Bayesian analysis of a reduced-form air quality model

Kristen M Foley,[*,†] Brian J. Reich,[‡] and Sergey L. Napelenok[†]

National Exposure Research Laboratory, US Environmental Protection Agency, RTP, NC, USA, and Department of Statistics, North Carolina State University, Raleigh, NC, USA

E-mail: foley.kristen@epa.gov

Abstract

Numerical air quality models are being used for assessing emission control strategies for improving ambient pollution levels across the globe. This paper applies probabilistic modeling to evaluate the effectiveness of emission reduction scenarios aimed at lowering ground-level ozone concentrations. A Bayesian hierarchical model is used to combine air quality model output and monitoring data in order to characterize the impact of emissions reductions while accounting for different degrees of uncertainty in the modeled emissions inputs. The probabilistic model predictions are weighted based on population density in order to better quantify the societal benefits/disbenefits of four hypothetical emission reduction scenarios in which domain-wide NO_x emissions from various sectors are reduced individually and then simultaneously. Cross validation analysis shows the statistical model performs well compared to observed ozone levels. Accounting for the variability and uncertainty in the emissions and atmospheric systems being modeled is shown to impact how emission reduction scenarios would be ranked, compared to standard methodology.

[*]To whom correspondence should be addressed
[†]National Exposure Research Laboratory, US Environmental Protection Agency, RTP, NC, USA
[‡]Department of Statistics, North Carolina State University, Raleigh, NC, USA

Introduction

The United States Environmental Protection Agency (US EPA) sets national ambient air quality standards (NAAQS) for six major pollutants to protect against their harmful effects on human health (primary standards) and vegetation, ecosystems, visibility, climate, etc. (secondary standards). Similar regulations exist in Europe and elsewhere. One of the six major pollutants regulated by the NAAQS is ozone. Human exposure studies have repeatedly shown compelling evidence of human health effects (e.g. increased medical visits for asthmatics) directly attributable to acute exposures to ozone (*1*). In addition, satellite data and model simulations have recently been used to estimate the annual damage to US crops due to an increase in background surface ozone levels to be in the hundreds of millions of dollars (*2*, *3*). Such findings have led EPA to propose stricter ozone standards which would require states to develop and implement emission control strategies for lowering ambient ozone levels. Since environmental regulations to improve air quality can be costly, federal, state and local agencies use deterministic air quality models (AQM) to compare the potential gain in public health and environmental protection from different control strategy options, prior to implementing them.

Deterministic AQM allow scientists and policy makers to explore "what if" scenarios, such as the impact of more stringent fuel economy standards for motor vehicles. While there have been significant advances in the sophistication and accuracy in air quality modeling in past decades, AQM will always be based on imperfect knowledge of the atmosphere and pollutant emissions. Furthermore, certain aspects of the modeling system are inherently more uncertain than others. For example, estimated NO_x emissions from power plants are typically based on direct observations, whereas VOC emissions from anthropogenic sources are known with much less accuracy (*4*). Thus, the impact of these data gaps on the decision-making process needs to be transparent and clearly communicated.

Probabilistic modeling can be used to quantitatively incorporate information on uncertainty and/or variability in the AQM outputs in order to improve the information leading to a particular decision. Rather than using a single AQM simulation as the "best estimate" for a specific outcome (e.g. bright line test of whether or not an area will attain the national ozone standard under a given emission control), a probabilistic approach can provide information on the probability of the event occurring (*5*). A probabilistic modeling approach can provide insight on whether a particular managment option is more likely to be successful compared to another and by how much. Implementing probabilistic methods in a regional or global-scale air quality modeling system is particularly challenging, because the models are highly complex and require input from multiple other deterministic models (e.g. meteorological models, global models for boundary conditions, models of anthropogenic and biogenic emissions).

The first step in developing a probabilistic modeling approach is to identify and quantify the main sources of uncertainty in model inputs and options. Several studies have used sensitivity analyses to determine what model parameters and inputs have the largest impact on modeled ozone concentrations (*6–8*). These findings and the development of reduced form air quality models (RFMs) have been used to efficiently generate probabilistic estimates of ozone levels based on an ensemble of model predictions that represent uncertainty in emissions and meteorology inputs, model settings, and boundary conditions (*4*, *9*). Evaluation of these probabilistic estimates against observations demonstrates that even assuming large uncertainty in specific model inputs does not necessarily account for the differences between modeled and observed ozone levels (*9*, *10*). A limitation of these approaches is that the uncertainty attributed to inputs is often a simplification or approximation. In addition, it is not computationally feasible to simulate the ozone values under all possible, scientifically valid model configurations even when using RFMs.

