Atmospheric Pollution Research
www.atmospolres.com

Using synoptic classification to evaluate an operational air quality forecasting system in Atlanta

Yongtao Hu [1], Michael E. Chang [2], Armistead G. Russell [1], M. Talat Odman [1]

[1] School of Civil and Environmental Engineering, Georgia Institute of Technology, 311 Ferst Drive, Atlanta, GA 30332, USA
[2] Brook Byers Institute for Sustainable Systems, Georgia Institute of Technology, Atlanta, Georgia, USA

ABSTRACT

Since 2006, a team of forecasters in Georgia (USA) has been using the high–resolution air quality forecasting system (Hi–Res) as an aid for making ozone (O_3) and fine particulate matter ($PM_{2.5}$) forecasts. Here, we examine Hi–Res's O_3 and $PM_{2.5}$ forecasting performance for the Atlanta metropolitan area during the summers of 2006–2009. A classificatory evaluation approach was adopted. The spatial synoptic classification (SSC) calendar for Atlanta was used to cluster the forecasting days into typical summer weather types of dry moderate, dry tropical, moist moderate, moist tropical, and a transition class. The forecasting days were also classified according to emissions conditions as special weekdays (Monday and Friday), typical weekdays and weekends/holidays. Evaluation of forecasts during 2006–2009 shows that O_3 performance was worse on moist days and better on dry days. This is an important concern for forecasters since a sizeable number of days that exceeded the National Ambient Air Quality Standard (NAAQS) for O_3 were observed under moist tropical weather type during the period. On the other hand, $PM_{2.5}$ performance during 2006–2008 was opposite – worse on dry days, especially on dry tropical days, and better on moist days. This too is a concern since higher concentrations of $PM_{2.5}$ were observed to occur on dry days. In 2009, $PM_{2.5}$ forecasting performance on dry days was improved significantly by integrating a new secondary organic aerosol (SOA) module into the system. As a result, the differences in $PM_{2.5}$ forecasting performance between dry and moist days were diminished. Other results of this study, suggest that a relatively larger forecasting error on weekends/holidays may be due to higher uncertainties in emission estimates on those days. To a lesser extent, this was also true on special weekdays because of the greater variations in rush hour emissions relative to typical weekdays.

Keywords:
Air quality forecasting
AQI
Classificatory evaluation

Article History:
Received: 07 April 2010
Revised: 28 August 2010
Accepted: 02 September 2010

Corresponding Author:
Yongtao Hu
Tel: +1-404-894-1854
Fax: +1-404-894-8266
E-mail: yh29@mail.gatech.edu

doi: 10.5094/APR.2010.035

1. Introduction

Next day forecast of the Air Quality Index (AQI) (U.S.EPA, 1999) are provided daily by state and local organizations in more than 300 major cities across the United States. These forecasts are compiled and disseminated through the EPA's AIRNow website (http://www.airnow.gov). The information is used to advise the general public, particularly sensitive groups (e.g. children, the elderly and those with pre–existing cardio respiratory diseases), to avoid or limit exposure on poor air quality days [e.g. when air pollutant concentrations are predicted to exceed National Ambient Air Quality Standards (NAAQS) for clean air]. Conceivably, such information could also be used to modify emissions in the short term to reduce air pollution. The forecasts are usually for ozone (O_3) and/or fine particulate matter (i.e., $PM_{2.5}$: particulate matter with an aerodynamic diameter less than 2.5 µm), and are often based on predictions from statistical models and/or numerical models running up to two days in advance (U.S.EPA, 2003). Both O_3 and $PM_{2.5}$ are photochemically formed from anthropogenic and biogenic precursor emissions, though $PM_{2.5}$ can also be emitted directly from sources like combustion processes and wind-blown dust. Once in the atmosphere, both O_3 and $PM_{2.5}$ can be transported long distances.

Forecasters that based their decisions on statistical models trained with historical air quality and meteorological data (McCollister and Wilson, 1975; Robeson and Steyn, 1990; Comrie, 1997; Ryan et al., 2000; Schlink et al., 2003; Goyal et al., 2006) found these approaches worked best under persistent meteorological conditions, but performed poorly in forecasting the onset and/or termination of pollution episodes (Comrie, 1997; McMillan et al., 2005; Eder et al., 2010). Forecasts based on numerical air quality models were first developed about a decade ago (Chang and Cardelino, 2000). These three–dimensional Eulerian chemical transport models are based on first principles and rely on detailed meteorological and emissions fields. Several such air quality forecasting systems are currently operational in the United States at either the national (e.g. covering the continental U.S. with a 12–km horizontal grid–spacing (Otte et al., 2005; Eder et al., 2006; Eder et al., 2009) or regional scales covering one or more metropolitan areas [e.g. Houston and Dallas, Texas (Byun et al., 2005), Pacific Northwest (Vaughan et al., 2004), New York–New Jersey (Cai et al., 2008) and Georgia (Odman et al., 2007)]. Such systems generally achieve good forecasting accuracy for O_3 (Eder et al., 2006; Eder et al., 2009; Hu et al., 2009), but $PM_{2.5}$ accuracy has been poor. Most recently though, a newly developed secondary organic aerosol (SOA) approach (Baek, 2009) has proven quite successful in improving $PM_{2.5}$ forecast accuracy (Hu et al., 2009).

The performance of both statistical and numerical models can be evaluated by comparing predictions with observations. Error metrics provide useful information for using modeling/forecasting results wisely. However, as traditional evaluation tends to evaluate the overall performance of a numerical modeling system (Russell and Dennis, 2000; Hogrefe et al., 2001; Fine et al., 2003; Fiore et al., 2003; Hogrefe et al., 2006; Morris et al., 2006; Eder and Yu, 2006; Appel et al., 2007; Appel et al., 2008; Dennis et al., 2010), detailed information is often lacking for use by local forecasters (Mueller, 2009). How to effectively guide the usage of the operational forecasting products from numerical models in assisting local specific AQI forecasting is a rising issue (Eder et al., 2010). Classificatory evaluation techniques based on clustering analysis are often location specific and provide additional information that complements traditional methods (Beaver and Palazoglu, 2006; Gilliam et al., 2006; Appel et al., 2007). In this paper, we introduce a new approach for evaluating operational air quality forecasting systems. Our primary aim is to identify location specific forecasting bias. This innovative method is used to evaluate the performance of the Hi–Res air quality forecasting system (Odman et al., 2007) in Atlanta for four summers during 2006–2009.

