A REAL-TIME EULERIAN PHOTOCHEMICAL MODEL FORECAST SYSTEM

Overview and Initial Ozone Forecast Performance in the Northeast U.S. Corridor

BY JOHN N. MCHENRY, WILLIAM F. RYAN, NELSON L. SEAMAN, CARLIE J.COATS JR., JANUSZ PUDYKIEWICZ, SARAV ARUNACHALAM, AND JEFFERY M. VUKOVICH

For the first time, a real-time Eulerian photochemical modeling system is used to successfully forecast ozone in the northeast United States.

Recently, there has been much interest in the development of chemical weather forecasting capabilities. For example, the U.S. Weather Research Program has recommended the adoption of a strategy that would lead toward such a goal (Dabberdt et al. 2000), and the National Science Foundation had earlier hosted a town meeting on the subject (McHenry 1999). This followed a report by the Board on Atmospheric Sciences and Climate of the National Research Council that concluded that the discipline of *routine* operational forecasting is a "necessary next step" in order to advance the state of science in environmental modeling (National Research Council Board on Atmospheric Sciences and Climate 1998). A first large step in this process is presented in this paper. A numerical forecast model is used to forecast tropospheric O_3 in real time, and the operational utility of this model is compared to current forecast practice.

Tropospheric O_3 has been recognized as a harmful pollutant for many years (Haagen-Smit et al. 1951; Lippman 1989; Heck et al. 1982; Bascom et al. 1996a,b)

AFFILIATIONS: MCHENRY, COATS, ARUNACHALAM, AND VUKOVICH—MCNC Environmental Modeling Center, North Carolina Supercomputing Center, Research Triangle Park, North Carolina; RYAN AND SEAMAN—Department of Meteorology, The Pennsylvania State University, State College, Pennsylvania; PUDYKIEWICZ—Meteorological Service of Canada, Dorval, Quebec, Canada
ADDITIONAL AFFILIATIONS: MCHENRY, COATS, AND VUKOVICH—Baron Advanced Meteorological Systems, Research Triangle Park, North Carolina; SEAMAN—NOAA/NWS/Office of Science and Technology, Suitland, Maryland; ARUNACHALAM—Carolina Environmental Program, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina
CORRESPONDING AUTHOR: John N. McHenry, Chief Scientist, Baron Advanced Meteorological Systems, North Carolina Supercomputing Center, 3021 Cornwallis Rd., Research Triangle Park, NC 27709-2889
E-mail: john.mchenry@baronams.com
DOI: 10.1175/BAMS-85-4-525
In final form 10 November 2003

and, since 1970, has been designated a "criteria pollutant" for which health standards are in place (Federal Register 1971; U.S. EPA 1986; originally a 1-h-average concentration of 125 ppbv). More stringent O_3 standards, based on an 8-h-average concentration of 85 ppbv, were promulgated in 1997 (Federal Register 1997). With increased awareness of the health effects of O_3, there has been an increase in the number of localities issuing air quality forecasts. Utilizing state- or local-agency-based forecast professionals (e.g., http://daq.state.nc.us/airaware/ozone/), over 300 cities nationwide presently issue daily air quality forecasts and public health warnings during the summer season (AIRNow 2003).

Unhealthy levels of O_3 in the eastern United States are a warm-season, episodic phenomenon. Regionwide episodic pollution control programs could be effective in reducing the extent of unhealthy O_3 concentrations, and many metropolitan areas have instituted "Ozone Action Day" (OAD) programs in addition to their public health warnings. OAD programs, sponsored by a public–private partnership of government agencies and large employers, support a variety of actions, such as free public transportation or liberal leave policies, that are taken when pollution episodes are expected to occur (AIRNow 2003). In order for health warnings and OAD programs to be effective, timely and accurate forecasts are required to be skillful on the metropolitan scale and to be issued to the public with a 24–36-h lead time.

The processes by which O_3 is formed in the troposphere are complex and challenging to forecast. Photochemical O_3 is a secondary pollutant. It is not emitted directly to the atmosphere but is formed via a complex set of reactions involving volatile organic compounds (VOCs), oxides of nitrogen (NO_x; or the sum of NO + NO_2), and ultraviolet (UV) radiation (Crutzen 1979). The reactions that form O_3 constitute a nonlinear system, with the rate of O_3 production dependent on the relative concentrations of its precursors (Liu et al. 1987), whose tropospheric lifetimes vary by many orders of magnitude. Given that the average residence time of ozone is on the order of a month or more (Hobbs 2000), it is not surprising that ozone varies on seasonal, synoptic, diurnal, and subdiurnal time scales (Rao et al. 1997; Hogrefe et al. 2001).

Until recently, numerical air quality models were not able, within the time constraints imposed by operational forecasting, to provide forecasts on a routine basis. While a few attempts have been made to use standard regulatory models to predict single-metropolitan-domain O_3 in near–real time with limited success (e.g., Chang and Cardelino 2000), most such models, which have been in use for decades, are almost exclusively relied upon in the United States to examine the effects of proposed pollution control strategies using historical case studies (Seinfeld 1988; Russell and Dennis 2000). In the mid- to late 1990s, efforts were begun in Canada (Pudykiewicz et al. 1997, 2003), Australia (Manins 2001), and Europe (van Aalst and de Leeuw 1997; Jakobs et al. 2001) to implement regional numerical air quality prediction (NAQP) capabilities in those locations. However, none to date in the United States has been successfully deployed as a forecast model with the regional or mesoscale approach reported here.

The tasks required for a reliable, accurate NAQP system are formidable. The NAQP system must include an accurate mesoscale meteorological model providing skillful forecasts of temperature and mixing ratio (chemical reactions and emissions rates are sensitive to both), incoming solar radiation (clouds), boundary layer depth and stability, and wind speed/direction. Wind fields pose a special problem in the eastern United States, where embayments and irregular coastlines are located near high-emission sources. Further, the meteorological model must also be coupled to an accurate emissions model using up-to-date emissions inventories, and both models must provide data to the photochemical model, which itself must execute in time to deliver forecast products for operational use, nominally in the early afternoon with a 36-h computational lead time. This is needed for a single-day forecast since most operational ozone forecasts are now based on the new 8-h-average concentration using forward averaging to define the 8-h period. Moreover, chemical boundary and initial conditions pose a challenge, as recent studies have shown that transported O_3 and precursors are often key components of local O_3 (Ryan et al. 1998; Dickerson et al. 1995; Knapp et al. 1998).

As a result of the difficulties in numerical modeling approaches, forecasts to support OAD programs have relied on alternative approaches. Although O_3 formation, when considered explicitly, is extremely complex, it is also true that maximum O_3 concentrations are well correlated to weather conditions (Clark and Karl 1982; Wolff and Lioy 1978; Robeson and Steyn 1990). For example, a significant amount of variation in metropolitan-scale peak O_3 can be explained by a small subset of meteorological predictors. For the Philadelphia (PHL), Pennsylvania, metropolitan area, maximum surface temperature explains 62% of the variance in summer season peak O_3 (Ryan 2002a). The association of peak O_3 with meteorological predictors lends itself to statistical forecast

schemes. Statistical approaches include standard multiple regression (Ryan et al. 2000), nonlinear regression (Hubbard and Cobourn 1997), neural networks (Ruiz-Suarez and Mayora-Ibarra 1995), classification and regression tree schemes (CART; Burrows et al. 1995), and hybrid approaches (Liu and Johnson 2002). Comparative studies have shown consistent results between approaches and adequate skill overall (Cobourn et al. 2002; Comrie 1997). In the PHL metropolitan area for the 1997–2002 summer seasons, the mean absolute error (MAE) of statistical forecast guidance for peak metropolitan scale O_3, based on multiple linear regression forecast guidance, was 13.4 ppbv with an rmse of 17.0 ppbv (Ryan 2002a).

While overall skill is adequate for statistical models, there are systemic shortcomings that limit their usefulness. Because forecast predictors are confined to meteorological variables, the underlying physical processes, based on the complex interplay of precursor emissions and chemistry, are not fully considered. While the use of O_3 precursor concentrations as predictors is possible (e.g., Liu and Johnson 2002), forecasting these variables at the time scales of interest is not possible in many locations because of limited observations and large local-scale variations in concentrations. Constraints on the skill of commonly used meteorological predictors also reduces forecast accuracy. While the typical explained variance for regression algorithms applied to historical data is in the range of 75%–80%, in operational practice it is usually on the order of 65%–75% (Ryan 2002a). The strong temperature–O_3 relationship, while allowing for reasonable skill overall, is a limiting factor in the critical high end of the O_3 distribution. This is because, in general, warm temperatures are necessary but not sufficient for high O_3. For the PHL forecasts over the period 1996–2002, 78% of the OAD forecasts verified for peak $O_3 \geq 125$ ppbv averaged over an hour at a given surface monitor (denoted a "1-h exceedance"), but only 52% of all observed 1-h exceedances carried such a forecast.

The NAQP system described in this paper overcomes the difficulties of the current statistical/heuristic approaches practiced by state or local agencies issuing operational ozone forecasts. Further, by evaluating its use in the northeastern United States during a high regional-O_3 episode that occurred in early August of 2001, this paper shows that the present system has skill sufficient to match or improve upon these "tried and true" methods. First, the model components, their configurations, and the software engineering advancements developed to allow operational use are described. Next, the overall performance of the photochemical model in the northeast United States is reported in the context of the meteorological evolution of the August 2001 episode. Third, the skill of the forecast model is compared to an operational statistical-model-based forecast in PHL; and fourth, results are contrasted against two different operational forecast methods available in the New England states during the episode, using persistence as a baseline measure of skill. Finally, the concluding section provides a summary and discusses the NAQP system in relation to the emerging U.S. national air quality forecasting capability within the National Oceanic and Atmospheric Administration (NOAA 2003a).

NAQP SYSTEM OVERVIEW. The NAQP system consists of a set of three one-way coupled models that run routinely on a parallel microprocessor supercomputer (McHenry et al. 2000, 2001). The component models are the fifth-generation Pennsylvania State University (PSU)–National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5), version 3.4 (Grellet al. 1994), the Sparse-Matrix Operator Kernel for Emissions (SMOKE; Coats 1996; Houyoux et al. 2000) model, and the Multiscale Air Quality Simulation Platform—Real Time (MAQSIP-RT; McHenry et al. 1999) photochemical model. Though the system has been run in real time since the summer of 1998 and includes ongoing improvements (McHenry and Coats 2003), the components described compose the version in operation under NOAA's "early start" chemical weather forecasting initiative begun during the summer of 2001 (NOAA 2003a).

Meteorological model. The prognostic meteorological model, MM5, was configured using 31 vertical sigma layers with its top at 100 hPa. The Kain–Fritsch (KF) deep convection scheme (Kain and Fritsch 1993) and a simple water–ice explicit moisture scheme (Dudhia 1989) were used to represent clouds, while the five-layer soil model represented the land surface (Dudhia 1996). The Medium-Range Forecast (MRF) scheme (Hong and Pan 1996) was used to represent PBL processes. MM5 has shown good skill in weather conditions conducive to high O_3 in the northeastern United States (Seaman and Michelson 2000), and its performance will not be formally evaluated here.

Emissions model. SMOKE is designed to take advantage of the linear-operator transform nature of the computational operations common to emissions modeling by using sparse matrices. This enables SMOKE to be two to three orders of magnitude faster

than traditional emissions modeling approaches while at the same time maintaining the full integrity of the emissions inputs. In 2001, these inputs included foundational databases for point and area sources (U.S. EPA 1999; Steiner et al. 1994; Houyoux et al. 2000), mobile-source vehicle miles traveled (Houyoux et al. 2000), and biogenic sources. The submodels within SMOKE include a model that computes point-source plume rise for each point-source stack in the inventory, a mobile-source model that determines appropriate vehicle emission factors (U.S. EPA 2003a), and the Biogenic Emissions Inventory System version 2 (BEIS2; U.S. EPA 1995), which utilized land-use data generated from the Biogenic Emissions Landcover Database version 3 (BELD3; Pierce et al. 1998) to estimate biogenic VOC and NO emissions. All of these submodel calculations are sensitive to meteorology, and thus prognostic MM5 data are provided in order to produce real-time emissions *forecasts.* Because emissions are of first-order importance in achieving accurate air quality simulations, significant effort was devoted to ensuring that all of the foundational emissions databases were the most recent and reliable available. This required a postprocessor that combines the outputs from several SMOKE executions since several different inventories were involved for point, area, and mobile sources. SMOKE and its postprocessor are pictured schematically in Fig. 1.

Photochemical model. MAQSIP-RT is the operational prognostic version of the more general MAQSIP photochemical model. MAQSIP (Odman and Ingram 1996; Kasibhatla and Chameides 2000; Hogrefe et al. 2001) was originally designed and developed as the prototype (Coats et al. 1995) for the Environmental Protection Agency (EPA) Community Multiscale Air Quality model (CMAQ; Byun and Ching 1999), and SMOKE was developed under a companion effort. In 2001, MAQSIP-RT differed from MAQSIP primarily in that dry deposition velocities were calculated online (following Wesley 1989), a relaxation stratospheric top boundary condition for ozone was implemented, a flexible one-way grid nesting scheme was available, and optimization for efficient execution on microprocessor-based parallel computer systems had been developed. The dry deposition scheme was enhanced to allow mapping from the most recent U.S. Geological Survey (USGS) 24-category land-use data available in MM5 to its 11-category Regional Acid Deposition Model (RADM) scheme (Chang et al. 1987).

