

An evaluation of real-time air quality forecasts and their urban emissions over eastern Texas during the summer of 2006

Second Texas Air Quality Study field study

S. McKeen,^{1,2} G. Grell,^{1,3} S. Peckham,^{1,3} J. Wilczak,⁴ I. Djalalova,^{1,4} E.-Y. Hsie,^{1,2} G. Frost,^{1,2} J. Peischl,^{1,2} J. Schwarz,^{1,2} R. Spackman,^{1,2} J. Holloway,^{1,2} J. de Gouw,² C. Warneke,^{1,2} W. Gong,⁵ V. Bouchet,⁶ S. Gaudreault,⁶ J. Racine,⁶ J. McHenry,⁷ J. McQueen,⁸ P. Lee,⁹ Y. Tang,⁸ G. R. Carmichael,¹⁰ and R. Mathur¹¹

Received 31 December 2008; revised 15 March 2009; accepted 25 March 2009; published 18 June 2009.

[1] Forecasts of ozone (O_3) and particulate matter (diameter less than $2.5 \mu m$, $PM_{2.5}$) from seven air quality forecast models (AQFMs) are statistically evaluated against observations collected during August and September of 2006 (49 days) through the Aerometric Information Retrieval Now (AIRNow) network throughout eastern Texas and adjoining states. Ensemble O_3 and $PM_{2.5}$ forecasts created by combining the seven separate forecasts with equal weighting, and simple bias-corrected forecasts, are also evaluated in terms of standard statistical measures, threshold statistics, and variance analysis. For O_3 the models and ensemble generally show statistical skill relative to persistence for the entire region, but fail to predict high- O_3 events in the Houston region. For $PM_{2.5}$, none of the models, or ensemble, shows statistical skill, and all but one model have significant low bias. Comprehensive comparisons with the full suite of chemical and aerosol measurements collected aboard the NOAA WP-3 aircraft during the summer 2006 Second Texas Air Quality Study and the Gulf of Mexico Atmospheric Composition and Climate Study (TexAQS II/GoMACCS) field study are performed to help diagnose sources of model bias at the surface. Aircraft flights specifically designed for sampling of Houston and Dallas urban plumes are used to determine model and observed upwind or background biases, and downwind excess concentrations that are used to infer relative emission rates. Relative emissions from the U.S. Environmental Protection Agency 1999 National Emission Inventory (NEI-99) version 3 emissions inventory (used in two of the model forecasts) are evaluated on the basis of comparisons between observed and model concentration difference ratios. Model comparisons demonstrate that concentration difference ratios yield a reasonably accurate measure (within 25%) of relative input emissions. Boundary layer height and wind data are combined with the observed up-wind and downwind concentration differences to estimate absolute emissions. When the NEI-99 inventory is modified to include observed NO_y emissions from continuous monitors and expected NO_x decreases from mobile sources between 1999 and 2006, good agreement is found with those derived from the observations for both Houston and Dallas. However, the emission inventories consistently overpredict the ratio of CO to NO_y . The ratios of ethylene and aromatics to NO_y are reasonably consistent with observations over Dallas, but are significantly underpredicted for Houston. Excess

¹Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, Colorado, USA.

²Chemical Sciences Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado, USA.

³Global Systems Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado, USA.

⁴Physical Sciences Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado, USA.

Copyright 2009 by the American Geophysical Union.
0148-0227/09/2008JD011697\$09.00

⁵Science and Technology Branch, Environment Canada, Downsview, Ontario, Canada.

⁶Meteorological Service of Canada, Environment Canada, Dorval, Quebec, Canada.

⁷Baron Advanced Meteorological Systems, Research Triangle Park, North Carolina, USA.

⁸Environmental Modeling Center, National Center for Environmental Prediction, National Weather Service, National Oceanic and Atmospheric Administration, Camp Springs, Maryland, USA.

⁹Air Resources Laboratory, National Oceanic and Atmospheric Administration, Silver Spring, Maryland, USA.

¹⁰Center for Global and Regional Environmental Research, University of Iowa, Iowa City, Iowa, USA.

¹¹National Exposure Research Laboratory, Environmental Protection Agency, Research Triangle Park, North Carolina, USA.

ratios of PM_{2.5} to NO_y reasonably match observations for most models but the organic carbon fraction of PM_{2.5} is significantly underpredicted, pointing to compensating error between secondary organic aerosol (SOA) formation and primary emissions within the models' photochemistry and emissions. Rapid SOA formation associated with both Houston and Dallas is inferred to occur within 1 to 3 h downwind of the urban centers, and none of the models reproduce this feature.

Citation: McKeen, S., et al. (2009), An evaluation of real-time air quality forecasts and their urban emissions over eastern Texas during the summer of 2006 Second Texas Air Quality Study field study, *J. Geophys. Res.*, 114, D00F11, doi:10.1029/2008JD011697.

1. Introduction

[2] Owing to the known human health effects of surface level O₃ and PM_{2.5} exposure, the need for air quality forecasts to the general public has prompted significant interest in the research, development, and deployment of real-time air quality forecast models (AQFMs) over the past several years. An important aspect of any long-term forecasting effort is establishing reliability on the basis of objective statistical evaluations that provide quantitative performance baselines and gauges for subsequent forecast model improvements. Like precipitation and observable weather, pollution near the ground is the result of transport and conversion processes occurring throughout the troposphere. Comparisons with upper air data are therefore necessary in order to adequately evaluate model predictions of low-altitude pollution. Unlike widely available upper air data for weather forecast model evaluations, detailed upper air photochemical and aerosol data necessary to evaluate air quality models are limited to intensive, and regionally specific field campaigns. The work presented here is an attempt to expedite the dissemination of high-quality upper air data collected during the TexAQS II field intensive to willing participants within the AQ forecasting community, and to optimize the scientific return of the model/measurement comparisons through multimodel and model ensemble evaluations.

[3] As part of the 2006 Second Texas Air Quality Study/Gulf of Mexico Atmospheric Composition and Climate Study (TexAQS II/GoMACCS), [Parrish *et al.*, 2009] conducted over eastern Texas and the Gulf of Mexico during the late summer of 2006, six operational and research institutions contributed their real-time forecast results to a central facility at the National Oceanic and Atmospheric Administration (NOAA) Environmental Sciences Research Laboratory (ESRL) Chemical Sciences Division (CSD). The models and research centers include: two versions (12 km and 36 km horizontal resolution) WRF/Chem model (Weather Research and Forecast model/Chemistry version 2.2) implemented by NOAA/ESRL Global Systems Division (GSD), results from both the operational CHRONOS (Canadian Hemispheric and Regional Ozone and NO_x System) and research AURAMS (A Unified Regional Air-quality Modeling System) models provided by Environment Canada, the NWS/NCEP (National Weather Service/National Center for Environmental Prediction) CMAQ/NAM (Community Multiscale Air Quality Model/North American Mesoscale) developmental forecasts, three forecasts (45, 15, and 5 km horizontal resolutions) from Baron AMS (Baron Advanced Meteorological System, Inc.) Corporation, and the University of

Iowa 12-km horizontal grid spaced STEM-2K3 (Sulfur Transport and Emissions Model 2003) forecasts. A more complete description of the models is given in section 2.

[4] Three-dimensional hourly fields of several key meteorological, radiation, and gas-phase atmospheric constituents from these 10 AQFMS were stored and graphically compared with corresponding real-time measurements from various measurement platforms available during the TexAQS II intensive. Day-to-day time series of comparisons at 14 surface and 11 wind-profiler sites in east Texas, as collected in real time, can be viewed at the NOAA/ESRL Physical Sciences Division (PSD) Internet Web address <http://www.etl.noaa.gov/programs/2006/texaqs/verification/>. Surface O₃ and PM_{2.5} forecasts for each model, as well as ensemble, bias-corrected ensemble, and Kalman-filter corrected ensemble forecasts are available for comparison at the surface sites in east Texas. To our knowledge, these were the first PM_{2.5} ensemble, bias-corrected and Kalman-filter ensemble AQ forecasts ever available in real time. Previous versions of these same models provided the basis for multimodel evaluations and ensemble studies specific to the summer of 2004 New England Air Quality Study International Consortium for Atmospheric/Research on Transport and Transformation (ICARTT/NEAQS) field study in the northeast United States and southern Canada [McKeen *et al.*, 2005; Pagowski *et al.*, 2005; Pagowski *et al.*, 2006; McKeen *et al.*, 2007; Wilczak *et al.*, 2006; Delle Monache *et al.*, 2008]. Model performance statistics derived from that 2004 study are compared and contrasted with the 2006 forecast results for surface O₃ and PM_{2.5} using observations from the U.S. Environmental Protection Agency (EPA) AIRNow (Aerometric Information Retrieval Now) network within section 3.

[5] Data used for upper air model evaluations were collected aboard the NOAA WP-3D Orion research aircraft stationed out of Houston and operating in east Texas and the Gulf of Mexico between 31 August and 12 October 2006. The scientific mission, instrument payload, and deployment strategies of the aircraft are given by Parrish *et al.* [2009] and the TexAQS II/GoMACCs planning document, available at the Internet Web address: <http://www.esrl.noaa.gov/csd/2006/> with specific references to air quality model verification associated with various science objectives within those documents. Section 4 presents direct comparisons between aircraft and model forecasts illustrating some deficiencies and inconsistencies within the various models. Owing to the large uncertainties and impact on O₃ and PM_{2.5} forecasts, one of the main objectives of TexAQS II involves emissions verification. Flight plans were designed and implemented to characterize the photochemical and aerosol composition upwind and downwind of the Houston

Table 1. Air Quality Forecast Model Name and Horizontal Grid Spacing, Abbreviated Mnemonic, Abbreviated Chemical Mechanism, Base Year of Anthropogenic Emissions Inventory, and Internet Web Addresses of the Seven Models^a

AQ Forecast Model (Horizontal Grid Spacing)	Mnemonic	Chemical Mechanism	Base Year Emissions	Internet Web Address
AURAMS (28 km)	AUR	ADOM-II	2001	http://www.msc-smc.ec.gc.ca/research/icartt/aurams_e.html
BAMS (15 km)	BAM	CBM-4	2001,2004	http://www.baronams.com/projects/SECMEP/index.html
CHRONOS (21 km)	CHR	ADOM-II	2001	http://www.msc-smc.ec.gc.ca/aq_smog/chronos_e.cfm
CMAQ/NAM (12 km)	CMQ	CBM-4	2001,2004	http://wwwt.emc.ncep.noaa.gov/mmb/aq/
STEM-2K3 (12 km)	STM	SAPRC-99	2001,2004	http://nas.cger.uiowa.edu/MIRAGE/mirage-2k6.html
WRF/Chem (36 km)	WRF-36	RADM2	1999	http://ruc.fsl.noaa.gov/wrf/WG11/
WRF/Chem (12 km)	WRF-12	RADM2	1999	http://ruc.fsl.noaa.gov/wrf/WG11/
Ensemble mean ^b	ENS			http://www.etl.noaa.gov/programs/2006/txaqs/verification/

^aAbbreviated mnemonic denotes individual models in Figures 2, 3, 5, 6, 8, 12, 13, and 14. See text for details of the anthropogenic and biogenic emissions. The Web addresses are for further information or details regarding each AQFM.

^bEnsemble results at 14 surface sites calculated in real time during the summer of 2006 are contained within this TexAQS II evaluation Web site.

and Dallas urban centers, allowing relative emission rates and, under certain assumptions, absolute emission rates to be determined. In section 5 model forecast upwind/downwind concentrations of several key gas-phase and aerosol constituents are compared to observations for eight flights around Houston, and two flights around Dallas during September of 2006. On the basis of a common definition of background (or upwind) concentration, relative emission rates for both models and observations are derived. The model emission ratios for several species, relative to NO_y, show consistent patterns of bias for Dallas and Houston when compared to observed ratios. The implications for air quality forecasts of O₃ and PM_{2.5} are discussed in section 6.

2. Air Quality Forecast Models

[6] Though a total of 10 nearly continuous model forecasts were available between 12 August and 30 September 2006, only the results of seven models having complete temporal coverage and a common spatial overlap are used. Table 1 summarizes some basic features (grid spacing, chemical mechanism, base year of the anthropogenic emissions inventory), and the Web address corresponding to each AQFM that links to further references, real-time forecast products, or TexAQS II applications. The domains of data availability for these models are shown in Figure 1a. The region of overlap for all of the models shown in Figure 1a is determined by the STEM-12 km and BAMS-15 km

model domain limits. Detailed background information on the models as they relate to both O₃ and PM_{2.5} forecasts are provided by McKeen et al. [2005] and McKeen et al. [2007], respectively. A brief summary of the modeling systems, including the underlying emission inventories used for this study, and updates that they may have undergone between the 2004 and 2006 forecast seasons are given in Appendix A.

[7] Participants within the TexAQS II forecast model evaluation project voluntarily provided hourly forecasts of O₃, PM_{2.5}, several precursor gases and meteorological variables, usually 48-h forecasts, in near real time. During planning of the project a collective agreement was reached to limit model information supplied to NOAA/ESRL/CSD to forecast fields only. Though emission inventories were acknowledged as important components needing evaluation, four factors prevented their inclusion within the centralized data set: some inventories were unavailable until the beginning of the experiment, some groups have emissions coupled to meteorological fields, the proprietary nature of some inventories, and the limited resources available for this unfunded project. The nature of the comparison study is informal by design, with a focus on identifying inconsistencies between measurements and the majority of forecast models. In this work we include the emission inventory of one forecast model (WRF/Chem) in order to demonstrate the importance of including emission inventories within future model evaluation studies, and to

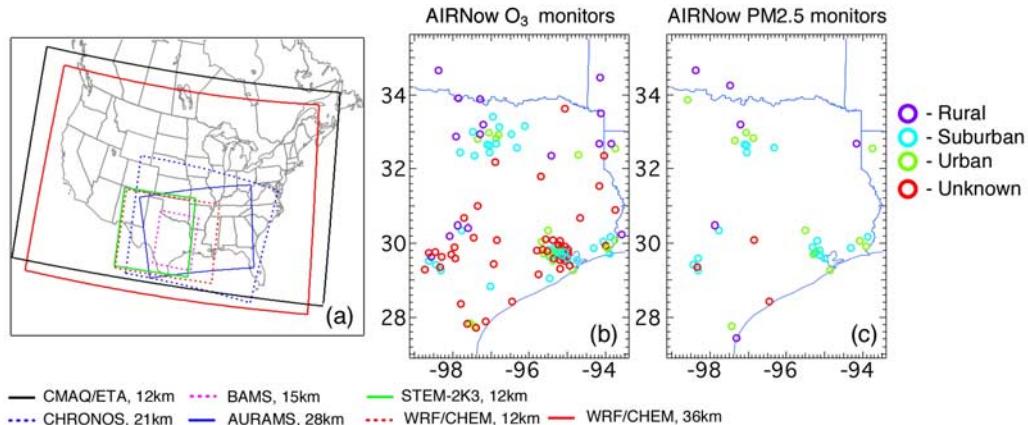


Figure 1. (a) Continental map of North America showing the model boundaries of the seven models used in the analysis and ensemble average. The domain of model overlap and the location of (b) 119 AIRNow O₃ monitors and (c) 38 PM_{2.5} monitors that reported data during the summer of 2006.

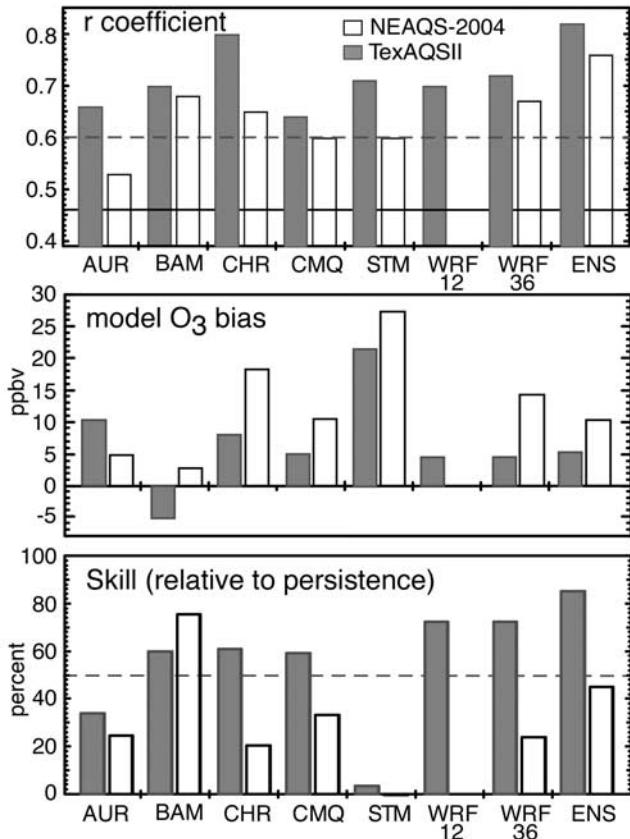


Figure 2. Comparison between the TexAQS II and NEAQS-2004 maximum 8-h average O₃ evaluation statistics for the median values of the (top) *r* correlation coefficient, (middle) mean bias in ppbv, and (bottom) forecast skill in percent for the seven AQFMs and the ensemble mean. Only comparisons at monitor locations having 30 or more days of available data are used (342 for NEAQS-2004 and 119 for TexAQS II). The value from persistence forecasts are shown as a horizontal dashed (TexAQS II) or solid (NEAQS-2004) lines, and are both 0 ppbv for bias and 50% for skill.

explore techniques for directly relating model forecast and observed concentrations to emission estimates.

[8] It is important to note that model results are based solely on the forecasts collected in real time during the summer of 2006. All of the AQFMs have since undergone modifications in numerical formulation or emissions inventory, and hence supersede the models used in this analysis. This evaluation merely represents a snapshot of the rapidly evolving field of air quality forecasting. The performance of current modeling systems cannot be inferred from the results presented here.

3. Model Statistics Based on Surface AIRNow O₃ and PM_{2.5} Observations

[9] In this section we focus on model/observation comparisons of two quantities of particular regulatory interest: daily maximum 8-h average O₃, and 24-h average PM_{2.5}. Daily maximum 8-h average O₃ data and hourly updated PM_{2.5} data for all available AIRNow monitors (<http://www.epa.gov/airnow/2006>) within the domain of overlap in Figure 1 are used in the analysis. The 24-h period starting from 0000 eastern daylight time (EDT) are used for the PM_{2.5} average comparisons. The location of the 119 stations within the domain of model overlap is shown in Figure 1b along with some information on surrounding population density. Figure 1c is the equivalent plot for available AIRNow PM_{2.5} monitors within the domain of overlap. The AIRNow procedure for accepting an 8-h maximum concentration is quite rigorous. If more than 2 hourly averages are missing within a 24-h period a missing value for that day is reported. Consistent with previous model evaluations, we additionally require that a monitor have more than half of the available days available for comparison. Four stations within Figure 1b failed to have more than 25 daily maximum O₃ values available for the sampling period considered, and two stations within Figure 1c failed to have more than 25 days with 22 h or more of PM_{2.5} data. These stations were eliminated from the statistical comparisons, leaving 115 stations for O₃ and 36 stations for PM_{2.5} within the sample region.

[10] The 49-day period between 0000 UT 12 August 2006 and 0000 UT 30 September 2006 is the sampling period used in this analysis. The statistical evaluation is only for results from the first 28 h of the 0000 UT forecasts. Daily values of maximum 8-h average O₃ are calculated from the 0000 UT forecasts between 0400 UT and 28 h into the 0000 UT forecast to match the definition of daily maximums used within AIRNow. In the case of CMAQ/NAM (no 0000 UT forecast, instead starting at 0600 UT), the 0000 UT to 0600 UT O₃ and PM_{2.5} are filled with the previous day's forecast for those times. Likewise, most models are missing 1 or 2 days of forecasts owing to system disruptions, power outages or other reasons. Missing days of forecast data are filled with the second day of the 48-h forecast from the previous day, allowing all 49 days with coincident and ensemble results for O₃, and 48 days (starting 13 August 2006) for PM_{2.5}.

