USING NATIONAL AIR QUALITY FORECAST GUIDANCE TO DEVELOP LOCAL AIR QUALITY INDEX FORECASTS

BY BRIAN EDER, DAIWEN KANG, S. TRIVIKRAMA RAO, ROHIT MATHUR, SHAOCAI YU, TANYA OTTE, KEN SCHERE, RICHARD WAYLAND, SCOTT JACKSON, PAULA DAVIDSON, JEFF MCQUEEN, AND GEORGE BRIDGERS

Application of a customized model evaluation protocol—which takes into consideration the perspective and needs of the local forecaster—demonstrates the efficacy of the National Air Quality Forecast Capability.

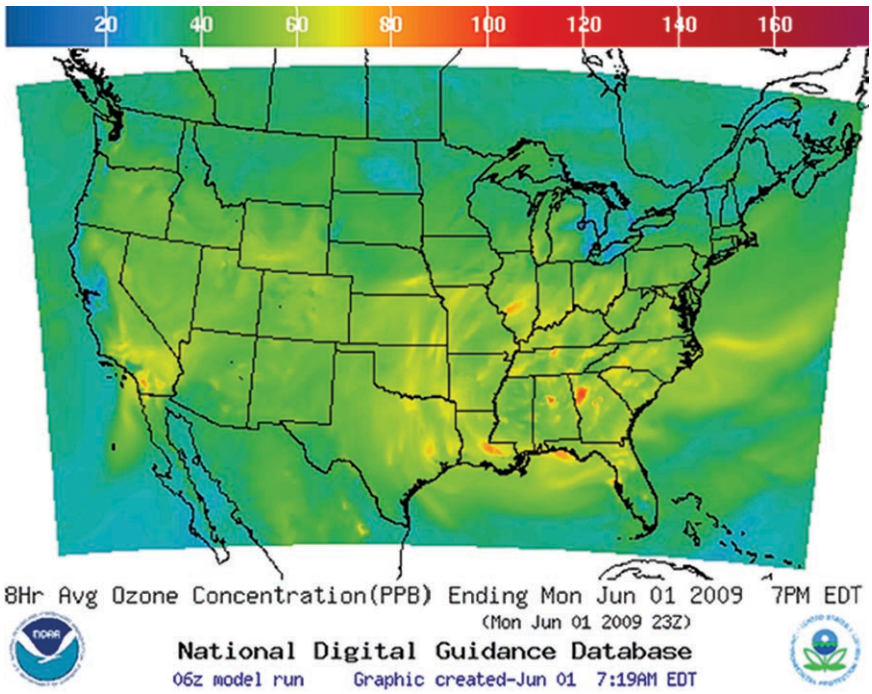


FIG. 1. An example of NOAA's Air Quality Forecast National Digital Guidance Database for the average 8-h O_3 concentrations, depicting the NAM–CMAQ modeling domain (available at www.weather.gov/aq).

Though still relatively new, the science of air quality forecasting has progressed greatly since the 1960s, when basic forecasts of air stagnation or pollution potential were first provided by the Weather Bureau [the predecessor of the National Weather Service (NWS)], which utilized numerical weather prediction models to forecast conditions conducive to poor air quality (Korshover 1976). Societal awareness of the impacts of poor air quality on human health and welfare increased markedly during the 1970s, coinciding with ▶

the formation of the U.S. Environmental Protection Agency (EPA) in 1970 and the promulgation of the Clean Air Act Amendments of 1970 and 1977. During this and the following decade, numerous state and municipal air quality agencies developed basic forecasting approaches in an attempt to reduce exposure, thereby reducing the impact of poor air quality on human health. These forecasts, which generally focused on urban scales, were often based on statistical models that forecast pollutant concentrations based on equations that had been trained or fitted to historical air quality [most often ozone (O_3)] and meteorological data (McCollister and Wilson 1975; Aron and Aron 1978; Wolff and Lioy 1978; Lin 1982; Robeson and Steyn 1990). These approaches worked best under persistent meteorological conditions, but generally performed poorly in forecasting the onset and/or termination of a pollution episode.

The prevalence of these air quality forecasts, as well as their level of sophistication, increased considerably during the 1990s as states and municipalities began establishing Ozone Action Day (OAD) programs (see, e.g., Ryan 1995; Hubbard and Cobourn 1998; Gaza 1998; Davis and Speckman 1999; Ryan et al. 2000). These diverse programs, which are supported by both public and private stakeholders, are designed to educate the public about O_3 and the dangers associated with exposure to elevated concentrations thereof and to inform them when air may be unhealthy, thereby enabling them to take action to limit their exposure. While the criteria used to establish OADs vary with states and municipalities, their intent is the same: to encourage voluntary reductions in O_3 precursor emissions through various approaches, such as offering free public transportation, telecommuting, and liberal leave policies, thereby aiding the states in avoiding exceedances of the National Ambient Air Quality Standards established by the EPA.

As the number of state and local air quality agencies issuing O_3 and/or OAD forecasts increased, so did the need for a centralized, coordinated effort to collect, standardize, and disseminate the forecasts on a national basis. Accordingly, the Office of Air Quality Planning and Standards within the EPA (in conjunction with state, local, and tribal air agencies, the National Park Service, and Environment Canada) developed the AIRNow program in 1998 (www.airnow.gov). AIRNow was developed not only to serve as a web-based clearinghouse designed to facilitate submittal of forecasts by state and local air quality agencies, but also to provide a centralized, nationwide repository for near-real-time O_3 and particulate matter data collected by the various agencies.

The Air Quality Index (AQI), a key component of the AIRNow program, was developed by the EPA in an effort to make these air quality forecasts as easily understood as weather forecasts. The AQI, which focuses on health effects humans may experience within a few hours or days after breathing polluted air, is calculated by the EPA based on the five major air pollutants regulated by the Clean Air Act, including O_3. Initially covering 20 states in the Northeast and Midwest, the AIRNow program currently serves all (or most) of 47 states, and areas of Canada, encompassing over 300 metropolitan areas. Through this expansion and maturation process, it became increasingly obvious that key stakeholders in the AIRNow program, specifically state and local air quality forecasters, would greatly benefit from a federally generated, operational O_3 forecast program (Dabberdt et al. 2004, 2006). Accordingly, Congress mandated, under provisions of the Energy Policy Act of 2002, that the National Oceanic and Atmospheric Administration (NOAA) develop the capability to provide operational forecast guidance for O_3 on a nationwide basis. As a result, NOAA, working in collaboration with the EPA, has developed, tested, and implemented the first stages of the National Air Quality Forecast Capability (NAQFC). Currently, the NAQFC (discussed in more detail in the sidebar) provides next-day forecasts of hourly maximum 1- and 8-h O_3 concentrations (ppb) and smoke concentrations ($\mu g\ m^{-3}$, not discussed) on a 12-km grid covering the contiguous United States. An example of the gridded output, which is disseminated through the NWS's National Digital Guidance Database, is presented in Fig. 1 (see title page).

AFFILIATIONS: Eder, Rao, Mathur, Otte, and Schere—National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina; Kang and Yu—Science and Technology Corporation, Hampton, Virginia; Wayland and Jackson—Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina; Davidson and McQueen—National Oceanic and Atmospheric Administration/National Weather Service, Silver Spring, Maryland; Bridgers—Division of Air Quality, North Carolina Department of Environmental and Natural Resources, Raleigh, North Carolina
CORRESPONDING AUTHOR: Brian Eder, Mail Drop E243-04, U.S. EPA, Research Triangle Park, NC 27711
E-mail: eder.brian@epa.gov

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/2009BAMS2734.1

In final form 22 June 2009
©2010 American Meteorological Society

It needs to be stressed that the NAQFC provides air quality forecast *guidance* to air quality forecasters (analogously to numerical weather prediction models providing guidance to weather forecasters), thereby supplementing, not replacing, the numerous and varied techniques that state and local agencies have employed through the years. Additionally, public dissemination of the NAQFC's forecasts via NOAA's web site (www.weather.gov/aq/) affords populations that are not covered by state and local air quality forecast

Table 1. Air quality index categories with their O_3 concentration breakpoints and precautionary health advice.

