Dynamic Evaluation of Regional Air Quality Model's Response to Emission Reductions in the Presence of Uncertain Emission Inventories

Sergey L. Napelenok[1*], Kristen M. Foley[1], Daiwen Kang[2], Rohit Mathur[1], Thomas Pierce[1], S. Trivikrama Rao[1]

[1]Atmospheric Modeling and Analysis Division, National Exposure Research Laboratory, United States Environmental Protection Agency, Research Triangle Park, North Carolina

[2]Computer Sciences Corporation, Research Triangle Park, North Carolina

Abstract

A method is presented and applied for evaluating an air quality model's changes in pollutant concentrations stemming from changes in emissions while explicitly accounting for the uncertainties in the base emission inventory. Specifically, the Community Multiscale Air Quality (CMAQ) model is evaluated for its ability to simulate the change in ozone (O_3) levels in response to significant reductions in nitric oxide ($NO_x = NO + NO_2$) emissions from the NO_x State Implementation Plan (SIP) Call and vehicle fleet turnover between the years of 2002 and 2005. The dynamic model evaluation (i.e. the evaluation of a model's ability to predict changes in pollutant levels given changes in emissions) differs from previous approaches by explicitly accounting for known uncertainties in the NO_x emissions inventories. Uncertainty in three sectors of NO_x emissions is considered – area sources, mobile sources, and point sources – and is propagated using sensitivity coefficients calculated by the decoupled direct method in three dimensions (DDM-3D). The change in O_3 levels between 2002 and 2005 is estimated based on

differences in the empirical distributions of the modeled and observed data during the two years. Results indicate that the CMAQ model is able to reproduce the observed change in daily maximum 8-hr average O_3 levels at more than two-thirds of Air Quality System (AQS) monitoring locations when a relatively moderate amount of uncertainty (50%) is assumed in area and mobile emissions of NO_x together with a low amount of uncertainty (3%) in the utility sector (elevated point sources) emissions. The impact of other sources of uncertainty in the model is also briefly explored.

Key words: dynamic model evaluation; CMAQ; direct decoupled method; DDM; air quality modeling; ozone; uncertainty; sensitivity

* Corresponding author. Email: napelenok.sergey@epa.gov; Phone: +1-919-541-1135; Fax: +1-919-541-1379; Address: EPA (MD-E243-04), RTP, NC 27711.

1. Introduction

Regional air quality models (RAQMs) are an integral part of air quality management. Their applications include developing national strategies for improving air quality (Moran, 2005), assessing the regional-scale transport of pollutants (Rao et al., 2008), and informing international treaties on the hemispheric transport of air pollutants (Holloway et al., 2003). When applied for these tasks, RAQMs typically employ "current" and "future" emission scenarios, with the commonly examined metric of interest being the change in the concentration or deposition of air pollutants attributable to the change in emissions.

46 The integrity of an RAQM is ascertained through comprehensive model evaluation by
47 comparing model outputs with observations (Dennis et al., 2010). In the context of air quality
48 management, where the model's response to varying inputs is of concern, it is essential to
49 evaluate the model's ability to accurately predict changes in pollutant concentrations (i.e., model
50 response) due to changes in emission inputs or meteorology. Evaluating the model's response is
51 referred to here as dynamic model evaluation. In contrast to the operational evaluation, in which
52 model outputs are usually paired in space and time with available observations, dynamic
53 evaluation is more difficult to perform, because a fully controlled dynamic air quality scenario is
54 nearly impossible to observe and quantify in nature. Rather, the model's ability to correctly
55 respond to different input conditions can be evaluated by examining modeling periods separated
56 by substantial and quantifiable differences in the input variables of interest, namely, emissions
57 and/or meteorological conditions.
58 One frequently evaluated scenario satisfying the above prerequisites is the period
59 between 2002 and 2005 in the eastern United States when substantial reductions of nitrogen
60 oxide ($NO_x = NO + NO_2$) emissions were achieved through the U.S. Environmental Protection
61 Agency's (US-EPA) NO_x State Implementation Plan (SIP) Call (US-EPA, 2005). Additionally,
62 reductions in mobile emissions due to vehicle fleet turnover and introduction of lower emitting
63 automobiles also occurred during this period. Several previous dynamic evaluation studies that
64 focused on this period (Gilliland et al., 2008; Godowitch et al., 2010; Pierce et al., 2010) found a
65 smaller change in model predictions of peak surface ozone (O_3) concentrations compared to the
66 observed change. These studies reported that the model predicts around 40-80% of the observed
67 change in the daily maximum 8-hr average O_3, and hypothesized that a lower magnitude
68 response in the model predictions may be attributable to errors in NO_x emission inputs used in

69 the simulations. However, potential uncertainties in the NO_x emission inventories were not
70 explicitly accounted for in these dynamic evaluation studies.
71 The objective of the work presented here is to perform a dynamic evaluation of the
72 Community Multiscale Air Quality (CMAQ) modeling system (Byun and Schere, 2006)
73 covering the NO_x SIP Call period while accounting for uncertainties in NO_x emission inputs.
74 Specifically, the change in the modeled daily maximum 8-hr average O_3 concentrations between
75 the summers of 2002 and 2005 in the eastern United States is evaluated against the change seen
76 in the observations. The proposed approach is used to investigate the extent to which the
77 previously noted model's insufficient response to changes in NO_x emissions is attributable to
78 uncertainties in these model inputs.
79 Various methods have been developed to propagate uncertainty in input parameters
80 through RAQMs. These typically include some variation of the "brute force" Monte Carlo
81 simulations, in which the model is rerun many times with varying inputs (Hanna et al., 2001;
82 Boynard et al., 2011), as well as studies where a reduced form of the model is developed based
83 on calculated sensitivity coefficients (Digar and Cohan, 2010; Tian et al., 2010), which are then
84 coupled with randomly sampled distributions of uncertainties. Each method culminates in an
85 ensemble of model predictions from which inferences on the impact of input uncertainty can be
86 ascertained. To dynamically evaluate CMAQ in light of uncertainties in NO_x emissions inputs,
87 the sensitivity-based approach to propagate uncertainty is used here as it is significantly more
88 computationally efficient (Pinder et al., 2009), because the random sampling calculations are
89 much less computationally demanding than running a regional photochemical simulation model
90 many times. Thus, the sensitivity-based approach allows for the development of large member
91 ensembles for two summer periods with a lower computational burden. Furthermore, the method

92 presented here extends previous work by combining the sensitivity-based uncertainty analysis
93 with a dynamic evaluation.
94
95 **2. Method**
96
97 *2.1. Modeling System and Observations*
98 Model simulations were conducted with a 12 km horizontal grid covering the eastern
99 United States nested within a 36 km grid over the entire continental United States, including
100 parts of Canada and Mexico. The modeled period spanned the period from 1 June to 31 August
101 in both 2002 and 2005. CMAQ version 4.7.1, with decoupled direct method in three dimensions
102 (DDM-3D) (Napelenok et al., 2008), was used to calculate O_3 concentrations and sensitivities of
103 O_3 to three emission sectors (area sources, mobile sources, and point sources) of NO_x.
104 Atmospheric chemistry was simulated with the latest available version of the carbon bond
105 mechanism (CB05) (Sarwar et al., 2008). Meteorological inputs were supplied by MM5 version
106 3.6.3 (Grell et al., 1994) configured with the standard physics options (see Godowitch et al.,
107 2010). Boundary conditions for the larger domain were specified based on outputs of the global
108 model GEOS-Chem (Bey et al., 2001). Emissions were developed using the Sparse Matrix
109 Operator Kernel Emissions (SMOKE) processor (http://www.smoke-model.org) version 2.4
110 based on temporally and spatially resolved wildfire, electricity generating units, and mobile
111 sources. Domain-wide NO_x emissions reflected reductions from the SIP Call as well as
112 reductions in the mobile sector. On average, point source emissions of NO_x were reduced by
113 22% and mobile source emissions were reduced by 18% between the summer of 2002 and the

summer 2005 in the modeling domain; area sources, which included biogenic NO emissions, remained relatively unchanged over these two years (0.2% reduction).