[11] Latitudes and longitudes of the AIRNow O₃ and PM_{2.5} monitoring stations falling within the domain of model overlap are mapped into the grid coordinates of each model. Comparisons with observations assume that model values are uniform over a model grid, and observed O₃ values are compared with model grid values that contain each monitor. Thus, no spatial interpolation is performed on the models, but depending on model resolution, O₃ observations from several monitors could be evaluated against results from only one model grid. Mean ensemble model data sets are constructed from the results of the seven AQFMs on the common time base for each monitor location by taking the arithmetic average of all seven model hourly O₃ forecasts. Although alternative ensemble formulations are possible, such as the ensemble determined from the median model [e.g., Galmarini *et al.*, 2004], or the weighted approach presented by Pagowski *et al.* [2005], ensemble statistics are not very sensitive to the exact formulation of the ensemble based on previous ensemble comparisons [McKeen *et al.*, 2005].

3.1. Standard or Bulk Statistics for O₃ and Comparison to 2004 ICARTT/NEAQS Results

[12] Standard statistical measures [see, e.g., Tilmes *et al.*, 2002] representing overall median conditions for the

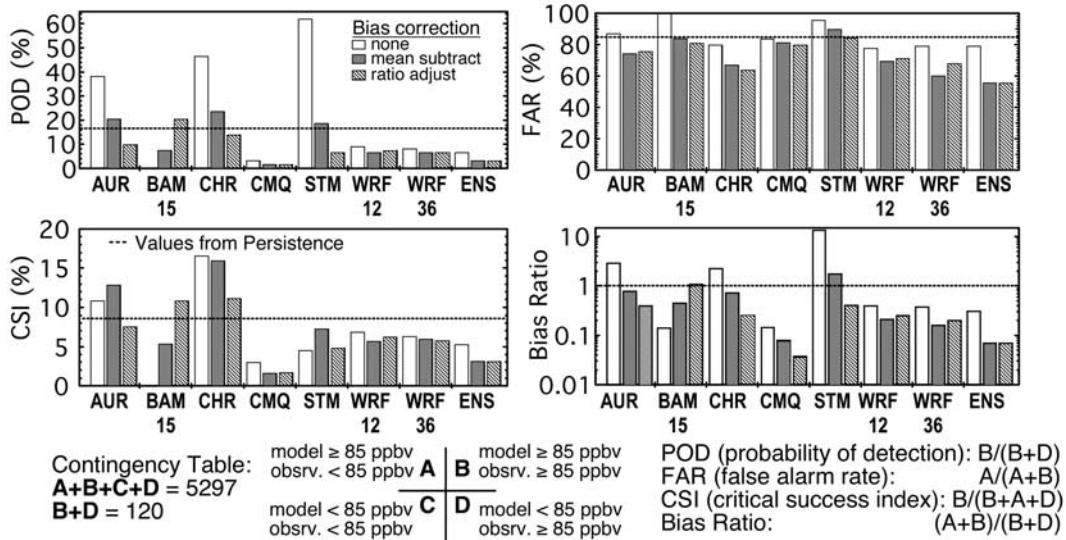


Figure 3. Categorical statistics for the 12 August to 30 September 2006 TexAQS II O_3 evaluations based on 85-ppbv threshold limit for maximum 8-h average O_3 . The contingency table and definition of the four metrics are given at the bottom. Statistics are given for the uncorrected model and for two bias correction techniques. Dashed horizontal lines are statistics derived from the persistence forecast.

domain of model overlap are given in Figure 2, which shows the median values of the correlation coefficient,

comparisons only encompass a third of the land area of the NEAQS-2004 study, high- O_3 events, and weather patterns

$$r(i) = \frac{\sum_{\text{days}} (O_3^{\text{modl}}(i, \text{day}) - O_3^{\text{modl}}(i, \text{avg})) (O_3^{\text{obs}}(i, \text{day}) - O_3^{\text{obs}}(i, \text{avg}))}{\sqrt{\sum_{\text{days}} (O_3^{\text{modl}}(i, \text{day}) - O_3^{\text{modl}}(i, \text{avg}))^2 \sum_{\text{days}} (O_3^{\text{obs}}(i, \text{day}) - O_3^{\text{obs}}(i, \text{avg}))^2}}, \quad (1)$$

the median value of the mean bias,

$$\text{Mean Bias}(i) = \left(\frac{1}{N_{\text{days}}} \right) \sum_{\text{days}} [O_3^{\text{modl}}(i, \text{day}) - O_3^{\text{obs}}(i, \text{day})], \quad (2)$$

and the model skill, defined as the fraction of points with root mean square error (RMSE) less than that of persistence (previous day's observations), where

$$\text{RMSE}(i) = \sqrt{\left(\frac{1}{N_{\text{days}}} \right) \sum_{\text{days}} (O_3^{\text{modl}}(i, \text{day}) - O_3^{\text{obs}}(i, \text{day}))^2}, \quad (3)$$

and i refers to O_3 monitor i ($i = 1$ to 115), N_{days} refers to number of observing days at each site, “obs” refers to observed, and “modl” refers to model values (or previous day's observations in the case of persistence calculations). These three quantities are shown for the seven AQFMs, the ensemble mean for the maximum 8-h average O_3 , and for both the TexAQS II as well as those reported by *McKeen et al.* [2005] for the 2004 ICARTT/NEAQS study period. Because of the large sampling sizes in both data sets, all statistical parameters, including the r coefficient from the persistence forecasts, are significantly different between the 2006 and 2004 studies at confidence levels greater than 95% [*McKeen et al.*, 2005].

[13] The average O_3 monitor densities (~ 3 per 10,000 km 2) are nearly identical in both studies, but the TexAQS II

associated with frontal passages in general, tend to occur over the entire TexAQS II study region with a weekly periodicity (J. Wilczak et al., Evaluation of meteorology and surface ozone in NMM-CMAQ and WRF-Chem simulations during the TexAQS II field program, submitted to *Journal of Geophysical Research*, 2008), whereas conditions during the ICARTT/NEAQS-2004 study showed less temporal regularity and coherence over the much larger study area. This is a probable reason for the higher r coefficients for the persistence forecasts, as well as the model forecasts to some degree. On the other hand all models have undergone significant modifications attempting to improve O_3 forecasts (see Appendix A for examples), and all models but WRF/Chem use updated emission inventories. Figure 2 also shows the improvement gained by the model ensemble compared to any individual model for both study periods, and for the given standard statistical parameters.

3.2. Threshold or Categorical O_3 Statistics for the Models and Ensemble

[14] Figure 2 provides an analysis of median or bulk conditions for O_3 , but a main justification for real-time O_3 forecasts is the information they provide for public pollution exposure and health advisories [e.g., *Dabberdt et al.*, 2004]. For health-related issues the useful information is not contained in the bulk statistics, but rather in the occurrence of maximum 8-h average O_3 values greater than a particular threshold value. Two threshold values that have previously

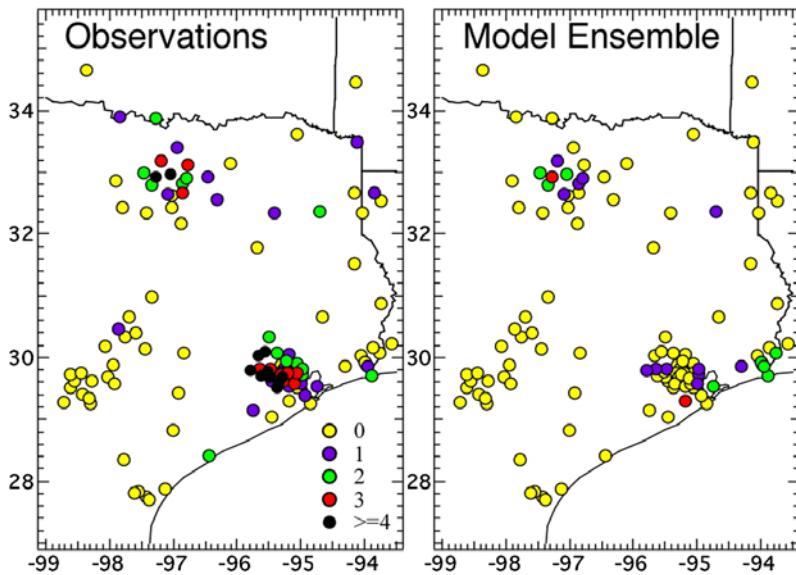


Figure 4. Number of occurrences of maximum 8-h average O_3 greater than 85 ppbv between 12 August and 30 September 2006 for (left) the observations and (right) the seven-model ensemble forecast.

been considered are the 125-ppbv maximum 1 h average O_3 limit, and the 85-ppbv maximum 8-h average O_3 limit [McHenry *et al.*, 2004; Kang *et al.*, 2005], but more recent regulatory measures are based on the 8-h average maximum, so only threshold statistics for the 85 ppbv maximum 8-h averages are presented here. The O_3 threshold statistics are detailed by McKeen *et al.* [2005] and are based on fractional occurrence of model versus measured O_3 within the contingency table shown in Figure 3, similar to threshold statistics for precipitation comparisons used in meteorological forecast models.

[15] Figure 3 shows four quantities that are used in the threshold statistical analysis; probability of detection (POD), false alarm rate (FAR), critical success index (CSI), and the bias ratio. For a perfect model the POD would be 100%, FAR would be zero, CSI would be 100%, and the bias ratio would be unity. Figure 3 shows these threshold statistics for the 7 AQFMs, the ensemble, and the impact of two bias correction techniques applied to each monitor comparison; the standard mean subtraction bias removal, and the multiplicative ratio adjustment correction [McKeen *et al.*, 2005]. The statistical values from the persistence forecast are also shown. Bias removal from a forecast tends to move the entire group of model points vertically relative to the 85-ppbv limit of the contingency table, giving preference to either a high POD or low FAR at the expense of the other. The CSI represents a balanced measure of the combined FAR and POD if the importance of model and observed exceedances are equally weighted. The bias ratio represents a bulk measure of the model threshold exceedance ratio relative to the observed ratio. It should be noted that only 120 observations exceed the 85-ppbv maximum 8-h average threshold (out of 5297 total points), and thus the threshold analysis is for a relatively small sample at the very tail of the observed O_3 population distribution.

[16] The threshold statistics shown in Figure 3 can be compared to the equivalent results for the ICARTT/NEAQS-2004 study [McKeen *et al.*, 2005, Figure 7a]. First,

the persistence forecast has decreased FAR (95% to 85%) and increased POD (5% to 17%) and CSI (2.7% to 8.5%). Similar to the increase in r coefficients for 2006 compared to 2004 discussed above, this increase in persistence reliability is probably due to more coherence in the meteorological forcing associated with TexAQS II. Bias ratios from 2004 to 2006 are quite different between corresponding models with six out of seven models plus the ensemble having bias ratios much larger than unity in 2004. The ensemble in particular (but also the WRF/Chem and CMAQ models) shows a fundamental inconsistency between the biases derived from the bulk statistics (all positive in Figure 2), and the number of 85-ppbv exceedances (underestimated by more than 50% in Figure 3) so that bias corrections under these conditions reduce PODs below unacceptably low values. The Baron Advanced Meteorological Systems, Inc. (BAMS) model is the only model with a negative bias in Figure 2, so bias correction works in the right direction to increase low POD with a multiplicative bias correction giving a CSI slightly above persistence. Five of the seven models, like persistence, show increased CSI from 2004 to 2006, but only two models have CSI better than persistence in 2006 as opposed to four during 2004. The ensemble in particular shows decreased CSI (15% to 5%) with bias corrections to the ensemble making CSI worse, not better as in the 2004 study. The one threshold quantity that is similar between all models, ensemble, and for both study periods is the FAR, which is never less than 75% for any uncorrected case and limits the magnitude of CSI. This may be due in part to a fundamental limitation in the AQFM comparisons with hourly monitor data that may be specific to a given site, ambient traffic or local land-use conditions not characterized by coarse-grid models. Short of correcting the underlying deficiencies within the models, more sophisticated correction strategies, such as ensemble-based Kalman filtering and probability forecasts, or direct data assimilation within the AQFMS are needed to circumvent the high FAR values found in this study as well as the 2004 evaluation.

[17] The tendency for most of the AQFMs to underpredict the number of 85-ppbv O₃ exceedances is also shown in Figure 4, where the number of exceedances for the ensemble model is used in the comparison, since the ensemble represents the outcome of “most models” and also provides the best bulk statistical comparisons in Figure 2. While most of the observed exceedances occur in the vicinity of the Houston/Galveston region, very few exceedances occur in this region for the model ensemble. This contrasts with the Dallas/Fort Worth region, which shows some spatial correlation between model ensemble and observed exceedance numbers, though total exceedance numbers in this region are also somewhat underpredicted by the ensemble. The large discrepancy in the Houston region relative to Dallas/Fort Worth suggests something unique about the photochemistry and/or meteorology near Houston that the AQFMs fail to reproduce. Comparisons with aircraft data given below (sections 4 and 5) suggest that highly reactive anthropogenic VOC emissions (ethylene in particular) are significantly underpredicted relative to NO_x emissions in the Houston region, whereas inferred ethylene/NO_x emission ratios for the Dallas/Forth Worth region (based on only one available upwind/downwind transect pair) are more consistent between all of the AQFMs and observations.

3.3. Standard or Bulk Statistics for PM_{2.5} and Comparison to 2004 ICARTT/NEAQS Results

[18] Similar to the statistics presented for O₃ above, bulk statistical measures for real-time AQFM predictions of PM_{2.5} are derived for the 13 August to 30 September 2006 period for the 36 AIRNow PM_{2.5} monitors shown in Figure 1c, and compared with equivalent statistics for the ICARTT/NEAQS study period in the northeast United States and southeastern Canada during the summer of 2004 [McKeen et al., 2007]. Current regulatory interest is focused on 24-h average concentrations, which are used in the AIRNow comparisons here (0400 UTC to 2800 UTC model forecasts) for the 13 August to 30 September 2006 period. McKeen et al. [2007] used log-transformed concentrations in the analysis of the AIRNow PM_{2.5} data within the 2004 study based on the probability distribution functions (and chi-square analysis) of that data set. For consistency in the comparisons, log-transformed concentrations are also applied to the 2006 AIRNow PM_{2.5} analysis, and used to determine *r* coefficients and ratio-equivalent RMSE, and model bias is viewed in terms of model to observed ratios.

[19] Figure 5 illustrates the median *r* coefficients, model bias, and forecast skill for the seven AQFMs and their corresponding statistics for the NEAQS-2004 study. Similar to the case for O₃, *r* coefficients for persistence are higher (and persistence RMSEs are lower) for 2006 compared to 2004, illustrating the different sampling conditions between the two cases. But exactly the opposite as O₃, all of the AQFMs show reduced statistical performance for PM_{2.5} in 2006. Although the ensemble represents the best forecast for both 2004 and 2006, it is unable to improve upon the persistence forecasts in 2006, and like all but one model is biased low by 14% or more in the median.

[20] The decreased performance and low bias suggests that fundamental PM_{2.5} processes are missing within the models during TexAQS II. Indeed, Saharan dust transport

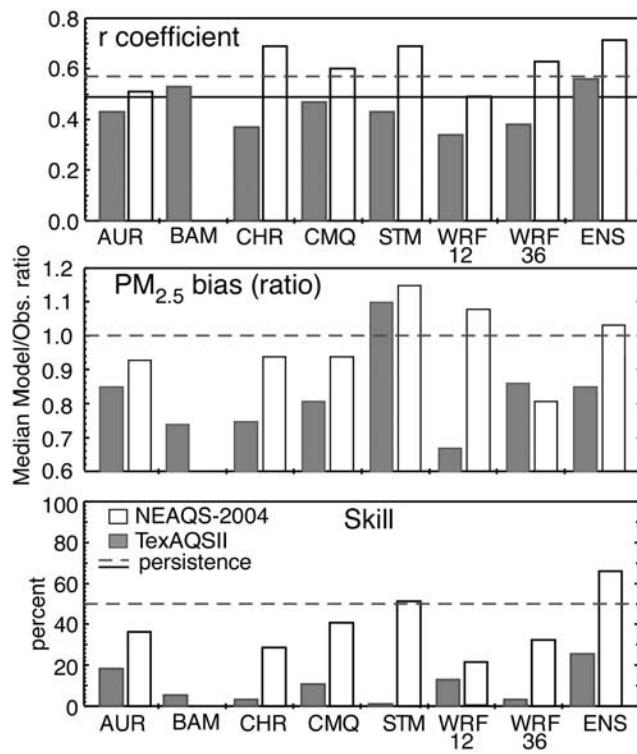


Figure 5. Comparison between the TexAQS II and NEAQS-2004 24-h average PM_{2.5} evaluation statistics for the median values of (top) the *r* correlation coefficient, (middle) bias as model/observed ratio, and (bottom) forecast skill in percent for the seven AQFMs and the ensemble mean. Only comparisons at monitor locations having 30 or more days of available data are used (118 for NEAQS-2004 and 38 for TexAQS II). The values from persistence forecasts are shown as horizontal dashed (TexAQS II) or solid (NEAQS-2004) lines and are both 1.0 for bias and 50% for skill.

events (i.e., 27–30 August 2006 and 20–24 September 2006) influenced coastal PM_{2.5} monitors, and to a lesser degree monitors as far as 150 km north of the Gulf of Mexico (e.g., Conroe, Texas). Only one exceedance of the 24-h average 35 $\mu\text{g}/\text{m}^3$ regulatory limit occurred for the study period and domain considered in TexAQS II, and it was associated with a Saharan dust event (Corpus Christi, Texas, 29 August 2006). But removing 11 days of comparisons (period of impact plus 1 day) has only a minor effect on the model statistics (ensemble model: median *r* coefficient = 0.60, median bias ratio = 0.83, skill = 28%), suggesting another cause of the reduced model PM_{2.5} performance. One possibility is a collective misrepresentation of nighttime planetary boundary layer (PBL) dynamics and depositional loss. For NEAQS-2004, McKeen et al. [2007] showed that the diurnal trends for urban and suburban locations were not reproduced adequately, and that the 24-h model averages were heavily skewed toward nighttime concentrations for most models, while the observations were not. This does not appear to be a factor for TexAQS II, since statistics based only on midday 8-h averages (not shown) are nearly identical to those shown in Figure 5 for TexAQS II, but were significantly different (much improved

skill with an 11% median ensemble low bias) for NEAQS-2004. PM_{2.5} sources not explicitly included in the models, such as regional-scale biomass burning, long-range transport from wild fires, or organic aerosol (OA, the sum of primary emitted organic aerosol and SOA) underestimates from anthropogenic sources discussed further below, could also be factors in the reduced PM_{2.5} model performance. But compared to O₃, PM_{2.5} does not appear to be a serious air quality concern in Texas during summer (overall median 24-h average = 10 $\mu\text{g}/\text{m}^3$) except for the occasional Saharan dust events along the coast.

4. Comparison of Model Forecasts With NOAA WP-3D Aircraft Data

[21] During the summer of 2006 the TexAQS II/GoMACCS intensive field study included several mobile platforms; five aircraft, smart balloon sondes, a mobile solar occultation flux van, and a research vessel. The NOAA WP-3D aircraft and the research vessel *Ronald H. Brown* included detailed photochemistry, aerosol composition and size measurements within their payloads. The NOAA WP-3D aircraft spent a significant fraction of its allotted flight hours between 300 and 700 m above the ground, with 10 daytime flights between 13 and 29 September 2006 consisting of upwind and several downwind transects of the Houston and Dallas urban areas. With several vertical profiles included in each flight, the WP-3D composition data provide valuable observations with which to compare upper air forecast model predictions, as well as diagnostic information to help explain the biases in surface O₃ and PM_{2.5} noted in sections 3.1 and 3.3. Five additional flights between 5 and 12 October 2006 are not included in this analysis since some of the model forecasts terminated at the end of the official ozone season on 30 September 2006.

