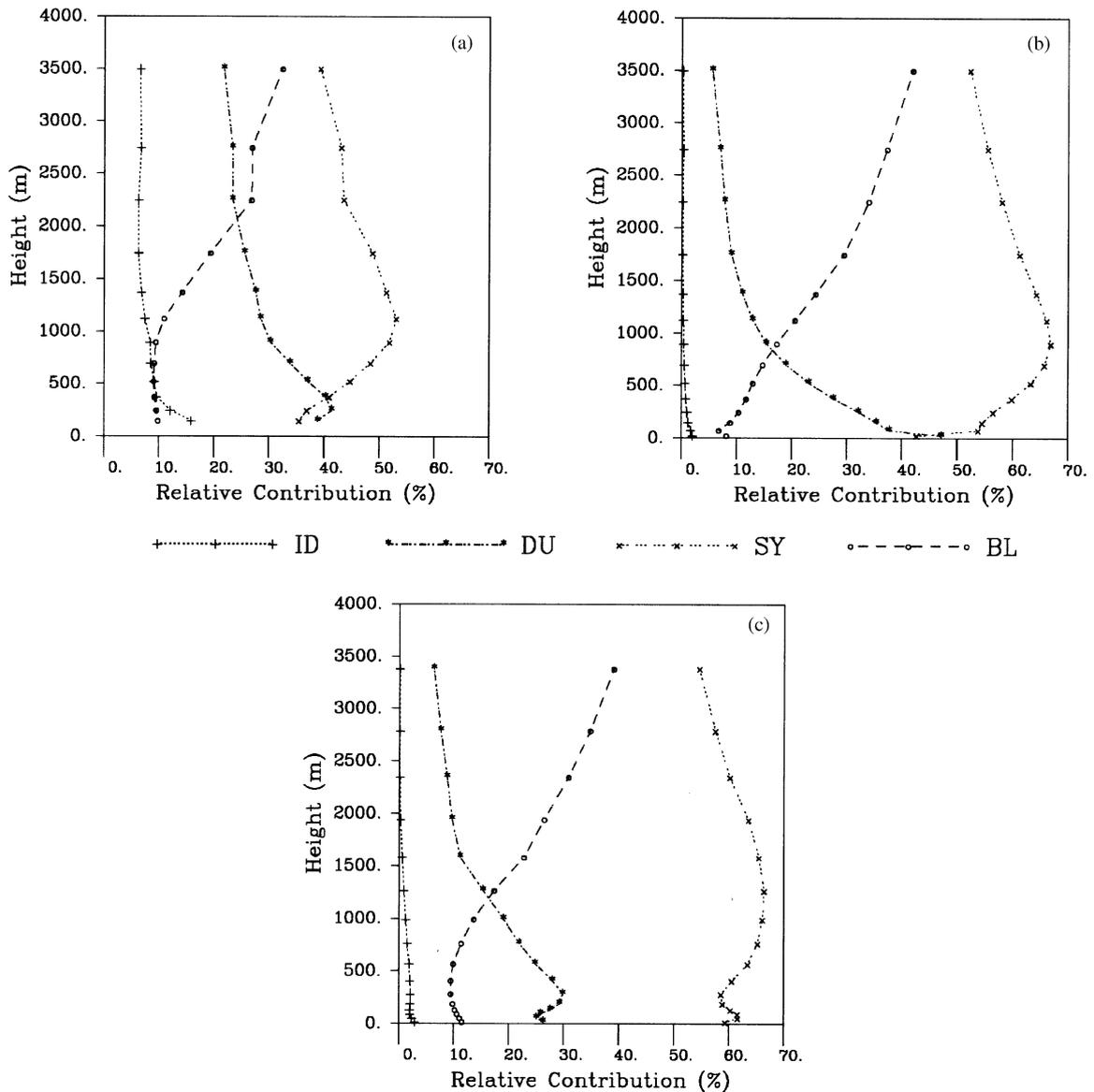
Fig. 7. Profiles of the relative contribution of the variances of the component time series to the sum of the component variances for wind speed observations and model predictions. (a) observations. (b) RAMS3b. (c) MM5.

observed variances and the spatially averaged correlation coefficients for wind speed for 36 and 12 km RAMS3b. From Table 4, it can be seen that the ratio of modeled to observed variance is increased for the 12 km grid relative to the 36 km grid for all components. The largest increase is on the diurnal scale and not on the intra-day time scale. The correlations between observed and predicted time series are nearly identical for the 12 and 36 km grids. In short, these results indicate that more—but not all—of the observed variability on shorter time scale could be simulated with the increased model resolution, but that the predictive

capability of the model as measured by correlation was not improved.

