JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 112, D12S14, doi:10.1029/2006JD007715, 2007

A detailed evaluation of the Eta-CMAQ forecast model performance for O_3, its related precursors, and meteorological parameters during the 2004 ICARTT study

Shaocai Yu,[1,2,3] Rohit Mathur,[1,2] Kenneth Schere,[1,2] Daiwen Kang,[1,2,3] Jonathan Pleim,[1,2] and Tanya L. Otte[1,2]

Received 29 June 2006; revised 26 January 2007; accepted 8 February 2007; published 24 May 2007.

[1] The Eta-Community Multiscale Air Quality (CMAQ) model's forecast performance for ozone (O_3), its precursors, and meteorological parameters has been assessed over the eastern United States with the observations obtained by aircraft, ship, ozonesonde, and lidar and two surface networks (AIRNOW and AIRMAP) during the 2004 International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) study. The results at the AIRNOW sites show that the model was able to reproduce the day-to-day variations of observed daily maximum 8-hour O_3 and captured the majority (73%) of observed daily maximum 8-hour O_3 within a factor of 1.5 with normalized mean bias of 22%. The model in general reproduced O_3 vertical distributions on most of the days at low altitudes, but consistent overestimations above ~6 km are evident because of a combination of effects related to the specifications of lateral boundary conditions from the Global Forecast System (GFS) as well as the model's coarse vertical resolution in the upper free troposphere. The model captured the vertical variation patterns of the observed values for other parameters (HNO_3, SO_2, NO_2, HCHO, and NO_y_sum (NO_y_sum = NO + NO_2 + HNO_3 + PAN)) with some exceptions, depending on the studied areas and air mass characteristics. The consistent underestimation of CO by ~30% from surface to high altitudes is partly attributed to the inadequate representation of the transport of pollution associated with Alaska forest fires from outside the domain. The model exhibited good performance for marine or continental clear airflows from the east/north/northwest/south and southwest flows influenced only by Boston city plumes but overestimation for southeast flows influenced by the long-range transport of urban plumes from both New York City and Boston.

Citation: Yu, S., R. Mathur, K. Schere, D. Kang, J. Pleim, and T. L. Otte (2007), A detailed evaluation of the Eta-CMAQ forecast model performance for O_3, its related precursors, and meteorological parameters during the 2004 ICARTT study, *J. Geophys. Res.*, *112*, D12S14, doi:10.1029/2006JD007715.

1. Introduction

[2] Ozone (O_3) pollution is a major concern in the United States since it can adversely affect human and ecosystem health. Tropospheric O_3 is generated in the presence of solar ultraviolet radiation through a complex series of photochemical reactions involving many volatile organic compounds (VOC) and nitrogen oxides (NO_x), which originate either from anthropogenic sources (e.g., industry and vehicle emissions) or biogenic sources (e.g., forest and soil). Harmful levels of O_3 concentrations are typically observed during high pressure, hot, sunny and stagnant atmospheric conditions at the locations with substantial VOC and NO_x concentrations. According to the revised 8-hour National Ambient Air Quality Standard (NAAQS) for O_3 (0.08 ppm) promulgated by the U.S. EPA in 1997, *U.S. Environmental Protection Agency* (*U.S. EPA*) [2005] estimated that about 160 million Americans are exposed annually to the daily maximum 8-hour O_3 concentrations that exceed this new NAAQS. Therefore it is desirable for local air quality agencies to accurately forecast ozone concentrations to alert the public of the onset, severity and duration of unhealthy air and to encourage people to help limit outdoor activities and reduce emissions-producing activities (e.g., reduce automobile usage).

[3] Real-time forecasting systems for O_3 with regional-scale air quality models have been developed and deployed for several years [*U.S. EPA*, 1999; *McHenry et al.*, 2004; *Cope et al.*, 2004; *McKeen et al.*, 2005; *Kang et al.*, 2005; *Otte et al.*, 2005]. *McKeen et al.* [2005] statistically eval-

[1]Atmospheric Sciences Modeling Division, Air Resources Laboratory, NOAA, Research Triangle Park, North Carolina, USA.
[2]U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA.
[3]On assignment from Science and Technology Corporation, Hampton, Virginia, USA.

0148-0227/07/2006JD007715$09.00

uated the real-time forecasts of O_3 from seven air quality forecast models over the eastern United States and southern Canada during the summer of 2004, and concluded that relative to any individual model, the ensembles, which were based on the mean and median of the seven models, had higher correlation coefficients, lower root-mean-square errors (RMSE), and better threshold statistics, pointing to ensemble modeling as a better real-time O_3 forecast tool.

[4] The regional air quality in New England was a focus of the 2004 International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) study. Two of the major goals of the 2004 ICARTT study were to link surface air quality with the important features of transport and chemistry that occur above the surface and to determine the relative importance of local pollution compared to long-range transport in shaping local air quality. As summarized by *Russell and Dennis* [2000], model evaluation studies for O_3 and its precursors are severely limited by a lack of data both aloft and at the surface. The performance requirements for current generation air quality models has increased significantly as a result of exceedingly complex issues that the models are now being used to simulate [*Mathur et al.*, 2005]. As an example, continuous applications of air quality models for operational forecasting have placed new demands on performance requirements for such models, which need to accurately simulate the pollutant distributions under varying dynamical and chemical conditions. The 2004 ICARTT experiment provided a comprehensive set of measurements of chemical constituents and meteorological variables, both from surface and aircraft based platforms, which can be used to examine in detail the performance of air quality models from a multipollutant perspective, both in terms of their surface concentrations as well as vertical structure. Such detailed information on model performance in turn helps in identifying deficiencies in existing models, and provides guidance for further model enhancements and consequently the development of more robust operational models. Additionally, since forecasts from operational air quality models are increasingly being used to provide in-field guidance to support the planning of field experiments [e.g., *McKeen et al.*, 2005], complementary analysis of the detailed 3-D structure of various pollutants from the models in conjunction with the field measurements can help improve understanding of regional pollution episodes. In this study, the National Weather Service's (NWS) operational mesoscale forecast Eta model is used to supply meteorological input to the Community Multiscale Air Quality (CMAQ) model (Eta-CMAQ model suite); the models are then used to provide predictions of O_3 and related chemical species in the forecast mode. The objectives of this study are two-fold. First, the temporal and spatial performance of the Eta-CMAQ forecast model for O_3 is evaluated against the observations from the Air Quality System (AQS) network over the eastern United States. Second, the ability of the Eta-CMAQ model to predict air quality and the meteorological conditions dictating episodes of high O_3 horizontally and vertically is comprehensively examined on the basis of the extensive measurements obtained by aircraft, ship, ozonesonde, and lidar during the 2004 ICARTT field experiment. Recommendations for further research and analysis in the pursuit of improved O_3 forecasts are also provided on the basis of the current evaluation. Note that the summer of 2004 in the eastern United States exhibited very few O_3 "episodes" or exceedances because of unusually cool and wet conditions (i.e., temperatures either below or much below normal, and precipitation either above or much above normal (http://www.ncdc.noaa.gov/oa/climate/research/2004/aug/aug04.html)), associated with continental polar air masses during July, and the influence of several hurricanes during August. Therefore the model performance presented here is probably not climatologically representative of summertime conditions, but is unique to the summer of 2004.

2. Description of the Eta-CMAQ Forecast Model and Observational Database

[5] The Eta-CMAQ air quality forecasting system [*Otte et al.*, 2005], created by linking the Eta model [*Rogers et al.*, 1996] and the CMAQ Modeling System [*Byun and Ching*, 1999; *Byun and Schere*, 2006], was applied over a domain encompassing the eastern United States (Figure 1c) during the summer 2004. The linkage of the two modeling systems is described in detail by *Otte et al.* [2005]. A series of postprocessors interpolates the Eta model output fields in the horizontal and in the vertical onto a coordinate structure and map projection that are compatible with CMAQ. In this application, both CMAQ and Eta model domains have horizontal grid spacing of 12 km. Twenty-two layers of variable thickness are specified on a sigma vertical coordinate system to resolve the atmosphere between the surface and 100 hPa. The thickness of layer 1 is about 38 m. The lateral boundary conditions are horizontally constant and are specified by continental "clean" profile for O_3 and other trace gases; the vertical variations are based on climatology [*Byun and Ching*, 1999]. To improve representation of O_3 in the free troposphere and possible effects related to stratospheric intrusion, the O_3 lateral boundary conditions above altitudes of 6 km were augmented using O_3 forecast results from NCEP's Global Forecast System (GFS). The Eta-CMAQ model forecast provides twice-daily 48-hour gridded O_3 predictions as air quality forecast guidance for the United States. The primary Eta-CMAQ model forecast for the next day is based on the current day's 1200 UTC Eta simulation cycle. The target forecast period is 0400 UTC (local midnight) to next day's 0300 UTC (local midnight). The current forecast cycle is initialized using the prediction from the previous forecast cycle [*Otte et al.*, 2005]. The emissions are projected to 2004 from the 2001 U.S. EPA national emission inventory [*Pouliot*, 2005]. The Carbon Bond chemical mechanism (version 4.2) has been used to represent photochemical reaction pathways.

[6] The hourly, near real-time observed O_3 data at 614 sites in the eastern United States are available from the U.S. EPA's Air Quality System (AQS) network (Figure 1), resulting in nearly 1.2 million total hourly O_3 observations for the study period (see Table 1). Four sites of Atmospheric Investigation, Regional Modeling, Analysis, and Prediction (AIRMAP) [*DeBell et al.*, 2004; *Mao and Talbot*, 2004] provided continuous measurements of O_3 and related photochemical species as well as meteorological parameters during the study; the sites include Castle Springs (CS) (43.73°N, 71.33°W), New Hampshire (NH), Isle of Schoals

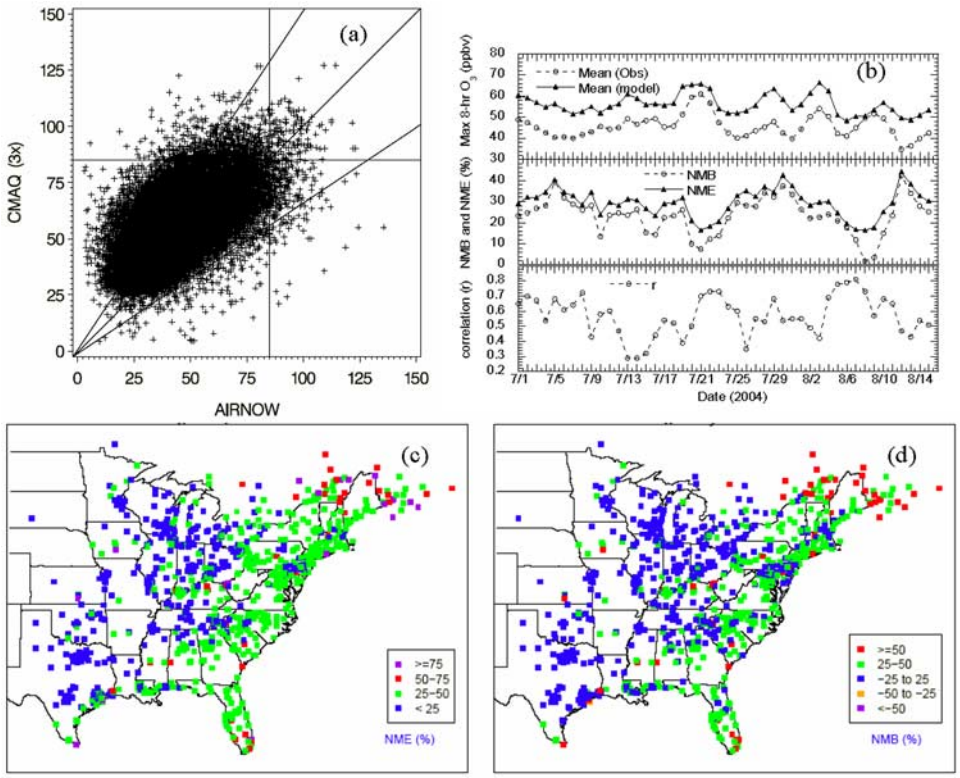

Figure 1. Comparison of the modeled and observed daily maximum 8-hour O_3 concentrations at the AIRNow monitoring sites. (a) Scatterplot (ppbv) (the 1:1, 1.5:1 and 1:1.5 lines are shown for reference); (b) daily variation of domain-wide mean, NME, NMB and correlation (r); and spatial distributions of (c) NME and (d) NMB during the period 1 July to 15 August 2004.

(IS) (42.99°N, 69.33°W), Maine, Mount Washington Observatory (MWO) (44.27°N, 71.30°W), NH, and Thompson Farm (TF) (43.11°N, 70.95°W), NH. From 1 July to 15 August 2004, measurements of vertical profiles of O_3, its related chemical species (CO, NO, NO_2, H_2O_2, CH_2O, HNO_3, SO_2, PAN, isoprene, toluene), and meteorological parameters (liquid water content, water vapor, temperature, wind speed and direction, and pressure) were carried out by instrumented aircraft (NOAA P-3 and NASA DC-8), ozonesonde and ship-based lidar deployed as part of the 2004 ICARTT field experiment. The detailed instrumentation and protocols for measurements are described at http://www.esrl.noaa.gov/csd/ICARTT/fieldoperations/. The flight tracks of P-3, DC-8, ship and locations where daily ozonesondes were launched are presented in Figure 2. The model performance during the period of 1 July to 15 August 2004 is examined in this study on the basis of the 1200 UTC model run for the target forecast period.