The observed O_3 concentrations for this study were obtained from the EPA's Air Quality System (AQS, http://www.epa.gov/air/data/aqsdb.html). Data from over 700 monitoring stations were available within the 12 km modeling domain during the two modeling periods. AQS O_3 data was processed to calculate daily maximum 8-hour average mixing ratios according to standard regulatory procedures (Office of the Federal Register, 1997). Monitoring sites with less than 80 complete days of observations out of a total possible of 92 were excluded from the analysis.

2.2 Uncertainty Propagation

Calculated DDM-3D sensitivity coefficients were used to estimate the O_3 response to perturbations in the uncertain inputs of the three NO_x emissions categories (area, mobile, point) through Taylor series expansion (Morgan and Henrion, 1990; Hakami et al., 2003). Generally, pollutant concentration as a function of any one perturbation can be reconstructed using the following:

$$C_j\left(\bar{\mathbf{x}},t\right)=C_0\left(\bar{\mathbf{x}},t\right)+\Delta\varepsilon_j S_j^{(1)}\left(\bar{\mathbf{x}},t\right)+\frac{1}{2}\Delta\varepsilon_j^2 S_{j,j}^{(2)}\left(\bar{\mathbf{x}},t\right)+h.o.t., \qquad (1)$$

where $C_j\left(\bar{\mathbf{x}},t\right)$ is the concentration due to a specific perturbation j at time t and location $\bar{\mathbf{x}}$; $C_0\left(\bar{\mathbf{x}},t\right)$ is base, unperturbed concentration; $\Delta\varepsilon_j$ is the fractional perturbation of the parameter j; $S_j^{(1)}\left(\bar{\mathbf{x}},t\right)$ and $S_{j,j}^{(2)}\left(\bar{\mathbf{x}},t\right)$ are the first and second order sensitivity coefficients, and *h.o.t.* are higher order terms with little impact on the approximation.

136 To account for uncertainty in several parameters at once – area, mobile and point

137 emissions – and dropping the higher order terms, the Taylor Series was expanded as follows:

$$
\begin{aligned}
C_{a+m+p}\left(\bar{\mathbf{x}},t\right) &\approx C_0\left(\bar{\mathbf{x}},t\right) \\
&+ \Delta\varepsilon_a S_a^{(1)}\left(\bar{\mathbf{x}},t\right) + \frac{1}{2}\Delta\varepsilon_a^2 S_{a,a}^{(2)}\left(\bar{\mathbf{x}},t\right) \\
&+ \Delta\varepsilon_m S_m^{(1)}\left(\bar{\mathbf{x}},t\right) + \frac{1}{2}\Delta\varepsilon_m^2 S_{m,m}^{(2)}\left(\bar{\mathbf{x}},t\right) \\
&+ \Delta\varepsilon_p S_p^{(1)}\left(\bar{\mathbf{x}},t\right) + \frac{1}{2}\Delta\varepsilon_p^2 S_{p,p}^{(2)}\left(\bar{\mathbf{x}},t\right) \\
&+ \Delta\varepsilon_a \Delta\varepsilon_m S_{a,m}^{(2)}\left(\bar{\mathbf{x}},t\right) \\
&+ \Delta\varepsilon_m \Delta\varepsilon_p S_{m,p}^{(2)}\left(\bar{\mathbf{x}},t\right) \\
&+ \Delta\varepsilon_p \Delta\varepsilon_a S_{p,a}^{(2)}\left(\bar{\mathbf{x}},t\right)
\end{aligned}
\qquad (2)
$$

139 where the subscripts *a, m*, and *p*, represent area, mobile, and point NO_x emissions

140 respectively.

141 The formulation in Equation 2 allows for efficient recalculation of O_3 predictions based

142 on perturbations in the three uncertain input parameters. An ensemble of model predictions was

143 developed for the two modeling periods (summers of 2002 and 2005) by randomly sampling

144 (with replacement) from a continous uniform distribution for each perturbation parameter, $\Delta\varepsilon_a$,

145 $\Delta\varepsilon_m$, and $\Delta\varepsilon_p$. Uniform distribution was chosen to not precsribe the shape of the distribution

146 and specify only the lower and upper bounds (Cullen and Frey, 1999). One hundred such

147 samples were found to be sufficient; higher sample sizes did not have an impact on the

148 interpretation of results.

149 The uncertainty ranges used to sample $\Delta\varepsilon_j$ were: ± 3% in point sources; ±50% in mobile

150 sources; and ± 50% in area sources. The low uncertainty in the emissions from point sources is

151 due to the prevalence of the continous emission monitoring systems (CEMS) on the majority of

152 these sources. Uncertainty in the other two sectors was based on the previous efforts that have
153 attempted to quantify emissions uncertainty. For example, Hanna et al. (2001) suggested
154 uncertainty ranges of approximately 20-50% for major point sources and 35-100% for other
155 categories. Comparisons of different methods of calculating mobile emissions have shown a
156 variability of approximately 35-47% (Parrish, 2006; Dallmann and Harley, 2010). NO_x
157 inventories constructed from inverse modeling studies based on satellite and ground-based
158 observations coupled with regional chemical transport models also show a wide range of
159 variability when compared to a priori estimates (Konovalov et al., 2006; Deguillaume et al.,
160 2007; Napelenok et al., 2008). The Taylor Series approach (Equation 2) allowed for flexibility
161 in selecting uncertainty ranges due to the trivial computational costs associated with the
162 sampling. The sensitivity of the final results to the choice of these uncertainty ranges is
163 discussed in more detail below.
164
165 *2.3 Dynamic evaluation metrics*
166 The model output data generated in this experiment consisted of two 100-member
167 ensembles of daily maximum 8-hr average ozone concentrations: one for the 2002 summer
168 season and one for the 2005 summer season, each with 100 Monte Carlo samples of the
169 uncertainty ranges for the three NO_x emission sectors (Figure 1). As noted before, a sample size
170 of 100 was found to be sufficient in the ensemble analysis. These ensembles are compared to
171 base model simulations and observational values. Observational data consisted of June through
172 August time series data for each monitoring site and each year. All data were first processed to
173 calculate the daily maximum 8-hr average O_3 values. Since the two modeling seasons were
174 separated by differing meteorology and mismatched temporal emission patterns (in addition to

different emission magnitudes), the dynamic signal should not be discerned by pairing model and observational data in time. Instead, the data were paired in space only (at each observational site.)