4.5. Implications to the use of meteorological models in air quality simulations

The above results illustrate the limitations and strengths of meteorological input data used to drive photochemical air-quality models. Out of the meteorological variables investigated in this study, temperature fields had the highest correlations, while wind speed predictions showed the weakest correlations with
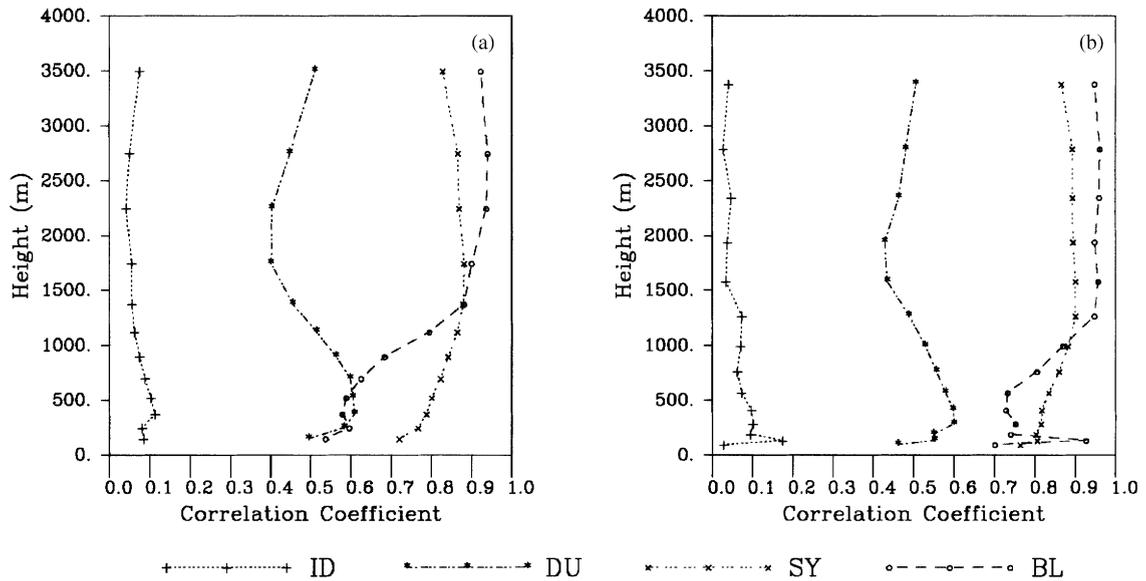
Fig. 8. Profiles of correlation coefficients between observed and predicted time series of wind speed on different spectral components. (a) Correlations between observations and RAMS3b. (b) Correlations between observations and MM5.

Table 4
Ratio of modeled to observed variances for different components and correlations between observed and predicted time series for different components for wind speed for both 12 km grid spacing and 36 km grid spacing RAMS3b simulations. All values are spatial averages

		Original	Intra-day	Diurnal	Synoptic	Baseline
Ratio of variances (mod/obs)	36 km	0.81	0.06	0.78	1.39	1.52
	12 km	0.89	0.07	0.92	1.43	1.54
Correlation (mod-obs)	36 km	0.59	0.09	0.62	0.73	0.68
	12 km	0.60	0.09	0.63	0.73	0.67

observations. The lack of agreement between meteorological model outputs and observations on the intra-day and diurnal time scales has important implications to air-quality predictions from photochemical models that use these meteorological input fields. First, since ozone fluctuations on the intra-day time scale are influenced, among other factors, by wind speed fluctuations on this time scale, ozone changes on the intra-day time scale cannot be simulated accurately. Second, while information about clouds derived from observations is used to calculate the photolysis rates in UAM-V, the absence of day-to-day variations in the heating/cooling cycle due to the absence of clouds in the RAMS3b simulation might impact the day-to-day variability in the evolution of the planetary boundary layer, which, in turn, will influence the day-to-day variability of the amplitude of the diurnal ozone component.

More generally, the features with longer temporal and larger spatial scales are captured well by the models.

While this is not surprising given that 4DDA was used in both simulations, it is also clear that the grid spacing used in this analysis as well as the formulation of model physics is better able to handle larger scale features such as the jet stream and geostrophic wind. In turn, the use of 4DDA can damp-out features with higher frequency temporal–spatial characteristics, because significant smoothing is applied to the analyzed fields prior to their use in 4DDA. In addition, due to the finite grid spacing, fluctuations with smaller spatial and temporal scales are not simulated properly, neither in terms of their energy nor their temporal variation. While a part of the poor performance on the shorter time scales might be attributable to an inadequate formulation of model dynamics and, therefore, theoretically could be improved, deterministic grid-based models cannot predict fluctuations at the sub-grid scale. This fundamental shortcoming should be viewed as the inherent uncertainty associated with the grid-based models. The

amount of the inherent uncertainty theoretically decreases as grid resolution increases, assuming that a perfect formulation of the model dynamics on all scales is possible. However, as discussed above, the better model performance on the longer time scales appears to be partly attributable to the use of 4DDA; thus, an increased grid resolution alone might reduce the magnitude of the inherent uncertainty but not necessarily improve overall model performance unless appropriate 4DDA is also carried out at this increased resolution.