3. Results and Discussion

3.1. Spatial and Temporal Evaluation of O_3 Over the Eastern U.S. Domain at the AQS Sites

[7] Table 1 summarizes the evaluation results for the hourly, daily maximum 1-hour and daily maximum 8-hour O_3 concentrations for two cases; one using all data and the other only using an O_3 threshold of 40 ppbv. As can be seen, the Normalized Mean Bias (NMB) and Normalized Mean Error (NME) when only data with $O_3 > 40$ ppbv is considered range from 6.1 to 11.9% and 18.2 to 21.5%, respectively, much lower than those when all data are considered, indicating that the overestimations in the low O_3 concentration range contribute significantly to the over-

Table 1. Operational Evaluation for O_3 Concentrations on the Basis of the AQS Data Over the Eastern United States

		Domain Mean, ppb						
	Number	Obs	Model	RMSE, ppbv	MB, ppbv	NMB, %	NME, %	R
				All data				
Hourly	1,170,000	28.1	39.7	19.4	11.5	40.9	54.8	0.64
Max 1-hour	40,189	51.9	60.4	16.9	8.5	16.4	25.3	0.61
Max 8-hour	40,189	45.7	56.1	16.6	10.4	22.6	28.8	0.60
				For $O_3 > 40$ ppbv				
Hourly	297,981	52.8	56.0	14.9	3.2	6.1	21.5	0.42
Max 1-hour	24,943	61.9	66.2	14.8	4.3	7.0	18.2	0.47
Max 8-hour	24,943	54.9	61.5	13.9	6.5	11.9	19.7	0.45

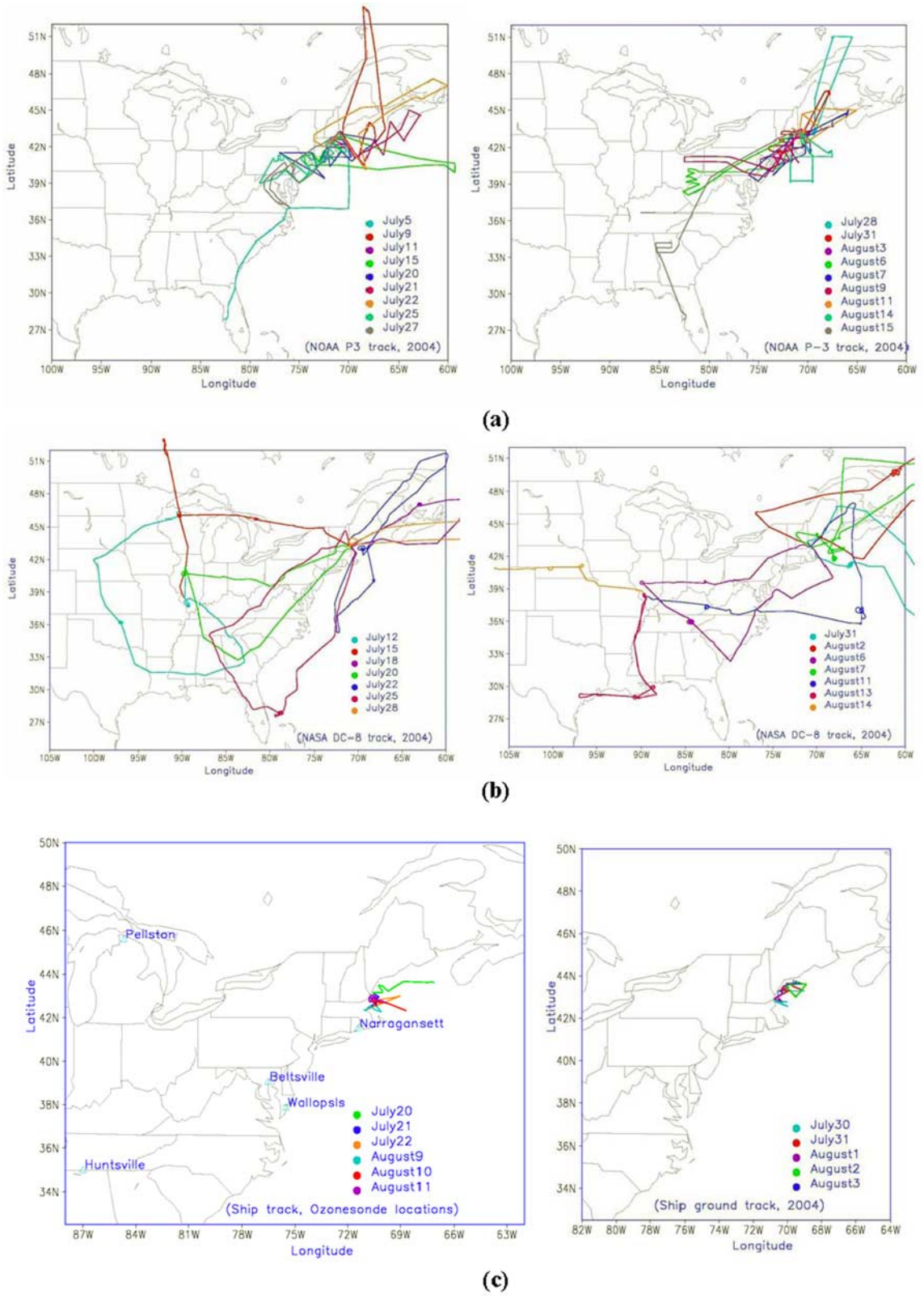

Figure 2. Tracks of (a) P-3, (b) DC-8 and (c) ship tracks and ozonesonde locations.

all overestimations. The recommended performance criteria for O_3 by *U.S. EPA* [1991] are mean normalized bias ±5 to ±15%; mean normalized gross error 30% to 35%; unpaired peak estimation accuracy: ±15 to ±20%. Table 1 shows that for the case with a 40 ppbv threshold for O_3, the Normalized Mean Bias (NMB) and Normalized Mean Error (NME) values for maximum 1-hour (maximum 8-hour) O_3 are 7.0% (11.9%) and 18.2% (19.7%), respectively, close to the

Table 2. Flight Observation Summary for WP-3 Aircraft

Date	Observation Summary[a]
9 Jul	Shortly after takeoff at 1535 UTC, the P-3 encountered the fresh Boston plume, which was not well mixed and confined to low altitudes, above the ocean close to Cape Ann with the CO and SO_2 concentrations as high as 200 ppbv and 4 ppbv, respectively.
11 Jul	On 11 July at night, after takeoff at 2300 UTC, the P-3 encountered the Boston plume as the aircraft turned to the south flying to the west of Boston (see Figure 2) and intercepted a significant biomass-burning layer at ~2700 m in the later (~2400 UTC) flight to southeast of Boston during a spiral climb.
15 Jul	A very well defined NY city plume was sampled extensively by flight by making cross sections and spirals through the plume after ~1830 UTC for understanding of ozone evolution in an urban plume.
20 Jul	The aircraft encountered a fresh NY city plume immediately downwind of NY, which reached up to ~1800 m, and the biomass-burning plume at ~3000 m between 1700 and 1800 UTC. The aging NY plume was intercepted over the Gulf of Maine with CO reaching 270 ppbv.
21 Jul	An aged (1.5 to 2.5 days old) NY city plume over the Gulf of Maine was intercepted in the lower troposphere (~270 to ~1000 m) in the outbound northeasterly flight between 1430 and 1530 UTC. A biomass-burning plume above Cape Cod at ~3000 m was encountered at 2000 UTC.
22 Jul	A progressively more aged NY city plume from 21 July over the Bay of Fundy reaching beyond Cape Breton and Prince Edward Island was sampled.
25 Jul	The flight looked at the outflow of Sunday emissions from Boston and NY cities and the downwind plume with high SO_2 (>4 ppb) for Montour power plant in central Pennsylvania in the NW of the power plant during 1600 and 1730 UTC. There was a widespread signature of biomass burning.
27 Jul	The flight aimed to look at the pollution accumulation ahead of the cold front. The pollution upwind and downwind of the Washington and Baltimore metropolitan area was sampled by P-3 during 1730 to 1830 UTC with very high SO_2 (>5 ppb) and CO (>180 ppb) concentrations.
28 Jul	The conveyor belt outflow without a clear signature of the anthropogenic pollution export was sampled with low concentrations for all species as there was a stationary front. There was a biomass burning plume over Quebec beyond of the model domain.
31 Jul	On 31 July at night, the P-3 encountered the NYC plume at ~1 August, 0130 UTC, with high SO_2 (>5 ppb) and CO (>180 ppb).
3 Aug	On 3 August at night, the P-3 encountered the NYC plume over southwestern Connecticut at ~700 m during 0400 to 0430 UTC with high SO_2 (>5 ppb) and CO (>180 ppb).
6 Aug	The P-3 encountered the plume of Ohio valley power plants at ~1000 m during 1530 and 2030 UTC with high SO_2 (>5 ppbv) and low O_3 (<60 ppbv) concentrations.
7 Aug	On 7 August at night, the P-3 encountered the NYC and Boston plumes at ~700 m during 8 August, 0100 to 0430 UTC, with high SO_2 (>5 ppb) and CO (>180 ppb).
9 Aug	On 9 August at night, the P-3 encountered the plume of Ohio valley power plants at ~1000 m during 10 August, 0030 and 0330 UTC, and the NYC and Boston plumes at ~700 m during 10 August, 0430 to 0630 UTC, with high SO_2 (>5 ppb) and CO (>180 ppb).
11 Aug	On 11 August at night, the P-3 encountered NYC plume at ~700 m during 0230 to 1030 UTC with high SO_2 (>5 ppb) and CO (>180 ppb).
14 Aug	It was a cloudy day across the whole eastern United States under the influence of Hurricane Charley. The P-3 encountered NYC plume at ~200 m during 1630 to 1730 UTC with high SO_2 (>5 ppb) and CO (>180 ppb).
15 Aug	It was still cloudy along eastern coast. The P-3 encountered Atlanta plume at ~700 m during 1820 to 2000 UTC with high SO_2 (>5 ppb) and CO (>180 ppb).

[a]Based on flight information presented at http://www.esrl.noaa.gov/csd/ICARTT/fieldoperations/fomp.shtml.

Table 3. Flight Observation Summary for DC-8 Aircraft

Date	Observation Summary[a]
15 Jul 2004	Characterization of Asian pollution, Alaskan fires, and anthropogenic pollution. The flight occurred above the cloud at ~8 km during 1730 and 2000 UTC in the NW of Boston city.
18 Jul 2004	Characterization of North American pollution outflow, possible characterization of Alaskan fires, and a flyby over the NOAA ship *Ronald H. Brown* in the NE of Boston city.
20 Jul 2004	Characterization of smoke from Alaskan fires transported over the United States and boundary layer pollution over the southeast and midwest. There were some scattered clouds.
22 Jul 2004	Sampling polluted boundary layer outflow along the eastern seaboard both to the north and south of Pease. Intercomparison between the NASA DC-8 and NOAA WP-3D aircraft.
25 Jul 2004	Convective outflow from southeast United States and map Ohio River Valley emissions in northerly flow under flight. The DC-8 flew above the clouds at 8 km during 1830 to 1930 UTC.
28 Jul 2004	Sample the structure and chemical evolution of the U.S. continental outflow out over the Atlantic Ocean. Most of time was beyond the model domain.
31 Jul 2004	Aged air sampling/recirculation, low-level outflow, P-3 intercomparison, and possible Asian influences.
2 Aug 2004	Sample low-level North American outflow and aged air pollution aloft and conduct a flyby over the ground Appledore Island air quality station.
6 Aug 2004	Flew over the Ohio River Valley. The DC-8 flew above the clouds at 10 km during1230 to 1330 UTC but below the cloud at 200 m during 1930 and 2000 UTC.
7 Aug 2004	Sample North American outflow, a stratospheric intrusion, and perform P-3 intercomparison.
11 Aug 2004	North America (NA) and warm conveyor belt (WCB) lifting, frontal crossing and low-level pollution.
13 Aug 2004	Outflow from major industrial cities (Houston and New Orleans) with clear skies for most of time except the period of 2130 to 2200 UTC.
14 Aug 2004	Flight above the cloud over Missouri-Kansas during 1900 and 2000 UTC.

[a]Based on flight information presented at http://www.esrl.noaa.gov/csd/ICARTT/fieldoperations/fomp.shtml.

performance criteria for the unpaired peak O_3. Additional insight into the air quality forecast (AQF) modeling system's positive bias (over prediction) and error (scatter) can be gained from Figure 1a, which shows that the model reproduced the majority (73.1%) of the observed daily maximum 8-hour O_3 concentrations within a factor of 1.5, but generally overestimated the observations in the low O_3 concentration range (<50 ppbv). It is believed that the points in the low O_3 concentration ranges below the background level reflect depletion of O_3 from deposition or titration by NO [*Lin et al.*, 2000]. The overestimation of the observations in the low O_3 concentration ranges could be indicative of titration by NO in urban plumes that the model does not resolve because most of the AQS sites are located in urban areas. This is further supported by the fact that most of sites with MB (mean bias) > 40 ppb in the low O_3 concentration range (<50 ppb) for the maximum 8-hour O_3 in Figure 1a are located within major metropolitan regions and along the Washington, D.C./New York City/Boston urban corridor (not shown). In order to investigate the AQF system's performance over time, the values of mean, NMB and correlation coefficients were calculated (domain wide averages) and plotted as daily time series for the daily maximum 8-hour O_3 concentrations (Figure 1b). Although the forecasts tracked the general temporal pattern well, the overestimation discussed above was prevalent throughout the four month period. The domain-wide mean values of NMB and NME during the ICARTT period for maximum 8-hour O_3 are 22.6% and 28.8%, respectively. The model had the best performance on 8 August (NMB = 1.9%, correlation coefficient (r) = 0.73) and the worst performance on 12 August (NMB = 42.4%, r = 0.47). A close inspection of the synoptic-scale meteorological conditions (not shown) reveals that on 8 August, the majority of the domain was dominated by high pressure and clear sky (conditions that are conducive to O_3 formation), whereas on 12 August, an active cold front stretched from the north to south accompanied by convective cloud cover and precipitation through the domain under low pressure. As shown by the diagnostic analysis [*Mathur et al.*, 2004], the significant overestimation in areas of cloud cover is mainly caused by the unrealistic vertical transport of excessive amounts of high O_3 concentrations near the tropopause to the ground associated with downward entrainment in CMAQ's convective cloud scheme. Spatially, the model performed better over the western region with NMB of ±25% than eastern coastal region of the domain with NMB > 25% (Figures 1c and 1d). The largest overestimation of the observed daily maximum 8-hour O_3 concentrations was in the northeast (NMB> +50% and NME > 50%), mainly caused by very low observed O_3 concentrations, which typically coincide with nonconducive meteorological conditions (i.e., cloud cover, precipitation and cool temperatures). Biases and errors associated with the maximum 1-hour O_3 (not shown) follow a similar pattern to the maximum 8-hour O_3.

3.2. Evaluation of Vertical Profiles for O_3, Its Related Species, and Meteorological Parameters

[8] Comparisons of modeled vertical profiles with aircraft, ozonesonde, and ship-based lidar observed vertical profiles provide an assessment of the ability of the model to simulate the vertical structure of air pollutants and meteorological fields. Following *Mathur et al.* [2005], modeled results were extracted by ''flying'' the aircraft through the 3-D modeling domain by mapping the locations of the aircraft to the model grid indices (column, row, and layer). Hourly resolved model outputs were linearly interpolated to the corresponding observational times. The flight tracks of aircraft show that measurements onboard the P-3 cover a regional area over the northeast around New York and Boston (Figure 2a) from 0 to ~5 km altitudes, whereas the DC-8 aircraft covers a broader regional area over the eastern United States (see Figure 2b) between 0 and 12 km altitudes. All DC-8 measurements were conducted in the daytime (~0700 to ~1900 LT), and P-3 also conducted most of its measurements during the daytime except on 11 and 31 July and 3, 7, 9, and 11 August when the P-3 measurements were conducted at the night (~2000 to ~0600 LT). Tables 2 and 3 present missions for the different flights and summaries of conditions encountered. To compare the modeled and observed vertical profiles, the observed and modeled data were grouped according to the model layer for each day and each flight. Thus these vertical profiles may be regarded as representing the mean conditions along the flight track for each day. Figures 3–7 present modeled and observed vertical daily profiles for O_3, its related species, and meteorological parameters during the ICARTT period. The temporal variations of modeled and observed JNO_2 (photolysis rates of NO_2) along the flight tracks for the daytime only are shown in Figures 8 and 9.