Common to both MAQSIP and MAQSIP-RT a modified Carbon-Bond 4 (CBM-4) chemistry mechanism was employed, representing about 150 gas-phase reactions and 35 chemical species (Gery et al. 1989). This version includes updated kinetic data for the CO+OH reaction (DeMore et al. 1994) and for peroxyacetyl nitrate (PAN) chemistry (Chang et al. 1996), an updated condensed isoprene chemical mechanism based on Carter (1996), and modifications to the chemical pathways of the universal peroxy radical operators (XO2 and XO2N) used in CBM-4. Kasibhatla et al. (1997) provide a detailed analysis of these modifications. Scalar advection was handled with a positive definite advection scheme (Bott 1989). Since mass inconsistencies may arise when advection schemes differ between the meteorological and photochemical models, or when the photochemical model interpolates meteorological wind fields (both an issue here), a normalization step was used whereby the concentrations computed at the end of the model

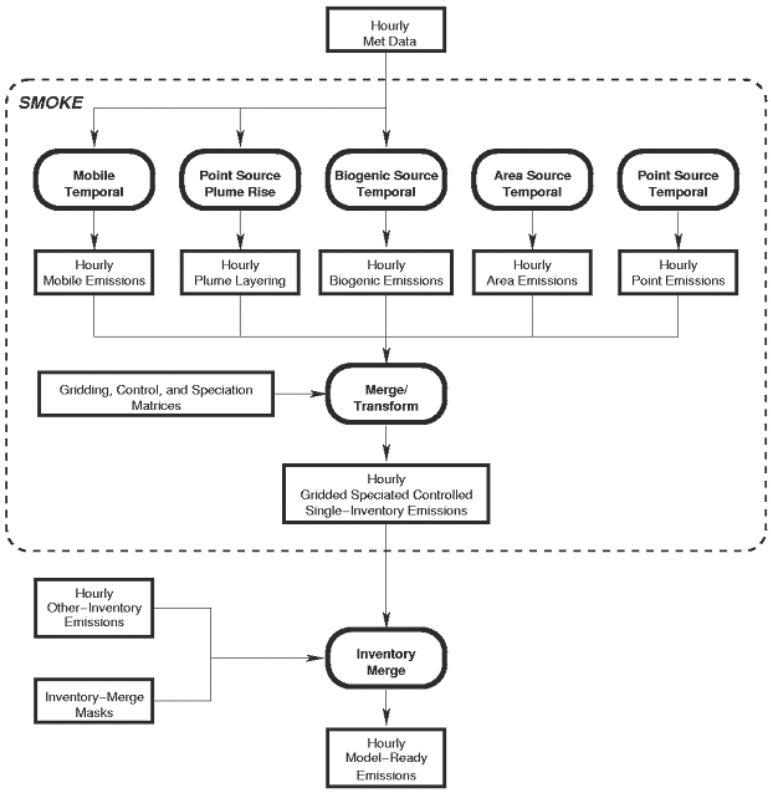


FIG. 1. SMOKE emissions model data flow diagram detailing processing for (dotted box) one complete inventory and (left, below dotted box) showing merge postprocessor.

advection step are adjusted based on the advection of a tracer of uniform density. Advantages and disadvantages of this approach are discussed in Odman and Russell (2000).

Clear- and cloudy-sky photolysis rate calculations were loosely based on Chang et al. (1987). Clear-sky actinic fluxes are precalculated using the delta-Eddington radiative transfer model described in Madronich (1987) and are adjusted in the presence of prognostic clouds using MM5 forecast deep convection and grid-scale nonconvective clouds. Cloud processes are treated using a chemical transport/scavenging version of the KF scheme that replicates each deep convective cloud in MM5 (McHenry et al. 1996), a shallow convection mixing/venting scheme (McHenry and Binkowski 1996), and a grid-scale explicit moisture scheme. All of these submodels compute aqueous-phase dissolution/dissociation along with wet deposition in the case of precipitating clouds. Aqueous kinetic reactions (pertinent to acid rain formation but thought to be of secondary importance for ozone photochemistry) are not included for efficiency. Vertical turbulent mixing was determined using K theory following the approach of Chang et al. (1987) in which formulations of the eddy diffusivities differ depending on the stability class of the boundary layer. MM5-prognosed PBL depths are used to identify the bottom of the free troposphere. Within the free troposphere, turbulent mixing follows Blackadar (1976). MAQSIP-RT is deployed with the identical vertical coordinate and layer structure as MM5.

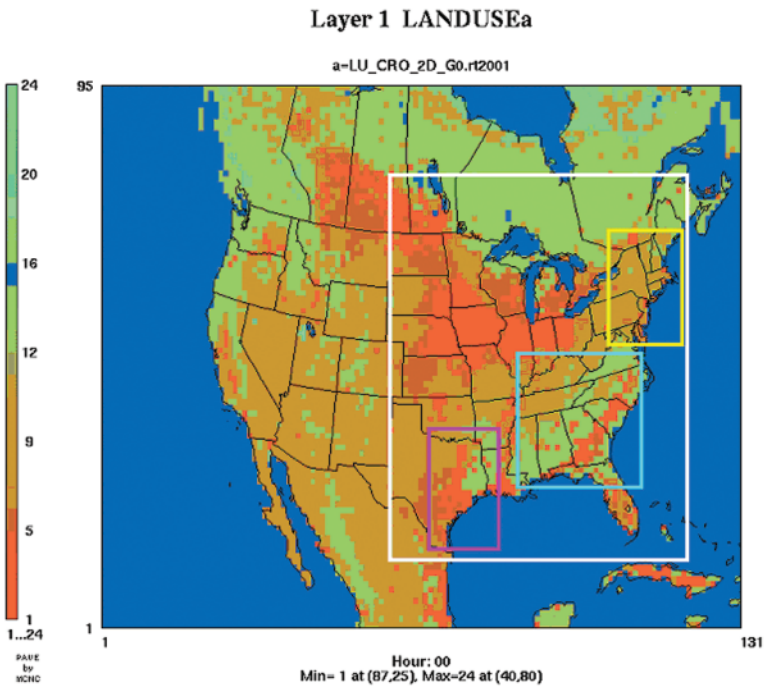


Fig. 2. MAQSIP-RT and SMOKE 45- and 15-km domains in reference to MM5 45-km domain. USGS land-cover categories are shown. The northeast U.S. 15-km domain is outlined in yellow. The MM5 15- and 5-km domains are not shown.

Model linkage. The NAQP system is linked with a software module, MM5 Coupler (MCPL), that drops directly into MM5, computes all the meteorological variables needed by SMOKE and MAQSIP-RT, windows them to the air quality modeling domains, and writes them to files (Coats et al. 1998). These files are readable via either disk-based or parallel virtual machine (PVM; Oak Ridge National Laboratories 2003) implementations of an input/output applications programming interface (Coats 2003) that overlays the network Common Data Form (netCDF; Rew and Davis 1990). MCPL is both highly configurable and efficient, consuming just 0.1%–0.2% of the MM5 run time.

System execution. MM5 is run as a unified, triply nested (45-, 15-, and 5-km grid spacing) multidomain model, whereas both SMOKE and MAQSIP-RT are run as telescoped single-domain executions, also at 45-, 15-, and 5-km grid spacings (Fig. 2). The system is run twice daily, at 0000 and 1200 UTC, although the 5-km domains run only at 0000 UTC because of time constraints. The most recent MAQSIP-RT run, typically 12 h old, is used to provide ("cycled") initial conditions for the current run, because of the lack of any routine real-time chemical analyses of the atmosphere. Kalnay et al. (1998) provide a good discussion of the rationale behind the growing need for such analyses. MM5, however, is reinitialized every 12 h with new analysis data derived from the Eta Data Assimilation System (EDAS) running at the National Centers for Environmental Prediction (NCEP).

Coarser MAQSIP-RT domains provide boundary conditions to finer domains, except on the outermost domain, where monthly ozone climatology is used (Logan 1999). The 100-hPa top boundary also uses this climatology. The SMOKE/MAQSIP-RT domains are currently designed as "windows" into the corresponding MM5 domains, since atmospheric chemistry is several times as expensive to compute as meteorology per unit area. The boundary of the 45-km domain is placed well away from major emissions re-

gions to ensure that polluted transport into the forecast region is of minimal influence on the forecast. The northeast 15-km domain is shown in yellow in Fig. 2, and the 1200 UTC MAQSIP-RT forecasts from this domain, valid the following day (24–36-h forecast), are evaluated in this paper.

THE 1–10 AUGUST 2001 NORTHEAST U.S. O_3 EPISODE: METEOROLOGICAL DESCRIPTION AND OVERALL MAQSIP-RT 1-H-AVERAGE PERFORMANCE IN THE NORTHEAST 15-KM DOMAIN. In the eastern United States, O_3-conducive weather conditions are often associated with an upper-level ridge with its main axis just west of the region of interest (Vukovich et al. 1977; Vukovich and Fishman 1986; Ryan et al. 1998). The region downstream (east) of the ridge axis is associated with large-scale subsidence that enhances photochemical activity by suppressing clouds and vertical mixing. Surface high pressure is located east of the upper-level ridge axis, and weak pressure gradients result in light surface winds. The ridge orientation also induces sustained downsloping winds along the eastern slopes of the Appalachians, causing adiabatic warming in the boundary layer and placing the NE United States downwind of concentrated NO_x sources in the industrialized Midwest and Ohio River valley. Ryan et al. (1998) describe this long-range transport phenomenon.

The O_3 episode of early August 2001 resembled the scenario described above on the synoptic scale but differed in mesoscale features leading to regional variations in peak O_3 concentrations (Ryan 2002b). While an upper-level ridge was centered west of the region (Fig. 3), a quasi-stationary offshore low and a series of weak disturbances transiting New England caused the ridge to stall and oscillate until pushing eastward late in the episode. The weather conditions associated with this oscillating ridge included stagnations and rapid reversals in flow (1–2 August), dissipating fron-

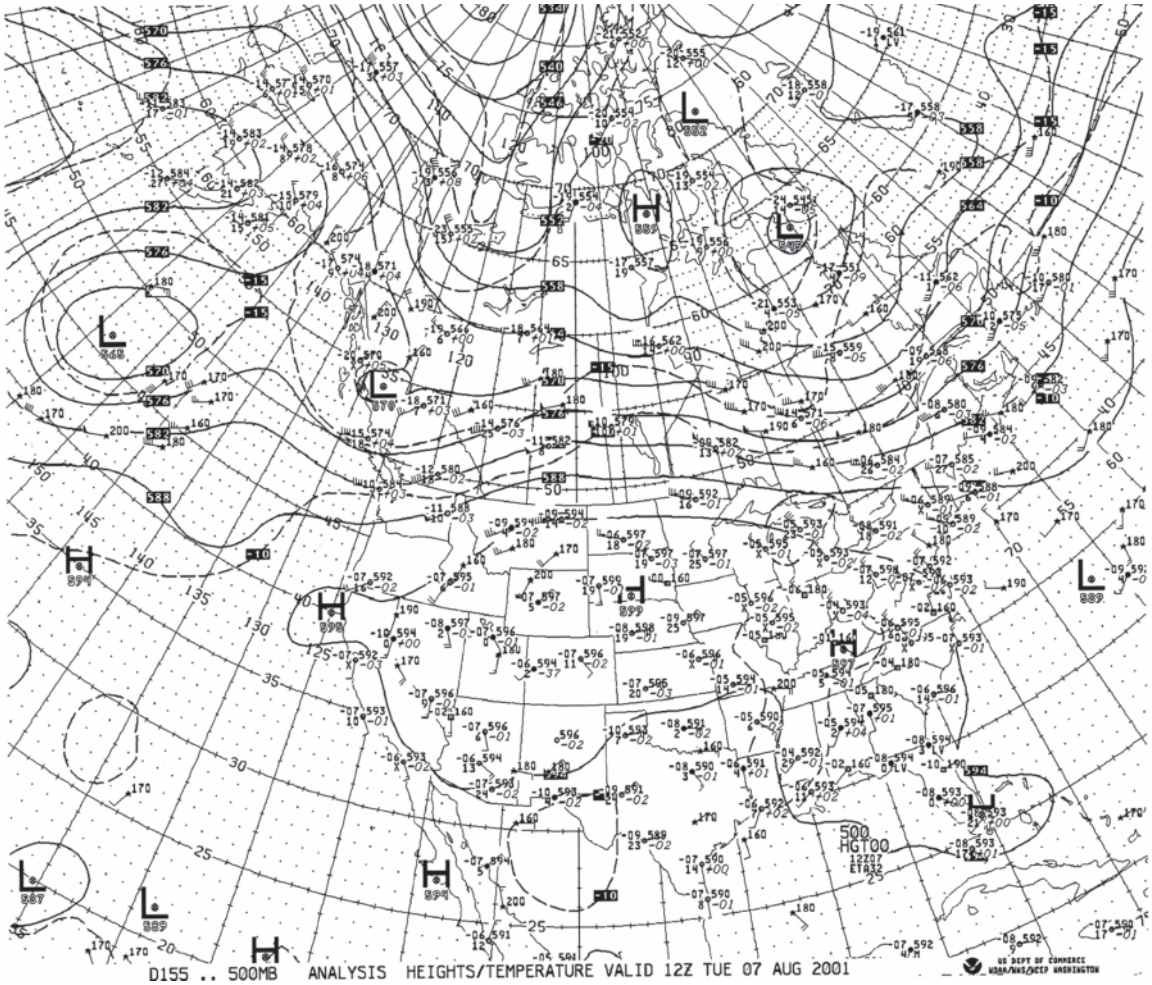


FIG. 3. 50-hPa analysis, prepared by NCEP, for 1200 UTC 7 Aug 2001. Solid contours are geopotential heights (dam); dashed contours are temperature (°C). Station data follow the standard convention.

tal boundaries (3–5 August), and a westerly transport high-O_3 period (6–9 August) ending with a strong convective event (10 August). These conditions cover the gamut of challenging high O_3 forecasts encountered in this region.

