JOURNAL OF GEOPHYSICAL RESEARCH: ATMOSPHERES, VOL. 118, 1005–1019, doi:10.1029/2012JD018100, 2013

Constraining ozone-precursor responsiveness using ambient measurements

Antara Digar,[1] Daniel S. Cohan,[1] Xue Xiao,[1] Kristen M. Foley,[2] Bonyoung Koo,[3] and Greg Yarwood[3]

Received 16 May 2012; revised 7 November 2012; accepted 25 November 2012; published 30 January 2013.

[1] This study develops probabilistic estimates of ozone (O_3) sensitivities to precursor emissions by incorporating uncertainties in photochemical modeling and evaluating model performance based on ground-level observations of O_3 and oxides of nitrogen (NO_x). Uncertainties in model formulations and input parameters are jointly considered to identify factors that strongly influence O_3 concentrations and sensitivities in the Dallas-Fort Worth region in Texas. Weightings based on a Bayesian inference technique and screenings based on model performance and statistical tests of significance are used to generate probabilistic representation of O_3 response to emissions and model input parameters. Adjusted (observation-constrained) results favor simulations using the sixth version of the carbon bond chemical mechanism (CB6) and scaled-up emissions of NO_x, dampening the overall sensitivity of O_3 to NO_x and increasing the sensitivity of O_3 to volatile organic compounds in the study region. This approach of using observations to adjust and constrain model simulations can provide probabilistic representations of pollutant responsiveness to emission controls that complement the results obtained from deterministic air-quality modeling.

Citation: Digar, A., D. S. Cohan, X. Xiao, K. M. Foley, B. Koo, and G. Yarwood (2013), Constraining ozone-precursor responsiveness using ambient measurements, *J. Geophys. Res. Atmos.*, *118*, 1005–1019, doi:10.1029/2012JD018100.

1. Introduction

[2] Secondary air pollutants like ozone (O_3) are formed as a result of complex nonlinear chemistry between various primary pollutants emitted directly into the atmosphere due to anthropogenic and natural activities. Understanding the responses of ambient pollutant concentrations to emission changes (sensitivity) is therefore crucial for the development of effective pollution abatement strategies. Photochemical models are used to estimate the sensitivity of secondary air pollutants to their precursor emissions, and thus serve as useful tools for determining the amount of emission reduction needed to attain ambient air-quality standards and informing the selection of control strategies.

[3] Models for informing air-quality management are typically run deterministically with a single best-available setting for model formulation and inputs. However, there has been a growing interest in probabilistic representations of model results that account for model uncertainty [*Dennis et al.*, 2010; *Hogrefe and Rao*, 2001]. Uncertainties in pollutant-emission sensitivity may arise from choices of numerical representations of atmospheric processes such as chemical mechanism, vertical mixing scheme, horizontal transport, and emission model (*structural uncertainty*), and/or from the values of input parameters such as emission rates, reaction rate constants, boundary conditions, and deposition velocities (*parametric uncertainty*) [*Deguillaume et al.*, 2008; *Fine et al.*, 2003; *Pinder et al.*, 2009].

[4] Recent work by *Digar and Cohan* [2010] and *Tian et al.* [2010] introduced efficient Monte Carlo techniques for characterizing *parametric* uncertainties in O_3 and particulate matter (PM) responses to emission controls. *Pinder et al.* [2009] jointly considered *parametric* and *structural* uncertainties to develop probabilistic estimates of O_3 concentrations. However, none of these studies evaluated the relative likelihoods of the various Monte Carlo cases.

[5] Previous work by *Bergin and Milford* [2000] had shown that a Bayesian inference approach can weight the relative likelihood of each Monte Carlo model formulation based on its performance in simulating observed concentrations, and thus yield probability distributions for predicting the actual values of pollutant-emission sensitivities as well as model inputs. That study used a simplified two-dimensional trajectory model, and only a handful of studies have applied Bayesian Monte Carlo approaches to characterize O_3 responsiveness in more computationally intensive three-dimensional regional models [*Beekmann and Derognat*, 2003; *Deguillaume et al.*, 2008].

[1]Department of Civil and Environmental Engineering, Rice University, Houston, Texas, USA.
[2]National Exposure Research Laboratory, U.S. Environmental Protection Agency, Durham, North Carolina, USA.
[3]ENVIRON International Corporation, Novato, California, USA.

Corresponding author: D. S. Cohan, Department of Civil and Environmental Engineering, Rice University, 6100 Main Street, MS-519, Houston, Texas 77005, USA. (cohan@rice.edu)

2169-897X/13/2012JD018100

[6] The aim of this study is to develop probabilistic representations of O_3 responsiveness to emission changes constrained by actual measurements of pollutant concentrations. The Monte Carlo Reduced Form Model (RFM) approach of *Digar and Cohan* [2010] has been used to generate a large ensemble of model predictions of O_3 concentrations and responsiveness to emission controls in the Dallas-Fort Worth (DFW) region of Texas, which is currently a nonattainment area for the 1997 eight hour O_3 National Ambient Air Quality Standard (NAAQS). The simulated concentrations of O_3 and its precursor nitrogen oxides ($NO_x \equiv NO$ and NO_2) are compared against observations to yield adjusted (*observation-constrained*) probabilistic representations of photochemical model inputs and output predictions. Use of both Bayesian and non-Bayesian statistical techniques allows us to evaluate the consistency of our results across various observational metrics and methods of comparison. Sections 2 and 3 describe the modeling and measurements used for this work, and section 4 describes the statistical methodology and metrics considered here. Important findings are elaborated in Results and Discussion (section 5), followed by the Conclusion.

2. Photochemical Model Description

2.1. Base Case Modeling

[7] The Comprehensive Air Quality Model with Extensions (CAMx) v5.32 [*ENVIRON*, 2010] is used here to study a 2006 summer episode in DFW spanning from 31 May to 2 July, which includes numerous (17) days with meteorological conditions favoring O_3 formation. This period was identified by the Texas Commission on Environmental Quality (TCEQ) based on its prevalence of observed 8 hour daily maximum O_3 concentrations exceeding the 8-hour O_3 1997 NAAQS of 84 ppb [*Texas Commission on Environmental Quality* (*TCEQ*), 2011a, 2011b]. Results for the first 5 days were neglected for model initialization. Sensitivity of O_3 to its precursor emissions is computed using the high-order decoupled direct method [*Dunker*, 1984; *Hakami et al*., 2003] within the CAMx model. The modeling domain covers 69 × 67 grids in the Eastern United States with a horizontal grid resolution of 36 km, encompassing nested finer domains of 12 (East Texas) and 4 km (DFW subdomain) spatial grid resolution (Figure 1). Back trajectory analysis shows southerly and easterly flow into the DFW region on a majority of episode days, with air flow into the DFW region coming overwhelmingly from within the 36 km model domain. The vertical configuration for the model domain consists of 28 layers of varying thickness, sufficient to examine the effect of vertical mixing within the typical planetary boundary layer height (for details, see Table 2-2 of *TCEQ*, 2011a, Appendix C).

[8] The CAMx model inputs (emissions, meteorological conditions, initial and boundary concentrations, chemical mechanism, and deposition scheme) were taken from the TCEQ's Base Case Modeling for the 8 hour O_3 State Implementation Plan (SIP) in DFW [*TCEQ*, 2011a]. The mobile emission (on-road and nonroad) inputs were obtained from the U.S. Environmental Protection Agency (EPA) MOBILE6.2 emission factor model, EPA's National Mobile Inventory Model (NMIM), and the Texas NONROAD (TexN) mobile source models, and were processed to a model-ready format by the Emissions Processing System version 3 (EPS3) [*ENVIRON*, 2007]. Base case biogenic emissions were derived from the Global Biosphere Emissions and Interactions System (GloBEIS3.13.1) model [*Yarwood et al*., 1999]. The Fifth Generation Meteorological Model (MM5 version 3.7.4) [*Dudhia*, 1993] was used to generate the meteorological inputs to CAMx including wind speed, wind direction, temperature, humidity, and so on (C. Emery et al., MM5 Q21 meteorological modeling of Texas for June 2006. Final report prepared for Texas Commission on Environmental Quality, unpublished report, 2009, http://www.tceq.texas.gov/assets/public/implementation/air/am/contracts/reports/mm/5820783986FY0802_June2006MM5_Final.pdf). [Details regarding the meteorological and emission modeling and their performance evaluations can be found in Appendixes A and B of *TCEQ*, 2011a]. Specifically, TCEQ found that benchmarks for error in wind direction (≤30°), wind speed (≤2 m/sec), and temperature (≤2 °C) were achieved 92%, 99%, and 92% of the time for the Dallas region. The base case model uses the Carbon Bond version 05 (CB05) chemical mechanism [*Yarwood et al*., 2005], a dry deposition scheme based on the works of *Wesely* [1989] and *Slinn and Slinn* [1980], and the global Model for Ozone and Related Chemical Tracers (MOZART) to generate episode-specific boundary condition concentrations for the coarse-grid (36 km) modeling domain [*ENVIRON*, 2008].

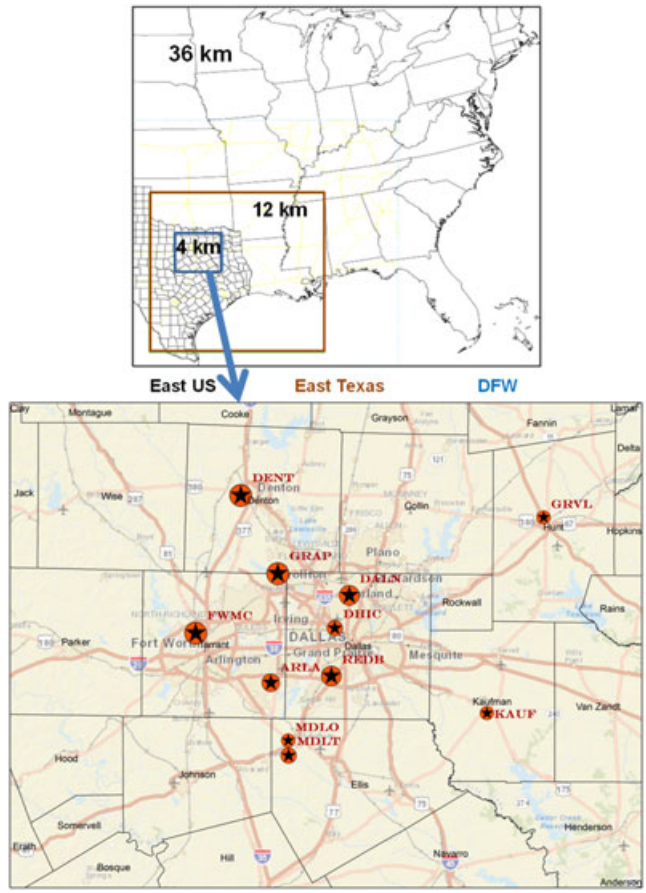

Figure 1. (top) Modeling domain used in *TCEQ* [2011a]. (bottom) Map showing the locations of the monitoring sites used in this study. The sizes of the circles are proportional to their 2006 eight hour O_3 design values.