[22] In late March of 2007 most of the data sets collected by the NOAA WP-3D were finalized and made available to experiment participants. A public Web site (<http://esrl.noaa.gov/csd/2006/modeleval/>) was constructed that overlays results from eight AQFMs with results from the NOAA WP-3D. Thousands of plots are available at this site showing every vertical profile and horizontal transect of 12 NOAA WP-3 flights for several chemical, meteorological, aerosol and radiation variables (O₃, CO, NO, NO₂, NO_y, HNO₃, PAN, NO_x, NO₃, N₂O₅, SO₂, NH₃, PM_{2.5} composition (SO₄²⁻, NH₄⁺, OA (organic aerosol, the sum of primary emitted organic aerosol and SOA), NO₃⁻, elemental carbon (EC)), total sulfur, total NH₃/NH₄⁺, isoprene, CH₂O, CH₃CHO, C₂H₄, toluene, xylenes, CH₃COOH, temperature, virtual potential temperature, H₂O vapor, relative humidity (RH), wind speed, wind direction, J^{N₂O₂}, HO₂). The following analysis is based upon the comparisons within the evaluation Web site with focus on CO, ethylene, NO_y, PM_{2.5}, and PM_{2.5} composition. The reader may consult this site for comparisons of the individual models or for variables that are not covered here. All of the models evaluated against AIRNow O₃ and PM_{2.5} surface data in section 3 are also evaluated in this section. However, several models were unable to provide results for all species. The BAMS model did not provide upper air PM_{2.5} information, the CMAQ/NAM model only provides composition information (except O₃) below 2.5 km and did not provide gas-

phase VOC information, while the CHRONOS model did not provide meteorological variables other than RH.

4.1. Observations and Details of Analysis

[23] Details of all the gas-phase, aerosol and actinic flux instrumentation aboard the NOAA WP-3 available during TexAQS II/GoMACCS and used in the following comparisons can be found within *Parrish et al.* [2009]. Only references to measurement techniques for the species considered in this work are provided here. One-second O₃, CO and NO_y measurements were made with the same NOAA WP-3 instrumentation as that of the ICARTT/NEAQS-2004 experiment [*Fehsenfeld et al.*, 2006]. Ethylene is measured at 10-s resolution with the photoacoustic laser spectroscopy technique described by *Kuster et al.* [2005]. Elemental carbon is determined by single particle soot photometer measurements [*Schwarz et al.*, 2006, 2008]. Size (and volume) distribution measurements of aerosol were made aboard the NOAA P-3 at 1-s time resolution by laser optical particle counters, similar to the measurements made during the ICARTT/NEAQS-2004 field project [*Brock et al.*, 2008]. The aerosol size cutoff is $\sim 1.0 \mu\text{m}$ diameter for this technique, a fixed refractive index for all particles is assumed, and a density of 1.6 g/m³ is assumed in order to convert size distribution data to mass mixing ratio. The PM_{2.5} comparisons therefore have large inherent uncertainties (estimated to be 30–40%), which must be considered in the analysis. Ten-second average aerosol composition (OA, sulfate, nitrate, ammonium) was measured by time of flight aerosol mass spectrometry (R. Bahreini et al., Organic aerosol formation in urban and industrial plumes near Houston and Dallas, Texas, submitted to *Journal of Geophysical Research*, 2008) with an upper cutoff diameter for this instrument of 1.0 μm . Like total PM_{2.5}, model predicted organic mass (organic PM_{2.5}) and observed organic aerosol (OA) are not precisely identical quantities. Analysis of the size distribution data coincident with AMS data suggests the differences between 1.0 μm and 2.5 μm cutoff diameter are typically much smaller than the 30 to 35% OA uncertainty attributed to the measurements (Bahreini et al., submitted manuscript, 2008). The two terms, organic PM_{2.5} and organic aerosol (OA), are therefore used interchangeably in the following discussions.

[24] Numerically, comparisons are done by flying the aircraft through each model domain using the three-dimensional model field specific for each flight, and for the nearest hour of model output. If the aircraft flies through a model grid cell, the observational average is calculated for the time spent in that grid, and the model value at the nearest hourly time slice for that grid is also recorded, regardless of the time spent in the grid cell. For species measured over sample times longer than 1 s (e.g., ethylene and aerosol composition) the averaging time may span locations within two model grids. In this case both model grid values are compared against the single average over the sample time. Similar to the surface comparisons described previously, there is no interpolation of model or observed data either in space or time in the comparisons. Further refinements to the comparisons could include a more rigorous way of limiting comparisons to well sampled grid cells, or weighting of averages according to time spent in grid. Here we rely on comparisons of median values or

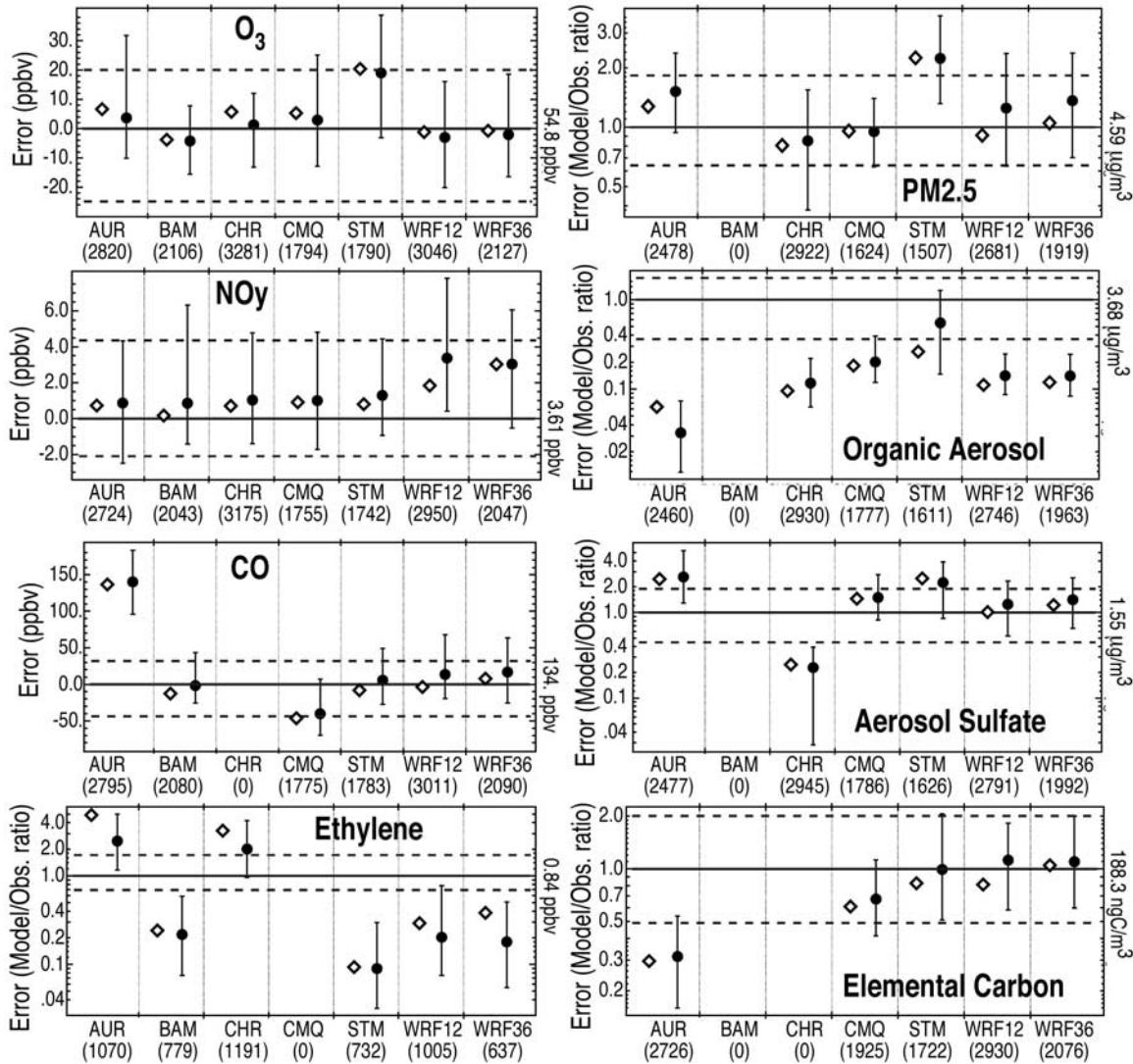


Figure 6. Solid circles show median biases (ppbv) for O₃, NO_y, CO, and median model/observed ratios for ethylene, PM_{2.5}, organic PM_{2.5} aerosol, sulfate PM_{2.5} aerosol, and EC PM_{2.5} aerosol for comparisons with the NOAA WP-3 aircraft data between 400 and 670 m altitude, 1000 to 1800 CDT, and all flights between 11 and 27 September 2006. Open diamonds are analogous biases for only the upwind transects between 13 and 29 September 2006 listed in Table 2. Solid horizontal lines represent perfect agreement, and dashed lines give the central two thirds of the sorted observation distribution relative to the observed median listed on the right of each plot. Model limits are the central two thirds of the sorted model error distributions. Numbers in parentheses gives the number of comparison points for each model.

median errors which should be relatively insensitive to these sampling and averaging issues.

4.2. Bias Statistics for the Daytime 400- to 670-m Window

[25] Along with the detailed time series comparisons, a set of summary statistics for each of the seven AQFMs, and for each of the variables, is also provided within the NOAA WP-3 aircraft comparison Web site. Figure 6 shows AQFM bias statistics for several gas-phase and aerosol species for daytime flight legs between 400 and 670 m above ground level (AGL). Focusing first on O₃, the aircraft biases in Figure 6 show close correspondence and correlation with the surface O₃ biases (Figure 2, middle). This close connection between aircraft and surface O₃ model bias implies

that the full suite of chemical measurements aboard the WP-3 can be used to diagnose sources of bias at the surface. The comparison between PM_{2.5} biases in Figure 6 with the corresponding surface biases (Figure 5, middle) shows no such correspondence. This is not surprising since the aircraft measurements are at midday while the PM_{2.5} surface statistics are for 24-h averages, and surface PM_{2.5} tends to have peaks in the early morning and late at night [McKeen et al., 2007].

[26] Also shown in Figure 6 are the bias statistics for only those transects that are upwind of either Dallas or Houston, which represents less than 10% of the complete data set. In most cases the biases determined from just the upwind transects are representative of the biases for all of the data in the selected height and time window, suggesting that

sources of the overall model biases are to a large degree independent of emissions or emission errors from the Dallas and Houston urban complexes. This is not absolutely true, since there is a case of recirculation affecting upwind values (27 September 2006 for Houston), Dallas's downwind becoming Houston's upwind (9/25/06), and the probability that emission errors for these urban corridors are endemic to all urban or anthropogenically perturbed regions. Nonetheless, Figure 6 is useful for assessing relative model performance in order to identify gross model errors and inconsistencies.

[27] It is useful to point out the reasons for certain obvious biases, since they have direct bearing on the emissions analysis below. The relatively high biases in NO_y for the WRF/Chem models appear to be related to underestimated PBL heights and weak mixing for many flight days, as discussed further below. The WRF/Chem results also tend to show the largest difference between upwind biases versus total biases in Figure 6, which is probably related to weak PBL dynamics. CO biases are very high for the AURAMS model, which is due to unrealistically high boundary specifications (~160 ppbv near the surface). The low CO bias for the CMAQ model was determined to be due to an error in the assignment of its deposition velocity. Both the AURAMS and CHRONOS models display very high background ethylene. Within the original ADOM mechanism, isoprene-OH reaction assumes ethylene as a direct product, and this original mechanism has not been changed to reflect updates to the isoprene oxidation mechanism in either of these models. The high PM_{2.5} biases for the STEM model are due to an error in NH₃/NH₄ partitioning resulting in all NH₃ being incorporated into the particulate phase.

[28] There are two features common to all models within Figure 6 that reflect uncertainties in emissions or photochemical processing. Ethylene biases for all models is higher for the upwind subset compared to the biases using all data, and all but the AURAMS and CHRONOS models have significantly lower ethylene compared to observations. This argues for a common underestimation of ethylene emissions, mostly from the Houston urban corridor where a majority of the sampling occurred. The other obvious deficiency common to all models in Figure 6 is the underprediction of OA. Observed dry aerosol mass is typically dominated by OA during this study, yet the models' total PM_{2.5} are greater than or match the measurements, with one model biased low and only by 20%. This underprediction of OA, and agreement in total PM_{2.5} points to a mismatch between observed and modeled aerosol composition. As noted previously [McKeen et al., 2007; Mathur et al., 2008] this discrepancy is explained by an overestimation of primary, unspeciated PM_{2.5} emissions compensating the lack of secondary OA formation within the models. The data within the evaluation Web site also confirms that the sum of sulfate, nitrate, ammonium, organic PM_{2.5} and EC is typically less than 50% of total PM_{2.5} for all models except STEM.

4.3. Flight by Flight Analysis: 25 September 2006 as an Example

[29] Figure 7 shows an example of the flight-by-flight information depicted in the public Web site for 25 September

2006. This was a day with upwind/downwind transects for both Dallas and Houston made by the NOAA WP-3. Results from the seven models and observations of water vapor and NO_y are shown for vertical profile number 8 in Figure 7c, and the NO_y comparisons for nearby 470-m horizontal transect number 6, about 80 km downwind of Dallas city center in Figure 7d. Satellite images for this day show clear skies under northerly flow, and all models likewise exhibit low RH and cloud-free conditions. Yet Figure 7c shows a large disparity in the PBL height as determined from the sharp water vapor decreases near 1.4 km. Comparing to PBL heights determined from a wind profiler network over the August–September 2006 time period, J. Wilczak et al. (submitted manuscript, 2008) show that WRF/Chem PBL heights are biased low and CMAQ/NAM PBL heights are biased high, similar to the results in Figure 7c. Moreover, the highly correlated and uniform PBL mixing seen in the observed NO_y and water vapor traces does not always occur in the models, with the WRF/Chem model appearing to mix NO_y too slowly, and the CMAQ/NAM and STEM models uniformly mixing NO_y to heights much higher than water vapor. NO_y model biases for the horizontal transects shown in Figure 7d correlate to some degree with the degree of vertical mixing shown in Figure 7c; the WRF/Chem and BAMS models are biased somewhat high and also have the lowest PBL heights from water vapor. Most models predict a single broad peak, and most models show offsets of NO_y plumes to the west, consistent with an underestimate of an observed westerly wind component not in the models (five minutes of transect time, or plume displacement, is ~30 km distance, corresponding to ~20° in wind direction error). Though Figures 7c and 7d are single examples of a single flight, they illustrate some practical considerations that must be kept in mind when comparing model results with aircraft observations, first that comparisons of absolute amounts and model biases can depend heavily on the model's formulation of PBL mixing and other aspects of fundamental meteorological accuracy, second that urban plumes in the models may not be fully sampled with direct aircraft observations. To compare the accuracy of emissions between model and observations, and the resulting downwind photochemistry, some form of normalization, such as relative increases in compounds downwind of the urban areas, must be considered.

4.4. Flight by Flight Analysis: Definition of Background and Excess Values

[30] In the upwind/downwind analysis of the Houston and Dallas urban plumes, every horizontal transect from each flight between 15 and 29 September 2006 is assigned a distance from either Houston city center (Harris County Courthouse, 29.7597°N latitude, 95.37065°W longitude) or Dallas city center (corner of Young and St. Paul streets, 32.77754°N latitude, 96.79568°W longitude) to the closest point of the transect, with upwind distances assigned negative and downwind distances positive. A “background” concentration for each species and for each flight is defined from the transect farthest upwind, and is taken to be the value at the 1/8 quantile of the sorted distribution. The uncertainty in this upwind concentration is defined to be half the difference between the 1/4 quantile and the minimum value of the sorted distribution for the transect,

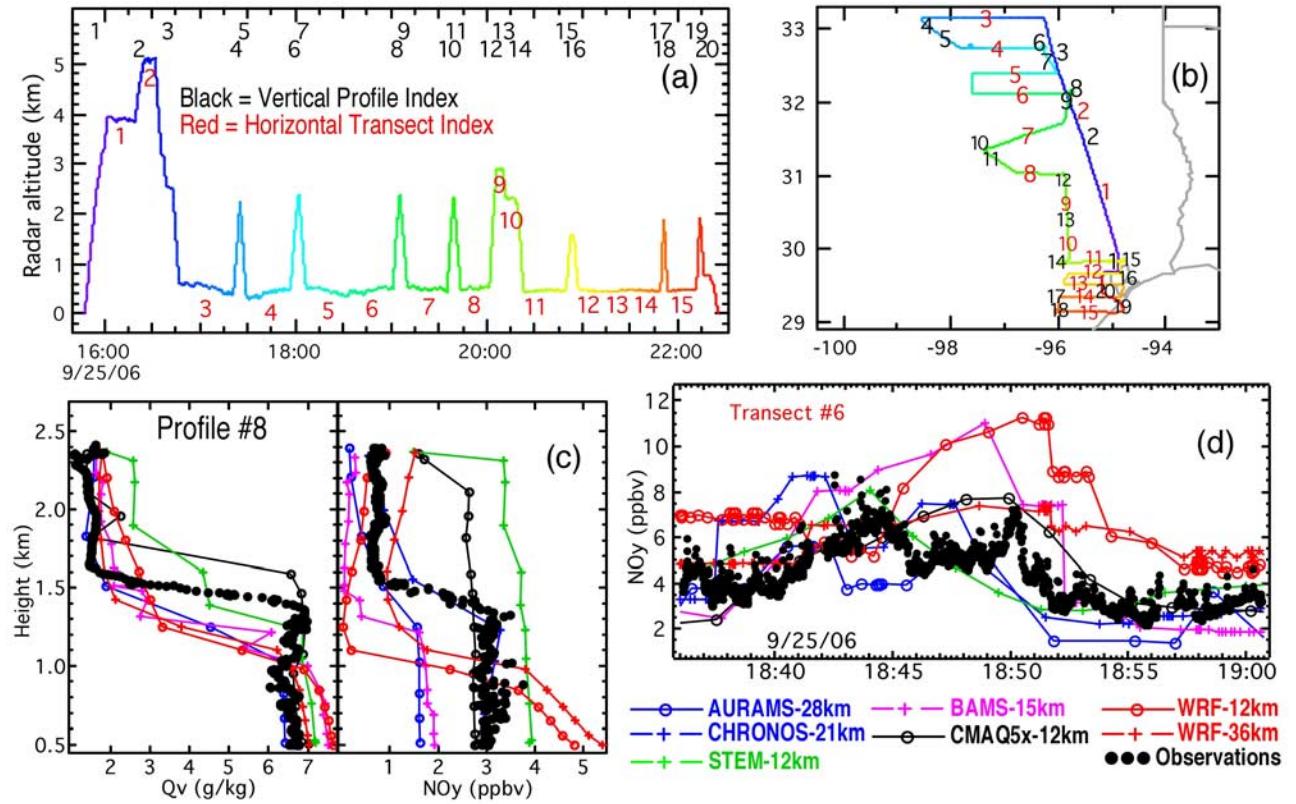


Figure 7. (a and b) Altitudes and locations of the center times for the vertical profiles (black numbers) and horizontal transects (red numbers) enumerated for the flight of 25 September 2006. The color scale denotes time. (c) Water vapor and NO_y vertical profiles for the models and observations for vertical profile number 8 centered at 1903:18 UTC. (d) Model and observed NO_y along horizontal transect number 6, about 80 km downwind of Dallas city center.

and is tracked for uncertainty propagation of ratios and fluxes presented later. Defining the background from the 1/8 quantile of the upwind transect is an attempt to filter out contributions from extraneous upwind sources occurring in both the observations and models. The background values are determined identically for all models and observations, however the coarser resolution models, WRF/Chem-36 km in particular, can have less than eight samples along a transect, in which case the background is taken to be the minimum value. Additionally, for flux calculations a plane is defined with orientation defined by the median of the aircraft-heading angle for each transect. The “excess flux” of a species along a transect is then defined as

$$\text{Excess Flux} = \int \rho(\chi - \chi_B) (\vec{v} \cdot \vec{n}) H(\chi, \chi_B) ds, \quad (4)$$

where ρ is air density (mol/m³), χ is volume mixing ratio, χ_B is the background value as just described, \vec{v} is the horizontal wind vector, \vec{n} is the unit vector normal to the transect plane, $H(\chi, \chi_B)$ is the Heavyside function (equal to 1 if $\chi > \chi_B$, otherwise equal to zero) and ds is an incremental distance parallel to the transect plane. Flux units are specified in mol/h (per vertical meter) for gas-phase species, and kilogram/h (per vertical meter) for

aerosol species. Similar to an excess flux, an “average excess above background” is defined,

$$\text{Average Excess Above Background} = \frac{\int H(\chi, \chi_B) \bullet (\chi - \chi_B) ds}{\int H(\chi, \chi_B) ds}. \quad (5)$$