2. Approach

2.1. Hi–Res air quality forecasting system

The high–resolution air quality forecasting system, Hi–Res, has been used by forecasters in the state of Georgia (USA) since 2006. Hi–Res uses the Weather Research and Forecasting model (WRF) (Michalakes et al., 2005) for predicting meteorological fields, the Sparse Matrix Operator Kernel Emissions model (SMOKE) (CEP, 2003) for emissions processing, and the Community Multi–Scale Air Quality (CMAQ) model (Byun and Schere, 2006) for simulating the chemistry and transport of air pollutants (Figure 1a). The emissions inventory used in Hi–Res as input to SMOKE is projected from a 2002 "typical year" inventory (MACTEC, 2005) to the current year using growth factors from the Economic Growth Analysis System (EGAS, http://www.epa.gov/ttn/ecas/egas5.htm) and control factors for the existing federal and local control strategies. Hi–Res nests its 4–km forecasting grid that covers Georgia in a 12–km mother grid that extends coverage to portions of neighboring states, and uses a 36–km outer grid over the eastern U.S. to provide air quality boundary conditions (Figure 1b). Hi–Res is run for two cycles each day: 00Z and 12Z. In each cycle, forecasts are developed for a 48–hour period that shifts the previous cycle's forecasts ahead by 12 hours. Hi–Res first simulates a 66–hour period starting from 00Z (or 12Z) on the 36–km grid. WRF is initialized and constrained at the boundaries using 00Z (or 12Z) 84–hour forecast products from the North American Mesoscale (NAM) model (http://nomads.ncdc.noaa.gov) while CMAQ is initialized from the previous forecasting cycle and uses "clean" boundary conditions for the 36–km grid (Figure 1c). Then, Hi–Res simulates the same 66–hour period on the 12–km grid, using NAM forecast products for WRF initialization and boundary constraints, and the 36–km air quality outputs for CMAQ boundary conditions. From the 12–km grid simulation, Hi–Res nests down to the 4–km grid for the last 51 hours that spans the 48–hour forecasting period of interest (Figure 1c). The simulations for each cycle take about 8 hours on a dedicated 4–core Intel Xeon processor on a 64–bit platform server.

Hi–Res air quality forecasts started as O_3 only forecasts for the summer of 2006. In 2007 and continuing to present day, both O_3 and $PM_{2.5}$ are forecasted (note: $PM_{2.5}$ was also modeled in 2006, but it was not reported in to the forecasting team). Forecast products are disseminated through the website http://forecast.ce.gatech.edu, as spatial distributions within the 4–km forecasting grid, and as temporal profiles at air quality monitoring sites, of not only O_3 and $PM_{2.5}$, but also AQI, winds, temperature, and precipitation. A report of representative single value forecasts for each major metro area in Georgia, including Atlanta, is also generated for use by the Georgia Environmental Protection Division (GA–EPD) as an aid for issuing official local AQI forecasts.

The Hi–Res system has evolved through the years but the model's underlying physical and chemical parameterizations have remained constant. For example, Hi–Res switched from single–cycle forecasting to two–cycle forecasting in 2008. In 2009, the 4–km domain was enlarged to cover the entire state of Georgia. At the beginning of each year, the emissions inventory is projected to the current year. In addition, before the O_3 season (May through September) of each year, WRF is updated to the latest release, though any physics options are intentionally kept the same. The same version of CMAQ (4.6) with extensions developed at the Georgia Institute of Technology (Hu et al., 2006), has been used during the past four years with no change in the model configurations. The one exception is that in 2009, we introduced a new secondary organic aerosol (SOA) module into the CMAQ model that includes known SOA formation pathways and multi-generational SOA producing reactions (Baek, 2009).

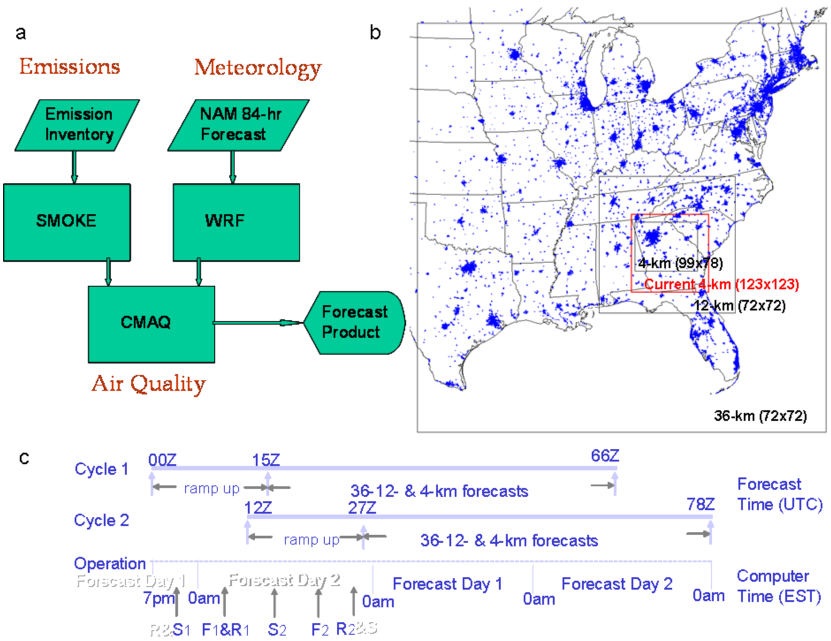


Figure 1. *The Hi–Res air quality forecasting system:* ***(a)*** *System components,* ***(b)*** *Modeling domains (urban areas are shown in blue), and* ***(c)*** *Forecasting cycles. S: start simulation, F: finish simulation, R: release products.*

2.2. Observations

Hourly O_3 and $PM_{2.5}$ observations were used for evaluating the representative single value forecasts for the Atlanta metropolitan area. Observations are available at eleven O_3 and seven $PM_{2.5}$ monitoring sites from the State and Local Air Monitoring Stations (SLAMS) network in Atlanta (Figure 2). The daily maxima of 8–hour average O_3 concentrations and 24–hr average $PM_{2.5}$ concentrations among the monitoring sites are used as the "single value" observations in Atlanta. These are then compared with the "single value" forecasts, corresponding to the maxima among the simulated values at the same locations.

