Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Atmospheric Environment 40 (2006) 5073–5086
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

New methods for evaluating meteorological models used in air quality applications

Robert C. Gilliam[a,*], Christian Hogrefe[b], S.T. Rao[a]

[a]*National Oceanic and Atmospheric Administration, Air Resources Laboratory, Atmospheric Sciences Modeling Division, In partnership with the United States Environmental Protection Agency, Mail Drop E243-03, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, USA*
[b]*Atmospheric Sciences Research Center, DEAS-ES351, University at Albany, Albany, NY 12222, USA*

Received 25 January 2005; received in revised form 8 July 2005; accepted 3 January 2006

Abstract

Meteorological models in conjunction with air quality models are being used to simulate the transport and fate of pollutants in the atmosphere. Hence, there is a need for an extensive evaluation of the entire modeling system. In this study, several new techniques to assess the performance of mesoscale meteorological models are introduced with an emphasis on evaluating the variables and processes that have the potential to influence the air quality predictions, since errors in the meteorological fields are passed on to the air quality model.

Model performance was diagnosed by examining the inter-correlation of observable variables in the atmosphere on distinct time scales: intraday, diurnal, and synoptic. It was found that the Mesoscale Model version 5 (MM5) model did replicate the observed relationship between intraday wind speed and temperature, intraday surface pressure and temperature, diurnal surface pressure and temperature as well as most of the correlations between variables on the synoptic timescale. However, a negative correlation between temperature and precipitation was evident in the observations on the intraday scale, but such relationship was not evident in the model output. Furthermore, the diurnal response of increasing wind speed with temperature was strong in the observed time series, but it was much weaker in the model. The correlation between diurnal changes in temperature and cloud fraction was consistently negative in the model whereas it was slightly positive in the observations.

Wind profilers were used to examine the simulated boundary layer wind structure. Of the twelve sites examined, the average distance error between the 24-h observed and modeled trajectory was approximately 150 km at height of 100 m above the surface. Errors in transport of this magnitude (100–200 km) can produce errors in air quality predictions. It is not the intent of this study to establish quantitative links between the performance of the specific meteorological simulation analyzed here and subsequent air quality simulations. Rather, the results presented here draw attention to errors and inconsistencies in the meteorology that are passed on to the air quality model which, in turn, have the potential to cause errors in air quality model predictions.

Keywords: Model evaluation; MM5; K–Z filter; Trajectories; Correlation analysis

*Corresponding author.
E-mail address: robert.gilliam@noaa.gov (R.C. Gilliam).

1352-2310/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2006.01.023

1. Introduction

Air quality modeling has expanded in sophistication and application during the past few decades. Meteorological and air quality modeling tools, such as the mesoscale model version 5 (MM5) model (Grell et al., 1994), the National Centers for Environmental Prediction (NCEP) Eta model (Janjic, 1994), and the community multiscale air quality (CMAQ) model (Byun and Ching, 1999), are being developed for air quality research applications (Bullock and Brehme, 2002), air quality forecasting (Otte et al., 2005; Vaughan et al., 2004), assessing climate impacts on air quality (Hogrefe et al., 2004), and for developing emission control strategies (Sistla et al., 2001a). Results from air quality simulations, which can have extensive ramifications (Chang and Hanna, 2004), are closely linked to the meteorological model that drives the chemical transport, diffusion, and chemical reactions in the air quality model (Pielke and Uliasz, 1998; Sistla et al., 2001b; Seaman, 2000). Thus, meteorological models need to be evaluated thoroughly along with the air quality models (Hogrefe et al., 2001a, b; Pielke, 1998).

The topic of meteorological model evaluation has been the focus of many air quality-related studies. A large majority of meteorological model evaluations center around how the model performs with regard to predicting surface-based measurements of temperature, wind speed, moisture, and precipitation. Typical evaluations take the point measurements and match them with the volume-averaged model results in space and time. Statistics such as mean bias, root-mean-squared-error, mean absolute error, and index of agreement are then calculated and used as metrics to judge model performance (Gego et al., 2005; Vaughan et al., 2004; Abraczinskas et al., 2004; Mass et al., 2003; Tesche, 2002; Emery, 2001; Saulo et al., 2001). While this is a straightforward method to compare observations with model results, and similar statistics are presented in this study, there are many other aspects of meteorological simulations that need to be closely examined from the standpoint of subsequent performance of air quality models.

In this study, several new approaches are introduced to examine the performance of the meteorological model with a focus on variables and processes that have the potential to affect subsequent air quality simulations that require these meteorological fields. The application of these techniques is illustrated by evaluating aspects of an annual simulation performed with the MM5, and brief discussions on how these evaluation results may affect subsequent air quality modeling results (e.g., CMAQ) are provided. First, traditional model performance statistics are examined for distinct subsets of data, i.e., grouping the data by season, topography, region, land-use, and synoptic flow patterns. Secondly, and the main focus of this study, the inter-correlation among observed variables (standard Pearson correlation between each variable and all other variables) in the atmosphere are compared with those of the simulated variables. This approach may prove to be useful in diagnosing how the meteorological model compares with the real atmosphere as a system, where many measurable quantities have direct and indirect relationships with one another. This method is taken a step further by spectrally decomposing the meteorological variable time series to provide further insight into how well the model "as a system" is performing in reproducing the features imbedded in various time scales. Lastly, profiler measurements are compared with simulated vertical wind profiles on a diurnal basis over a period of several months. It is expected that this examination can elucidate not only how well the planetary boundary layer (PBL) wind is representated by the model, but also the representation of various boundary-layer processes. The trajectories are caculated using observed and simulated wind at multiple levels in the PBL, and can provide a quantitative estimate of how errors in the PBL wind can affect the transport of pollutants in air quality models. The use of all of these methods is illustrated with annual MM5 simulations at 36 and 12 km horizontal grid cell dimensions.

2. Method of approach

2.1. Meteorological simulations

The primary meteorological model simulation used in this study was an annual MM5 version 3.6.1 model run that covered the Continental United States, with a horizontal grid spacing of 36 km. The study utilizes a similar 12 km MM5 version 3.6.3 simulation that was executed for a full year, except that this model domain covered only the eastern part of the United States. The two model simulations were executed using the same physics options: the Pleim–Xiu land surface model (Pleim and Xiu,

1995), the Pleim–Chang PBL model (Pleim and Chang, 1992), the rapid radiative transfer model (RRTM) long-wave radiation (Mlawer et al., 1997), the Kain–Fritsch (Kain, 2004) cumulus parameterization, and the Reisner 2 microphysics (Reisner et al., 1998) scheme. One exception is that the Pleim–Xiu soil moisture and temperature initialization was modified slightly so as not to recycle the soil temperature (reinitialized using Eta analysis) between 5.5-day run segments; the moisture was the only recycled soil property. This was done in an attempt to lessen a previously recorded cold bias of the 36 km simulation. The convective scheme was updated to a newer version (Kain–Fritsch, version 2) for the 12 km simulation. Additionally, both simulations used the four-dimensional data assimilation (FDDA) analysis nudging option available in MM5, which forces the model to remain in close agreement with analysis fields. It should be noted that the main reason for developing these annual simulations is for regulatory applications conducted by the US EPA. Therefore, the model setup has been carefully constructed to optimize overall performance through many sensitivity tests and represents a standard meteorological configuration for the US EPA's air quality applications. A more detailed description of the MM5 configuration is given by McNally (2003).

