Efficient Probabilistic Estimates of Surface Ozone Concentration Using an Ensemble of Model Configurations and Direct Sensitivity Calculations

Robert W. Pinder, Robert C. Gilliam, K. Wyat Appel, Sergey L. Napelenok, Kristen M. Foley, Alice B. Gilliland

National Exposure Research Laboratory, U.S. Environmental Protection Agency

Research Triangle Park, North Carolina, USA

ABSTRACT

Because all models are a simplification of the phenomenon they aim to represent, it is often more useful to estimate the probability of an event rather than a single "best" model result. Previous air quality ensemble approaches have used computationally expensive simulations of separately developed modeling systems. We present an efficient method to generate ensembles with hundreds of members based on several structural configurations of a single air quality modeling system. We use the Decoupled Direct Method in 3D (DDM-3D) to directly calculate how ozone concentrations change as a result of changes in input parameters. The modeled probability estimate is compared to observations and is shown to have a high level of skill and improved resolution and sharpness. This approach can help resolve the practical limits of incorporating uncertainty estimation into deterministic air quality management modeling applications.

1. Introduction

Atmospheric chemical and physical processes are complex. Air quality models offer a simplified representation of the fate and transport of air pollutants that can be used to manage and gain insight into air quality problems. Such models rely on parameterization in order to make the mathematical solutions tractable and the results intelligible. Despite these simplifications, air quality models remain computationally intensive. Continental-scale spatial domains are necessary in order to capture long-range transport of ozone. On such a scale, sources of air pollution range from organic gases from natural forests to thousands of automobile tailpipes trapped in suburban congestion. Direct measurement of all of these sources is impossibly expensive, so most applications rely on relatively few observations and many extrapolations.

Because of the necessary parameterizations and extensive data requirements, it is not possible to eliminate uncertainty in air quality modeling. However, air quality modeling is still a critical component of air quality management decision support tools. When weighing the societal benefits of different air quality management options, policy-makers need quantitative information about the relative risks and likelihood of success to guide their decisions. Developing an air quality modeling approach that can estimate both the probability of an event, as well as a single "best-estimate", would advance the current air quality tools available for these management decisions.

Many have shown in air quality forecasting applications that an ensemble of deterministic models can be used to estimate a probabilistic range [1, 2]. Previous work has explored structural uncertainty related to the form of the model [3-6], parametric uncertainty related to the inputs to the model [7] or both [8-11].

The challenge with these approaches is that air quality models require significant input data and computational resources to complete a single simulation. Some of the more successful ensemble results have been generated using a collection of modeling systems developed by independent institutions [3-5]. Developing, maintaining, and meeting the input data requirements for multiple modeling systems is time consuming, expensive, and redundant. Is it possible to achieve similar success with an ensemble of multiple configurations of a single air quality model?

Additionally, traditional ensemble modeling efforts still do not fully address the range of possible uncertainties involving emissions, which is one of the largest parametric uncertainties. Practical constraints have limited the degree to which emission uncertainties, or other parametric uncertainties, can be addressed. For example, the emissions data for a one month of simulation of the Eastern United States domain typically has 26 species, 720 hourly time steps, and a horizontal grid of 205 by 199 cells with 9 vertical layers, yielding on the order of 10^9 input values. These models are computationally intensive, so repeated sampling of the uncertain input space is costly. In theory, to consider uncertainty in all of the inputs, ensemble populations need to be large. Improved sampling methods can reduce the number of simulations [12, 13], but for many air quality practitioners, such additional computational expenses are not feasible. Is it possible to more efficiently create an ensemble of members that reflect the uncertainty in the input data?

In the approach presented here, our goal is to develop an engineering approach that can accurately and efficiently estimate the probability of exceeding an ozone concentration threshold. We use several configurations of a single air quality model and the Decoupled

Direct Method in 3-D (numerous references available in Cohan et al. [14]) to simultaneously calculate the impact of uncertainty in the model form and input values. Finally, we apply methods from the air quality forecasting community and observed ozone concentrations to evaluate the quality ensemble for use as air quality management and decision making tools.

2. Methods

Our method for efficiently developing an ensemble of air quality simulations consists of seven steps shown in Figure 1. The models, relevant datasets, and each of the steps are described in detail below.

2.1 Air Quality Model and Inputs Description

We employ the Community Multiscale Air Quality (CMAQ) model to simulate the chemical concentrations and deposition over a continental-scale region using an Eulerian grid structure [15]. The inputs include chemical emissions and a representation of the atmosphere from a numerical weather simulation model. Important processes are dispersion, gas-phase chemistry, aerosol thermodynamics and mass-transfer, and deposition. The horizontal resolution is 12 x 12 km and there are 14 vertical layers from the surface to 100 hPa. Meteorological inputs are from the PSU/NCAR 5th generation mesoscale model, also known as MM5 [16]. We simulate the time period from June 24, 2002 to July 28, 2002. The first seven days are excluded from the analysis to eliminate sensitivity to initial conditions. The spatial domain includes most of the Eastern United States (Figure S2 in Supplemental Information). Emissions are generated using the SMOKE emissions processing system (http://www.smoke-model.org/version2.3.2/html/ch02s16.html). Year 2002 specific emissions data for motor vehicles are from MOBILE 6 (http://www.epa.gov/otaq/m6.htm); power plant emissions are from Continuous Emission Monitors (http://www.epa.gov/camddataandmaps/). Biogenic volatile organic carbon and NO_x emissions are simulated using BEIS v.3.13 [17] and are derived using the same meteorological fields as the air quality simulations. All other emission sources are from the 2001 National Emission Inventory (http://www.epa.gov/ttn/chief/net/critsummary.html).