When these meteorological models are used to drive air quality simulations, air quality predictions cannot be expected to be accurate for scales which are not captured by all the input fields used in the photochemical model. In other words, even if the performance of the meteorological model on the intra-day time scale were improved potentially by decreasing horizontal grid dimensions to 4, 1 km or smaller (assuming a perfect formulation of model physics), the resulting air quality predictions might not become more accurate because other model inputs (e.g. emissions, land-use patterns, soil moisture, etc.) are not accurately resolved to the corresponding spatial scales. On the other hand, the time scales of best performance for both state-of-science meteorological models (i.e., synoptic and longer-term) were also shown to be of great importance in photochemical modeling for policy-making (Hogrefe et al., 2000). This implies that, for regulatory purposes, there may be little gain from increased horizontal grid resolution of the photochemical modeling systems, especially if an increased spatial resolution comes at the expense of extended simulation periods covering a larger domain.

5. Summary

In this study, model predictions of the layer 1 temperature, water vapor mixing ratio, and wind speed used in seasonal photochemical simulations were compared with all available surface observations. For this comparison, both traditional evaluation statistics as well as a spectral decomposition technique were used. Traditional evaluation statistics indicate that—out of the investigated variables—model performance is best for temperature and worst for wind speed. With the application of the spectral decomposition technique for model evaluation, it becomes evident that model performance is time scale specific and, therefore, the outcome of model evaluation on different time scales can be tied to the model formulation of the relevant processes on these time scales. The results show that predictions from both state-of-science meteorological models do not capture either the amount of energy or the temporal evolution of fluctuations on the intra-day

time scale. Correlations between the observed and simulated time series of meteorological variables are highest on the longer time scales. The better model performance on the synoptic and longer-term time scales strengthens the findings of Hogrefe et al. (2000) that photochemical modeling systems (i.e., both meteorological and chemical models) are best suited for characterizing average patterns over regional domains, rather than episodic (1–2 days) modeling for specific receptor locations.

References

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 (submitted for publication).

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.

ENVIRON, 1997. User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx). Available from 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.

Grell, G.A., Dudhia, J., Stauffer, D., 1994. A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note, NCAR/TN-398 + STR.

Hanna, S.R., 1994. Mesoscale meteorological model evaluation techniques with emphasis on needs of air quality models. In:

Pielke, R.A., Pearce R.P. (Eds.), Mesoscale Modeling of the Atmosphere. Meteorological Monographs, Vol. 25. American Meteorological Society, Boston, MA 02108, pp. 47–58.

Hanna, S.T., Chang, J.C., 1992. Representativeness of wind measurements on a mesoscale grid with station separations of 312 m to 10 km. Boundary Layer Meteorology 60, 309–324.

Hogrefe, C., Rao, S.T., 2000. Evaluating meteorological input variables for seasonal photochemical modeling on different time scales. Proceedings of the 11th Joint Conference of the American Meteorological Society and the Air & Waste Management Association on the Applications of Air Pollution Meteorology, Long Beach, CA, pp. 132–137.

Hogrefe, C., Rao, S.T., Kasibhatla, P., Hao, W., Sistla, G., Mathur, R., McHenry, J., 2001. Evaluating the performance of regional-scale photochemical modeling systems: Part II—ozone predictions. Atmospheric Environment 35, 4175–4188.

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.

Kain, J.S., Fritsch, J.M., 1990. A one-dimensional entraining/detraining plume model and its application in convective parameterization. Journal of the Atmospheric Sciences 47, 2784–2802.

Kain, J.S., Fritsch, J.M., 1993. convective parameterization for mesoscale models: the Kain–Fritsch scheme. In: Emanuel K., Raymond D.J. (Eds.), The Representation of Cumulus Convection in Numerical Models, Meteorological Monographs, Vol. 24. American Meteorological Society, Boston, MA 02108, pp.165–170.

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

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

Kuo, H.L., 1974. Further studies of the parameterization of the influence of cumulus convection on large-scale flow. Journal of the Atmospheric Sciences 31, 1232–1240.

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. (Office of Science and Technology, NYSDEC, 50 Wolf Rd., Albany, NY 12233-3259).

Lyons, W.A., Tremback, C., Pielke, R.A., 1995. Applications of the Regional Atmospheric Modeling System (RAMS) to provide input to photochemical grid models for the Lake Michigan Ozone Study (LMOS). Journal of Applied Meteorology 34, 1762–1786.

McNally, D., Tesche, T.W., 1993. MAPS Sample Products. Alpine Geophysics, 16225 W. 74th Dr., Golden, CO 80403.

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 Environmental Programs, Research Triangle Park, NC 27709, 83 pp.

Olerud, D., Wheeler, N., 1997. Investigation of the Regional Atmospheric Modeling System (RAMS) for the Ozone Transport Assessment Group (OTAG). Report to the Southeast Modeling Center, MCNC Environmental Programs, Research Triangle Park, NC 27709.

Pielke, R.A., Uliasz, M., 1998. Use of meteorological models as input to regional and mesoscale air quality models-limitations and strengths. Atmospheric Environment 32, 1455–1466.

Rao, S.T., Ku, J.-Y., Berman, S., Zhang, K., Mao, H., 2000. 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., 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.

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

SMRAQ, 1997b. Meteorological conclusions for SMRAQ project. Available online at http://envpro.ncsc.org/SMRAQ/meteorology/reports/met_smraq_concl2.htm.

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).

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), 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 8-hour 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.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bulletin of the American Meteorological Society 63, 1309–1313.

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