2.2. Model Uncertainty Scenarios

[9] This study jointly considers uncertainties in both model formulation (*structural uncertainties*) and in model input parameters (*parametric uncertainties*).

2.2.1. Structural Scenarios

[10] Past studies have shown that alternate chemistry and emission models strongly influence photochemical sensitivities [*Pinder et al*., 2009; *Fine et al*., 2003; *Russell and Dennis*, 2000; *Bergin et al.*, 1999]. Uncertainty in the representation of dry deposition can also impact O_3 modeling [*Mallet and Sportisse*, 2006]. TCEQ had previously funded work that showed that boundary conditions from alternate global models could significantly influence background O_3 concentrations, but had not explored the impact on O_3 sensitivities [*ENVIRON*, 2009]. Therefore, structural scenarios were constructed by choosing either the Base Case setting explained earlier (section 2.1) or the alternate setting described later for each of four features: chemical mechanism, biogenic emissions model, dry deposition scheme, and boundary conditions model. Although additional structural uncertainties, most notably in the meteorological model, also influence O_3 sensitivities, alternate meteorological simulations and other inputs were not available within the scope of this study.

2.2.1.1. Alternate Chemical Mechanism

[11] In this setting, the 2005 version of the Carbon Bond chemical mechanism (CB05) in the base model is replaced by the sixth version (CB6) [*Yarwood et al*., 2010]. In CB6, several long-lived, abundant organic compounds, namely, propane, acetone, benzene, and ethyne (acetylene), are added explicitly to improve oxidant formation from these compounds as they are oxidized slowly at the regional scale. Compared to the CB05 mechanism, CB6 increases the number of model species (from 51 to 76) and the number of reactions (from 156 to 218). We adjust the rate constant for the reaction ($OH + NO_2$) in CB6 to be consistent with the most recent findings of *Mollner et al*. [2010] (CB6 also includes several updates for organic and inorganic aerosol chemistry). Detailed discussion of the differences between CB05 and CB6 is provided by (D. S. Cohan et al., Factors influencing ozone-precursor response in Texas attainment modeling, final report, Texas Air Quality Research Program, Project 10-008, 2011, unpublished report, http://aqrp.ceer.utexas.edu/projectinfo%5C10-008%5C10-008%20Final%20Report.pdf.)

2.2.1.2. Alternate Biogenic Emissions

[12] The GloBEIS-derived biogenics inventory is replaced by alternate biogenic emissions (BIO) from the Model of Emissions of Gases and Aerosols from Nature (MEGAN) [*Guenther et al*., 2006], which uses updated land cover data based on satellite and ground observations. *Guenther et al*. [2006] reports that the global annual isoprene emission, as estimated by MEGAN, approximately ranges from 500 to 750 Tg. Strong differences (about a factor of 2) between biogenic emission estimates from BEIS and MEGAN have been documented by *Carlton and Baker* [2011]. For the 12 km CAMx modeling domain overall, MEGAN estimated 47% lower biogenic NO_x emissions (ENO_x) and 24% higher biogenic nonmethane volatile organic compound (NMVOC) emissions than GloBEIS (for detailed differences, see *Cohan et al*., unpublished report, 2011). Within the DFW region, both models estimated only about 10 tpd of biogenic ENO_x, but MEGAN estimated twice as much biogenic NMVOC as GloBEIS.

2.2.1.3. Alternate Dry Deposition Scheme

[13] The original base case that uses land-use inputs and a dry deposition scheme (DEP) based on the work of *Wesely* [1989] and *Slinn and Slinn* [1980] is replaced here by an updated approach [*Zhang et al*., 2001, 2003]. The Zhang scheme incorporates vegetation density effects via leaf area index, possesses an updated representation of nonstomatal deposition pathways, has more land-use categories, and has been tested extensively through its use in daily air-quality forecasting.

2.2.1.4. Alternate Boundary Conditions

[14] Here, the MOZART boundary conditions used in the base case model are replaced by alternate boundary conditions (BC) from the GEOS-Chem global model [*Bey et al*., 2001] that exhibit higher O_3 concentrations (0.7–8 ppb) than MOZART at all model layers and the differences increase aloft (Figure 2) [*Cohan et al*., unpublished report, 2011].

2.2.2. Parametric Uncertainties

[15] For *parametric* uncertainties, we target the model input parameters identified by *Digar and Cohan* [2010] as most likely to influence model predictions of O_3 concentrations and their sensitivities to NO_x and volatile organic compound (VOC) emission controls. These parameters include specific emission rates, reaction rate constants, and boundary conditions (Table 1). The uncertainty estimates were derived from an extensive literature review of experimental and model-based studies. In particular, for the uncertain reaction rate of $NO_2 + OH$, we used a factor of uncertainty given by National Aeronautics and Space Administration's (NASA's) Jet Propulsion Laboratory (JPL) report [*Sander et al*., 2006] that compiles findings from multiple laboratory-based experiments. For the rest of the input parameters, we applied uncertainties based on the findings by *Deguillaume et al*. [2007] and *Hanna et al*. [2001] that compute uncertainty based on Monte Carlo simulations of air-quality models at different ranges of input perturbations, predecided (a priori) based on expert elicitation. *Deguillaume et al*. [2007, 2008] also used Bayesian analysis to constrain the a priori input distribution by comparing the Monte Carlo outputs with actual measurements.

[16] Sections 4.1.1 and 4.1.2 describe additional screening that was conducted to further narrow the structural cases and input parameters that most influence O_3 concentrations and sensitivities for the episode considered here.

3. Ground-Level Measurements of Ozone and Its Precursors

[17] Measurement data were obtained from the U.S. Environmental Protection Agency's (EPA's) Air Quality System (AQS) database for ground-level concentrations of O_3 and NO_x. These monitors record hourly concentrations of ambient air pollutants through a nationwide monitoring network (http://www.epa.gov/ttn/airs/airsaqs/index.htm); the monitors in Texas are operated by TCEQ. The raw data were then postprocessed to obtain daily maximum 8 hour O_3 and 24 hour average NO_x concentrations at all the monitors that fall within the nine-county DFW nonattainment area (based on 1997 8-hour O_3 NAAQS): Denton, Collin, Parker, Tarrant, Dallas, Rockwall, Kaufman, Johnson, and Ellis Counties. We considered 11 monitors that measure both O_3 and NO_x concentrations (Figure 1).

[18] Measurements of O_3 are conducted by well-established techniques; thus, instrumental error is relatively small [*EPA*,

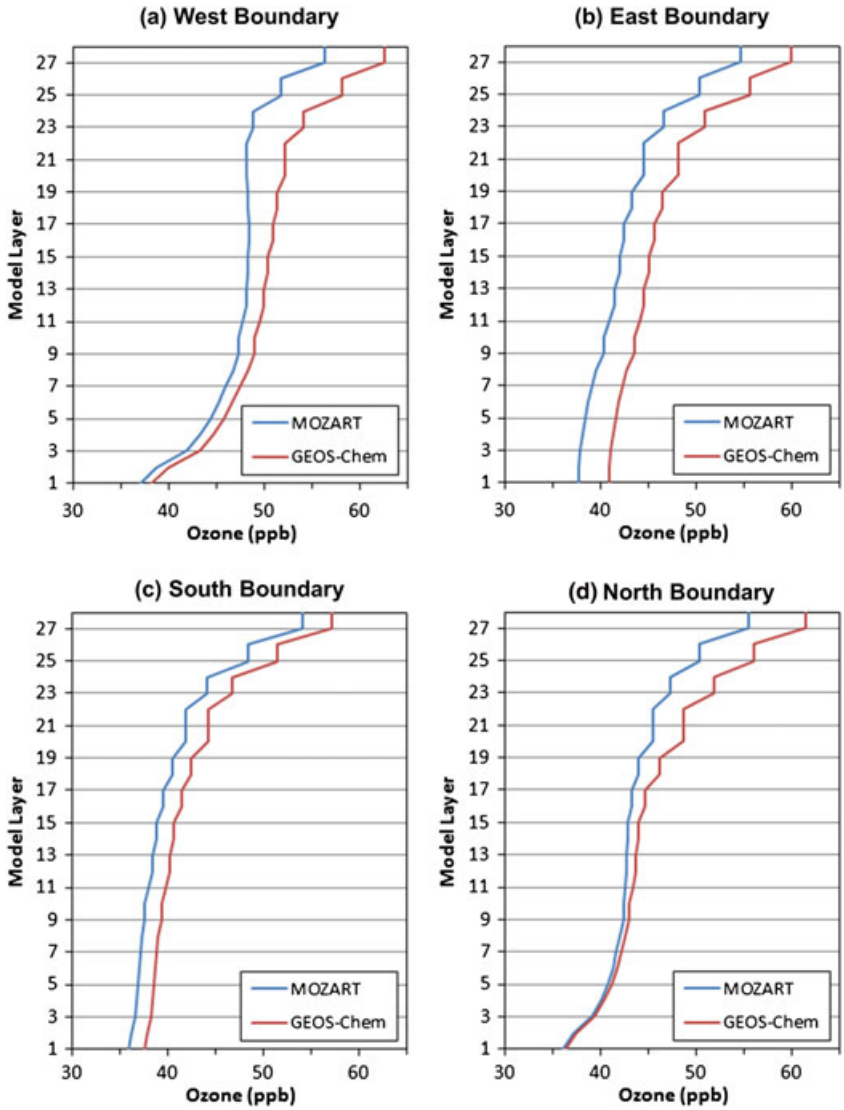

Figure 2. Episode-average ozone boundary conditions (ppb) by model layer estimated from the MOZART and GEOS-CHEM global model simulations.

$$\beta = \frac{[NO] + [NO_2]}{[NO] + [NO_2] + 0.95[PAN] + 0.35[HNO_3] + [N_2O_5] + [PNA] + [HONO] + [NO_3]} \quad (1)$$

2006]. However, due to lack of direct measurement technique for nitrogen dioxide (NO_2), NO_x measurements tend to have significant instrumental bias and monitor interferences [*Demerjian*, 2000; *Dunlea et al.*, 2007]. NO_x concentrations are, therefore, bias-corrected for interference with other nitrogen species. We apply a bias-correction factor (β) adapted from *Lamsal et al.* [2008] computed using modeled species concentrations to correct reported NO_x observations:

where PAN is peroxy acetyl nitrate and PNA is peroxy nitric acid. The factor β was computed for each monitor based on the episode average of the daily 24 hour mean concentration of the modeled NO_y (= NO_x +HNO_3 + PAN +HONO+N_2O_5) species.