The Higher-Order Decoupled Direct Method in three dimensions [14, 18] is implemented in CMAQ version 4.5 (CMAQ-DDM-3D [19]) for CB-IV and SAPRC99 chemical mechanisms. Our study is focused on the Atlanta and Birmingham metropolitan areas; therefore, for the nested CMAQ-DDM-3D simulations, we use a sub-domain over the Southeastern United States (map in Figure 2). This sub-domain includes a region-wide episode of high ozone during the first two weeks and a period of variable clouds and precipitation that cause low ozone concentrations during the third and fourth week. The emission inputs to CMAQ-DDM-3D are as described above; the meteorological inputs, boundary conditions and initial conditions are from the structural uncertainty simulations, described in Section 2.2.

Our analysis focuses on the hourly-average ozone concentrations reported at Air Quality System (AQS) monitoring stations (http://www.epa.gov/air/data/aqsdb.html) in the Birmingham and Atlanta metropolitan areas. The location of these 38 monitors and the location of the 97 monitors used to estimate the uncertainty in the boundary conditions are shown in Section S1 in the Supplemental Information.

2.2 Structural Uncertainty

Structural uncertainty refers to a lack of knowledge about the fundamental mechanisms underlying the environmental process. This kind of uncertainty can be addressed by an ensemble that includes multiple representations of a single process. To build our six structural uncertainty cases, we develop three different configurations of the meteorological model and two different gas-phase chemical mechanisms. After an analysis of previous meteorological model configurations and sensitivity tests, we determined that the choice of land surface model (LSM) and the planetary boundary layer model (PBLM) are among the meteorological factors that have the largest impact on the air quality simulation [20, 21]. These processes control the mixing depth, mixing intensity, and temporal evolution of the surface mixed layer. The three configurations are the Pleim-Xiu (P-X) LSM with the Asymmetric Convective Model (ACM) PBLM, the Noah LSM with Medium Range Forecast (MRF) PBLM, and Noah LSM with Mellor-Yamada-Janjic (M-Y-J) PBLM. Each meteorological simulation is constrained using analysis nudging that employs the three hourly National Centers for Environmental Prediction (NCEP) Eta Data Assimilation System (EDAS) analysis. This reduces simulation errors but restricts the variability between the three cases. We also select two different gas-phase chemical mechanisms, Carbon Bond IV (CB-IV) and SAPRC99. In the CMAQ implementation, these mechanisms have similar representation of the inorganic reactions; however, the grouping and reactivity of the organic gases differ significantly. The supplemental information contains the details of the model configurations, relevant references, and spatial maps of ozone anomalies.

2.3 Parametric Uncertainty

Parametric uncertainty refers to uncertainty in the inputs and parameter values. To address this type of uncertainty, CMAQ-DDM-3D is used to estimate a reduced form model of ozone concentration. While there are many sources of uncertainty, we limit this study to NO_x emissions (ENO_x), VOC emissions (EVOC), and ozone boundary conditions (BO_3). Previous work has shown these to be among the inputs that have the most significant impact on ozone concentration [22, 23]. Our calculations are with respect to temporally-invariant, domain-wide changes in these inputs. We select a range of scaling factors from 0.4-1.2 for uncertainty in NO_x emissions and a range of 0.5-1.5 for uncertainty in VOC emissions. Because these scaling factors represent domain-wide bias in the emission inventory over the entire simulation period, we selected from the lower range of emission uncertainty estimates for a particular source or location as estimated in [24]. Based on a previous study that found that the urban NO_x emissions are over estimated for this domain, NO_x emission uncertainty is not centered at zero [25]. For ozone boundary conditions we select a range of 0.62-1.12, based on the distribution of errors for 97 ozone monitoring sites near the border of our sub-domain, as described in the Supplemental Information.

To estimate the range of ozone concentrations that result from these uncertainties, we use CMAQ-DDM-3D to directly calculate the sensitivity of ozone concentration to a change in the input values. Following the approach and notation of Cohan et al [14], we first define the semi-normalized first-order sensitivity $\mathbf{S}^{(1)}$ and second-order sensitivity $\mathbf{S}^{(2)}$ as

$$\mathbf{S}_j^{(1)}(x,t) = \frac{\partial \mathbf{C}(x,t)}{\partial \varepsilon_j} \qquad (1)$$

$$\mathbf{S}^{(2)}_{j,k}(x,t) = \frac{\partial^2 \mathbf{C}(x,t)}{\partial \varepsilon_j \partial \varepsilon_k} \quad (2)$$

where ε_j is a constant scaling factor applied to the CMAQ inputs, and the subscript *j* and *k* refer to one of ENO_x, EVOC, or BO_3. The ozone concentration matrix, **C**, and sensitivity matrices, **S**, are calculated across the spatial domain, *x*, and for each hour of the simulation, *t*. We use a Taylor Series expansion to estimate the change in ozone concentration due to simultaneous changes in multiple parameters. Equation 3 describes the reduced form model, where *m* refers to BO_3, *j* refers to ENO_x, and *k* refers to EVOC:

$$\mathbf{C}_{\mathrm{j+k+m}} \approx \mathbf{C}_0 + \Delta\varepsilon_{\mathrm{m}}\mathbf{S}^{(1)}_{\mathrm{m}} + \Delta\varepsilon_{\mathrm{j}}\mathbf{S}^{(1)}_{\mathrm{j}} + \Delta\varepsilon_{\mathrm{k}}\mathbf{S}^{(1)}_{\mathrm{k}} + \\ + \frac{1}{2}\Delta\varepsilon_{\mathrm{j}}^2\mathbf{S}^{(2)}_{\mathrm{j}} + \frac{1}{2}\Delta\varepsilon_{\mathrm{k}}^2\mathbf{S}^{(2)}_{\mathrm{k}} + \Delta\varepsilon_{\mathrm{j}}\,\Delta\varepsilon_{\mathrm{k}}\mathbf{S}^{(2)}_{\mathrm{j,k}} \quad (3)$$

Terms higher than second-order are neglected, which is sufficient to reproduce the brute-force simulated ozone concentration to within a few percent [14]. We found that the cross sensitivities of emissions to boundary conditions and second-order boundary condition sensitivities are small, so they are excluded from this analysis.

The first- and second-order sensitivities are calculated using CMAQ-DDM-3D for ENO_x, EVOC, and BO_3 for each time and grid cell and for each of the six structural uncertainty cases over the SE US sub-domain. We use the reduced form model described by Equation 3 to calculate the ozone concentration at each hour and grid cell after the emissions inputs and the boundary conditions from the base model have each been increased or decreased by a constant factor.

2.4 Monte Carlo Methods

A single ensemble member consists of the 8-hour maximum ozone for each location at each day, given a level of NO_x emissions, VOC emissions, ozone boundary conditions, and a specific structural uncertainty case. First, we randomly select one of the six structural uncertainty cases. The NO_x emission, VOC emission, and ozone boundary condition scaling factor are sampled from a uniform distribution that spans the range of uncertainty described in Section 2.3. Using the sensitivities relevant to that structural uncertainty case, we calculate the ozone concentration at each monitoring location for each hour of the simulation. We then repeat this process 40,000 times to build a population of results, and find that a 1,000-member ensemble achieves sufficient convergence.

2.5 Evaluating the Ensemble Quality

To test the properties of this ensemble technique, we devise three test cases:

Structural Ensemble: A six-member ensemble that includes each of the structural uncertainty cases and no parametric uncertainty

Range Ensemble: A 54-member ensemble that includes the six structural cases each with eight combinations of the maximum and minimum range for each of the parametric uncertainty parameters ($2^3 \times 6 = 48$ members) plus an additional six structural case simulations using the central value of each parametric uncertainty distribution: BO_3 = 0.88, ENO_x = 0.8, EVOC = 1.0.

Full Ensemble: A 1,000-member ensemble that includes both structural and parametric uncertainty, generated using Monte Carlo methods described in Section 2.4

For n different parametric uncertainty ranges, the range ensemble requires $2^n + 1$ times more computation time than the structural case. For the full ensemble with α parameters with negligible second derivatives and β parameters with non-negligible second derivatives ($\alpha + \beta = n$), the number of sensitivity calculations is

$$\alpha + 2\beta + \sum_{i=1}^{\beta-1} i \qquad (4)$$

In the full ensemble where $\alpha = 1$ and $\beta = 2$, 6 sensitivity calculations (the **S** terms in Equation 3) are required. These calculations require 6.8 times more computational time than a single CMAQ run. The full ensemble is 1 – (6.8 / 9) = 24% more efficient than the range case; for an ensemble study where $n = \beta = 10$ the full ensemble is an order of magnitude more efficient.

For each of the ensemble members in each case for each monitoring location and day, we calculate the observed and simulated mean concentration from 9 am to 5 pm local time. These values are used for all statistical calculations described here. For each ensemble case at each location and day, we calculate the probability of exceeding the ozone concentration threshold of 68 ppb, which is equal to one standard deviation (14 ppb) greater than the observed mean ozone concentration (54 ppb). As a reference, we calculate the climatological frequency (0.17) as the mean observed frequency of exceeding this threshold over the 38 locations and 28 days in our modeling domain. For a given location and time, the ensemble estimated probability of exceeding the threshold is the number of ensemble members greater than 68 ppb divided by the total number of members. This assumes that each ensemble member has equal likelihood.

The probability that a particular location will exceed a threshold ozone concentration on a specific day can not be directly measured – we can only measure the outcome. To evaluate the model estimated probability, it is useful to define three properties of ensembles frequently used by the meteorological forecasting community: reliability, resolution, and sharpness [2]. *Reliability* refers to the skill of the ensemble-estimated probability. To calculate reliability, we aggregate together all of the locations and times with a similar estimated probability of exceeding a threshold and compare with the observed frequency. That is, for all locations and times where the model estimated probability of exceeding 68 ppb is 20%, is the observed frequency of exceeding 68 ppb at those times and locations also 20%? *Resolution* is a measure of differentiation – how well does the ensemble sort observed events into groups that are different from each other? Lastly, *sharpness* is the extent to which the ensemble deviates from the climatological average. Sharpness is an inherent feature that is not dependent on the observed values. We calculate sharpness as the absolute mean difference between the ensemble-estimated probability of exceeding 68 ppb and the climatological average. In a trivial example, if the ensemble estimated probability always equaled the domain average climatological frequency, then the ensemble estimate would be very reliable, but would not be useful at capturing the day to day and location to location changes. Such an ensemble estimated probability would have poor resolution and poor sharpness. A key challenge is to develop a probabilistic system that expresses all three properties.