[9] As shown in Figures 3a and 3b, the model generally reproduced the observed O_3 vertical structure on most days with the best performance on 25, 27, and 31 July and 6, 7, 9, and 11 August for the P-3 measurements and 18 and 22 July and 6, 7, 13, and 14 August for the DC-8 measurements, although it tended to overestimate in the upper layers, especially for DC-8 observations at altitudes >6 km. Noticeable among these are generally the better model performance during nighttime (11 and 31 July and 3, 7, 9, and 11 August) relative to P-3 observations. A close inspection of the temporal variations of modeled and observed O_3 along the flight tracks (not shown) reveals that the modeled overestimations of observed O_3 at the low altitudes for most days in Figures 3a and 3b occurred over the ocean regions. Comparisons of time-height variations in O_3 structure along the ship tracks (Figure 3c) indicates that the model predicted more uniform vertical O_3 profiles than the observations, and the overestimations increase with altitude on the basis of the lidar measurements over the ocean off the coast of New Hampshire (NH) and Maine (see Figure 2c). A close inspection of the synoptic-scale meteorological conditions (not shown) reveals widespread cloud cover and precipitation over the ocean off the coast of New Hampshire (NH) and Maine for nights of 16 and 17 July. The unrealistic vertical transport of excessive amount of high O_3 concentrations near the tropopause to the ground by the CMAQ's convective cloud scheme mainly caused the significant overestimations over the ocean for nighttimes of 16 and 17 July as shown in Figure 3c. Comparisons of vertical profiles of median O_3 concentrations at five sites on the basis of ozonesonde observations reveal a consistent model overestimation above ~6 km (Figure 3d), although the model reproduces the O_3 vertical profile well at the low altitudes, especially at the Pellston site. As discussed before,

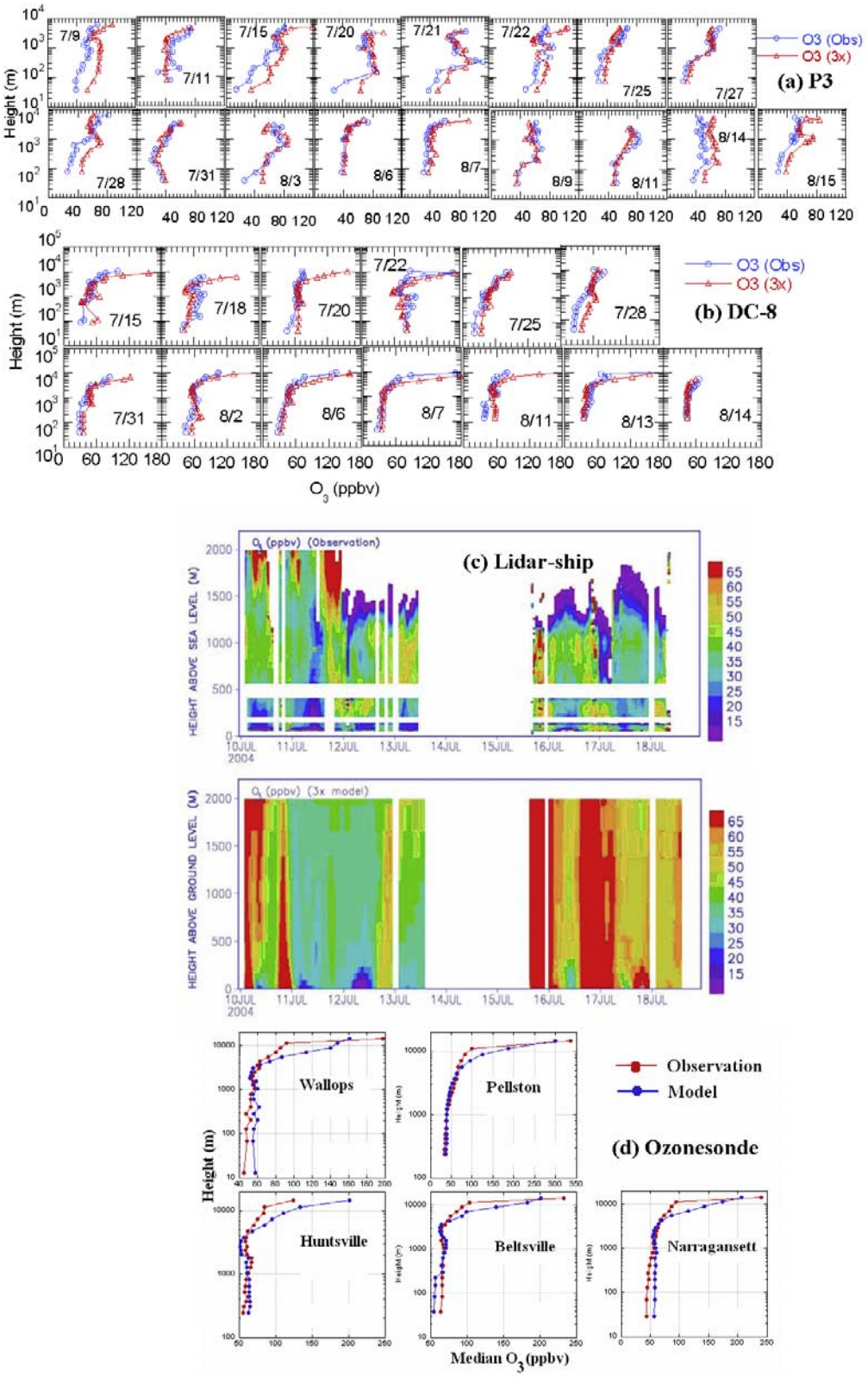

Figure 3. Comparison of vertical O_3 (ppbv) profiles for the model and observations from (a) P-3, (b) DC-8, (c) ship-Lidar and (d) ozonesonde during the ICARTT period.

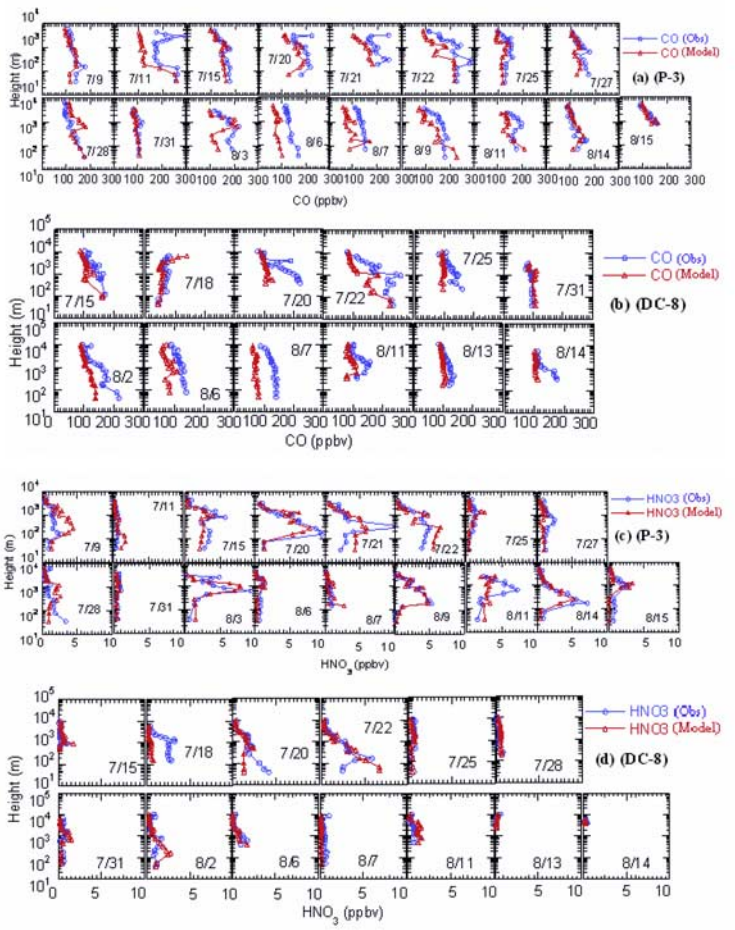

Figure 4. Comparison of vertical (a and b) CO and (c and d) HNO_3 (ppbv) profiles for the models and observations from P-3 (Figures 4a and 4c), and DC-8 (Figures 4b and 4d) during the ICARTT period.

this higher bias at the high altitude is attributed to the lateral boundary conditions derived by the GFS model and coarse model resolution in the free troposphere, and is consistent with the DC-8 comparisons shown in Figure 3b.

[10] The model's ability to simulate the vertical profiles for other parameters (CO, HNO_3, SO_2, NO, NO_2, HCHO, NO_y_sum (NO_y_sum = NO + NO_2 + HNO_3 + PAN) and NO_y (NO_y = NO + NO_2 + NO_3 + 2*N_2O_5 + HONO + HNO_3 + PNA + PAN + NTR)), as measured by the P-3 and DC-8 aircrafts, is illustrated in Figures 4–6. In general, the model captured the vertical variation patterns of the observed values for various species, with some exceptions, depending on the studied regions and air mass characteristics. Noticeable among these are the consistent underestimations for CO vertical profiles on most days except 15 and 31 July and 14 and 15 August for P-3 measurements, and 18 and 31 July for DC-8 measurements. The summary of Table 3 indicates that there was a widespread signature of biomass burning plume (i.e., the observed acetonitrile, the biomass burning plume tracer, was strongly enhanced) over the studied areas except these days, which were only significantly affected by the urban (New York, Boston or Washington and Baltimore) plumes. One of the reasons for this underestimation of CO can be attributed to the inadequate representation of the transport of pollution associated with biomass burning from outside the domain [*Mathur et al.*, 2005; *McKeen et al.*, 2002]. The significant underestimations of CO during 20 and 22 July 2005, further support this explanation as the aerosol index images from the TOMS satellite observations http://toms.gsfc.nasa.gov/) clearly show that the eastern United States was significantly influenced by pollutants from large Alaskan forest fires during these days. Tables 2 and 3 indicate that a progressively more aged NY city plume over the Bay of Fundy on 22 July with a widespread signature of biomass burning was sampled by both P-3 and DC-8, confirming the conclusion.

[11] Another noticeable discrepancy is the consistent underestimations of observed NO at altitudes greater than 6 km relative to DC-8 measurements (Figure 6b). This may be because the aircraft and lightning NO emissions are not included in the current model emission inventory. As shown in Figure 10, on 15 July, the DC-8 flight flew through convective clouds and there is a significant underestimation of observed NO at altitudes greater than 6 km relative to DC-8 measurements (Figure 6b). On 9, 21, 27, and 28 July, the P-3 encountered the fresh city plume (Boston or New

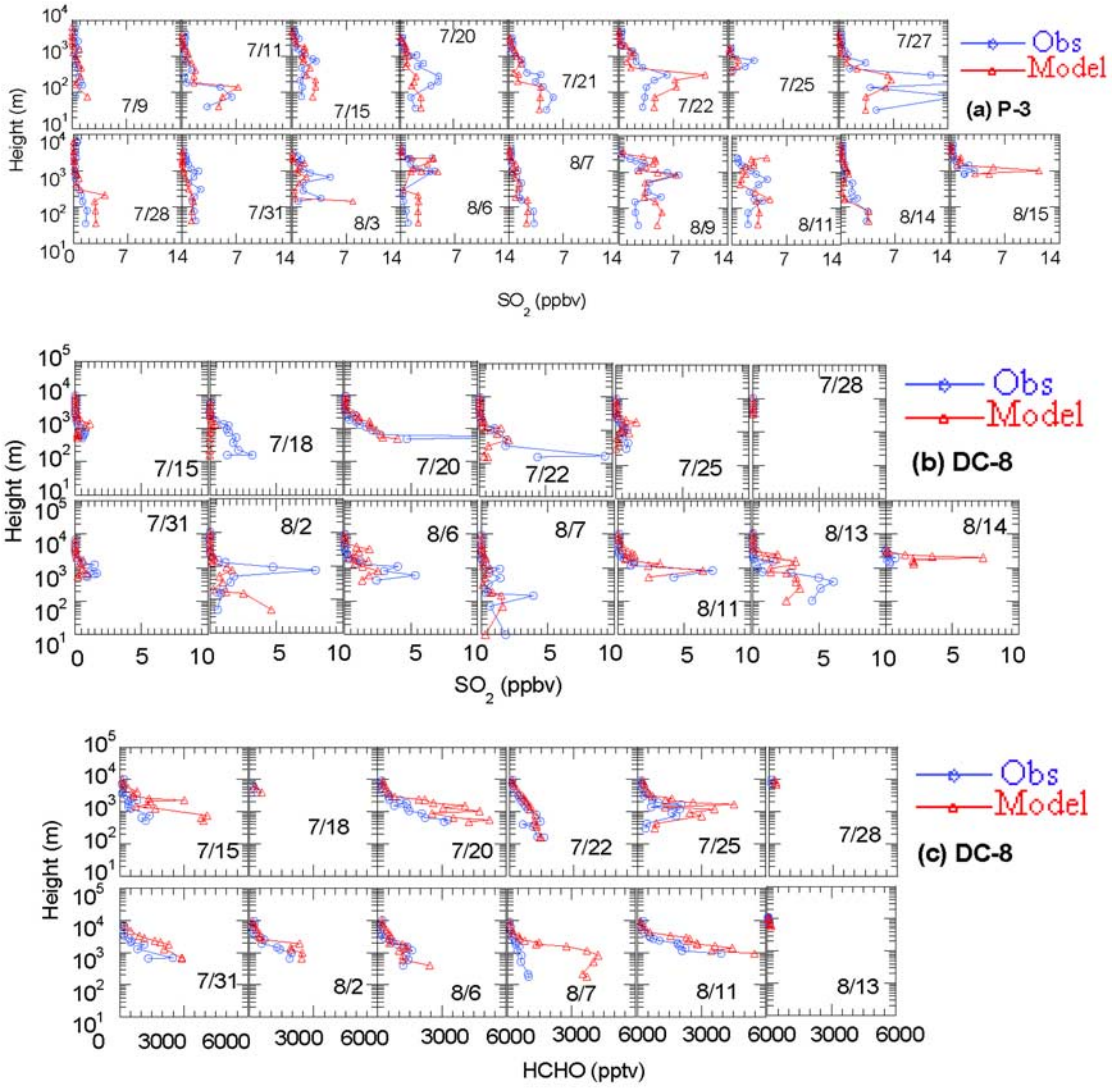

Figure 5. Same as Figure 4 but for SO_2 (ppbv) and HCHO (pptv).

York) shortly after takeoff as summarized in Table 3 with very high NO concentration at low altitudes. The model estimations also missed these high NO concentrations at low altitudes as shown in Figure 6a.

[12] As summarized in Table 3, the P-3 sampled the plume of Ohio Valley power plants at ~1000 m during 6 August from 1530 to 2030 UTC and 10 August from 0030 to 0330 UTC. Figures 3–6 show that the model reproduced the SO_2, NO, NO_2, HNO_3, O_3 and NOy_sum concentrations well relative to P-3 observations in the power plant plumes at this height for these two days. However, the model overestimated SO_2 in the NYC and Boston plumes at low altitudes <700 m for these two days. The modeled SO_2 concentrations are generally higher than the observations at the low altitude (<200 m) most of the time when the P-3 sampled the urban plumes of New York and Boston except 21 July and 7 August, indicating that the model may have overestimated some of emission sources of SO_2 from the New York and Boston areas.

[13] The point source emissions from power plants are often rich in SO_2 and NO_x and mobile sources (or urban plumes) are rich in CO and NO_x. On 27 July, the surface weather map showed convective activity associated with a surface cold front that stretched from the center of a surface low over the West Virginia–Pennsylvania state line to the southwest along the Appalachian Mountains with thunderstorms. There was pollution accumulation ahead of the cold front. The pollution upwind and downwind of the Washington and Baltimore metropolitan area between 600 and 2000 m altitudes was sampled by the P-3 during 1730 to 1830 UTC with very high SO_2 (>5 ppb), CO (>180 ppb), and HNO_3 (>3 ppb) but low O_3 (~60 ppb) concentrations. The model reproduced the low O_3 concentrations, but underestimated all other species (CO, HNO_3, NO, NO_2, NOy_sum) below 2 km for this pollution accumulation event ahead of the cold front as shown in Figures 3–6.