4. Method

4.1. Model Uncertainty Analysis

[19] This section details the methodology adopted for incorporating *structural* and *parametric* uncertainties in the photochemical air-quality modeling.

Table 1. Screening Test for the Selection of Uncertain Input Parameters

Parameter[a]	Uncertainty[b] (1σ)	Reference	Impact[c] on O_3 Concentration	Impact[c] on O_3 Sensitivity to ANO_x	Impact[c] on O_3 Sensitivity to AVOC
Emission Rates		*Deguillaume et al.*, 2007			
Domain-wide NO_x	0.336		0.105	−0.463	0.496
Domain-wide biogenic VOC	0.405		0.026	0.216	−0.319
Domain-wide AVOC	0.336		0.006	0.073	−0.150
Reaction Rate Constants					
All photolysis frequencies	0.347	*Hanna et al.*, 2001	0.091	0.401	0.091
$R(OH+NO_2)$	0.131	*Sander et al.*, 2006	−0.017	−0.057	0.029
$R(NO+O_3)$	0.095	*Hanna et al.*, 2001	−0.023	−0.058	−0.024
R(all VOCs + OH)	0.095	*Deguillaume et al.*, 2007	0.003	0.021	0.014
Boundary Conditions		*Deguillaume et al.*, 2007			
$BC(O_3)$	0.203		0.036	0.006	−0.042
$BC(NO_x)$	0.549		0.002	−0.001	−0.001
$BC(HNO_3)$	0.549		0.001	−0.000	−0.000
BC(PAN)	0.549		0.008	−0.003	−0.002
BC(HONO)	0.549		0.000	−0.000	−0.000
$BC(N_2O_5)$	0.549		0.000	−0.000	0.000

[a]Parameters selected based on the impact analysis by *Digar and Cohan* [2010] and *Digar et al.* [2011].
[b]All distributions are assumed lognormal.
[c]*Impact factor*: The fractional change in concentrations and first-order sensitivity of ozone to emissions, due to a 1σ change in an input parameter as detailed in section 4.1.2. Uncertainty factors are based on ±2σ (i.e., 95%) confidence interval. Underlined terms were chosen for the parametric Monte Carlo sampling.
ANO_x, anthropogenic NO_x; AVOC, anthropogenic volatile organic compound.

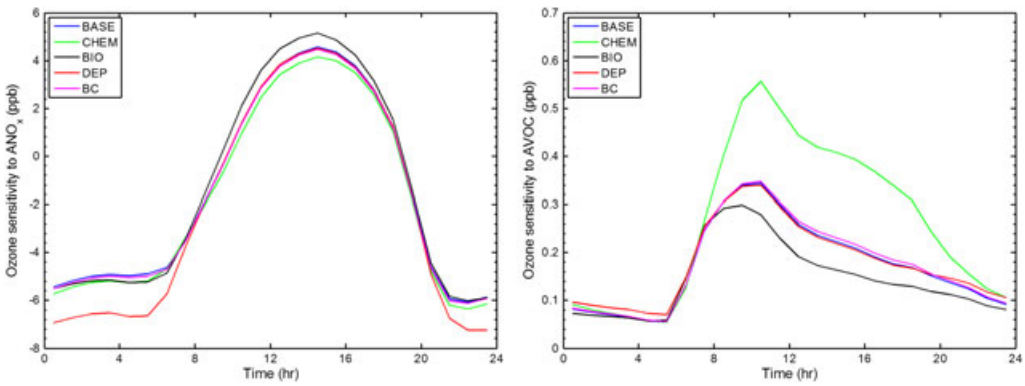

Figure 3. Diurnal profile of ozone sensitivities to DFW ANO_x (left) and AVOC (right) emissions, averaged over the episode and the grid cells covering the DFW region.

4.1.1. Screening for Structural Uncertainty

[20] To assess the effect of model *structural uncertainty*, we first run the photochemical model with the base-case scenario (BASE) and then with each of the alternate assumptions of atmospheric processes detailed in section 2.2.1., which include alternate chemical mechanism (CHEM), biogenic emission inventory (BIO), dry deposition scheme (DEP), and boundary conditions (BC). Figure 3 shows how the diurnal patterns of DFW O_3 sensitivities to DFW anthropogenic emissions change with each of these different model assumptions. Here, "sensitivity" denotes the seminormalized local first-order sensitivity coefficient $S_j^{(1)}\left(=\frac{\partial C}{\partial \varphi_j}\right)$ that measures the responsiveness of concentrations (C) to fractional perturbation $\varphi_j\ \left(=\frac{p_j}{P_j}\right)$ in precursor j, where P_j is the unperturbed input value (base), P_j is new perturbed parameter value, and φ_j is a scaling variable with a nominal value of 1 [*Cohan et al.*, 2005]. The unit of sensitivity is, therefore, the same as that of concentrations. Afternoon O_3 in DFW is primarily NO_x limited in all of the structural cases, with O_3 about an order of magnitude more sensitive to DFW anthropogenic NO_x (ANO_x) than anthropogenic VOC (AVOC). In general, use of MEGAN biogenic emission increases O_3 sensitivities to ANO_x (S_{NO_x}) and decreases sensitivity to AVOC (S_{VOC}) relative to the base case during daytime because of its stronger biogenic VOC emissions (EBVOC). The alternate CB6 chemical mechanism also affected daytime O_3 sensitivities but in the opposite direction, yielding stronger sensitivities to AVOC, though conditions remain predominantly NO_x sensitive under either structural configuration. The alternate BC case did not significantly affect O_3 sensitivities, and DEP affected sensitivities mostly during night.

[21] To select the most important structural factors that influence predictions of O_3 concentrations, we compare each structural scenario against the observations. For screening

the factors that most strongly affect O_3 sensitivities, we compare each alternate scenario against the base-case simulation results. The statistical measures that serve as the bases for the comparisons are as follows:

$$\text{Root Mean Square Error (ppb)} : \text{RMSE} = \sqrt{\frac{\sum_{j=1}^{N}(Y_j - O_j)^2}{N}} \quad (2)$$

$$\text{Mean Bias (ppb)} : \text{BIAS} = \frac{1}{N}\sum_{j=1}^{N}(Y_j - O_j) \quad (3)$$

$$\text{Normalized Mean Bias (\%)} : \text{NMB} = \frac{\sum_{j=1}^{N}(Y_j - O_j)}{\sum_{j=1}^{N} O_j} \cdot 100\% \quad (4)$$

$$\text{Normalized Mean Bias (\%)} : \text{NMB} = \frac{\sum_{j=1}^{N}|Y_j - O_j|}{\sum_{j=1}^{N} O_j} \cdot 100\% \quad (5)$$

where N is the number of observations (site/days), and Y_j denotes concentrations or sensitivities from each of the model structural cases considered earlier. For comparisons of concentrations, O_j represents the observations; for comparisons of sensitivities, O_j represents the base-case simulation results.

[22] The comparison results (Table 2) show that alternate chemical mechanism (CB6 vs. CB05) and biogenic model (MEGAN vs. GloBEIS) most strongly influence the predicted O_3 concentrations and sensitivities. Therefore, we build an ensemble of models with the following *structural*

Table 2. Screening Test for the Selection of Uncertain Model Structural Assumptions[a]

	Structural Cases				
	Base	CHEM	BIO	DEP	BC
Comparison of each structural case against the observations for 8 hour O_3 concentration in DFW					
RMS (ppb)	13.01	13.21	13.63	12.95	13.01
BIAS (ppb)	−0.61	4.59	−1.06	1.88	0.02
NMB (%)	−1.04	7.83	−1.82	3.22	0.04
NME (%)	17.79	16.88	18.85	17.08	17.76
Comparison of each alternate case against the base case for 8 hour DFW O_3 sensitivity to DFW ANO_x					
RMS (ppb)	—	0.79	1.37	0.12	0.16
BIAS (ppb)	—	−0.40	0.75	−0.01	−0.09
NMB (%)	—	−12.07	22.81	−0.25	−2.81
NME (%)	—	15.35	26.08	2.13	2.83
Comparison of each alternate case against the base case for 8-hour DFW O_3 sensitivity to DFW AVOC					
RMS (ppb)	—	0.44	0.17	0.02	0.02
BIAS (ppb)	—	0.26	−0.08	−0.00	0.01
NMB (%)	—	63.35	−19.33	−0.80	1.88
NME (%)	—	63.45	21.90	2.04	2.14

[a]ANO_x, anthropogenic NO_x; AVOC, anthropogenic volatile organic compound; BC, boundary condition; BIO, alternate biogenic emissions; CHEM, alternate chemical mechanism; DEP, dry deposition scheme; DFW, Dallas-Fort Worth; NMB, normalized mean bias; NME, normalized mean error; RMS, root mean square.

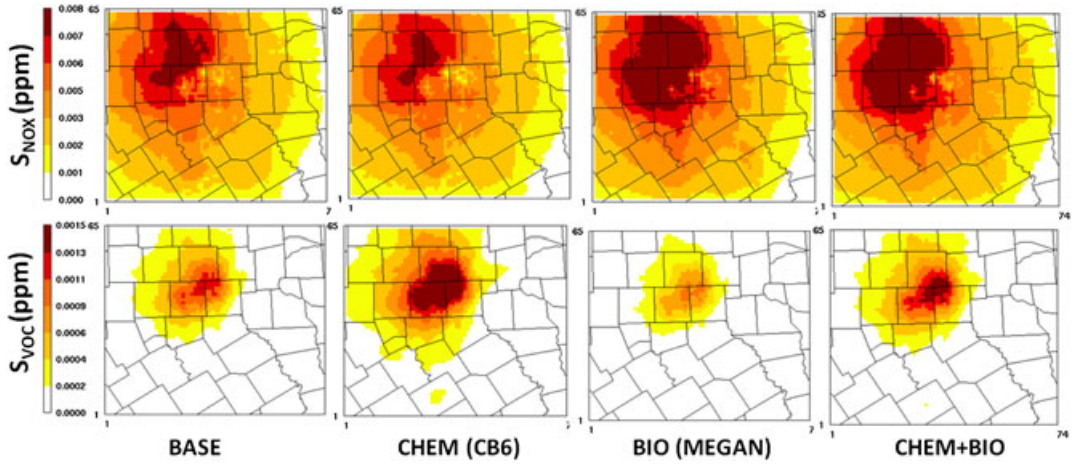

Figure 4. Sensitivity of 8 hour O_3 to anthropogenic NO_x and VOC emissions from DFW for different structural model scenarios under default settings of input parameters. Episode-average results are shown for the 4 km resolution domain.

members: (1) BASE, (2) CHEM, (3) BIO, and (4) a combination of alternate chemical mechanism (CB6) and biogenics (MEGAN) (hereafter abbreviated as CHEM + BIO). Figure 4 shows the spatial plots for O_3 sensitivities to each of these four structural members. NO_x-limited conditions for daily maximum 8 hour O_3 persist even in the urban center regardless of which structural scenario is considered.