When calculated with respect to the climatological probabilities, the Brier Skill Score (BSS) is often used as a scalar representation of the skill of the ensemble. It is expressed as a percentage improvement over climatological probability. The BSS can be calculated as the difference between the resolution score and the reliability score (here a low reliability score implies high reliability). The derivation of this decomposition is described by Wilks [2]. These metrics are used to assess the quality of the three ensemble test cases (structural, range, and full).

Another test of the quality of the uncertainty estimates is the spread-skill relationship, defined as the correlation of the spread of the ensemble with the error in the mean estimate. When the uncertainty is large, the ensemble members should diverge and the spread should increase. We calculate the spread-skill relationship as the correlation between the standard deviation of the ensemble members and the mean absolute error in the ozone concentration.

3. Results

3.1 Ozone Concentrations and Sensitivity

The differences between the structural uncertainty simulations are small. The average ozone concentration anomaly, defined as the difference from the mean of all six structural cases, is shown in the Supplemental Information Figure S2. The largest difference is due to the chemical mechanism: SAPRC99 has higher ozone concentrations compared to CB-IV. The different meteorological options impact the spatial distribution of the ozone. There are also minor structural differences in the sensitivity results, shown in Figure 2. The highest NO_x sensitivity gradients are in and around the urban core of large cities, such as Atlanta. The urban center has considerably lower NO_x emission sensitivity, but higher VOC sensitivity. This feature is more exaggerated in the SAPRC99 chemical mechanism simulations. The absolute value of the VOC emission sensitivities (first-order, second-order, and cross-sensitivities) are largest in the MRF configurations. The diversity in the results is due to structural differences in the simulation of key physical and chemical processes. However, compared to the total uncertainty and un-captured variability, this diversity is small.

Each of the three ensemble test cases (structural, range, and full), have differing properties. For illustration, the observed concentrations and ensemble values for a single ozone monitoring site near Atlanta, Georgia is shown in Figure 3. The structural ensemble (blue open circles) is biased high; over all of the monitoring locations, only 24% of the observations fall within the bounds of the highest and lowest ensemble members. The range ensemble (orange) has the largest spread, with 64% of observed values falling within the 25^{th} and 75^{th} inter-quartile range and 96% are between the highest and lowest members. The full ensemble (grey box-whisker) has less spread than the range case but captures 42% of all observed values within the 25^{th} and 75^{th} inter-quartile range and 94% are between the highest and lowest members.

3.2 Ensemble Evaluation Metrics: Resolution, Reliability, & Sharpness

The reliability diagram (Figure 4) is used to assess the skill of the ensemble estimated probability: for all times and locations when the ensemble estimated probability of exceeding the threshold is *p* (the x-axis), is the observed frequency of exceeding the threshold at those times and locations also *p* (the y-axis)? The x-axis of the diagram is ensemble estimated probability of exceeding the threshold, grouped into seven ascending bins. For each of the times and locations grouped in these bins, the observed frequency of exceeding the threshold is plotted on the y-axis. The grey shaded area represents ensemble estimates that have overall skill greater than a climatological estimate, denoted by the dotted lines as the observed probability of exceeding the threshold (0.17). The structural ensemble (blue) overestimates the probability of exceeding the threshold. The range ensemble (orange) and full ensemble (black) have similarly improved performance. The range ensemble performs better for higher probabilities, while the full ensemble performs better for lower probabilities. Further analysis using the Rank Histogram is available in Section S3 of the Supplemental Information.

The BSS and its relative decomposition into the resolution and reliability components underscore the key differences between the full and range cases. As shown in Table 1, both cases have relatively similar BSS, but the decomposition is quite different. The full case has lower reliability and higher bias, due to asymmetry in the relative impact of changes in BO_3, ENO_x, and EVOC. However, the full case has higher resolution and is also sharper than the range case. With more individual members that further resolve the uncertainty space, the full ensemble can more effectively differentiate between high and low ozone events (Supplemental Information Figure S4). Finally, the full ensemble has the highest spread-skill correlation at 0.45, followed by the structural and range cases with 0.39 and 0.37, respectively.

4. Discussion

We have demonstrated that a single air quality modeling system can combine aspects of structural and parametric uncertainty to provide reliable estimates of the probability of exceeding an ozone threshold concentration. By directly calculating the sensitivities, it is possible to generate large ensembles that have similar skill but greater resolution and sharpness compared to simply simulating the bounds of the uncertainty range. Future work should focus on improving the calibration and reliability, potentially by pairing this method with an ensemble weighting scheme (such as Raftery et al. [26]). Future work should also explore additional sensitivities such as chemical rate parameters and spatially variable emissions sensitivities.

Acknowledgements: We thank Jenise Swall, Tanya Otte, Jonathon Pleim, Steve Howard, Lara Reynolds, Chris Frey, and Haluk Ozkaynak for their helpful comments and suggestions. **Disclaimer**: The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to administrative review and approved for publication.