2.2. Observations

Several observational data sets were used for comparison with the model simulations. The TDL observational data set, provided by the Data Support Section at the National Center for Atmospheric Research (NCAR-DSS), was used for the hourly surface meteorology. All surface observations have undergone a quality control process to remove anomalous records. The wind profiler observations used in the trajectory analysis were obtained from the NOAA Forecast Systems Lab's Meteorological Assimilation Data Ingestion System (MADIS). These profiler observations were matched (space and time) with the MM5 output and are used for the trajectory analysis.

2.3. Data extraction, sub-setting and conversion

Model performance statistics were calculated using a recently developed model evaluation toolkit that utilizes database technology to store observation and corresponding model values, as well as other characteristics: station location, elevation, land-use, observation network, and date/time. This metadata can be used as criteria to extract specific data subsets from the model evaluation database. In Section 3.1, this database method was used to calculate model performance statistics by season, geography, region, MM5 land-use category, and synoptic pattern. Geographical separation of the statistics was done according to the elevation of the site: coastal (elevation <25 m), inland ($25\,\text{m} <$ elevation < 350 m), and mountains (elevation > 350 m). The synoptic patterns were determined by a map typing procedure (McKendry et al., 1995). Cluster 1 represents patterns where a Canadian high pressure was anchored over the northeast US Cluster 4 represents a synoptic pattern where a cold front passes off the east coast of the US, and high-pressure builds into the central US Cluster 7 corresponds to weak large-scale flow over the entire eastern US, which typically occurs during the summer. For these data subsets, many statistical measures were calculated, but only the mean bias, mean absolute error, and index of agreement are discussed in detail. See Fox (1981) and Willmott (1982) for details on the mean bias, mean absolute error, and index of agreement, which is also known as anomaly correlation (Wilks, 1995).

In Section 3.2, observation time series were extracted from the database. The time series included are temperature, dewpoint temperature, wind speed, u–v wind components, surface pressure, precipitation, and cloud fraction. All variables were extracted directly from the observation record with the exception of cloud fraction. Cloud fraction observations are recorded at multiple levels from laser beam ceilometers. In order to derive a single value of cloud fraction for the entire column, the maximum cloud fraction at all levels was used. The MM5 model does not output all of these specific variables, so a few had to be derived. Temperature and wind speed as well as wind components were directly extracted from the model output (May-September 2001, 36 km run). The dewpoint temperature was calculated using standard formulas that are dependent on dry bulb temperature and relative humidity. For precipitation, the rainwater mixing-ratio at the lowest model level was used. For surface pressure, the pressure perturbation at the lowest model grid level was used since this is the time varying pressure variable in the raw model output. Since cloud fraction was not a raw model output variable, the total amount of cloud water in

the atmospheric column was used. Even though some of these simulated variables (cloud water and precipitation rate) do not exactly match the observed variables (cloud fraction and hourly precipitation), in terms of units; the variables have a direct correlation. That is to say, changes in the rainfall rate correlate directly with changes in total hourly rainfall, and that changes in cloud fraction are directly correlated with changes in cloud water at a model grid point.

All observation and model time series were spectrally decomposed using the Kolmogorov–Zurbenko (K–Z) filter described by Rao et al. (1997) and Hogrefe et al. (2000). The filter was applied to decompose the time series into intraday (variations with period of 12 h or less), diurnal (1–2 days), and synoptic (2–21 days) components. A standard Pearson correlation (Wilks, 1995) between each observed variable and all other observed variables was calculated for the warm season (May–September). This was repeated for each filtered set of variables with exception of the seasonal component since only the warm season was considered. In addition, the same procedure was performed for the time series of simulated variables. To avoid confusion in the coming discussions, Pearson correlations will be often referred to as "correlation between variables", "intercorrelations among variables", or "relationship between variables".

3. Results

3.1. Statistical measures

Domain-wide statistics provide a general performance measure on how well the model simulation replicates the observed meteorology in the eastern United States. Table 1 is a compilation of several commonly used model performance statistics for 2 m temperature, for both the annual 36 and 12 km MM5 simulations. Although the 36 km model domain covered the Continental United States, only observation-model pairs that were within the 12 km domain were used for the 36 km domain statistics. The first categorization is seasonal. For all seasons, the 36 km simulation is biased cold (−0.24 to −1.36 K), while the 12 km simulation is biased cold during the winter (−0.91 K), but biased warm (0.41–0.69 K) in the spring and summer. The cold bias is, at least in part, a result of the deep soil temperature reaction to changes in air temperature on the synoptic time scale. For example, during an extended cold period (several weeks), the deep soil temperature cools in the model. When the synoptic pattern then shifts to a warm period of several days to a week, the deep soil temperature is too slow to respond (warm). This increases the temperature gradient between soil levels, especially during daytime heating period, and hence, the associated heat flux from the upper to the lower layer of the two-layer soil model increases. Because of the surface energy-budget constraints, the simulated heat flux from the surface soil layer to lower atmosphere is lessened, resulting in overall lower simulated 2 m temperatures.

The mean absolute error decreases in both the 36 and 12 km simulations from the winter (2.36 and 2.38 K) to spring (1.81 and 2.02 K), and from the

Table 1
Model performance statistics calculated for 2 m temperature for various subsets of data

2-m temperature						
Data subsets	36-km MM5			12-km MM5		
	BIAS	MAE	IOA	BIAS	MAE	IOA
Seasonal						
Winter	−1.36	2.36	0.94	−0.91	2.38	0.95
Spring	−0.37	1.81	0.93	0.69	2.02	0.93
Summer	−0.24	1.67	0.93	0.41	1.99	0.92
Fall	−0.35	1.76	0.97	−0.06	2.28	0.93
Geographical[a]						
Coastal	−0.35	1.74	0.93	0.33	1.88	0.92
Inland	−0.29	1.66	0.94	0.56	1.95	0.93
Mountain	−0.26	1.93	0.93	0.43	2.23	0.93
Regional[a]						
Northeast	−0.65	2.24	0.92	0.67	2.43	0.93
Midwest	−1.06	2.22	0.90	0.52	1.88	0.91
Deep South	−0.41	1.65	0.90	0.41	1.91	0.89
Southeast	−0.49	1.78	0.92	0.40	2.17	0.91
Land-use[a]						
Agriculture	−0.20	1.67	0.94	0.51	1.94	0.93
Plains	−0.82	2.03	0.92	−0.32	2.40	0.93
Forested	−0.25	1.71	0.93	0.73	2.03	0.92
Desert	−0.29	1.67	0.93	0.66	1.93	0.92
Urban	−0.21	1.70	0.91	0.22	1.61	0.94
Water	−0.29	1.83	0.93	0.28	1.98	0.67
Synoptic cluster[a]						
Cluster-1	−0.46	1.82	0.95	0.36	2.03	0.95
Cluster-4	−0.35	1.85	0.98	0.23	2.28	0.98
Cluster-7	−0.55	1.96	0.97	−0.17	2.13	0.97

The annual 36 and 12-km MM5 simulation for 2001 was used for the calculations.