Here we extend upon previous RFM approaches by using observational data and a Bayesian statistical model to calibrate probabilistic estimates generated from a reduced form regional air quality model. The uncertainty in specific emissions inputs and boundary conditions is updated based on ozone observations. Posterior estimates of ozone concentration are then used to compare the probability of success of different NO_x emission reduction scenarios in attaining the ozone NAAQS for a case study in the southeast US during the summer of 2005. Results are also weighted by population density to better quantify the societal benefits of the different emissions reductions in terms of human exposure.

Methods

Modeling system and observations

The Community Multiscale Air Quality (CMAQ) model (*11*) version 4.7.1 (*12*) was used to simulate ozone concentrations in the southeastern United States. The model simulation was run from July 1, 2005 to September 30, 2005 using an Eulerian grid structure with a 12 km by 12km horizontal grid spacing and 14 vertical layers from the surface to 100 hPa. The inputs include meteorological fields developed using the fifth generation mesoscale model (MM5) version 3.6.3 (*13*) and anthropogenic pollutant emissions based on the 2001 National Emissions Inventory (NEI; `http://www.epa.gov/ttn/chief/emch/index.html\#2001`) processed using the SMOKE processor, version 2.3.2 (`http://www.smoke-model.org`). To estimate emissions for 2005, the 2001 NEI was updated with year 2005 specific emissions data for electric generating units equipped with Continuous Emission Monitoring systems (CEMS), mobile emissions processed by MOBILE 6 (`http://www.epa.gov/otaq/m6.htm`), and meteorologically adjusted biogenic emissions from the Biogenic Emission Inventory System (BEIS) 3.13 (*14*). The Higher-Order Decoupled Direct Method in three dimensions (DDM-3D) (*15*) is implemented in CMAQ for the Statewide Air Pollution Research Center (SAPRC99) gas-phase chemical mechanism (*16*). Boundary and initial conditions were specified from the output of a simulation with 36km grid resolution covering the entire contiguous US.

Ozone predictions from the CMAQ model are paired in time and space with hourly average ozone observations obtained from EPA's Air Quality System (AQS; `http://www.epa.gov/ttn/airs/airsaqs/`). This analysis focuses on the daily maximum eight-hour average ozone concentrations (MD8 O3) at 307 monitoring stations in the southeastern US. The MD8 is the averaging metric of interest, because it is used for determining compliance with the EPA's ozone standards. Specifically, the current EPA standard for ozone is based on the average of the annual fourth highest MD8 ozone concentration across three summers, referred to as the ozone design value. The results presented here focus on the ozone levels for one summer, but could easily be extended to multiple years for a regulatory application.

Reduced-form CMAQ model

A reduced form of the CMAQ model was developed based on sensitivity coefficients calculated using DDM-3D (*17*) updated for the CMAQ model version 4.7.1. DDM-3D was used to calculate first-order and second-order semi-normalized sensitivity coefficients to perturbations in a set of input parameters. Taylor series expansion was then applied to approximate the ozone concentrations, C, as a function of perturbations in the set of chosen parameters. Generally, the reduced-form CMAQ model for one species and d parameters can be represented by:

$$C(\mathbf{s},t|\alpha) \quad = \quad C_0(\mathbf{s},t)+\sum_{j=1}^{d}\alpha_j S_j^{(1)}(\mathbf{s},t)+\frac{1}{2}\sum_{j=1}^{d}\alpha_j^2 S_j^{(2)}(\mathbf{s},t)+\sum_{k<j}\alpha_j\alpha_k S_{jk}^{(12)}(\mathbf{s},t) \tag{1}$$

where $C(\mathbf{s},t|\alpha)$ is the species concentration due to a specific set of perturbations $\alpha = (\alpha_1,\ldots,\alpha_d)$ at location s on day t, $C_0(\mathbf{s},t)$ is unperturbed concentrations from the base simulation, α_j is the perturbation in input parameter j, and $S_j^{(1)}(\mathbf{s},t)$, $S_j^{(2)}(\mathbf{s},t)$, $S_{jk}^{(12)}(\mathbf{s},t)$ are first-order, second-order, and cross-sensitivities, respectively. For example for a 10% decrease in NO_x emissions, $\alpha_j = -0.10$. In application, the coefficients α can represent a combination of input uncertainty and emissions reductions from a control as detailed in the Results section below. More details on the use of sensitivity coefficients in Taylor series expansions including examples for other air quality applications can be found elsewhere (*4*, *10*).