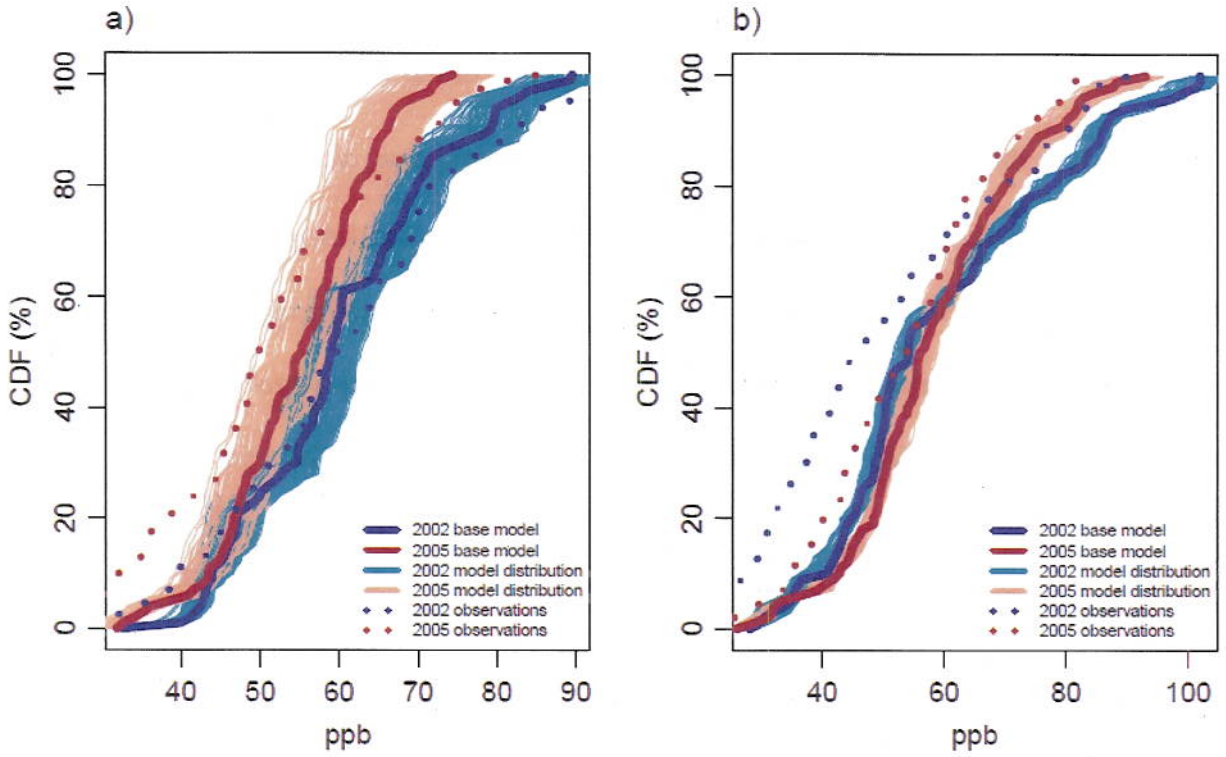


Figure 1. Base model, observations, and model ensemble empirical cumulative distributions of daily maximum 8-hr average ozone concentrations for 2002 and 2005 at two AQS sites: a) Terre Haute, IN (AQS#181670024); and b) Detroit, MI (AQS#261630019). All ensembles were constructed based on ±50% uncertainty in emissions of area and mobile NO_x and ±3% uncertainty in emissions of point NO_x. The wide spread of the ensemble at the Terre Haute site indicates greater sensitivity to NO_x emissions in comparison to the site in Detroit.

The ability of the CMAQ model to correctly predict the change in daily maximum 8-hr average O_3 values was evaluated by estimating the difference between the empirical distributions of ozone model predictions in the two years. The difference was quantified using two measures: the root mean square deviation (RMSD) and mean deviation (MD) calculations. The RMSD was calculated as follows:

$$RMSD(2002,2005) = \sqrt{\frac{\sum_{p=p_l}^{p_u}\left(\hat{X}^{2002}(p) - \hat{X}^{2005}(p)\right)^2}{N}}, \quad (3)$$

where N is the cardinality of $\{p_l,..., p_u\}$ and $\hat{X}^{year}(p)$ is the p^{th} sample quantile of the data set. For example, $\hat{X}^{2002}(0)$ equals the minimum value of the 2002 ozone predictions, $\hat{X}^{2002}(50)$ equals the median, $\hat{X}^{2002}(100)$ equals the maximum value, etc.

The RMSD measures the average distance between two distributions, but it does not indicate whether the difference tends to be positive or negative. Since the model simulation needs to capture the magnitude and direction of a change in ozone values due to changes in emissions, the mean deviation between the two distributions was also calculated:

$$MD(2002,2005) = \frac{\sum_{p=p_l}^{p_u}\left(\hat{X}^{2002}(p) - \hat{X}^{2005}(p)\right)}{N} \quad (4)$$

RMSD and MD were calculated for two cases: one where p_l = 0% and p_u = 100% to quantify the difference in the full range of O_3 concentrations; and one where p_l = 95% and p_u = 100% to quantify the difference in the more policy-relevant higher ozone values.

These metrics were calculated for the difference between the observational data sets in 2002 and 2005, and for the difference between the base modeled data sets in the same years. Additionally, an ensemble of differences between modeled O_3 values for 2002 and 2005 was calculated by randomly sampling model ensemble members from each simulation year and then calculating the RMSD and MD for each simulation pair. The 95% confidence intervals based on the ensemble distributions for these metrics were then used to establish criteria for acceptable model performance. The model is said to have successfully captured the observed change in O_3 at any one site if the observed RMSD and MD metrics both fell within these estimated

211 confidence intervals. For example, at the AQS site near Terre Haute, IN the observed change
212 was estimated by RMSD=10.0 ppb and MD = 9.5 ppb (using p_l =0%, p_u = 100%) (Figure 2a, b).
213 The 95% confidence interval based on the ensemble of model results was (3.34, 13.5) ppb for the
214 RMSD and (-0.6, 12.6)ppb for the MD. Thus, the model was able to capture the observed
215 change in the ozone distribution across these years, considering ±50% uncertainty in the inputs
216 of mobile and area sources, and ±3% uncertainty in point sources of NO_x emissions. Use of both
217 the RMSD and the MD provides a stringent test for the model since each metric highlights a
218 different attribute of the difference between the two empirical distributions. This is illustrated by
219 an AQS site outside of Detroit, MI (Figure2c, d). At this site, the observed RMSD of 6.6 ppb
220 was within the model range of (4.5, 6.8) ppb, but the observed MD of -3.3 pbb was well outside
221 the model range of (-0.7, 3.4) ppb. The behavior in MD at Detroit was caused by the fact that at
222 the lower part of the distributions of both modeled and observed values (10%-60% range), the O_3
223 values were higher in 2005, and at the higher end of the distribution (60-100% range), the O_3
224 values were higher in 2002 (Figure 1b). At the same time, the model underestimated the
225 difference at the low end of the distribution and overestimated the difference at the high end,
226 leading to a range of MDs much greater than the MD of observations (Figure 2d).