[a]Model performance metrics were calculated for dates covering the May 2001–September 2001 period.

spring to summer (1.67 and 1.99 K). The 12 km simulated temperature has a larger mean absolute error (MAE) than the corresponding 36 km simulation each season, indicating that there is a lack of improvement in the near surface temperature field from the finer scale simulation. It should be stressed again that the two simulations were not only different in grid cell size, but the soil temperature re-initialization was done by the Eta model every 5.5 days, and an updated convective scheme was used. The index of agreement (IOA), which is a measure of how well the model represents the pattern of perturbations about a mean value (Wilks, 1995), indicates little difference in the 2 m temperature patterns of the 36 and 12 km simulations. Gego et al. (2005) examined the improvement in the simulated 2 m temperature by decreasing the grid size from 12 to 4 km. They found an overall improvement for the afternoon and early evening hours, but a slight degradation in the early to late morning. Mass et al. (2002) also found only slight improvements as grid spacing was decreased from 36 to 12 km, but little improvement from 12 to 4 km.

The next group of statistics (geographical, regional, land-use, and synoptic clusters) in Table 1 was calculated from data subsets over the US for the warm season (from May to September). Mean bias (MB) calculations indicate the 36 and 12 km simulations are biased about the same magnitude, but opposite in sign for most of the subsets. The 36 km simulation has a general MB between −0.25 and −0.50 K, while the 12 km simulation has a MB between 0.25 and 0.50 K. It is apparent that the temperature bias in the 36 km simulation is not a result of geographical considerations, or synoptic patterns, but does reflect a regional/land-use influence. This is identified by the fact that all MB are similar (−0.25 to −0.50 K) with the exception of the MB in the Midwest (−1.06 K). Additionally, the temperature bias as a function of land-use is virtually the same (approx. −0.25 K) except for locations classified as plains/rangeland (−0.82 K), which is the predominant land-use pattern in the Midwest and central United States. This connection implies that the MM5 had difficulties in simulating the 2 m temperature because of a mischaracterization of the "plains" land-use, since the agriculture land-use, which is also extensive across the Midwest and central United States, has a more typical MB of −0.20 K.

To bring the preceding temperature statistics into perspective, a set of benchmarks were proposed by Emery (2001), and MB, MAE and IOA values fall within the guidelines (MAE < 2 K; −0.5 K < MB < 0.5 K; IOA > 0.80). Additionally, the temperature statistics compare well with Abraczinskas et al. (2004) who presented similar statistics for an annual 2002 MM5 simulation. Baker (2004) found daily variability in the statistics over an entire year ranged from 1.50 K (−2.00 K) to 3.00 K (0.50 K) for the MAE (MB). Mass et al. (2003) presented similar statistics for a forecast version of the MM5 over the northwest US, which is a difficult region to simulate because of the topography. The MAE was consistently near 2.25 K, and the MB ranged from 1.00 K at night to −1.00 K during the day.

Errors in the 2 m temperature are of interest because they can effect the concentration of certain types of pollutants estimated by air quality models. Gas-phase chemistry, biogenic and mobile source emissions, gas/particle partitioning of semi-volatile organic compounds, and ammonium nitrate are affected by temperature. In fact, observed aerosol nitrate increases significantly at night during the winter, when relative humidity rises above 80% (Nenes et al., 1998). The cold bias results in an overprediction of the relative humidity if the mixing ratio is well simulated. This relative humidity overprediction, could contribute to the overprediction of aerosol nitrate in the winter.

Table 2 presents the same statistic subsets as Table 1 except for the 10 m wind speed. The overall performance of the 2001 MM5 in terms of simulated seasonal 10 m wind speed is almost identical to the 36 km 2002 annual simulation evaluated by Abraczinskas et al. (2004). The 36 km simulation has slightly weaker 10 m winds than the observations indicate: winter ($-0.15\,m\,s^{-1}$), spring ($-0.10\,m\,s^{-1}$), and summer ($-0.19\,m\,s^{-1}$). The mean absolute errors (approx. $1.3\,m\,s^{-1}$) are well below the limit of $2\,m\,s^{-1}$ proposed by Emery (2001), as are the biases (within $\pm 0.5\,m\,s^{-1}$). The IOA is slightly lower than the benchmark of 0.60 (Emery, 2001), especially in the summertime. This can be attributed to the difficulty in simulating light and variable winds near the surface when the synoptic forcing is weak during the warm season. Similar to the 2 m temperature, the wind speed statistics calculated from the 12 km simulation indicate little improvement over the 36 km simulation. The wind speed bias for both 36 and 12 km simulation are similar to the annual 2002 MM5 simulation evaluated by Abraczinskas et al. (2004).

It is difficult to quantify how errors in the wind speed would affect air quality simulations, but

Table 2
Model performance statistics calculated for 10 m wind speed for various subsets of data

10-m wind speed						
Data subsets	36-km MM5			12-km MM5		
	BIAS	MAE	IOA	BIAS	MAE	IOA
Seasonal						
Winter	−0.15	1.34	0.55	−0.08	1.29	0.59
Spring	−0.10	1.34	0.58	0.05	1.26	0.61
Summer	−0.19	1.22	0.47	−0.10	1.15	0.49
Fall	−0.24	1.31	0.56	−0.00	1.27	0.59
Geographical[a]						
Coastal	−0.12	1.38	0.48	0.06	1.31	0.51
Inland	−0.04	1.17	0.55	0.06	1.11	0.57
Mountain	−0.40	1.42	0.55	−0.25	1.34	0.58
Regional[a]						
Northeast	0.06	1.31	0.43	−0.12	1.21	0.46
Midwest	−0.01	1.53	0.70	0.15	1.41	0.72
Deep South	−0.45	1.34	0.53	−0.39	1.20	0.57
Southeast	−0.04	1.34	0.51	−0.17	1.24	0.58
Land-use[a]						
Agriculture	−0.06	1.23	0.61	0.06	1.16	0.63
Plains	−0.76	1.54	0.51	−0.67	1.45	0.54
Forested	−0.17	1.18	0.45	0.00	1.10	0.48
Desert	−0.17	1.14	0.57	−0.02	1.08	0.61
Urban	−0.34	1.18	0.50	−0.41	1.16	0.52
Water	0.49	1.55	0.46	−0.66	1.62	0.48
Synoptic cluster[a]						
Cluster-1	−0.18	1.30	0.55	0.07	1.22	0.58
Cluster-4	−0.06	1.23	0.53	0.06	1.20	0.55
Cluster-7	−0.10	1.27	0.52	−0.02	1.23	0.55

The annual 36 and 12-km MM5 simulation for 2001 was used for the calculations.