AQI	Maximum 8-h concentration (ppb)	Precautionary health advice
Good (0–50)	0–64	No health impacts are expected when air quality is in this range
Moderate (51–100)	65–84	Unusually sensitive people should consider limiting prolonged outdoor exertion
Unhealthy for sensitive groups (101–150)	85–104	The following groups should limit prolonged outdoor exertion: people with lung disease, such as asthma; children and older adults; people who are active outdoors
Unhealthy (151–200)	105–124	The following groups should avoid prolonged outdoor exertion: people with lung disease, such as asthma; children and older adults; people who are active outdoors; and everyone else should limit prolonged outdoor exertion
Very unhealthy (201–300)	> 125	The following groups should avoid all outdoor exertion: people with lung disease, such as asthma; children and older adults; people who are active outdoors; and everyone else should limit outdoor exertion.

THE NATIONAL AIR QUALITY FORECAST CAPABILITY

The NAQFC, developed collaboratively by NOAA and the EPA, couples the North American Mesoscale (NAM) meteorological model [currently the Weather Research and Forecasting Nonhydrostatic Mesoscale Model (WRF–NMM; Janjic 2003)] with the Community Multiscale Air Quality (CMAQ) modeling system (Byun and Schere 2006). The WRF–NMM model generates forecast meteorological fields that are input into CMAQ using a "one way" or "offline" coupling (i.e., there are no feedbacks from CMAQ to WRF–NMM). Software developed by the National Centers for Environmental Prediction (NCEP) is then used to perform bilinear interpolations and nearest-neighbor mappings of the WRF–NMM fields to the CMAQ domain. The preprocessor to CMAQ (PREMAQ) then performs the remaining coordinate transformations, computes additional meteorological fields, and prepares the meteorological dependent emissions (Otte et al. 2005).

The base emissions, which are processed using the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system (Houyoux et al. 2000), are based on the EPA's 2001 National Emissions Inventory (except as noted below) for area source emissions, nonroad emissions, and point source emissions from sources not associated with electric generating units (EGUs). (The base emissions for California and Georgia are augmented with data provided by their state governments.) Additionally, the latest national emission inventories from Canada (1995) and Mexico (1999) are included to complete the continental dataset. For the EGU point sources, continuous emission monitoring (CEM) data from 2005 are updated using projections from the Department of Energy's Annual Energy Outlook released in January 2007. For mobile source emissions, vehicle miles traveled (VMT) data are projected to the forecast year (2007) and are used with forecast year fleet data to initialize EPA's MOBILE6 model (U.S. EPA 2003). Biogenic emissions of VOCs and NO are processed using Biogenic Emission Inventory System (BEIS) version 3.13 (Schwede et al. 2005), which is part of PREMAQ.

The NAQFC provides forecasts of O_3 concentrations (hourly, maximum 1 and 8 h ppb) for the CONUS using 12-km horizontal grid spacing on a Lambert conformal map projection. The CMAQ forecasts use 22 layers extending from the surface to 100 hPa adapted from the hybrid sigma-pressure vertical coordinate used by the WRF–NMM. The layers used in the NAQFC are selected such that CMAQ's layer interfaces are coincident with a subset of the WRF–NMM layer interfaces to minimize the effects of vertical interpolation. The forecast guidance for the next day's surface-layer O_3 is based on the current day's 1200 UTC NAM cycle. The target forecast period is local midnight through local midnight. An additional 7 h are required beyond local midnight to calculate peak 8-h average O_3 concentrations. As a result, a 48-h forecast is needed (based on the 1200 UTC initialization) to obtain the target 24-h forecast period.

agencies with comparable air quality information, allowing them to make informed decisions in order to limit their exposure.

An operational evaluation of the 2007 experimental version of the NAQFC was recently completed, which examined the model's performance during the summer of 2007 over the contiguous United States (CONUS; Eder et al. 2009). The evaluation, which covered the 4-month period (June–September), used strict monitor-to-gridcell matching criteria with O_3 observations obtained from the EPA's Air Quality System (AQS). Results were comparable to, or better than, previous configurations, despite the expansion of the experimental forecast domain into many areas dominated by complex terrain (Eder et al. 2005, 2006). Focusing on standard-type statistics associated with forecast concentrations (i.e., mean and normalized mean bias and error), this evaluation was designed to aid model developers in improving the quality of their forecasts, while supporting, in part, the NAQFC's transition into operational status that occurred on 18 September 2007.

While such an evaluation approach is important, it is equally if not more important to evaluate the NAQFC from the perspective of the local forecaster; that is, with data obtained for specific areas and using the "same circumstances as the model's intended application" as recommended by the 11th Prospectus Development Team of the U.S. Weather Research Program in their report "Meteorological research needs for improved air quality forecasting" (see Dabberdt et al. 2004). Accordingly, the purpose of this article is to demonstrate the efficacy of the NAQFC, from the perspective of local forecasters, thereby encouraging its use by state and local agencies. Such an approach has required the development of a new evaluation protocol: one that examines the utility of the NAQFC in forecasting AQI levels rather than O_3 concentration levels; focuses on the use of categorical-type statistics (i.e., false alarm rates, hit rates, and critical success indices) related to exceedances and nonexceedances that are critical to local forecast agencies; and, most challenging, examines performance over a local forecast region, such as an air shed or metropolitan statistical area (MSA), rather than one based on matched model grid cells and AQS monitors.

After brief discussions summarizing the AIRNow program and its AQI, this paper will demonstrate the development of this evaluation protocol for the Charlotte, North Carolina, MSA for the summer of 2007. In addition to comparing model forecasts with observations, model forecasts will be compared, *for perspective*, with the local forecasts provided by the North Carolina Department of Environment and Natural Resources (NCDENR). This approach will then be applied and likewise compared to local air quality forecasts from four additional MSAs across the nation, including Atlanta, Georgia; Dallas–Fort Worth and Houston, Texas; and Washington, D.C. Each of the local forecasts were prepared and disseminated differently, using a variety of techniques (i.e., statistical models, local deterministic models, etc.), the details of which are beyond the scope of this work.

AIR QUALITY INDEX AND AIRNOW.

The AQI is a dimensionless, six-color-coded index (ranging from 0 to 500) developed by the EPA to provide the public with simple information concerning local air quality and its potential impact on health. It is analogous to the wind chill and heat indices developed by the NWS decades earlier. The AQI focuses on health effects that various subsets of the general population (i.e., the very young or old, asthmatics, etc.) may experience after breathing polluted air. Although EPA calculates the AQI for the five major air pollutants regulated by the Clean Air Act [i.e., O_3, fine particulate matter (PM2.5), carbon monoxide (CO), sulfur dioxide (SO2), and nitrogen dioxide (NO_2)], this demonstration will focus on the AQI based on O_3 because this is the only criteria pollutant currently forecast by the NAQFC. The conversion from O_3 concentration (ppb) to AQI is shown below [Eq. (1)], with precautionary statements associated with each of the five categories provided in Table 1:

$$I_p = \frac{I_{\mathrm{Hi}} - I_{\mathrm{Lo}}}{\mathrm{BP}_{\mathrm{Hi}} - \mathrm{BP}_{\mathrm{Lo}}}(C_p - \mathrm{BP}_{\mathrm{Lo}}) + I_{\mathrm{Lo}}, \qquad (1)$$

where I_p is the index for the pollutant (O_3 in this case), C_p is the rounded concentration of the pollutant, $\mathrm{BP}_{\mathrm{Hi}}$ is the breakpoint that is $\geq C_p$, $\mathrm{BP}_{\mathrm{Lo}}$ is the breakpoint that is $\leq C_p$, I_{Hi} is the AQI value corresponding to $\mathrm{BP}_{\mathrm{Hi}}$, and I_{Lo} is the AQI value corresponding to $\mathrm{BP}_{\mathrm{Lo}}$.