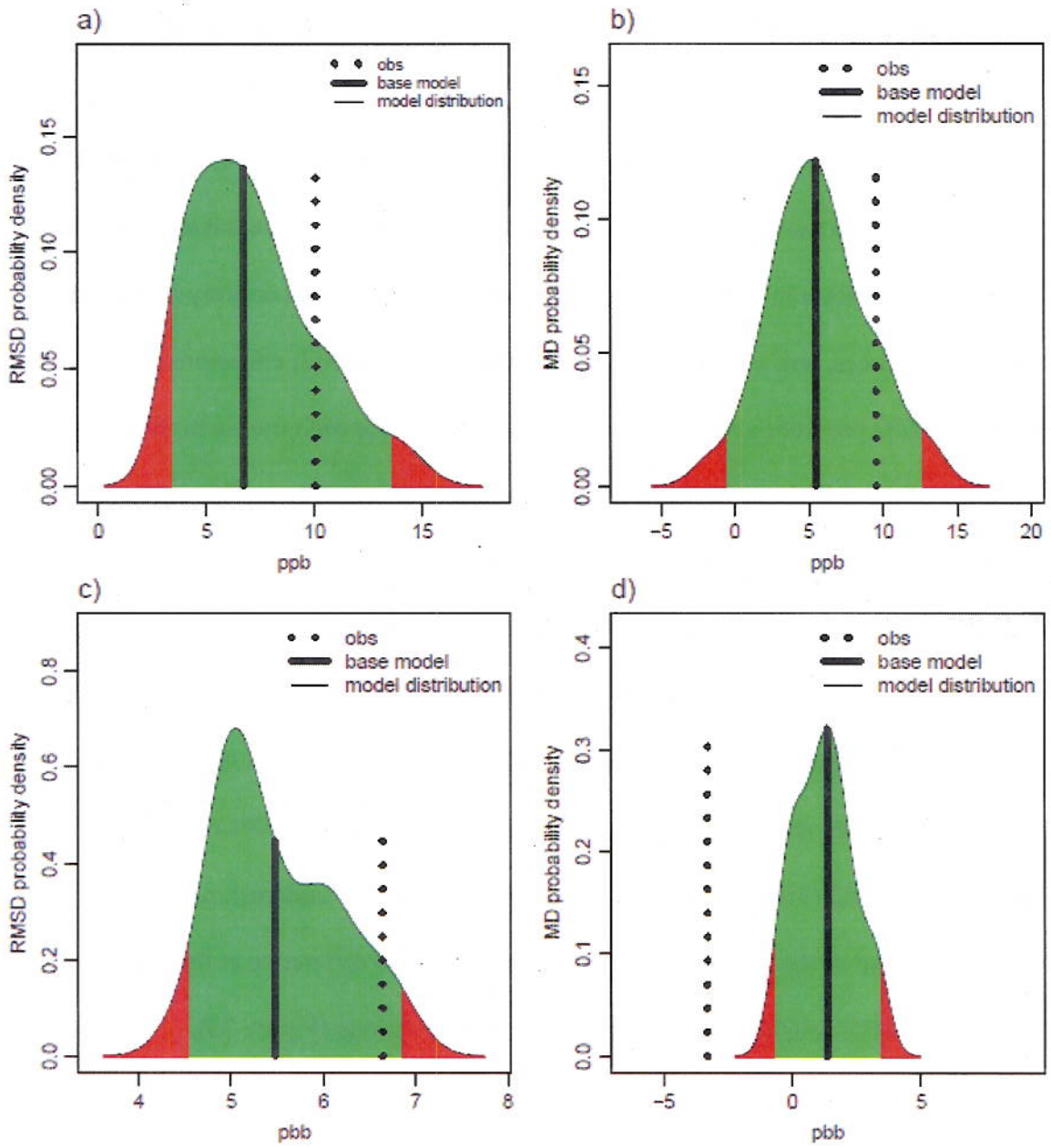


Figure 2. Root mean square deviation (RMSD) and mean deviation (MD) at two AQS sites: a,b) Terre Haute, IN (AQS#181670024); and c,d) Detroit, MI (AQS#261630019) for modeled and observed values indicated by solid and dotted vertical lines respectively, as well as the distribution of the two metrics for the ensemble of model simulations. Both the observation-based RMSD and MD fall within the 95% confidence interval of the ensemble distribution at Terre Haute (indicated by green shading), but the observation-based MD for Detroit is much lower than the model ensemble distribution. The model ensemble was developed based on ±50% uncertainty in area and mobile emissions of NO_x and ±3% uncertainty in point emissions of NO_x.

3. Results

238 *3.1 Base model performance*

239 Before examining the model's response to changes in emissions between 2002 and 2005,
240 model performance for the base case simulations for the two years was assessed (Table 1). In
241 2002, normalized mean error (NME) was 16.6%, and normalized mean bias (NMB) was 0.8%.
242 The results for 2005 showed NME of 17.6% and NMB of 2.6%. These results are similar to the
243 model performance metrics reported in other studies for this domain (Eder and Yu, 2006; Appel
244 et al., 2007).

245 **Table 1. Operational model evaluation for daily maximum 8-hr average ozone concentrations in 2002 and**
246 **2005. Green and red site designation corresponds to the designation shown in Figure 3, and shows the subset**
247 **of the sites that fall within the 95% confidence interval of the model ensemble distribution based on a ±50%**
248 **uncertainty in area and mobile source emissions of NO_x and ±3% point source emissions of NO_x (green), and**
249 **those that did not (red).**

	N	$Mean_{obs}$	$Mean_{model}$	RMSE (ppb)	NME (%)	MB (ppb)	NMB (%)	r
2002 All	**61379**	**54.6**	**55.1**	**11.9**	**16.6**	**0.4**	**0.8**	**0.80**
Green sites	39634	54.8	55.2	11.2	15.8	0.4	0.7	0.81
Red sites	21745	54.4	54.9	13.1	18.3	0.5	0.9	0.77
2005 All	**61126**	**49.9**	**52.5**	**11.4**	**17.6**	**2.6**	**5.2**	**0.75**
Green sites	39456	49.9	52.4	10.9	16.8	2.5	4.9	0.77
Red sites	21670	50.0	52.9	12.4	19.0	2.9	5.8	0.72