4.1.2. Screening for Parametric Uncertainty

[23] Uncertainties in input parameters (*parametric uncertainties*) are characterized by Monte Carlo analysis, where values of input parameters are selected randomly from the probability distribution assumed for each input based on their standard deviations. For computational efficiency, we use an RFM to compute adjusted concentrations (C^*) and sensitivities ($S_j^{(1)}$) based on the uncertainties in input parameters using the relationships given by *Cohan et al.* [2005] and *Digar and Cohan* [2010],

$$C^* = C_0 + \sum_j \varphi_j S_j^{(1)} + \sum_k \varphi_k S_k^{(1)} + \frac{1}{2}\sum_j \varphi_j^2 S_{j,j}^{(2)} + \frac{1}{2}\sum_k \varphi_k^2 S_{k,k}^{(2)} + \sum_{j,k} \varphi_j \varphi_k S_{j,k}^{(2)} \quad (6)$$

$$S_j^{(1)*} = \left(1 + \varphi_j\right)\left(S_j^{(1)} + \varphi_j S_j^{(1)} + \sum_k \varphi_k S_{j,k}^{(2)}\right) \quad (7)$$

where C_0 is the concentration modeled under default setting of the parameters, and φ_j and φ_k are the perturbations in parameters j and k, respectively. $S_j^{(1)} = \frac{\partial C}{\partial \varphi_j}$ and $S_j^{(2)} = \frac{\partial^2 C}{\partial \varphi_j^2}$ denote seminormalized first- and second-order sensitivities of concentrations to the parameter j. $S_{j,k}^{(2)} = \frac{\partial^2 C}{\partial \varphi_j \partial \varphi_k}$ denotes cross-sensitivity between two input parameters j and k. In the RFMs, the value of each φ is restricted to within a 2-sigma range for that parameter to avoid extreme values of input parameters.

[24] As discussed in section 2.2, we use a suite of uncertain model input parameters listed in Table 1. Each parameter was assumed to have a lognormal probability distribution, characterized by the uncertainty value (1σ) reported in Table 1. To screen parameters that strongly influence O_3 concentrations and sensitivity to emissions, we perform an impact analysis where relevant "impact factors" were evaluated as follows:

$$\textit{Impact factor } (IF) \textit{ for the influence of parameter } j \textit{ on concentration } C : IF_C = \frac{\varphi_j S_j^{(1)}}{C} \quad (8)$$

$$\textit{Impact factor } (IF) \textit{ for the influence of parameter } k \textit{ on sensitivity } S_j^{(1)} ; IF_S = \frac{\varphi_k S_{j,k}^{(2)}}{S_j^{(1)}} \quad (9)$$

[25] Although there was considerable overlap in the selected parameters, there were also some differences in those found to have a greater impact on concentrations and the two sensitivities (Table 1). Domain-wide ENO_x and EBVOC, photolysis rates (hν), and the reaction rate constants $R(NO_2 + OH)$ and $R(NO + O_3)$ significantly impacted all three categories. Meanwhile, boundary conditions (BC) of NO_y were not major influences on any of the results. However, the $BC(O_3)$ parameter significantly impacted concentrations and sensitivity to VOC, but not to NO_x, whereas anthropogenic VOC emissions (EAVOC) impacted sensitivities, but not concentrations.

4.1.3. Joint Consideration of Structural and Parametric Uncertainty

[26] We construct an ensemble consisting of the four targeted structural members based on the screening test in Table 2 (BASE, CHEM, BIO, and CHEM + BIO), each coupled with 1000 Monte Carlo samplings from the probability distributions for the selected model input parameters underlined in Table 1. Total sample size of the final ensemble was, therefore, 4000. The final set of *parametric* factors considered in this study is summarized as follows:

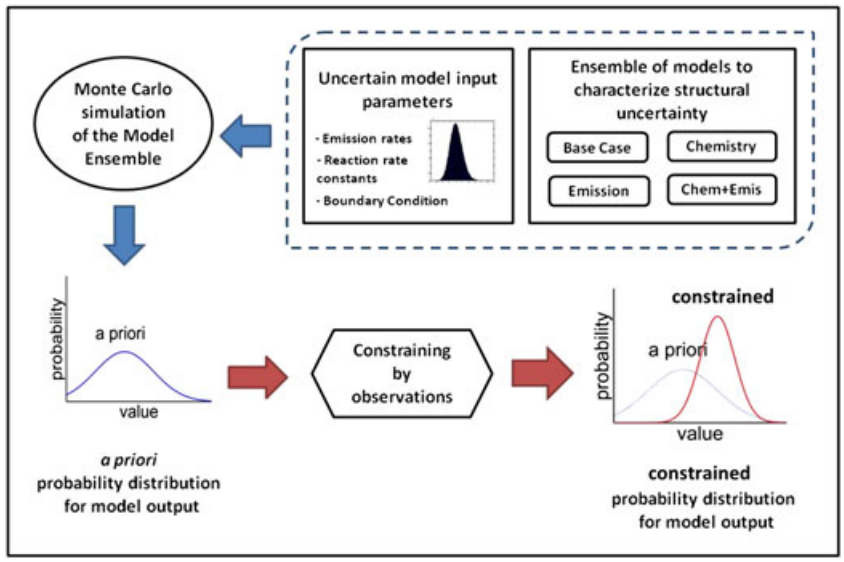

Figure 5. Flowchart for the observation-constrained Monte Carlo analysis.

For O_3 concentration: ENO_x, EBVOCs, photolysis frequencies, $R(NO_2 + OH)$, $R(NO + O_3)$, and $BC(O_3)$

For O_3 sensitivity to ANO_x emissions: ENO_x, EBVOCs, EAVOC, photolysis frequencies, $R(NO_2 + OH)$, $R(NO + O_3)$, and R(all VOCs + OH)

For O_3 sensitivity to EAVOC: ENO_x, EBVOCs, EAVOC, photolysis frequencies, $R(NO_2 + OH)$, $R(NO + O_3)$, R (all VOCs + OH), and $BC(O_3)$.

4.2. Constraining Model Predictions Using Measurements

[27] A key limitation of the traditional Monte Carlo analysis of the model ensemble [e.g., *Pinder et al*., 2009; *Digar et al*., 2011] is the assumption that each of the cases is equally likely. This study uses actual observations to prioritize cases that show good agreement with measured concentrations over those that do not perform well. Figure 5 shows the framework of the observation-constrained Monte Carlo analysis. Concentration estimates from each of the 4000 simulations are compared with actual measurements at the monitors to evaluate the adjusted (*observation-constrained*) probability distribution of the ensemble. Various techniques are used to weight (Bayesian) or screen (model performance and hypothesis testing) the best-performing model cases to characterize adjusted probability distributions of pollutant concentrations and sensitivities. The methods and observation metrics used in this study are elaborated below.

4.2.1. Metric 1: Bayesian Analysis

[28] A Bayesian inference approach [*Bergin and Milford*, 2000; *Deguillaume et al*., 2007] is applied to assign relative weightings to each case based on its performance in simulating observed O_3 and NO_x. For evaluating the likelihood of model prediction ($C^*_{m,n}$) for the m^{th} simulation of the n^{th} observation ($n = 1, 2, \ldots, N$, where N denotes total number of observations), a Gaussian likelihood function is used (as defined by *Bergin and Milford*, 2000). Errors (θ) in the observed O_3 and NO_x concentrations are assumed to be independent and normally distributed with mean zero. The likelihood of model prediction $C^*_{m,n}$ given observation O_n can be expressed as

$$L\left(C^*_{m,n}|O_n\right) = \frac{1}{\theta\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left[\frac{\left(O_n - C^*_{m,n}\right)^2}{\theta^2}\right]\right\} \quad (10)$$

[29] The total likelihood for simulation m given all observations of a species can then be computed by the product of its likelihoods for individual observations. Errors in the observed O_3 and NO_x concentrations are assumed to be independent; thus, $L\left(C^*_m|O\right) = \prod_{n=1}^{N} L(C^*_{m,n}|O_n)$. $L\left(C^*_m|O\right)$ is computed separately for O_3 and NO_x, then multiplied together to get the overall likelihood based on both species. Finally, Bayes theorem is applied to compute the a posteriori probability distribution (p') based on the a priori probabilities $\left(p\left(C^*_m\right) = 1/M\right)$ and the likelihoods computed above.

$$p'\left(C^*_m|O\right) = \frac{L\left(C^*_m|O\right)p\left(C^*_m\right)}{\sum_{m=1}^{M} L\left(C^*_m|O\right)p\left(C^*_m\right)} = \frac{L\left(C^*_m|O\right)}{\sum_{m=1}^{M} L\left(C^*_m|O\right)} \quad (11)$$

[30] The mean (μ') and standard deviation (σ') of the resulting posterior ensemble distribution can be computed by

$$\mu' = \sum_{j=1}^{M}\left(Y_j p'_j\right) \quad (12)$$

$$\sigma' = \sqrt{\sum_{j=1}^{M}\left(Y_j - \mu'\right)^2 p'_j} \quad (13)$$

where Y_j denotes the j^{th} value of the simulation and p'_j denotes the respective posterior probability for that iteration [obtained from equation (10)], and M is the total size of the ensemble (= 4000).