Notice that plume width is normalized out of this definition, and the most upwind transect may have an average excess due to contributions from upwind sources and the definition of the background value. Figure 8 shows the background, average excess, and excess flux for NO_y for the 400- to 650-m AGL flight legs downwind (south) of Dallas on 25 September 2006. The remaining comparisons for the other days of the campaign are available on the Web site. The uncertainty associated with the background (~ 0.6 ppbv for the observations) is significantly less than the average excess values determined for Dallas, and average excess values tend to agree between models and observations for the three transects greater than 15 km downwind of Dallas city center. The model excess fluxes, on the other hand, show more variability and departure from the observations due to errors in the mean wind speed perpendicular to the transect plane (see Web site for details). The high excess NO_y flux over Dallas (10 km downwind of city center in Figure 8) is a factor of 2.5 higher than the next downwind

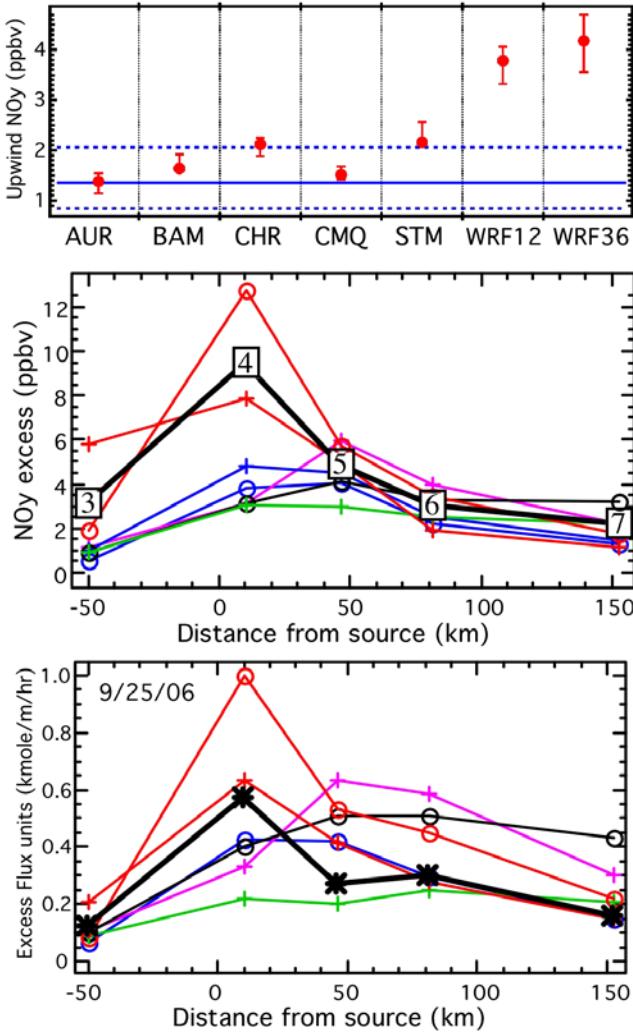


Figure 8. (top) Upwind (or background) values assigned to NO_y for the 450- to 600-m AGL transects over Dallas on 25 September 2006. The assigned value for the observations is the solid horizontal line, with uncertainty limits shown as dashed lines. Model values are red circles with limits showing the 1/4 quantile (upper) and minimum value (lower). (middle) Average NO_y excess above background for observations (black line with transect number in open boxes) and models. Line designations for the models are the same as in Figure 7. (bottom) Excess NO_y flux above background.

transect, or any other transect downwind of either Houston or Dallas. This singular, anomalous observation point is best explained by incomplete vertical mixing of fresh emissions in the heart of Dallas under light wind conditions. The various models, on the other hand, often show high excess fluxes and concentrations directly over urban areas, similar to the behavior of WRF/Chem on this particular day.

5. Comparison of Emissions and Emission Ratios Derived From Upwind/Downwind Transects

[31] Given the fact that direct emissions of all individual sources over large urban areas are impossible to quantify, the logical next step is to estimate them by examining

integrated fluxes going into and out of a predefined urban region. Here, the WP-3 aircraft data are used to determine fluxes in a manner nearly identical to the mass balance approach used by Ryerson *et al.* [1998] to determine NO_x and SO₂ fluxes from power generation facilities. Integrated fluxes for isolated sources determined in this manner are found to be quite consistent with observed emissions (i.e., CEMS measurements), and the assumption that this consistency holds for large source regions is implicit in the fluxes derived here. Extending equation (4) in the vertical, the total excess PBL flux perpendicular to a vertical plane defined by the aircraft heading for a particular transect downwind of an urban area is

$$\text{Excess PBL Flux} = \int_0^{Z_{bl}} \int \rho(\chi - \chi_B) (\vec{v} \cdot \vec{n}) H(\chi, \chi_B) ds dz,$$

where Z_{bl} is the boundary layer height. Under the regular assumptions of the mass balance approach ($\rho \chi (\vec{v} \cdot \vec{n})$) at transect height equals the average over the PBL depth) the excess PBL flux is just the excess flux of equation (4) times the PBL depth,

$$\text{Excess PBL Flux} \approx Z_{bl} \cdot \int \rho(\chi - \chi_B) (\vec{v} \cdot \vec{n}) H(\chi, \chi_B) ds. \quad (6)$$

Assuming the vertical transport across the PBL is negligible, the excess PBL flux of an inert species will equal its emission rate over the urban area. The urban areas used here are large enough that a downwind transect could intercept plumes emitted 2 to 4 h previously. For emitted species with photochemical loss or deposition the excess PBL fluxes can only approximate lower limits to actual emissions. Section 5.1 presents some excess PBL flux calculations derived from the observations, and comparisons with the emission inventories used by the WRF/Chem model. Flux determinations from the models are also possible, but Figure 7c shows that some models are complicated by the mismatch in PBL height and species mixing depth. Section 5.2 analyzes observed and modeled excess concentration ratios, which circumvents many of the uncertainties in the absolute flux determinations, providing more useful and direct comparisons with emission inventory estimates.

5.1. Estimates of Observed Excess PBL Fluxes Using the Mass Balance Approach

[32] For each upwind/downwind flight pattern over Dallas and Houston during September of 2006 a single transect between 300 and 670 m is assigned to represent the maximum urban impact from either the Houston or Dallas urban corridors. Eight transects for Houston and two transects for Dallas are identified (Table 2), all occurring between 1630 and 2100 UTC. Considering wind speeds and directions, emissions from the maximum impact transects likely occurred between 1430 and 2100 UTC, which is the period just after the morning traffic peak until 1600 local time. For all flight patterns except for 25 September over Dallas, this transect is taken to be the one with the highest excess NO_y flux value between 10 and 50 km downwind of

Table 2. Downwind/Upwind Flight Pattern Days, Upwind Transect Number and Center Time, Maximum Urban Impact Transect Number and Center Time, Assigned PBL Height, Mean Flow Wind Speed, and Direction for the 10 Cases Between 13 and 29 September 2006^a

Date	Urban Area	Upwind Transect, Center Time	Urban Transect, Center Time	PBL Height (km)	Wind Speed (m/s) and Direction
13 September	Dallas	7, 1951:39	9, 2101:51	1.2 ± 0.20	5.0, northerly
15 September	Houston	2, 1641:40	7, 1914:24	1.0 ± 0.20	4.5, southeasterly
19 September	Houston	2, 1629:11	4, 1657:47	1.0 ± 0.15	7.5, northeasterly
20 September	Houston	3, 1652:28	6, 1809:21	1.2 ± 0.40	5.0, easterly
21 September	Houston	3, 1753:45	5, 1843:48	1.5 ± 0.50	9.5, southerly
25 September	Dallas	3, 1704:18	5, 1820:39	1.4 ± 0.20	3.5, northerly
25 September	Houston	11, 2032:04	13, 2123:01	1.1 ± 0.20	7.0, northerly
26 September	Houston	2, 1635:06	4, 1808:50	0.9 ± 0.15	3.5, northeasterly
27 September	Houston	3, 1946:41	5, 2037:43	1.2 ± 0.15	5.5, southerly
29 September	Houston	3, 2054:18	5, 2158:51	1.0 ± 0.15	7.0, southeasterly

^aAll times are given as UTC.

the city center. As explained in section 4.4, the 25 September 2006 transect 10 km downwind of Dallas is anomalously high owing to light winds and incomplete vertical mixing, so the next transect 45 km downwind is used. PBL heights for each of the maximum urban impact transects are also assigned, and are based on discontinuities in the RH profiles from the aircraft (see, e.g., Figure 7c), if those profiles are within an hour of the transect, otherwise PBL heights are determined from wind profiler data at Cleburne (for Dallas on 13 September) or Arcola (for Houston on 20 and 21 September) that are available at <http://www.etl.noaa.gov/programs/2006/texaqs/verification/>. Uncertainties in PBL height are set to a minimum of 150 m but also

include variability in nearby WP-3 vertical profiles or wind profiler determinations within 1 h of the maximum urban impact transect. Table 2 gives the relevant information regarding the assigned background transect, maximum impact transect used in the PBL excess flux calculations, PBL heights and uncertainties, and observed mean flow conditions.

5.1.1. Primary Emission Estimates for Six Species by Flight

[33] With PBL heights (and uncertainties) from Table 2 and the background mixing ratios (and uncertainties) defined above, Figure 9 shows the calculation of equation (6) for the 10 maximum urban impact transects available during

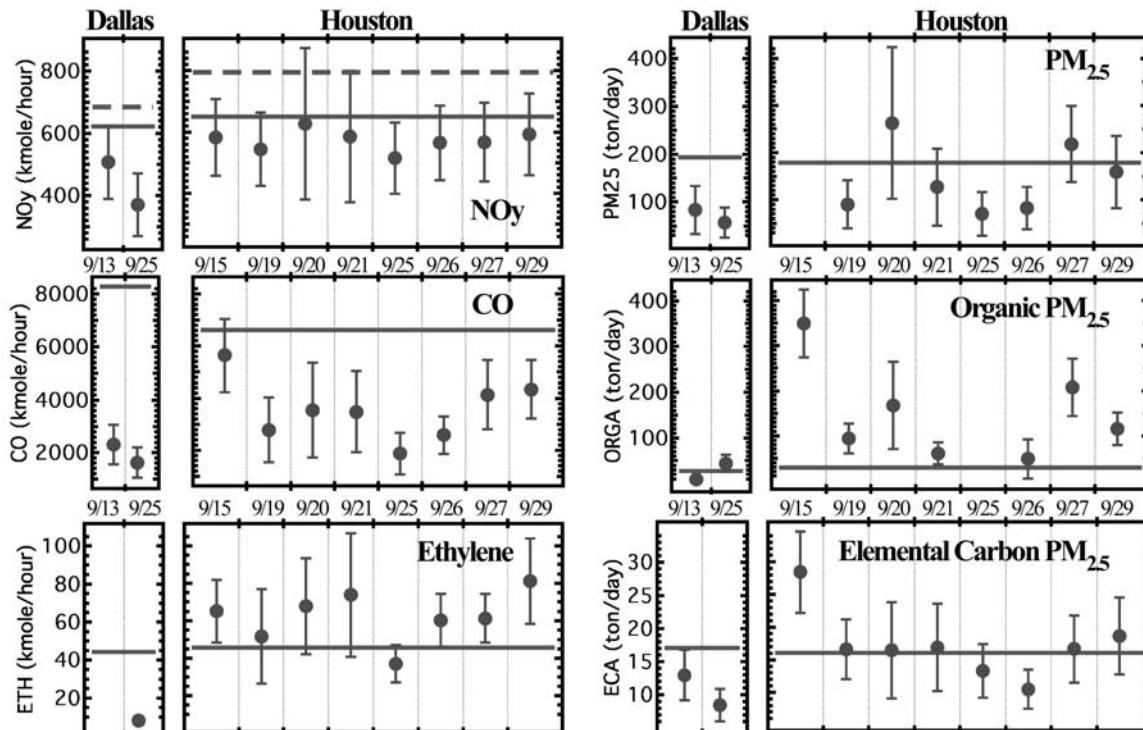


Figure 9. Excess PBL flux for the maximum urban impact transects (Table 2) with 1σ error estimates due to PBL and background uncertainties. Dashed horizontal bars are noontime emissions used by the WRF/Chem models for the Dallas (32.4°N to 33.3°N latitude, 97.6°W to 96.4°W longitude) and Houston (29.4°N to 30.1°N latitude, 95.8°W to 94.9°W longitude) urban corridors. Solid horizontal line is for the same inventory with point source emissions taken from July 2004 CEMS observations.

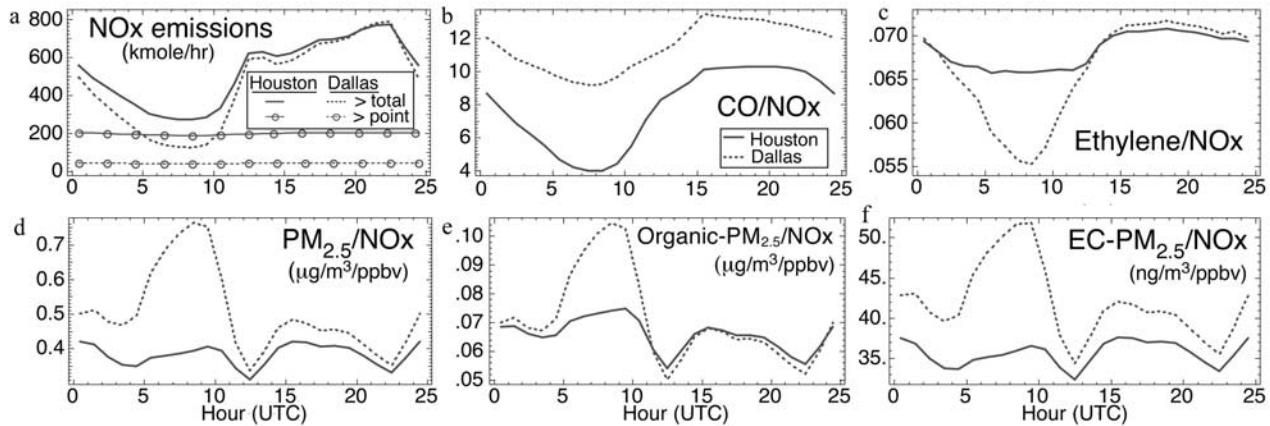


Figure 10. (a) Diurnal variation of total and point NO_x emissions and total emission ratios of (b) CO, (c) ethylene, (d) PM2.5, (e) organic PM2.5, and (f) elemental carbon PM2.5 relative to NO_x for WRF/Chem emissions inventory and for the Dallas and Houston urban corridors used in Figure 9.

September of 2006, and for six species that have primary emissions within the models. For comparison purposes the emissions inventory used by the WRF/Chem models are summed over encompassing areas (29.4°N to 30.1°N latitude, 95.8°W to 94.9°W longitude for Houston, and 32.4°N to 33.3°N latitude, 97.6°W to 96.4°W longitude for Dallas) and noontime emissions from this inventory are included in Figure 9. The Dallas area comparisons are limited to two flights, but overestimates of WRF/Chem emissions for all species except OA are implied. The excess fluxes for Houston on 25 September 2006 appear lower than on other days for most species, but results imply that the emissions of all species except organic PM_{2.5}, based on the WRF/Chem inventory, are overestimated on this day. Some qualitative statements regarding the emissions inventory for the Houston region can also be made: CO emission are too high relative to NO_y, ethylene emissions are too low for Houston, emissions of organic aerosol do not account for the observations, and EC and total PM_{2.5} emissions appear consistent with observations within a large scatter (factor of 1.8).

[34] One day, 15 September 2006, sticks out as being exceptionally high for CO, OA and EC (as well as aerosol NH₄ and O₃, not shown). The time series of the maximum urban impact transect for this day shows broad elevated levels of these species at the eastern edge merging with the Houston plume, but with relatively reduced enhancements of NO_y, SO₂, and anthropogenic VOC. These differences suggest the additional influence of an aged continental air mass from the east or southeast influencing the northeastern most reaches of this day's flight with a possible biogenic burning signature. The same signatures of high CO, OA, and EC are also apparent in air masses originating from western Louisiana on 21 September 2006 for three transects more than 100 km north of Houston. While biomass burning sources originating east of Texas are not expected to influence the near-source analysis presented here, they may influence the broader region, contribute to the model bias statistics with aircraft data (OA in Figure 6) and possibly the surface AIRNow PM_{2.5} bias statistics as well.

[35] Uncertainties in the temporal allocation of the emissions inventory also limit comparisons of absolute fluxes. Figure 10 shows a ~30% difference relative to noontime values for NO_x emissions over the 1430 to 2100 UTC period, which is when the aircraft did most of its urban sampling. On the other hand, model emission ratios show much smaller relative differences during this time period, and argues for a higher certainty and usefulness for comparisons of emission ratios as opposed to comparisons of absolute rates.

5.1.2. Comparison of Observed Excess NO_y PBL Fluxes With Emission Inventories

[36] Observed NO_y excess fluxes for Houston (Figure 9) are fairly consistent for all days, and this consistency is implicit in further analysis below, where emission and concentration difference ratios relative to NO_y are used in comparisons. Thus a quantitative evaluation and understanding of NO_y inventory emissions is needed as a basis for the evaluation of other species. For purposes of comparison an observed median value of 575 kmol/h ($\pm 20\%$) for the Houston corridor is assigned from Figure 9. For Dallas/Fort Worth a similar estimate of 440 kmol/h can be assigned, but with an associated high uncertainty ($\pm 50\%$). The base noontime NEI-99 emissions inventory is ~40% too high for Houston, and 57% too high for Dallas/Fort Worth. To understand the sources of these discrepancies it is useful to consider the apportionment of NO_y emissions into the source categories typically used within the inventories. Table 3 summarizes the four subcategories (on-road, off-road, area and point) of emissions within the NEI-99 inventory for Dallas and Houston and for several gas-phase and aerosol species. The NO_y and SO₂ point emissions within Table 3 incorporate updates with the 2004 CEMs observed emissions, shown as the solid line for NO_y in Figure 9. The difference in noontime NO_y emissions from the CEMS update is ~150 kmol/h for Houston, which reduces the original 40% discrepancy with observations to 13%. This NO_y decrease reduced the point source fraction of total emissions from ~44% to 31%, owing to the significant reductions in NO_x emissions from EGUs between 1999 and 2004 in eastern Texas instituted within

Table 3. The 1100–1200 CDT Ozone Season Day Emission Sums for Houston and Dallas From the NEI-99 Version 3 Inventory Updated With July 2004 CEMS Point Source Emissions of NO_x and SO₂^a

	Dallas					Houston				
	Total	On-Road	Nonroad	Area	Point	Total	On-Road	Nonroad	Area	Point
NO _x	621.4	426.8	134.9	18.8	40.9	649.9	313.5	121.4	15.4	199.6
CO	8285.	4919.	3209.	83.2	74.1	6610.	3617.	2630.	117.4	245.1
SO ₂	49.2	12.7	12.2	0.5	23.7	215.9	8.2	22.1	0.2	185.4
NH ₃	115.6	44.6	0.6	67.6	2.9	76.1	33.4	0.5	38.1	4.0
Ethylene	44.2	29.8	12.7	0.9	0.9	45.8	22.0	9.2	1.3	13.2
Acetylene	10.1	7.4	2.1	0.2	0.4	9.0	5.5	1.4	0.2	1.9
Formaldehyde	12.9	5.7	4.1	0.2	2.9	13.6	4.2	2.9	0.1	6.5
Acetaldehyde	2.4	1.3	0.9	0.0	0.1	2.2	1.0	0.6	0.0	0.6
Toluene	27.1	10.4	2.6	11.1	3.0	18.3	7.7	2.1	6.8	1.6
Xylenes	12.3	5.3	1.8	3.7	1.4	8.7	3.9	1.5	2.6	0.7
Total PM _{2.5}	193.6	11.0	16.4	159.	7.3	178.4	8.1	19.4	125.1	25.8
Primary PM _{2.5}	145.7	1.0	0.4	141.	3.6	126.9	0.8	0.7	112.4	13.0
Sulfate PM _{2.5}	3.2	0.2	0.5	1.3	1.3	6.3	0.1	1.3	1.1	3.8
Nitrate PM _{2.5}	0.3	0.0	0.0	0.2	0.0	0.3	0.0	0.0	0.2	0.1
Organic PM _{2.5}	27.3	4.0	5.4	15.7	2.2	28.8	2.9	5.5	11.8	8.6
Elemental C PM _{2.5}	17.1	5.7	10.0	1.2	0.1	16.1	4.0	10.7	1.0	0.3

^aThe summation limits for Houston are 29.4°N to 30.1°N latitude, 95.8°W to 94.9°W longitude and 32.4°N to 33.3°N latitude, 97.6°W to 96.4°W longitude for Dallas. Units are kmol/h for gas-phase species and short-ton/d for PM_{2.5} species. Primary PM_{2.5} signifies the unspeciated PM_{2.5} component.

state implementation and federal regulatory programs [e.g., Frost *et al.*, 2006; Kim *et al.*, 2006]. For Dallas/Fort Worth the original point source fraction was much smaller (~16%) so reduced NO_x point emissions between 1999 and 2004 has less of an impact on the emission discrepancy for this region with the modified inventory still high by ~46%.