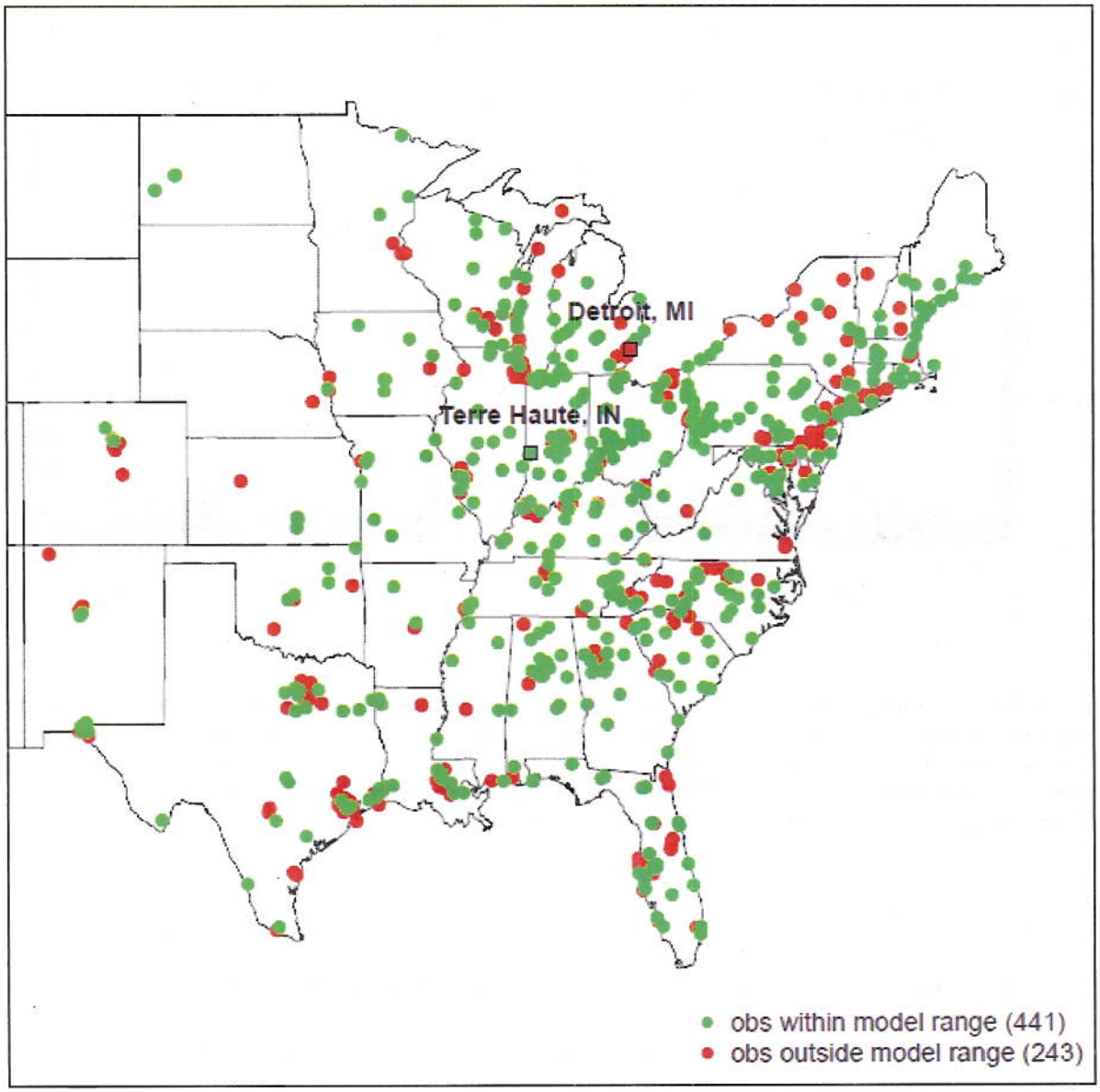
250

251 *3.2 Spatial Distribution of the Dynamic Signal in the Presence of NOx Emissions Uncertainty*

252 Under a moderate amount of emission uncertainty (±50% in area and mobile sources of
253 NO_x and ±3% in point sources of NO_x), the model ensemble was able to capture the observed
254 change in the ozone distribution at a majority of AQS sites (Figure 3). The results were
255 consistent when the RMSD and MD were calculated across all percentiles, $p = 0 - 100\%$, and
256 when using only the upper end of the distribution, $p = 95 - 100\%$. This similarity suggests that
257 the model performs equally well in predicting both high ozone values and overall O_3 when

258 accounting for uncertainties in NO_x emissions inputs. No discernable pattern was found in the
259 spatial distribution of sites for which the model ensemble did not encompass the observed
260 change in the O_3 distribution. Therefore, it is likely that several factors contributed to the poor
261 performance at these sites that are not related to errors in the NO_x emissions. For example, the
262 model has been shown to have errors associated with the transport of O_3 and its precursors in the
263 area of the I-95 corridor in the Northeast (Godowitch et al., 2010). Some of the sites where
264 model performance was poor also included various urban areas where O_3 is frequently less
265 sensitive to NO_x emissions. In fact, the base model performance was relatively poor at the sites
266 where the observed change was not captured by the model ensemble (Table 1).
267 The spread of the ensemble of predicted RSMD and MD values was evaluated using the
268 Talagrand diagram (Hamill, 2001). This evaluation shows that the observed RMSD and MD
269 values tend to fall outside of the range of model predicted values (Figure 4.) The Talagrand
270 diagrams are characterized by a U-shape, indicating a statistically overconfident model (i.e. the
271 range of model values is too narrow). In addition, the observed values are often larger than the
272 largest ensemble member shown by the far right bin of the histogram. This supports the earlier
273 findings that the model response to the changes in NO_x emissions is too low compared to
274 observations (Gilliland et al., 2008). This also provides evidence that the uncertainty ranges for
275 the emission inputs of NO_x were too narrow and/or that other sources of model uncertainty (e.g.
276 boundary conditions, VOC emissions inventory, chemistry, and meteorology), as well as those
277 associated with capturing the inherent variability in point measurements with grid-average model
278 predictions, are still present. These sources of uncertainties have been identified in the past and
279 work has been progressing to address them (McKeen et al., 1991; Hanna et al., 2005; Vautard et
280 al., 2006).

281



282

283 **Figure 3. Geographical extent of the ability of CMAQ to capture the observed change in daily maximum 8-hr**
284 **average ozone between the summers of 2002 and 2005. Green points represent AQS sites that fell within the**
285 **95% confidence interval of the model ensemble distributions of RMSD and MD based on ±50% uncertainty**
286 **in area and mobile source emissions of NO_x and ±3% point source emissions of NO_x. Red points are AQS**
287 **sites where the observed change was not captured by the model ensemble.**

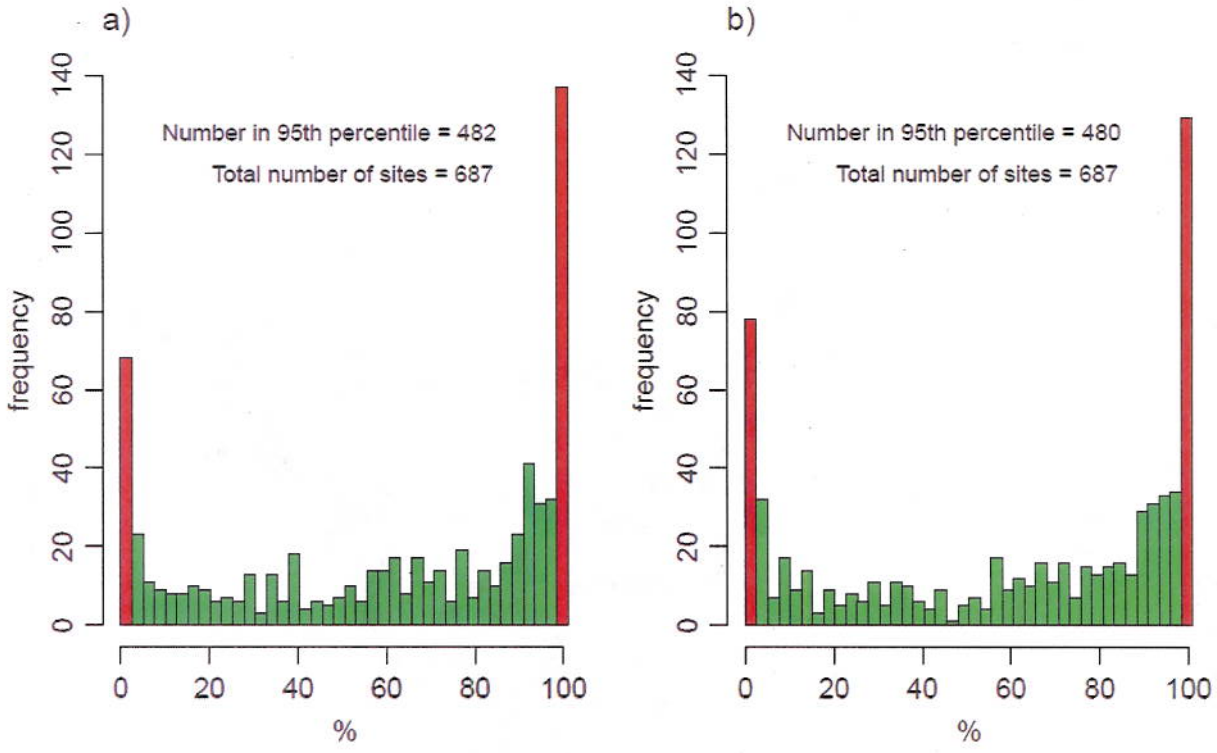


Figure 4. Talagrand diagrams for a) root mean square deviations (RMSD) and b) mean deviations (MD) of the change in daily maximum 8-hr average ozone concentrations between 2002 and 2005. Plots show the percentile of the observed metric with respect to the ensemble of modeled values based on ±50% uncertainty in area and mobile source emissions of NO_x and ±3% point source emissions of NO_x. Each bar represents 2.5% and the green bars show the 95% confidence interval.