[14] The model shows good performance for HNO_3 most of the time except 9, 21, 22, 27, and 28 July and 11 August relative to P-3 observations and 18, 20, and 22 July relative

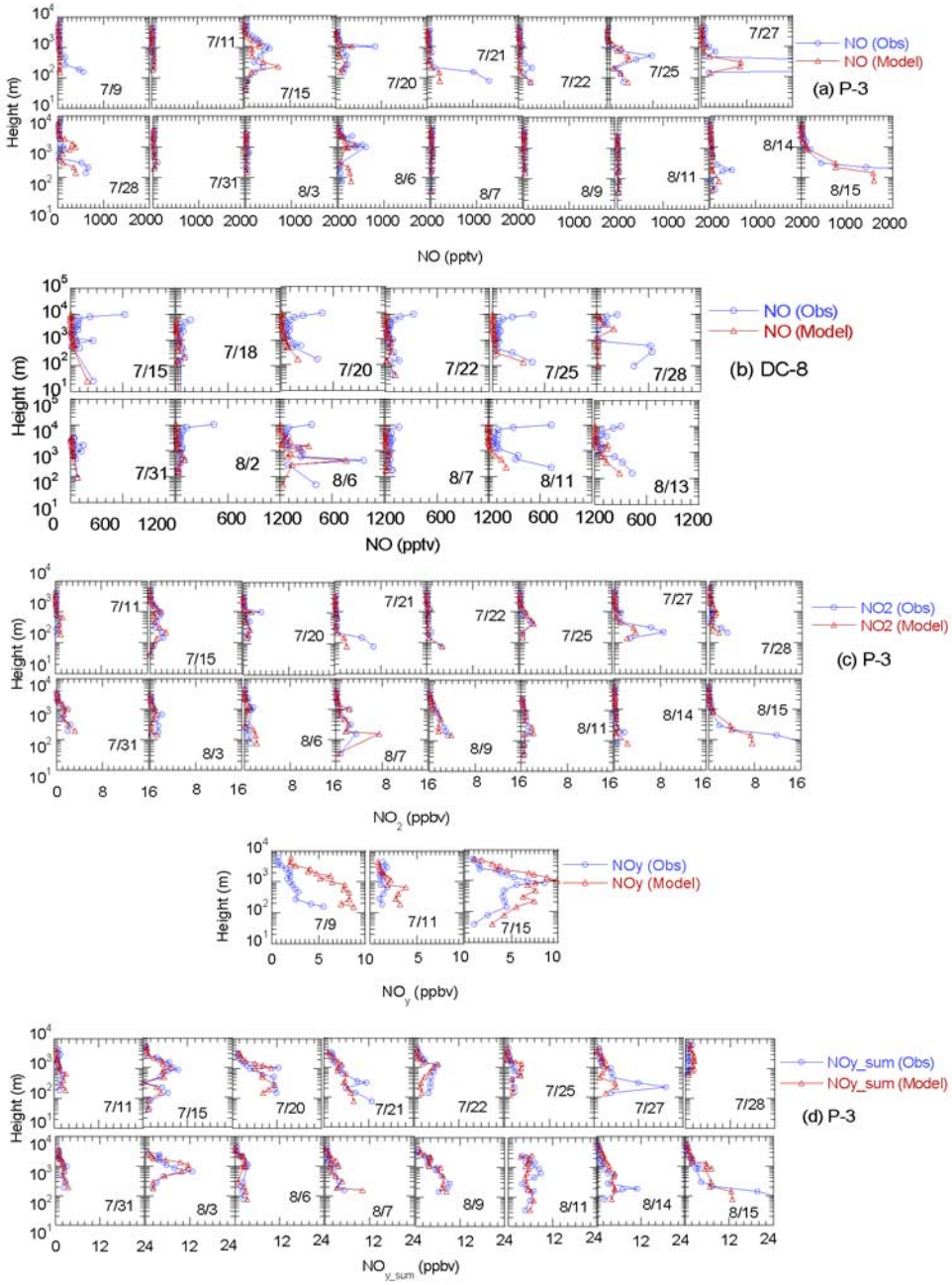

Figure 6. Comparison of daily vertical (a and b) NO, (c) NO_2, NO_y and (d) NO_{y_sum} profiles for the models and observations from the aircrafts P-3 (Figures 6a, 6c, and 6d) and DC-8 (Figure 6b) during the ICARTT period.

to DC-8 observations as shown in Figures 4c and 4d. The model overestimated the HNO_3 concentrations at the low altitudes in the air masses containing fresh plumes such as 9 and 21–22 July. The model performance for NO_y_sum is generally very good most of the time except 21 and 27 July and 14 and 15 August at the low altitudes as shown in Figure 6. NO_2 follows the same pattern of NO_y_sum for the model performance. The very good model performance of NOy_sum combined with consistent overestimations of NOy on 11 and 15 July in Figure 6 reveals that the model overestimated the sum of NO_3, N_2O_5, HONO, PNA and NTR. One of the possible reasons for these overestimations is the uncertainties associated with atmospheric sinks for the modeled terminal organic nitrate species represented by the lumped species called NTR in the CB IV chemical mechanism.

[15] Figure 7 reveals that the Eta model reproduced the vertical profiles of observed water vapor and wind speed very well most of the time and is in better agreement with the DC-8 observations. Specifically, the model consistently

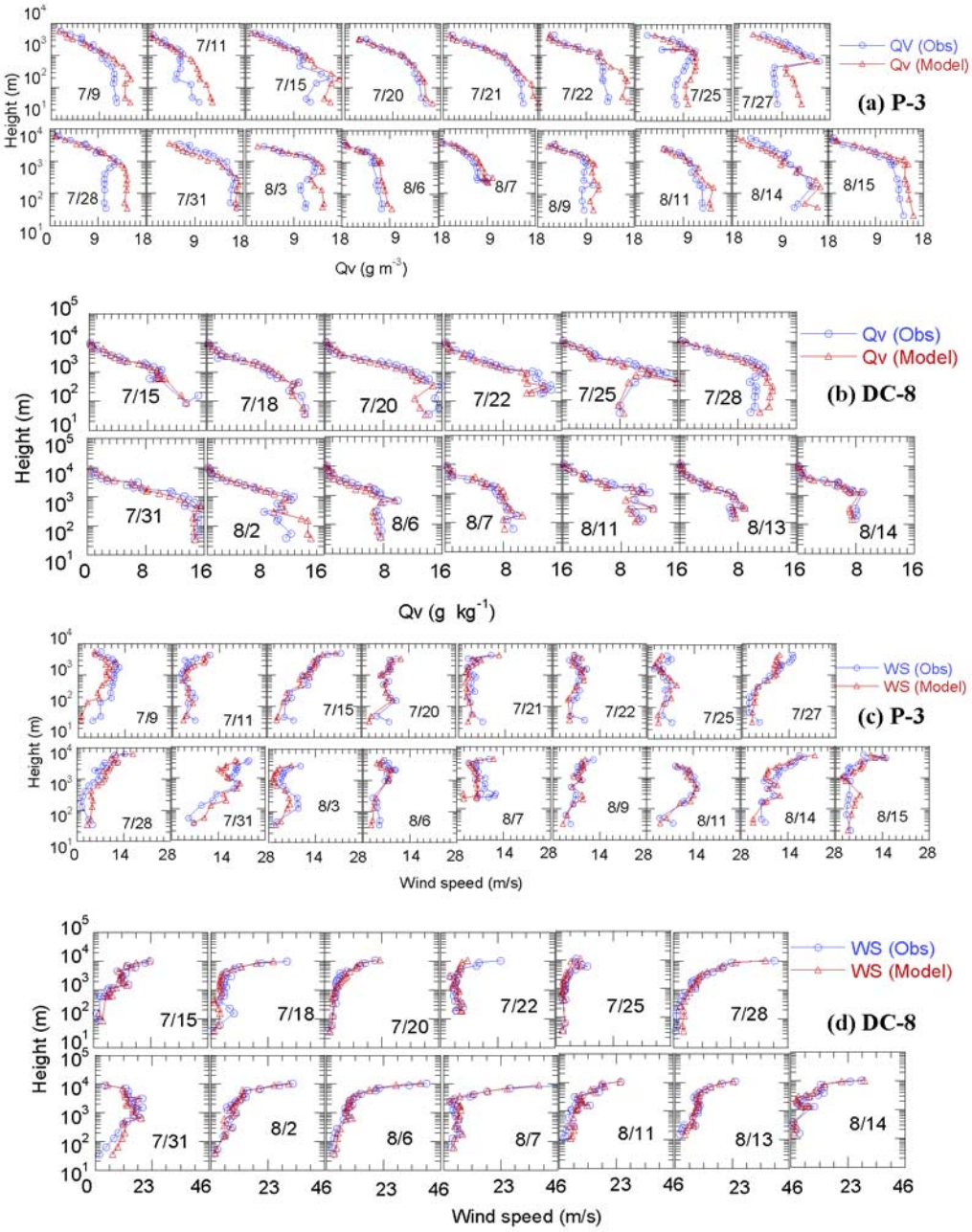

Figure 7. Same as Figure 4 but for water vapor and wind speed (WS).

overestimated water vapor at low altitudes relative to P-3 observations, especially on 27 and 28 July when there was a surface cold front across the northeastern domain. The model also overestimated water vapor at low altitudes relative to the DC-8 observations on 28 July, indicating that the model did not reproduce moisture well for the cold front system. The model seems to consistently underestimate the wind speed slightly in layer 1 most of the time except on 31 July and 3 August relative to the P-3 observations. The model also tracked the vertical variations of temperatures, pressures and wind directions very well most of the time (not shown).

[16] Cloud and aerosol can significantly affect photolysis rates of NO_2 (JNO_2) by enhancing and reducing the UV actinic flux, depending on their optical properties, solar zenith angle, and the position of the layer of interest relative to the observation point. Figures 8 and 9 show that the model captured the temporal and vertical variations of the observed JNO_2 very well for some periods but was weighted too low or too high for other periods along the flight tracks for each day. Upon a closer inspection of visible satellite images and aircraft observations, it is noted that the model generally captured the observed JNO_2 very well during the cloud-free periods, but underestimated the JNO_2 values by 20–90% (see Figure 9) when there was a

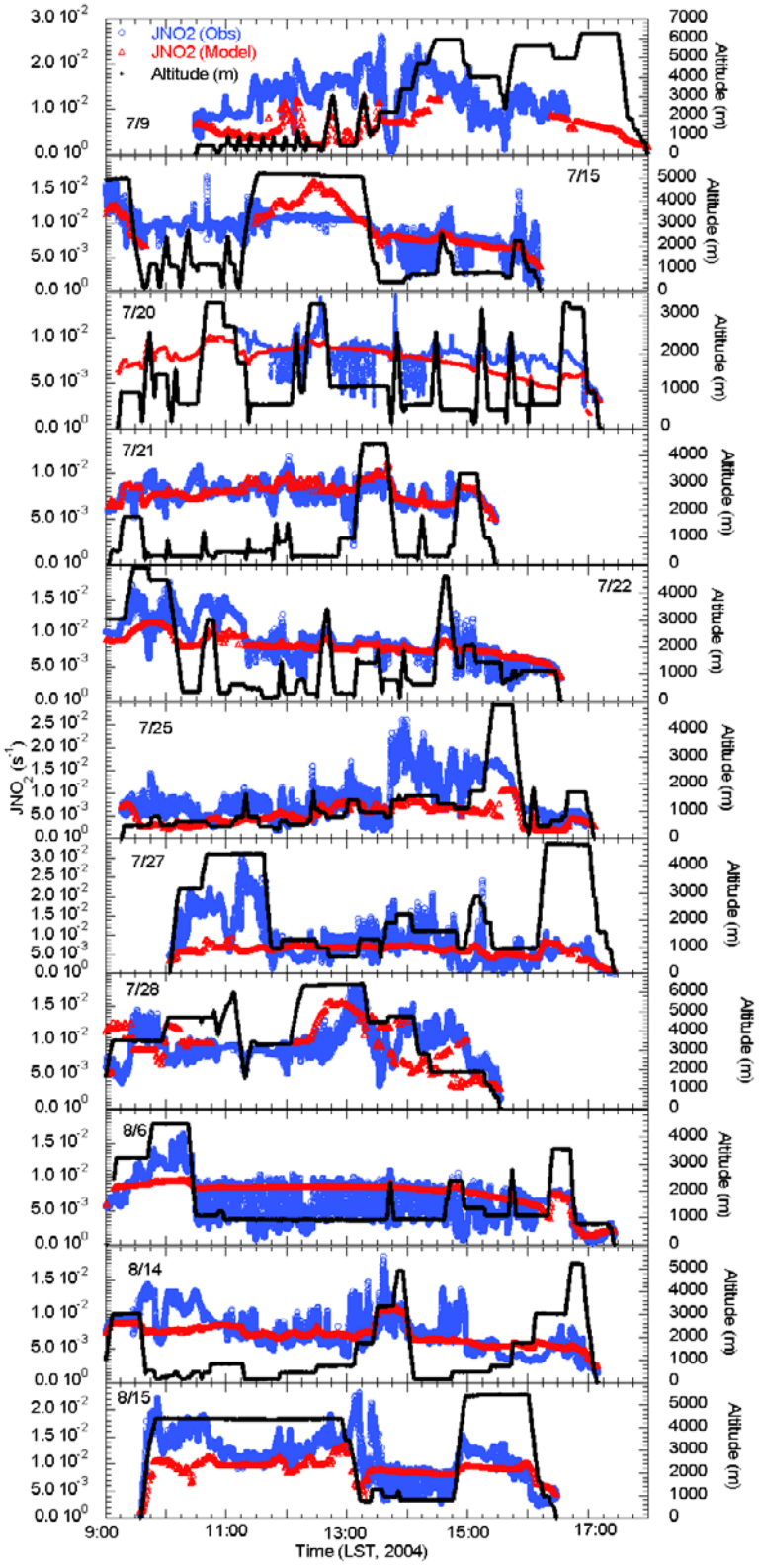

Figure 8. Time series comparison of the modeled and observed JNO_2 along the P-3 tracks for each day.