Supporting Information Available including a description of boundary condition uncertainty, meteorological simulations, and additional evaluation metrics.

Figure 1. Flowchart to generate ensemble including structural and parametric uncertainty using a direct calculation of higher-order sensitivities

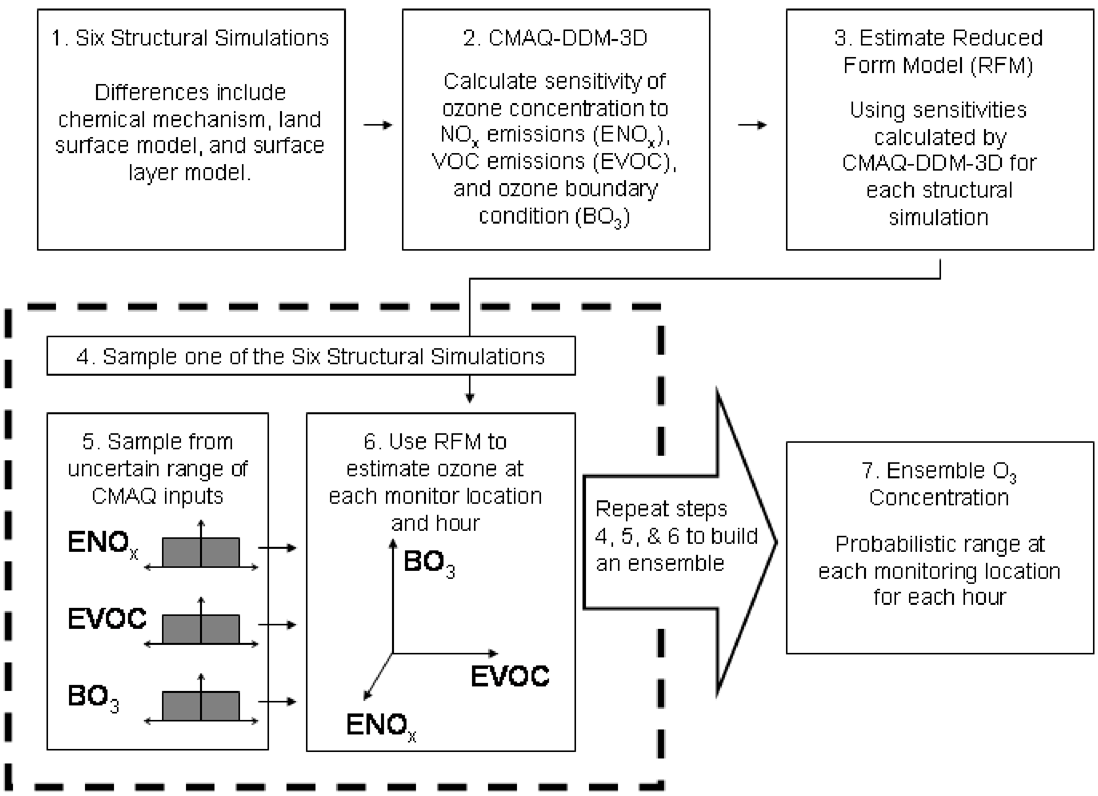

Figure 2. The first-order ($S_j^{(1)}$), second-order ($S_j^{(2)}$), and cross-sensitivity ($S_{jk}^{(2)}$) of the ozone concentration to ozone boundary concentrations (BO_3), NO_x emissions (ENO_x), and VOC emissions (EVOC) for each of the six structural uncertainty cases, averaged over the entire simulation period.