3.3 Impact of Uncertainty Assumptions

The presented methodology for evaluating a regional air quality model was designed to allow for flexibility in the assumptions made about the range of uncertainty in the model input parameters. The results presented above were based on the assumption of relatively moderate amount of uncertainty in emissions of area and mobile source NO_x (±50%) and a low amount of uncertainty in point source emissions (±3%). As discussed previously, higher levels of emissions uncertainty have been suggested in the literature. Therefore, the dynamic evaluation presented in section 3.2 was repeated using a range of different uncertainty estimates (Figure 5). As expected, accounting for uncertainty in area and mobile sources of emissions of NO_x

improved the model's ability to capture the observed change in daily maximum 8-hr average O_3. At 100% uncertainty in area and mobile NO_x emissions, the observed change in the ozone distribution was captured at more than 90% of the sites.

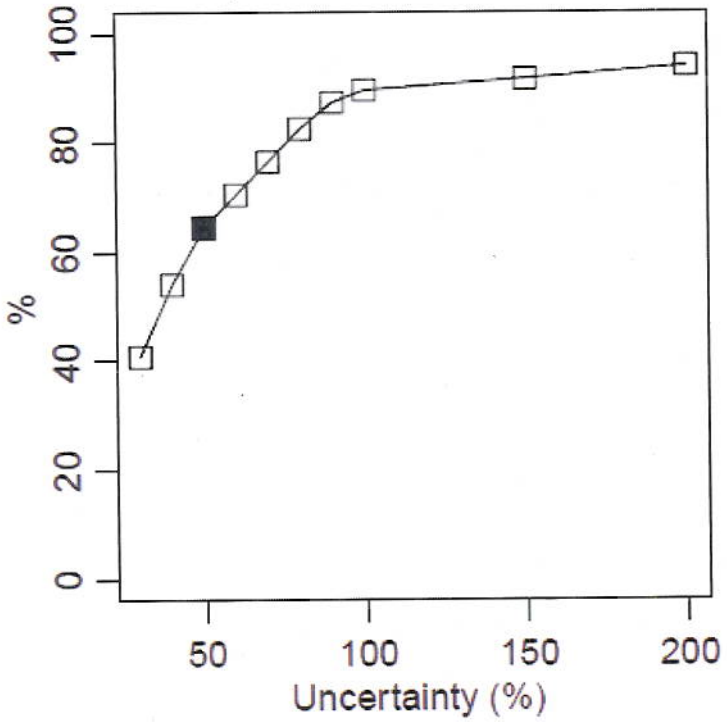
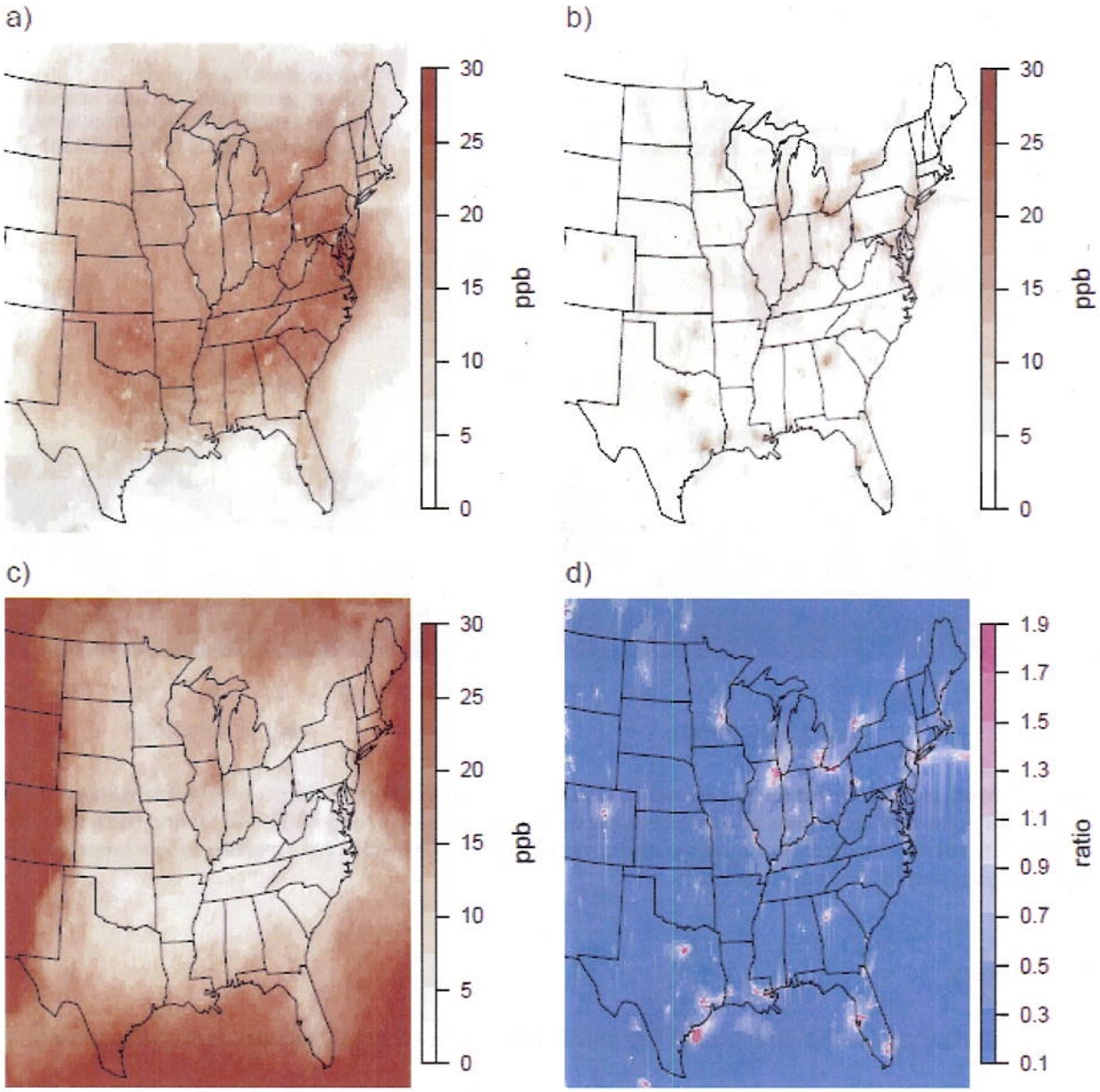


Figure 5. Ability of the model ensemble to capture the observed change in daily maximum 8-hr average O_3 values between the summers of 2002 and 2005 shown as a percentage of all site where the observations fall within the 95th percentile of the model ensemble (y-axis) as a function of the uncertainty in area and mobile sources of emissions of NO_x (x-axis). The uncertainty in area and mobile sources are assumed equal for simplicity of presentation. In each case, the uncertainty in point emissions of NO_x was assumed to be ±3%. The shaded squared indicates the ±50% uncertainty case.