[37] NO_y emission changes from 1999 to 2006 in the on-road, nonroad, and area sources categories within Table 3 probably occurred as well. Bishop and Stedman [2008] report on-road NO_y emissions decreases between 48% and 68% for Denver, Phoenix, West Los Angeles and Chicago between 1999 and 2007, owing largely to improvements in the emission control systems of newer vehicles. Though analogous trend data are not available for Houston or Dallas, reductions in on-road NO_y emissions from 1999 to 2007 by TCEQ for Dallas/Fort Worth (http://www.tceq.state.tx.us/assets/public/implementation/air/am/docs/dfw/p1/DFW_2007_Modeling_Report_20040825.pdf) were also estimated at 50%, also owing to emission control systems on newer vehicles, fleet turnover rates, and projected vehicle miles traveled. Table 3 shows that on-road emissions constitute 68% of the total NO_y within the NEI inventory for Dallas/Fort Worth and 48% of the total for Houston after CEMS correction. Thus, if one assumes a 40% reduction in on-road emissions at both locations, the modified NEI-99 inventory would be 3.0% low compared to the derived emissions for Houston, and 19% high for Dallas/Fort Worth. Given the uncertainties inherent to the derived emissions, and in the inventories themselves, this can be considered very good agreement. Only the WRF/Chem model used the NEI-99 emissions. All other models used on-road emissions based on the NEI-2001 inventory. It is therefore probable that those models also overestimated NO_y emissions by relying on an inventory 5 years older than date used in the analysis.

5.2. Comparisons of Average Excess Above Background Ratios

[38] Though Figure 9 shows that while the mass balance approach to determining emissions may be qualitatively useful for inventory comparisons, uncertainties in PBL

height, temporal allocation of emissions, the inertness of the species, and other mass balance assumptions limit the usefulness of these absolute comparisons. Moreover, only the emissions inventory used by WRF/Chem was made available for analysis, preventing model-to-model emission inventory comparisons or assessments. The best alternative is to use relative concentration differences of the models and observations as surrogates for relative emission rates. Figure 10 shows the diurnal variation of CO, ethylene, and PM_{2.5} ratios relative to NO_x within the WRF/Chem inventory. For the 1430 to 2100 UTC period the temporal variability in emission ratios (~10% in the worst case, CO/NO_x) is reduced considerably relative to absolute NO_x. Rather than relying on mass balance assumptions of a well-mixed and contained PBL, the implicit assumption made here is simply that concentration ratio differences are independent of transport processes. The dependence of concentration differences on species reactivity or deposition cannot be eliminated but can be minimized by considering the freshest urban transects.

[39] The public Web site provides concentration difference ratio plots on a flight-by-flight basis for 18 species relative to NO_y and CO. Figure 11 shows examples of CO, ethylene and PM_{2.5} difference ratios relative to NO_y for the 25 September 2006 transects downwind of Dallas shown in Figures 6 and 7. Because of uncertainties in the backgrounds, uncertainties in the ratio determinations for the upwind legs and transects >60 km downwind limit the model/observation comparisons. However, these comparisons illustrate trends for 25 September 2006 that are apparent in other flights and in the summary statistics discussed below: (1) the CO/NO_y ratios for all models are a factor of 2 to 3 higher than the observations, (2) for Dallas the C₂H₄/NO_y ratios are close to observations, but the Canadian models exhibit high ethylene ratios due to the biogenic C₂H₄ component within the AURAMS and CHRONOS models, and (3) organic PM_{2.5}/NO_y ratios are significantly underpredicted, particularly the farther downwind the transect is from Dallas. It should also be noted that combining the NO_y emissions used by WRF/Chem for the adopted Dallas region (688.9 kmol/h) along

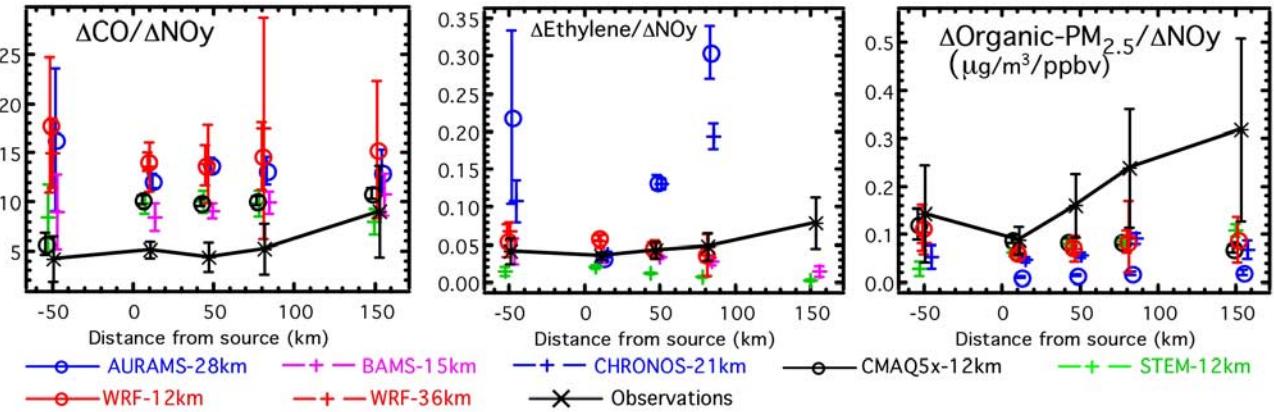


Figure 11. (left) CO, (middle) ethylene, and (right) organic PM_{2.5} concentration difference ratios relative to NO_y for the 25 September 2006 transects upwind and downwind of Dallas. Model labels are the same as in Figure 7. Only model and observed ratios with less than 80% background uncertainty are shown.

with the area totals in Table 3 yields emission ratios of 12.0 for CO/NO_y, 0.064 for ethylene/NO_y, and 0.060 $\mu\text{g}/\text{m}^3/\text{ppbv}$ for organic PM_{2.5}/NO_y. These ratios correspond closely to the WRF/Chem ratios for the two transects closest to Dallas city center in Figure 11.

[40] Summary difference ratio statistics for transects that are close to the city centers adopted for Houston and Dallas are also calculated and available at the comparison Web site. Difference ratios for 21 species relative to NO_y, CO, SO₂, PM_{2.5}, and EC are provided, but here we focus only on

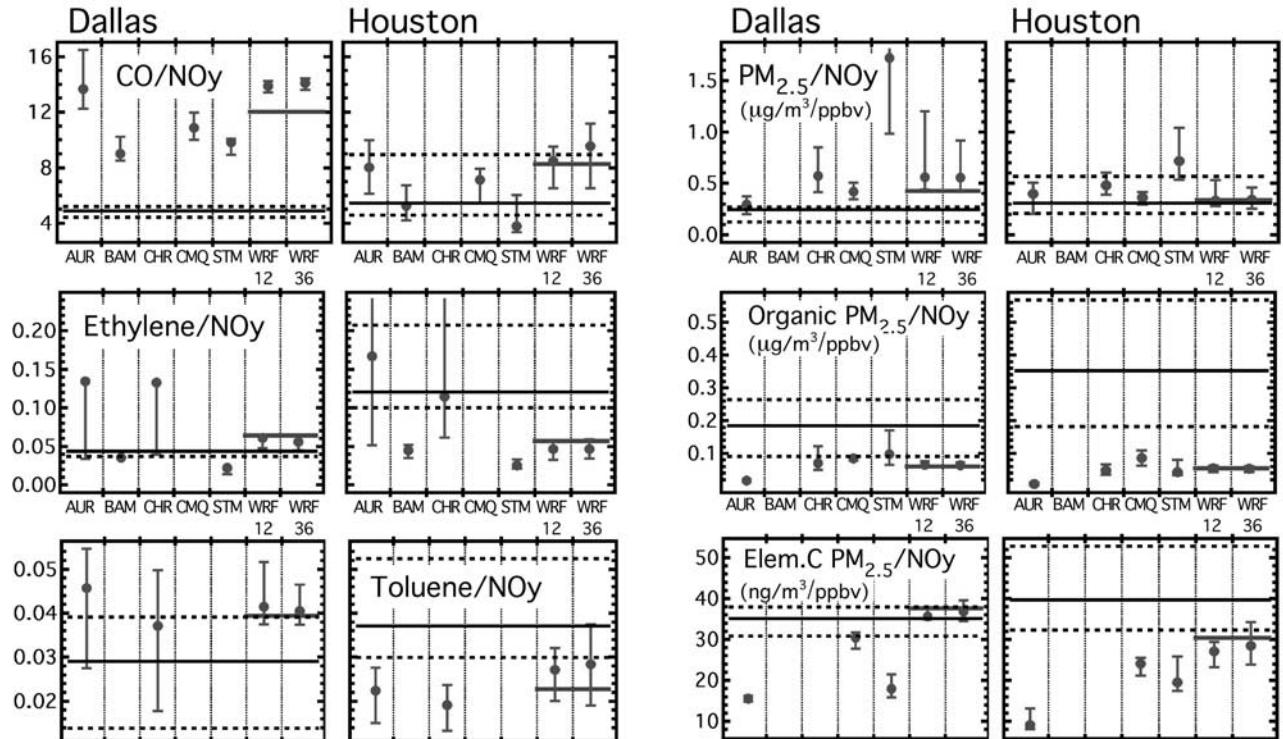


Figure 12. CO, ethylene, toluene, PM_{2.5}, organic PM_{2.5}, and elemental carbon PM_{2.5} concentration difference ratios relative to NO_y for the Dallas and Houston urban centers and for the seven forecast models (circles) and observations. Observed medians are solid lines across entire graph, and dashed lines are uncertainty limits for observations. For Dallas, uncertainty limits are maximum and minimum values of four possible transects. For Houston, uncertainty limits are 1/6 and 5/6 quantiles of sorted distributions for 29 possible transects. Solid lines overlaid on the WRF/Chem results only are noontime ratios from the emissions inventory used in those models. For PM_{2.5} aerosol species the inventory emission ratios ([ton/d]/[kmol/h]) have been multiplied by 1.515 to convert to ($\mu\text{g}/\text{m}^3$)/(ppbv) (assuming 1000 mbar and 30°C).

ratios of a few key species relative to NO_y. Figure 12 shows the concentration difference ratios for the seven models compared to observations, and for only those transects determined to be between 10 km upwind and 50 km downwind of the city centers. Noontime emission ratios from the inventory used by WRF/Chem are also shown. For Dallas there is a maximum of four values (minimum and maximums are shown in the limits), while a maximum of 29 values are available for Houston (1/6 and 5/6 quantiles of sorted distribution shown as limits).

5.2.1. Average Excess Above Background Ratios Versus Emission Ratios for the WRF/Chem Model

[41] For nearly all species the emission inventory ratios from the WRF/Chem model are reproduced quite well by the upwind/downwind difference ratios calculated from the concentrations, giving confidence that the difference ratio calculations for the other models and observations are also representative of their underlying emission inventory ratios. There are a couple WRF/Chem species that show explainable differences with the inventory ratios. CO/NO_y concentration difference ratios are higher than inventory ratios for both Dallas and Houston owing to photochemical CO formation from VOC oxidation (~16% of the emission ratio for Dallas, much less for Houston). Likewise for PM_{2.5} over Dallas, the WRF/Chem models show some nitrate and sulfate formation contributing to PM_{2.5}, making the concentration difference ratios larger than the inventory. Though Figure 12 shows true toluene/NO_y emission ratios within the inventory, the toluene species calculated within WRF/Chem is a lumped species that includes other aromatics. The inventory ratios for this lumped species should be increased by 24 and 40% for Dallas and Houston respectively. These adjustments put the lumped emission ratios ~20% and 10% above the WRF/Chem concentration difference ratios for Dallas and Houston, respectively, indicating a small amount of photochemical loss between emission and sampling, and similar to C₂H₄. The organic PM_{2.5}/NO_y and EC/NO_y ratios are unaffected by photochemistry within WRF/Chem, and concentration difference ratios match emission ratios for these species typically within 15%. The difference ratios for the species shown in Figure 12 relative to CO, PM_{2.5} and SO₂ and EC (see Web site) show comparable agreement with inventory ratios for NO_y. Considering all available species, and the permutations of possible excess ratios, a 25% uncertainty estimate is assigned to the concentration difference ratios when using them as surrogates for emission ratios within the near-source analysis.

5.2.2. CO/NO_y Emission Ratios

[42] The CO/NO_y concentration difference ratios in Figure 12 show a clear disparity between all models and observations for Dallas, and vehicle emissions are the dominant source of these species within the inventories. The WRF/Chem and AURAMS ratios are relatively higher than the other models owing to their older inventories and on-road speciation profiles having higher CO/NO_y emission ratios in vehicles produced before 2001 [Parrish, 2006]. But the other models with newer on-road speciation profiles overestimate the CO/NO_y ratios by nearly a factor of 2. This factor of 2 to 3 is entirely consistent with the results of G. J. Frost et al. (Observational evaluation of mobile source emissions, manuscript in preparation, 2009) who indepen-

dently derived an identical estimate for on-road CO emission overestimation using tunnel measurements and CO/CO₂ analysis. For Houston the models tend to do much better owing to the higher fraction of NO_y from point sources. The WRF/Chem inventory requires ~30% reduction in CO/NO_y to match observed ratios in Houston, but since NO_y emissions should be reduced by ~40% from the original NEI-99 inventory, a 70% drop in Houston CO emissions is needed to be consistent with observations.

[43] Table 3 shows that on-road emission ratios of CO/NO_x for both Houston and Dallas/Fort Worth are a factor of 2 higher than the observed ratio of ~5, and nonroad ratios, accounting for a significant fraction (~40%) of the total emissions for both regions, are a factor of 4 higher. While the on-road fraction clearly needs to be reduced within the inventories, either the CO/NO_x emission ratio within the nonroad emissions is way too high, or the relative activity between on-road and nonroad emissions needs to be adjusted to reconcile both CO and NO_x emissions, particularly in the Dallas region.

5.2.3. C₂H₄/NO_y and Toluene/NO_y Emission Ratios

[44] Ethylene/NO_y ratios for the AURAMS and CHRONOS models in Figure 12 are complicated by the biogenic contribution to ethylene, although the lower limits within Figure 12 for these models are representative of concentrated urban plumes and consistent with the other models (see Figure 11). Ethylene, toluene and xylene (not shown, see Web page) ratios to NO_y show a very similar pattern with all models slightly overpredicting Dallas ratios, significantly underpredicting Houston ratios, and failing to reproduce the higher ratios for Houston compared to Dallas. The ethylene/CO ratios (not shown, see Web site) are quite similar for both Houston and Dallas (for the available non-Canadian models) with concentration difference ratios very similar to the emission inventory data reported for Boston/New York and Los Angeles by Warneke et al. [2007], and a factor of 2 to 3.5 lower than observed ratios. Underestimated alkene emissions within the inventories, primarily from petrochemical industry sources, have been documented for aircraft observations from the TexAQS 2000 study [Ryerson et al., 2003; Wert et al., 2003], yet updated emission profiles for those sources have not been included within the model inventories considered here. On the basis of surface observations at the LaPorte site near Houston, Karl et al. [2003] also showed that aromatic compounds are severely underpredicted within the emission inventories. Thus the low model ethylene and toluene ratios shown for Houston in Figure 12 are no surprise. Toluene ratios for Dallas are on the order of 30 to 50% too high for the AQFMs, which implies toluene emission overpredictions on the order of 70 to 90%, assuming a 40% CEMS modified NEI-99 NO_y overestimate for Dallas/Fort Worth. This is somewhat lower than the factor of 2 to 3 overestimate from Warneke et al. [2007] based on toluene/CO ratios in the two U.S. urban corridors they examined, but the observed Δtoluene/ΔCO ratio of 0.006 for both Dallas and Houston (see Web page) are a factor of 2 higher than the ratios observed in that study.

[45] Toluene/NO_y and Toluene/VOC emission ratios during 2006 are probably less than what is assumed in NEI-99 inventory. This is because the NEI-99 inventory assumed nonoxygenated fuel use for all gasoline powered vehicles, while reformulated gas (RFG) with reduced aromatic con-

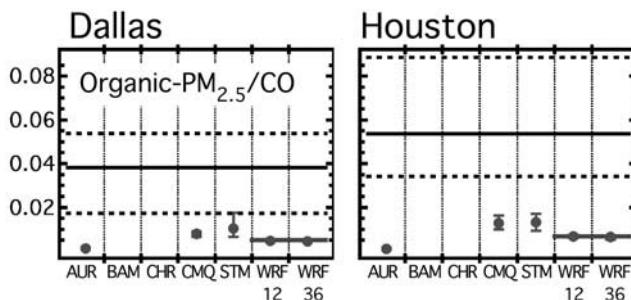


Figure 13. Organic PM_{2.5} concentration difference ratio relative to CO for the Dallas and Houston urban centers and for the seven forecast models and observations. Meanings of lines and symbols are the same as in Figure 12.

tent were used in both Houston and Dallas during the summer of 2006. For evaporative fuel losses the toluene/VOC ratio decreases dramatically, from 4.1% to 1.9% with the substitution of a 10% ethanol blend RFG. For gas vehicle exhaust this ratio is reduced from 10.4% to 9.5%. Annually averaged, exhaust VOC is between 58% and 88% larger than evaporative VOC emissions within NEI-99, depending on road and population designation within the inventory. Assuming a fixed VOC/NO_y ratio, the toluene/NO_y ratio of on-road and nonroad sources would then be reduced by 25% to 28%, and total emissions would be reduced by roughly half this amount (Table 3). Inventory toluene/NO_y ratios are then close to observations for Dallas (~20% high), but make the disparity for Houston even worse.

[46] Although Table 3 shows Houston having a larger ethylene fraction from point emissions compared to Dallas, the underestimate of ethylene from the petrochemical and refining operations in southeast Texas still needs to be modified to fit observations. Several flights show observed ethylene spikes downwind of particular sources (e.g., Sweeny, 21 September, transect 3), downwind of the Houston Ship Channel region (e.g., 27 September 2006, transect 12), and downwind of refining and petrochemical regions outside of Houston (e.g., Texas City, 15 September 2006, transect 3 and Beaumont, 20 September 2006, transect 2) that are not apparent in any of the models. Such detailed and systematic estimates have been made with in situ aircraft data collected during the 2000 field study [Xie and Berkowitz, 2007]. J. Mellqvist et al. (Measurements of industrial emissions of alkenes in Texas using the Solar Occultation Flux method, submitted to *Journal of Geophysical Research*, 2009) compare ethylene and propylene emission fluxes using a solar occultation flux (SOF) instrument with inventory data for individual facilities and petrochemical complexes during the TexAQS II study. Model simulations that incorporate this information within their inventories are obviously needed to assess the impact such top-down emissions information may have on forecast O₃.