The sensitivity coefficients produced by DDM-3D vary in space and time, providing a computationally efficient calculation of ozone under different perturbations in emissions inputs through the RFM. For example, in urban centers, NO_x emissions frequently act as a sink of ozone resulting in negative sensitivity to sectors involving NO_x emissions (see supplemental Figure S1). In this analysis we consider sensitivity to d = 6 inputs: mobile-source NO_x emissions (MNO_x), point-source NO_x emissions (PNO_x; e.g. power plants, industrial boilers), other NO_x emissions (ONO_x; e.g. construction equipment, large ships, biogenic soil sources), anthropogenic VOCs emissions from all sectors (AVOC), biogenic VOCs emissions (BVOC), and ozone boundary conditions (O3 BC).

Statistical model

A Bayesian hierarchical framework can be used to estimate the optimal perturbations in emission and boundary conditions in (1) for accurately predicting observed ozone concentrations while accounting for additional sources of bias and error in the deterministic model. Let $y(\mathbf{s},t)$ be the observed AQS ozone measurement at spatial location $\mathbf{s}$ on day t. We model $y(\mathbf{s},t)$ as:

$$y(\mathbf{s},t) = \beta_0(\mathbf{s}) + \beta_1(\mathbf{s})C(\mathbf{s},t|\alpha) + w(\mathbf{s},t), \tag{2}$$

where $\beta_0(\mathbf{s})$ and $\beta_1(\mathbf{s})$ are additive and multiplicative biases, respectively, $C(\mathbf{s},t|\alpha)$ is the reduced-form CMAQ model in (1) where the sensitivity coefficients are treated as known covariates and the α_js are unknown scaling factors, and the residual errors are independently and identically distributed as $w(\mathbf{s},t) \sim \mathrm{N}(0,\sigma_e^2)$. The spatial bias terms account for systematic differences between the CMAQ output and observed data during these summer months. For example, errors in the meteorological inputs (e.g. a temperature bias) can induce systematic biases in the ozone predictions. In addition, there may be other discrepancies, because the CMAQ output represents a grid cell average, while the AQS observation measures ozone at a point location. The bias terms are modeled as spatial Gaussian processes with mean $\mathrm{E}[\beta_i(\mathbf{s})] = b_i$ and stationary exponential spatial covariance $\mathrm{Cov}[\beta_i(\mathbf{s}),\beta_i(\mathbf{s}')] = \tau_i^2 \exp(-||\mathbf{s}-\mathbf{s}'||/\rho_i)$. We denote this spatial model as $\beta_i \sim \mathrm{GP}(b_i,\tau_i^2,\rho_i)$, $i = 0,1$.

We use uninformative priors for the model's hyperparameters to determine the amount of information in the data about these parameters. The means have priors $b_i \sim \mathrm{N}(0,100^2)$ and the variances have priors τ_i^2, $\sigma_e^2 \sim$ InvGamma(0.1,0.1). After scaling the spatial locations to $[0,1]^2$, the spatial range parameters have priors $\log(\rho_i) \sim$ N(-2,1), to have prior median 0.14 and prior 95% interval (0.02,0.96). The scaling factor hyperparameters have priors $\alpha_j \overset{iid}{\sim} \mathrm{N}(0, 10^2)$. (Note that sensitivity tests showed that the final model fit was not sensitive to the choice of prior for the α_js.) We also compare a set of RFM results using informative priors for the α_js based on previous studies. This is discussed further in the following section.

The Bayesian hierarchical model is fit using Markov chain Monte Carlo (MCMC) (*18*) sampling implemented in the statistical language `R` (`http://www.r-project.org/`). The likelihood parameters β_0 and β_1 have Gaussian full conditionals and are updated as blocks using Gibbs sampling. Similarly, the hyperparameters b_i, τ_i^2, σ_e^2 are updated from their full conditionals using Gibbs sampling (*19*). The scaling factors α and spatial ranges $\log(\rho_i)$ do not have conjugate full conditionals and are updated using Metropolis sampling. The algorithm is tuned following the recommendations in (*20*). For the analysis in the Results section below, we have sampled 25,000 draws and have discarded the first 10,000 as burn-in. Convergence was monitored using trace plots and autocorrelations of several representative parameters. For the full model, this algorithm takes a few hours on an ordinary PC.