3.3 Sensitivity to VOC Emissions and Boundary Conditions

To investigate possible causes for the narrow range of model predicted values, indicated by the Talagrand diagram (Figure 4), two additional sources of uncertainty that may have influenced the model's ability to accurately predict a change in ozone during this time period were considered. DDM-3D sensitivity of O_3 to emissions of Volatile Organic Compounds (VOC) and to boundary conditions of all modeled species was calculated for a sample two week

322 period in July 2005 (Figure 6). The modeling domain showed several large areas where ozone
323 sensitivity to VOC emissions was greater than the sensitivity to NO_x emissions (Figure 6d)
324 during this period. Higher VOC sensitivity would suggest a VOC-limited ozone formation
325 regime where perturbations in NO_x emissions would have little impact.
326 Many of the sites with observed RMSD and MD metrics that fell outside the estimated
327 95% confidence intervals exhibited higher relative sensitivity to these two additional parameters
328 than to emissions of NO_x. For example, at the poor performing site in Detroit, MI, sensitivity to
329 VOC emissions accounted for a higher contribution to ozone formation than the sensitivity to
330 NO_x emissions (Figure 7) on most days. At the site at Terre Haute, IN, for which the model was
331 able to capture the observed change in O_3 well, the sensitivity to VOC emissions was lower with
332 one notable exception on July 14th. This day was characterized by low predicted O_3
333 concentrations. At both of these sites, as well as all others in the domain, model results were
334 also highly sensitive to boundary conditions. These results help explain the narrower range of
335 the ensemble distributions seen at the Detroit site (Figure 1b), assuming these findings can be
336 extended to the full summer episodes of the two years. Furthermore, since NO_x emissions were
337 greatly reduced in 2005 from their 2002 levels, it is likely that VOC sensitivity became relatively
338 more important in 2005, leading to possible O_3 formation regime changes and transitioning some
339 NO_x-limited regions to become VOC-limited (e.g. the northeastern urban corridor along the
340 interstate 95).
341



342

343 **Figure 6. Average sensitivities of daily maximum 8-hr average ozone to a) emissions of NO_x, b) emissions of**
344 **VOCs, c) boundary conditions, and d) the ratio of VOC to NO_x sensitivity between 3 – 15, July 2005.**
345 **Regions where VOC sensitivity is higher approximate a VOC-limited ozone formation regime.**

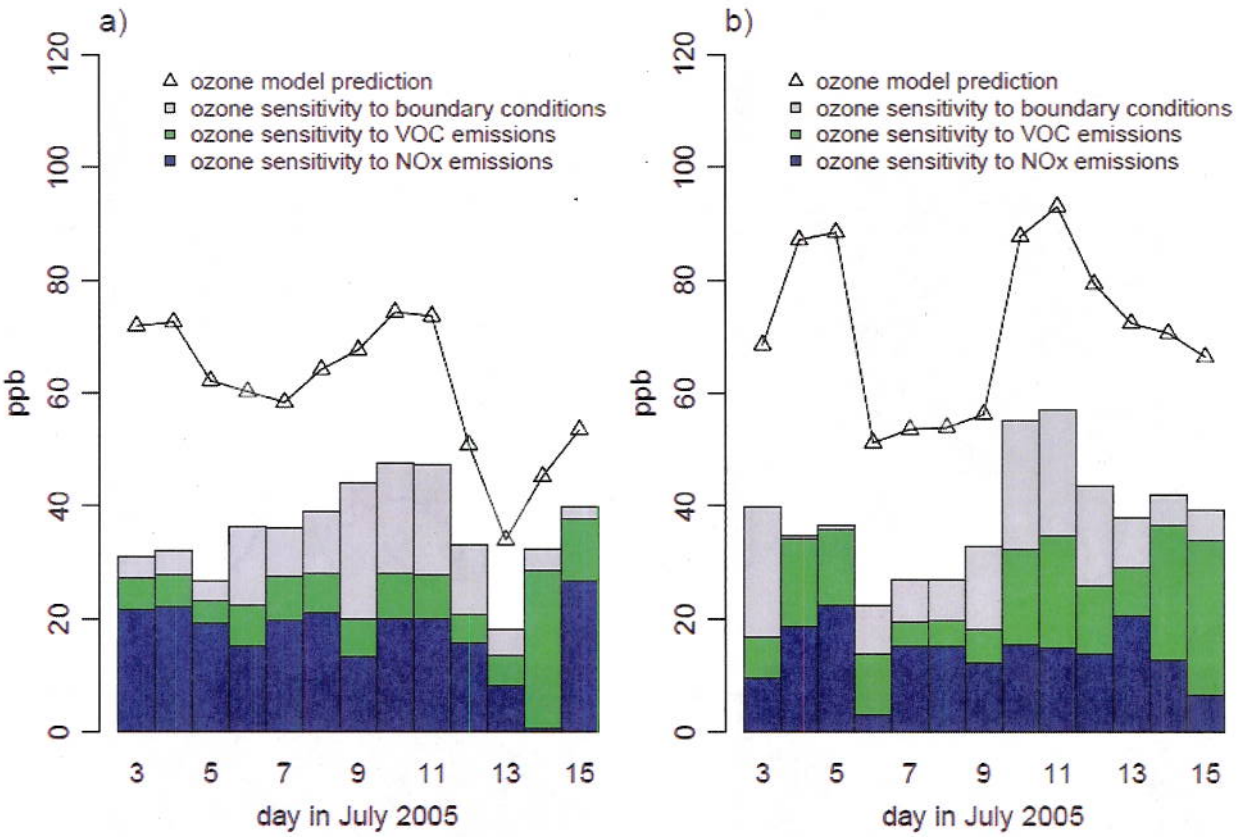


Figure 7. Modeled concentration and sensitivities of daily maximum 8-hr average ozone to emissions of NO_x, VOCs, and boundary conditions at two AQS sites: a) Terre Haute, IN (AQS#181670024); and b) Detroit, MI (AQS#261630019). The Detroit site shows higher sensitivity to VOC emissions resulting in less responsiveness to the propagation of NO_x uncertainty.

4. Summary

An analysis of the change in O_3 concentrations due to large reductions in NO_x emissions in the Eastern U.S. between the summers of 2002 and 2005 was performed while accounting for uncertainty in these emissions. The full distribution of observed and modeled O_3 concentrations was analyzed, as well as the more policy-relevant higher end of the distributions. The difference between the two study years was estimated by comparing modeled and observed O_3 distributions at AQS sites without pairing the data in time. This approach was used to account for the fact that weekly emissions patterns and meteorological drivers are different between the two years. A model performance criterion was developed based on the model's ability to capture the observed

361 O_3 change (estimated by the RMSD and MD metrics), while accounting for uncertainties in NO_x
362 emission inputs. Many standard statistical tests exist for testing for differences between two
363 empirical distributions (e.g. the nonparametric Kolmogorov-Smirnov test). In contrast, the
364 interest in this application is in testing for the difference of differences between two pairs of
365 distributions. Specifically, it is important to test not only whether or not the air quality model
366 can simulate that there is some change in the ozone distribution as a response to changes in
367 emissions, but whether the model can capture the magnitude and direction of this change. This
368 complication motivates the novelty of the proposed test metrics.

369 Assuming ±50% uncertainty in NO_x emissions from area and mobile sources, and ±3%
370 uncertainty in point sources, the ensemble of model predictions was able to capture the observed
371 change in O_3 levels at about two-thirds of AQS sites based on the 95% confidence interval
372 chosen here. Several factors can explain the poor ensemble performance. The main contributor
373 was likely the fact that errors in the dynamic evaluation cannot be attributed to NO_x emissions
374 uncertainty alone. Evidence was found for significant presence of VOC-limited regions in the
375 domain, and VOC emissions are likely to have the same, if not higher, level of uncertainty.
376 Furthermore, boundary conditions were found to have a substantial impact on O_3 formation, and,
377 although untested in this experiment, uncertainty in meteorological inputs was also likely
378 present. Since uncertainty in the emission inventory is difficult to quantify explicitly, it was
379 shown that ensemble performance is sensitive to the selection of the uncertainty levels in
380 emissions of mobile and area sector NO_x.

381 In summary, the methodology presented in this study illustrates the impact of propagating
382 assumed levels of uncertainty in one set (emissions of NO_x) of many uncertain model input
383 variables. It shows that accounting for even one uncertain parameter influences the

384 interpretation of model's responsiveness to changes in emissions as well as base case model
385 performance evaluation. Ideally, a more comprehensive analysis of uncertainties in all model
386 input variables would be extremely useful, and our work is progressing towards this purpose.
387
388 **Disclaimer and Acknowledgements**
389 Although this manuscript has been reviewed and approved for publication, it does not
390 necessarily reflect the policy or views of the US Environmental Protection Agency. The authors
391 would like to thank Jenise Swall, Brian Eder, Rob Pinder, as well the anonymous reviewers for
392 their constructive comments on the manuscript.
393

References

Appel, K. W., Gilliland, A. B., Sarwar, G., Gilliam, R. C. (2007). "Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model performance Part I - Ozone." *Atmos. Environ.* 41(40): 9603-9615.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q. B., Liu, H. G. Y., Mickley, L. J., Schultz, M. G. (2001). "Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation." *J. Geophys. Res.-Atmos.* 106(D19): 23073-23095.