[a]Model performance metrics were calculated for dates covering the May 2001–September 2001 period.

indirect effects can be inferred. For example, near-surface wind passed on to the air quality model is used in combination with micrometeorological variables to calculate eddy diffusion that dictates mixing of pollutants. Furthermore, an error in the model's wind speed and direction can cause material to be advected to locations that might be much different from reality. For example, a wind speed bias of only $0.50\,m\,s^{-1}$ can cause transport of pollutants to be off by 43 km over the course of the day, which is greater than the grid spacing of most air quality models. This subject is examined in more detail in Section 3.3 where observed and simulated vertical wind profilers are compared.

3.2. Inter-relationships among simulated and observed variables

Numerical weather models are constructed to simulate many observable quantities in the atmosphere. These variables, which describe the behavior of the atmosphere as a system, are related either directly or indirectly to many physical processes. Thus, it stands to reason that if a model is simulating atmospheric processes accurately, the inter-correlation among the simulated variables will be identical to that found in the observations. In this section, we compare the correlation among a set of observed variables spectrally decomposed into intraday, diurnal, and synoptic time scales with that of the same simulated variables. As an illustration, Table 3 presents the observed (left side of Table 3) and simulated (right side of Table 3) correlation matrices for each of these time scales at a single site in central North Carolina (Raleigh-Durham international airport, referred to in text as KRDU). Because of the large number of inter-correlations, only the more important relationships (i.e., those that relate to air-quality modeling) will be discussed.

Intraday variations in temperature and wind are induced by small-scale processes, so it is not surprising that studies such as Hogrefe et al. (2001a) found that these signals are more difficult for MM5 to simulate. The correlation matrix for the intraday variation (Table 3, Panel IN) indicates that the positive correlation between wind speed and temperature is similar in both the model (0.19) and observations (0.25). This implies that the response of near-surface wind to temperature, or vice versa, on a temporal scale of a few hours is well modeled. An example of an underlying physical process that drives this correlation is the momentum transport from aloft to the surface as the temperature warms. Alternatively, if the surface layer experiences shear induced mixing due to a mesoscale feature at night, the wind increases, as does the temperature. In Fig. 1(A), the modeled and the observed correlations for 131 sites in the eastern US are compared. Overall, the model has a stronger correlation (bias of 0.17 over all stations) between these variables than that in the observations and the standard deviation of the difference between all modeled and observed correlations is 0.10.

The intraday correlation between surface pressure and temperature is also consistent between observations (−0.14) and the model (−0.16) at the KRDU site (Table 3). This negative relationship is intuitive

Table 3
Observation (left column) and 36 km MM5 (right column) inter-correlation (Pearson correlation) among temperature (tmp), dewpoint temperature (dew), wind speed (ws), u–v wind components (u/v), surface pressure (sp), hourly precipitation (pcp1), and cloud fraction (cf) for intraday (IN), diurnal (DI), synoptic (SY)

IN	tmp	dew	ws	u	v	sp	pcp1	cf		tmp	dew	ws	u	v	sp	pcp1	cf
tmp	1.00								tmp	1.00							
dew	0.06	1.00							dew	−0.11	1.00						
ws	0.19	−0.03	1.00						ws	0.25	−0.36	1.00					
u	0.03	0.03	0.03	1.00					u	0.03	−0.12	0.12	1.00				
v	0.14	0.05	0.25	0.11	1.00				v	0.06	0.09	0.03	0.16	1.00			
sp	−0.14	0.01	−0.02	0.09	−0.08	1.00			sp	−0.16	−0.01	−0.05	−0.06	−0.06	1.00		
pcp1	−0.22	−0.02	−0.07	−0.08	−0.06	0.10	1.00		pcp1	0.00	0.00	0.01	0.01	0.01	−0.01	1.00	
cf	−0.02	0.01	0.04	0.02	0.01	0.01	0.00	1.00	cf	−0.02	−0.04	0.00	0.07	0.01	0.01	0.13	1.00

DI	tmp	dew	ws	u	v	sp	pcp1	cf		tmp	dew	ws	u	v	sp	pcp1	cf
tmp	1.00								tmp	1.00							
dew	0.04	1.00							dew	−0.16	1.00						
ws	0.62	−0.09	1.00						ws	0.30	−0.17	1.00					
u	0.07	0.19	0.10	1.00					u	−0.21	−0.06	−0.10	1.00				
v	0.10	0.20	0.08	0.33	1.00				v	0.09	0.21	0.06	−0.05	1.00			
sp	−0.27	−0.13	0.00	0.01	−0.17	1.00			sp	−0.38	−0.24	−0.08	0.13	−0.29	1.00		
pcp1	−0.05	0.06	−0.02	−0.01	−0.03	−0.06	1.00		pcp1	−0.02	0.02	−0.02	−0.05	−0.03	−0.04	1.00	
cf	0.20	−0.05	0.10	−0.03	−0.04	0.04	−0.06	1.00	cf	−0.10	−0.03	−0.08	0.03	−0.15	0.03	0.42	1.00

SY	tmp	dew	ws	u	v	sp	pcp1	cf		tmp	dew	ws	u	v	sp	pcp1	cf
tmp	1.00								tmp	1.00							
dew	0.76	1.00							dew	0.67	1.00						
ws	0.11	0.16	1.00						ws	−0.25	−0.23	1.00					
u	0.42	0.41	0.22	1.00					u	0.29	0.22	−0.06	1.00				
v	0.49	0.47	0.30	0.72	1.00				v	0.42	0.51	−0.25	0.28	1.00			
sp	−0.21	−0.32	−0.16	−0.36	−0.16	1.00			sp	−0.26	−0.39	−0.02	−0.41	−0.13	1.00		
pcp1	−0.04	0.25	0.19	0.06	0.04	−0.20	1.00		pcp1	−0.23	0.13	−0.07	−0.07	0.02	−0.19	1.00	
cf	−0.12	0.28	0.14	−0.01	0.04	−0.15	0.30	1.00	cf	−0.29	0.18	−0.13	−0.10	0.01	−0.17	0.83	1.00

These correlations were calculated for the warm season period from May 2001 to September 2001.

and consistent with physical constraints; warming air near the surface rises and results in lower surface pressure (or vice-versa). This relationship is similar across all sites in the eastern US (not shown here); the standard deviation of all correlation differences is 0.10.

The intraday relationship between observed temperature and precipitation is inconsistent between observations (−0.22) and model (0.00) at the KRDU site (Table 3). As expected, the observations reveal that rapid decreases in temperature are incident to rainfall events. However, the model shows no correlation between these two variables. The non-filtered time series were inspected during a few periods of precipitation (not shown), and the observations show a distinct decrease in temperature when precipitation was measured, while the model does not. In Fig. 1(B) is the correlation comparison that includes the 131 sites, all of which show a similar difference between the observed and simulated precipitation-temperature correlation.