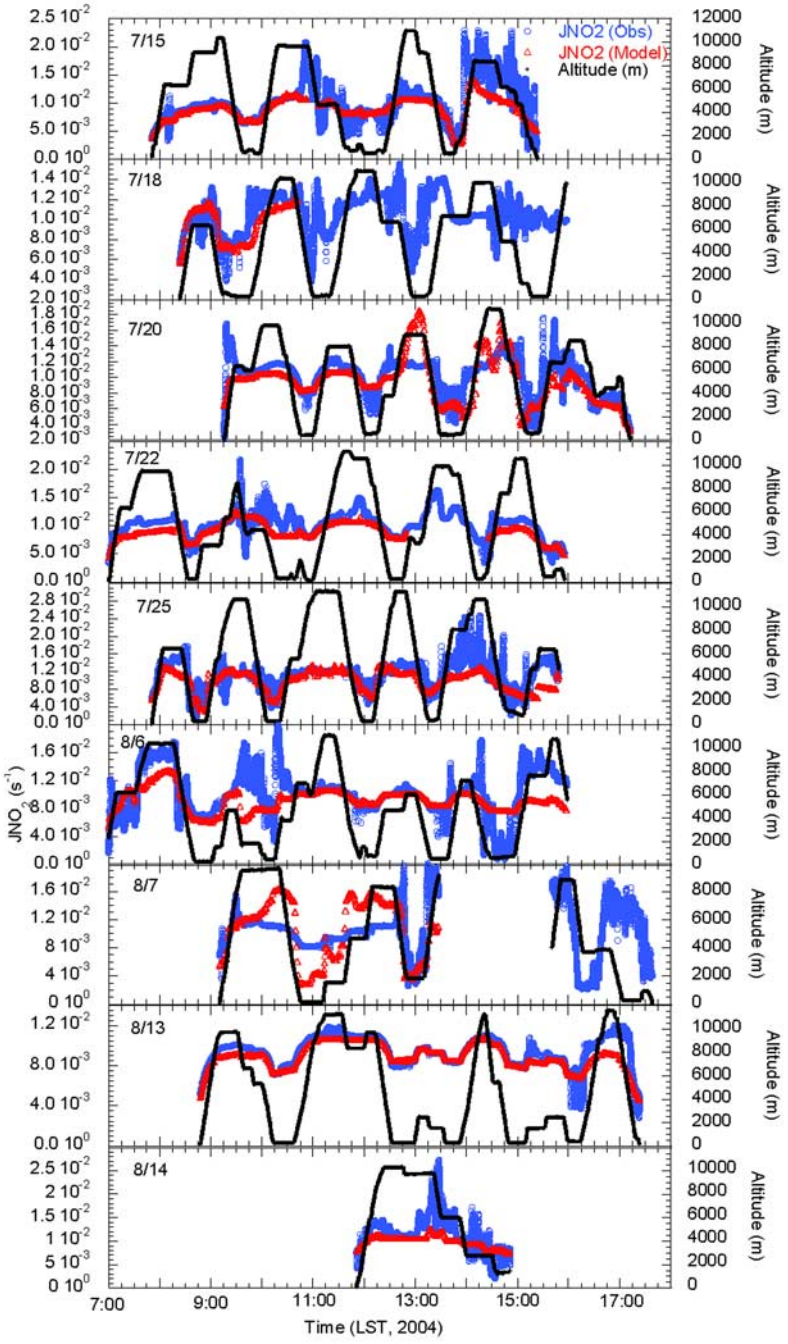

Figure 9. Same as Figure 8 but along the DC-8 tracks.

solid cloud deck below the aircraft such as the period of 1400 to 1500 LT on 15 July (see Figure 10) and overestimated JNO_2 values significantly when solid cloud deck is above the aircraft such as the period of 1400 to 1500 LT on 6 August (see Figure 9). Note that the cloud location information in Figure 10 was obtained from the observational documents (see Table 3). The poor model performance for JNO_2 during cloudy periods is tied to both misplacements of cloud cover (not shown), and potential inaccuracies in estimation of attenuation of photolysis rates for cloudy conditions. As summarized in Table 3, the P-3 encountered the plume of Ohio valley power plants (e.g., cloud-free polluted conditions) at ~1000 m during 1030 and 1530 LT on 6 August. A very large fluctuation of JNO_2 values varying from 6.9×10^{-4} to 1.1×10^{-1} s^{-1} at this altitude was observed because of the significant effects of the strongly scattering aerosols within the power plant plume (see Figure 8). The relatively constant modeled JNO_2 values of approximately 8.6×10^{-3} s^{-1} during the power plant plume sampling period indicate that the model generally overestimated the observed JNO_2 without capturing its fluctuations within the power plant plume.

3.3 Time Series Comparisons Over the Ocean Surface With the *Ronald H. Brown* Ship Observations

[17] The cruise tracks of the NOAA ship *Ronald H. Brown* in Figure 2 indicate that most of ship's time was

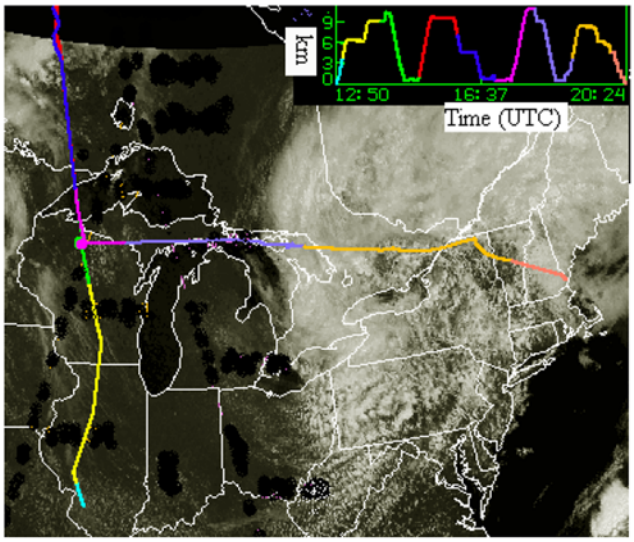

Figure 10. Visible satellite image (GOES-12) for 1630 UTC, 15 July 2004, with DC-8 track overlaid and the height along the flight track shown in the embedded figure (adapted from http://www-air.larc.nasa.gov/missions/intexna/DC8overlay.htm).

spent sampling along the coast of New Hampshire, Massachusetts and Maine. Anthropogenic sources from the Washington, D.C./New York City/Boston urban corridor and biogenic emissions in New Hampshire and Maine significantly impact the sampled air masses along the coast of New England. *Driscoll et al.* [2003] found that NO_x emissions in the northeast United States are primarily from the transportation (54%), electric utilities (25%) and industrial sources (11%). The time series and scatterplots of the model predictions and observations for each parameter (O_3, CO, NO_y, NO_2, NO, SO_2, Isoprene, wind speed, wind direction, RH, photolysis rates for O_3 (JO_3) and NO_2 (JNO_2)) along the ship tracks during the ICARTT period are shown in Figure 11. The air mass flow patterns sampled in the Gulf of Maine can be divided into two groups for our study period as shown in Figure 12b. One is the offshore flow from the southwest and west, and another is the relatively clear marine and continental flow from east, south, north and northwest as summarized in Table 4. The air masses in the southwest offshore flows had passed over the urban New York/Boston corridor during the previous 2–24 hours before being sampled at the ship. *Angevine et al.* [2004] showed that transit times from Boston and NYC to the regions of ship measurements (Gulf of Maine) were approximately 2–3 hours and 12 hours, respectively. These southwest offshore flows led to high pollution episodes along the New England coast. The sampled air masses in the westerly flows were typically about 2–4 hours downwind of Boston and the surrounding forested areas. As indicated in Figure 11 and Table 5, the urban plumes from NYC and Boston in the southwest/west offshore flows were clearly seen above the background concentrations for each species on days 10, 15–17, and 20–23 July, 29 July to 1 August, 3–4, 8–12, and 16–17 August, whereas the clear marine or continental flows from the east/north/northwest/south mainly impacted the ship observational areas on 11–13, 18, and 25–28 July and 5–7 August days characterized by low concentrations for O_3, CO, NO_y, NO_2, NO, SO_2. Note that because of unusually cool and wet conditions with temperatures either below or much below normal and precipitation either above or much above normal over the eastern United States during the summer of 2004, O_3 concentrations are not very high, even during the pollution episodes. There was very good model performance for the clear marine or continental flows from the east/north/northwest/south on days 11–13, 18, and 25–28 July and 5–7 August for each species (O_3, NO_y, NO_2, NO, SO_2) as shown in Figure 11 and Table 4. This suggests that the model can simulate the background environments for clear marine or continental flows very well.

[18] The model overestimated the observed O_3 in all southwest/west offshore flows except on days 16–17 July and 4 and 8–11 August. Two case studies for southwest offshore flows on 20 and 30 July are shown in Figures 12 and 13, respectively. On these days, long-range transport of urban plumes from both NYC and Boston region significantly impacted the atmosphere over the Gulf of Maine during the late afternoon as illustrated in Figure 12c and 13b. Comparison with the ozonesonde observations in Figures 12d and 13c shows that on these two days, the model simulated O_3 at 1840 UTC very well from ground to high altitudes (~2 km) at Narragansett, RI, but consistently overestimated morning O_3 (by ~50 ppb) between the surface and ~2 km over the ocean as shown by the ship ozonesonde. The model reproduced surface O_3 concentrations at ~2200 UTC at the Portsmouth site when the ship arrived back in Portsmouth as shown in Figures 12c and 13b for both 20 and 30 July. Figure 13 also reveals that the model actually simulated the surface high O_3 (>80 ppb) very well at the AQS coastal sites in NH and Maine such as Reid State Park (Maine), Cape Elizabeth (Maine), Portsmouth (NH), Odiorne State Park (NH) and Newbury (MA). On the other hand, a case study for the southwest offshore flow on 10 August illustrated in Figure 14 reveals that the transport of only Boston urban plume impacted the ship observational region (Gulf of Maine). The relatively weak wind speeds (see Figure 11) later during the day reveal that sea breeze carried polluted air from the coastal waters inland into New Hampshire (see Figure 14b); similar features were found during 2002 Northeast Air Quality Study (NEAQS) [*Angevine et al.*, 2004]. The model reproduced this episode very well as shown in Figures 14a and 14b. The significant increases of observed CO, NO_y, NO_2, and NO in Figure 11 during this day also strongly indicate the fingerprint of the fresh urban plumes directly from the Boston city. The model captured the buildup of these species well although it tended to overestimate their concentrations. Compared to the ozonesonde profiles, Figure 14c shows that the model O_3 vertical profiles between surface and ~2 km are close to the observations both at Narragansett, RI, and on the *Ronald H. Brown* ship although the model results are slightly lower than observations by ~10 ppb. The better model performance for O_3 + NO_2 than for O_3 at low concentrations as shown in Figure 14a that reveals that the model exaggerated the effects of NO titration on O_3 as inferred from the O_3

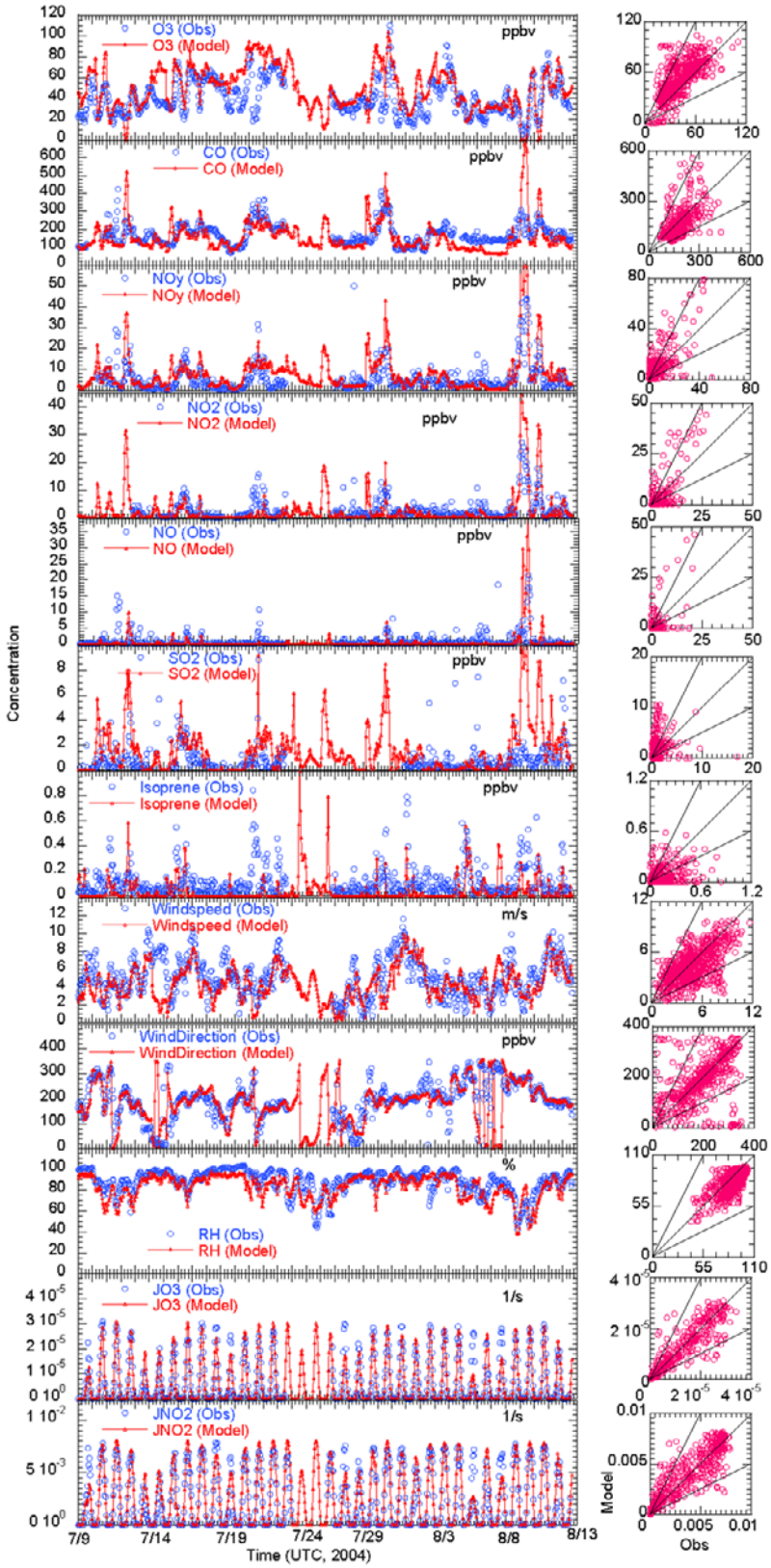

Figure 11. Time series and scatterplots (the 2:1, 1:1 and 1:2 lines are shown for reference) of model predictions and observations for different species and meteorological parameters on the basis of ship measurements.

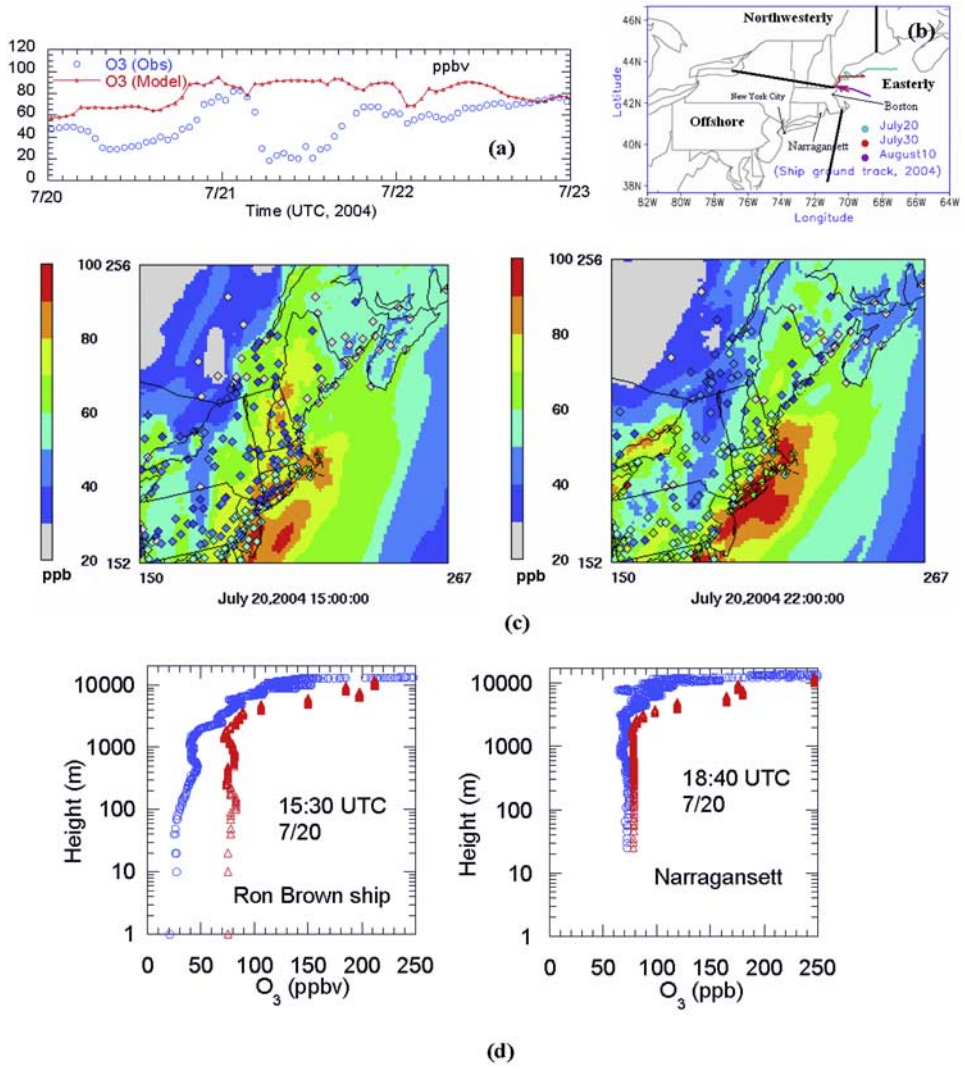

Figure 12. On 20 July 2004 for ship, (a) time series of modeled and observed O_3, (b) ship tracks on 20 and 30 July and 10 August and three flow directions, (c) the model simulation results for O_3 concentration (ppb) with AQS observed data overlaid (diamond) at 1500 and 2200 UTC (20 July 2004), and (d) vertical profiles of model and ozonesonde for O_3 on *Ronald H. Brown* ship (1530 UTC) and at Narragansett, RI (1840 UTC).