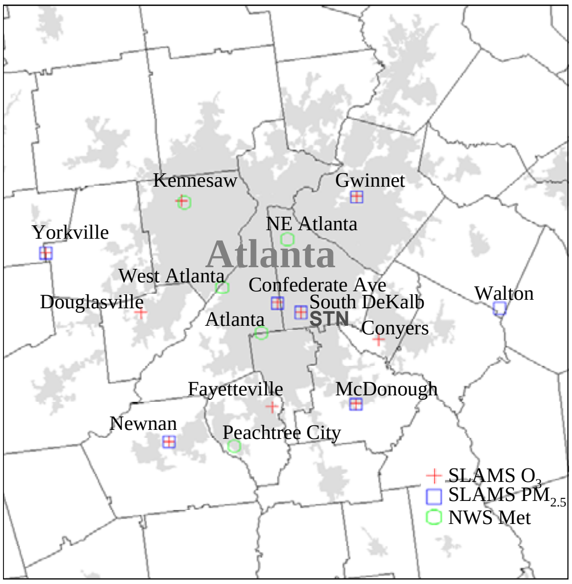


Figure 2. *Air quality and meteorology monitoring sites in Atlanta, GA area.*

Hourly readings of surface meteorological variables were also collected from five National Weather Service (NWS) monitoring sites located in Atlanta (Figure 2). Five–site averages of the daily high temperature and daily average specific humidity at 2 m, and wind speed at 10 m are obtained as "single value" meteorological observations in Atlanta, to be compared with their corresponding composites from simulations at the same five locations.

2.3. Classificatory evaluation

Performance is evaluated using the aforementioned composite observations for each day of the four summers (May 1 through September 30) during 2006–2009. Metrics of mean normalized bias (MNB) and error (MNE) are adopted to evaluate the discrepancy between the forecasts and the observations for O_3 and $PM_{2.5}$. Mean bias (MB) and error (ME) are used for evaluating the meteorological variables. Mean of forecasts (MF) and mean of observations (MO) are also calculated. These metrics are defined as follows:

$$MNB = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(f_i - o_i\right)}{o_i}\times 100\% \qquad (1)$$

$$MNE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|f_i - o_i\right|}{o_i}\times 100\% \qquad (2)$$

$$MB = \frac{1}{N}\sum_{i=1}^{N}\left(f_i - o_i\right) \qquad (3)$$

$$ME = \frac{1}{N}\sum_{i=1}^{N}\left|f_i - o_i\right| \qquad (4)$$

$$MF = \frac{1}{N}\sum_{i=1}^{N} f_i \qquad (5)$$

$$MO = \frac{1}{N}\sum_{i=1}^{N} o_i \qquad (6)$$

where f_i and O_i are the "single value" forecast and observation, respectively, of the i th valid forecast–observation pair, and N is the total number of such valid pairs of interest.

Forecasting performance is examined at first in overall and then at the classified cluster levels. Two types of classificatory evaluations are conducted here: (1) sub–setting according to synoptic classification, and (2) sub–setting according to emissions conditions. Cumulative percentage distributions are then constructed using the performance metrics such as MNE for each sub–set and compared between sub–sets.

Synoptic classification. Spatial synoptic classification (SSC) is used (Kalkstein et al., 1996; Sheridan, 2002) to cluster the synoptic environment in the Atlanta area for each day during the study periods. SSC has been applied to studies of heat wave mortality (Rainham et al., 2005), epidemiology (Goldberg, 2007), O_3 variability (Davis et al., 2009) and aerosol variability (Power et al., 2006). It has also been applied to study the heat–island–initiated (Dixon and Mote, 2003) and aerosol–associated (Lacke et al., 2009) precipitation patterns in Atlanta. SSC classifies each day at a location into one of six weather types, i.e. dry polar (DP), dry moderate (DM), dry tropical (DT), moist polar (MP), moist moderate (MM), moist tropical (MT), or transition (TR). In order to establish SSC, seed days that typify a particular weather type at a particular location are first pre–selected and then evaluated through discriminant analysis to measure the separation among the weather types with respect to multiple variables simultaneously. A distinct combination of these multiple variables such as temperature, dew point depression, mean cloud cover, mean sea level pressure, diurnal temperature range, and diurnal dew point range, is the so–called seed–day criteria. It determines each SSC and represents typical surface meteorological characteristics of a certain air mass at the location. Depending on the evaluation results, the SSC approach is re–run with modified seed–day criteria and reselected seed days until modification is no longer needed. SSC produces year–round weather–type classifications. SSC calendars are available for stations nationwide and are updated daily on a website (http://sheridan.geog.kent .edu/ssc.html). The SSC calendar for Atlanta, as summarized in Table 1, shows that there were significantly more moist days (MM and MT) and less dry days (DM and DT) in the summer of 2009 than years 2006–2008. It also shows that weather types of dry polar and moist polar rarely occur in Atlanta during summertime.

Table 1. *Spatial synoptic classification in Atlanta during summers of 2006-2009*

SSC	DP	DM	DT	MP	MM	MT	TR	MISSING	TOTAL
2006	4	60	15	2	18	45	7	2	153
2007	3	48	20	0	24	33	19	6	153
2008	2	54	9	1	22	44	15	6	153
2009	4	28	4	1	46	62	7	1	153
Total	13	190	48	4	110	184	48	15	612

Emissions conditions. Emission rates of pollutants such as NO_x change during the week (Beirle et al., 2003; Kaynak et al., 2009). Compared to weekdays, they decrease on weekends, even though power plant emissions remain relatively constant. In urbanized areas such as Atlanta, different rush hour patterns, hence different emission rates, are observed on Monday and Friday than other weekdays. Most importantly, emission estimation uncertainties differ among the days of the week for a specific area and these uncertainties also vary between areas (Kaynak et al., 2009). Therefore, we group each day of the study periods into one of three types: special weekdays (Monday and Friday), typical weekdays (Tuesday through Thursday), and weekends/holidays, to examine the forecasting performance in terms of difference in emissions estimation errors.