As seen in Table 1, an AQI value of 100 corresponds to the National Ambient Air Quality Standard (NAAQS) for O_3, which prior to 2008 was set at 85 ppb (averaged over eight consecutive hours) by the EPA (Office of the Federal Register 1999). AQI values between 0 and 50 are considered "good" air quality, and values between 51 and 100 indicate "moderate" air quality. Values above 100 are considered to be unhealthy; first, for certain sensitive segments of the population (101–150), and then for the entire population as AQI values exceed 150. It should be noted that the EPA recently strengthened the NAAQS for

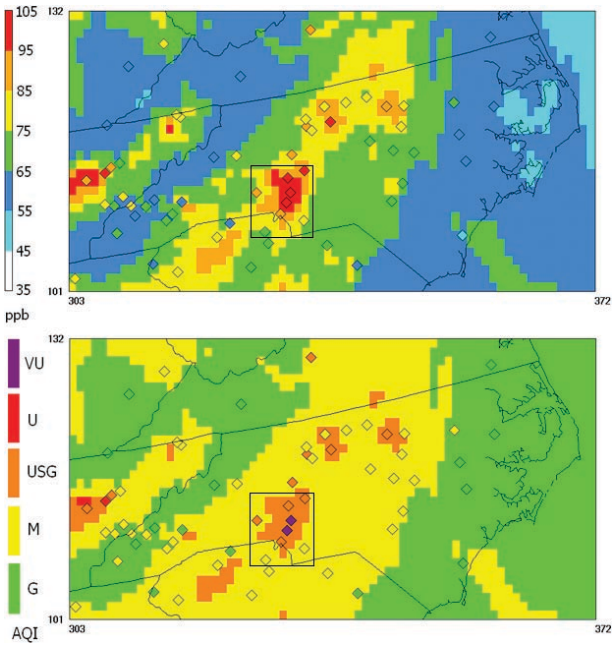


Fig. 2. Modified output of the NAQFC centered on North Carolina for 15 Aug 2007 with AQS observations superimposed as diamonds. (top) Maximum 8-h O_3 concentrations (ppb) and (bottom) corresponding AQI values (VU = very unhealthy, U = unhealthy, USG = unhealthy for sensitive groups, M = moderate, G = good). The approximate Charlotte MSA is outlined.

O_3 and changed the AQI breakpoints to reflect the new primary standard (Office of the Federal Register 2008). Because this demonstration utilizes modeled and observed data from 2007, it will use the older breakpoints; however, impacts of the newer standards on this demonstration will be discussed later.

A LOCAL DEMONSTRATION OF THE NAQFC. In order to demonstrate the efficacy of the NAQFC from the perspective of a state/local forecaster, a case study closely examining the model's performance for the Charlotte MSA during the summer of 2007 is presented. The Charlotte MSA was selected for two reasons. First, the agency responsible for the forecasts—NCDENR—was able to provide forecasts of O_3 concentrations to this research as well as AQI values, thereby facilitating this evaluation. Second, the MSA itself is conveniently defined by eight counties, making it easier to match model output using Federal Information Processing Standards (FIPS) codes.

An example of data obtained from the NAQFC for 1 day (August 15) is provided in Fig. 2 (top panel), providing the forecast maximum 8-h O_3 concentration (ppb) focusing on North Carolina. Note the high concentrations (O_3 > 85 ppb) forecast from sections of upstate South Carolina stretching into the Piedmont of North Carolina, with the highest values in close proximity to Charlotte (denoted by box). Superimposed on this forecast map are observations (diamonds) obtained from the AQS, which indicate that the model performed fairly well on this day with respect to concentration values. While this output is useful in terms of providing general forecast guidance, the forecasters may be equally interested in forecast values of the AQI, rather than concentrations, as these are typically disseminated to the general public. Accordingly, output from the NAQFC can be easily converted to AQI, as shown in Fig. 2 (bottom panel), which presents the same forecast–observation configuration. Note that the NAQFC forecast Code Orange (*unhealthy air quality for sensitive groups*) conditions for the greater Charlotte area, when, in fact, two monitors indicated Code Purple (*very unhealthy air quality*) conditions.

To more closely examine the model's performance, specific to the Charlotte area, we must first identify its corresponding MSA, which is defined by eight counties (seven in North Carolina, one in South Carolina) as shown in Fig. 3 and listed in the appendix. These eight counties are contained within 103 of the NAQFC's 12-km grid cells. Within this MSA there are eight AQS monitors; again, with seven in North Carolina and one in South Carolina. Data from our example day are provided in Table 2, which presents the eight modeled and observed maximum 8-h O_3 concentrations and corresponding AQI pairs and their color codes, as calculated from Eq. (1). Also

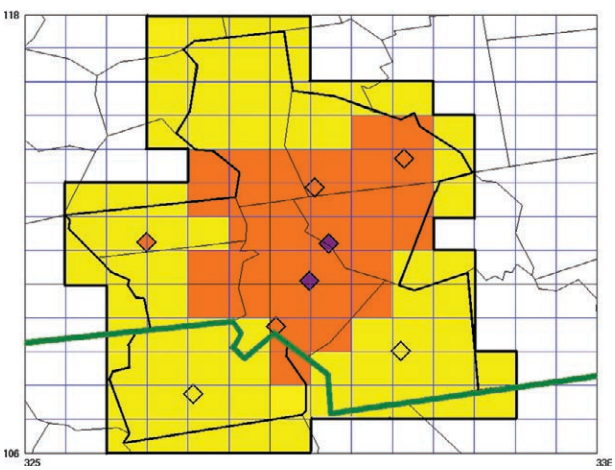
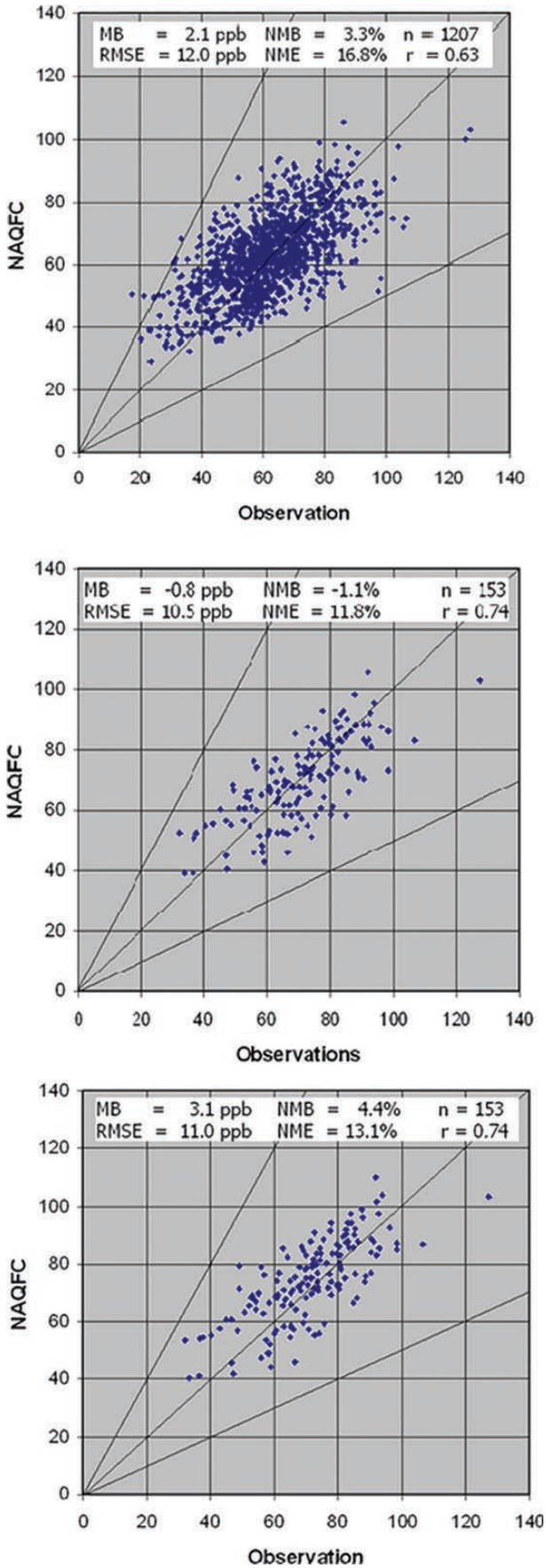