[31] The observation metric chosen for the Bayesian analysis is highly aggregated, as were the metrics used by *Bergin and Milford* [2000] and *Deguillaume et al*. [2008]. Here, episode averages of the daily 8 hour O_3 and of the 24 hour NO_x concentrations at each of the 11 monitors were considered ($N = 11$). The consideration of episode-average concentrations on a site-by-site basis tests the ability of each model case to simulate overall levels and spatial patterns in O_3 and NO_x, even if errors in simulating meteorology or emissions variability may have obscured day-to-day comparisons. The errors in the observed episode averages are assumed to be independent across space because we do not expect calibration or representativeness errors to be spatially correlated at these temporal scales.

[32] Errors and uncertainty in applying measurement data to evaluate model results can arise from instrumental error and from the use of a point measurement to represent a model grid-cell average concentration. The resulting uncertainty can be quantified jointly by examining the variability between pollutant concentrations measured by multiple monitors within the same grid cell. Analysis using 5 years (centered on our base case model year 2006, i.e., 2004–2008) of data for the summer O_3 season (May to September) showed that the error (θ) characterizing the standard deviation of

Table 3. Statistics for Evaluating Model Performance in Metric 2[a]

Performance Statistics	Formula	Screening Criteria
MNGE	$\frac{1}{N}\sum_1^N\left(\frac{\lvert Model-Obs\rvert}{Obs}\right) * 100\%$	$-5\% < \text{MNGE} < +5\%$
MNB	$\frac{1}{N}\sum_1^N\left(\frac{Model-Obs}{Obs}\right) * 100\%$	$\text{MNB} < 30\%$
UPA[b]	$\frac{Model_{max}-Obs_{max}}{Obs_{max}} * 100\%$	$-15\% < \text{UPA} < +15\%$

[a]Mean normalized gross error (MNGE) and mean normalized bias (MNB) were computed for model results (Model) when O_3 observations (Obs) were greater than the recommended threshold of 60 ppb [*EPA*, 2006].
[b]UPA, unpaired peak accuracy.

differences between observed 8 hour O_3 values at three pairs of sites falling in the same grid location ranged from 3.0 to 7.2 ppb; for bias-corrected 24 hour NO_x observations, θ ranged from 2.2 to 8.2 ppb. Because these estimates are based on a limited number of site pairs, to be conservative, we choose the maximum values of these ranges (i.e., θ = 7.2 and 8.2 ppb for 8 hour O_3 and 24 hour NO_x, respectively).

4.2.2. Metric 2: Screening Based on Model Performance

[33] An alternate approach to developing *observation-constrained* distributions is to retain only cases that meet specified performance criteria [e.g., *Mallet and Sportisse*, 2006]. Because the base modeling used here was developed for an SIP attainment plan, we formulate a new metric [metric 2 (M2)] that screens the 4000 cases based on the three-model performance evaluation criteria recommended by the EPA [1999, 2007] for determining the acceptability of an O_3 SIP model (Table 3). This metric uses all available valid observations of daily 8 hour O_3 at each monitor (N = 289). Mean normalized bias (MNB) and mean normalized gross error (MNGE) were computed for model results (Model) when O_3 observations (Obs) were greater than the recommended threshold of 60 ppb [*EPA*, 2006]. The screened cases were assigned equal weights to develop the adjusted (*observation-constrained*) distribution.

4.2.3. Metric 3: Screening Based on Nonparametric Test

[34] Statistical nonparametric tests of significance like the Cramér-von Mises (CvM) criterion and the Kolmogorov-Smirnov test have been used to test for general differences in predicted and observed distributions of air-quality data [*Holland and Fitz-Simons*, 1982; *Taylor et al*., 1987]. The CvM criterion [*Anderson*, 1962] provides a nonparametric test of the null hypothesis (H_0) that two samples are drawn from the same (unspecified) distribution. In the CvM two-sample test, the test statistics T is computed as follows:

$$T = \frac{AB}{(A+B)^2}\left\{\sum_{i=1}^{A}[F_A(x_i) - G_B(x_i)]^2 + \sum_{j=1}^{B}[F_A(y_j) - G_B(y_j)]^2\right\} \quad (14a)$$

where $F_A(x)$ and $G_B(y)$ are the empirical distribution functions of the two samples $x = x_1, x_2, \ldots, x_A$ (representing model predictions) and $y = y_1, y_2, \ldots, y_B$ (representing observations) of size A and B, respectively. Note that $G_B(x_i)$ denotes the relative frequency that the observed concentration is at most x_i (i.e., sum of all the elements in the sample less than x_i, divided by the sample size B), and $F_A(y_j)$ denotes the relative frequency that the modeled concentration is at most y_j.

[35] The null hypothesis is rejected when T is large, indicating that the two samples are significantly different. The advantage of this method is that it assesses whether there are any differences in the modeled and observed probability distributions, not just differences in the means of the two samples (e.g., differences in the variance and/or the tail of the samples). Note that the CvM criterion does not pair observations in time and space, but instead indicates whether the distribution of model predictions is consistent with the distribution of observations. For our case, the two samples represent the modeled and observed distribution of pollutant concentrations, and the sample size for the two distributions are equal here (i.e., A = B = N, where N denotes total number of observations). Therefore, equation (14a) reduces to the form

$$T = \frac{1}{4}\left\{\sum_{i=1}^{N}[F_A(x_i) - G_B(x_i)]^2 + \sum_{j=1}^{N}[F_A(y_j) - G_B(y_{ji})]^2\right\} \quad (14b)$$

[36] The test statistic T is computed for each of the 4000 members of the model ensemble, separately for available 8 hour O_3 (N = 289) and 24 hour NO_x (N = 303) concentrations using equation (14b). Next, we compute the p value associated with each test statistic (T), defined as the probability of observing a test statistic greater than or equal to T, if H_0 is true. A small T will result in a large p value, indicating that there is not sufficient evidence to reject the null hypothesis (H_0). Screening is then applied to select Monte Carlo cases that generate p values greater than the 10% significance level, i.e., α = 0.1, below which we reject the null hypothesis. We select only those cases that satisfy this test for both of the observational constraints (O_3 and NO_x).

4.3. Adjusted Ozone Sensitivity

[37] To characterize adjusted O_3 response to emission changes, we use the RFM given in equation (7) to generate the a priori (equal-weighted) probability density of O_3 sensitivity to any emission j for each of the 4 structural cases based on the 1000 samplings of input parameters k. Because pollutant sensitivities cannot be directly evaluated, the *observation-constrained* O_3 sensitivities for the full ensemble (all 4000 cases) are estimated based on the model's performance in reproducing observed concentrations. Therefore, for metric 1 (M1), we assume that the a posteriori probabilities estimated for O_3 concentrations by equation (11) can also be

Table 4. Observation-Constrained Probability of the Structural Ensemble Members[a]

Metrics	Constrained by Measurements of	Base	CHEM	BIO	CHEM + BIO
Metric 1 (Bayesian)	O_3 and NO_x	19.37%	35.37%	16.14%	29.12%
	O_3	25.68%	29.55%	21.66%	23.10%
	NO_x	25.44%	24.73%	25.17%	24.66%
Metric 2 (EPA)	O_3	16.14%	33.69%	17.99%	32.19%
	O_3 and NO_x	12.92%	37.08%	16.97%	33.03%
Metric 3 (CvM)	O_3	15.65%	34.60%	18.84%	30.91%
	NO_x	25.32%	25.32%	24.70%	24.66%

[a]BIO, alternate biogenic emissions; CHEM, alternate chemical mechanism; CvM, Cramér-von Mises.

applied to adjust the a priori probability distribution of O_3 sensitivities; for M2 and metric 3(M3), we assign equal probability to sensitivities from each simulation that passed the respective screening test, and zero to the remaining cases.

5. Results and Discussion

[38] In this section, results for input parameter values, O_3 concentrations, and sensitivities to emissions are presented to show how the adjusted (*observation-constrained*) probability distributions generated by application of the three observational metrics differ from the a priori (equal-weighted) distribution. The evaluation of the quality of the final three adjusted model ensembles has been discussed elsewhere [*Digar*, 2012].

[39] Application of M1 (Bayesian weightings) to our ensemble of 4000 simulations assigns half of the total weight to the 496 best-performing model simulations. Most of the spread in weightings results from evaluation against O_3 observations rather than against NO_x observations; however, the multiplication of weightings by equation (10) leads the joint weightings to differ substantially from those that would have resulted from considering O_3 alone (Table 4).

[40] M2 screened 1134 cases that satisfied all three of the EPA's recommended model performance criteria detailed in Table 3. This selection was mainly restricted by the bias term (MNB), which was satisfied by 1137 cases. The other two criteria, namely, the unpaired peak accuracy and MNGE, selected nearly all of the 4000 cases, rejecting only 15% and 1% of cases, respectively. M3, which selects cases based on the CvM two-sample test, selects 766 model cases that satisfy the test for both O_3 and NO_x observations. Screening based on O_3 or NO_x observations alone would have selected 1003 and 2457 cases, respectively.

Table 5. Performance of the Base Model and Observation-Constrained Model Ensemble Means against Observed 8 Hour O_3 at All Sites and Days in Dallas-Fort Worth[a]

Statistics	Base Model	Metric 1	Metric 2	Metric 3
NMB (%)	−6.08	−0.690	4.52	1.03
NME (%)	17.74	15.76	15.84	15.73
Correlation	0.704	0.720	0.716	0.714
R^2	0.495	0.518	0.513	0.510

[a]NMB, normalized mean bias; NME, normalized mean error.

[41] Accuracy of the ensemble-mean prediction is tested by evaluating the normalized mean bias, the normalized mean error (NME), the correlation, and the regression coefficients of the ensemble mean with 8 hour O_3 observations for all sites and days ($N = 289$) in the DFW region (Table 5). As expected, the model performance improves when the ensemble is constrained based on the observations. All the observational metrics help to minimize the model bias and error, and to some extent increase the overall correlation and regression (Table 5). The base-case model underpredicts O_3 concentrations by 6%. The non-Bayesian metrics, on the other hand, tend to slightly overpredict O_3 (M2 by 4.5% and M3 by 1%), although they reduce the overall error by 11%.

[42] To further evaluate the performance of the ensemble in simulating episode-average conditions (similar to the scenario used in M1) at a given location, results for observation-constrained O_3 concentrations are probed for the 11 DFW monitors (Table 6). In an effort to correct the overall underprediction of the base model, the ensemble weightings increased the mean O_3 concentration at all the sites. As a result, the posterior adjustments significantly improved the prediction accuracy for monitors that had greater negative bias in the base-case modeling scenario (Table 6).