3. Results

3.1. Overall performance during 2006–2009

Forecasting performance remained steady, with MNE's below 30% and around 40%, respectively, for O_3 and $PM_{2.5}$ for the years 2006–2008 (Table 2). In 2009, O_3 bias increased and the $PM_{2.5}$ performance improved dramatically. Compared to other years, 2009 was cooler by ~1.0 K and more humid by 1 g kg^{-1} in specific humidity (Table 3), and was much cleaner in terms of both O_3 and $PM_{2.5}$ (Table 2). The cleaner air in Atlanta is attributable to the economic down–turn and subsequent reduction in emissions and the much higher rainfall. Atlanta experienced three consecutive drought years during 2006–2008, but in 2009 it received the highest amount of rainfall since 1948 (http://www.srh.noaa.gov/ffc/?n=rainfall_scorecard). Hi–Res performed best at predicting both temperature and specific humidity in 2009 (Table 3), and had comparable performance to the other years in predicting the wind speed.

Table 2. *Forecasting performance statistics for O_3 and $PM_{2.5}$*

Variable		O_3	$PM_{2.5}$
2006	MO	69 ppb	24.4 µg m^{-3}
	MF	76 ppb	14.8 µg m^{-3}
	MNB	11%	-34%
	MNE	29%	38%
2007	MO	69 ppb	25.8 µg m^{-3}
	MF	70 ppb	14.6 µg m^{-3}
	MNB	8.5%	-37%
	MNE	19%	44%
2008	MO	61 ppb	20.6 µg m^{-3}
	MF	68 ppb	11.9 µg m^{-3}
	MNB	17%	-38%
	MNE	23%	42%
2009	MO	52 ppb	16.0 µg m^{-3}
	MF	63 ppb	16.2 µg m^{-3}
	MNB	28%	8%
	MNE	30%	25%
Overall	MO	63 ppb	21.9 µg m^{-3}
	MF	69 ppb	14.4 µg m^{-3}
	MNB	17%	-25%
	MNE	24%	37%

The slightly high bias of O_3 in 2009 is likely linked to the emissions model's standard growth factors not capturing the recent economic downturn and its reducing effect on NO_x emissions. The reasons for $PM_{2.5}$ performance improvement in 2009 are complex. CMAQ historically has significantly under–predicted organic species. But, better prediction of temperature and humidity, and the enhanced SOA module all contributed to improved performance in 2009. Despite a year in which more rainfall occurred (both in actual and in the model), the simulated $PM_{2.5}$ average of 16.2 µg m^{-3} (versus the observed 16.0 µg m^{-3}) was much higher than previous years' simulations (Table 2). In order to look at the $PM_{2.5}$ performance closely at its component level, we

Table 3. *Forecasting performance statistics for meteorological variables at surface*

Variable		Temperature (K)	Specific Humidity (g kg^{-1})	Wind speed (m s^{-1})
2006	MO	303.5	12.4	2.2
	MF	303.0	11.7	3.0
	MB	-0.50	-0.74	0.8
	ME	1.57	1.39	1.0
2007	MO	304.4	12.1	1.9
	MF	303.3	11.5	3.1
	MB	-1.16	-0.59	1.2
	ME	1.90	1.22	1.2
2008	MO	303.4	12.3	2.3
	MF	302.6	11.4	3.1
	MB	-0.80	-0.89	0.8
	ME	1.94	1.33	0.9
2009	MO	302.7	13.1	2.2
	MF	303.1	12.8	3.0
	MB	0.37	-0.36	0.8
	ME	1.47	0.96	0.9
Overall	MO	303.5	12.5	2.2
	MF	303.0	11.8	3.1
	MB	-0.53	-0.65	0.9
	ME	1.72	1.23	1.0

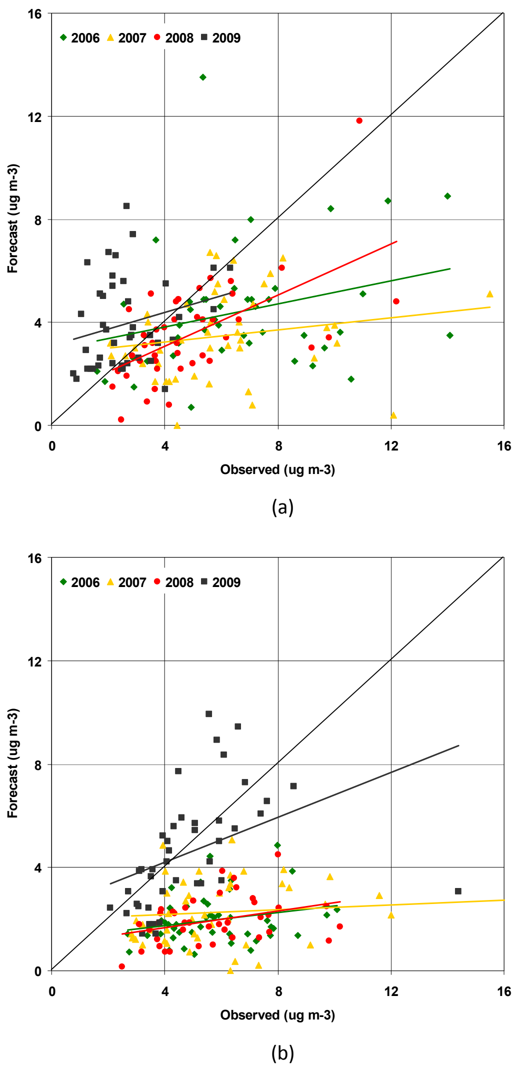


Figure 3. *Forecasting performance of **(a)** fine particulate sulfate and **(b)** fine particulate organic carbon at South DeKalb, a U.S. EPA Speciation Trends Network (STN) site.*

used 24–h measurements of speciated $PM_{2.5}$ from the Speciation Trends Network (STN) site located at South DeKalb (Figure 2) collected every third day and compared them with model predictions corresponding to the same location. The slightly higher bias of sulfate predictions in 2009 implies a slight overestimation of SO_2 emissions relative to other years (Figure 3a). Our analyses support that the significant improvement of organic carbon (OC) predictions was due to the enhanced SOA module and helped improve the $PM_{2.5}$ performance, albeit anthropogenic volatile organic compounds (VOCs) and carbonaceous emissions might be overestimated (Figure 3b).

3.2. Linking performance to weather types

Days classified as dry polar are the lowest in temperature with clear and dry air. Moist polar days are cloudy with humid and cool air. Since Atlanta has few days in these two classifications during the summer season, we drop them out from our following analysis. The air on dry moderate (DM) days is mild and dry, while dry tropical (DT) days are the hottest and driest (Table 4). Compared to moist polar days, moist moderate (MM) days are warmer and more humid, while moist tropical (MT) air is warm and humid.