Fig. 3. The Charlotte MSA as defined by eight counties (bold black lines) and 103 NAQFC grid cells (those containing color denoting forecast AQI values) for 15 Aug 2007 with observed AQI values superimposed as diamonds for eight AQS monitors located within the MSA.



included is the maximum concentration (and AQI) of all eight paired monitors and corresponding grid cells, as well as the maximum for all 103 grid cells. (Note that, for this example, the highest value of all 103 grid cells happened to also be one of the eight grid cells that contain a monitor.) This type of conversion and analysis was performed on a daily basis from May through September for the Charlotte MSA, resulting in a possible 1224 (8 monitors × 153 days) observations. Results from this evaluation were then examined using a sequence of approaches as described below.

The first evaluation approach uses *every* single modeled–observed concentration pair, where the observations are matched *directly* with their corresponding grid cells (n = 1207) and corresponding date. This approach is most similar to performance evaluations performed in the past (Eder et al. 2006, 2009) and is by far the most stringent. Examination of Fig. 4 (top panel) reveals that even with this strict approach, the NAQFC performed fairly well for the Charlotte MSA, since the normalized mean error (NME) and normalized mean bias (NMB) were 16.8% and 3.3%, respectively. The correlation coefficient was 0.63 and all except a handful of NAQFC predictions were within a factor of 2 of the actual values. It is apparent from Fig. 4 that the NAQFC tends to underpredict the higher values and overpredict the lower values, continuing a trend seen in earlier model configurations (Eder et al. 2006).

The second approach uses the *maximum* of the eight monitors and the *maximum* of the eight grid cells, *for each day*, where the monitors are not necessarily matched with their corresponding grid cells (n = 153). This somewhat more relaxed evaluation approach is often used by local and state forecasters forecasting for an area like the Charlotte MSA. Such an approach allows for "pattern matching" for one key characteristic (i.e., maximum 8-h O_3 concentration), even when the modeled and observed patterns are not

Fig. 4. Scatterplot and discrete statistics of the maximum 8-h O_3 concentrations (ppb) associated with eight paired AQS observations and corresponding NAQFC grid cells for the summer of 2007 for the Charlotte MSA. (top) All eight observations are matched directly with their grid cells; (middle) only the maximum of the eight AQS observations are matched with maximum from the NAQFC using only the eight grid cells that contain monitors (though not necessarily matched); and (bottom) maximum of the eight AQS observations are matched with maximum from the NAQFC using all 103 grid cells that define the Charlotte MSA. For reference, 1:2 and 2:1 ratio lines are provided.

Table 2. Observed and modeled maximum 8-h O_3 concentrations (ppb) and corresponding AQI values [color code calculated from Eq. (1)] for the eight paired AQS monitors and corresponding NAQFC grid cells for 15 Aug 2007. Also included is the maximum concentration (and AQI) of all eight paired monitors and corresponding grid cells

AQS monitor		37 109 0004	37 119 0041	37 119 1005	37 119 1009	37 159 0021	37 159 0022	37 179 0003	45 109 0006	Maximum
O_3	Model	72.3	103.1	92.1	100.3	87.3	97.6	77.2	71.7	103.1
	Observation	92.8	127.1	88.8	125.5	102.6	103.9	78.8	71.5	127.1
AQI	Model	70	147	119	140	107	134	83	68	147
	Observation	122	202	111	201	146	150	87	67	202

coincident in space and time. Figure 4 (middle panel) reveals that with this approach, the NAQFC appears to have performed "better," as evidenced by the reduction in NME (11.8% compared to 16.8% with the first approach) and NMB (–1.1%, compared to 3.3%). Also note that the correlation has increased from 0.63 to 0.74 and that all of the NAQFC predictions were within a factor of 2 of the actual values (in fact, every prediction but one was within a factor of 1.5).

The third approach may be more applicable to those forecasters that may choose to use the NAQFC output directly. It uses the *maximum* of the eight monitors and the *maximum* of *all* grid cells encompassing the MSA (e.g., 103 grid cells), for each day (again, $n = 153$). Results from this approach are fairly comparable to the second approach (Fig. 4, bottom panel) although values associated with the NMB would always be greater because we use the highest modeled value of 103 grid cells instead of 8.

Each of these three approaches utilizes O_3 concentration data. The final step in this demonstration is to convert the concentrations from the third approach into AQI values (Fig. 5, top panel) and then compare the NAQFC's performance to that of the local forecast provided by NCDENR (Fig. 5, bottom panel). As seen in the scatterplots, which highlight the color codes of the AQI, the NAQFC performed comparably to the local forecast in terms of correlation (0.74 for NAQFC and 0.73 for NCDENR) and error (NME = 25.1% for NAQFC and 21.5% for NCDENR). The model did, however, overpredict (NMB = 9.3%), while the human forecast was basically unbiased (NMB = –0.1%). Note that the human forecast sample size was slightly smaller ($n = 150$), because it was unavailable for 3 days during the period (17 July and 29 and 30 September).

The performance of both the NAQFC and NCDENR forecasts, utilizing this last approach, was examined on a daily basis throughout the summer

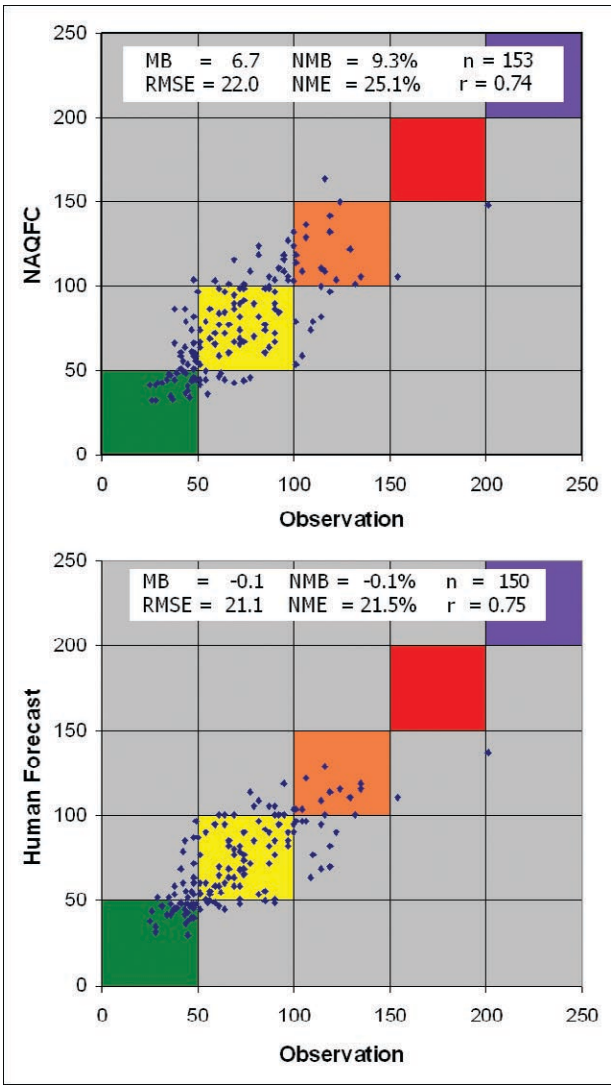


Fig. 5. Scatterplot and discrete statistics of the maximum AQI values associated with eight AQS observations and (top) the maximum AQI values from the 103 NAQFC model grid cells comprising the Charlotte MSA for the summer of 2007, and (bottom) the maximum AQI value forecast by the NCDENR for the Charlotte MSA.