1–2 August: Onset of higher O_3 and airmass differences. On 1 August an upper-level trough was just offshore, with surface high pressure centered over central Maryland. As a result, O_3 concentrations peaked in the mid-Atlantic west of the Interstate-95 (I-95) cor-

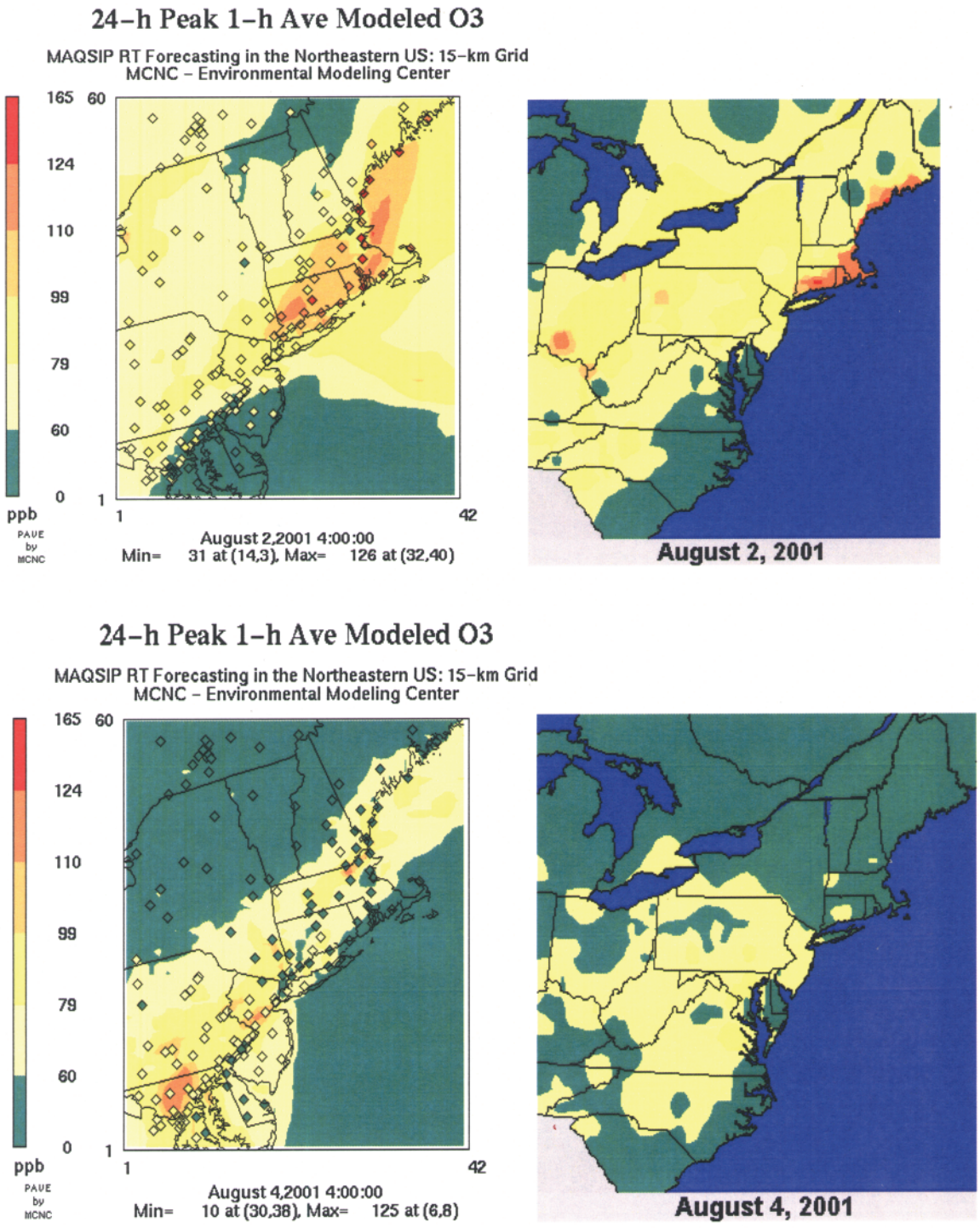


Fig. 4. (left) Model-predicted and (right) observed peak 1-h-avg O_3 for (top) 2 Aug and (bottom) 4 Aug 2001. Observed O_3 courtesy of EPA AIRNow (http://www.epa.gov/airnow).

ridor beneath the center of high pressure. Early on 2 August, an area of low pressure developed southeast of Cape Hatteras (HAT), North Carolina, at the base of the departing trough. Onshore flow was enhanced as the center of high pressure moved offshore, providing a cooler, cleaner maritime air mass to the southern mid-Atlantic, while in New England winds recirculated as high pressure passed to the south. This resulted in a steep south-to-north O_3 gradient, with highest concentrations found across Connecticut and coastal New England. As shown in Fig. 4 (top), which depicts forecast versus observed peak 1-h-average surface O_3, this was well resolved by the forecast model.

3–5 August: Frontal boundaries. The first in a series of short waves crossed New England early on 3 August driving a cold front, preceded by a prefrontal trough, over the eastern Great Lakes. These boundaries focused convection across central and western Pennsylvania, upstate New York, and New Hampshire. East of the active convection, a stable prefrontal warm sector concentrated high O_3 along a line from Maryland through north-central Connecticut to eastern Massachusetts. As its upper-level support moved rapidly offshore on 4 August, the low-level frontal boundary became quasi-stationary along a line from Portland (PSM), Maine, to Pittsburgh (PIT), Pennsylvania. Significant cloud cover occurred east of the I-95 corridor as the upper-level low lingered offshore, while west of this circulation widespread convection developed by midafternoon. The effect of this cloudiness, which suppressed O_3 production, was not well forecast, leading to overprediction in the model (Fig. 4, bottom).

6–10 August: The "high tide of summer." The frontal boundary dissipated over the region on 6 August, and temperatures warmed as the upper-level ridge pushed slightly east. Warm air advection resulted in a very stable atmosphere with a strong midlevel cap and surface temperatures exceeding 32°C along the I-95 corridor. In addition, an Appalachian lee trough (ALT; Weisman 1990) developed across the mid-Atlantic states. Convergence and limited vertical mixing are associated with the ALT, allowing it to focus precursor emissions and thus ozone formation (Seaman and Michelson 2000).

The synoptic-scale pattern remained quasi-stationary on 7 August with the exception of a short-wave disturbance crossing northern New England. Boundary layer temperatures increased regionwide on the order of 3°–4°C over the preceding 24-h period. While scattered unhealthy concentrations occurred as far south as the Delmarva Peninsula, the highest O_3 was concentrated along the southern New England coast, extending eastward into Cape Cod. The following day hot weather continued, although a "back door" cold front dropped quickly across eastern New England reaching just north of Providence (PVD), Rhode Island, by 1800 UTC 8 August. This boundary pushed the most conducive O_3 weather to just southeast of the I-95 corridor.

As the short wave departed on 9 August, the upper-level ridge oscillated eastward. Boundary layer winds backed to the west-southwest, the band of highest O_3 became oriented directly along the I-95 corridor, and peak concentrations rose. A vigorous cold front approached the region on 10 August. Widespread convection associated with the front reached a line from Boston (BOS), Massachusetts, through western Long Island, New York, to just south of Trenton (TTN), New Jersey, by 1800 UTC. Unhealthy O_3 concentrations were restricted to areas east of the oncoming convection. The arrival of a frontal boundary is typical of the termination phase of extended high-O_3 events, and the timing and extent of prefrontal convection is often a critical forecast question.

MAQSIP-RT was able to follow the changes in O_3 concentrations, predicting higher peak 1-h-average concentrations on 7 August (Fig. 5, top) and 9 August and lower concentrations on 8 August (Fig. 5, bottom) and 10 August. The location of the regional O_3 maximum was also well forecast. Peak O_3 concentrations are found along and east of the I-95 corridor, with a maximum in New England, on 7 August. As boundary layer winds shifted to northwest on 8 August, concentrations fell in New England but remained higher across the southern mid-Atlantic. The center of the elevated O_3 region then oscillated westward on 9 August and again was well placed by the forecast model (not shown).

While the location and day-to-day movement of the elevated-O_3 region is well handled by MAQSIP-RT, suggesting that the MM5 forecast wind fields were generally adequate overall, the finer-scale location and magnitude of the peak concentrations within the region were less accurate. In the PHL area on 7 August, concentrations are underpredicted at nearly all locations. This was due, in part, to underpredictions of upwind O_3 at rural and higher-elevation monitors across central Pennsylvania (Fig. 6, top). On 9 August, while the core high O_3 region is accurately retrograded westward along the I-95 corridor, the model predicted a stronger southwesterly wind component that pushed the plume north and prematurely cleaned out the southwestern New England coast. Figure 6 (bottom) reveals this as a slight leftward phase shift

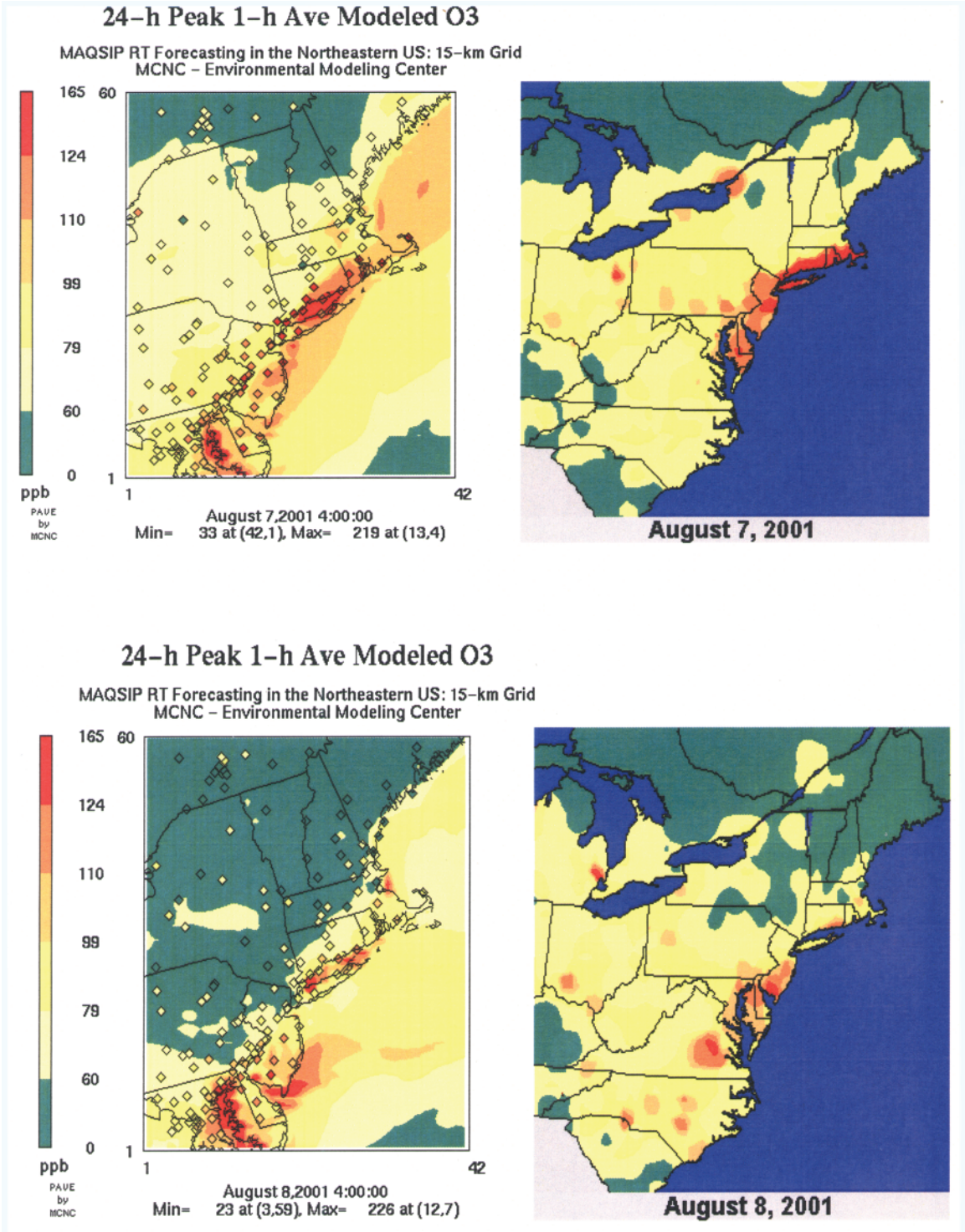


Fig. 5. As in Fig. 4, but for (top) 7 Aug and (bottom) 8 Aug 2001.

in the time series of hourly domainwide mean forecast versus observed O_3.