Meteorological forecasting error is largest for specific humidity on "tropical" days (MT and DT). Apparently, the system is better at forecasting temperatures on moderate days but not hotter (DT) and cooler (MM) days (Table 4). However, in general, the differences are small between weather types in terms of forecasting performance for all three examined meteorological variables.

When classified by weather, distinctive differences are found in observed O_3 and $PM_{2.5}$ concentrations (e.g. compare the cumulative percentage distributions in Figure 4). Analysis of variance (ANOVA) showed *P* values less than 10^{-10} for both O_3 and $PM_{2.5}$ observations when grouped by weather types: this suggests a significant association between O_3/$PM_{2.5}$ observations and weather type. Dry tropical days are the most polluted and moist moderate days are the cleanest (Table 4). Dry moderate days are relatively cleaner than dry tropical days, but contributed the most number of days exceeding the NAAQS. Moist tropical days are considered moderately clean, but still exceed the NAAQS quite often. Transition (TR) days are defined as days in which one weather type yields to another, based on large shifts in pressure, dew point, and wind over the course of the day. Because of this characteristic, air quality status on transition days was dependent on the weather types of the contiguous days surrounding it. There are quite a number of NAAQS exceedence days classified as TR (Table 4). The significance of the differences is also indicated by the ANVOA tests for the O_3 and $PM_{2.5}$ forecasting performances (as MNE) among weather types. The calculated *P* values are less than 10^{-10} and 0.03, respectively, for the O_3 and $PM_{2.5}$ forecasting performances, suggesting a strong variation of O_3 performance, but much weaker variation of $PM_{2.5}$ performance, by weather types.

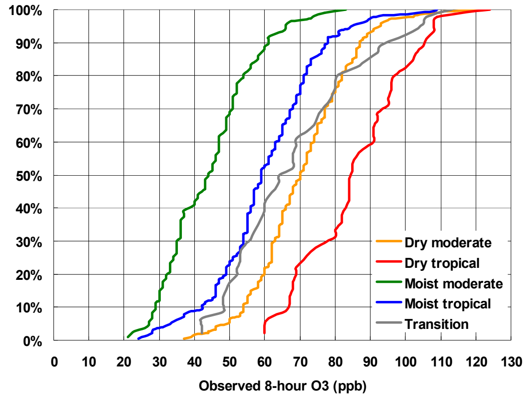
Because of the updated SOA module in CMAQ and the dramatic change in classification distribution among weather types in 2009, we separated the summer of 2009 from the other years for O_3 and $PM_{2.5}$ evaluations. O_3 forecasting errors (as MNE) show the opposite of observations in cumulative percentage distribution patterns among weather types (Figure 5a and 5c versus Figure 4a). Moist (MT and MM) days have lower observed O_3 concentrations but larger forecasting errors. Dry (DT and DM) days have higher observed O_3 concentrations but lower forecasting errors. This is expected because dry days are clear and moist days are cloudy and O_3 photochemistry is less well established in the model in the presence of clouds. This results in larger uncertainty in O_3 predictions on moist days. Forecasters would be advised to pay closer attention to moist tropical days since there were a large fraction of NAAQS exceedence days in this weather class and a larger error is expected from the forecasting model. O_3 performance is found to be worse in 2009 than during 2006–2008 for moist weather types (Figure 5a versus 5c). At the same time, the forecasting performance for meteorological variables such as specific humidity and temperature in 2009 is better than the other years. Therefore, most likely, this degradation in O_3 performance was due to increased errors/biases in 2009 emissions. Note that the forecasting errors for transition days are fairly good compared to other weather types. This suggests that the numerical modeling system captures the meteorology on transition days at least as well as on the days of other weather types.

$PM_{2.5}$ performance, however, differs significantly between 2006–2008 and 2009 in the cumulative percentage distribution patterns among weather types (Figure 5b versus 5d). Distribution patterns of 2009 show poorer performance on moist days and better performance on dry days, similar to the O_3 performance distribution patterns for both periods. In contrast, 2006–2008's distribution patterns show the opposite to some extent; in particular, dry tropical days performed the poorest (despite the highest observed concentrations for this weather type and that a normalized metric is adopted for calculating error), while days associated with other weather types, especially moist days, had much better performance (Figure 5b). The hypothesis here is that under clear conditions on dry days, such as dry tropical days, strong photochemistry leads to large amounts of SOA in the particle phase of the air in the area, but this mechanism was missing from the previous CMAQ model used for the 2006–2008 forecasting. The new SOA module introduced into the model in 2009 includes the missing SOA formation pathways. This advancement in the model was more significant on dry days than on moist days, though it improved the $PM_{2.5}$ performance under all weather types (Table 4).

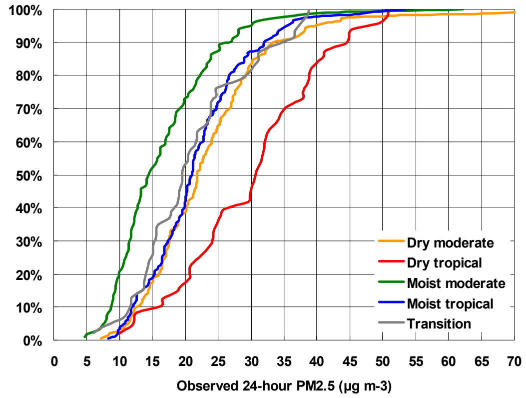
Table 4. *Forecasting error versus observation for spatial synoptic classifications*

	SSC [a]	DM	DT	MM	MT	TR
O_3	MO (ppb)	71	86	44	60	67
	Number of days > 75 ppb	65	33	2	25	15
	MNE 2006-2008	14%	18%	39%	27%	14%
	MNE 2009	13%	8%	48%	26%	19%
$PM_{2.5}$	MO (μg m^{-3})	23.5	30	16.7	21.8	21.3
	Number of days > 35 μg m^{-3}	17	15	3	10	5
	MNE 2006-2008	40%	44%	47%	38%	42%
	MNE 2009	17%	19%	30%	25%	34%
Temperature	MO (K)	303.4	307.5	300.7	304.9	304.1
	ME (K)	1.63	2.19	1.97	1.46	1.80
Specific Humidity	MO (g kg^{-1})	10.7	11.9	13.7	14.4	11.8
	ME (g kg^{-1})	1.20	1.27	1.04	1.37	1.26
Wind speed	MO (m s^{-1})	2.1	1.9	2.3	2.0	2.5
	ME (m s^{-1})	1.0	0.9	1.2	1.0	1.1