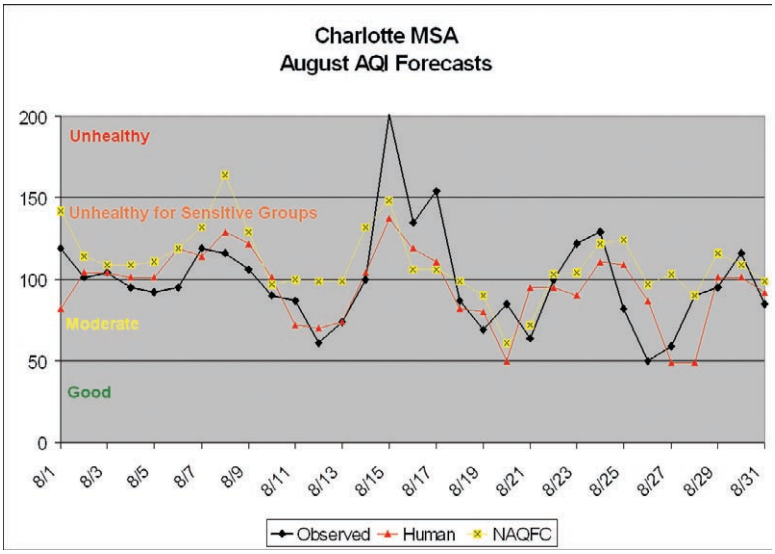


Fig. 6. Daily time series for Aug 2007 of observed and forecast (NCDNER and NAQFC) AQI indices for the Charlotte MSA.

of 2007. Results for 1 month (August) are depicted in Fig. 6. As seen in the Fig. 6, both forecast approaches track the observations quite well (as would be expected given their high correlations), but the NAQFC overpredicted on all but 7 days during the month, while the NCDENR equally over- (15 days) and underpredicted (16 days). Bias adjustment approaches to address such systematic over- or underpredictions associated with the NAQFC have been developed by Kang et al. (2008) but are beyond the scope of this paper.

Categorical analysis. In addition to the discrete statistics discussed above (of both concentration and AQI forecasts), a series of categorical statistics were also calculated to facilitate a comparison of the NAQFC forecast to that of the local human forecast for the Charlotte MSA. These statistics were calculated with the third approach (i.e., using the maximum of the eight monitors and the maximum of all grid cells encompassing the MSA, for each day).

The first categorical metric examined is the category hit rate (cH_i). This metric, which is also known as the probability of detection (POD), is defined as

$$cH_i = \frac{N_f^i}{N_{obs}^i}, \qquad (2)$$

where i is the AQI index (1, 2, 3, 4, 5) category (or the color scheme, i.e., green, yellow, orange, red, purple), and N_f^i is the number of correctly forecast instances in the ith category and N_{obs}^i is the number of observed instances in the ith category. Values of cH_i for both the NAQFC and the NCDENR are presented in Fig. 7. For the 2007 summer period, the NCDENR forecast was more successful at forecasting both code green (62.0%–53.3%) and code yellow days (69.7%–60.5%) but was not as successful as the NAQFC at forecasting code orange days (50.0%–61.5%). Note that there was only one code red day (17 August) and one code purple day (15 August) observed in the Charlotte MSA during the summer, and neither was forecast correctly by either the NAQFC or the NCDENR.

The exceedance hit rate (eH) is defined as follows:

$$eH = \frac{N_{fo}}{N_{fo} + N_o}, \qquad (3)$$

where N_{fo} is the number of both observed and forecast exceedances (AQI ≥ 3) and N_o is the number of observed, but not forecast, exceedances. The eH for the NAQFC (63.0%) was considerable higher than that of the human forecaster (52.0%). Note that these percentages are nearly identical to the code orange cH because of the absence of correctly forecast code red and code purple days.

The next categorical metric examined is the exceedance false alarm rate (eFAR), which is defined as

$$eFAR = \frac{N_f}{N_f + N_{fo}}, \qquad (4)$$

where N_f is the number of forecast but not observed exceedances (AQI ≥ 3) and N_{fo} is defined as above.

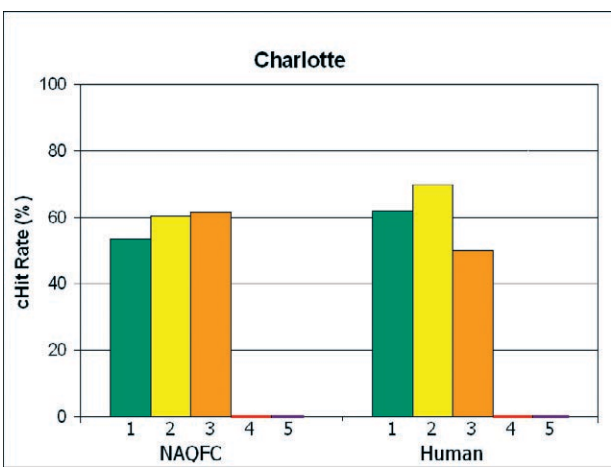


Fig. 7. Categorical hit rate for each AQI category for the Charlotte MSA. Note that neither the NAQFC nor the human forecaster correctly forecast the one code red and the one category purple experienced by the MSA.

The eFAR for NAQFC was 64.1%, while that of the NCDENR forecast was considerably better at 52.8%. It is not surprising that the NAQFC produced higher eH and eFAR, given its general tendency to overpredict and the fact that we include all grid cells (103 in this case) that encompass the Charlotte MSA.

The final categorical metric calculated is the exceedance critical success index (eCSI). It is a more comprehensive index in that it is a function (nonlinear) of both the eH and eFAR. Accordingly, the eCSI is often a better indicator of the "relative worth" of different forecast techniques (Schaefer 1990). It is calculated as

$$\mathrm{eCSI} = \frac{N_{\mathrm{fo}}}{N_{\mathrm{fo}} + N_o + N_f}, \tag{5}$$

with each parameter defined earlier. The eCSI is a fairly stringent index in that it ignores the large number of nonexceedance forecasts that verified correctly. The eCSI for the NAQFC was 39.6%, which was slightly better than the NCDENR of 35.7%.

Application to other MSAs. Having demonstrated the NAQFC's efficacy in producing local AQI forecast guidance for the Charlotte MSA, we now apply the same approach to four other MSAs, namely, Atlanta, Dallas–Fort Worth and Houston (including Galveston and Brazoria), and Washington, D.C., using the four categorical metrics cH, eH, eFAR, and eCSI. (the appendix lists information about these MSAs and their local forecast agencies). Examination of Fig. 8, which presents the cH for each of the MSAs, reveals a variety of outcomes. In three of the MSAs (Washington, Houston, and Atlanta) the human forecast was better than the NAQFC in predicting code green and code yellow days (following the trend from Charlotte), while the NAQFC proved slightly better in the Dallas–Fort Worth MSA. The NAQFC generally outperformed the local forecast in predicting code orange in both Washington and Atlanta and in forecasting code red in Atlanta and Houston for this particular study period. The human forecast was better in the Dallas–Fort Worth area for the higher categories, especially code red days. Note that other than the one code purple day observed in Charlotte, there were no code purple days in the other MSAs.