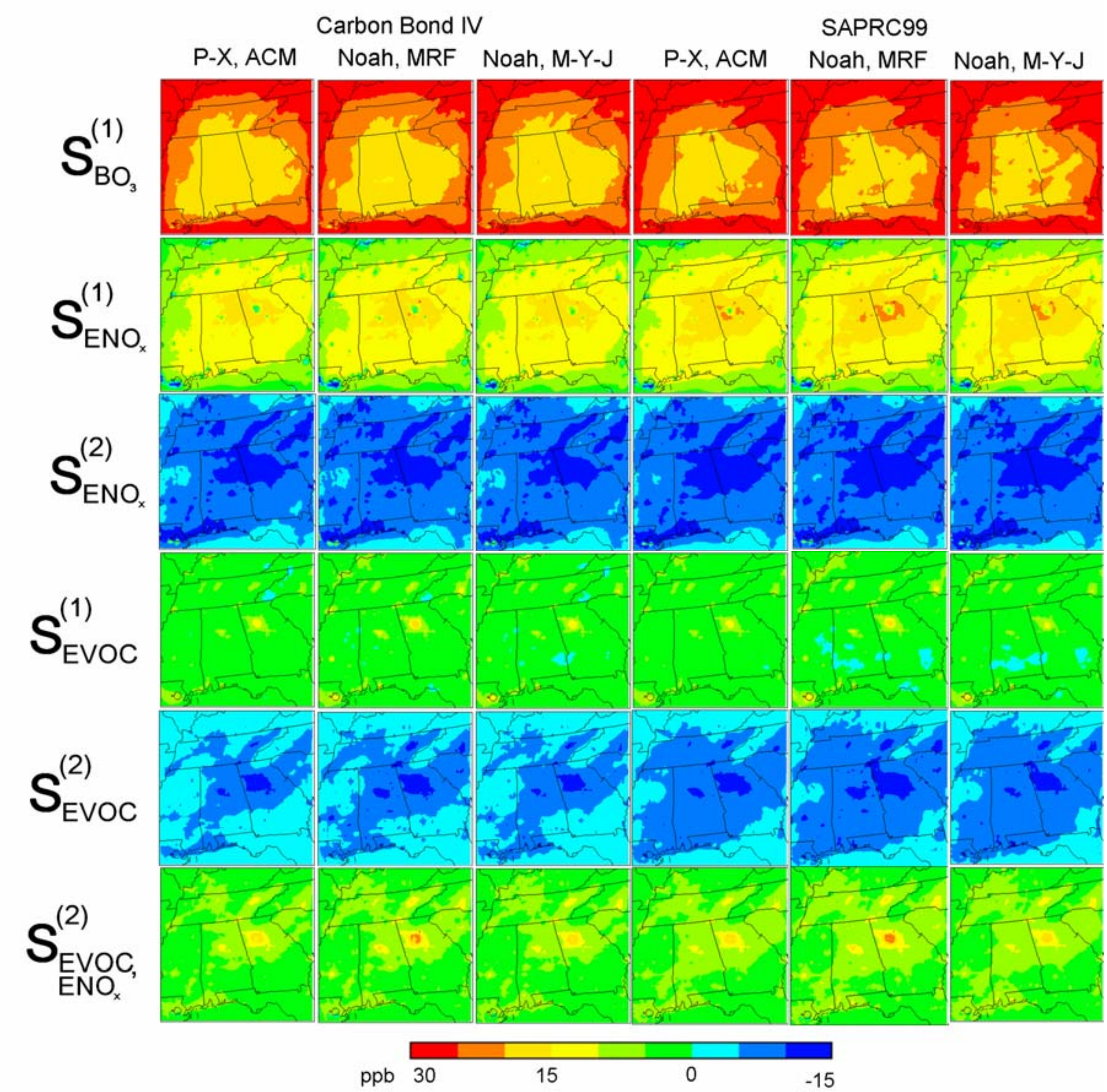

Figure 3. For a location near sub-urban Atlanta, observed (red) and ensemble ozone concentrations as estimated by including only structural uncertainty (blue), both structural and the bounds of the parametric uncertainty (orange: box = 25th and 75th percentile, crosses = range), and 1,000 member ensemble composed of sampling both the structural cases and from the range of uncertain input parameters (grey: box = 25th and 75th percentile, dashed line = range, open circles = outliers).