The inter-correlations on the diurnal time scale between the meteorological variables are presented in Table 3, (Panel DI). The relationship between temperature and wind speed on the diurnal timescale is well understood; along with a daytime increase in temperature is an increase in near-surface wind that results from boundary-layer growth and turbulent mixing. Similarly, the decrease in wind speed at night is a result of cooling and stabilization of the boundary layer. Observations show a relatively strong correlation (0.62) between temperature and wind speed, while it is much weaker in the model (0.30). An examination of the diurnal filtered time series indicates that the observed temperature and wind speed signals are in

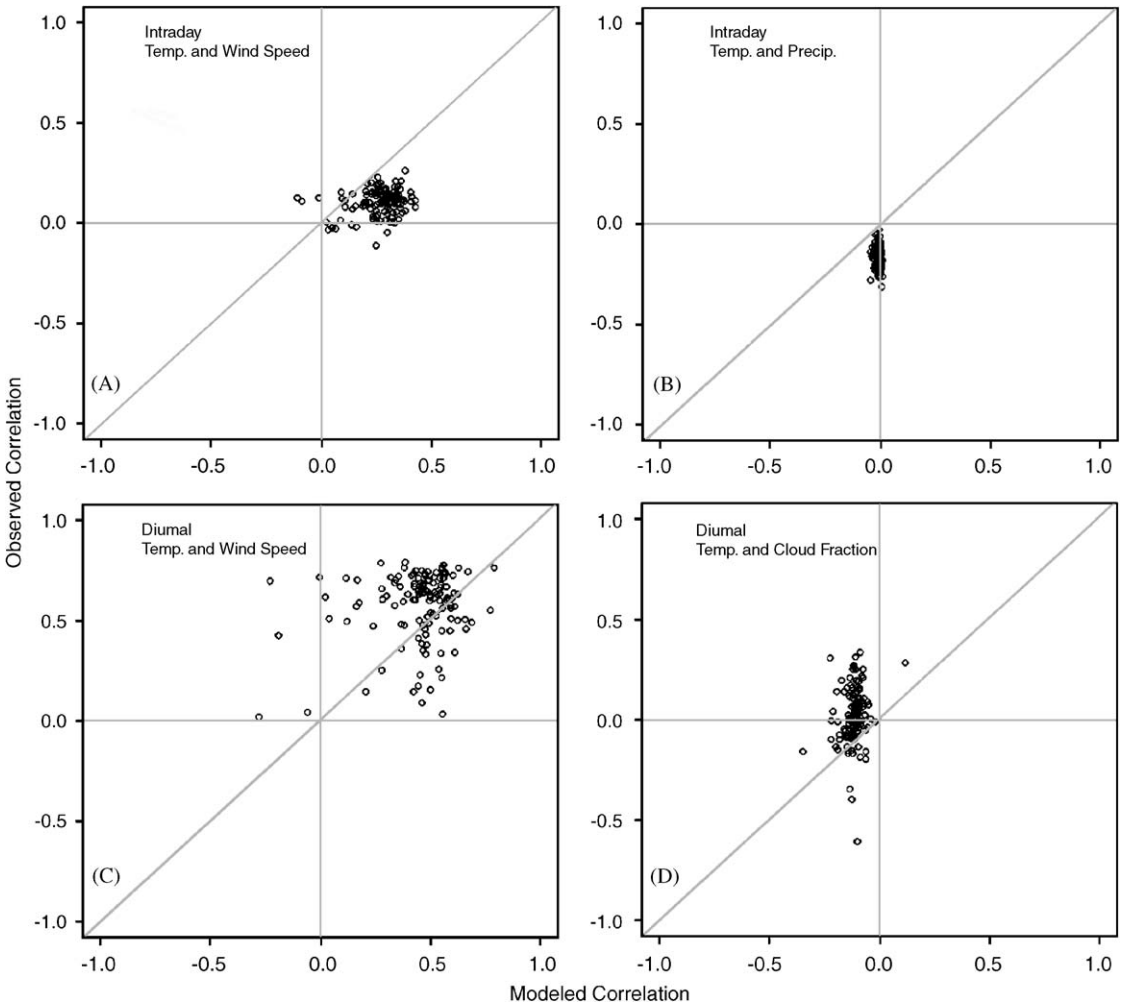

Fig. 1. Scatter plots of observed and simulated correlations between: (A) intraday 2 m temperature and 10 m wind speed, (B) intraday 2 m temperature and precipitation, (C) diurnal 2 m temperature and 10 m wind speed, and (D) diurnal cloud cover and temperature.

phase while the modeled signals are slightly out of phase by a few hours in many instances, thereby affecting the correlations among the simulated variables. Additionally, the correlation between simulated wind speed and observed wind speed is only around 0.33 for the diurnal component. These inconsistencies are attributable to errors in the model representation of the boundary-layer mixing processes, and the resulting PBL growth and collapse. A study by Dennis et al. (2004) found persistent errors in the air quality simulation during PBL transition periods, which may be in part attributed to this inconsistency. In addition, a comprehensive examination of the diurnal 2 m temperature and 10 m wind speed simulated by five MM5 PBL parameterizations was conducted by Zhang and Zheng (2004). While the Pleim–Xiu PBL scheme was not evaluated, this study found that the temperature and wind were out of phase in many of the other MM5 PBL schemes, and hypothesized that the rapid downward momentum-transport is not correctly modeled. In Fig. 1(C), indicates that the majority of sites (approx. 75%) have a similar error (much stronger correlation in observations than in the model).

Another correlation that should be consistent between the model results and observations on a diurnal time scale is temperature and cloud fraction. It is typical in the southeast US for cloud cover to increase as daytime temperatures rise. The observations show a positive correlation (0.20), while the model has a slight negative correlation (−0.10). In fact, Fig. 1(D) indicates that the model has nearly the same correlation for all sites in the eastern US,

while the observations substantial variability. This difference may be related to the 36 km simulations inability to resolve the smaller-scale fair-weather cumulus cloud 9 structures well and when the model simulates clouds, they are sizable enough to affect the radiation to the extent that the near-surface temperature cools.

On the other hand, the model results (−0.38) are similar to the observations (−0.27) in replicating the temperature and surface pressure correlation on the diurnal timescale; a negative relationship exists, as expected since there is a pronounced diurnal surface pressure change associated with daytime heating. A scatterplot (not shown) that compares observed and simulated correlation of temperature and pressure for the 131 sites in the eastern US similarly indicate very good agreement for a majority of sites (approx. 75%). However, observations from the 25% of remaining sites indicate little correlation between temperature and pressure (between −0.10 and 0.10), while the model generally has a stronger negative correlation (between −0.50 and −0.10).

It has been shown in another study, which examined model performance on different time scales (Hogrefe et al., 2005), that the annual MM5 simulation can reproduce the synoptic-scale variations of variables like the near-surface temperature and wind quite well. The model examined in this study used analysis nudging by the NCEP Eta data analysis system (EDAS), which further ensures that the simulation accurately propagates weather systems into and out of the model domain. It is expected that observed and model inter-correlations among variables will be similar, especially those most influenced by the synoptic-scale forcing. The inter-correlations for both the observations and model are presented in Table 3. Unlike previous correlation matrices, dewpoint temperature and dry bulb temperature are highly correlated in the observations and the simulation results, 0.76 and 0.67, respectively. This is a fundamental response to synoptic-scale air mass changes; cooler air masses are inherently dry and warmer air masses are moist.