The NAQFC's predictions of higher-category AQIs is generally better, as reflected in its higher eH rates (Fig. 9, top panel), which exceed 80% in Atlanta, Dallas–Fort Worth, and Washington. When compared to the human forecasts the NAQFC performed better in Atlanta, Charlotte, and Washington, and equally well in Dallas–Fort Worth but worse in Houston. A negative consequence of the NAQFC's ability to capture more of the higher-category AQIs is its propensity to forecast more eFARs than the human forecast. As seen in Fig. 9 (middle panel), the NAQFC forecast resulted in higher false alarm rates than the local forecaster in each MSA. For some of the MSAs (Charlotte, as seen earlier in Dallas–Fort Worth and Houston), the differences were small (< 5%), while in Atlanta and Washington they were large (> 15%).

Finally, the more comprehensive eCSI (which again ignores the large number of nonexceedance forecasts that verified correctly) indicates that the NAQFC performed either slightly better or comparable to the local forecasts in Atlanta, Charlotte (discussed earlier), and Washington, but more poorly in Dallas–Fort Worth and Houston (Fig. 9, bottom panel).

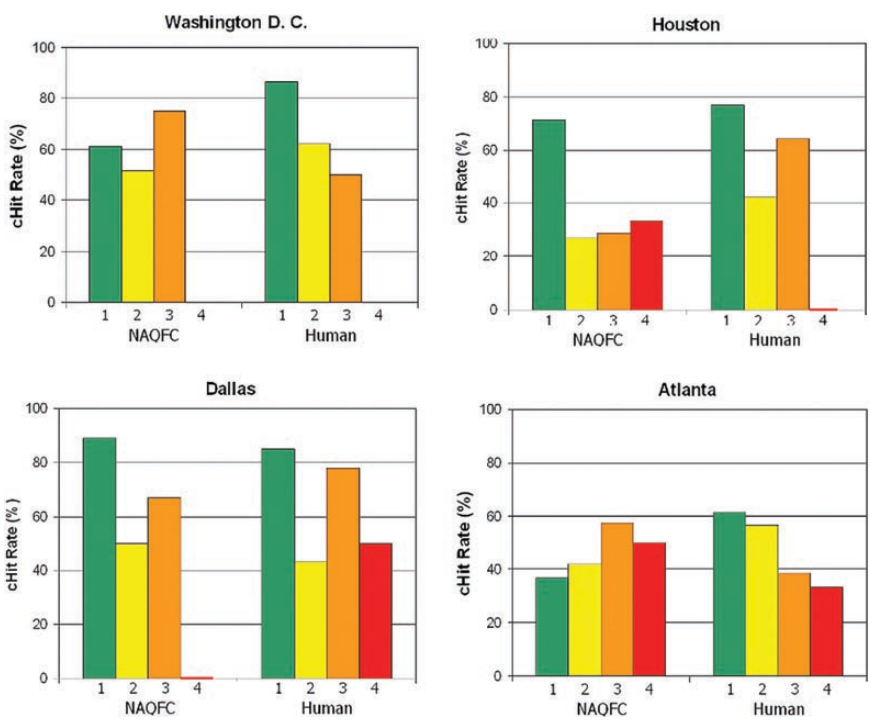


Fig. 8. Categorical hit rates for each AQI category for the Washington, Houston, Atlanta, and Dallas–Fort Worth MSAs.

New AQI breakpoints. As discussed earlier, this demonstration is based on the AQI and its breakpoints established by the EPA in 1997 (Office of the Federal Register 1999) that were in effect through 2007. Following extensive research, including a review of over 1700 studies, the EPA concluded that the 1997 primary standard and its associated AQI did not adequately protect public health with an adequate margin of safety (Office of the Federal Register 2008). Consequently, the EPA recently (12 March 2008) revised the NAAQS for O_3 from 80 (effectively 85 ppb because of rounding) to 75 ppb. As a result, starting with the 2008 O_3 season, the states have been encouraged by the EPA to use the revised AQI with the "100 level" (the upper end of the moderate or acceptable category) based on the new 75 ppb standard, with proportional changes to the other AQI values. The impacts of these changes on the NAQFC's performance are discussed briefly here, where eH, eFAR, and eCSI are recalculated for each MSA based on the new AQI. As seen in Fig. 10, results associated with eH (top panel) are mixed, because the NAQFC eH based on the revised standard was better for Charlotte and Houston, comparable for Atlanta, and somewhat worse for Dallas–Fort Worth and Washington. Results for eFAR (middle panel) and eCSI (bottom panel) are more conclusive in that the NAQFC forecasts produced considerable lower eFAR percentages and higher eCSI percentages based on the revised standard. This "improvement" is not surprising, given that the NAQFC has historically forecast O_3 concentrations in the middle of the distribution with more skill than those concentrations on either end of the distribution (Eder et al. 2006).

These results should be encouraging to state and local forecasters, given that the new standard and hence new AQI effectively increases the frequency of exceedance days and lengthens the O_3 forecast season, thereby further taxing their resources, which in many cases are limited or even decreasing given the recent economic trends.

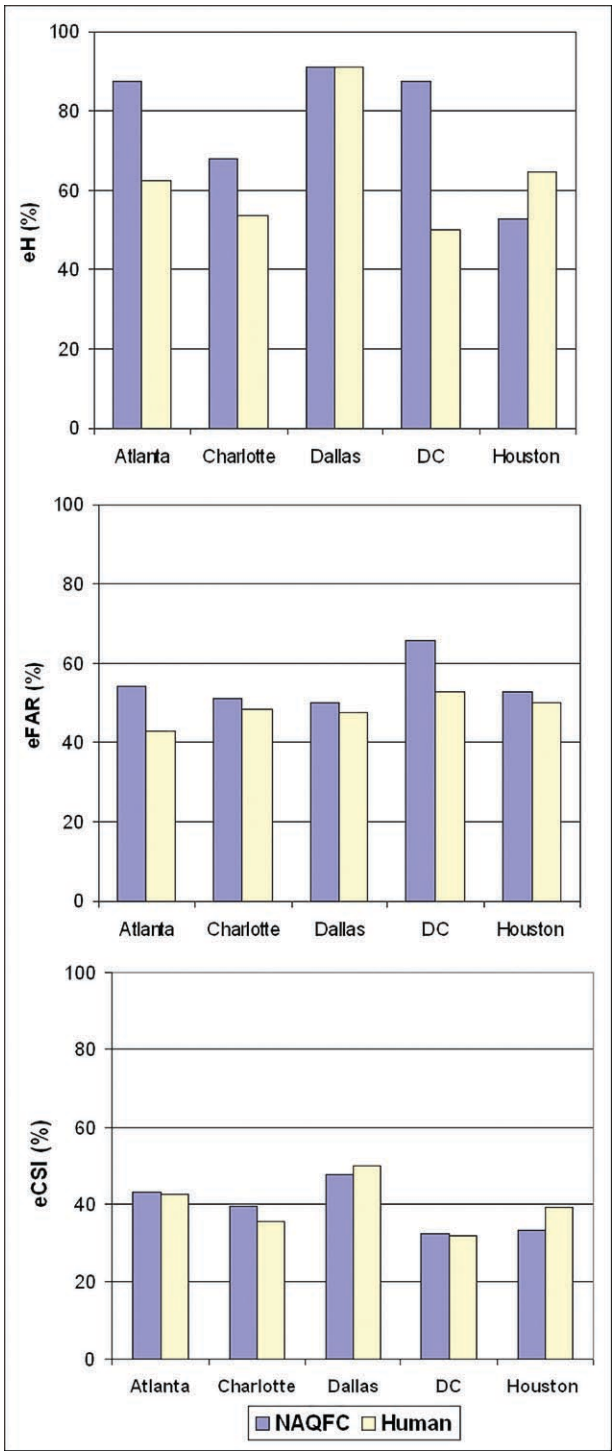


Fig. 9. (top) Exceedance hit rate, (middle) exceedance false alarm rate, and (bottom) exceedance critical success index for each MSA.