[a] *Statistics are for 2006-2009 summers where not specified.*



(a)



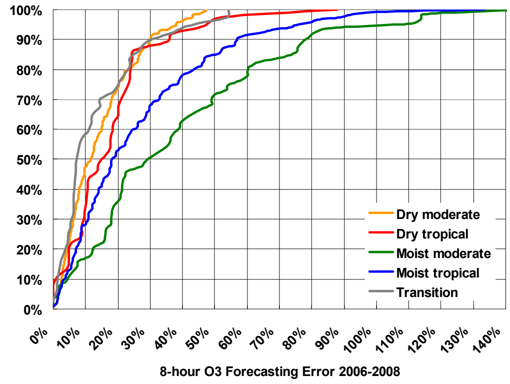
(b)

Figure 4. *Cumulative percentage distributions of O_3* ***(a)*** *and $PM_{2.5}$* ***(b)*** *observations for weather types.*

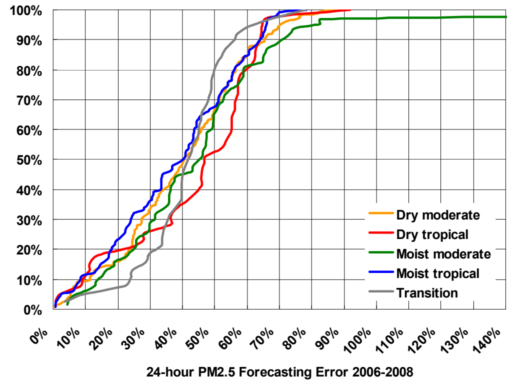
3.3. Linking performance to emissions conditions

Forecasting errors differ among the three emission classes (special weekdays, typical weekdays and weekends/holidays), though the differences are small in general (Table 5). Slightly larger differences are found for $PM_{2.5}$ forecasting error. For both O_3 and $PM_{2.5}$, best performance is found on typical weekdays followed by special weekdays. Performance is the worst on weekends/holidays, implying higher uncertainties in emissions estimates during those days (Table 5). Uncertainties in weekend emission estimates likely impact Monday simulations.

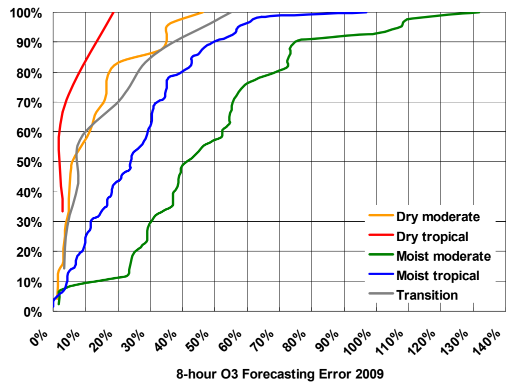
The cumulative percentage distributions of forecasting error (Figure 6) also show small performance differences among the three types of days, especially for O_3. There is a 62%, 65%, and 65% chance of having an above average O_3 forecasting accuracy (MNE of 24%), respectively, on special weekdays, typical weekdays, and weekends/holidays (Figure 6a). However, there is a higher probability, i.e. 20%, on weekends/holidays, versus 17% and 15% on typical and special weekdays, respectively, to have a much worse O_3 performance (i.e. MNE larger than 40%). Note that the difficulty in making accurate emissions estimates for the special rush hour patterns on special weekdays may have reduced the chance to have the best O_3 performance (MNE less than 20%) during these days. The chances are 53%, 57%, and 40% to achieve an above average $PM_{2.5}$ forecasting accuracy (MNE of 37%), respectively, on special weekdays, typical weekdays and weekends/holidays (Figure 6b). $PM_{2.5}$ forecasting has always had a lower chance of achieving a given performance on weekends/ holidays compared to weekdays (Figure 6b).



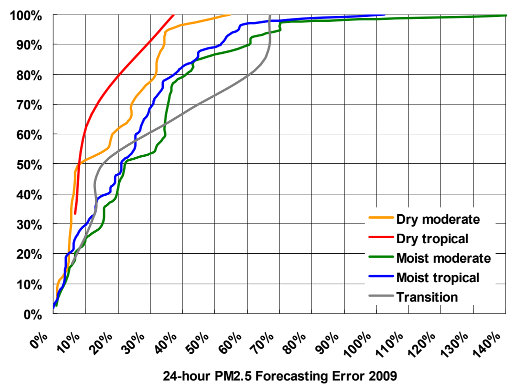
(a)



(b)



(c)

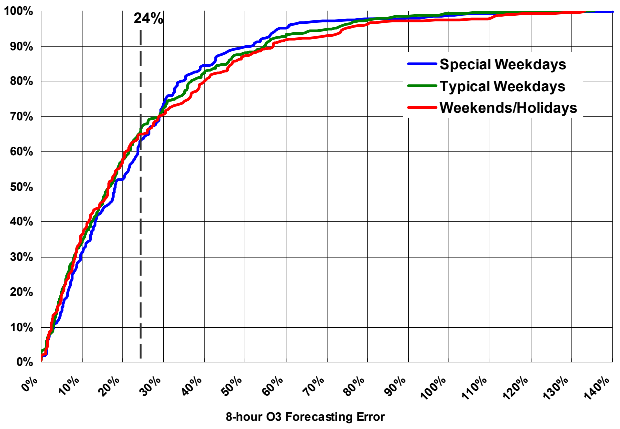


(d)

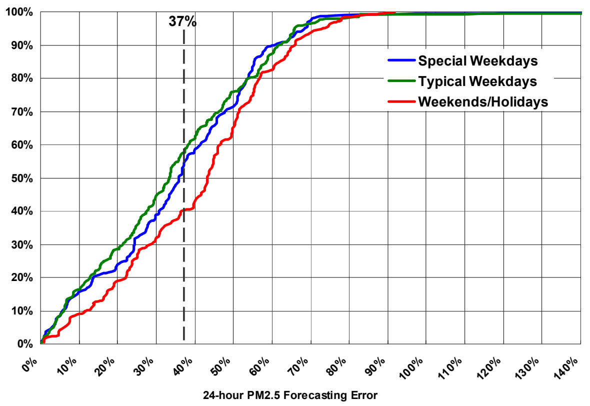
Figure 5. *Cumulative percentage distributions of forecasting error for weather types:* ***(a)*** *O_3 during 2006-2008,* ***(b)*** *$PM_{2.5}$ during 2006-2008,* ***(c)*** *O_3 in 2009, and* ***(d)*** *$PM_{2.5}$ in 2009.*