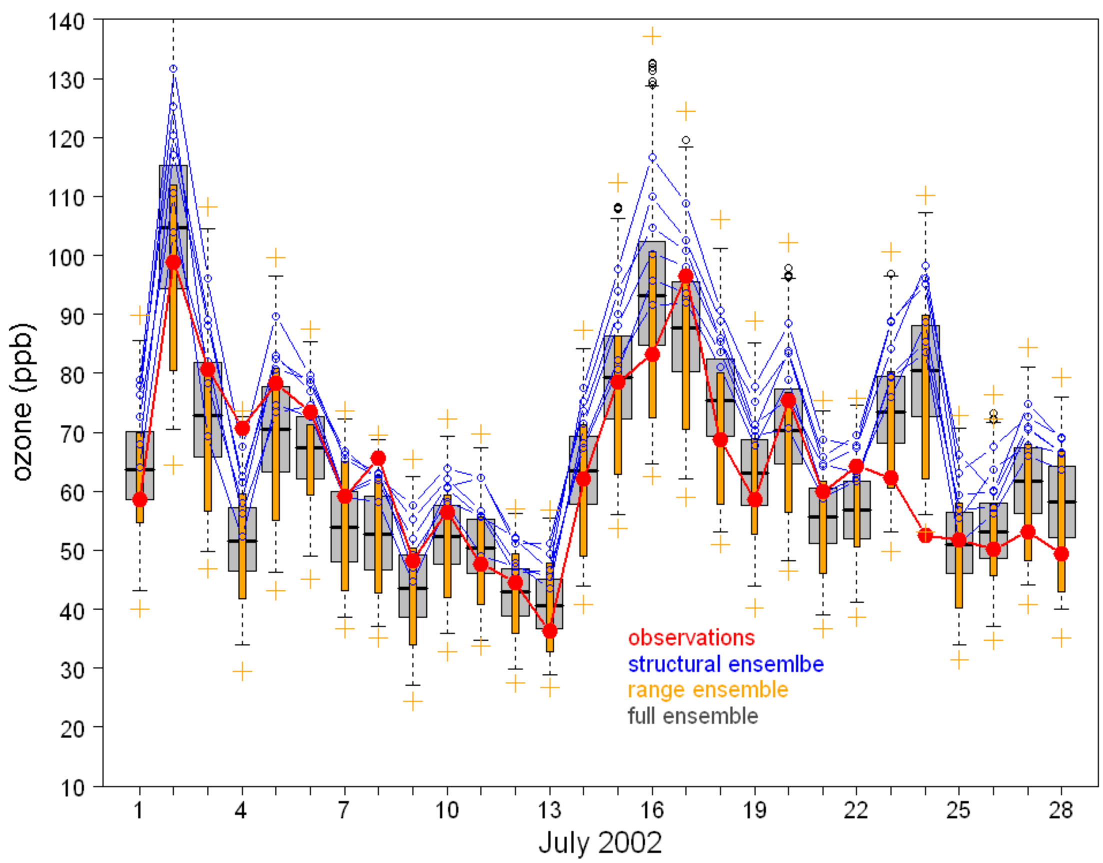

Figure 4. The Reliability Diagram

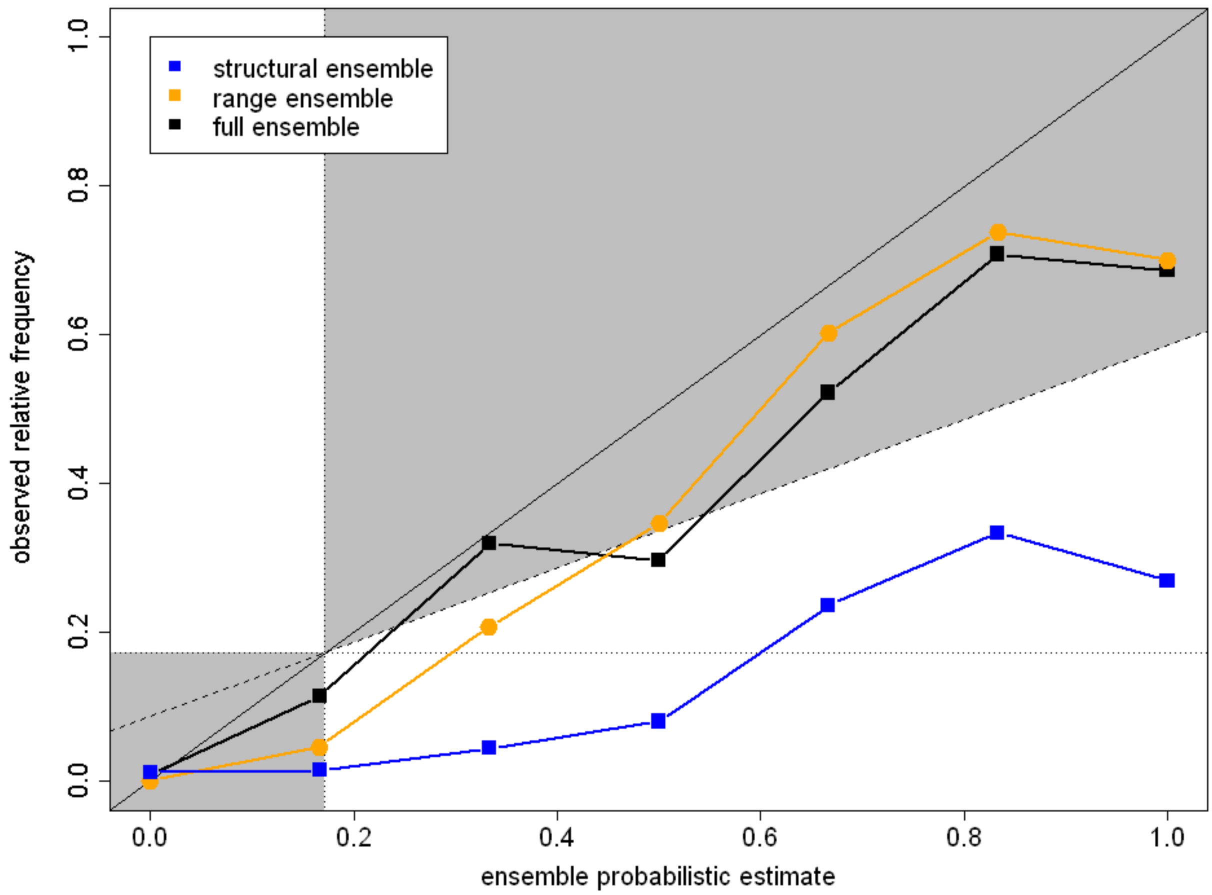

Table 1. Ensemble performance metrics

Case	Structural Ensemble	Range Ensemble	Full Ensemble
Mean Correlation	0.73	0.70	0.71
Mean Error (ppb)	11	1.3	3.2
Mean Standard Deviation (ppb)	4.2	11	7.1
Spread-skill Correlation	0.39	0.37	0.45
Brier Skill Score	-0.15	0.36	0.35
Reliability	0.47	0.019	0.14
Resolution	0.32	0.38	0.49
Sharpness	0.36	0.19	0.24

References:

1. Leith, C. E., Theoretical Skill of Monte Carlo Forecasts. *Monthly Weather Review* **1974,** *102*, (6), 9.

2. Wilks, D. S., *Statistical Methods in the Atmospheric Sciences*. Elsevier: 2006; p 1.

3. van Loon, M.; Vautard, R.; 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., Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble. *Atmospheric Environment* **2007,** *41*, (10), 2083-2097.

4. 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., Is regional air quality model diversity representative of uncertainty for ozone simulation? *Geophysical Research Letters* **2006,** *33*, (24).

5. McKeen, S.; Wilczak, J.; Grell, G.; Djalalova, I.; Peckham, S.; Hsie, E. Y.; Gong, W.; Bouchet, V.; Menard, S.; Moffet, R.; McHenry, J.; McQueen, J.; Tang, Y.; Carmichael, G. R.; Pagowski, M.; Chan, A.; Dye, T.; Frost, G.; Lee, P.; Mathur, R., Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004. *Journal of Geophysical Research-Atmospheres* **2005,** *110*, (D21).