Another process well represented by the model (mod) when compared to observations (obs) is the positive correlation between the wind components and temperature (0.49 obs, 0.43 mod). A positive v-wind component (southerly wind) is associated with warm advection while a negative v-wind component is associated with cold advection. The correlation between pressure and temperature is also similar between the observations and model (approx. −0.25), a case where decreasing pressure is associated with southerly wind, and warmer temperatures. Similar is the correlation between pressure and the v-wind component (−0.16 obs, −0.13 mod). A southerly wind is climatologically associated with falling pressure while a northerly wind is associated with high pressure rising. Precipitation and pressure correlations on the synoptic timescale are consistent; (−0.20 observation versus −0.19 model); low (high) pressure systems are mostly associated with precipitation (no precipitation).

Notable correlations not consistent between the observations and model results on the synoptic time scale are wind speed and temperature (0.11 observation versus −0.25 model), cloud fraction and precipitation (0.30 observation versus 0.83 model), and the wind speed and v-wind component (0.30 observation versus −0.25 model). The reasons for these differences are not apparent; thus further analysis and sensitivity studies will be required, which is out of the scope of this study.

3.3. Evaluation of the simulated planetary boundary layer wind

One aspect of meteorological simulations that is especially important to consider in air-quality applications is the representation of the wind throughout the depth of the PBL. These winds are the primary transport mechanisms for the pollutants released in the PBL. In this section, wind profile observations are compared to the model simulation at twelve sites in the eastern US; however, only a subset of these sites is specifically discussed. The average diurnal wind, calculated from a 2-month dataset (13 July 2004–15 September 2004) at several heights, is used to examine potential errors in simulated air pollution transport. The method generates conceptual trajectory using the mean observed and simulated wind starting at 00 UTC and ending at 23 UTC at each model level. There are a few assumptions in generating the trajectories: (1) the flow is horizontally homogenous, but changes hourly according to the diurnal mean wind calculated over the specified period (13 July 2004–15 September 2004), (2) turbulent processes and day-to-day variability are not considered, as the trajectory is calculated from the mean wind, and (3) the trajectory represents the flow at a constant height above the surface. These assumptions are rather restrictive, thus, the term

"conceptual trajectories"; however, the analysis can provide an easy to apply method to estimate potential errors in air pollution transport from simulated wind errors over limited areas. Additionally, the conceptual trajectory plots at the very least, can aid in assessing the ability of the model to simulate certain diurnal-dependent mesoscale process in the PBL.

Fig. 2 displays the observed and simulated trajectories at 100 (red), 350 (blue) and 1000 (green) meters for a number of profiler sites. Fig. 2(A) is for a site near the coast in southern Maine. The observed trajectory indicates an initial southerly wind at 00 UTC that veers more westerly in the morning (12 UTC) through early afternoon, at which point the wind sharply becomes south once again through the afternoon into the evening. This cycle clearly represents the New England summertime land-sea breeze evolution. By comparing the general shape of the simulated and observed trajectory, much insight is gained in how well the model represents the diurnal wind variation above the surface. The overall change from southerly to westerly, then back to a southerly wind direction is well represented by the model, as is the timing of the shift from a westerly to southerly wind that occurs at 1600 UTC. The main error that will affect the transport at 100 m near this site is negative wind speed bias in the model, which can lead to pollutant transport uncertainty on the order of 100 km. The observed and modeled trajectories at 350 and 1000 m are in closer agreement indicating an improved wind simulation above the surface layer.

Figs. 2(C), (F), (G), (J) display the trajectories for a network of profiler sites around Cape Canaveral, FL. The spacing of these sites is 10–30 km or, from a model perspective, 1–3 grid points. All of the observed trajectories at these sites have a daily sea breeze signature, albeit quite different from site to site. Previous studies have examined sea breeze characteristics in this particular region in detail (Rao et al., 1999; Atkins and Wakimoto, 1994; Zhong and Takle, 1993). Many of the local variations can be seen in the observed trajectories. The simulated trajectories at False Cape (FCPFL in Fig. 2(C)) are very different from what was observed, likely because this site lies on a very complex maze of lagoons and small islands that are not resolved well by the 12-km scale model simulation. However, the model does compare well with the abrupt shifts in wind direction at the other sites. For example, at Merritt Island (MIDFL, Fig. 2(F)) and south cape (SCPFL, Fig. 2(J)) the sharp easterly wind shift in the early afternoon (~1600 UTC) exists in both the model and observation trajectory at 100 m. At Mosquito Lagoon (MLNFL, Fig. 2(G)) the model simulates the wind shift from southerly to northerly as the trajectories shows a looping pattern. This indicates that the model is simulating the northerly wind component of the sea breeze, which is a characteristic around coastal capes (Gilliam et al. 2004). The site south of Cape Canaveral (SCPFL, Fig. 2(J)) has a southerly sea breeze component; hence, it seems that relatively fine details are captured by the 12 km scale model. Nonetheless, the model has persistently weaker winds than indicated by the wind profilers. These weak simulated winds would cause the transport of pollutants in an air quality model to be offset by hundreds of kilometers, as the distance between 100 m model-observed trajectory endpoints for the Cape Canaveral sites is: 98 (MIDFL), 150 (MLNFL), 180 (SCPFL), and 220 km (FCPFL). In fact, the average displacement for all 12 sites in the eastern US at 100 m is 150 km, at 350 m it is about 100 km, and at 1000 m it is approximately 140 km. These displacement errors are consistent with Haagenson et al. (1987) and Kahl (1996) who found simulated trajectory errors to be in the range of 150–200 km per 24 h. When the grid spacing is considered, this analysis suggests that the transport of a pollutant is potentially displaced 10–15 grid cells for this 12-km simulation. Thus, pollutant transport uncertainty because of errors in the simulated wind field near the surface should be considered when this simulated meteorology is used in air quality simulations.

4. Summary

In this study, several approaches were introduced to examine meteorological model's performance with an emphasis on variables and processes that would affect subsequent air quality simulations. The application of these approaches was illustrated by evaluating an annual simulation from the MM5 modeling system. First, standard statistical measures of model performance were calculated for seasonal subsets of data. The statistics reveal a universal cold bias in the winter, which in turn, could influence the air quality model (e.g., overestimation of aerosol nitrate because of an overestimation of relative humidity). In general, the meteorological models performed better in terms of