[43] Detailed comparisons are illustrated for the Denton monitor (DENT), which recorded the highest 8 hour O_3 design values among all the DFW sites in 2006. Figure 6a shows the probability density functions (PDFs) of episode-average O_3 concentrations at Denton. The blue curve in Figure 6a depicts the a priori (equal-weighted) probability density. The other solid curves show the final *observation-constrained* distributions resulting from joint consideration of the full 4000 case ensemble under the 3 observational metrics. The deterministic model (BASE) underpredicts (62.0 ppb) the episode-average daily 8 hour O_3 observation of 70.1 ppb at Denton during the study period. The a priori equal-weighted ensemble predicts a mean concentration of 65.5 ppb with a standard deviation of 7.3 ppb (Table 6). Application of each of the 3 metrics narrowed the spread of the ensemble predictions, as can be seen by the curves in Figure 6a and the smaller standard deviations (~2 ppb) in Table 6, indicating greater confidence in the ensemble. M2 and M3 yielded ensemble-mean predictions of episode-average O_3 (69.04 and 68.85 ppb, respectively) that more closely matched observations at Denton. Testing showed that withholding a monitor from the observations used to constrain

Table 6. Comparison of Observed and Modeled Episode-Average 8 Hour Ozone Concentrations (ppb) at Dallas-Fort Worth Sites[a]

	2004–2006 O_3 Design Value	Observed O_3	Base Model O_3	Prior Ensemble ($\mu \pm \sigma$)	Observation-Constrained Ensemble ($\mu \pm \sigma$)		
					Metric 1	Metric 2	Metric 3
DENT	93.33	70.11	61.98	65.51 ± 7.33	65.53 ± 2.16	69.04 ± 2.03	68.85 ± 1.87
GRAP	90.67	68.79	63.20	66.62 ± 7.21	66.76 ± 2.21	70.02 ± 2.17	69.98 ± 2.03
FWMC	89.33	59.54	62.42	66.10 ± 7.17	66.23 ± 2.22	69.50 ± 2.17	69.43 ± 2.01
DALN	85.00	62.34	60.52	64.09 ± 7.00	64.25 ± 2.11	67.42 ± 1.99	67.36 ± 1.84
REDB	85.00	64.91	59.26	62.73 ± 7.07	62.75 ± 2.09	66.13 ± 1.91	65.94 ± 1.74
ARLA	83.33	65.51	61.63	65.12 ± 7.27	65.12 ± 2.17	68.61 ± 2.01	68.43 ± 1.86
DHIC	81.67	61.70	59.53	63.08 ± 6.75	63.33 ± 2.09	66.27 ± 2.04	66.26 ± 1.88
MDLT	80.50	62.03	56.57	59.77 ± 6.76	59.77 ± 2.00	63.01 ± 1.91	62.77 ± 1.75
MDLO	75.00	57.68	55.96	59.23 ± 6.50	59.35 ± 1.94	62.32 ± 1.82	62.19 ± 1.71
GRVL	75.00	61.02	53.96	57.04 ± 6.70	56.96 ± 2.10	60.29 ± 2.11	59.96 ± 1.79
KAUF	74.67	58.04	55.22	58.31 ± 6.88	58.20 ± 2.14	61.64 ± 2.14	61.29 ± 1.81

[a]μ and σ denote mean and standard deviation, respectively.

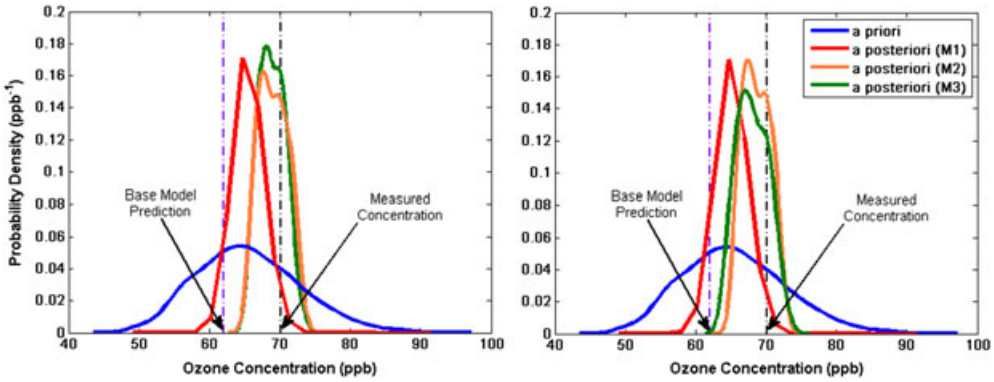

Figure 6. PDFs for episode-average 8 hour ozone concentration at Denton (a) when observations from Denton were used to constrain the a priori results and (b) when observations from Denton were withheld.

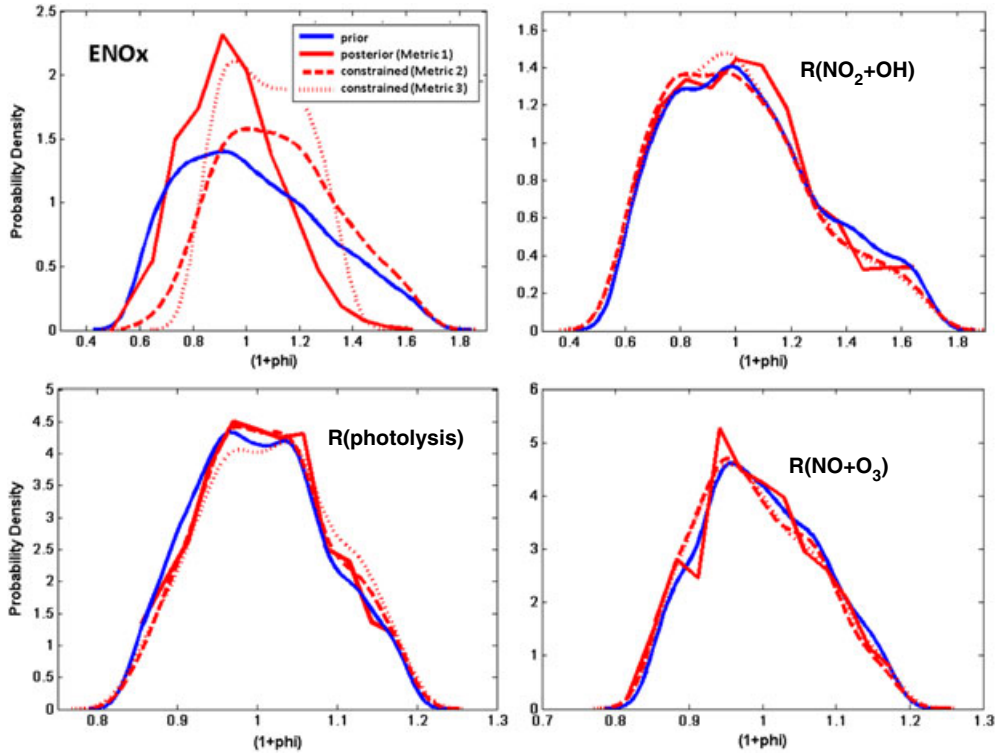

Figure 7. Prior and constrained distribution of scaling factors (1 + phi) for selected model input parameters.

the ensembles did not substantially alter the posterior results at that monitor (Figure 6b).

[44] Although each of these metrics uses different criteria and methods for comparing pollutant concentrations, they each yield similar allocation of adjusted probabilities among the structural scenarios (Table 4). For the study region and episode, application of each metric tends to prioritize model cases that use the CB6 chemical mechanism. For example, under M1, 384 of the 496 highest-weighted cases used CB6, lending to 64% of overall weight being placed on the CHEM and CHEM + BIO scenarios (Table 4). The CHEM and CHEM + BIO scenarios were also favored relative to their CB05 counterparts by M2 and M3. The metrics do not show a consistent preference between the MEGAN and GloBEIS biogenic inventories.

[45] Application of the three metrics also generated *observation-constrained* probability distributions for the scaling factors (1+φ) for the model input parameters listed in Table 1. Figure 7 shows the PDFs for some of the key parameters. The a priori PDFs are derived from the 1000 Monte Carlo cases randomly sampled from the truncated lognormal probability distributions assumed for each input parameter, and the adjusted PDFs are generated by applying the same weightings (M1) and screenings (M2 and M3) used for constraining O_3 concentrations. No significant differences were observed in the a priori and *observation-constrained*

distributions of model input parameters, except for ENO_x. Adjustment under M2 and M3 preferred slightly higher levels of NO_x emission, as indicated by the positive shifts in the adjusted PDFs in Figure 7a. M1 favored ENO_x levels close to the original values.

[46] We now examine how the relative sensitivities of O_3 to DFW ANO_x (S_{NO_X}) and to DFW AVOC (S_{VOC}) change when the observational constraints are considered. Although results have been presented here for Denton, similar trends were observed for the other sites in DFW as well. The base-case model (without incorporating uncertainties) predicts that at Denton, S_{NO_X} is 6.56 ppb and S_{VOC} is 0.83 ppb, indicating that DFW ANO_x controls are approximately 7.9 times as effective per ton as DFW AVOC controls for reducing episode-average 8 hour O_3 concentrations (Figure 8). The equal-weighted a priori ensemble yields a distribution of O_3 sensitivity results and indicates 93% likelihood that O_3 is more sensitive to DFW ANO_x than to DFW AVOC, but a 2.3% chance that reducing local ANO_x emission may actually increase episode-average 8 hour O_3 concentrations in the region. A sharp negative correlation is observed between O_3 sensitivities to NO_x and VOC, which leads to a large variability in the ratio

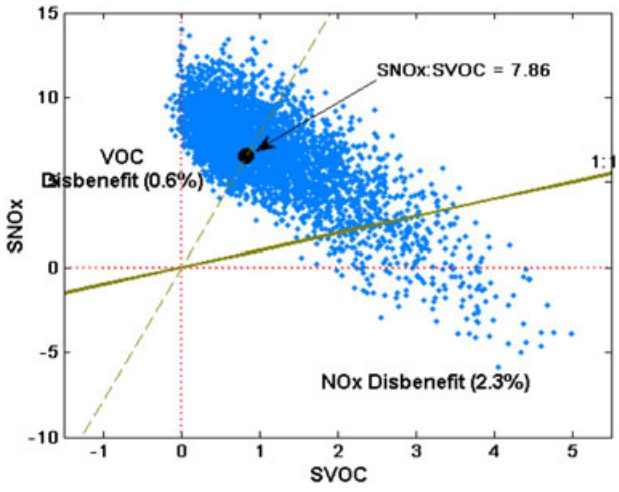

Figure 8. A priori episode-average 8 hour ozone sensitivity results at Denton.