Summary of 1-h-average skill. Table 1 provides a summary of day-to-day forecast performance by MAQSIP-RT for 1-h-average concentrations across the northeast 15-km domain. Results for 5 August are not included because of a computer failure at 1200 UTC on 4 August. The statistics in Table 1 are restricted to O_3 concentrations (observed and modeled) above a threshold of 60 ppbv, the range of greatest interest to operational forecasters. To perform the

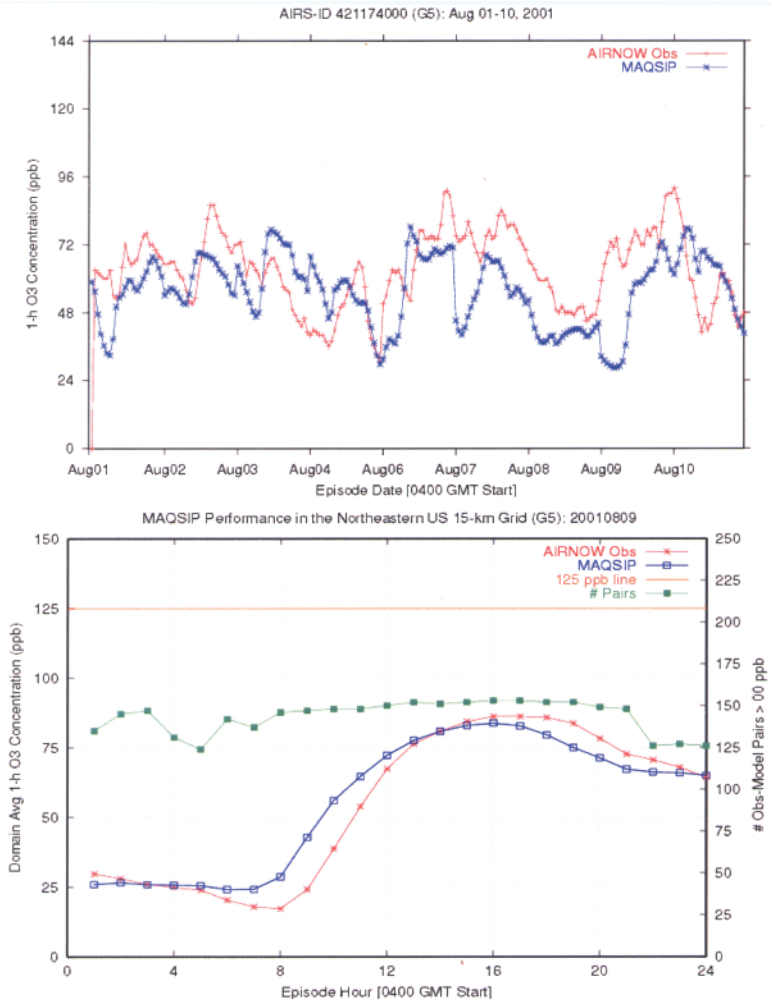


Fig. 6. Hourly predicted (blue line) and observed (red line) O_3 for (top) the Tioga monitor located in north-central Pennsylvania (41°38′41″N, 70°56′21″W), and domainwide mean hourly predicted and observed O_3 concentrations for 9 Aug 2001.

Table 1. MAQSIP-RT day-to-day model performance measures for 1–10 Aug for 1-h-avg concentrations across the 15-km NE domain. All measures are based on a 60-ppbv threshold.

Date	Bias (ppbv)	MAE (ppbv)	Gross error (%)	Rms error (ppbv)
1 Aug	–12.6	15.7	20.5	20.8
2 Aug	–9.0	12.3	15.9	16.2
3 Aug	0.5	11.0	15.5	14.0
4 Aug	12.5	17.9	26.8	20.8
6 Aug	–3.5	16.0	21.9	20.6
7 Aug	–14.4	17.4	20.2	23.1
8 Aug	–18.0	21.2	27.3	25.9
9 Aug	–7.7	20.6	19.5	20.6
10 Aug	2.6	15.9	17.2	15.9

analysis, 1-h-average O_3 concentration data were obtained from the U.S. EPA Air Quality System (AQS) database (EPA 2003b).

MAQSIP-RT is able, in forecast mode, to meet several key performance criteria for regulatory models that are exercised with analyzed, rather than forecast, meteorological fields. Gross error, in the 15%–27% range throughout the episode, meets the EPA performance criterion of 35% (EPA 1991). Model bias shows a good deal of day-to-day variation but overall is -9.7% when normalized, which is within the EPA performance criterion of ±5%–15%. Mean absolute error is in the 11–21-ppbv range, and rmse, which gives a rough estimate of forecast consistency, is in the 16–26-ppbv range. Model performance is consistent across the distribution range. In Fig. 7, 1-h-average predicted versus observed O_3 is given for the entire episode. There is a tendency to overpredict in the very low O_3 range, the overnight hours, but results are consistent with observations through the high range of the distribution during the afternoon.

COMPARISON WITH PEAK 1-H-AVERAGE OPERATIONAL STATISTICAL FORECASTS—PHL.

In Philadelphia, as in several large eastern cities, forecasts are provided and verified only on a metropolitan-wide scale using peak 1-h-average (within a 24-h period of time defined from local midnight to local midnight) O_3 as the predictand. This forecast may overestimate the area affected by poor air quality, but provides a margin of safety for public warnings where the spatial extent of the unhealthy air is both difficult to predict and measure. In addition, OAD pollution control strategies are only effective if they are applied at upwind source locations (typically lower in O_3) *and* downwind receptor locations. Metropolitan-wide peak O_3 forecasts may be effective for public outreach programs but are not particularly appropriate for the evaluation of a numerical forecast model (Hanna et al. 1996; Tesche et al. 1990). Thus, using this measure poses a fairly stiff test. However, for a model to be adopted by forecasters, it should show reasonable skill using the measure of interest.

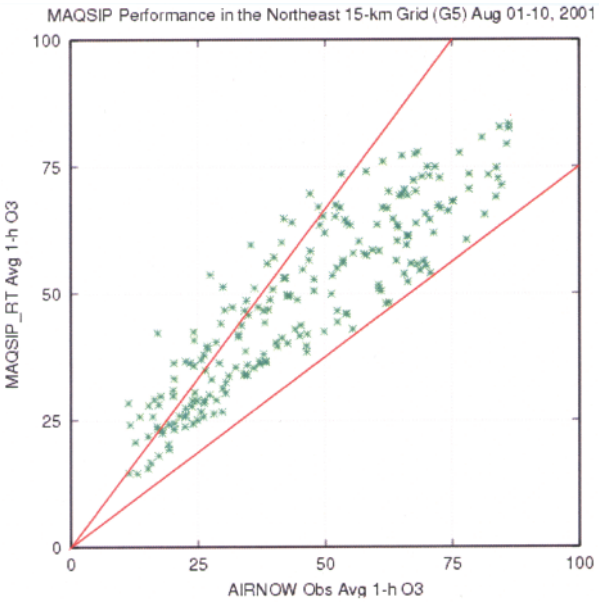


FIG. 7. Domain-averaged hourly observed and predicted O_3 concentrations for the periods 1–4 Aug and 6–10 Aug for the entire northeast U.S. 15-km model domain.

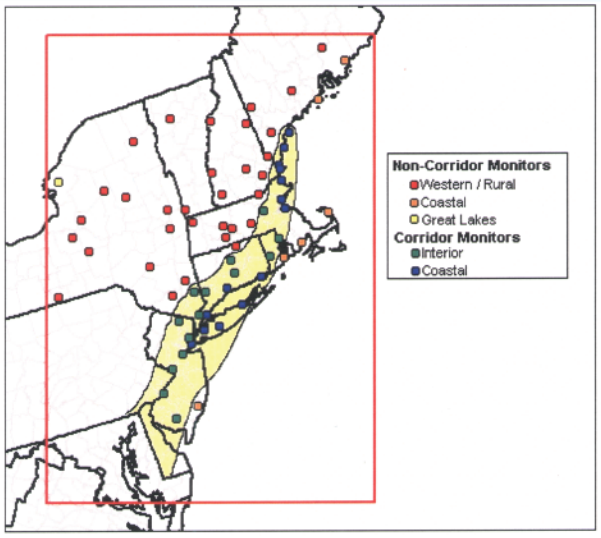


FIG. 8. Map of the northeastern United States showing 67 surface ozone monitor locations at which forecasts are issued daily. The 15-km MAQSIP-RT NE forecast domain is outlined in red. Corridor monitors are enclosed within the yellow "finger," with distinction between coastal (blue dots) and interior (green dots) corridor monitors shown. Western rural monitors are shown as red dots, coastal monitors as orange dots, and the single Great Lakes monitor as the yellow dot near the eastern shore of Lake Ontario.

The statistical model compared here has been used in the PHL forecast area in various forms since 1996 (see Ryan et al. 2000). The model is based on multiple linear regression techniques and trained with a historical database for the period 1991–2000. A set of three forecast algorithms was used in 2001, but they shared a similar set of predictors: maximum temperature, warm air advection in the boundary layer, wind speed at the surface and aloft, sky cover or relative humidity, solar zenith angle, and previous day's peak O_3. Meteorological inputs needed by the model are selected by the forecaster from all available NWP models and associated model output statistics (MOS) forecasts. Generally this includes both the NCEP Eta (Janjic 1994) and Aviation [(AVN) now Global Forecast System (GFS)] models; the data chosen for insertion are often a blend or consensus of the NWP model forecasts. Typically, credence is given to discussions posted by NCEP and local National Weather Service (NWS) forecast offices about the NWP model forecasts prior to selection of inputs to the statistical model. Over the period 1997–2001, the MAE of the statistical model is 12–16 ppbv, with an rmse of 15–19 ppbv, and positive biases ranging from 3 to 7 ppbv. The public forecast, modified by the operational forecasters, improved on the regression model by 13.5% in terms of MAE and 20% in terms of rmse.

For the 1–10 August period, the MAE for MAQSIP-RT was 12.1 ppbv and compared well with the statistical models (11.5–12.9 ppbv). The expert-modified public forecast was the best forecast with an MAE of 8.0 ppbv. MAQSIP-RT outperforms the raw statistical forecasts by consideration of the median absolute error (7.3 ppbv compared to 9.6–12.0 ppbv). As the difference in mean and median error suggests, day-to-day skill of both statistical and numerical models varied. The regression model was not able to resolve the northward extent of the advection of maritime air on 2 August and overpredicted in the range of 27–34 ppbv. On 4 August the regression model carried a better forecast of cloud cover and a subsequent reduction in temperature and so provided better forecasts than MAQSIP-RT, although it still retained an overprediction of 10–20 ppbv. Regression model skill was better than MAQSIP-RT on 7–8 August because MAQSIP-RT, as noted above, underpredicted across this region, but the regression model was less skillful on 9–10 August, with a tendency to underpredict. Overall, the skill of MAQSIP-RT was as good as or better than the raw statistical guidance.

COMPARISON WITH OTHER FORECAST METHODOLOGIES—NEW ENGLAND. In New England and the northern mid-Atlantic states,

forecasts are issued at monitor-specific locations rather than metropolitan-wide and use the 8-h-average O_3 as a predictand. This provides a unique database for comparing numerical model forecast performance with existing methods. The forecasts are verified with respect to peak 8-h-average O_3 (using forward 8-h averages defined from local midnight to local midnight) observed at a set of 67 monitors for which daily forecasts are issued (Fig. 8). The analysis compares forecast skill for the MAQSIP-RT with 1) persistence (PER) as a baseline measure of skill; 2) the official forecasts provided by air quality forecasting agencies across the Northeast (NEF); and 3) the operational Canadian Hemispheric and Regional Ozone and NO_x System (CHRONOS) model (CHR; Pudykiewicz et al. 1997).

The CHRONOS model is, like MAQSIP-RT, a three-dimensional Eulerian chemical transport model (Sirois et al. 1999; Environment Canada 2004). The emissions data used by CHRONOS are derived from the Canadian Emission Processing System (Moran et al. 1997), and meteorological fields including the horizontal wind components, vertical motions, temperature, clouds, and PBL parameters are calculated by the Global Environmental Multiscale model (GEM) (which is the operational weather prediction model of the Meteorological Service of Canada; Côté et al. 1998a,b). Unlike MAQSIP-RT, it is run once per day at 0000 UTC. CHRONOS graphical forecasts (Environment Canada 2003) were available to the NEFs during the episode of interest, whereas MAQSIP-RT results were available only to the authors and to NOAA researchers.

Domainwide analysis. In order to compare MAQSIP-RT with the three other forecast methods, numerical model data were bilinearly interpolated to the monitor locations. Including all 67 monitors shown in Fig. 8, bias, MAE, and rmse statistics are given in Figs. 9 and 10. These figures show that MAQSIP-RT performs as well or better than other forecast methods for both bias and error measures. Though NEF and MAQSIP-RT are statistically close for day-to-day bias, with each better on 4 of 9 days and one near tie (4 August), for the whole episode MAQSIP-RT is less biased than either CHRONOS or NEF. For both MAE and rmse, MAQSIP-RT performs best for day-to-day predictive success and for the entire episode. In Fig. 9, PER reflects a small episode bias because its large day-to-day negative and positive biases cancel; in addition, its errors are clearly largest (Fig. 10). Hence, all three operational methods improve on persistence when all 67 monitors are included in the database.