observations during the nighttime over the ocean. Upon closer inspection, it is noted that on other days with good model performance for southwest/west offshore flows, i.e., 16–17 July and 4 and 8–11 August, the ship observations in Gulf of Maine were significantly affected only by Boston city plumes according to the model simulations. As pointed out by *Angevine et al.* [2004], pollutant concentrations in stable layers over coastal water surfaces are allowed to remain high because of the lack of deposition or deep vertical mixing of the over-water trajectories. Our analyses suggest that for conditions involving southwest offshore flows impacted by long-range transport of NYC and Boston urban plumes, the model overestimated the O_3 concentrations over the ocean regions, but simulated the O_3 concentrations well over the ocean regions under the conditions impacted only by the Boston plumes. The transport patterns determine the model performance, indicating that the model does not simulate well the transport over land-ocean interface. This suggests additional investigation of the representations of boundary layer mixing and dry deposition over the ocean in the model. The large discrepancies between the model and observations for coastal grid cells where the model results are too high are due to the fact that the model's boundary layer mixing cannot resolve steep subgrid land-to-sea gradients.

[19] Figure 11 indicates that the model captured, with a good deal of fidelity, the temporal variations and broad synoptic changes seen in the observed wind speed, wind direction and relative humidity (RH) along the ship track most of the time, especially for RH, although with some occasional major excursions. The model reproduced the diurnal variations in the observed JNO_2 very well along the ship track most of the time, except on the peaks of 9, 18, 19, and 27 July and 5 and 8 August in which the model seriously overestimated the observations. Misplacements of cloud cover in the model results in the overestimations of

Table 4. Summary of Wind Fields Observed by High-Resolution Doppler Lidar (HRDL) on the *Ronald H. Brown* Ship and Model Performance for O_3 During the 2004 ICARTT Period

Day (2004)	Flow Types[a]	Model Performance for O_3
	Days With Offshore Flows (Southwesterly/Westerly)	
10 Jul[b]	Wly NWly all levels, LLJ (0400–0600 UT), BL wave (1200–2100 UT)	overprediction
15 Jul[b]	NEly winds shifting to Sly, UL easterly in BL, SWly, LLJ (2200–2400 UT)	overprediction
20 Jul	SSW to Sly flow in BL, Wly above	overprediction
21 Jul[b]	SWly winds shifting to Nly to Ely to Sly in BL, LLJ (0000–1000 UT), BL wave (1400–1700 UT)	overprediction
22 Jul[b]	SSWly for low-level winds, shifts from Wly to Ely to SWly >1 km, LLJ (2100–2400 UT)	overprediction
23 Jul[b]	SSW to Sly flow, LLJ events (0000–1500 UT)	overprediction
29 Jul[b]	Wly all levels then shift to Nly to Sly at surface, LLJ (2300 UT), BL waves	overprediction
30 Jul[b]	Wly and SWly flow all levels, LLJ (0000–0300 and 2000–2400 UT)	overprediction
31 Jul[b]	WSWly flow all levels, LLJ throughout, vels up to 20 m/s	overprediction
1 Aug[b]	W-SWly flow 15–20 m/s, LLJ (0600–0900 UT)	overprediction
3 Aug	SWly winds shifting to NEly, BL wave	overprediction
12 Aug	SWly SSWly flow all levels	overprediction
16 and 17 Jul[b]	sustained WSWly flow for 2.5 days, BL and LLJ events throughout	very good
4 Aug	Wly in BL, NW above	good
8–11 Aug[b]	W-SWly flow for 4 days, LLJs throughout and hi velocity shear	very good
	Days With Clear Marine and Continental Flows (Easterly/Northerly/Northwesterly)	
11 Jul[b]	Wly winds shifting to Nly to Ely, LLJ (0300–0600 UT)	very good
12 Jul[b]	Ely winds shifting to Sly to SWly, LLJ (0600 and 2300 UT)	very good
13 Jul[b]	SSWly winds shifting to Ely, LLJ (0300–0400, 0900–1100, and 2300 UT)	very good
18 Jul	SWly winds shifting to SEly	good
25–27 Jul	light predominantly Ely flow <4 m/s throughout	very good
28 Jul[b]	Nly winds shifting to Ely at 2000 UT and then to WNWly, LLJ (1100 UT)	very good
5 Aug	Wly and NWly 0–100 m, Nly 100–1000 m	very good
6 and 7 Aug[b]	Wly winds shifting to Ely then back via N	good

[a]Obtained from *Brewer* [2005].
[b]Nocturnal low-level jet (LLJ).

Table 5. Statistical Summaries of the Comparisons of the Model Results With the Observations at the Different Sites During the 2004 ICARTT Period (1 July to 15 August 2004)[a]

	$\langle C \rangle$[b]				
Parameters	Obs	Model	r	% Within a Factor of 1.5[c]	% Within a Factor of 2[c]
			Castle Springs (N = 1047)		
O_3	35.17	43.63	0.493	66.6	90.1
NO	0.14	0.05	0.222	12.1	22.5
CO	188.84	108.78	0.706	19.3	74.7
NO_Y	2.27	3.14	0.587	43.6	67.7
SO_2	1.16	0.87	0.388	29.6	45.8
JNO_2 (1/s)	3.18×10^{-3}	4.07×10^{-3}	0.820	49.6	63.4
Temperature, °C	19.65	19.78	0.867	100.0	100.0
RH, %	78.69	71.64	0.781	97.7	100.0
			Isle of Schoals (N = 1078)		
O_3	36.68	52.31	0.541	56.9	80.2
CO	171.70	121.15	0.610	60.9	90.3
NO	0.76	0.18	0.448	0.8	3.5
			Mount Washington (N = 1076)		
O_3	45.87	45.85	0.554	87.7	98.8
NO	3.64	0.01	−0.054	8.9	13.8
CO	152.43	95.19	0.301	46.7	84.3
NO_Y	4.04	2.23	−0.060	20.6	38.4
SO_2	0.74	0.30	−0.001	19.0	32.6
JNO_2 (1/s)	3.59×10^{-3}	4.43×10^{-3}	0.768	43.1	61.9
			Thompson Farm (N = 1067)		
O_3	28.80	41.68	0.751	48.1	73.8
NO	0.33	0.29	0.436	31.3	51.3
CO	173.07	154.66	0.593	77.7	98.5
NO_Y	3.93	7.26	0.321	28.8	51.6
SO_2	1.22	1.63	0.084	14.3	25.3
JNO_2 (1/s)	3.19×10^{-3}	3.90×10^{-3}	0.865	53.8	68.1
Temperature, °C	20.33	20.44	0.887	99.9	100.0
RH, %	80.97	75.18	0.829	98.5	100.0

[a]r is correlation coefficient between the model predictions and observations.
[b]$\langle C \rangle$ is the mean concentration (ppb).
[c]Percentages (%) are the percentages of the comparison points at which model results are within a factor of 1.5 and 2 of the observations. N is number of samples.

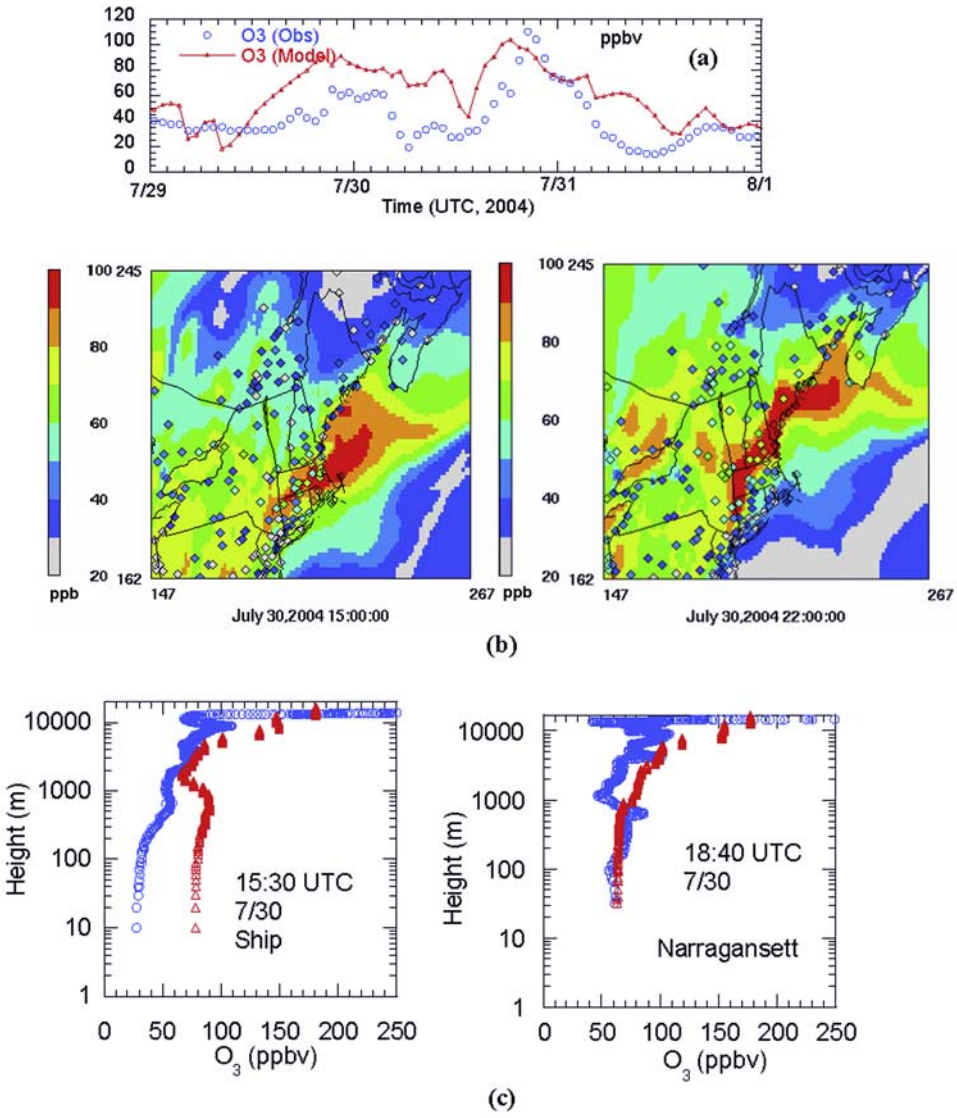

Figure 13. On 30 July 2004 for ship, (a) time series of modeled and observed O_3, (b) The model simulation results for O_3 concentration (ppb) with AQS observed data overlaid (diamond) on 1500 and 2200 UTC, and (c) vertical profiles of model and ozonesonde for O_3 on *Ronald H. Brown* ship (1530 UTC) and at Narragansett, RI (1840 UTC).

the observed JNO_2 (not shown), which can also contribute to the higher O_3 bias. The model performance for JO_3 follows those of JNO_2 as shown in Figure 11. The better model performance for JNO_2 along the ship track than for the aircraft may be because the aircraft flight paths covered much larger geographic areas and possibly encountered cloudy conditions more often.

3.4. Time Series Comparison and Diagnostic Evaluation at the AIRMAP Sites

[20] Figure 15 presents time series comparisons and scatterplots of the model predictions and observations for O_3, CO, NO, NO_y, SO_2, JNO_2, temperature (T) and RH at the CS site. Following *Yu et al.* [2003, 2006], the percentages of the comparison points where the model results are within a factor of 1.5 and 2, respectively, of the observations are listed in Table 5. The model captured the hourly variations and broad synoptic changes seen in the observations of each parameter (O_3, CO, NO_y, JNO_2, T and RH) (correlation coefficient>0.49, see Table 5) except NO and SO_2 at CS, IS and TF sites. The serious underestimation of NO, CO, NOy and SO_2 at the MWO site (the highest mountain (1916 m) in the northeastern United States), in part, reflects the inherent subgrid variability in their emissions and concentrations that are not adequately captured by the model grid structure. This is also due to the fact that usually the models misrepresent mountain sites because they essentially sample free tropospheric air while models can't resolve the terrain. Another possible reason for the underestimation at the mountain sites could be related to unrealistic downward mixing in CMAQ's convective scheme that may bring free tropospheric air to the mountain

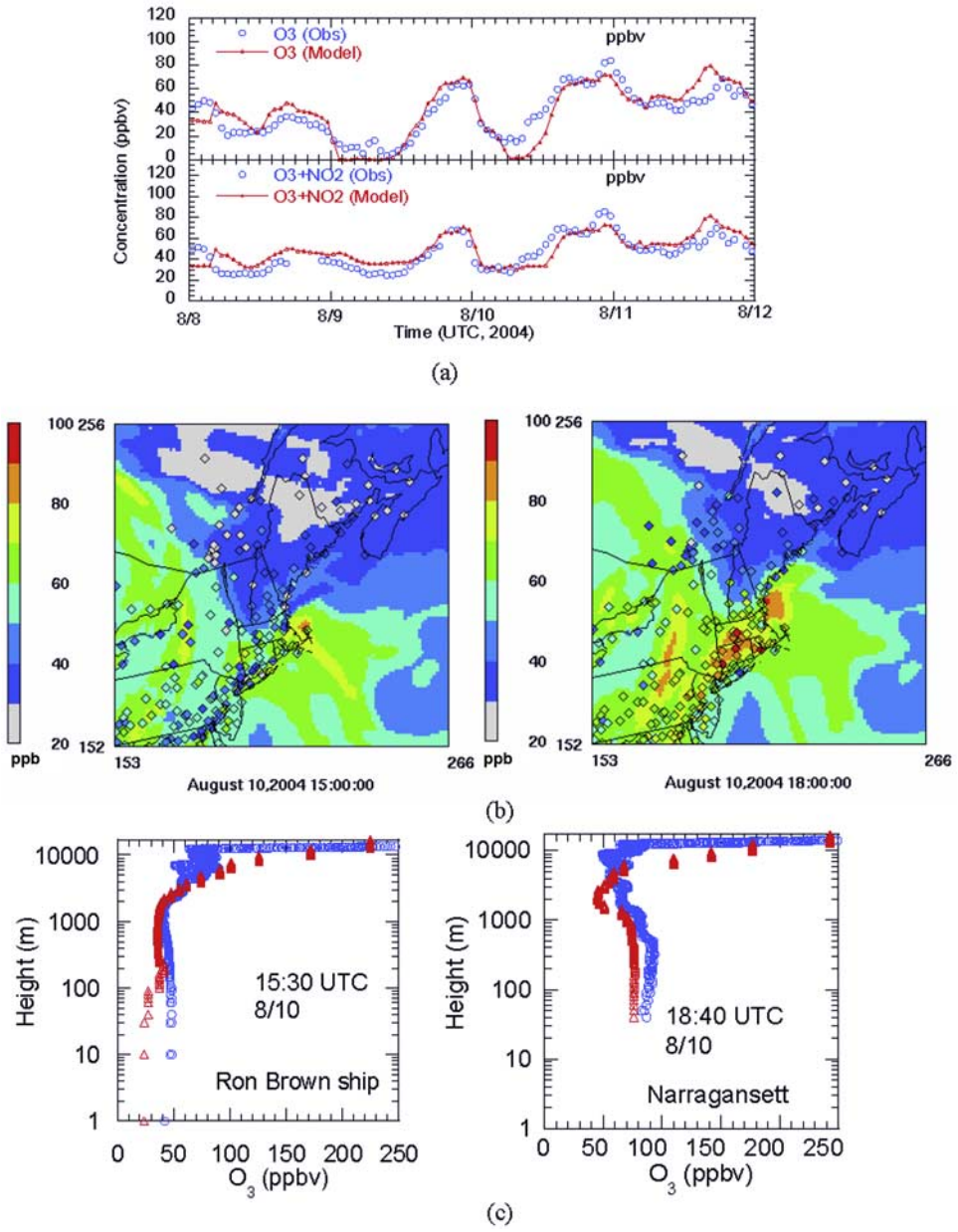

Figure 14. Same as Figure 13 but for 10 August 2004.