5.2.4. PM_{2.5}/NO_y, Organic PM_{2.5}/NO_y, and EC/NO_y Emission Ratios

[47] PM_{2.5}/NO_y ratios for the STEM model in Figure 12 are affected by biases related to NH₄/NH₃ partitioning errors mentioned previously. The CHRONOS and WRF/Chem PM_{2.5}/NO_y ratios are a factor of 2 higher than observed for Dallas, but for Houston the models are reasonably

consistent with each other. Except for STEM, model ratios and emission ratios are consistent (within 30%) within the observations. Despite the agreement for the total PM_{2.5}/NO_y, organic PM_{2.5}/NO_y ratios illustrate the glaring inconsistency between model and observed PM_{2.5} composition. For both Dallas and Houston the observations show the OA increases to be the dominant compositional fraction contributing to PM_{2.5} increases, whereas all models show a minimal contribution to total PM_{2.5} increases. Moreover, none of the models capture the higher PM_{2.5}/NO_y for Houston relative to Dallas, and if the low emission ratios for WRF/Chem are representative of true emission ratios, a strong photochemical source of organic PM_{2.5} missing from the models is required to explain the observations. Possible sources of missing OA in global- and regional-scale models have been the subject of recent inquiry [Heald et al., 2005; Volkamer et al., 2006] with no clear consensus on biogenic versus anthropogenic origin of the missing sources [Weber et al., 2007; de Gouw et al., 2008].

5.2.5. Organic PM_{2.5}/CO Emission Ratios

[48] The high correlation between OA and CO has been used to argue for an anthropogenic component to OA formation over the northeast United States [Sullivan et al., 2006; de Gouw et al., 2008], and over Atlanta [Weber et al., 2007]. The U.S. studies imply $\Delta\text{OA}/\Delta\text{CO}$ of $\sim 0.01 \mu\text{g/m}^3/\text{ppbv}$ from primary emissions [de Gouw et al., 2008] consistent with emission inventories [Bond et al., 2004], reaching peak values of $\sim 0.04 \mu\text{g/m}^3/\text{ppbv}$ on the time scale of 12 to 24 h. Figure 13 shows $\Delta\text{Organic PM}_{2.5}/\Delta\text{CO}$ excess ratios for Dallas and Houston, and for those transects between 10 km upwind and 50 km downwind of the assigned city centers. The model ratios scatter about the $0.01 \mu\text{g/m}^3/\text{ppbv}$ value, similar to the emission ratios used within WRF/Chem. The observations are close to or above the $0.04 \mu\text{g/m}^3/\text{ppbv}$ peak values of the U.S. urban data [de Gouw et al., 2008], but in sharp contrast, these levels are reached within 1 to 3 h of Dallas and Houston centers. This relatively fast OA formation rate is consistent with the observations of rapid PM_{2.5} formation downwind of the Houston ship channel in 2000 [Brock et al., 2003]. Rapid OA formation was also observed less than 50 km downwind of Mexico City by Volkamer et al. [2006], who used SOA's close correlation with glyoxal to argue that primary VOC oxidation steps must be responsible for much of the anthropogenic SOA. Though a large fraction of the model/observed discrepancy in Figure 13 is due to overestimates of CO emissions, one can infer (e.g., Figure 12) at least a factor of 2 underestimate (for CMAQ/NAM, CHRONOS and STEM) in SOA formation over Dallas, and larger discrepancies for Houston and other models.

5.2.6. O₃/NO_y and NO_x/NO_y Concentration Difference Ratios

[49] The O₃/NO_y and NO_x/NO_y ratios for Dallas and Houston are shown in Figure 14, and are useful for diagnosing discrepancies between model forecast and AIR-Now network surface observations. None of the models simulate the observed ratio of ~ 3.1 for Houston O₃/NO_y, while most of the models show better agreement with the ratio of ~ 2.0 for Dallas. The inferred increase in O₃ production efficiency for Houston relative Dallas (and other U.S. cities) is a well-known feature [e.g., Kleinman et al., 2002; Ryerson et al., 2003; Wert et al., 2003]. The low

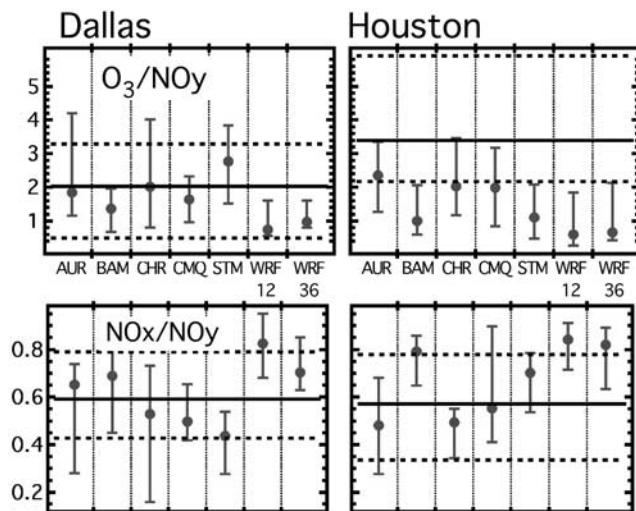


Figure 14. O_3 and CO concentration difference ratios relative to NO_y for the Dallas and Houston urban centers and for the seven forecast models and observations. Meanings of lines and symbols are the same as in Figure 12.

model O_3/NO_y ratios for Houston are consistent with low biases in ethylene/ NO_y and toluene/ NO_y noted above, suggesting that all models are unable to reproduce Houston O_3 formation because of missing sources of reactive VOC. Ratios of the VOC oxidation products CH_2O/NO_y and CH_3CHO/NO_y (see Web site) are also consistently low for Houston, in keeping with the low VOC emission hypothesis. Improvements to the persistent low model biases in O_3 and CH_2O over Houston for the TexAQS 2000 experiment have been noted in 3-D model studies that modify point source emissions of reactive alkenes to match inferred levels derived from aircraft and surface observations [Jiang and Fast, 2004; Byun et al., 2007]. Relative model biases in O_3/NO_y ratios are inversely related to biases in NO_x/NO_y . Assuming the photochemical mechanisms are reliable, reasons for low O_3/NO_y and high NO_x/NO_y for a given model could be due to either NO_x levels being too high and titrating available O_3 , or insufficient oxidation of NO_x (and associated O_3 formation) through reactive VOC emission underestimations. The WRF/Chem model ratios are biased low for both Dallas and Houston, and though our absolute flux estimate (Figure 9) implies Dallas NO_x emissions may be too high by ~50% in the original NEI-99 inventory, the reduced PBL heights noted earlier for WRF/Chem could also be contributing to high NO_x levels and excessive O_3 titration. This may also be a contributing factor for the BAMS model which often exhibits high NO_x and NO_y levels (e.g., Figure 7d). The high O_3/NO_y ratio for the STEM model over Dallas can be explained by 50% higher than observed upwind O_3 on both 13 and 25 September 2006 (see Web site), leading to higher peroxy-radical production, efficient NO_x oxidation and local O_3 formation relative to observations and other models.

6. Conclusions

[50] In this study the gas-phase and aerosol composition from seven real-time air quality forecast models (AQFMs)

are evaluated against observations from the AIRNow surface network and NOAA WP-3 aircraft data collected over eastern Texas during the TexAQS II field study. Forecast performance statistics for surface O_3 and $PM_{2.5}$ are presented for each model as well as the model ensemble, and these statistics are compared to previous real-time forecast evaluations during the ICARTT/NEAQS 2004 field study in New England. The flight plans of the NOAA WP-3 during TexAQS II allows for many opportunities to compare model predicted and observed composition upwind and downwind of Houston and Dallas in the 300 to 600 m height range during midday. The measurements are used to estimate daytime emissions for several key species integrated over broad (~100 km by 100 km) corridors of the two cities, and these emissions are compared to the emissions inventory used by the WRF/Chem modeling group for their real-time forecasts (U.S. EPA NEI-99). Emission inventories for the other forecast models are not available, so a technique is developed to derive integral average relative emission ratios from constituent mixing ratio or concentration forecast fields alone. For both the models and observations a background or upwind mixing value for each flight is defined, and an average value above background for each downwind transect is defined. Ratios of the excess mixing ratios are shown to accurately represent ratios within the emissions inventory for the WRF/Chem model to within 25%, allowing quantitative comparisons with similar ratios derived from observations. The WRF/Chem emission ratios, particularly with respect to NO_y , show characteristic biases similar to the results of other models, allowing a cursory assessment to be made for relative emissions used in present-day AQFMs in general.

[51] Surface maximum 8-h daily average O_3 forecasts for eastern Texas during the summer of 2006 show a marked improvement in correlations, bias and RMSE-based skill scores for all models compared to similar forecasts for New England during the summer of 2004. Though some of this may be due to the smaller region, and more spatially uniform meteorological forcing during the 2006 study, improvements in all the AQFM formulations and emissions have also occurred since 2004. As found in the 2004 study, the ensemble mean of the model forecasts outperforms any single model, and alternative ensemble strategies and bias correction techniques more applicable to an operational ensemble framework would be expected to improve O_3 forecasts further [Pagowski et al., 2006; Wilczak et al., 2006; Delle Monache et al., 2008]. For threshold statistics only the two Canadian AQFMs were able to forecast the 85-ppbv O_3 exceedances better than persistence. All other AQFMs as well as the ensemble mean showed far less skill at threshold exceedance predictions compared to New England in 2004. Considering the ensemble mean as the best, most representative realization of the model suite, the number of 85-ppbv O_3 exceedances for this forecast was severely underestimated for the Houston region, and much less so for the Dallas/Fort Worth region. This preferential underprediction for Houston is consistent with the low O_3 formation per NO_y emitted calculations (Figure 14) for all AQFMs. The low AQFM ethylene emission biases for Houston (Figure 12), and previous studies showing the importance of light alkenes to O_3 formation within Eulerian forecast models of Houston [Jiang and Fast, 2004; Byun et

al., 2007] strongly link an underestimation of light alkene emissions within the inventories to inabilities in predicting Houston O₃ exceedances.

[52] Statistical evaluations of the PM_{2.5} forecasts, based on 24-h averages, are much less reliable during TexAQS II compared to ICARTT/NEAQS-2004. All of the models except STEM-2K3 have low bias with a median bias ratio (model/obs.) of 85% for the ensemble mean. Correlations and RMSE-based skill for all models, and their ensemble, are much smaller in 2006 compared to 2004 and fall well below persistence. This statistical picture is not altered appreciably when periods of Saharan dust events along the Texas coastline are removed from the analysis. Though surface monitor PM_{2.5} levels were low during TexAQS II, the lower correlations, skill and low biases suggest a missing component to the PM_{2.5} forecasts. The daytime aircraft comparisons of PM_{2.5} yield a different picture, similar to aircraft comparisons in 2004, with most models showing positive bias compared to the PM volume measurements. This is despite the fact that all models severely underpredict OA, the dominant component of ambient PM_{2.5} measured during both field studies. As discussed in section 4.2, this discrepancy has been explained previously by an overestimation of primary, unspeciated PM_{2.5} emissions within the inventories that compensate the lack of secondary OA formation within the models. Another source of model low bias at the surface may be due to biomass burning sources not considered within the models. As discussed in relation to the observed PM and OC emission flux estimates (Figure 9, right), strong signatures of biogenic burning were observed on 21 and 25 September 2006. Isolated biomass burning signatures from the WP-3 aircraft have been reported on two other days as well [Schwarz *et al.*, 2008]. While the already high model PM_{2.5} biases within the aircraft comparisons argues against an additional burning source, it is nonetheless possible that the nine flight days within the aircraft analysis undersampled biomass burning conditions during the 49-day period of the surface analysis.

[53] Model statistics with the WP-3 aircraft data are examined in terms of all daytime data between 400 and 650 m AGL (11–27 September 2006) and also in terms of a subset of this collection, the 10% of data from upwind transects of either Houston or Dallas. For all species considered, and nearly all models, there is little difference between the bias in the total data set and the bias in only the upwind transects. This implies that most model discrepancy with aircraft observations is not due directly to urban emission specification, rather to model processes acting over a larger regional scale, such as boundary conditions, upwind larger-scale emissions, and parameterizations specific to each modeling system. Conversely, looking only at model/observed biases provides very little information from which to evaluate the urban emissions for Dallas and Houston specified within the models.

[54] Observed concentration difference ratios (relative to NO_y and CO) show that secondary organic aerosol (SOA) formation from Houston and Dallas occurs rapidly, within 3 h or 50 km downwind of the urban centers, and SOA formation is 40 to 80% more efficient for Houston compared to Dallas/Fort Worth. Preferential formation of SOA downwind of the industrial Houston ship channel is docu-

mented by Bahreini *et al.* (submitted manuscript, 2008), demonstrating that the difference in SOA production efficiency is due to Houston's unique petrochemical footprint. The precursors and photochemical/heterogeneous chemistry leading to SOA formation are not known with certainty and are likely due to a variety of biogenic and anthropogenic sources [Kroll and Seinfeld, 2008]. Comparing model and observed source apportionment from the sparse surface network over the United States is complicated by regional and seasonal SOA forcing from biomass burning and wild-fires and uncertainties in biogenic emissions [Yu *et al.*, 2007b]. Recent model estimates have North American biogenic SOA sources dominating over anthropogenic sources for the larger continental scale [Heald *et al.*, 2008; Fu *et al.*, 2008], while precursor volatility and emission-based estimates argue for a dominant anthropogenic contribution [Donahue *et al.*, 2009]. The fresh, urban emissions results presented here show that urban and industrial SOA sources are clearly missing in the AQFMs. Since most of the AIRNow surface PM_{2.5} monitors are located near the urban centers, the low bias seen in the AQFMs may also be related to the absence of these urban SOA sources.

[55] Point-by-point comparisons of AQFM results with aircraft data is always complicated by model errors in horizontal transport as well as deficiencies in the treatment of turbulent dynamics affecting vertical exchange within the PBL for the various constituents. A snapshot of AQFM/observation comparisons for 25 August 2006 (Figure 7c) under cloud-free conditions illustrates the large disparity in vertical profiles for water vapor and NO_y produced by the various models, with many of them showing inconsistencies between meteorological and constituent transport in terms of mixing depth and degree of mixing. PBL mixing is only one of the many model uncertainties that contributes to overall model bias or to compensating errors. More detailed analysis of the relationship between PBL depth and O₃ observed during TexAQS II are given by R. Banta *et al.* (Dependence of daily peak O₃ concentrations near Houston, Texas, on environmental factors: Wind speed, temperature, and boundary-layer depth, submitted to *Journal of Geophysical Research*, 2008), and the relationship between PBL biases and surface O₃ biases within the CMAQ/NAM and WRF/Chem models is covered by Wilczak *et al.* (submitted manuscript, 2008).

[56] The definition of an upwind background for each species and urban flight pattern, in conjunction with the standard assumption of complete and capped PBL mixing, allows integrated flux estimates to be made over areas on the order of 100 × 100 km² encompassing Houston and Dallas/Fort Worth. The estimated fluxes of NO_y show consistency within 20% for the eight flight days over Houston analyzed here, and ~50% consistency for the two flight days over Dallas/Fort Worth. Comparisons of the estimated fluxes with the NEI-99 emissions inventory show ~40% overestimation of NO_y emissions in the base inventory for Houston and ~57% higher fluxes for the Dallas/Fort Worth region. However, significant changes in NO_y emissions are known to have occurred for large point sources [Frost *et al.*, 2006], which accounts for ~27% of the disparity for Houston and 11% for Dallas/Fort Worth. Reductions of on-road NO_x emissions in major U.S. cities

of 50% or more between 1998 and 2007 have been observed [Bishop and Stedman, 2008], and similar changes have been estimated for Houston and Dallas/Fort Worth by TCEQ. If a 40 to 50% reduction in on-road NO_y emissions is assumed, the discrepancies in total NO_y emissions for Houston and Dallas/Fort Worth are reduced to within 20% of the observationally derived values. In light of such large year-to-year emission changes for NO_y, and other species, it is obviously important that AQFMS use the most recent emissions data possible.

[57] The technique used to derive emission ratio estimates, based on concentration differences for the observations and AQFMs, accurately reproduces the NEI-99 inventory ratios for the case of the WRF/Chem model results. Close to urban centers, the ratios obtained for fairly inert species within WRF/Chem (e.g., CO, SO₂, EC, OA and PM_{2.5}) are typically within 12% (see Web page), while concentration difference ratios reproduce input emission ratios typically within 25% for more reactive species. The concentration difference ratios clearly shows discrepancies between the emission inventories used within the AQFMs and data sampled by the NOAA WP-3 aircraft. The CO/NO_y emission discrepancy detailed by Frost et al. (manuscript in preparation, 2009) is a very clear case of inventory discrepancy affecting AQFM model results that is apparent with the concentration difference ratios shown here (Figure 12, top left), but is not apparent in traditional bias analysis of the aircraft data (e.g., Figure 6). Concentration difference ratios also point to inconsistencies for other directly emitted species (ethylene, aromatics, and speciated PM_{2.5}), as well as photochemically produced species (e.g., O₃, HNO₃, CH₂O, CH₃CHO) that need to be reconciled within the suite of AQFMs presented here.

[58] Several recommendations can be made regarding improvements needed within existing air quality forecast models, as well as future lines of research, on the basis of the evaluations presented in this study. Obvious deficiencies within each modeling framework have been noted, and in many cases those deficiencies have already been addressed by the individual forecast centers. Disparities in the treatment of PBL dynamics points to a need for more comprehensive evaluations of meteorological fields than what is performed here in order to establish a more quantitative connection between meteorological forecast skill and air quality forecast skill. Independent confirmation of observed emission rates are needed, particularly for NO_y. Satellite NO₂ column measurements in conjunction with meteorological reanalysis may allow this possibility. By normalizing out backgrounds and factoring out biases inherent to AQFM dynamics, concentration difference ratios appear to provide a useful tool in the diagnosis of near-urban source emissions, photochemistry and aerosol formation. However, the close correspondence between concentration difference ratios and emission ratios is demonstrated for only two of the seven models used within this study, and further verification of the technique, or the development of an alternative methodology is desirable. Future model evaluations using comprehensive aircraft data should also require emission inventory information from each modeling center in order to unequivocally explain discrepancies with observations, intermodel differences, and emission inventory biases determined by indirect means. Finally, the connection between

emission inventory errors and air quality forecast errors outlined in this paper still need confirmation. This is only possible if the emissions inventories are updated with the top-down information provided by intensive field programs such as TexAQS II.

Appendix A: Descriptions of Air Quality Models

A1. NOAA/ESRL/GSD WRF/CHEM

[59] The Weather Research and Forecasting (WRF) Chemical model is based upon the nonhydrostatic WRF community model developed at NCAR (National Center for Atmospheric Research) (<http://www.wrf-model.org>). Details of WRF/Chem are given by Grell et al. [2005] and Fast et al. [2006]. The real-time forecasts can be found at Internet Web address <http://www.frd.fsl.noaa.gov/aq/wrf>. This model system is “online” in the sense that all processes affecting the gas phase and aerosol species are calculated in lock step with the meteorological dynamics. Meteorological initial conditions are taken from the Rapid Update Cycle (RUC) model analysis fields generated at NOAA/ESRL/GSD, and lateral boundary conditions from the NCEP ETA-model forecast. Hourly output from WRF/Chem forecasts are started at 0000 UT of each day using a horizontal grid spacing of 36 km with an additional model forecast of 12 km horizontal resolution embedded within the 36-km domain using one-way nesting. The physics options within WRF/Chem included the Noah land-surface model, MM5 similarity surface layer, YSU boundary layer scheme, RRTM long-wave scheme, Dudhia (12-km forecast) or Goddard (36-km forecast) short-wave schemes, Thompson microphysics scheme, and a version of the Grell-Dévényi ensemble convection parameterization [Grell and Dévényi, 2002].

[60] Gas-phase chemistry is based upon the RADM2 mechanism of Stockwell et al. [1990] with updates to the original mechanism [Stockwell et al., 1995]. Lateral boundary conditions for O₃ and its precursors are prescribed identical to work by McKeen et al. [2002], and are based on averages of midlatitude aircraft profiles from several field studies over the eastern Pacific. Lateral boundary conditions are specified on inflow, and are the same on all sides of the 36-km simulations.