Table 7. Comparison of Prior and Observation-Constrained Predictions of Episode-Average Sensitivities of 8 Hour Ozone at Denton to Dallas-Fort Worth Emissions[a]

	S_{ANOx} (ppb)	S_{AVOC} (ppb)
	($\mu \pm \sigma$)	($\mu \pm \sigma$)
Prior	6.79 ± 2.59	1.09 ± 0.81
Metric 1	6.98 ± 2.19	1.03 ± 0.54
Metric 2	6.67 ± 3.01	1.35 ± 0.74
Metric 3	6.49 ± 2.83	1.28 ± 0.69

[a]μ and σ denote mean and standard deviation, respectively.

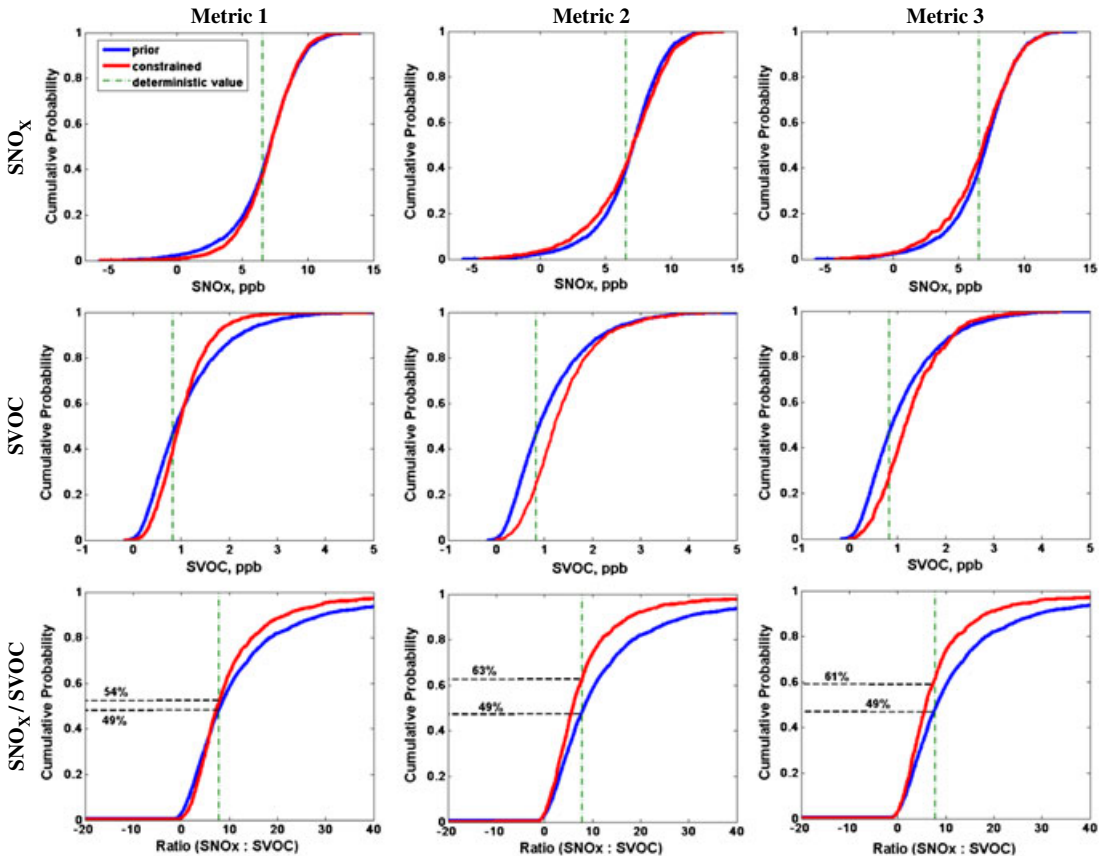

Figure 9. CDFs for ensemble predictions of episode-average sensitivity of 8 hour ozone at the Denton monitor under the three metrics.

of these two sensitivities. This reflects the tendency of changes in model inputs to push the O_3 formation regime toward being more NO_x limited or more VOC limited, and hence less sensitive to the other precursor.

[47] The observational metrics also yield adjusted distributions of O_3 sensitivity to DFW ANO_x and AVOC emissions. M1 does not substantially change the mean estimate; however, applications of M2 and M3 shift O_3 sensitivity toward slightly higher S_{VOC} and slightly lower S_{NO_X} than in the equal-weighted ensemble (Figure 9 and Table 7). This is also seen in the shift toward lower values of $S_{NO_X} : S_{VOC}$ under M2 and M3, even as predictions remain primarily NO_x limited (Figure 9). This is because most of the cases accepted by the M2 and M3 screenings used the alternate (CB6) chemical mechanism and higher domain-wide ENO_x (Table 4 and Figure 7), each of which makes O_3 slightly more sensitive to VOC compared to NO_x (Figures 3 and 4). M1 favored cases with CB6 (Table 4) but gave low weightings to cases with high ENO_x (Figure 7).

6. Conclusion

[48] In this study, measurements of O_3 and NO_x have been used to adjust probabilistic estimates of O_3 concentrations and O_3 responsiveness to NO_x and VOC emission changes in the DFW region. Three distinct observation-based approaches have been applied to weight or screen an ensemble of model simulations that employ alternate model assumptions (*structural uncertainty*) and model input values (*parametric uncertainty*).

[49] Screening analysis of *structural* uncertainties led to a focus on scenarios involving alternate choices for the biogenic emissions model and chemical mechanism. However, alternate meteorological scenarios were not available for this study, an important limitation that could be explored in further research. The omission of some key structural uncertainties such as meteorology led the ensemble spread to be too narrow (underdispersive) in simulating observed concentrations [*Cohan et al*., unpublished report, 2011], and may lead to errors in assessing the true values of the uncertain inputs considered here. For *parametric* uncertainties, impact analysis identified the specific emission rates, reaction rate constants, and boundary conditions that most influence O_3 concentrations and their sensitivities to NO_x and VOC emissions. Some parameters such as O_3 boundary conditions were found to impact concentrations far more strongly than sensitivities, whereas the converse was true for some other parameters such as EAVOC.

[50] Traditional Monte Carlo analysis of uncertain inputs or model ensembles yields probabilistic (a priori) estimates of model outputs but assumes that each of the scenarios is equally likely. This article has explored three of the many Bayesian and non-Bayesian approaches that could be used to adjust these a priori estimates by evaluating each case against observations. All three metrics tend to favor the CB6 chemical mechanism over CB05 for this region and episode, and two of the metrics favor scaling up NO_x emission rates. These resulted in enhanced O_3 responsiveness to VOC emission and dampened sensitivity to NO_x, although the region still remained predominantly NO_x limited.

[51] The Bayesian and non-Bayesian metrics introduced here are just three of the many that could be chosen for observation-constrained analysis, and each has its own strengths and weaknesses. Non-Bayesian M2 and M3 use more observed data to prioritize simulations compared to Bayesian M1, but disregard simulations that fail to meet the test criteria. The Bayesian metric, on the other hand, retains all cases but with varying weights based on their likelihood of estimating actual measurements. This metric, however, relies on the questionable assumption that the measurement errors are statistically independent and normally distributed. Alternate Bayesian metrics could be developed to avert this assumption or to use less aggregated observational data. M2 has the practical advantage of mimicking EPA's standard test criteria for acceptable attainment demonstration modeling, but there is little conceptual basis for treating all acceptable cases equally or excluding all others.

[52] A key assumption of this study is that performance of a model case in simulating observed concentrations provides an indicator for the reliability of the input choices and output sensitivity predictions associated with that case. Because ambient monitors observe concentrations but not sensitivities, this assumption is both necessary and yet unverifiable. Dynamic evaluation of how pollutant concentrations respond to emission changes over weekly (i.e., weekday vs. weekend) or interannual (e.g., before and after a major emission trend) time scales can provide a proxy for ground-truthing sensitivity estimates [*Dennis et al*., 2010; *Gilliland et al*., 2008; *Pierce et al*., 2010; *Yarwood et al*., 2003].

[53] Methods applied in this study could readily be extended to consider other regions (e.g., larger domains to include both urban and rural settings), episodes, and observational metrics. In particular, more disaggregated observations could be considered for the Bayesian metric; doing so would more fully capitalize on the spatial and temporal specificity of available data, but also tends to yield vastly different weightings among similar model cases due to the multiplicative nature of the Bayesian likelihood function [*Cohan et al*., unpublished report, 2011]. Future work could also consider observations taken aloft by aircraft, sondes, and satellites [*Henderson et al*., 2012; *Yang et al*., 2010]. Other model constraining methods such as Bayesian model averaging [*Raftery et al*., 2005] may be explored to consider errors in both the model and the measurements. Additional structural uncertainties such as use of alternate meteorological inputs or model formulations could expand the ensemble considered here.

[54] **Acknowledgments.** The preparation of this manuscript is based on work supported by the State of Texas through the Air Quality Research Program administered by The University of Texas at Austin by means of a grant from the TCEQ and by National Science Foundation grant #087386. Although this article has been reviewed by the EPA and TCEQ and approved for publication, it does not necessarily reflect the policies or views of either agency. The baseline modeling for the study was provided by TCEQ.

References

Anderson, T. W. (1962), On the distribution of the two-sample Cramer-von Mises criterion, *Ann. Math. Stat.*, *33*, 1148–1159.

Beekmann, M., and C. Derognat (2003), Monte Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the atmospheric pollution over the Paris area (ESQUIF) campaign, *J. Geophys. Res.*, *108*(8559), D17.

Bergin, M. S., and J. B. Milford (2000), Application of Bayesian Monte Carlo analysis to a Lagrangian photochemical air quality model, *Atmos. Environ.*, *34*(5), 781–792.

Bergin, M. S., G. S. Noblet, K. Petrini, et al. (1999), Formal uncertainty analysis of a Lagrangian photochemical air pollution model, *Environ. Sci. Technol.*, *33*, 1116–1126.

Bey, I., D. J. Jacob, R. M. Yantosca, J. A. Logan, B. D. Field, A. M. Fiore, Q. B. Li, H. G. Y. Liu, L. J. Mickley, M. G. Schultz (2001), Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, *J. Geophys. Res.*, *106*(D19), 23,073–23,095, doi:10.1029/2001JD000807.

Carlton, A. G., and K. R. Baker (2011), Photochemical modeling of the Ozark Isoprene Volcano: MEGAN, BEIS, and their impacts on air quality predictions, *Environ. Sci. Technol.*, *45*(10), 4438–4445.