site and which can then result in the simulated low concentrations of these species. Relatively large discrepancies between modeled and measured concentrations are noted for primary species, such as NO and SO_2. These are likely related to the discrepancies between modeled and observed wind speed and direction, which cause modeled plumes to be displaced leading to relatively larger error for primary species when the modeled and measured values are paired in space and time. The model underestimated CO by 20–50% consistently at each site, similar to those comparisons for the vertical profiles. As shown in Figure 16, the Eta model tended to overestimate the observed PBL heights at all times at Concord, NH, which were derived from radar wind profilers during the 2004 ICARTT study [*Pleim*, 2007]. The general overestimations of the PBL heights by the Eta model could also cause the underestimation of precursor concentrations at the surface sites. The model reproduced the observed temperatures with ~±5% errors and relative humidity (RH) with ~±10% errors at CS and TF sites.

[21] The analysis of photolysis rates of NO_2 focuses on daytime data by excluding data where $JNO_2 < 5 \times 10^{-5}$ s^{-1} following *Thornton et al.* [2002]. Table 5 indicates that the model reproduced 49.6%, 43.1% and 53.8% of observed JNO_2 values within a factor of 1.5 at the CS, MWO and TF sites, respectively. *DeMore et al.* [1997] suggest that about ±20% uncertainty in photolysis rates can be associated with uncertainty in the cross section and quantum yield data used in the calculation of JNO_2 values. The sensitivity tests of *Hanna et al.* [2001] indicate that a 50% uncertainty in JNO_2 could cause about a 40 ppbv, or a 20% uncertainty in predicted maximum O_3 concentration in their cases. Addi-

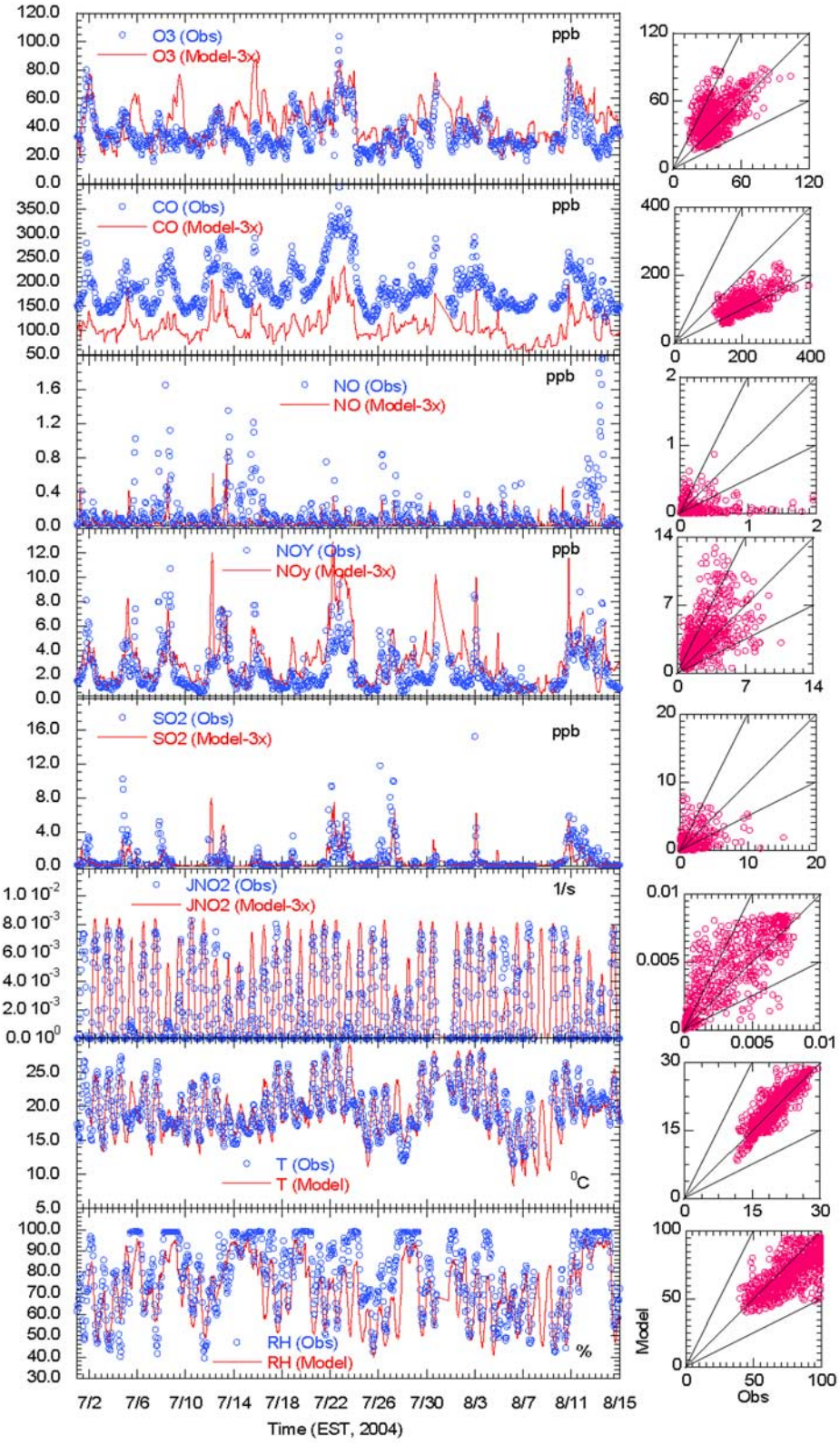

Figure 15. Time series and scatterplots of model predictions and observations for each parameter at the Castle Springs site.

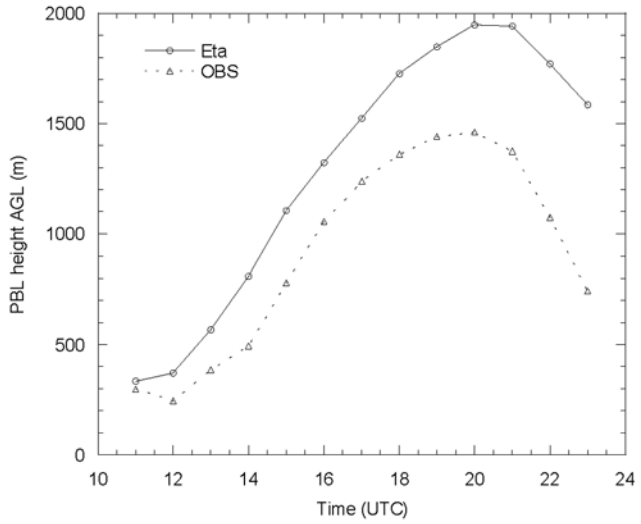

Figure 16. Comparison of PBL height of the Eta model and observation derived from a radar wind profiler at Concord, NH, averaged over the ICARTT period.

tional uncertainties in the model simulations can also arise from uncertainties and errors associated with the spatial and temporal representation of cloud fields in the model and their subsequent effects on photolysis attenuation.

[22] The upper limits of the ozone production efficiencies (ε_N) value can be estimated by the O_3-NO_z ($NO_z = NO_y - NO_x$) slope. *Jacob et al.* [1995] estimated NO_2 concentrations for daytime conditions during the Shenandoah Cloud and Photochemistry Experiment (SCAPE) by assuming the $NO/NO_2/O_3$ photostationary steady state (PSS) in order to obtain NO_x and NO_z concentrations. Following *Jacob et al.* [1995], *Griffin et al.* [2004] and *Kleinman et al.* [2004], NO_2 concentrations at the CS and TF sites were estimated on the basis of the PSS assumption in this study. On the basis of comparisons between observed and calculated NO_2 from the field program, *Kleinman et al.* [2004] estimated an accuracy of ±25% for the calculated NO_2 values by the PSS assumption for NO_x to NO_y ratios in fresh plumes. On the basis of the ship data in which NO_2 concentrations were observed (see section 3.3) in this study, it was found that the mean NO_2 concentrations for the observations and estimations by the PSS assumption are 2.41±2.87 and 3.36±4.36 ppbv, respectively, with correlation coefficient of 0.932 between them. The PSS assumption overestimated the observed NO_2 by 28% in this case. In this study, the PSS assumption is only used to estimate NO_2 concentrations at the CS and TF sites. The $[O_3]/[NO_x]$ values can be used to determine NO_x-sensitive and VOC-sensitive chemical regimes. *Arnold et al.* [2003] showed that $[O_3]/[NO_x]$ values >46 indicate strong NO_x-sensitive conditions, whereas values <14 indicate VOC-sensitive conditions. Table 6 summarizes the variations in the $[O_3]/[NO_x]$ ratio at the CS and TF sites. The results along the *Ronald H. Brown* ship tracks are also listed in Table 6 for comparison. As can be seen, the model generally reproduced the temporal variations in the observed $[O_3]/[NO_x]$ ratios across the different conditions represented at the three locations. Both model and observations show that the CS site is mainly under strongly NO_x-sensitive conditions (>66%), whereas the TF site and ship over the ocean are under neither strongly NO_x-sensitive nor VOC-sensitive conditions. Following *Arnold et al.* [2003], both modeled and observed O_3-NO_z slopes are obtained for only observational data with $[O_3]/[NO_x]>46$ at the CS and TF sites and on ship. There is significant correlation between O_3 and NO_z for both model predictions and observations ($r > 0.61$) at the three locations (see Figure 17 and Table 7). While the ε_N values of the model (5.2 to 6.4) and observation (8.5 to 10.7) at the CS and TF sites are close to the lower and higher bounds of the estimated ranges (5 to 10) of other investigators [*Olszyna et al.*, 1994; *Fiore et al.*, 2002] at rural sites in the eastern United States, respectively, the modeled ε_N value is about 40% lower than the observations. The modeled intercepts (background O_3) are consistently higher than the observations at each location. The results along the ship tracks over the ocean in Table 7 and Figure 17 reveal that the modeled ε_N value (3.6) is much lower than the corresponding observation (11.7) and both modeled and observed ε_N values are outside of estimation range (5 to 10) of other investigators at rural sites in the eastern United States. As suggested by *Chin et al.* [1994], the ε_N values estimated by the O_3-NO_z slopes are upper limits because NO_z species (primarily HNO_3) are removed from the atmosphere more rapidly than O_3. Figure 17 shows that compared to the observations, the model produced less O_3 at the high NO_z regime. The scatterplots of Figure 17 also reveal that the modeled NO_z concentrations were higher than the observations, indicating that the model chemistry produces more terminal oxidized nitrogen products than inferred from observations, thereby contributing in part to the noted underestimation of ε_N.

4. Summary and Recommendations

[23] A rigorous evaluation of the Eta-CMAQ forecast model performance for O_3, its related precursors, and meteorological parameters has been carried out over the eastern United States by comparing the model results with

Table 6. Statistical Summary of Number of Hours for Response Surface Indicator Ratios (O_3/NO_x) for Model and Observations at the CS, WMO, and TF Sites for All Days (Obs-Limited Hours) During the Period of 1 July to 15 August 2004[a]

	Castle Springs		Thompson Farm		*Ronald H. Brown* ship	
O_3/NO_x	Obs	Model	Obs	Model	Obs	Model
0–14	32 (7)	18 (4)	181 (38)	105 (22)	106 (35)	49 (16)
15–25	34 (7)	19 (4)	51 (11)	72 (15)	58 (19)	21 (7)
26–45	94 (20)	18 (4)	59 (12)	125 (26)	46 (15)	30 (10)
>46	312 (66)	417 (88)	188 (39)	177 (37)	93 (31)	103 (34)
Total hours	472 (100)	472 (100)	479 (100)	479 (100)	303 (100)	303 (100)

[a]The values in parentheses are the percentages (%).

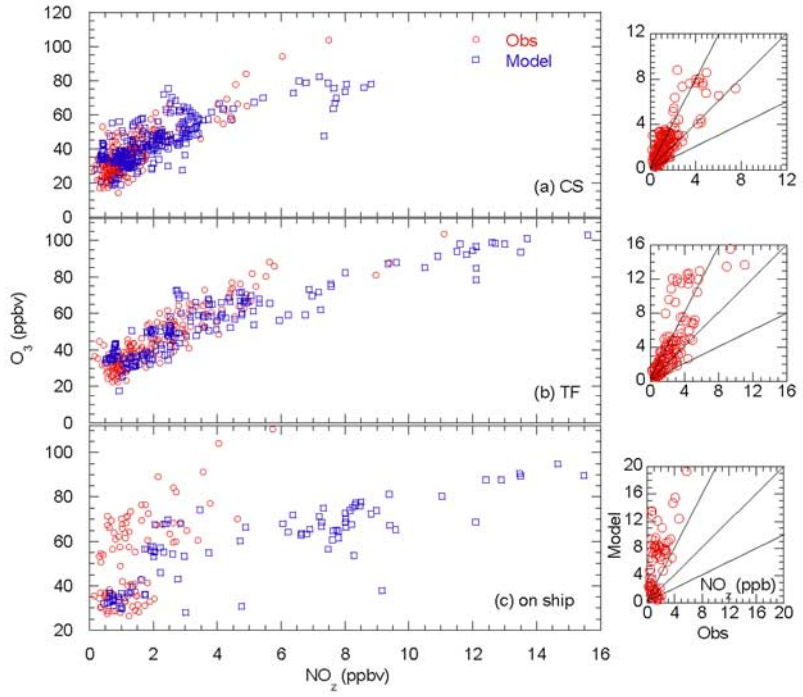

Figure 17. O_3 as a function of NO_z for the NO_x-limited conditions indicated by the observational data with $[O_3]/[NO_x]>46$ at (a) Castle Springs (CS), (b) Thompson Farm (TF), and (c) along ship tracks. Right plots are scatterplots of modeled and observed NO_z.

the observations using measurements obtained during the 2004 ICARTT study. The results at the AIRNOW surface sites show that the model was able to reproduce the day-to-day variations of observed daily maximum 8-hour O_3 and captured the majority (73%) of observed daily maximum 8-hour O_3 within a factor of 1.5 with NMB = 22%. The model significantly overestimated the O_3 concentrations in areas of cloud cover mainly caused by the unrealistic vertical transport in CMAQ's convective cloud scheme. On the basis of results from aircraft, ozonesonde and ship-based lidar observations, the model generally reproduced O_3 vertical structures most of the days at low altitudes with consistent overestimations above ~6 km due to the lateral boundary conditions derived by the GFS and coarse model resolution in the free troposphere. The model consistently underestimates CO by ~30% from surface to high altitudes because of the inadequate representation of the transport of pollution associated with biomass burning from outside the domain. The model captured the vertical variation patterns of the observed values for other parameters (HNO_3, SO_2, NO_2, HCHO, NOy_sum) with some exceptions, depending on the studied regions and air mass characteristics. The consistent underestimation of observed NO at altitudes > 6 km relative to DC-8 measurements is attributed, in part, to the exclusion of aircraft and lightning NO emissions in the real-time model emission inventory. The very good model performance of NO_y_sum relative to consistent overestimation of NOy reveals that the model overestimated sum of NO_3, N_2O_5, HONO, PNA and NTR. The model can generally capture the observed JNO_2 very well during the cloud-free periods, but underestimated the JNO_2 values by 20–90% when there was a solid cloud deck below the aircraft and overestimated JNO_2 values significantly when solid cloud deck was above the aircraft. The poor model performance for JNO_2 during cloudy periods is tied to both misplacements of cloud cover, potential inaccuracies in estimation of attenuation of photolysis rates for cloudy conditions. On the other hand, the model was able to reproduce the vertical profiles of observed water vapor and wind speed.