SUMMARY AND DISCUSSION. The purpose of this paper has been to demonstrate the efficacy of the NAQFC in providing AQI forecast guidance for O_3 to local forecasters, thereby demonstrating its utility to state and local agencies. Accordingly, this demonstration applied a specially designed evaluation protocol that assessed NAQFC predictions from the perspective of air quality forecasters, rather than the typical evaluation approach designed to assist air quality model developers. As such, this approach focused on AQI categories rather than O_3 concentrations and on categorical-type metrics that are critical to state and local air quality forecasters, who are required to forecast exceedances and nonexceedances of the NAAQS. Finally and most challenging, a protocol was used that examined performance over a

"local forecast region," such as an air shed or MSA, rather than one based on matched model grid cells and AQS monitors.

Results from this approach, which was demonstrated for the Charlotte MSA and subsequently applied to four additional MSAs during the summer of 2007, revealed that the performance of the NAQFC was generally comparable to the that of the local forecast agencies as indicated by the comprehensive critical success indices. The NAQFC's propensity to forecast higher-category AQIs did result in generally better hit rates when compared to the human forecaster, but not without consequences (i.e., higher false alarm rates). The tendency of the NAQFC to overpredict may be ameliorated somewhat, starting with the 2008 O_3 season (as states begin to base their "100 level" of the AQI on the new 75 ppb NAAQS), because the NAQFC has historically performed better when forecasting O_3 concentrations in the middle of the distribution (Eder et al. 2006).

When compared to the evolution of numerical weather prediction's incorporation into weather forecasting, which took decades, the success of this program, which involves numerical (deterministic) O_3 forecasts based on the North American Mesoscale (NAM) and Community Multiscale Air Quality (CMAQ) model, has been realized in a remarkably short time frame. First conceived as part of NOAA's New England Pilot Program in 2002, which served as a "test bed" for O_3 forecasting by including all the elements of a forecast system [i.e., emissions, meteorological and chemical models, and evaluation (Kang et al. 2005)], the NAQFC was initially deployed in 2004. In less than 5 yr, it has expanded in both its scope and coverage so that it now provides timely, accurate O_3 forecasts over the contiguous United States

The results presented here are designed to encourage all O_3 forecasters to utilize the NAQFC as part of their forecast preparation, whether they are "seasoned" veterans with numerous tools at their disposal or forecasters with limited experience and resources. Reliance on the NAQFC may become even more important now that the EPA has tightened the O_3 NAAQS, effectively increasing the number of exceedance days and lengthening the O_3 season, thereby further taxing state and local agencies' budgets that are limited or even decreasing. It is important to remember, however, that the NAQFC is intended to provide guidance to, and certainly not replace, the local forecaster. Just as with numerical weather forecasting, the reliability of a tailored O_3 forecast made at the local level will almost always be better than one derived from a continental-scale model.

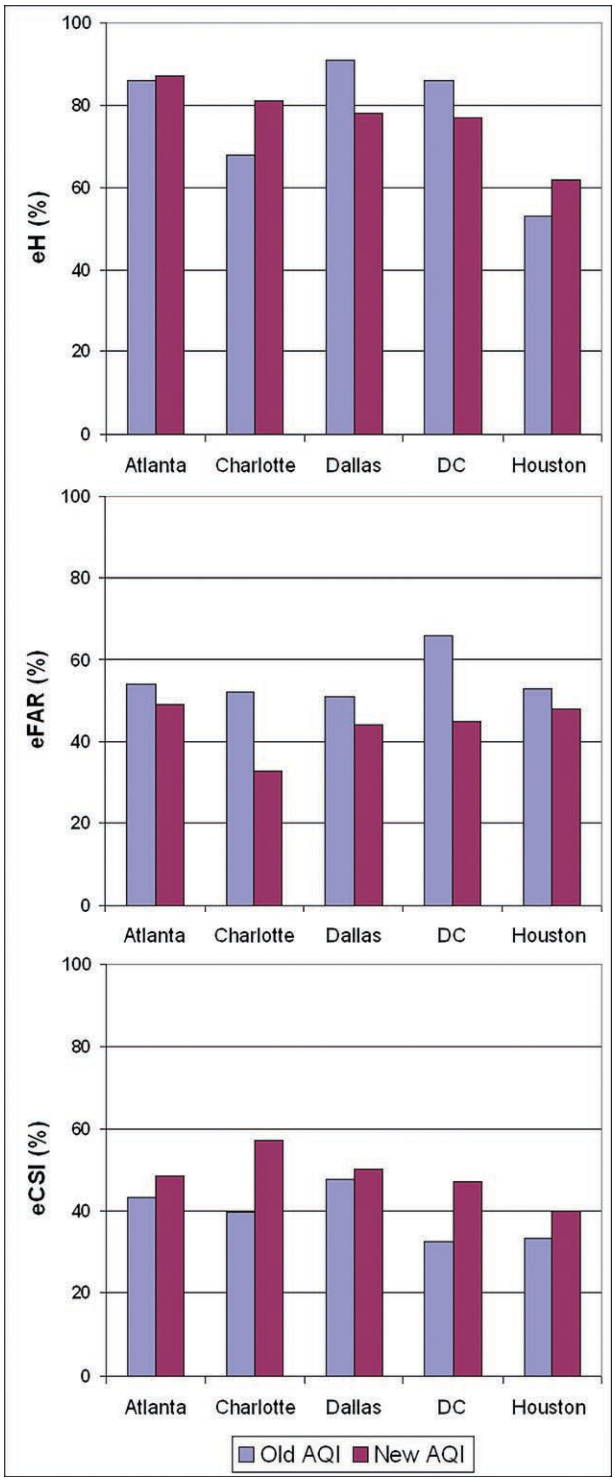


Fig. 10. (top) Exceedance hit rate, (middle) exceedance false alarm rate, and (bottom) exceedance critical success index for each MSA comparing the NAQFC forecasts based on the old and new AQI.

Future work. Continuing development and testing is aimed at expanding the NAQFC to provide O_3 predictions nationwide (including Hawaii and

Alaska) and to add predictions of fine particulate matter ($PM_{2.5}$). Just as with the O_3-based AQI forecasts demonstrated here, an AQI forecast based on $PM_{2.5}$ could also become available. In fact, it may be possible, using the numerous species potentially forecast by NAM–CMAQ, to move toward a more comprehensive AQI, similar to the index recently implemented by Environment Canada (www.ec.gc.ca/cas-aqhi), which they deem to be a personal health protection tool based on the synergistic risk associated with three pollutants: O_3, $PM_{2.5}$, and NO_2 (Stieb et al. 2008). Should the EPA decide in the future to move in that direction, a similar health or exposure index could be developed for the United States using output potentially available from WRF–CMAQ and disseminated through AIRNow.