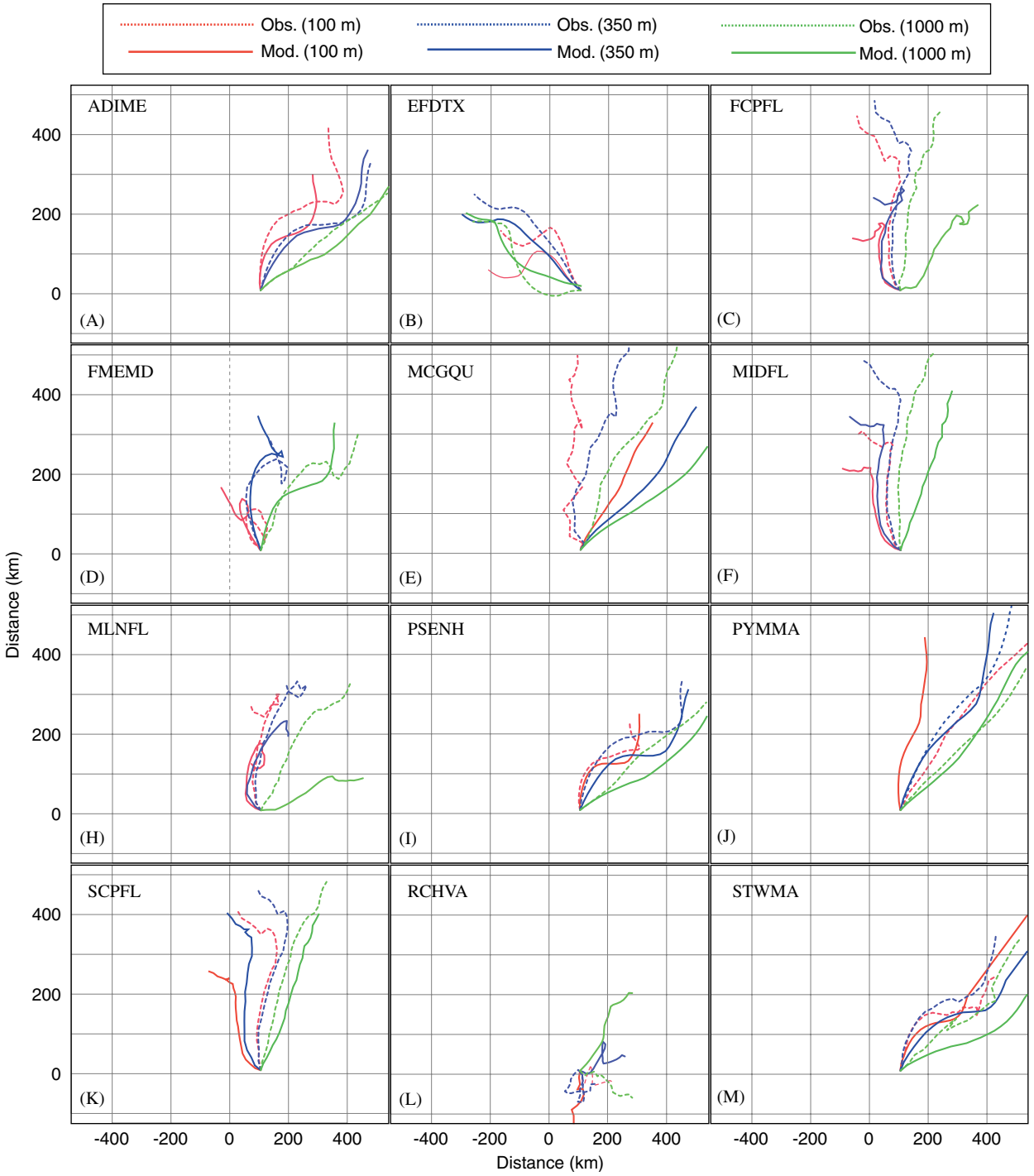

Fig. 2. Profiler-based observation and model trajectories at 12 sites in the Eastern US. The trajectories are calculated using two months of diurnally averaged wind speed and direction profiles. Starting point of trajectories is at 00 UTC. Three levels are plotted in different color with observation trajectories are the solid lines, and model trajectories are the dotted lines.

temperature during the warmer months. When the 12 km is compared with the 36 km simulation, the statistics indicate no improvement was achieved with the high-resolution modeling in simulating the 2 m temperature. The statistics calculated for the various subsets were valuable in diagnosing model

problems in that they help determine that the 2 m temperature simulation was similar across all subsets with the exception of the Midwest US and land-use that is categorized as plains or rangeland. Since the majority of the Midwest is classified as plains, this implies an error in the land-use characterization.

A new method was tested to diagnose model performance by examining the inter-correlation of observable variables in the atmosphere. Time series of the dry-bulb temperature, dewpoint temperature, wind speed, u–v wind components, pressure, precipitation, and cloud cover were first separated into intraday, diurnal, synoptic, and baseline components and then the correlation between each variable on each time scale was calculated. The inter-correlations were similar between the observations and model for many variable combinations, but several were dramatically different. In particular, the intraday response of temperature to precipitation was evident in the observations, but no such response was evident in the model output. The diurnal response of increasing wind with temperature was strong in the observed time series, but it is much weaker in the model. An examination of the modeled time series showed that temperature and wind speed was out of phase whereas it was in phase in the observed time series. These differences may cause errors in the air quality simulation of concentrations during boundary-layer transition periods. The diurnal response of temperature to cloud fraction was evident in the observations, but the model showed a negative correlation. This discrepancy could adversely effect the radiation at the surface and the air quality model's photochemical calculations.

Wind profilers were used to examine the simulated boundary-layer wind structure. A diurnal 2-month average vertical wind profile calculated for several sites across the eastern US were used to show mean-diurnal modeled and observed trajectories. Although the model captured many of the complex boundary layer variations throughout the day, the overall wind speed estimation by the model was less than those observed at most levels in the PBL. Of the 12 sites examined, the average deviation between the 24 h observed and modeled trajectory was about 150 km at 100 m above the surface. This trajectory deviation decreased at 350 and 1000 m above the surface, which indicates better model performance at levels away from the surface layer. Errors in transport of this magnitude (100–200 km) will have significant effects on the accuracy of the simulated concentration of ozone and fine particulates. A study is underway to relate the errors in the meteorological model presented here with errors in the concentrations simulated by the CMAQ air quality model.

Acknowledgements

The authors thank NOAA Forecast Systems Lab (FSL) and the National Center for Atmospheric Research (NCAR) for providing us the surface and profiler data used in the study.

Disclaimer: The research presented here was performed under the Memorandum of Understanding between the US Environmental Protection Agency (EPA) and the US Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) and under agreement number DW 13921548. This work constitutes a contribution to the NOAA Air Quality Program. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.

References

Abraczinskas, M.A., Olerud, D.T., Sims, A.P., 2004. Characterizing annual meteorological modeling performance for visibility improvement strategy modeling. In: The Southeastern US 13th Conference on the Applications of Air Pollution Meteorolgy with the Air and Waste Management Association, 23–27 August 2004, Vancouver, BC, Canada (Web address: http://ams.confex.com/ams/pdfpapers/80261.pdf).

Atkins, N.T., Wakimoto, R.M., 1994. Observations of the seabreeze front during cape: Part I: single-doppler, satellite, and cloud photogrammetry analysis. Monthly Weather Review 122, 1092–1114.

Baker, K., 2004. Meteorological modeling protocol for application to $PM_{2.5}$/haze/ozone modeling projects. Lake Michigan Air Directors Consortium, 12 January 2004, Des Plaines, IL, 9pp.

Bullock, O.R., Brehme, K.A., 2002. Atmospheric mercury simulation using the CMAQ model: formulation description and analysis of wet deposition results. Atmospheric Environment 36, 2135–2146.

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

Chang, J.C., Hanna, S.R., 2004. Air quality model performance evaluation. Meteorology and Atmospheric Physics 87, 167–196.

Dennis, R., Roselle, S., Gilliam, R., Arnold, J., 2004. High time-resolved comparisons for in-depth probing of CMAQ fine particle and gas predictions. In: The 27th NATO/CCMS International Technical Meeting on Air Pollution Modeling and Its Applications, 25–29 October 2004, Banff, Canada.