[24] The capability of the model to reproduce the observed pollutants over the ocean areas (Gulf of Maine) differed from day to day, depending on the offshore flow

Table 7. Correlations Between O_3 and NO_z for the NO_x-Limited Conditions Indicated by the Observational Data With [O3]/[NOx]>46 (Aged Air Masses) at the CS, WMO and TF Sites During the Period of 1 July to 15 August 2004[a]

Sites	Regression Equations
Castle Springs (N = 312)	Obs: $[O_3] = 10.7[NO_z] + 22.8$, r = 0.838
Castle Springs (N = 312)	Model: $[O_3] = 6.4[NO_z] + 30.1$, r = 0.784
Thompson Farm (N = 188)	Obs: $[O_3] = 8.5[NO_z] + 26.4$, r = 0.896
Thompson Farm (N = 188)	Model: $[O_3] = 5.2[NO_z] + 34.0$, r = 0.911
Ship over the ocean (N = 93)	Obs: $[O_3] = 11.7[NO_z] + 35.4$, r = 0.608
Ship over the ocean (N = 93)	Model: $[O_3] = 3.6[NO_z] + 38.7$, r = 0.833

[a]N is number of points, and r is correlation coefficient.

types and transport patterns, i.e., good performance for marine or continental clear flows from the east/north/northwest/south and southwest flows influenced only by Boston city plumes but overestimation for southeast flows influenced by the long-range transport plumes including both NYC and Boston. Time series comparisons at the AIRMAP sites indicate that the model captured the hourly variations and broad synoptic changes in the observations of different gas species (O_3, NO_2, CO, NO_y, PAN) except NO and SO_2 at each site, although there were occasional major excursions. The ε_N values of the model (5.2 to 6.4) and observations (8.5 to 10.7) at the CS and TF sites are close to the lower and higher bounds of the estimated ranges (5 to 10) of other investigators at rural sites in the eastern United States, respectively. However, the modeled ε_N value is about 40% lower than the observations. Since the majority of the eastern United States during the summer 2004 experienced unusually cool and wet conditions, the model performance presented here is probably not climatologically representative of summertime, but is unique to the summer of 2004.

[25] In light of the uncertainties in the photochemical mechanism, prognostic model forecasts of meteorological fields, and challenges associated with real-time specification of the day-to-day variability in emission patterns and magnitude, the overall performance of the Eta-CMAQ forecast model during the ICARTT period can be considered to be reasonable since the traditional performance metrics of error and bias in predicted concentrations are similar in magnitude to those from typical hind-cast applications with regional air quality models [e.g., *Russell and Dennis*, 2000]. Nevertheless, comparisons with the detailed measurements from ICARTT revealed several systematic trends in model errors and biases, addressing which can result in substantial improvements in the model's forecast skill. On the basis of our analysis the following recommendations for improvements in model processes and key input data are offered:

[26] 1. Since the prognostic model forecasts of meteorological fields, especially the accurate description of frontal passages, cloud cover, and wind fields, are key to forecasting regional air quality, improvements in the forecast of these variables will directly lead to improvements in the predictions of local pollution gradients as well as in representation of regional pollution patterns.

[27] 2. Limited comparisons with profiler data suggest overestimation of simulated boundary layer heights which may lead to excessive dilution of pollutant species, and suggest the need for improvements in representation of subgrid treatment of PBL processes and mixing therein.

[28] 3. The systematic overpredictions in simulated JNO_2 and O_3 under conditions of widespread cloudiness highlight the need for accurate representation of the effects of clouds on photolysis attenuation as well as accurate representation of spatial trends in cloudiness in the models.

[29] 4. The systematic overprediction in simulated free-tropospheric values of O_3 highlights the need for improved methods for lateral boundary conditions through consistent coupling with larger-scale models.

[30] 5. The systematic underprediction of NO at altitudes >6 km suggests the possible role of emissions from lightning and aircrafts, which have traditionally not been included in episodic regional-scale modeling but may potentially be important for continuous long-term applications. The NO underpredictions also highlight the need for greater scrutiny of the performance of current chemical mechanisms at the lower temperature range typical of the free troposphere.

[31] 6. Relatively large discrepancies between modeled and observed pollutant concentrations during events influenced by biomass burning highlight the need for accurate event-based representation of these emissions in real-time applications.

[32] **Acknowledgments.** The authors would like to thank S. T. Rao, R. Pinder, J. Godowitch, and the anonymous reviewers for the constructive and very helpful comments that led to a substantial strengthening of the content of the paper. We also thank Paula Davidson for programmatic support, Jeff Young, Jeff McQueen, Pius Lee, and Marina Tsidulko for collaboration and critical assistance in performing the forecast simulations. We are grateful to the 2004 ICARTT investigators for making their measurement data available. NOAA Environmental Technology Laboratory provided the Lidar O_3 and the High Resolution Doppler Lidar (HRDL) wind vertical profiles. The authors would like to thank Jim Wilczak and Laura Bianco for deriving and providing the PBL heights from radar wind profiler measurements. The research presented here was performed under the Memorandum of Understanding between the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Commerce's NOAA and under agreement DW13921548. This work constitutes a contribution to the NOAA Air Quality Program. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.

References

Angevine, W. M., C. J. Senff, A. B. White, E. J. Williams, J. Koermer, S. T. K. Miller, R. Talbot, P. E. Johnston, S. A. McKeen, and T. Downs (2004), Coastal boundary layer influence on pollutant transport in New England, *J. Appl. Meteorol.*, *43*, 1425–1437.

Arnold, J. R., R. L. Dennis, and G. S. Tonnesen (2003), Diagnostic evaluation of numerical air quality models with specialized ambient observations: Testing the Community Multiscale Air Quality modeling system (CMAQ) at selected SOS 95 ground sites, *Atmos. Environ.*, *37*, 1185–1198.

Brewer, A. (2005), Doppler lidar measurements on the Ronald H. Brown, paper presented at 2005 NOAA Research Vessel *Ronald H. Brown* Data Workshop, Boulder, Colo., 8–9 March.

Byun, D. W., and J. K. S. Ching (1999), Science algorithms of the EPA Models-3 Community Multi-scale Air Quality (CMAQ) modeling system, *EPA/600/R-99/030,* Off. of Res. and Dev., U. S. Environ. Prot. Agency, Washington, D. C.

Byun, D. W., and K. L. Schere (2006), Review of the governing equations, computational algorithms, and other components of the models-3 Community Multi-scale Air Quality (CMAQ) modeling system, *Appl. Mech. Rev.*, *59*, 51–77.

Chin, M., D. J. Jacob, J. W. Munger, D. D. Parrish, and B. G. Doddridge (1994), Relationship of ozone and carbon monoxide over North America, *J. Geophys. Res.*, *99*, 14,565–14,573.

Cope, M. E., et al. (2004), The Australian air quality forecasting system. Part I: Project description and early outcomes, *J. Appl. Meteorol.*, *43*, 649–662.

DeBell, L. J., M. Vozzella, R. W. Talbot, and J. E. Dobb (2004), Asian dust storm events of spring 2001 and associated pollutants observed in New England by the Atmospheric Investigation, Regional Modeling, Analysis and Prediction (AIRMAP) monitoring network, *J. Geophys. Res.*, *109*, D01304, doi:10.1029/2003JD003733.

DeMore, W. B., S. P. Sander, C. J. Howard, A. R. Ravishankara, D. M. Golden, C. E. Kolb, R. F. Hampson, M. J. Kurylo, and M. J. Molina (1997), Chemical kinetics and photochemical data for use in stratospheric modeling, evaluation 12, NASA Jet Propul. Lab., Calif. Inst. of Technol., Pasadena.

Driscoll, C. T., et al. (2003), Nitrogen pollution in the northeastern United States: Sources, effects, and management options, *Bioscience*, *53*, 357–374.

Fiore, A. M., D. J. Jacob, I. Bey, R. M. Yantosca, B. D. Field, A. C. Fusco, and J. G. Wilkinson (2002), Background ozone over the United States in summer: Origin, trend, and contribution to pollution episodes, *J. Geophys. Res.*, *107*(D15), 4275, doi:10.1029/2001JD000982.

Griffin, R., C. A. Johnson, R. W. Talbot, H. Mao, R. S. Russo, Y. Zhou, and B. C. Sive (2004), Quantification of ozone formation metrics at

Thompson Farm during the New England Air Quality Study (NEAQS) 2002, *J. Geophys. Res.*, *109*, D24302, doi:10.1029/2004JD005344.

Hanna, S. R., Z. Lu, H. C. Frey, N. Wheeler, J. Vukovich, S. Arunachalam, M. Fernau, and D. A. Hansen (2001), Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, *Atmos. Environ*, *35*, 891–903.

Jacob, D. J., L. W. Horowitz, J. W. Munger, B. G. Heikes, R. R. Dickerson, R. S. Artz, and W. C. Keene (1995), Seasonal transition for NO_x- to hydrocarbon-limited conditions for ozone production over the eastern United States in September, *J. Geophys. Res.*, *100*(D5), 9315–9324.

Kang, D., B. K. Eder, A. F. Stein, G. A. Grell, S. E. 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.

Kleinman, L. I., W. F. Ryan, P. H. Daum, S. R. Springston, Y.-N. Lee, and L. J. Nunnermacker (2004), An ozone episode in the Philadelphia metropolitan area, *J. Geophys. Res.*, *109*, D20302, doi:10.1029/2004JD004563.

Lin, C.-Y., D. J. Jacob, J. W. Munger, and A. M. Fiore (2000), Increasing background ozone in surface air over the United States, *Geophys. Res. Lett.*, *27*, 3465–3468.

Mao, H., and R. Talbot (2004), O_3 and CO in New England: Temporal variations and relationships, *J. Geophys. Res.*, *109*, D21304, doi:10.1029/2004JD004913.

Mathur, R., J. Pleim, K. Schere, J. Young, T. Otte, G. Pouliot, B. Eder, D. Kang, S. Yu, and H.-M. Lin (2004), Adaptation and application of the Community Multiscale Air Quality (CMAQ) modeling system for real-time air quality forecasting during the summer of 2004, paper presented at 2004 Models-3/CMAQ Conference, 18–20 Oct., Chapel Hill, N. C.

Mathur, R., et al. (2005), Multiscale air quality simulation platform (MAQSIP): Initial applications and performance for tropospheric ozone and particulate matter, *J. Geophys. Res.*, *110*, D13308, doi:10.1029/2004JD004918.

McHenry, J. N., W. F. Ryan, N. L. Seaman, C. J. Coats, J. Pudykeiwicz, S. Arunachalum, and J. M. Vukovich (2004), A real-time Eulerian photochemical model forecast system, *Bull. Am. Meteorol. Soc.*, *85*, 525–548.

McKeen, S. A., et al. (2002), Ozone production from Canadian wildfires during June and July of 1995, *J. Geophys. Res.*, *107*(D14), 4192, doi:10.1029/2001JD000697.

McKeen, S. A., et al. (2005), Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004, *J. Geophys. Res.*, *110*, D21307, doi:10.1029/2005JD005858.

Olszyna, K. J., E. M. Bailey, R. Simonaitis, and J. F. Meagher (1994), O_3 and NOy relationships at a rural site, *J. Geophys. Res.*, *99*, 14,557–14,563.

Otte, T. L., et al. (2005), Linking the Eta model with the Community Multiscale Air Quality (CMAQ) modeling system to build a national air quality forecasting system, *Weather Forecasting*, *20*, 367–384.

Pleim, J. E. (2007), A combined local and non-local closure model for the atmospheric boundary layer. Part 2: Application and evaluation in a mesoscale meteorology model, *J. Appl. Meteorol*, in press.

Pouliot, G. A. (2005), The emissions processing system for the Eta/CMAQ air quality forecast system, paper presented at 7th Conference on Atmospheric Chemistry, 85th AMS Annual Meeting, Am. Meteorol. Soc., San Diego, Calif.

Rogers, E., T. Black, D. Deaven, G. DiMego, Q. Zhao, M. Baldwin, N. Junker, and Y. Lin (1996), Changes to the operational "early" Eta Analysis/Forecast System at the National Centers for Environmental Prediction, *Weather Forecasting*, *11*, 391–413.

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

Thornton, J. A., et al. (2002), Ozone production rates as a function of NO_x abundances and HO_x production rates in the Nashville urban plume, *J. Geophys. Res.*, *107*(D12), 4146, doi:10.1029/2001JD000932.

U. S. Environmental Protection Agency (1991), Guideline for regulatory application of the urban airshed model, *USEPA Rep. EPA-450/4-91-013,* Off. of Air Qual. Planning and Stand., Research Triangle Park, N. C.

U. S. Environmental Protection Agency (1999), Guideline for developing an ozone forecasting program, *EPA-454/R-99-009*, Research Triangle Park, N. C.

U. S. Environmental Protection Agency (2005), Air quality criteria for ozone and related photochemical oxidants (first external review draft), *EPA/600/R-05/004aA-cA,* Washington, D. C.

Yu, S. C., P. S. Kasibhatla, D. L. Wright, S. E. Schwartz, R. McGraw, and A. Deng (2003), Moment-based simulation of microphysical properties of sulfate aerosols in the eastern United States: Model description, evaluation, and regional analysis, *J. Geophys. Res.*, *108*(D12), 4353, doi:10.1029/2002JD002890.

Yu, S. C., R. Mathur, D. Kang, K. Schere, B. Eder, and J. Pleim (2006), Performance and diagnostic evaluations of ozone estimations by the Eta-CMAQ air quality forecast system during the 2002 New England Air Quality Study (NEAQS), *J. Air Waste Manage. Assoc.*, *56*, 1459–1471.

D. Kang, R. Mathur, T. L. Otte, J. Pleim, K. Schere, and S. Yu, Atmospheric Sciences Modeling Division, Air Resources Laboratory, NOAA, Research Triangle Park, NC 27711, USA. (yu.shaocai@epa.gov)