Table 5. *Forecasting error (MNE) versus observation (MO) for three types of days*

Subset	Special weekdays		Typical weekdays		Weekends/Holidays	
O_3	24%	64 ppb	23%	64 ppb	25%	61 ppb
$PM_{2.5}$	37%	20.7 µg m^{-3}	35%	22.0 µg m^{-3}	41%	22.8 µg m^{-3}



(a)



(b)

Figure 6. *Cumulative percentage distributions of O_3* ***(a)*** *and $PM_{2.5}$* ***(b)*** *forecasting error (NME) for day of week groups.*

4. Summary

This study adopted a new approach to evaluate operational air quality forecasting performance for specific locations. The classificatory evaluation method used here clusters the forecasting days according to synoptic classifications and emissions conditions. This method complements the traditional air quality forecasting/ modeling evaluation methods by providing additional local specific guidance information on how to interpret air quality forecasting results.

Our study revealed distinct capabilities of the Hi–Res air quality forecasting system for predicting O_3 and $PM_{2.5}$ for the Atlanta area under different weather types during the summertime, including dry moderate, dry tropical, moist moderate, moist tropical and transition conditions. Smaller, but still significant, differences in prediction capabilities were found under different emissions conditions, i.e. special weekdays, typical weekdays and weekends/holidays. Results show that O_3 performance of the Hi–Res system in Atlanta was worse on moist days and better on dry days. This suggests that forecasters should focus their attention on moist tropical days since a sizable number of NAAQS exceedence days are expected under this weather type. During 2006–2008, $PM_{2.5}$ performance, as quantified using the MNE, was worse on dry days, especially on dry tropical days, than on moist days, despite the higher concentration of $PM_{2.5}$ on dry days. In 2009, $PM_{2.5}$ performance improved under all weather types, though most significantly on dry days, owing to the new SOA module. Our results also suggest that a relatively larger forecasting error is expected on weekend/holidays due to higher uncertainties in emission estimates on those days. While the Hi–Res forecasting system performed better on weekdays, on special weekdays the forecasting capability became slightly worse because of the rush hour pattern's impacts on emissions estimating accuracy.

Acknowledgements

The Hi-Res air quality forecasting operation is funded by the Georgia Environmental Protection Division.

References

Appel, K.W., Gilliland, A.B., Sarwar, G., Gilliam, R.C., 2007. Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: sensitivities impacting model performance Part I-Ozone. *Atmospheric Environment* 41, 9603-9615.

Appel, K.W., Bhave, P.V., Gilliland, A.B., Sarwar, G., Roselle, S.J., 2008. Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: sensitivities impacting model performance; Part II-particulate matter. *Atmospheric Environment* 42, 6057-6066.

Baek, J., 2009. *Improving Aerosol Simulations: Assessing and Improving Emissions and Secondary Organic Aerosol Formation in Air Quality Modeling*. Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, GA, 140 pp.

Beaver, S., Palazoglu, A., 2006. A cluster aggregation scheme for ozone episode selection in the San Francisco, CA Bay Area. *Atmospheric Environment* 40, 713-725.

Beirle, S., Platt, U., Wenig, M., Wagner, T., 2003. Weekly cycle of NO_2 by GOME measurements: a signature of anthropogenic sources. *Atmospheric Chemistry and Physics* 3, 2225-2232.

Byun, D., Schere, K.L., 2006. Review of the governing equations, computational algorithms, and other components of the models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Reviews* 59, 51-77.

Byun, D.W., Jang, M.D., Kim, S.T., Perma, R., 2005. Development and operational evaluation of the Eastern Texas Air Quality (ETAQ) forecasting system. *4th Annual CMAS Models-3 Users' Conference*, Friday Center, UNC-Chapel Hill, September 26-28, 2005.

Cai, C., Hogrefe, C., Katsafados, P., Kallos, G., Beauharnois, M., Schwab, J.J., Ren, X., Brune, W.H., Zhou, X., He, Y., Demerjian, K.L., 2008. Performance evaluation of an air quality forecast modeling system for a summer and winter season – photochemical oxidants and their precursors. *Atmospheric Environment* 42, 8585-8599.

CEP, 2003. *Sparse Matrix Operator Kernel Emissions Modeling System (SMOKE) User Manual*, Carolina Environmental Program - The University of North Carolina at Chapel Hill, Chapel Hill, NC.

Chang, M.E., Cardelino, C., 2000. Application of the urban airshed model to forecasting next-day peak ozone concentration in Atlanta, Georgia. *Journal of the Air and Waste Management Association* 50, 2010-2024.

Comrie, A.C., 1997. Comparing neural networks and regression models for ozone forecasting. *Journal of the Air and Waste Management Association* 47, 653-663.

Davis, R.E., Normile, C.P., Sitka, L., Hondula, D.M., Knight, D.B., Gawtry, S.P., Stenger, P.J., 2009. A comparison of trajectory and air mass approaches to examine ozone variability. *Atmospheric Environment* 44, 64-74.

Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S.T., Scheffe, R., Schere, K., Steyn, D., Venkatram, A., 2010. A framework for evaluating regional-scale numerical photochemical modeling systems. *Environmental Fluid Mechanics* 10, 471-489.

Dixon, P.G., Mote, T.L., 2003. Patterns and causes of Atlanta's urban heat island–initiated precipitation. *Journal of Applied Meteorology* 42, 1273-1284.

Eder, B., Yu, S., 2006. A performance evaluation of the 2004 release of Models-3 CMAQ. *Atmospheric Environment* 40, 4811-4824.

Eder, B., Kang, D., Rao, S.T., Mathur, R., Yu, S., Otte, T., Schere, K., Wayland, R., Jackson, S., Davidson, P., McQueen, J., Bridgers, G., 2010. Using national air quality forecast guidance to develop local air quality index forecasts. *Bulletin of the American Meteorological Society* 91, 313-326.

Eder, B., Kang, D., Mathur, R., Pleim, J., Yu, S., Otte, T., Pouliot, G., 2009. A performance evaluation of the National Air Quality Forecast Capability for the summer of 2007. *Atmospheric Environment* 43, 2312-2320.