[61] Emissions used within the WRF-Chem simulations are taken from the EPA 1999 National Emission Inventory (NEI, version 3, March 2004 release) [U.S. EPA, 2004a]. A detailed discussion of this inventory and its implementation within WRF/Chem is given by Frost et al. [2006]. Briefly, hourly emissions of NO_x, SO₂, CO, speciated VOCs, NH₃, speciated PM_{2.5}, and total PM₁₀ were prepared for an average day in the 1999 summer ozone season (May–September) on a 4 km × 4 km grid. These emissions were then projected onto the 12- and 36-km grids used in the WRF-Chem simulations.

[62] Biogenic emissions of isoprene, monoterpenes, other VOC (OVOC), and nitrogen emissions by the soil are specified at reference temperature and PAR (photosynthetic active radiation) according to Guenther et al. [1994] for deciduous, coniferous and mixed forest. Agricultural and grassland NO fluxes are those reported by Simpson et al. [1995]. Temperature and light dependence of isoprene

emissions are taken from *Guenther et al.* [1995], while the temperature dependence of monoterpenes, soil NO and OVOC are those of *Simpson et al.* [1995]. Emissions are applied over a surface grid according to the single WRF land-use category assigned to that grid, and the temperature dependence of the emissions is tied to the surface temperature. Similar to the anthropogenic sources, the emissions of monoterpenes and OVOC are disaggregated into the RADM2 species classes.

A2. Environment Canada CHRONOS and AURAMS Models

[63] CHRONOS (Canadian Hemispheric and Regional Ozone and NO_x System) is the original Canadian national AQFM designed for O₃ and PM forecasts. CHRONOS has been used for operational forecasting since 2001, and is based upon the chemical-transport model of *Pudykiewicz et al.* [1997]. Real-time forecasts and limited information about this model can be found at Internet address http://www.weatheroffice.gc.ca/chronos/index_e.html. CHRONOS is an offline model on a polar stereographic grid (standard latitude 60°N) covering contiguous North America (Canada and USA) with the subregion for data used in this study shown in Figure 1. It is currently driven by meteorological fields (at hourly intervals) calculated by the 15-km regional version of the Global Environmental Model (GEM), the operational weather prediction model of the Environment Canada [*Côté et al.*, 1998a, 1998b]. Several upgrades to the GEM model used during TexAQS II are given by *Mailhot et al.* [2006]. Hourly output is obtained from 48-h forecasts at a horizontal grid spacing of 21 km started at 0000 and 1200 UT of each day. CHRONOS uses the ADOM-II (Acid Deposition and Oxidant Model–2) chemical mechanism, which is based upon the lumped approach of *Lurmann et al.* [1986] with updates to kinetic rates and reaction pathways from *Atkinson et al.* [1992]. CHRONOS simulates PM2.5 concentration using a bulk approach. It takes into account bulk primary PM emission and simplified bulk secondary formation of sulfate, nitrate, ammonium and secondary organics.

[64] Anthropogenic emissions are prepared with SMOKE (Sparse Matrix Operator Kernel Emissions, version 2.2) using 2001 base year U.S. emissions (NEI-2001, version 3), 2000 base year emissions for Canada, and 1999 base year emissions for Mexico, all available through the U.S. EPA (<ftp://ftp.epa.gov/EmisInventory/2001v2CAP/>). Biogenic emissions are processed in-line and based on the BEIS2 (Biogenic Emissions Inventory System-2) and the BELD (Biogenic Emissions Land-use Database, version 3) surface vegetation characterization described by *Pierce et al.* [1998].

[65] The AURAMS model is similar to CHRONOS in that it was built upon the CHRONOS AQFM. It was designed as an episodic, regional particulate matter modeling system. AURAMS is also an off-line model and is driven by the meteorological fields (at 15-min intervals) from a 24-km limited-area version of GEM, which is in turn driven by the regional GEM forecast. For the purpose of the TexAQS II study, daily 48-h forecasts, starting at 0000 UT, were run with AURAMS using a horizontal grid spacing of 28 km for the domain given in Figure 1. AURAMS employs the same ADOM II gas-phase chemical mechanism

but, in addition, has a full, size-resolved and chemically resolved, representation of aerosol microphysics, gas-aerosol interaction processes, and cloud-phase gas and aerosol processing [*Gong et al.*, 2003, 2006]. Updates to AURAMS from the model used in the 2004 evaluation include: treating CO as a dynamic species with prescribed emissions, using instantaneous aerosol yield based on work by *Jiang* [2003] for SOA formation, separating total organic PM_{2.5} aerosol into primary organic aerosol (OA) and secondary OA, and using the gridded ozonesonde climatology of *Logan* [1999] for ozone lateral boundary conditions. The anthropogenic emissions of gaseous precursors are identical to that of CHRONOS. The biogenic emission processing in AURAMS is also similar, but used the BEIS-3.03 emission assignments between vegetative categories and specific biogenic VOC.

A3. BAMS Multiscale Air Quality Simulation Platform-Real-Time Models

[66] Multiscale Air Quality Simulation Platform-Real-Time (MAQSIP-RT) [*McHenry et al.*, 2004] is an off-line chemical-transport model that has been applied for real-time ozone forecasting since 1998 (J. N. McHenry et al., Real-time nested mesoscale forecasts of lower tropospheric ozone using a highly optimized coupled model numerical prediction system, paper presented at Symposium on Interdisciplinary Issues in Atmospheric Chemistry, American Meteorological Society, Dallas, Texas, 1999). During 2006, MAQSIP-RT utilized the WRF (version2) mesoscale model for meteorological information. WRF forecasts of near-surface meteorology are also used to compute all emissions components that are meteorologically modulated within the SMOKE (Sparse Matrix Operator Kernel Emissions, version 2.5-RT) emissions processing/modeling system (C. J. Coats Jr., High-performance algorithms in the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system, paper presented at Ninth AMS Joint Conference on Applications of Air Pollution Meteorology with A&WMA, American Meteorological Society, Atlanta, Georgia, 1996). These include all biogenic emissions, point-source plume rise, all mobile source emissions, and evaporative VOC emissions from stationary (fuel storage tank) sources. During summer 2006, twice-daily forecasts were provided for a 45-km domain, the 15-km domain shown in Figure 1a and a 5-km domain. The WRF initial and boundary conditions are derived from NCEP's operational GFS model. MAQSIP-RT is configured with a modified Carbon-Bond 4 (CBM-4b) chemistry mechanism [*Gery et al.*, 1989], with updates listed in the ICARTT/NEAQS-2004 MAQSIP-RT model description [*McKeen et al.*, 2005].

[67] For summer 2006 the EPA NEI Version 3 (2001) point, area, and nonroad anthropogenic emission inventories [U.S. EPA, 2004b] were used as the base inventory for the SMOKE processing system. This was augmented by applying reduction factors for major NO_x point sources based on U.S. DOE (Department of Energy) fuel use data trends for Electric Generating Units obtained from EPA (G. A. Pouliot, The emissions processing system for the Eta/CMAQ air quality forecast system, paper presented at 7th Conference on Atmospheric Chemistry, American Meteorological Society, San Diego, California, 2005, available at http://ams.confex.com/ams/Annual2005/techprogram/programexpanded_

257.htm), and, on a statewide basis (except for Arizona, Oklahoma and Texas), project the 2001 major point source NO_x emissions to the year 2004. For Texas point sources two additional inventories provided by TCEQ were used to augment the base inventory; a statewide NO_x point source inventory based on CEMS measurements for 2003 was applied to all three domains, and a special “upset-VOC” inventory adjunct for the Houston area was applied to the 5-km domain. The “upset-VOC” inventory involves the addition of 149 ton/d of olefin to point VOC sources in the Houston area based on TexAQS 2000 measurements and CMAQ model analysis [Byun et al., 2007]. Except for this additional Houston point source modification, the temporal and VOC speciation profiles, the spatial surrogate database, and cross-reference tables were the same as those used in the ICARTT/NEAQS-2004 study. A 2000 base year Canadian inventory available through the U.S. EPA [U.S. EPA, 2004c] was used in the 45-km domain. For on-road (mobile) emissions, SMOKEv2.5-RT includes the MOBILE6 [U.S. EPA, 2003] modeling system for estimating motor vehicle emissions, which was implemented with updated year 2004 VMT (Vehicle Miles Traveled) and VMT mix for seven Highway Performance Monitoring System roadway classes (interstate freeway, urban freeway, principal arterial, minor arterial, major collector, minor collector, and locals). Biogenic emissions modeling utilized the Biogenic Emissions Inventory System (BEIS) version 3.09 (J. Vukovich and T. Pierce, Implementation of BEIS3 within the SMOKE Modeling Framework, presented at the Emissions Inventory Conference, Environmental Protection Agency, Atlanta, Georgia, 2002, available at <http://www.epa.gov/ttn/chief/conference/ei11/index.html#ses-10>) with land use obtained from the Biogenic Emissions Landuse Database version 3 (BELD3) [Pierce et al., 1998].

A4. University of Iowa STEM-2K3 Model

[68] The University of Iowa STEM model (Sulfur Transport and dEposition Model) was initially used for simulating sulfur dioxide (SO₂) transport and transformation in the atmosphere [Carmichael et al., 1986]. This model has been generalized to simulate regional air quality [Tang et al., 2003; Carmichael et al., 2003; Tang et al., 2004], and the latest version is used here. During the TexAQS II experiment, STEM-2K3 provided a multiscale forecast, including a primary domain with 60-km horizontal resolution covering the contiguous United States, and a nested 12-km domain covering Texas. This model system was driven off-line by a corresponding nested WRF-ARW meteorological model (WSM 3-class simple ice scheme, RRTM long-wave, Dudhia short-wave, Monin-Obukhov surface layer, Noah land surface, YSU boundary layer, and Grell-Dévényi ensemble cumulus cloud scheme) using GFS data for initial and boundary conditions.

[69] STEM-2K3 employs the SAPRC-99 [Carter, 2000] gaseous mechanism, aerosol thermodynamics module SCAPE II (Simulating Composition of Atmospheric Particles at Equilibrium) [Kim et al., 1993a, 1993b; Kim and Seinfeld, 1995] and NCAR Tropospheric Ultraviolet-Visible (TUV) radiation model [Madronich and Flocke, 1999]. In this experiment, the STEM-2K3 model included aerosols in four size bins (in diameter): 0.1–0.3 μm, 0.3–1.0 μm, 1.0–2.5 μm, and 2.5–10 μm [Tang et al., 2004]. Thirty

photolysis frequencies required for the SAPRC-99 mechanism are explicitly treated online taking into account the influence of aerosols and clouds [Tang et al., 2003].

[70] STEM-2K3 used time-varying lateral and top boundary conditions provided by RAQMS global chemical transport model forecasts [Pierce et al., 2007] with 2-degree horizontal resolution, and updates every 6 h. During this experiment, the STEM-2K3 used the EPA NEI-2001 emission inventory (compiled by Jeffrey M. Vukovich, Baron AMS, Inc.), and is compatible with the inventory developed for the BAMS MAQSIP-RT 15-km horizontal resolution model described above. This includes the CEMS-based point emissions for NO_x. BEIS2 biogenic emissions were employed in this study. The RAQMS global model has its own emissions [Pierce et al., 2007], which are mainly based on the EDGAR inventory [Olivier et al., 1996].

A5. NWS/NCEP CMAQ/NAM Model

[71] Surface O₃ forecasts from the CMAQ/NAM are based on the off-line photochemical-transport CMAQ model [Byun and Schere, 2006], meteorological fields derived from the NWS/NCEP NAM forecasts, and emissions processing also based on NAM forecast meteorological fields (J. McQueen et al., Development and evaluation of the NOAA/EPA prototype air quality model prediction system, paper presented at 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, American Meteorological Society, Seattle, Washington, 2004). In 2006 the Eta model was replaced by the Weather Research and Forecasting Nonhydrostatic Mesoscale Model (WRF-NMM) as the operational NAM. An interface component, PREMAQ, which facilitates the transformation of NAM-derived meteorological fields to conform to the CMAQ grid structure, coordinate system, and input data format, has been developed [Otte et al., 2005]. During 2006, three sets of forecast simulations were performed with this CMAQ-NAM system: (1) operational O₃ forecasts over the eastern United States, (2) experimental O₃ forecasts over the contiguous United States, and (3) developmental forecasts for both O₃ and fine particulate matter (PM_{2.5}) over the contiguous United States. All simulations were performed with a horizontal grid resolution of 12 km and the vertical extent from the surface to 100 mbar was discretized with 22 layers of variable thickness. The CBM-4 mechanism is used in the photochemical calculations, which includes several improvements and additions to the original Gery et al. [1989] formulation mentioned previously and detailed by Byun and Ching [1999].

[72] The operational runs were based on CMAQ and interface module configurations described previously [Yu et al., 2007a; Otte et al., 2005; McKeen et al., 2005]. However, to rectify deficiencies in forecast results noted in previous years, several modifications to process modules, coupling between the meteorological and chemical calculations, and input data specification were tested in the developmental CMAQ-NAM configuration as summarized below. Both the horizontal and vertical grid and coordinate systems used in the WRF-NMM are different from those traditionally employed in CMAQ. To reduce errors associated with interpolation of meteorological data from the WRF-NMM coordinate and grid structure to that of CMAQ, modifications were introduced to the system to improve

coupling in the vertical direction such that the CMAQ calculations are performed with the same hybrid sigma-P vertical coordinate system that is utilized in the WRF-NMM. To reduce the systematic overestimation of O₃ in the free troposphere, static lateral boundary conditions based on “clean” tropospheric conditions were used. Additionally, the cloud scheme was modified to reduce the impacts of unrealistic downward entrainment of higher O₃ from above cloud. Additional aspects of the CMAQ model configuration used in the 2006 forecast applications are summarized by R. Mathur et al. (The Community Multiscale Air Quality (CMAQ) model: Model configuration and enhancements for 2006, available at http://www.nws.noaa.gov/ost/air_quality/2006/fghome06.htm).

[73] The emission processing and incorporation of both anthropogenic and biogenic emissions into the CMAQ model is based on the Sparse Matrix Operator Kernel Emissions (SMOKE) system, which is also used with the Baron AMS MAQSIP-RT model described above. The primary differences are that the mobile emissions are based on the MOBILE6 emissions model [U.S. EPA, 2003], and that the meteorologically dependent components of the SMOKE emissions modeling system have been incorporated into the single interface program PREMAQ, and use the NAM forecast fields to determine the meteorologically modulated emissions. The emission inventories used by the CMAQ-NAM system were updated to represent the 2006 forecast period. NO_x and SO₂ emissions from Electric Generation Units (EGUs) were projected to 2006 using the 2004 Continuous Emissions Monitoring (CEM) data and 2004 to 2006 projection estimates derived from the annual energy outlook by the Department of Energy (<http://www.eia.doe.gov/oiaf/aoe>). Additionally, monthly temporal profiles were created on a state-by-state basis using the 2004 CEM for EGU emissions and used to represent the monthly variations in EGU emissions during the 2006 forecast period. For other pollutants and non-EGU point sources, base year 2001 emission estimates were used. Since MOBILE6 is computationally expensive and inefficient for real-time applications, mobile source emissions were estimated using approximations to the MOBILE6 model. In this approach MOBILE6 was used to create retrospective emissions over an 8-week period over the air quality forecast grid using the 2006 projected vehicle miles traveled (VMT) data and 2006 vehicle fleet information. Least squares regressions relating the emissions to variations in temperature were then developed for each grid cell at each hour of the week and for each emitted species (G. A. Pouliot, presented paper, 2005). Consequently, mobile emissions could then be readily estimated in the forecast system using the temperature fields from the NAM model. Area source emissions were based on the 2001 National Emissions Inventory, version 3, while BEIS3.12 (T. Pierce et al., Integration of the Biogenic Emission Inventory System (BEIS3) into the Community Multiscale Air Quality Modeling System, paper presented at 12th Joint Conference on the Applications of Air Pollution Meteorology with the A&WMA, American Meteorological Society, Norfolk, Virginia, 2002) was used to estimate the biogenic emissions.

[74] **Acknowledgments.** This research is partially funded by Early Start Funding from the NOAA Office of Atmospheric Research, the NOAA

Air Quality Program, and the NOAA/NWS Office of Science and Technology. Credit for program support and management is given to Paula Davidson (NOAA/NWS/OST), Steve Fine (NOAA/ARL), and Jim Meagher (NOAA/ESRL/CSD). Michael Trainer (NOAA/ESRL/CSD) designed and managed the WP-3 flight plans that were instrumental in this work. The thoughtful comments of two reviewers are highly appreciated. This work would not be possible without the measurements collected on board the NOAA WP-3 aircraft by the following people within NOAA/ESRL/CSD: aerosol volume measurements by Chuck Brock, aerosol mass spectrometer measurements by Ann Middlebrook and Roya Bahreini, NO_y and NO_x measurements by Tom Ryerson. Computational and logistic assistance from the following individuals and organizations is also gratefully appreciated: Carlie Coats (Baron AMS), Ted Smith (Baron AMS), Don Olerud (Baron AMS), Jeff Vukovich (Baron AMS) for emissions used by BAMS and the University of Iowa, Sophie Cousineau (EC/MSC), L.-P. Crevier (EC/MSC), Hugo Landry (EC/SMC), Mike Moran (S&T/EC), Paul Makar (S&T/EC), Balbir Pabla (S&T/EC), and the NOAA/FSL High Performance Computing Facility. The U.S. Environmental Protection Agency through its Office of Research and Development collaborated in the research described here. It has been subjected to Agency review and approved for publication.

References

- Atkinson, R., D. L. Baulch, R. A. Cox, R. F. Hampson, J. A. Kerr, and J. Troe (1992), Evaluated kinetic and photochemical data for atmospheric chemistry: Supplement IV, *Atmos. Environ., Ser. A*, 26, 1187–1230.
- Bishop, G., and D. Stedman (2008), A decade of on-road emissions measurements, *Environ. Sci. Technol.*, 42, 1651–1656, doi:10.1021/es702413b.
- Bond, T. C., D. G. Streets, K. F. Yarber, S. M. Nelson, J. H. Woo, and Z. Klimont (2004), A technology-based global inventory of black and organic carbon emissions from combustion, *J. Geophys. Res.*, 109, D14203, doi:10.1029/2003JD003697.
- Brock, C. A., et al. (2003), Particle growth in urban and industrial plumes in Texas, *J. Geophys. Res.*, 108(D3), 4111, doi:10.1029/2002JD002746.
- Brock, C. A., et al. (2008), Sources of particulate matter in the northeastern United States in summer: 2. Evolution of chemical and microphysical properties, *J. Geophys. Res.*, 113, D08302, doi:10.1029/2007JD009241.
- Byun, D. W., and J. K. S. Ching (Eds.) (1999), Science algorithms of the EPA Models-3 Community Multiscale 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 Multiscale Air Quality (CMAQ) modeling system, *Appl. Mech. Rev.*, 59, 51–77, doi:10.1115/1.2128636.
- Byun, D. W., S.-T. Kim, and S.-B. Kim (2007), Evaluation of air quality models for the simulation of a high ozone episode in the Houston metropolitan area, *Atmos. Environ.*, 41, 837–853, doi:10.1016/j.atmosenv.2006.08.038.
- Carmichael, G. R., L. K. Peters, and T. Kitada (1986), A second generation model for regional-scale transport chemistry and deposition, *Atmos. Environ.*, 20, 173–188, doi:10.1016/0004-6981(86)90218-0.
- Carmichael, G. R., et al. (2003), Regional-Scale chemical transport modeling in support of intensive field experiments: Overview and analysis of the TRACE-P Observations, *J. Geophys. Res.*, 108(D21), 8823, doi:10.1029/2002JD003117.
- Carter, W. (2000), Documentation of the SAPRC-99 chemical mechanism for VOC reactivity assessment, final report, Univ. of Calif.-Riverside, Riverside.
- Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth (1998a), The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation, *Mon. Weather Rev.*, 126, 1373–1395, doi:10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO;2.
- Côté, J., J.-G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth (1998b), The operational CMC/MRB Global Environmental Multiscale (GEM) model. Part II: Results, *Mon. Weather Rev.*, 126, 1397–1418, doi:10.1175/1520-0493(1998)126<1397:TOCMGE>2.0.CO;2.
- Dabberdt, W. F., et al. (2004), Meteorological research needs for improved air quality forecasting, *Bull. Am. Meteorol. Soc.*, 85, 563–586, doi:10.1175/BAMS-85-4-563.
- de Gouw, J. A., et al. (2008), Sources of particulate matter in the northeastern United States: 1. Direct emissions and secondary formation of organic matter in urban plumes, *J. Geophys. Res.*, 113, D08301, doi:10.1029/2007JD009243.
- Delle Monache, L., J. Wilczak, S. McKeen, G. Grell, M. Pagowski, S. Peckham, R. Stull, J. McHenry, and J. McQueen (2008), A Kalman-filter bias correction method applied to deterministic, ensemble averaged, and probabilistic forecasts of surface ozone, *Tellus, Ser. B*, 60(2), 238–249, doi:10.1111/j.1600-0889.2007.00332.x.