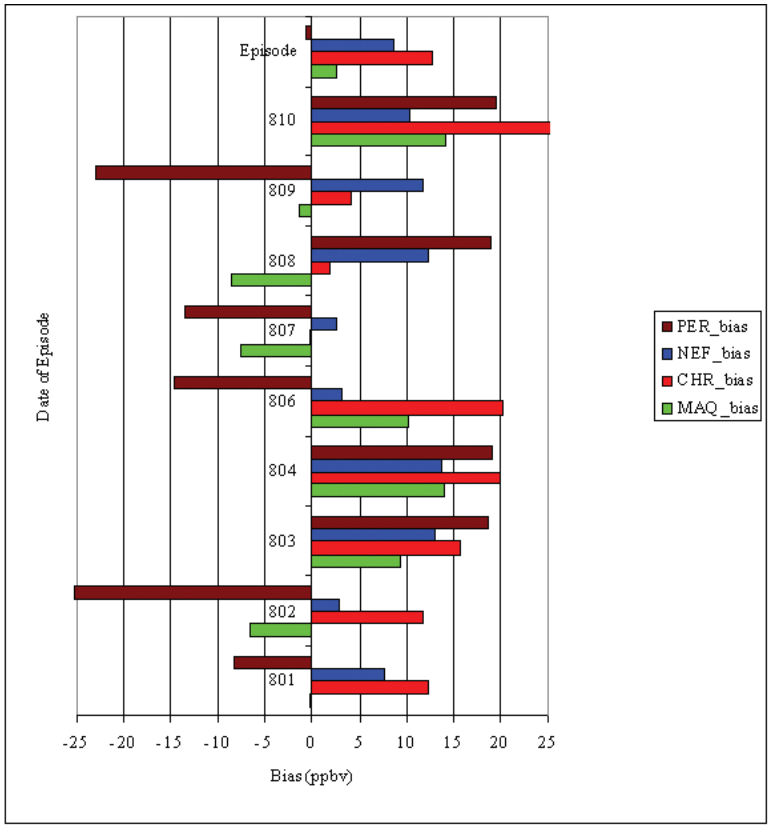
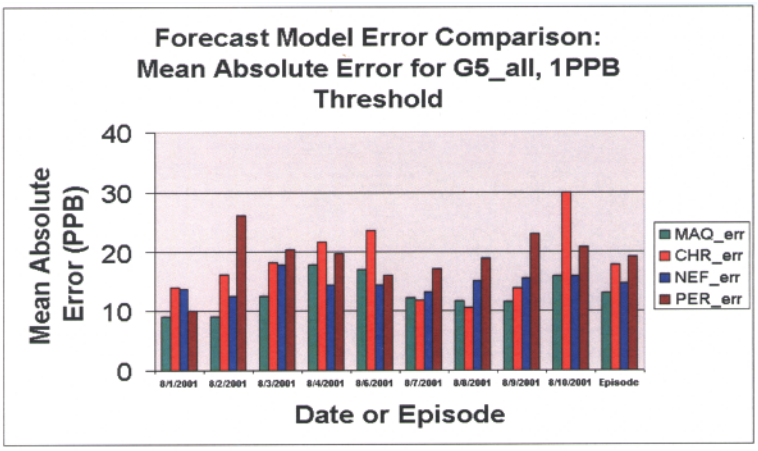


FIG. 9. Overall episode and daily bias statistics for peak 8-h-avg O_3 at all 67 forecasted monitors for the 1–10 Aug 2001 episode. MAQSIP-RT (MAQ) is given in green, NE forecasts (NEF) in blue, CHRONOS (CHR) in red, and persistence (PER) in brown. All measures are in ppbv.

Subregional analysis. In addition to domainwide skill, it is also useful to consider model performance for subregional subsets of the database. These subsets are identified in Fig. 8 and include monitors along the heavily populated Interstate-95 corridor (green dots), coastal monitors (blue plus orange dots), rural monitors to the west and north of the corridor (red dots), and one Great Lakes monitor (yellow dot). These subregions contain differences in emissions and meteorological regimes and pose forecasting challenges. The western rural monitors (WRMs) are

outside the main belt of I-95 corridor emissions and experienced lower O_3 concentrations, with mean peak 8-h averages of 57.9 ppbv compared to 77.9 ppbv for the interior-corridor monitors (ICMs). Since they are typically upwind of the corridor during extended high-O_3 episodes, they can inform the downwind forecaster (Ryan et al. 1998). The ICMs are located near concentrated emissions of O_3 precursors and observe the highest O_3 levels in the region. Coastal monitors (CMs) may be located near urban centers, but, because they are close to the land–sea interface, they are subject to rapid changes in O_3 concentrations not well handled by the statistical models; the Great Lakes monitor (GLM) is similar. Figure 11 shows the relationship between the different subregions and the response of the MAQSIP-RT. The WRMs tend to congregate in the lower end of the distribution, while the CMs show an extremely wide range of distribution. At the GLM, the MAQSIP-RT forecasts were biased low for this episode.

Forecast performance [including index of agreement (IA)] for the various methods and subregions is given in Table 2. For the WRM subregion, all methods tend to overpredict peak O_3, with biases ranging from 4.4 ppbv for MAQSIP-RT to 13.2 ppbv for CHRONOS. For all measures, MAQSIP-RT performs best, with an MAE of only 11.3 ppbv and a much-reduced bias of 4.4 ppbv. For the CMs, the relatively high error by the PER method is an indication of the diffi-



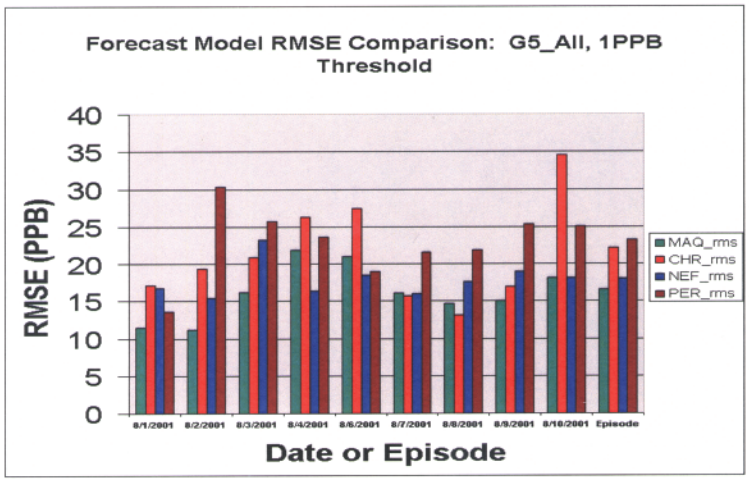


Fig. 10. As in Fig. 9, but for (top) MAE and (bottom) rmse.

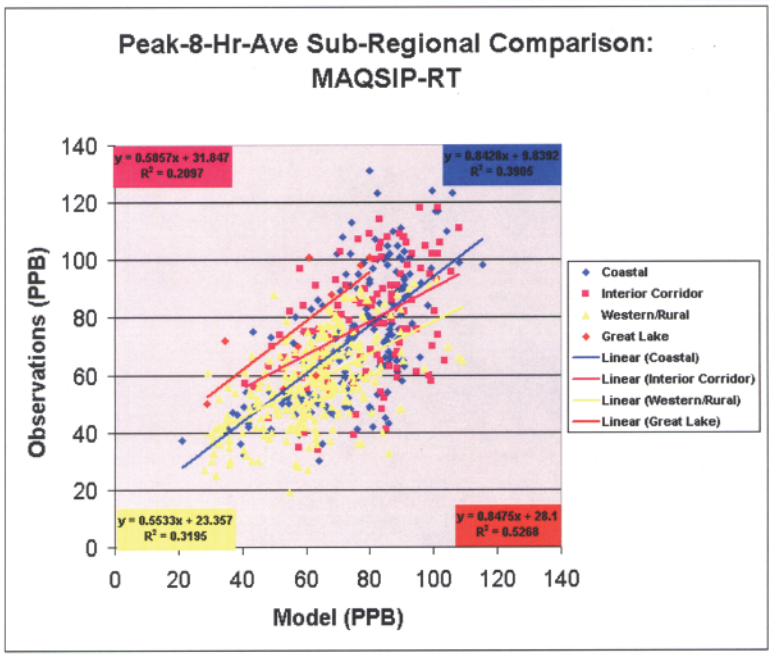


Fig. 11. Scatterplot of MAQSIP-RT forecasts and observations for forecasted monitors grouped by western rural monitors (yellow), interior corridor monitors (purple), coastal monitors (blue), and the Great Lakes monitor (red). Observations are plotted as a function of the model forecast; best-linear-fit statistics are given in the border. The location of the monitors is shown in Fig. 8.

Table 2. Forecast performance measures for selected subregions (all measures in ppbv).				
Subregion	**MAQSIP**	**CHRONOS**	**NE forecasts**	**Persistence**
Coastal	Monitors = 20	N=177	Mean peak O_3 = 72.8	
Bias	+1.9	+9.3	+5.8	–1.6
MAE	13.8	18.7	16.4	24.4
Rmse	17.2	23.4	20.3	29.1
Index of agreement	*0.77*	*0.66*	*0.71*	*0.45*
Western rural	Monitors = 32	N=277	Mean peak O_3 = 57.9	
Bias	+4.4	+13.2	+11.6	–0.7
MAE	11.3	16.1	14.3	14.8
Rmse	14.9	20.0	17.2	17.8
Index of agreement	*0.74*	*0.70*	*0.68*	*0.58*
I-95 corridor interior	Monitors = 20	N=118	Mean peak O_3 = 77.9	
Bias	+0.7	+17.7	+6.7	–12.6
MAE	15.0	20.0	13.2	21.6
Rmse	18.4	24.8	16.3	24.8
Index of agreement	*0.68*	*0.61*	*0.77*	*0.44*

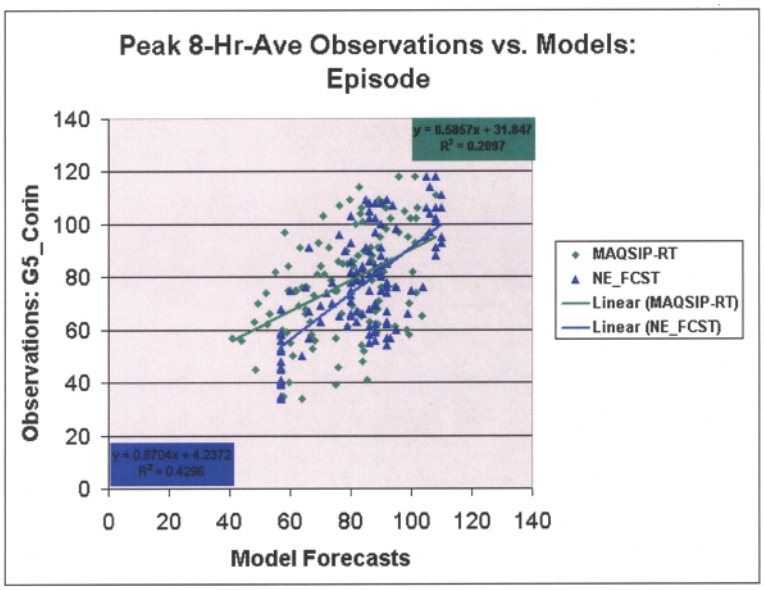


Fig. 12. Scatterplot of peak 8-h-avg O_3 for the interior-corridor forecast monitors for MAQSIP-RT and northeast U.S. forecasts. Observations are plotted as a function of model forecasts; best-linear-fit statistics are given in the border.

culty in forecasting along the land–sea boundary. In this subregion, MAQSIP-RT provides the most consistent forecast. Within the interior I-95 corridor, however, the NEF is slightly better than MAQSIP-RT. This reflects model difficulties in resolving O_3 in a region of steep emissions gradients as well as the role of forecaster experience in determining the extent of peak O_3 concentrations in the near-urban environment. The NEF shows better skill at the lower concentrations (Fig. 12) and slightly less scatter at the higher concentrations. In the corridor interior there is a large difference in bias between the two numerical methods.

Threshold analysis. Air quality forecasts are issued to the public using color codes (AIRNow 2003), which are simpler to present to the public and allow for clear health messages; see Table 3. The boundary between any two colors, for example, yellow to orange, represents a health-exposure threshold. Using these color categories, Figs. 13, 14, and 15 provide spatial comparisons of the three operational forecast methods against observations for 2, 7, and 9 August. Here, the forecast maximum 8-h-average exposure in parts per billion by volume is plotted using forward 8-h averages. By presenting forecast results in these terms, the numerical model output can be interpreted using the same health implications underlying the official forecasts.

Thus, in addition to the domainwide and subregional discrete evaluation just presented, contingency-table-based statistics—similar to those used for assessing quantitative precipitation forecasts (QPFs)—are also important metrics of forecast performance (Tables 4 and 5; Murphy and Winkler 1987). Accuracy (A) measures how often the forecasts were correct either above or below the threshold. Bias (B) determines whether the same *fraction* of events are both forecast and observed. If $B = 1$, then the forecast is unbiased. If $B < 1$ there is a tendency to underpredict, and if $B > 1$ there is a tendency to overpredict. The false alarm *ratio* (F) measures the percentage of forecast high O_3 events that turn out to be false alarms. The probability of detection (POD) or "hit" rate (H) is a measure of how often a high threshold occurrence is actually predicted to occur, and the Heidke skill score (HSS) computes the percentage improvement in forecast accuracy as compared to random chance, its range being ±1, with a random forecast equal to zero. Another useful measure, the critical success index (CSI) or "threat" score, measures the percentage of cases that are correctly forecast out of those either forecast or observed. For this measure, the range is [0–1], 1 being perfect. Finally, the Pierce skill score (PSS), or Hansen and Kuipers discriminant, is a measure of

Table 3. Color-coded ozone forecast concentration thresholds for forecast 8-h-avg O_3.

Color code	8-h-avg threshold (ppbv)	U.S. EPA health message (AIRNow 2003)
Green	< 65	*Good:* No health impacts are expected in this range.
Yellow	65–84	*Moderate:* Unusually sensitive people should consider limiting prolonged outdoor exertion.
Orange	85–104	*Unhealthy for sensitive groups:* Active children and adults, and people with respiratory disease, such as asthma, should limit prolonged outdoor exertion.
Red	105–124	*Very unhealthy:* Active children and adults, and people with respiratory disease, such as asthma, should avoid prolonged outdoor exertion; everyone else, especially children, should limit prolonged outdoor exertion.
Purple	125	*Hazardous:* Same message as *very unhealthy*.