Emery, C.A., 2001. Enhanced meteorological modeling and performance evaluation for two Texas ozone episodes. Prepared for the Texas Natural Resource Conservation Commission, by ENVIRON International Corporation.

Fox, D.G., 1981. Judging air quality model performance. Bulletin of the American Meteorological Society 62, 599–609.

Gego, E., Hogrefe, C., Kallos, G., Voudouri, A., Irwin, J.S., Rao, S.T., 2005. Examination of model predictions at different horizontal grid resolutions. Environmental Fluid Mechanics 5, 63–85.

Gilliam, R.G., Raman, S., Niyogi, D., 2004. Observational and numerical study on the influence of large-scale flow direction and coastline shape on sea-breeze evolution. Boundary Layer Meteorology 111, 275–300.

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, 138pp.

Haagenson, P.L., Kuo, Y.-H., Skumanich, M., Seaman, N.L., 1987. Tracer verification of trajectory models. Journal of Applied Meteorology 26, 410–426.

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.

Hogrefe, C., Rao, S.T., Kasibhatla, P., Kallos, G., Tremback, C.J., Hao, W., Olerud, D., Xiu, A., McHenry, J., Alapaty, K., 2001a. Evaluating the performance of regional-scale photochemical modeling systems: Part I–meteorological predictions. Atmospheric Environment 35, 4159–4174.

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

Hogrefe, C., Biswas, J., Lynn, B., Civerolo, K., Ku, J.-Y., Rosenthal, J., Rosenzweig, C., Goldberg, R., Kinney, P.L., 2004. Simulating regional-scale ozone climatology over the eastern United States: model evaluation results. Atmospheric Environment 38, 2627–2638.

Hogrefe, C., Porter, P.S., Gego, E., Gilliland, A., Gilliam, R., Swall, J., Irwin, J., Rao, S.T., 2005. Temporal features in observed and predicted $pm_{2.5}$ concentrations over the eastern United States. Atmospheric Environment (the paper in this same special edition).

Janjic, Z.I., 1994. The step-mountain Eta coordinate model: further developments of the convection, viscous sublayer and turbulence closure schemes. Monthly Weather Review 122, 928–945.

Kahl, J.D.W., 1996. On the prediction of trajectory model error. Atmospheric Environment 30, 2945–2957.

Kain, J.S., 2004. The Kain–Fritsch convective parameterization: an update. Journal of Applied Meteorology 43, 170–181.

Mass, C.F., Ovens, D., Westrick, K., Colle, B.A., 2002. Does increasing horizontal resolution produce more skillful forecasts? Bulletin of the American Meterological Society 83, 407–430.

Mass, C.F., Albright, M., Ovens, D., Steed, R., MacIver, M., Grimit, E., Eckel, T., Lamb, B., Vaughan, J., Westrick, K., Storck, P., Colman, B., Hill, C., Maykut, N., Gilroy, M., Ferguson, S.A., Yetter, J., Sierchio, J.M., Bowman, C., Stender, R., Wilson, R., Brown, W., 2003. Regional environmental prediction over the Pacific Northwest. Bulletin of the American Meteorological Society 84, 1353–1366.

McKendry, I.G., Steyn, D.G., McBean, G., 1995. Validation of synoptic circulation patterns simulated by the Canadian climate center general circulation model for Western North America. Atmosphere Ocean 33, 809–825.

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

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J., Clough, S.A., 1997. RRTM, a validated correlated-k model for the longwave. Journal of Geophysical Research 102, 16663–16682.

Nenes, A., Pandis, S.N., Pilinis, C., 1998. ISORROPIA: a new thermodynamic equilibrium model for multiphase multicomponent inorganic aerosols. Aquatic Geochemistry 4, 123–138.

Otte, T.L., Pouliot, G., Pleim, J.E., Young, J.O., Schere, K.L., Wong, D.C., Lee, P.C.S., Tsidulko, M., McQueen, J.T., Davidson, P., Mathur, R., Chuang, H.-Y., DiMego, G., Seaman, N.L., 2005. Linking the eta model with the community multiscale air quality (CMAQ) modeling system to build a real-time national air quality forecasting system. Weather and Forecasting 20, 367–384.

Pielke, R.A., 1998. The need to assess uncertainty in air quality evaluations. Atmospheric Environment 32, 1467–1468.

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.

Pleim, J.E., Chang, J., 1992. A non-local closure method for vertical mixing in the convective boundary layer. Atmospheric Environment 26, 965–981.

Pleim, J.E., Xiu, A., 1995. Development and testing of a surface flux and planetary boundary layer model for application in mesoscale models. Journal of Applied Meteorology 34, 16–32.

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.

Rao, P.A., Fuelberg, H.E., Droegemeier, K.K., 1999. High-Resolution modeling of the cape canaveral area land–water circulations and associated features. Monthly Weather Review 127, 1808–1821.

Reisner, J., Rasmussen, R.M., Bruintjes, R.T., 1998. Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. Quarterly Journal of the Royal Meteorological Society 124B, 1071–1107.

Saulo, A.C., Seluchi, M., Campetella, C., Ferreira, L., 2001. Error evaluation of NCEP and LAHM regional model daily forecasts over Southern South America. Weather and Forecasting 16, 697–712.

Seaman, N.L., 2000. Meteorological modeling for air-quality assessments. Atmospheric Environment 34, 2231–2259.

Sistla, G., Zhou, N., Hao, W., Ku, J.-Y., Rao, S.T., Bornstein, R., Freedman, F., Thunis, P., 2001a. Effects of uncertainties in meteorological inputs on urban airshed model predictions and ozone control strategies. Atmospheric Environment 30, 2011–2025.

Sistla, G., Hao, W., Ku, J.-Y., Kallos, G., Zhang, K., Mao, H., Rao, S.T., 2001b. An operational evaluation of two regional-scale ozone air quality modeling systems over the eastern United States. Bulletin of the American Meteorological Society 82, 945–963.

Tesche, T.W., 2002. Operational evaluation of the MM5 meteorological model over the continental United States: protocol for annual and episodic evaluation. Prepared for the US Environmental Protection Agency, Office of Air Quality Planning and Standards, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences. Academic, New York, pp. 277–281.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bulletin of the American Meteorological Society 63, 1309–1369.

Vaughan, J., Lamb, B., Frei, C., Wilson, R., Bowman, C., Figueroa-Kaminsky, C., Otterson, S., Boyer, M., Mass, C., Albright, M., Koenig, J., Collingwood, A., Gilroy, M., Maykut, N., 2004. A numerical daily air quality forecast system for The Pacific Northwest. Bulletin of the American Meteorological Society 85, 549–561.

Zhang, D.-L., Zheng, W.-Z., 2004. Diurnal cycles of surface winds and temperatures as simulated by five boundary layer parameterizations. Journal of Applied Meteorology 43, 157–169.

Zhong, S., Takle, E.S., 1993. The effects of large-scale winds on the sea-land breeze circulations in an area of complex coastal heating. Journal of Applied Meteorology 32, 1181–1195.