6. Mallet, V.; Sportisse, B., Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach to ozone modeling. *Journal of Geophysical Research -- Atmospheres* **2006,** *111*.

7. Zhang, F. Q.; Bei, N. F.; Nielsen-Gammon, J. W.; Li, G. H.; Zhang, R. Y.; Stuart, A.; Aksoy, A., Impacts of meteorological uncertainties on ozone pollution predictability estimated through meteorological and photochemical ensemble forecasts. *Journal of Geophysical Research-Atmospheres* **2007,** *112*, (D4).

8. Delle Monache, L.; Hacker, J. P.; Zhou, Y. M.; Deng, X. X.; Stull, R. B., Probabilistic aspects of meteorological and ozone regional ensemble forecasts. *Journal of Geophysical Research-Atmospheres* **2006,** *111*, (D24).

9. Mallet, V.; Sportisse, B., Ensemble-based air quality forecasts: A multimodel approach applied to ozone. *Journal of Geophysical Research-Atmospheres* **2006,** *111*, (D18).

10. Delle Monache, L.; Deng, X. X.; Zhou, Y. M.; Stull, R., Ozone ensemble forecasts: 1. A new ensemble design. *Journal of Geophysical Research-Atmospheres* **2006,** *111*, (D5).

11. Delle Monache, L.; Stull, R. B., An ensemble air-quality forecast over western Europe during an ozone episode. *Atmospheric Environment* **2003,** *37*, (25), 3469-3474.

12. McKay, M. D.; Conover, W. J.; Beckman, R. J., A comparison of three methods for selection values of input variables in the analysis of output from a computer code. *Technometrics* **1979,** *22*, (2), 6.

13. Tatang, M. A.; Pan, W. W.; Prinn, R. G.; McRae, G. J., An efficient method for parametric uncertainty analysis of numerical geophysical models. *Journal of Geophysical Research-Atmospheres* **1997,** *102*, (D18), 21925-21932.

14. Cohan, D. S.; Hakami, A.; Hu, Y. T.; Russell, A. G., Nonlinear response of ozone to emissions: Source apportionment and sensitivity analysis. *Environmental Science & Technology* **2005,** *39*, (17), 6739-6748.

15. Byun, D. W.; Schere, K. L., Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Reviews* **2006,** *59*, 51-77.

16. Grell, G. A.; Dudhia, J.; Stauffer, D. R. *A Description of the Fifth-generation Penn State/NCAR Mesoscale Model (MM5)*; 1994; p 122.

17. Schwede, D.; Pouliot, G.; Pierce, T., Changes to the Biogenic Emissions Inventory System Version 3 (BEIS3). In *4th CMAS Models-3 Users' Conference*, Chapel Hill, North Carolina, USA, 2005.

18. Hakami, A.; Odman, M. T.; Russell, A. G., High-order, direct sensitivity analysis of multidimensional air quality models. *Environmental Science & Technology* **2003,** *37*, (11), 2442-2452.

19. Napelenok, S. L.; Cohan, D. S.; Odman, M. T.; Tonse, S., Extension and evaluation of sensitivity analysis capabilities in a photochemical model. *Environmental Modelling and Software* **2008,** 23 (8), 994-999

20. Gilliam, R. C.; Hogrefe, C.; Rao, S. T., New methods for evaluating meteorological models used in air quality applications. *Atmospheric Environment* **2006,** *40*, (26), 5073-5086.

21. Hogrefe, C.; Rao, S. T.; Kasibhatla, P.; Kallos, G.; Tremback, C. J.; Hao, W.; Olerud, D.; Xiu, A.; McHenry, J. N.; Alapaty, K., Evaluating the performance of regional-scale photochemical modeling systems: Part I -- meteorological predictions. *Atmospheric Environment* **2001,** *35*, 15.

22. Hakami, A.; Seinfeld, J. H.; Chai, T. F.; Tang, Y. H.; Carmichael, G. R.; Sandu, A., Adjoint sensitivity analysis of ozone nonattainment over the continental United States. *Environmental Science & Technology* **2006,** *40*, (12), 3855-3864.

23. Martien, P. T.; Harley, R. A., Adjoint sensitivity analysis for a three-dimensional photochemical model: Application to Southern California. *Environmental Science & Technology* **2006,** *40*, (13), 4200-4210.

24. Miller, C. A.; Hidy, G.; Hales, J.; Kolb, C. E.; Werner, A. S.; Haneke, B.; Parrish, D.; Frey, H. C.; Rojas-Bracho, L.; Deslauriers, M.; Pennell, B.; Mobley, J. D., Air emission inventories in North America: A critical assessment. *Journal of the Air & Waste Management Association* **2006,** *56*, (8), 1115-1129.

25. Napelenok, S.N.; Pinder R.W.; Gilliland, A.B.; Martin, R.V. A method for evaluating spatially-resolved NOx emissions using Kalman filter inversion, direct sensitivities, and space-based NO2 observations. *Atmospheric Chemistry and Physics Discussions* **2008**, 8, 6469-6499

26. Raftery, A. E.; Gneiting, T.; Balabdaoui, F.; Polakowski, M., Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review* **2005,** *133*, (5), 1155-1174.