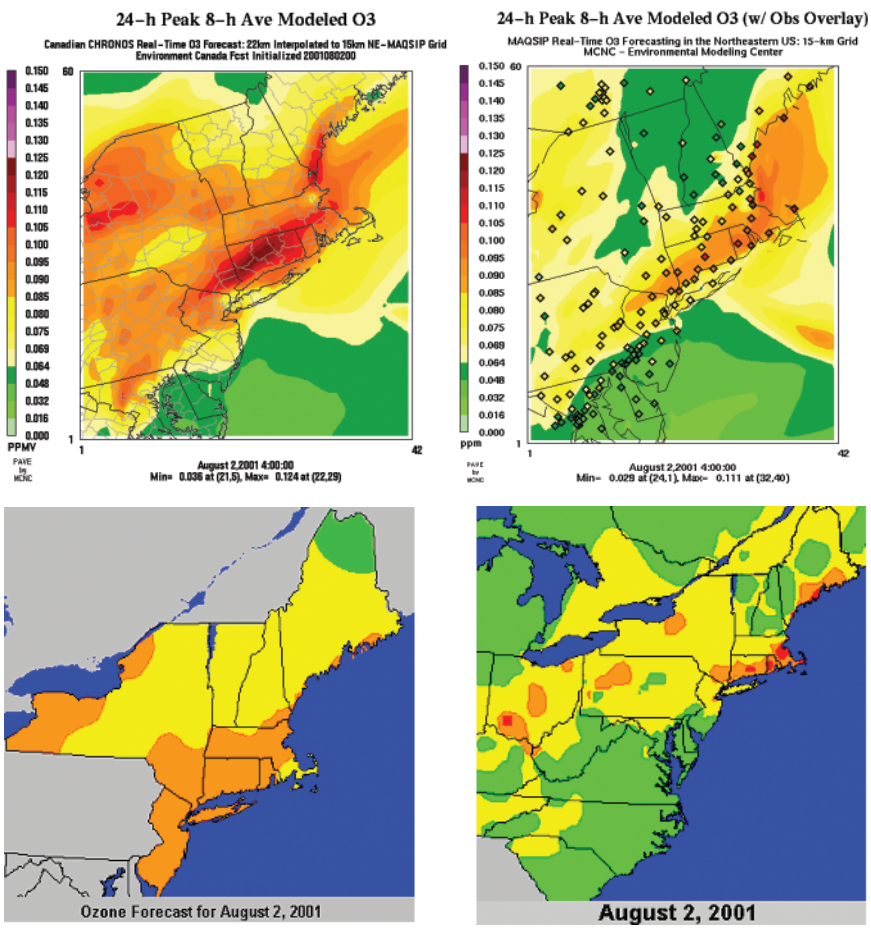


Fig. 13. Spatial peak 8-h-avg 24-h ozone forecasts [(top left) CHRONOS; (top right) MAQSIP-RT; (bottom left) official northeast U.S. forecast] vs (bottom right; courtesy EPA AIRNow) gridded observations for 2 Aug 2001, using identical color scales following U.S. EPA color scale (Table 3).

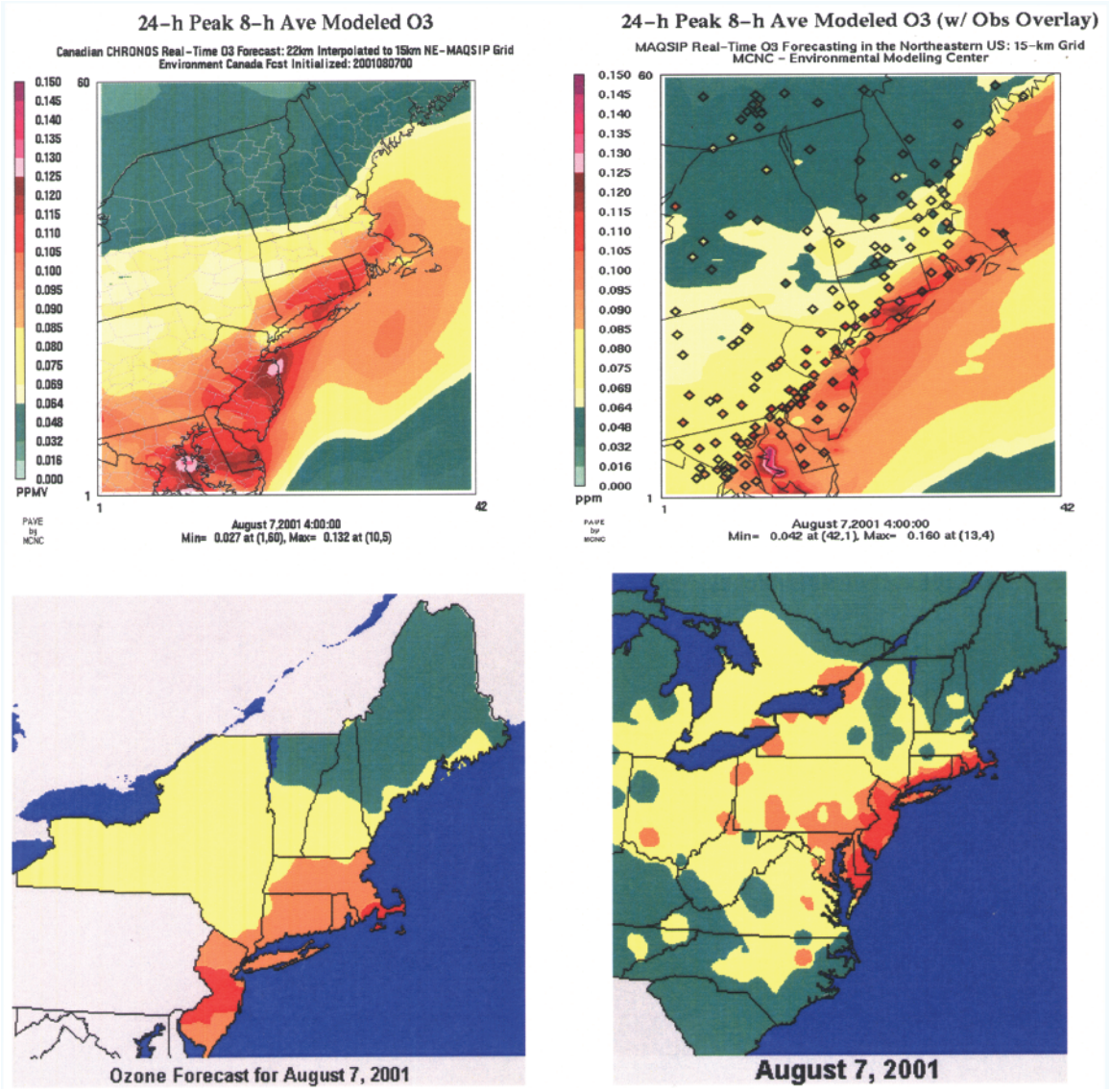


Fig. 14. Same as Fig. 13, except for 7 Aug 2001.

skill obtained by the difference between the hit rate and the false alarm *rate*, where the false alarm rate is the counterpart to the hit rate. The range of this measure is ±1. If the PSS is greater than zero, then the hit rate exceeds the false alarm rate and the forecast has some skill. Note, however, that if the number of correct "negatives" is large, as it is in this case, the PSS tends toward the POD.

Table 4. Contingency table for threshold forecasts.

		Observed	
		Yes	No
Forecast	Yes	a	b
	No	c	d

The results of forecast skill for a threshold of 8-h-average O_3 exceeding 85 ppbv over all 67 monitors are given in Table 6. This threshold was chosen because 85 ppbv over 8 h is the measured level at which a monitor is considered in violation of the new 8-h-average U.S. National Ambient Air Quality Standard (NAAQS) for O_3 and represents the cutoff between relatively good (green/yellow) and relatively poor (orange/red/purple) air quality. Further, agencies in New England (Connecticut DEP 2003; Desimone 2003, personal communication; Giuliano 2003, personal communication) use this cutoff to issue OAD advisories (in contrast to the PHL metro practice noted above).

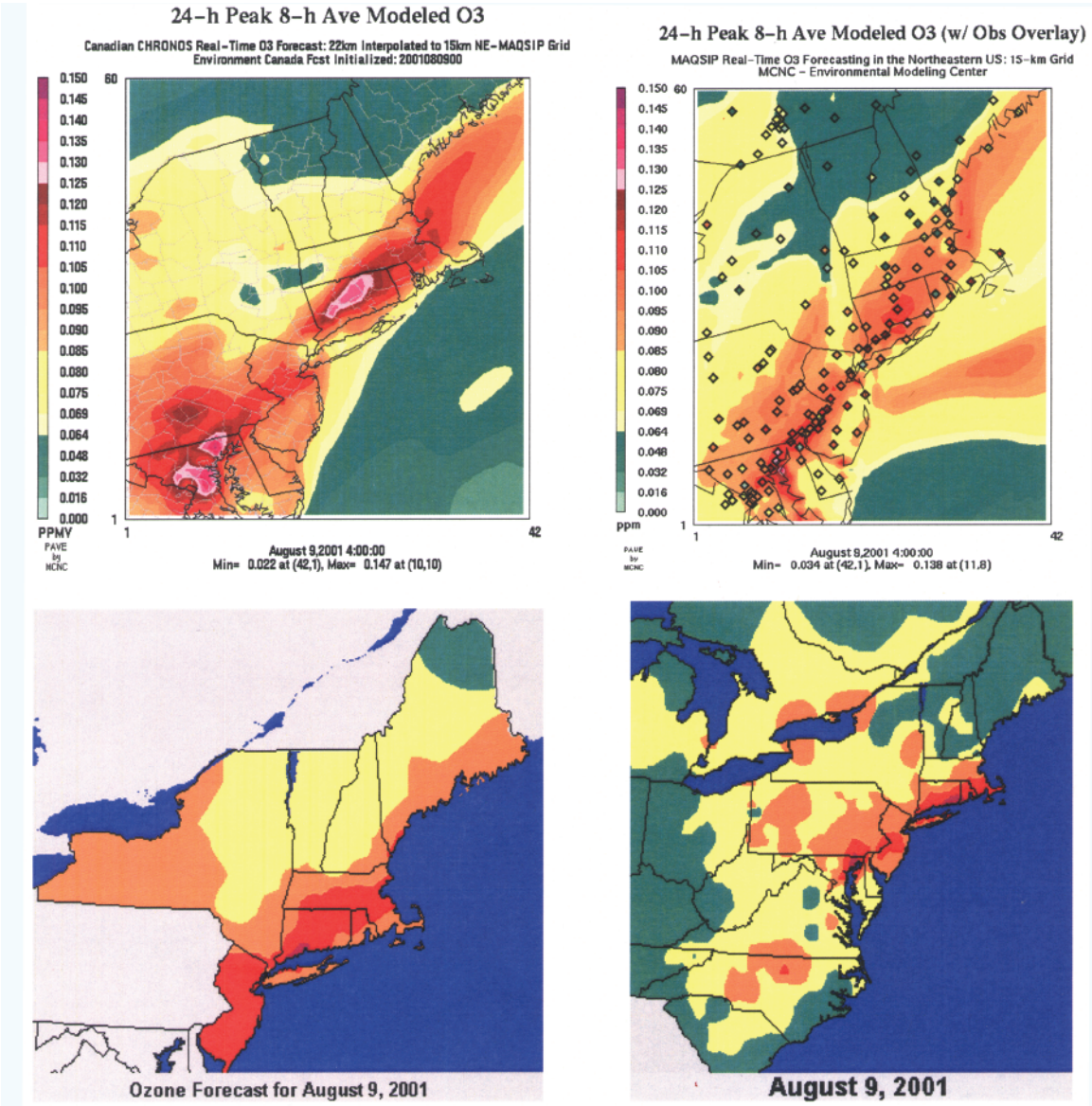


Fig. 15. Same as Fig. 13, except for 9 Aug 2001.

Overall, the MASIP-RT and NEF forecasts have the highest skill. With the exception of the PER bias score, MAQSIP-RT has better accuracy, bias, hit rate, and false alarm ratio scores than the other methods, with NEF second. The PER bias score is discounted, however, for the same negative-bias/positive-bias cancellation reasons noted above. Although PER is usually a good predictor for peak O_3, it does not provide much skill in either the critical high end of the O_3 distribution or in situations where there are frequent but subtle wind shifts, as occurred during this episode. The NEF forecasts have better HSS and CSI measures, with MAQSIP-RT second. NEF is just marginally better than MAQSIP-RT with respect to the PSS. Note, however, that the NEF forecasts, while reasonably skillful overall, tend to have a very high bias (1.61) compared to MAQSIP (1.06). This suggests a degree of "hedging" by the forecasters in the higher O_3 cases to avoid missing an unhealthy warning (orange/red/purple). The result is a higher false alarm ratio (0.22 to 0.13) compared to MAQSIP-RT. Reflecting its overall high bias, the CHRONOS model exhibits a high false alarm ratio.

SUMMARY AND DISCUSSION. The focus of this paper has been the description and evaluation of a numerical air quality prediction (NAQP) system running as part of a NOAA-funded pilot study. The NAQP system couples the MM5, the SMOKE model, and the MAQSIP-RT photochemical oxidant model.

Table 5. Threshold skill measures.

Skill measure	Symbol	Formula
Accuracy	*A*	$\frac{a+d}{a+b+c+d}$
Bias	*B*	$\frac{a+b}{a+c}$
False alarm ratio	*F*	$\frac{b}{a+b}$
Probability of detection (POD)	*H*	$\frac{a}{a+c}$
Heidke skill score	HSS	$\frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)}$
Critical success index	CSI	$\frac{a}{a+b+c}$
Pierce skill score	PSS	$\frac{a}{(a+c)}-\frac{b}{(b+d)}$

Table 6. Skill score results for forecast methods. "Blend" refers to a 50–50 weighted average of both numerical forecasts.

	MAQSIP-RT	CHR	NEF	PER	Blend
H	0.49	0.28	0.45	0.19	0.49
F	0.13	0.26	0.22	0.14	0.19
B	1.06	1.47	1.61	0.98	1.48
A	0.80	0.68	0.77	0.76	0.80
CSI	0.34	0.20	0.39	0.10	0.41
PSS	0.37	0.12	0.38	0.05	0.41
HSS	0.38	0.13	0.42	0.05	0.46

The routine use of this coupled forecast system represents a major advance in the field of air quality forecasting. During summer 2002, for example, the system was run for 5 months (1 May–30 September) without interruption and participated in the New England Air Quality Study (NEAQS; NOAA 2003b). Early results from that field program are encouraging and will be reported in a paper to follow.