Cohan, D. S., A. Hakami, Y. Hu, and A. G. Russell (2005), Nonlinear response of ozone to emissions: Source apportionment and sensitivity analysis, *Environ. Sci. Technol.*, *39*(17), 6739–6748.

Deguillaume, L., M. Beekmann, and L. Menut (2007), Bayesian Monte Carlo analysis applied to regional-scale inverse emission modeling for reactive trace gases, *J. Geophys. Res.*, *112*, D02307.

Deguillaume, L., M. Beekmann, and C. Derognat (2008), Uncertainty evaluation of ozone production and its sensitivity to emission changes over the Ile-de-France region during summer periods, *J. Geophys. Res.*, *113*, D02304.

Demerjian, K. L. (2000), A review of national monitoring networks in North America, *Atmos. Environ.*, *34*(12–14), 1861–1884, doi:10.1016/S1352-2310(99)00452-5.

Dennis, R., et al. (2010), A framework for evaluating regional-scale numerical photochemical modeling systems, *Environ. Fluid Mech.*, *10* (4), 471–489.

Digar, A. (2012), Uncertainty in regional air quality modeling, Ph.D. thesis, Civil and Environmental Engineering, Rice University, Houston, Tex.

Digar, A., and D. S. Cohan (2010), Efficient characterization of pollutant-emission response under parametric uncertainty, *Environ. Sci. Technol.*, *44*(17), 6724–6730.

Digar, A., D. S. Cohan, D. D. Cox, B. Kim, and J. W. Boylan (2011), Likelihood of achieving air quality targets under model uncertainties, *Environ. Sci. Technol.*, *45*(1), 189–196.

Dudhia, J. (1993), A nonhydrostatic version of the Penn State–NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front, *Mon. Weather Rev.*, *121*(5), 1493–1513.

Dunker, A. M. (1984), The decoupled direct method for calculating sensitivity coefficients in chemical kinetics, *J. Chem. Phys.*, *81*(5), 2385–2393.

Dunlea, E. J., et al. (2007), Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment, *Atmos. Chem. Phys.*, *7*(10), 2691–2704.

ENVIRON (2007), User's Guide to Emissions Processor, Version 3. ENVIRON International Corporation., Novato, Calif.

ENVIRON (2008), Boundary conditions and fire emissions modeling, final report to the Texas Commission on Environmental Quality (TCEQ). ENVIRON International Corporation., Novato, Calif.

ENVIRON (2009), Updated boundary conditions. Final report prepared for Texas Commission on Environmental Quality, Unpublished report available at http://www.tceq.texas.gov/assets/public/implementation/air/am/contracts/reports/pm/5820784005FY0916-20090730-environ-updated_bc.pdf.

ENVIRON (2010), User's guide—comprehensive air-quality model with extensions, Version 5.30., ENVIRON International Corporation, Novato, Calif.

Fine, J., L. Vuilleumier, S. Reynolds, P. Roth, and N. Brown (2003), Evaluating uncertainties in regional photochemical air quality modeling, *Annu. Rev. Environ. Resour.*, *28*, 59–106.

Gilliland, A. B., C. Hogrefe, R. W. Pinder, J. M. Godowitch, K. L. Foley, and S. T. Rao (2008), Dynamic evaluation of regional air quality models: Assessing changes in O-3 stemming from changes in emissions and meteorology, *Atmos Environ*, *42*(20), 5110–5123.

Guenther, A., T. Carl, P. Harley, C. Wiedinmyer, P. I. Palmer, and C. Geron (2006), Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), *Atmos. Chem. Phys.*, *6*(11), 3181–3210.

Hanna, S. R., Z. G. Lu, H. C. Frey, N. Wheeler, J. Vukovich, S. Arunachalam, M. Fernau, and D. A. Hansen (2001), Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, *Atmos. Environ.*, *35*(5), 891–903, doi:10.1016/S1352-2310(00).00367-8.

Hakami, A., M. T. Odman, and A. G. Russell (2003), High-order, direct sensitivity analysis of multidimensional air quality models, *Environ. Sci. Technol.*, *37*(11), 2442–2452.

Henderson, B. H., et al. (2012), Combining Bayesian methods and aircraft observations to constrain the $HO + NO_2$ reaction rate, *Atmos. Chem. Phys.*, *12*, 653–667. doi:10.5194/acp-12-653-2012.

Hogrefe, C., and S. T. Rao (2001), Demonstrating attainment of the air quality standards: Integration of observations and model predictions into the probabilistic framework, *J Air Waste Manage. Assoc. 51*(7), 1060–1072.

Holland, D. M., and T. Fitz-Simons (1982), Fitting statistical distributions to air quality data by the maximum likelihood method, *Atmos. Environ.*, *16* (5), 1071–1076.

Lamsal, L. N., et al. (2008), Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument, *J. Geophys. Res.*, *113*(D16), D16308.

Mallet, V. and B. Sportisse (2006), Ensemble-based air quality forecasts: A multimodal approach applied to ozone, *J. Geophys. Res.*, *111*, D18302.

Mollner, A. K., et al. (2010), Rate of gas phase association of hydroxyl radical and nitrogen dioxide, *Science*, *330*(6004), 646–649.

Pierce, T., C. Hogrefe, S. T. Rao, P. S. Porter, and J. Ku (2010), Dynamic evaluation of a regional air quality model: Assessing the emissions-induced weekly ozone cycle, *Atmos. Environ.*, *44*(29), 3583–3596.

Pinder, R. W., R. C. Gilliam, K. W. Appel, S. L. Napelenok, K. M. Foley, and A. B. Gilliland (2009), Efficient probabilistic estimates of surface ozone concentration using an ensemble of model configurations and direct sensitivity calculations, *Environ. Sci. Technol.*, *43*(7), 2388–2393.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005), Using Bayesian model averaging to calibrate forecast ensembles, *Mon. Weather Rev.*, *133*, 1155–1174.

Russell, A. G., and R. Dennis (2000), NARSTO critical review of photochemical models and modeling. *Atmos. Environ.*, *34*(12–14), 2283–2324.

Sander, S. P., et al. (2006), Chemical kinetics and photochemical data for use in atmospheric studies, Evaluation Number 15, NASA Jet Propulsion Laboratory, available at http://jpldataeval.jpl.nasa.gov/pdf/JPL_15._AllInOne.pdf, NASA JPL.

Slinn, S. A., and W. G. N. Slinn (1980), Predictions for particle deposition on natural waters, *Atmos. Environ.*, *14*(9), 1013–1016.

Taylor, J. A., R. W. Simpson, and A. J. Jakeman (1987), Statistical modeling of restricted pollutant data sets to assess compliance with air quality criteria, *Environ. Monit. Assess.*, *9*(1), 29–46.

Texas Commission on Environmental Quality (2011a), Dallas-Fort Worth attainment demonstration SIP revision for the 1997 eight-hour ozone standard nonattainment area. Prepared for Texas Commission on Environmental Quality as part of the revisions to the State of Texas air quality implementation plan for the control of ozone air pollution in Dallas-Fort Worth eight-hour ozone nonattainment area, Unpublished government document available at http://www.tceq.texas.gov/airquality/sip/dfw_revisions.html.

Texas Commission on Environmental Quality (2011b), Dallas-Fort Worth reasonable further progress state implementation plan revision for the 1997 eight-hour Ozone standard. Prepared for Texas Commission on Environmental Quality as part of the revisions to the State of Texas air quality implementation plan for the control of ozone air pollution in Dallas-Fort Worth 1997 eight-hour ozone nonattainment area, Unpublished government document available at http://www.tceq.texas.gov/airquality/sip/dfw_revisions.html.

Tian, D., M. S. Bergin, D. S. Cohan, S. L. Napelenok, Y. Hu, M. E. Chang, and A. G. Russell (2010). Uncertainty analysis of ozone formation and response to emission controls using higher-order sensitivities. *J. Air Waste Manage. Assoc.*, *60*, 797–804.

U.S. Environmental Protection Agency (1999), Draft report on the use of models and other analyses in attainment demonstrations for the 8-hr Ozone NAAQS, U.S. Environmental Protection Agency, Research Triangle Park, N.C.

U.S. Environmental Protection Agency (2006), Air Quality Criteria for Ozone and Related Photochemical Oxidants, Government report available at http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=149923.

U.S. Environmental Protection Agency (2007), Guidance on the use of models and other analyses for demonstrating attainment of air quality goals for Ozone, PM2.5, and Regional Haze, Government report available at http://www.epa.gov/scram001/guidance/guide/final-03-pm-rh-guidance.pdf.

Wesely, M. L. (1989), Paramaterization of surface resistances to gaseous dry deposition in regional-scale numerical models, *Atmos. Environ.*, *23*(6), 1293–1304.

Yang, Q., et al. (2010), A study of tropospheric ozone column enhancements over North America using satellite data and global chemical transport model, *J. Geophys. Res.*, *115*, D08302, doi:10.1029/2009JD012616.

Yarwood, G., G. Wilson, C. Emery, and A. Guenther (1999), Development of the GloBEIS—a state of the science biogenic emissions modeling system. Final report to the Texas Natural Resource Conservation Commission, Austin, Tex., Report available at http://www.camx.com/publ/pdfs/cb05_final_report_120805.aspx.

Yarwood, G., T. E. Stoeckenius, J. G. Heiken, A. M. Dunker (2003), Modeling weekday/weekend ozone differences in the Los Angeles region for 1997, *J. Air Waste Manage. Assoc.*, *53*(7), 864–875.

Yarwood, G., S. Rao, M. Yocke, G. Z. Whitten, and S. Reyes (2005), Updates to the Carbon Bond chemical mechanism: CB05. Final report

to the U.S. Environmental Protection Agency, Report available at http://www.camx.com/publ/pdfs/cb05_final_report_120805.aspx.

Yarwood, G., G. Z. Whitten, and J. Jung (2010), Development, evaluation and testing of Version 6 of the Carbon Bond chemical mechanism (CB6), Final report prepared for Texas Commission on Environmental Quality, Unpublished final report from ENVIRON to TCEQ under Work Order No. 582-7-84005-FY10-26.

Zhang, L., J. R. Brook, and R. Vet (2003), A revised parameterization for gaseous dry deposition in air-quality models, *Atmos. Chem. Phys.*, *3*, 2067–2082.

Zhang, L. M., S. Gong, J. Padro, and L. Barrie (2001), A size-segregated particle dry deposition scheme for an atmospheric aerosol module, *Atmos. Environ*, *35*(3), 549–560.