Boynard, A., Beekmann, M., Foret, G., Ung, A., Szopa, S., Schmechtig, C., Coman, A. (2011). "An ensemble assessment of regional ozone model uncertainty with an explicit error representation." *Atmos. Environ.* 45(3): 784-793.

Byun, D., Schere, K. L. (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(1-6): 51-77.

Cullen, A. C., Frey, H. C. (1999). Probabilistic Techniques in Exposure Assessment. Plenum Publishing Corporation. pp 59-78.

Dallmann, T. R., Harley, R. A. (2010). "Evaluation of mobile source emission trends in the United States." *J. Geophys. Res.* 115(D14): D14305.

Deguillaume, L., Beekmann, M., Menut, L. (2007). "Bayesian Monte Carlo analysis applied to regional-scale inverse emission modeling for reactive trace gases." *J. Geophys. Res.* 112(D2): D02307.

Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S. T., Scheffe, R., Schere, K., Steyn, D., Venkatram, A. (2010). "A framework for evaluating regional-scale numerical photochemical modeling systems." *Environ. Fluid Mech.* 10(4): 471-489.

Digar, A., Cohan, D. S. (2010). "Efficient Characterization of Pollutant-Emission Response under Parametric Uncertainty." *Environ. Sci. Technol.* 44(17): 6724-6730.

Eder, B., Yu, S. C. (2006). "A performance evaluation of the 2004 release of Models-3 CMAQ." *Atmos. Environ.* 40(26): 4811-4824.

Gilliland, A. B., Hogrefe, C., Pinder, R. W., Godowitch, J. M., Foley, K. L., Rao, S. T. (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.

Godowitch, J. M., Pouliot, G. A., Rao, S. T. (2010). "Assessing multi-year changes in modeled and observed urban NOX concentrations from a dynamic model evaluation perspective." *Atmos. Environ.* 44(24): 2894-2901.

Grell, G. A., Dudhia, J., Stauffer, D. R. (1994) "A description of the 5th generation Penn State/NCAR Mesoscale Model (MM5)." NCAR Technical Note. NCAR/TN-398+STR.

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

Hamill, T. M. (2001). "Interpretation of rank histograms for verifying ensemble forecasts." *Mon Weather Rev* 129(3): 550-560.

Hanna, S. R., Lu, Z. G., Frey, H. C., Wheeler, N., Vukovich, J., Arunachalam, S., Fernau, M., Hansen, D. A. (2001). "Uncertainties in predicted ozone concentrations due to input

438 uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG
439 domain." *Atmos. Environ.* 35(5): 891-903.
440 Hanna, S. R., Russell, A. G., Wilkinson, J. G., Vukovich, J., Hansen, D. A. (2005). "Monte Carlo
441 estimation of uncertainties in BEIS3 emission outputs and their effects on uncertainties in
442 chemical transport model predictions." *J. Geophys. Res.-Atmos.* 110(D1).
443 Holloway, T., Fiore, A., Hastings, M. G. (2003). "Intercontinental Transport of air pollution:
444 Will emerging science lead to a new hemispheric treaty?" *Environ. Sci. Technol.* 37(20):
445 4535-4542.
446 Konovalov, I. B., Beekmann, M., Richter, A., Burrows, J. P. (2006). "Inverse modelling of the
447 spatial distribution of NOx emissions on a continental scale using satellite data." *Atmos.*
448 *Chem. Phys.* 6: 1747-1770.
449 McKeen, S. A., Hsie, E. Y., Trainer, M., Tallamraju, R., Liu, S. C. (1991). "A Regional Model
450 Study of the Ozone Budget in the Eastern United States." *J. Geophys. Res.* 96(D6):
451 10809-10845.
452 Moran, M. D. (2005) "Current and proposed emission control programs: How will acid
453 deposition be affected?" Environment Canada. 99-162.
454 Morgan, M. G., Henrion, M. (1990). Uncertainty: A Guide to Dealing with Uncertainty in
455 Quantitative Risk and Polic Analysis. Cambridge University Press. pp 172-220.
456 Napelenok, S. L., Cohan, D. S., Odman, M. T., Tonse, S. (2008). "Extension and evaluation of
457 sensitivity analysis capabilities in a photochemical model." *Environ. Modell. Softw.*
458 23(8): 994-999.
459 Napelenok, S. L., Pinder, R. W., Gilliland, A. B., Martin, R. V. (2008). "A method for evaluating
460 spatially-resolved NOx emissions using Kalman filter inversion, direct sensitivities, and
461 space-based NO2 observations." *Atmos. Chem. Phys.* 8(18): 5603-5614.
462 Parrish, D. D. (2006). "Critical evaluation of US on-road vehicle emission inventories." *Atmos.*
463 *Environ.* 40(13): 2288-2300.
464 Pierce, T., Hogrefe, C., Trivikrama Rao, S., Porter, P. S., Ku, J.-Y. (2010). "Dynamic evaluation
465 of a regional air quality model: Assessing the emissions-induced weekly ozone cycle."
466 *Atmos. Environ.* 44(29): 3583-3596.
467 Pinder, R. W., Gilliam, R. C., Appel, K. W., Napelenok, S. L., Foley, K. M., Gilliland, A. B.
468 (2009). "Efficient Probabilistic Estimates of Surface Ozone Concentration Using an
469 Ensemble of Model Configurations and Direct Sensitivity Calculations." *Environ. Sci.*
470 *Technol.* 43(7): 2388-2393.
471 Rao, S. T., Hogrefe, C., Holloway, T., Kallos, G. (2008). Long-range transport of atmospheric
472 pollutants and transboundary pollution. World Atlas of Atmospheric Pollution. Anthem
473 Press. 114p.
474 Office of the Federal Register (1997). Protection of the Environment, 40 CFR Appendix I to Part
475 50. Title 40, Volume 2.
476 Sarwar, G., Luecken, D., Yarwood, G., Whitten, G. Z., Carter, W. P. L. (2008). "Impact of an
477 updated carbon bond mechanism on predictions from the CMAQ modeling system:
478 Preliminary assessment." *J Appl Meteorol Clim* 47(1): 3-14.
479 Tian, D., Cohan, D. S., Napelenok, S., Bergin, M., Hu, Y. T., Chang, M., Russell, A. G. (2010).
480 "Uncertainty Analysis of Ozone Formation and Response to Emission Controls Using
481 Higher-Order Sensitivities." *J. Air Waste Manage. Assoc.* 60(7): 797-804.
482 US-EPA (2005) "Evaluating O3 control programs in the Eastern United States: Focus on the
483 NOx Budget Trading Program." EPA-454-K-05-001, pp 9-14.

Vautard, R., Van Loon, M., Schaap, M., Bergstrom, R., Bessagnet, B., Brandt, J., Builtjes, P. J. H., Christensen, J. H., Cuvelier, C., Graff, A., Jonson, J. E., Krol, M., Langner, J., Roberts, P., Rouil, L., Stern, R., Tarrason, L., Thunis, P., Vignati, E., White, L., Wind, P. (2006). "Is regional air quality model diversity representative of uncertainty for ozone simulation?" *Geophys. Res. Lett.* 33(24).