- Donahue, N. M., A. L. Robinson, and S. N. Pandis (2009), Atmospheric organic particulate matter: From smoke to secondary organic aerosol, *Atmos. Environ.*, **43**, 94–106, doi:10.1016/j.atmosev.2008.09.055.
- Fast, J. D., W. I. Gustafson Jr., R. C. Easter, R. A. Zaveri, J. C. Barnard, E. G. Chapman, G. A. Grell, and S. E. Peckham (2006), Evolution of ozone, particulates, and aerosols direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model, *J. Geophys. Res.*, **111**, D21305, doi:10.1029/2005JD006721.
- Fehsenfeld, F. C., et al. (2006), International Consortium for Atmospheric Research on Transport and Transformation (ICARTT): North America to Europe—Overview of the 2004 summer field study, *J. Geophys. Res.*, **111**, D23S01, doi:10.1029/2006JD007829.
- Frost, G. J., et al. (2006), Effects of changing power plant emissions on ozone in the eastern United States, *J. Geophys. Res.*, **111**, D12306, doi:10.1029/2005JD006354.
- Fu, T. M., D. J. Jacob, F. Wittrock, J. P. Burrows, M. Vrekoussis, and D. K. Henze (2008), Global budgets of atmospheric glyoxal and methylglyoxal, and implications for formation of secondary organic aerosols, *J. Geophys. Res.*, **113**, D15303, doi:10.1029/2007JD009505.
- Galmarini, S., et al. (2004), Ensemble dispersion forecasting—Part I: Concept, approach and indicators, *Atmos. Environ.*, **38**, 4607–4617, doi:10.1016/j.atmosev.2004.05.030.
- Gery, M. W., G. Z. Whitten, J. P. Killus, and M. C. Dodge (1989), A photochemical kinetics mechanism for urban and regional scale computer models, *J. Geophys. Res.*, **94**, 12,925–12,956, doi:10.1029/JD094iD10p12925.
- Gong, S. L., et al. (2003), Canadian Aerosol Module: A size segregated simulation of atmospheric aerosol processes for climate and air quality models: 1. Module development, *J. Geophys. Res.*, **108**(D1), 4007, doi:10.1029/2001JD002002.
- Gong, W., et al. (2006), Cloud processing of gases and aerosols in a regional air quality model (AURAMS), *Atmos. Res.*, **82**, 248–275, doi:10.1016/j.atmosres.2005.10.012.
- Grell, G. A., and D. Dévényi (2002), A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, *Geophys. Res. Lett.*, **29**(14), 1693, doi:10.1029/2002GL015311.
- Grell, G. A., S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. Skamarock, and B. Eder (2005), Fully coupled “online” chemistry within the WRF model, *Atmos. Environ.*, **39**, 6957–6975, doi:10.1016/j.atmosev.2005.04.027.
- Guenther, A., P. Zimmerman, and M. Wildermuth (1994), Natural volatile organic compound emission rate estimates for U.S. woodland landscapes, *Atmos. Environ.*, **28**, 1197–1210, doi:10.1016/1352-2310(94)90297-6.
- Guenther, A. B., et al. (1995), A global model of natural volatile organic compound emissions, *J. Geophys. Res.*, **100**, 8873–8892, doi:10.1029/94JD02950.
- Heald, C. L., D. J. Jacob, R. J. Park, L. M. Russel, B. J. Huebert, J. H. Seinfeld, H. Liao, and R. J. Weber (2005), A large organic aerosol source in the free troposphere missing from current models, *Geophys. Res. Lett.*, **32**, L18809, doi:10.1029/2005GL023831.
- Heald, C. L., et al. (2008), Predicted change in global secondary organic aerosol concentrations in response to future climate, emissions, and land use change, *J. Geophys. Res.*, **113**, D05211, doi:10.1029/2007JD009092.
- Jiang, G. F., and J. D. Fast (2004), Modeling the effects of VOC and NO_x emission sources on ozone formation in Houston during the TexAQS 2000 field campaign, *Atmos. Environ.*, **38**, 5071–5085, doi:10.1016/j.atmosev.2004.06.012.
- Jiang, W. (2003), Instantaneous secondary organic aerosol yields and their comparison with overall aerosol yields for aromatic and biogenic hydrocarbons, *Atmos. Environ.*, **37**, 5439–5444, doi:10.1016/j.atmosev.2003.09.018.
- 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**(12), 1782–1796.
- Karl, T., T. Jobson, W. C. Kuster, E. Williams, J. Stutz, R. Shetter, S. R. Hall, P. Goldan, F. Fehsenfeld, and W. Lindinger (2003), Use of proton-transfer-reaction mass spectrometry to characterize volatile organic compound sources at the La Porte super site during the Texas Air Quality Study 2000, *J. Geophys. Res.*, **108**(D16), 4508, doi:10.1029/2002JD003333.
- Kim, S.-W., A. Heckel, S. A. McKeen, G. J. Frost, E.-Y. Hsie, M. K. Trainer, A. Richter, J. Burrows, S. E. Peckham, and G. A. Grell (2006), Satellite-observed U.S. power plant NO_x emission reductions and their impact on air quality, *Geophys. Res. Lett.*, **33**, L22812, doi:10.1029/2006GL027749.
- Kim, Y. P., and J. H. Seinfeld (1995), Atmospheric gas-aerosol equilibrium III: Thermodynamics of crustal elements Ca²⁺, K⁺, Mg²⁺, *Aerosol Sci. Technol.*, **22**, 93–110, doi:10.1080/02786829408959730.
- Kim, Y. P., J. H. Seinfeld, and P. Saxena (1993a), Atmospheric gas-aerosol equilibrium I: Thermodynamic model, *Aerosol Sci. Technol.*, **19**, 151–181.
- Kim, Y. P., J. H. Seinfeld, and P. Saxena (1993b), Atmospheric gas-aerosol equilibrium II: Analysis of common approximations and activity coefficient calculation methods, *Aerosol Sci. Technol.*, **19**, 182–198, doi:10.1080/02786829308959629.
- Kleinman, L. I., P. H. Daum, D. Imre, Y.-N. Lee, L. J. Nunnermacker, S. R. Springston, J. Weinstein-Lloyd, and J. Rudolph (2002), Ozone production rate and hydrocarbon reactivity in 5 urban areas: A cause of high ozone concentration in Houston, *Geophys. Res. Lett.*, **29**(10), 1467, doi:10.1029/2001GL014569.
- Kroll, J. H., and J. H. Seinfeld (2008), Chemistry of secondary organic aerosol: Formation and evolution of low-volatility organics in the atmosphere, *Atmos. Environ.*, **42**, 3593–3624, doi:10.1016/j.atmosev.2008.01.003.
- Kuster, W. C., F. J. M. Harren, and J. A. deGouw (2005), Inter-comparison of laser photoacoustic spectroscopy and gas chromatography techniques for measurements of ethene in the atmosphere, *Environ. Sci. Technol.*, **39**, 4581–4585, doi:10.1021/es0504385.
- 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, doi:10.1029/1998JD100096.
- Lurmann, F. W., A. C. Lloyd, and R. Atkinson (1986), A chemical mechanism for use in long-range transport/acid deposition computer modeling, *J. Geophys. Res.*, **91**, 10,905–10,936, doi:10.1029/JD091iD10p10905.
- Madronich, S., and S. Flocke (1999), The role of solar radiation in atmospheric chemistry, in *Handbook of Environmental Chemistry*, edited by P. Boule, pp. 1–26, Springer, Heidelberg, Germany.
- Mailhot, J., et al. (2006), The 15-km version of the Canadian regional forecast system, *Atmos. Ocean*, **43**, 134–149. (Available at <http://cmos.metapress.com/content/f0702m0x637t53l/fulltext.pdf>)
- Mathur, R., S. Yu, D. Kang, and K. L. Schere (2008), Assessment of the wintertime performance of developmental particulate matter forecasts with the Eta-Community Multiscale Air Quality modeling system, *J. Geophys. Res.*, **113**, D02303, doi:10.1029/2007JD008580.
- McHenry, J. N., W. F. Ryan, N. L. Seaman, C. J. Coates Jr., J. Pudykiewicz, S. Arunachalam, and J. M. Vukovich (2004), A real-time Eulerian photochemical model forecast system, *Bull. Am. Meteorol. Soc.*, **85**, 525–548, doi:10.1175/BAMS-85-4-525.
- McKeen, S. A., G. Wotawa, D. D. Parrish, J. S. Holloway, M. P. Buhr, G. Hubler, F. C. Fehsenfeld, and J. F. Meagher (2002), Ozone production from Canadian wildfires during June and July of 1995, *J. Geophys. Res.*, **107**(D14), 4192, doi:10.1029/2001JD000697.
- McKeen, S., 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.
- McKeen, S., et al. (2007), The evaluation of several PM2.5 forecast models using data collected during the ICARTT/NEAQS 2004 field study, *J. Geophys. Res.*, **112**, D10S20, doi:10.1029/2006JD007608.
- Olivier, J. G. J., A. F. Bouwman, C. W. M. Van der Maas, J. J. M. Berdowski, J. P. J. Bloos, A. J. H. Visschedijk, P. Y. J. Zandveld, and J. L. Haverlag (1996), Description of EDGAR Version 2.0.: A set of global emission inventories of greenhouse gases and ozone-depleting substances for all anthropogenic and most natural sources on a per country basis and on 1° × 1° grid, *RIVM/TNO Rep. 771060002*, Natl. Inst. of Public Health and the Environ., Bilthoven, Netherlands.
- 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 Forecast.*, **20**, 367–384, doi:10.1175/WAF855.1.
- Pagowski, M., et al. (2005), A simple method to improve ensemble-based ozone forecasts, *Geophys. Res. Lett.*, **32**, L07814, doi:10.1029/2004GL022305.
- Pagowski, M., G. A. Grell, D. Dévenyi, S. Peckham, S. A. McKeen, W. Gong, L. Delle Monache, J. N. McHenry, J. McQueen, and P. Lee (2006), Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts, *Atmos. Environ.*, **40**, 3240–3250, doi:10.1016/j.atmosev.2006.02.006.
- Parrish, D. D. (2006), Critical evaluation of US on-road vehicle emission inventories, *Atmos. Environ.*, **40**, 2288–2300, doi:10.1016/j.atmosev.2005.11.033.
- Parrish, D. D., et al. (2009), Overview of the Second Texas Air Quality Study (TexAQS II) and the Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS), *J. Geophys. Res.*, doi:10.1029/2009JD011842, in press.
- Pierce, R. B., et al. (2007), Chemical data assimilation estimates of continental U. S. ozone and nitrogen budgets during the Intercontinental Chemical Transport Experiment—North America, *J. Geophys. Res.*, **112**, D12S21, doi:10.1029/2006JD007722.
- Pierce, T., C. Geron, L. Bender, R. Dennis, G. Tonneson, and A. Guenther (1998), Influence of increased isoprene emissions on regional ozone modeling, *J. Geophys. Res.*, **103**, 25,611–25,629, doi:10.1029/98JD01804.

- Pudykiewicz, J., A. Kallaur, and P. K. Smolarkiewicz (1997), Semi-Lagrangian modelling of tropospheric ozone, *Tellus, Ser. B*, **49**, 231–258.
- Ryerson, T. B., et al. (1998), Emissions lifetimes and ozone formation in power plant plumes, *J. Geophys. Res.*, **103**, 22,569–22,583, doi:10.1029/98JD01620.
- Ryerson, T. B., et al. (2003), Effect of petrochemical industrial emissions of reactive alkenes and NO_x on tropospheric ozone formation in Houston, Texas, *J. Geophys. Res.*, **108**(D8), 4249, doi:10.1029/2002JD003070.
- Schwarz, J. P., et al. (2006), Single-particle measurements of midlatitude black carbon and light-scattering aerosols from the boundary layer to the lower stratosphere, *J. Geophys. Res.*, **111**, D16207, doi:10.1029/2006JD007076.
- Schwarz, J. P., et al. (2008), Measurement of the mixing state, mass, and optical size of individual black carbon particles in urban and biomass burning emissions, *Geophys. Res. Lett.*, **35**, L13810, doi:10.1029/2008GL033968.
- Simpson, D., A. Guenther, C. N. Hewitt, and R. Steinbrecher (1995), Biogenic emissions in Europe: 1. estimates and uncertainties, *J. Geophys. Res.*, **100**, 22,875–22,890, doi:10.1029/95JD02368.
- Stockwell, W. R., P. Middleton, J. S. Chang, and X. Tang (1990), The second generation regional acid deposition model chemical mechanism for regional air quality modeling, *J. Geophys. Res.*, **95**, 16,343–16,367, doi:10.1029/JD095iD10p16343.
- Stockwell, W. R., P. Middleton, J. S. Chang, and X. Tang (1995), The effect of acetyl peroxy-peroxy radical reactions on peroxyacetyl nitrate and ozone concentrations, *Atmos. Environ.*, **29**, 1591–1599, doi:10.1016/1352-2310(95)00093-E.
- Sullivan, A. P., R. E. Peltier, C. A. Brock, J. A. de Gouw, J. S. Holloway, C. Warneke, A. G. Wollny, and R. J. Weber (2006), Airborne measurements of carbonaceous aerosol soluble in water over northeastern United States: Method development and an investigation into water-soluble organic carbon sources, *J. Geophys. Res.*, **111**, D23S46, doi:10.1029/2006JD007072.
- Tang, Y., et al. (2003), Impacts of aerosols and clouds on photolysis frequencies and photochemistry during TRACE-P: 2. Three-dimensional study using a regional chemical transport model, *J. Geophys. Res.*, **108**(D21), 8822, doi:10.1029/2002JD003100.
- Tang, Y., et al. (2004), Three-dimensional simulations of inorganic aerosol distributions in East Asia during spring 2001, *J. Geophys. Res.*, **109**, D19S23, doi:10.1029/2003JD004201.
- Tilmes, S., et al. (2002), Comparison of five Eulerian air pollution forecasting systems for the summer of 1999 using the German ozone monitoring data, *J. Atmos. Chem.*, **42**, 91–121, doi:10.1023/A:1015753032760.
- U.S. Environmental Protection Agency (U.S. EPA) (2003), User's guide to MOBILE6.1 and MOBILE6.2 Mobile Source Emission Factor Model, *EPA4420-R-03-010*, 262 pp., Washington, D. C. (Available at <http://www.epa.gov/oms/m6.htm>)
- U.S. Environmental Protection Agency (U.S. EPA) (2004a), EPA clearinghouse for inventories and emissions factors: 1999 National Emission Inventory documentation and data—Final version 3.0, report, Washington, D. C. (Available at <http://www.epa.gov/ttn/chief/net/1999inventory.html>)
- U.S. Environmental Protection Agency (U.S. EPA) (2004b), Interstate air quality rule: Notice of data availability, *CAIR Docket OAR-2003-0053*, Washington, D. C. (Available at <http://www.epa.gov/oar/interstateairquality/rule.html>)
- U.S. Environmental Protection Agency (U.S. EPA) (2004c), EPA clearinghouse for inventories and emissions factors: North American Emissions Inventories—Canada, report, Washington, D. C. (Available at <http://www.epa.gov/ttn/chief/net/canada.html#data>)
- Volkamer, R., J. L. Jimenez, F. S. Martini, K. Dzepina, Q. Zhang, D. Salcedo, L. T. Molina, D. R. Worsnop, and M. J. Molina (2006), Secondary organic aerosol formation from anthropogenic air pollution: Rapid and higher than expected, *Geophys. Res. Lett.*, **33**, L17811, doi:10.1029/2006GL026899.
- Warneke, C., et al. (2007), Determination of urban volatile organic compound emission ratios and comparison with an emissions database, *J. Geophys. Res.*, **112**, D10S47, doi:10.1029/2006JD007930.
- Weber, R. J., et al. (2007), A study of secondary organic aerosol formation in the anthropogenic-influenced southeastern United States, *J. Geophys. Res.*, **112**, D13302, doi:10.1029/2007JD008408.
- Wert, B. P., et al. (2003), Signatures of terminal alkene oxidation in airborne formaldehyde measurements during TexAQS 2000, *J. Geophys. Res.*, **108**(D3), 4104, doi:10.1029/2002JD002502.
- Wilczak, J., et al. (2006), Bias-corrected ensemble and probabilistic forecasts of surface ozone over eastern North America during the summer of 2004, *J. Geophys. Res.*, **111**, D23S28, doi:10.1029/2006JD007598.
- Xie, Y., and C. M. Berkowitz (2007), The use of conditional probability functions and potential source contribution functions to identify source regions and advection pathways of hydrocarbon emissions in Houston, Texas, *Atmos. Environ.*, **41**, 5831–5847, doi:10.1016/j.atmosenv.2007.03.049.
- Yu, S., R. Mathur, K. Schere, D. Kang, J. Pleim, and T. Otte (2007a), A detailed evaluation of the Eta-CMAQ forecast model performance for O₃, its related precursors, and meteorological parameters during the 2004 ICARTT study, *J. Geophys. Res.*, **112**, D12S14, doi:10.1029/2006JD007715.
- Yu, S., P. V. Bhate, R. L. Dennis, and R. Mathur (2007b), Seasonal and regional variations of primary and secondary organic aerosols over the continental United States: Semi-empirical estimates and model evaluation, *Environ. Sci. Technol.*, **41**, 4690–4697, doi:10.1021/es061535g.
- V. Bouchet, S. Gaudreault, and J. Racine, Meteorological Service of Canada, Environment Canada, 2121 Route Transcanadienne, Dorval, QC H9P 1J3, Canada.
- G. R. Carmichael, Center for Global and Regional Environmental Research, University of Iowa, Iowa City, IA 52242, USA.
- J. de Gouw, G. Frost, J. Holloway, E.-Y. Hsie, S. McKeen, J. Peischl, J. Schwarz, R. Spackman, and C. Warneke, Chemical Sciences Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Mail Code R/CSD4, 325 Broadway, Boulder, CO 80305, USA. (stuart.a.mckeен@noaa.gov)
- I. Djalalova and J. Wilczak, Physical Sciences Division, Environmental Science Research Laboratory, National Oceanic and Atmospheric Administration, Mail Code R/GSD1, 325 Broadway, Boulder, CO 80305, USA.
- W. Gong, Science and Technology Branch, Environment Canada, 4905 Dufferin Street, Downsview, ON M3H 5T4, Canada.
- G. Grell and S. Peckham, Global Sciences Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Mail Code R/GSD1, 325 Broadway, Boulder, CO 80305, USA.
- P. Lee, Air Resources Laboratory, National Oceanic and Atmospheric Administration, Silver Spring Metro Center 3, Room 3316, 1315 East-West Highway, Silver Spring, MD 20910, USA.
- R. Mathur, National Exposure Research Laboratory, Environmental Protection Agency, 109 T. W. Alexander Drive, Research Triangle Park, NC 27711, USA.
- J. McHenry, Baron Advanced Meteorological Systems, North Carolina Supercomputing Center, 3021 Cornwallis Road, Research Triangle Park, NC 27709, USA.
- J. McQueen and Y. Tang, Environmental Modeling Center, National Center for Environmental Prediction, National Weather Service, National Oceanic and Atmospheric Administration, 5200 Auth Road, W/NP2, Camp Springs, MD 20746, USA.