The skill of the forecast model is evaluated with respect to a variety of other forecasting approaches that include standard statistical methods as well as another numerical model, for a northeastern U.S. ozone episode that occurred 1–10 August 2001. Overall model performance by MAQSIP-RT was quite good. The model, executed in real time as a forecast system, met or exceeded EPA performance criteria for regulatory air quality models. Moreover, its performance was consistent with current benchmark statistical forecast methods with respect to metropolitan-wide peak 1-h-average O_3 in the PHL area, although not as well as forecasts modified using expert analysis.

When discretely evaluated against monitor-specific forecasts in the northeast United States, MAQSIP-RT improved on expert forecasts, persistence, and the CHRONOS numerical model by a variety of traditional (bias, MAE, rmse, IA) measures, taken over the whole set of monitors. In addition, MAQSIP-RT performed best in two key subregions, the WRMs that "define" the regional background O_3 concentrations and the CMs that are often subject to abrupt airmass changes. The expert forecasts were slightly better in the interior of the I-95 corridor, reflecting model difficulties in resolving the effects of steep near-urban precursor gradients as well as forecaster experience in this environment.

Because air quality forecasts are issued to the public in the form of color codes representing exposure levels, forecast skill at high O_3 thresholds are an important measure of performance. For a variety of threshold forecast skill measures, MAQSIP-RT outperformed the CHRONOS model and provided results similar to the expert New England forecasters for an 8-h average of 85 ppbv as the threshold. This threshold represents the cutoff between relatively good and relatively poor air quality and is used to trigger Ozone Action Day advisories in New England. The strengths of MAQSIP-RT with respect to high-O_3 threshold forecasts are the lack of a systematic bias and a relatively low false alarm rate. Overall, these results suggest that MAQSIP-RT, if used as the only source of forecast information, *would have provided OAD and health advisories statistically indistinguishable from those issued by the expert New England forecasters for this episode.*

Spatial O_3 patterns, of special concern to forecasters, are well resolved in cases of airmass changes. These cases are of interest in the northeastern United States because of the frequent interactions of maritime and continental air masses across the high-emissions urban corridor. While regional-scale O_3 was generally well predicted, in some cases significant underpredictions occurred that affect downstream peak O_3 forecasts in adjacent metropolitan areas. In particular, O_3 across rural Pennsylvania was underpredicted on 6–8 August with poor forecast skill following in the PHL area. As configured in 2001, MAQSIP-RT used 12-h forecasts as the initial chemistry fields for the following run. To the extent a prior run underpredicted key trace gases, this underprediction was carried over to the subsequent run. This approach was modified in 2002 by assimilating real-time surface O_3 data from the EPA AIRNow Data Management Center (AIRNow 2002). The effects of the data assimilation on the 2002 forecasts are being studied.

Because NAQP systems are so complex, there are numerous opportunities for improvements. For example, differences in the numerical or theoretical treatment of similar physical processes between the meteorological and chemical models may be a source of error, and work has begun on the implications (Byun 1999a,b). Further, the development of combined numerical-statistical approaches, much like the MOS systems used in NWP, can be expected to result in objective forecast improvements. Additionally, ensemble NAQPs, which could involve combinations of different meteorological, emissions, and atmospheric chemistry models as well as integrated meteorological–chemical models, such as the Weather Research and Forecast Chemistry model (WRF-Chemistry 2004), could prove extremely useful in the future. In the near term, the advent of NOAA's new operational air quality forecasting system (Davidson 2003), planned for debut in September of 2004, connecting the NCEP Eta Model to the EPA CMAQ model should provide an independent "ensemble member." Table 6 presents the threshold statistics for a 50–50 blend of the MAQSIP-RT system with CHRONOS for the episode evaluated, and it can be seen that this blend improves some, but not all, of the calculated performance measures.

Overall, MAQSIP-RT can be expected to add value to current operational ozone forecasts designed to help the public avoid unhealthy exposure to high ozone levels and to reduce metropolitan region precursor emissions through OAD programs, though forecast accuracy may differ for regions of the United States outside the NE in which the relative roles of transport, in situ production, and precursor emissions vary. Further, the system reported here represents a good example of the new and expanding role of the private sector in providing value-added products to the public (Dutton 2002). In combination with the emerging NOAA effort, the MAQSIP-RT system should help "stimulate effective public and private-sector cooperation" (Dutton 2002) in developing the best possible numerical guidance to those in operational air quality forecast decision-making roles and to the general public.

ACKNOWLEDGMENTS. This work was funded in part by the National Oceanic and Atmospheric Administration under PO 43EANR130149 to MCNC as part of NOAA's "Early Start" Air Quality Forecasting Pilot Program; in part by the North Carolina Department of Environment and Natural Resources; and in part by Capitol Broadcasting Company (CBC) who contributed to the MM5 implementation. The authors would like to thank the following individuals for their contributions to this work: David Rogers and Pai-Yei Whung of NOAA's Office of Weather and Air Quality; Ken Schere of NOAA's Air Resources Laboratory; Ted Smith of Baron Advanced Meteorological Systems for his work on forecast system design and product delivery; Dave Conroy of EPA Region I for providing the NEF dataset; Annette Lario-Gibbs, Glenn Hunter, and Dave Stauffer of The Pennsylvania State University's Department of Meteorology for early discussions about system design; Greg Fishel of CBC and Gary Lackmann of the North Carolina State University's Marine, Earth and Atmospheric Sciences Department for MM5 performance discussions; the many state-agency forecasters who use and evaluate the forecasts; and three anonymous reviewers who helped significantly improve the manuscript. Although some of the research described in this article was funded in part by the U.S. Environmental Protection Agency through Grant R825211 to MCNC, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

REFERENCES

AIRNow, cited 2003: US EPA Office of Air Quality Planning and Standards ozone forecast, action day, and public health alert clearinghouse. [Available online at www.epa.gov/airnow.]

Bascom, R., and Coauthors, 1996a: Health effects of outdoor air pollution. 1. *Amer. J. Respir. Crit. Care Med.,* **153,** 3–50.

——, and Coauthors, 1996b: Health effects of outdoor air pollution, 2. *Amer. J. Respir. Crit. Care Med.,* **153,** 477–498.

Blackadar, A. K., 1976: Modeling the nocturnal boundary layer. *Proc. Third Symp. on Atmospheric Turbulence, Diffusion and Air Quality,* Raleigh, NC, Amer. Meteor. Soc., 46–49.

Bott, A., 1989: A positive definite advection scheme obtained by nonlinear renormalization of the advective fluxes. *Mon. Wea. Rev.,* **117,** 1006–1015.

Burrows, W. R., M. Benjamin, S. Beauchamp, E. R. Lord, D. McCollor, and B. Thomson, 1995: CART decision-tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada. *J. Appl. Meteor.,* **34,** 1848–1862.

Byun, D. W., 1999a: Dynamically consistent formulations in meteorological and air quality models for multiscale atmospheric studies. Part I: Governing equations in a generalized coordinate system. *J. Atmos. Sci.,* **56,** 3789–3807.

——, 1999b: Dynamically consistent formulations in meteorological and air quality models for multiscale atmospheric studies. Part II: Mass conservation issues. *J. Atmos. Sci.,* **56,** 3808–3820.

——, and J. S. Ching, Eds., 1999: Science algorithms of the EPA-Models-3 Community Multiscale Air Quality (CMAQ) modeling system. EPA Rep. EPA-600/R-99/030, U.S. EPA, Office of Research and Development, Washington, DC, 758 pp.

Carter, W. P. L., 1996: Condensed atmospheric photooxidation mechanism for isoprene. *Atmos. Environ.,* **30,** 4275–4290.

Chang, J. S., R. A. Brost, I. S. A. Isaksen, S. Madronich, P. Middleton, W. R. Stockwell, and C. J.Walcek, 1987: A three-dimensional Eulerian acid deposition model: Physical concepts and formulation. *J. Geophys. Res.,* **92,** 14 681–14 700.

——, Y. Li, M. Beauharnois, H.-C. Huang, C.-H. Lu, and G. Wojcik, 1996: SAQM user's guide. California Air Resources Board, Sacramento, CA, 500 pp.

Chang, M. E., and C. Cardelino, 2000: Application of the Urban Airshed Model to forecasting next-day peak ozone concentrations in Atlanta, Georgia. *J. Air Waste Manage. Assoc.,* **50,** 2010–2024.

Clark, T. L., and T. R. Karl, 1982: Application of prognostic meteorological variables to forecasts of daily maximum one-hour ozone concentrations in the northeastern United States. *J. Appl. Meteor.,* **21,** 1662–1671.

Coats, C. J., Jr., 1996: High-performance algorithms in the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system. Preprints, *Ninth Joint Conf. on Applications of Air Pollution Meteorology with A&WMA,* Atlanta, GA, Amer. Meteor. Soc., 584–588.

——, cited 2003: The EDSS/Models-3 I/O API. Baron Advanced Meteorological Systems, Research Triangle Park, NC. [Available online at www.baronams.com/products/ioapi/.]

——, A. F. Hanna, D. Hwang, and D. W. Byun, 1995: Model engineering concepts for air quality models in an integrated modeling system. *Proc. AWMA Specialty Conf. on Regional Photochemical Measurement and Modeling Studies,* San Diego, CA, Air and Waste Management Association, 213–222.

——, J. N. McHenry, A. Lario-Gibbs, and C. D. Peters-Lidard, 1998: MCPL: A drop-in MM5-V2 module suitable for coupling MM5 to parallel environmental models; with lessons learned for the design of the weather research and forecasting (WRF) model. *Proc. Eighth PSU/NCAR Mesoscale Model Users' Workshop,* Boulder, CO, Mesoscale and Microscale Meteorology Division, NCAR, 117–120.

Cobourn, G. W., L. Dolcine, M. French, and M. Hubbard, 2000: A comparison of non linear regression and neural network models for ground-level ozone forecasting. *J. Air Waste Manage. Assoc.,* **50,** 1999–2009.

Comrie, A. C., 1997: Comparing neural networks and regression models for ozone forecasting. *J. Air Waste Manage. Assoc.,* **47,** 653–663.

Connecticut DEP, cited 2003: Air quality index, tips for ozone action days. Connecticut Department of Environmental Protection, Hartford, CT. [Available online at www.dep.state.ct.us/updates/oz/aqi.asp.]

Côté, J., J.-G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998a: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part II: Results. *Mon. Wea. Rev.,* **126,** 1397–1418.

——, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998b: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.,* **126,** 1373–1395.

Crutzen, P. J., 1979: The role of NO and NO_2 in the chemistry of the troposphere and stratosphere. *Annu. Rev. Earth Planet. Sci.,* **7,** 443–472.

Dabberdt, W. F., and Coauthors, 2000: Forecast issues in the urban zone: Report of the 10th prospectus development team of the U.S. Weather Research Program. *Bull. Amer. Meteor. Soc.,* **81,** 2047–2063.

Davidson, P., cited 2003: National air quality forecast capability: First steps toward implementation.

Program Overview, National Air Quality Forecast Capability, National Weather Service, Office of Science and Technology. [Available online at www.nws.noaa.gov/ost/air_quality/Davidson.pdf.]

DeMore, W. B., and Coauthors, 1994: Chemical kinetics and photochemical data for use in stratospheric modeling, Evaluation number 11. NASA Jet Propulsion Laboratory Publ. 84-26, Pasadena, CA, 273 pp.

Dickerson, R. R., B. G. Doddridge, and K. Rhoads, 1995: Large scale pollution of the atmosphere over the remote North Atlantic: Evidence from Bermuda. *J. Geophys. Res.,* **100,** 8945–8952.

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment by a mesoscale three-dimensional model. *J. Atmos. Sci.,* **46,** 3077–3107.

——, 1996: A multilayer soil temperature model for MM5. *Proc. Sixth PSU/NCAR Mesoscale Model Users' Workshop,* Boulder, CO, Mesoscale and Microscale Meteorology Division, NCAR, 49–50.

Dutton, J., 2002: Opportunities and priorities in a new era for weather and climate services. *Bull. Amer. Meteor. Soc.,* **83,** 1303–1311.

Environment Canada, cited 2003: CHRONOS graphical forecast results. [Available online at www.msc-smc.ec.gc.ca/AQ_SMOG/chronos_e.cfm.]

——, cited 2004: CHRONOS description. [Available online at www.cmc.ec.gc.ca/~arqidor/chronos_description/chronos_description.html.]

EPA, 1986: Air quality criteria for ozone and other photochemical pollutants, Vol. II. EPA Rep. EPA-600/8-84/020bF, Environmental Criteria and Assessment Office, Research Triangle Park, NC, 460 pp.

——, 1991: Guideline for regulatory application of the Urban Airshed Model. EPA Rep. EPA/450/4-91/013, Research Triangle Park, NC, 108 pp.

——, 1995: Urban Airshed Model (UAM) biogenic emission inventory system: Version 2 (BEIS 2) user's guide. EPA Contract 68-D3-0034, Source Receptor Analysis Branch, Research Triangle Park, NC, 16 pp.

——, cited 1999: National Emissions Inventory 1999 Version 1. Technology Transfer Network, Washington, DC. [Available online at www.epa.gov/ttn/chief/net/index.html#1999.]

——, cited 2003a: The Mobile 5b Vehicle Emissions Modeling Software. Office of Transportation and Air Quality. [Available online at www.epa.gov/otaq/m5.htm.]