Eder, B., Kang, D., Mathur, R., Yu, S., Schere, K., 2006. An operational evaluation of the Eta-CMAQ air quality forecast model. *Atmospheric Environment* 40, 4894-4905.

Fine, J., Vuilleumier, L., Reynolds, S., Roth, P., Brown, N., 2003. Evaluating uncertainties in regional photochemical air quality modeling. *Annual Review of Environment and Resources* 28, 59-106.

Fiore, A.M., Jacob, D.J., Mathur, R., Martin, R.V., 2003. Application of empirical orthogonal function to evaluate ozone simulations with regional and global models. *Journal of Geophysical Research D: Atmospheres* 108, ACH 10-1 - ACH 10-10.

Gilliam, R.C., Hogrefe, C., Rao, S.T., 2006. New methods for evaluating meteorological models used in air quality applications. *Atmospheric Environment* 40, 5073–5086.

Goldberg, M.S., 2007. On the interpretation of epidemiological studies of ambient air pollution. *Journal of Exposure Science and Environmental Epidemiology* 17, S66-S70.

Goyal, P., Chan, A.T., Jaiswal, N., 2006. Statistical models for the prediction of respirable suspended particulate matter in urban cities. *Atmospheric Environment* 40, 2068-2077.

Hogrefe, C., Porter, P.S., Gego, E., Gilliland, A., Gilliam, R., Swall, J., Irwin, J., Rao, S.T., 2006. Temporal features in observed and simulated meteorology and air quality over the Eastern United States. *Atmospheric Environment* 40, 5041-5055.

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

Hu, Y., Odman, M.T., Chang, M.E., Russell, A.G., 2009. Comprehensive evaluation on air quality forecasting ability of Hi-Res in southeastern United States. 8th Annual Community Modeling and Analysis System (CMAS) Conference, October 19-21, 2009, Chapel Hill, NC.

Hu, Y., Odman, M.T., Russell, A.G., 2006. Mass conservation in the Community Multiscale Air Quality model. *Atmospheric Environment* 40, 1199-1204.

Kalkstein, L.S., Nichols, M.C., Barthel, C.D., Greene, J.S., 1996. A new spatial synoptic classification: application to air-mass analysis. *International Journal of Climatology* 16, 983-1004.

Kaynak, B., Hu, Y., Martin, R.V., Sioris, C.E., Russell, A.G., 2009. Comparison of weekly cycle of NO_2 satellite retrievals and NO_x emission inventories for the continental United States. *Journal of Geophysical Research D: Atmospheres* 114, art. no. D05302.

Lacke, M.C., Mote, T.L., Shepherd, J.M., 2009. Aerosols and associated precipitation patterns in Atlanta. *Atmospheric Environment* 43, 4359–4373.

MACTEC, 2005. Documentation of the Revised 2002 Base Year, Revised 2018, and Initial 2009 Emission Inventories for VISTAS, edited, Visibility Improvement State and Tribal Association of the Southeast (VISTAS).

McCollister, G.M., Wilson, K.R., 1975. Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants. *Atmospheric Environment* 9, 417-423.

McMillan, N., Bortnick, S.M., Irwin, M.E., Berliner, L.M., 2005. A hierarchical Bayesian model to estimate and forecast ozone through space and time. *Atmospheric Environment* 39, 1373-1382.

Michalakes, J., Dudhia, J., Gill, D., Henderson, T., Klemp, J., Skamarock, W., Wang, W., 2005. The Weather Research and Forecast Model: Software Architecture and Performance. *Proceedings of the Eleventh ECMWF Workshop on the Use of High Performance Computing in Meteorology*, Eds. Walter Zwieflhofer and George Mozdzynski, World Scientific, pp. 156 – 168.

Morris, R.E., Koo, B., Guenther, A., Yarwood, G., McNally, D., Tesche, T.W., Tonnesen, G., Boylan, J., Brewer, P., 2006. Model sensitivity evaluation for organic carbon using two multi-pollutant air quality models that simulate regional haze in the southeastern United States. *Atmospheric Environment* 40, 4960-4972.

Mueller, S.F., 2009. Model representation of local air quality characteristics. *Journal of Applied Meteorology and Climatology* 48, 945-961.

Odman, M.T., Hu, Y., Chang, M.E., Russell, A.G., 2007. Forecasting ozone and $PM_{2.5}$ in southeastern U.S. *Developments in Environmental Science* 6, 219-288.

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 national air quality forecasting system. *Weather and Forecasting* 20, 367-384.

Power, H.C., Sheridan, S.C., Senkbeil, J.C., 2006. Synoptic climatological influences on the spatial and temporal variability of aerosols over North America. *International Journal of Climatology* 26, 723-741.

Rainham, D.G.C., Smoyer-Tomic, K.E., Sheridan, S.C., Burnett, R.T., 2005. Synoptic weather patterns and modification of the association between air pollution and human mortality. *International Journal of Environmental Health Research* 15, 347-360.

Robeson, S.M., Steyn, D.G., 1990. Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. *Atmospheric Environment-Part B Urban Atmosphere* 24, 303-312.

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

Ryan, W.F., Piety, C.A., Luebehusen, E.D., 2000. Air quality forecasts in the Mid-Atlantic region: current practice and benchmark skill. *Weather and Forecasting* 15, 46-60.

Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondracek, J., Richter, M., Dostal, M., Bertucco, L., Kolehmainen, M., Doyle, M., 2003. A rigorous inter-comparison of ground-level ozone predictions. *Atmospheric Environment* 37, 3237-3253.

Sheridan, S.C., 2002. The redevelopment of a weather-type classification scheme for North America. *International Journal of Climatology* 22, 51-68.

U.S.EPA, 2003. Guidelines for Developing an Air Quality (Ozone and $PM_{2.5}$) Forecasting Program, edited, USEPA Office of Air Quality Planning and Standards, Information Transfer and Program Integration Division, AIRNow Program, Research Triangle Park, NC, EPA-456/R-03/002, June 2003.

U.S.EPA, 1999. 40 CFR Part 58 Air Quality Index Reporting; Final Rule edited, 42529-42549 Federal Register / Vol. 64, No. 149 / Wednesday, August 1, 1999 / Rules and Regulations.

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 American Meteorological Society* 85, 549-560.