ACKNOWLEDGMENTS. The authors would like to thank the following agencies for providing forecasted and observed O_3 data to the AIRNow program: the Georgia Department of Natural Resources, the North Carolina Department of Environmental and Natural Resources, the Texas Commission on Environmental Quality, and Washington D.C.'s Metropolitan Washington Council of Governments.

The U.S. Environmental Protection Agency, through its Office of Research and Development, funded and managed the research described here. It has been subjected to the agency's administrative review and approved for publication.

APPENDIX: METROPOLITAN STATISTICAL AREA DESCRIPTIONS.

Metropolitan statistical area, forecasting agency (estimated 1 Jul 2006 population)	Federal Information Processing Standard (FIPS) code	FIPS name
Charlotte, NC, North Carolina Department of Environmental and Natural Resources (1,583,016)	37025	Cabarrus
	37071	Gaston
	37097	Iredell
	37109	Lincoln
	37119	Mecklenburg
	37159	Rowan
	37179	Union
	45091	York (SC)
Atlanta, GA, Georgia Department of Natural Resources (5,138,223)	13247	Rockdale
	13151	Henry
	13113	Fayette
	13121	Fulton
	13063	Clayton
	13067	Cobb
	13057	Cherokee
	13117	Forsyth
	13223	Paulding
	13135	Gwinett
	13089	Dekalb
	13077	Coweta
	13097	Douglas
Dallas–Fort Worth, TX, Texas Commission on Environmental Quality (6,003,967)	48113	Dallas
	48439	Tarrant
	48085	Collin
	48121	Denton
	48397	Rockwall
	48231	Hunt
	48257	Kaufman
	48213	Henderson
	48139	Ellis
	48251	Johnson
	48221	Hood
	48367	Parker

Houston, TX, Texas Commission on Environmental Quality (5,539,949)	48201 48167 48039 48157 48473 48339 48291 48071	Harris Galveston Brazoria Fort Bend Walter Montgomery Liberty Chambers
Washington, D.C., Washington D.C.'s Metropolitan Washington Council of Governments (5,290,400)	11001 51013 51059 51107 51153 51179 51510 51610 51600 51683 51685 24009 24017 24021 24031 24033	DC Arlington (VA) Fairfax (VA) Loudoun (VA) Prince Williams (VA) Strafford (VA) Alexandria (VA) Falls Church (VA) Fairfax(VA) Manassas (VA) Manassas Park (VA) Calvert (MD) Charles(MD) Frederick (MD) Montgomery (MD) Prince George (MD)

REFERENCES

Aron, R., and I. Aron, 1978: Statistical forecasting models: Carbon monoxide concentrations in the Los Angeles basin. *J. Air Pollut. Control Assoc.*, **28**, 681–684.

Byun, D., and K. Schere, 2006: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality modeling system. *Appl. Mech. Rev.*, **59**, 51–77.

Dabberdt, W., and Coauthors, 2004: Meteorological research needs for improved air quality forecasting: Report of the 11th Prospectus Development Team of the U.S. Weather Research Program. *Bull. Amer. Meteor. Soc.*, **85**, 563–586.

——, and Coauthors, 2006: USWRP workshop on air quality forecasting. *Bull. Amer. Meteor. Soc.*, **87**, 215–221.

Davis, J., and P. Speckman, 1999: A model for predicting maximum and 8 h average ozone in Houston. *Atmos. Environ.*, **33**, 2487–2500.

Eder, B., D. Kang, R. Mathur, S. Yu, and K. Schere, 2005: An evaluation of the Eta-CMAQ air quality forecast model for the summer of 2005. *AWMA Specialty Conference on Environmental Data Analysis*, Chicago, IL, Air and Waste Management Association.

——, ——, ——, ——, and ——, 2006: An operational evaluation of the Eta-CMAQ air quality forecast model. *Atmos. Environ.*, **40**, 4894–4905.

——, ——, ——, J. Pleim, S. Yu, T. Otte, and G. Pouliot, 2009: A performance evaluation of the National Air Quality Forecast capability for the summer of 2007. *Atmos. Environ.*, **43**, 2312–2320.

Gaza, R., 1998: Mesoscale meteorology and high ozone in the northeast United States. *J. Appl. Meteor.*, **37**, 961–967.

Houyoux, M., J. Vukovich, C. Coats Jr., N. Wheeler, and P. Kasibhatla, 2000: Emission inventory development and processing for the seasonal model for regional air quality (SMRAQ) project. *J. Geophys. Res.*, **105**, 9079–9090.

Hubbard, M., and W. Cobourn, 1998: Development of a regression model to forecast ground-level ozone concentrations in Jefferson County, Kentucky. *Atmos. Environ.*, **32**, 2637–2647.

Janjic, Z., 2003: A nonhydrostatic model based on a new approach. *Meteor. Atmos. Phys.*, **82**, 271–285.

Kang, D., B. Eder, A. Stein, G. Grell, S. Peckham, and J. McHenry, 2005: The New England Air Quality Forecasting Pilot Program: Development of an evaluation protocol and performance benchmark. *J. Air Waste Manage. Assoc.*, **55**, 1782–1796.

——, R. Mathur, S. T. Rao, and S. Yu, 2008: Bias adjustment techniques for improving ozone air quality forecasts. *J. Geophys. Res.*, **113**, D23308, doi:10.1029/2008JD010151.

Korshover, J., 1976: Climatology of stagnating anticyclones east of the Rocky Mountains, 1936–1975. U.S. Department of Commerce Tech. Rep. ARL-55, 26 pp.

Lin, Y., 1982: Oxidant prediction by discriminant analysis in the South coast air basin of California. *Atmos. Environ.*, **16**, 135–143.

McCollister, G., and K. Wilson, 1975: Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants. *Atmos. Environ.,* ***9***, 417–423.

Office of the Federal Register, 1999: Air quality index reporting. Final Rule, 40 CFR Part 58, 4 August 1999, *Federal Register,* Vol. 64, No. 149, 42 529–42 549.

——, 2008: Air quality index reporting. Final Rule, 40 CFR Parts 50 and 58, 27 March 2008, *Federal Register*, Vol. 73, No. 60, 16 436–16 514.

Otte, T., and Coauthors, 2005: Linking the Eta model with the Community Multi-scale Air Quality (CMAQ) modeling system to build a national air quality forecasting system. *Wea. Forecasting,* **20**, 367–384.

Robeson, S., and D. Steyn, 1990: Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. *Atmos. Environ.,* **24B**, 303–312.

Ryan, W., 1995: Forecasting severe ozone episodes in the Baltimore metropolitan area. *Atmos. Environ.,* **29**, 2387–2398.

——, R. Piety, and E. Luebehusen, 2000: Air quality forecasts in the Mid-Atlantic region: Current practice and benchmark skill. *Wea. Forecasting*, **15**, 46–60.

Schaefer, J., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575.

Schwede, D., G. Pouliot, and T. Pierce, 2005: Changes to the Biogenic Emissions Inventory System, Version 3 (BEIS3). *Extended Abstract, Fourth CMAS Models-3 Users' Conf.*, Chapel Hill, NC, Community Modeling and Analysis System. [Available online at www.cmascenter.org/conference/2005/abstracts/2_7.pdf.]

Stieb, D., R. Burnett, M. Smith-Doiron, O. Brion, H. Shin, and V. Economou, 2008: A new multipollutant, no-threshold air quality health index based on short-term associations observed in daily time-series analysis. *J. Air Waste Manage. Assoc.*, **58**, 435–450.

U.S. EPA, 2003: User's guide to MOBILE6.1 and MOBILE6.2. Publication EPA420-R-03-010, 262 pp.

Wolff, G., and P. Lioy, 1978: An empirical model for forecasting maximum daily ozone levels in the northeastern U.S. *J. Air Pollut. Control Assoc.*, **28**, 1035–1038.