——, cited 2003b: Air Quality System (AQS) database of ambient air quality data. Technology Transfer Network, Washington, DC. [Available online at www.epa.gov/ttn/airs/airsaqs/archive%20data/downloadaqsdata.htm.]

EPA Region III, cited 2003: Mid-Atlantic Air Protection, Ozone Action Days. [Available online at www.epa.gov/reg3artd/airquality/actdays.htm.]

Federal Register, 1971: National primary and secondary ambient air quality standards. *Fed. Regist.,* **36,** 8186–8201.

——, 1997: National ambient air quality standards for ozone, Final rule 40 CFR Part 50, July 18, 1997. *Fed. Regist.,* **62,** 1–37.

Gery, M. W., and Coauthors, 1989: A photochemical kinetics mechanism for urban and regional scale computer models. *J. Geophys. Res.,* **94,** 12 295–12 356.

Grell, G. A., J. Dudhia, and D. R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech. Note NCAR/TN-398+STR, NCAR, Boulder, CO, 122 pp.

Haagen-Smit, A. J., E. F. Darley, M. Zaitlin, H. Hull, and W. Noble, 1951: Investigation of injury to plants from air pollution in the Los Angeles area. *Plant Physiol.,* **27,** 18–24.

Hanna, S. R., G. E. Moore, and M. E. Fernau, 1996: Evaluation of photochemical grid models (UAM-IV, UAM-V, and the ROM/UAM-IV couple) using data from the Lake Michigan Ozone Study (LMOS). *Atmos. Environ.,* **30,** 3265–3279.

Heck, W. W., O. C. Taylor, R. Adams, J. Miller, E. Preston, and L. Weinstein, 1982: Assessment of crop loss from ozone. *J. Air Pollut. Control Assoc.,* **32,** 353–361.

Hobbs, P. V., 2000: *Introduction to Atmospheric Chemistry.* Cambridge University Press, 262 pp.

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

Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.,* **124,** 2322–2339.

Houyoux, M. R., J. M. Vukovich, C. J. Coats Jr., N. W. Wheeler, and P. S. Kasibhatla, 2000: Emission inventory development and processing for the seasonal model for regional air quality (SMRAQ) project. *J. Geophys. Res.,* **105** (D7), 9079–9090.

Hubbard, M. C., and W. G. Cobourn, 1997: Development of a regression model to forecast ground-level ozone concentrations in Louisville, Kentucky. *Atmos. Environ.,* **32,** 2637–2647.

Jakobs, H. J., S. Tilmes, A. Heidegger, K. Nester, and G. Smiatek, 2001: Short-term ozone forecasting with a network model system during summer 1999. *J. Atmos. Chem.,* **42,** 23–40.

Janjic, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, vis-

cous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.,* **122,** 927–945.

Kain, J. S., and J. M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain–Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models,* K. A. Emanuel and D. J. Raymond, Eds., Amer. Meteor. Soc., 165–170.

Kalnay, E., S. Lord, and R. D. McPherson, 1998: Maturity of operational numerical weather prediction: Medium range. *Bull. Amer. Meteor. Soc.,* **79,** 2753–2769.

Kasibhatla, P., and W. L. Chameides, 2000: Seasonal modeling of regional ozone pollution in the eastern United States. *Geophys. Res. Lett.,* **27,** 1415–1418.

——, ——, B. Duncan, M. Houyoux, C. Jang, R. Mathur, T. Odman, and A. Xiu, 1997: Impact of inert organic nitrate formation on ground-level ozone in a regional air quality model using the Carbon Bond Mechanism 4. *Geophys. Res. Lett.,* **24,** 3205–3208.

Knapp, K. G., B. B. Balsley, M. L. Jensen, H. P. Hanson, and J. W. Birks, 1998: Observations of the transport of polluted air masses from the northeastern United States to Cape Sable Island, Nova Scotia, Canada, during the 1993 NARE summer intensive. *J. Geophys. Res.,* **103,** 13 399–13 411.

Lippman, M., 1989: Health effects of ozone: A critical review. *J. Air Waste Manage. Assoc.,* **39,** 672–695.

Liu, P.-W. G., and R. Johnson, 2002: Forecasting peak daily ozone levels—1. A regression with time series errors model having a principal component trigger to fit 1991 ozone levels. *J. Air Waste Manage. Assoc.,* **52,** 1064–1074.

Liu, S. C., M. Trainer, F. C. Fehsenfeld, D. D. Parrish, E. J. Williams, D. W. Fahey, G. Hubler, and P. C. Murphy, 1987: Ozone production in the rural troposphere and the implications for regional and global ozone distributions. *J. Geophys. Res.,* **92,** 4191–4207.

Logan, J. A., 1999: An analysis of ozonesonde data for the troposphere: Recommendations for testing 3-D models, and development of a gridded climatology for tropospheric ozone. *J. Geophys. Res.,* **104,** 16 115–16 149.

Madronich, S., 1987: Photodissociation in the atmosphere, 1. Actinic flux and the effects of ground reflections and clouds. *J. Geophys. Res.,* **92** (D8), 9740–9752.

Manins, P., Ed., 2001: Air quality forecasting for Australia's major cities, final report. Australian Department of Environment and Heritage, CSIRO Atmospheric Research, 341 pp.

McHenry, J. N., cited 1999: Recommended hallmarks of NSF's environmental research priorities for the coming decade. *Text of Speech before the National Science Board, Task Force on the Environment, Town Hall Meeting,* National Science Foundation, March 7, 1999, Arlington, VA. [Available online at www.baronams.com/pub_files/index.html.]

——, and F. S. Binkowski, 1996: Development and testing of a Betts–Miller based shallow convective cloud chemistry scheme for incorporation into multiscale air quality models. Preprints, *Ninth Joint Conf. on Applications of Air Pollution Meteorology with the Air and Waste Management Association,* Atlanta, GA, Amer. Meteor. Soc., 254–259.

——, and C. J. Coats, 2003: Improved representation of cloud/actinic flux interaction in multiscale photochemical models. Preprints, *Fifth Conf. on Atmospheric Chemistry: Gases, Aerosols, and Clouds,* Long Beach, CA, Amer. Meteor. Soc., 83d Annual Meeting CD-ROM, 2.1.

——, J. S. Kain, and J. Pleim, 1996: Development and testing of a Kain–Fritsch based deep convective cloud chemistry scheme for incorporation into multiscale air quality models. Preprints, *Ninth Joint Conf. on Applications of Air Pollution Meteorology with the Air and Waste Management Association,* Atlanta, GA, Amer. Meteor. Soc., 207–212.

——, N. Seaman, C. J. Coats, A. Lario-Gibbs, J. Vukovich, N. Wheeler, and E. Hayes, 1999: Real-time nested mesoscale forecasts of lower tropospheric ozone using a highly optimized coupled model numerical prediction system. Preprints, *Symp. on Interdisciplinary Issues in Atmospheric Chemistry,* Dallas, TX, Amer. Meteor. Soc., 125–128.

——, ——, ——, D. Stauffer, A. Lario-Gibbs, J. Vukovich, E. Hayes, and N. Wheeler, 2000: The NCSC-PSU numerical air quality prediction project: Initial evaluation, status, and prospects. Preprints, *Symp. on Atmospheric Chemistry Issues in the 21st Century,* Long Beach, CA, Amer. Meteor. Soc., 95–102.

——, C. J. Coats, B. Cameron, J. Vukovich, A. Trayanov, and T. Smith, 2001: High-resolution real-time ozone forecasts for the August–September Texas AQS-2000 (Houston) field study: Forecast process and preliminary evaluation. Preprints, *Millenium Symp. on Atmospheric Chemistry,* Albuquerque, NM, Amer. Meteor. Soc., 186–193.

Moran, M. D., and Coauthors, 1997: An overview of CEPS1.0, Version 1.0 of the Canadian Emissions Processing System for regional-scale air quality models. *Proc. Seventh AWMA Emission Inventory Symp.,* Research Triangle Park, NC, Air and Waste Management Association, 95–106.

Murphy, A. H., and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.,* **115,** 1330–1338.

National Research Council Board on Atmospheric Sciences and Climate, 1998: *The Atmospheric Sciences:*

Entering the Twenty-First Century. National Academy Press, 364 pp.

NOAA, cited 2003a: Chemical Weather Research and Development Program. NOAA Forecast Systems Laboratory, Boulder, CO. [Available online at www-frd.fsl.noaa.gov/aq.]

——, cited 2003b: New England Air Quality Study (NEAQS). NOAA Forecast Systems Laboratory, Boulder, CO. [Available online at www.al.noaa.gov/NEAQS.]

Oak Ridge National Laboratories, cited 2003: PVM: Parallel Virtual Machine. [Available online at www.epm.ornl.gov/pvm/.]

Odman, M. T., and C. L. Ingram, 1996: The Multiscale Air Quality Simulation Platform (MAQSIP): Source code documentation and validation. Environmental Programs, Rep. ENV-96TR002-v1.0, MCNC-North Carolina Supercomputing Center, Research Triangle Park, NC, 83 pp.

——, and A. G. Russell, 2000: Mass conservative coupling of non-hydrostatic meteorological models with air quality models. *Air Pollution Modeling and Its Application XIII,* S.-E. Gryning and E. Batchvarova, Eds., Kluwer, 651–660.

Pierce, T., E. Kinnee, and C. Geron, 1998: Development of a 1-km resolved vegetation cover database for regional air quality modeling. Preprints, *23d Conf. on Agricultural and Forest Meteorology,* Albuquerque, NM, Amer. Meteor. Soc., 329–333.

Pudykiewicz, J., A. Kallaur, and P. K. Smolarkiewicz, 1997: Semi-Lagrangian modeling of tropospheric ozone. *Tellus,* **49B,** 231–248.

——, and Coauthors, 2003: Operational air quality forecasting in Canada: Numerical model guidance for ground-level ozone and particulate matter. Preprints, *Fifth Conf. on Atmospheric Chemistry: Gases, Aerosols, and Clouds,* Long Beach, CA, Amer. Meteor. Soc., 83d Annual Meeting CD-ROM, 3.2.

Rao, S. T., I. G. Zurbenko, R. Neagu, P. S. Porter, J. Y. Ku, and R. F. Henry, 1997: Space and time scales in ambient ozone data. *Bull. Amer. Meteor. Soc.,* **78,** 2153–2165.

Rew, R. K., and G. P. Davis, 1990: NetCDF: An interface for scientific data access. *IEEE Comput. Graph. Appl.,* **10,** 76–82.

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

Ruiz-Suarez, J. C., and O. A. Mayora-Ibarra, 1995: Short-term ozone forecasting by artificial neural networks. *Adv. Eng. Software,* **23,** 143–149.

Russell, A., and R. Dennis, 2000: NARSTO critical review of photochemical models and modeling. *Atmos. Environ.,* **34,** 2283–2324.

Ryan, W. F., cited 2002a: Air quality forecast report—Philadelphia forecast area, 2001. [Available online at www.meteo.psu.edu/~wfryan/phl_2001_final_report.htm.]

——, cited 2002b: Summary of the 2001 mid-Atlantic ozone season. [Available online at www.atmos.umd.edu/~forecaster/episodes/summary2001_0222.htm.]

——, B. G. Doddridge, R. R. Dickerson, R. M. Morales, and K. A. Hallock, 1998: Pollutant transport during a regional O_3 episode in the mid-Atlantic states. *J. Air Waste Manage. Assoc.,* **48,** 786–797.

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

Seaman, N. L., and S. A. Michelson, 2000: Mesoscale meteorological structure of a high-ozone episode during the 1995 NARSTO-Northeast study. *J. Appl. Meteor.,* **39,** 384–398.

Seinfeld, J. H., 1988: Ozone air quality models: A critical review. *J. Air Pollut. Control Assoc.,* **38,** 616–645.

Sirois, A., J. Pudykiewicz, and A. Kallaur, 1999: A comparison between simulated and observed ozone mixing-ratios in eastern North America. *J. Geophys. Res.,* **104,** 21 397–21 423.

Steiner, C., M. Causley, and M. Yocke, 1994: Minerals Management Service Outer Continental Shelf activity database (MOAD), user's guide. U.S. Department of the Interior, Minerals Management Service Gulf of Mexico OCS Region, OCS Study MMS 94-0018, 22 pp.

Tesche, T. W., P. Georgopoulos, J. H. Seinfeld, F. Lurmann, and P. M. Roth, 1990: Improvements in procedures for evaluating photochemical models. CARB Rep. A832-103, California Air Resources Board, Sacramento, CA, 164 pp.

Van Aalst, R. M., and F. A. A. M. de Leeuw, Eds., 1997: National ozone forecasting systems and international data exchange in northwest Europe. Working Group on Data Exchange and Forecasting Ozone episodes in northwest Europe Tech. Rep. 9, European Environment Agency, Bilthoven, Netherlands, 50 pp.

Vukovich, F. M., and J. Fishman, 1986: The climatology of summertime O_3 and SO_2 (1997–1981). *Atmos. Environ.,* **20,** 2423–2433.

——, W. D. Bach Jr., B. W. Crissman, and W. J. King, 1977: On the relationship between high ozone in the rural surface layer and high pressure systems. *Atmos. Environ.,* **11,** 967–983.

Weisman, R. A., 1990: An observational study of warm season southern Appalachian lee troughs. Part I: Boundary layer circulation. *Mon. Wea. Rev.,* **118,** 950–963.

Wesley, M. L., 1989: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmos. Environ.,* **23,** 1293–1304.

Wolff, G. T., and P. J. Lioy, 1978: An empirical model for forecasting maximum daily ozone levels in the northeastern United States. *Atmos. Environ.,* **11,** 967–983.

WRF-Chemistry, cited 2004: Weather Research and Forecasting (WRF) Model. Working group 11: Atmospheric chemistry. [Available online at http://www.wrf-model.org/WG11/.]