Model Comparisons

Bayesian model comparison and validation

Using the modeled and observed data described above, we fit several special cases of the statistical model to determine the effect of each model component on predictive performance (Table 1). We compare the fits of the base CMAQ model with and without the residual error term (Models 1, 2). We also compare a RFM using informative priors for α based on previous studies and without any Bayesian updating, i.e., using samples from the informative prior directly as predictions (Model 3). Following Digar et al. (*4*), log Normal distributions were used (e.g. for mobile NO_x sources: $\alpha_{eMNO_X} + 1 \sim$ logN(0,σ_u), for uncertainty level σ_u). The uncertainty parameters used to sample α were: 0.25, 0.15, and 0.50 in mobile, point and all other NO_x sources, respectively (*21*); 0.33, and 0.40 in anthropogenic and biogenic VOCs, respectively (*4*); 0.15 in ozone boundary conditions (based on model - observed differences at the AQS monitoring sites near the boundary of the model simulation). Finally, we apply several versions of the Bayesian RFM described in the previous section by fitting the model with no bias ($\beta_0 = 0$, $\beta_1 = 1$), constant bias, and spatially varying bias terms (Models 4-6).

Table 1 presents five-fold cross-validation results comparing the six models. For each of the five subsamples, or folds, complete days are randomly selected to be withheld from the model fitting as testing data. Using this approach, all observations are used for both training and validation, with each observation being used for validation exactly once. For each site/day in the training dataset, the MCMC algorithm produces several equally-likely samples for each parameter in the statistical model. For each sample, we make a prediction following Equation (2). This provides many samples from the predictive distribution for $y(\mathbf{s},t)$, which are used to summarize various characteristics of the posterior distribution. For example, a 90% prediction interval is computed by taking the 5th and 95th percentiles of these samples.

We compute the root mean squared error RMSE= $(\sum_t \sum_j [\hat{y}(\mathbf{s}_j,t) - y(\mathbf{s}_j,t)]^2/N)^{\frac{1}{2}}$, where $y(\mathbf{s}_j,t)$ are observed values, $\hat{y}(\mathbf{s}_j,t)$ are the posterior predictive means for the fit without data from day t, and N is the total number of observations in the dataset. Mean bias is calculated in a similar fashion. We also compare the average predictive standard deviation and the coverage probabilities of 90% prediction intervals for the testing observations. To compare classification of extreme events, we compute the Brier score for exceedance probabilities for a threshold c = 75ppb, that is,

$$BS_c = \frac{1}{N}\sum_t \sum_j (P[y^*(\mathbf{s}_j,t) > c] - I[y(\mathbf{s}_j,t) > c])^2 \tag{3}$$

where $P[y^*(\mathbf{s}_j,t) > c]$ is the posterior predictive probability that $y(\mathbf{s}_j,t) > c$ and $I[y(\mathbf{s}_j,t) > c]$ equals 0 or 1 indicating whether or not the corresponding observation was greater than c. Models with small Brier scores are preferred.

The comparison in Table 1 reveals several things. First, the base CMAQ model (Model 1) tends to over estimate ozone concentrations for this domain and time period. Adding mean zero independent Gaussian errors to the base model (Model 2) can be used to create prediction intervals with proper coverage (COV90=90) but that are very wide (SD=10.16) and still biased high. Ideally we wish to have estimates that are well calibrated with low variance, so that they provide reliable information with low uncertainty. The RFM based on informative priors with no Bayesian updating (Model 3) retains the high bias seen in Models 1 and 2. Model 3 does have the best Brier score for a threshold of 75ppb, the current ozone NAAQS, however it is very poorly calibrated, with only 56% of the cross validation data falling within the 90% predictive intervals. This reinforces the idea that there are remaining sources of uncertainty and/or bias not captured by the uncalibrated RFM. In terms of the Bayesian RFMs, we see that the model with spatially varying bias terms has the best performance. Thus, results from Model 6 will be the focus of the remaining analysis.

Validation of probability of exceedance of ozone standard

The cross validation analysis showed the Bayesian RFM used in this case study is well calibrated across the entire range of observed ozone levels (see supplemental Figure S2 for additional evaluation). However the final application of the model in this study is based on only the fourth highest MD8 value at a given location rather than the entire summer ozone distribution across all sites. Consequently, there is need for additional evaluation of the model results based on the performance for this specific metric of interest. Reliability diagrams (*22*) are used to validate the posterior probability predictions of the daily MD8 and the fourth highest concentration exceeding the current ozone standard of 75ppb (Figure 1). Reliability, or conditional bias, summarizes the conditional distribution of the observations for specific predicted values. Model predicted probabilities of the daily MD8 concentration exceeding a threshold are paired with the observed MD8 value for each site/day. Similarly, the predicted probabilities of the fourth highest ozone value exceeding a threshold are paired with the fourth highest observed value at each site. For both metrics, the predicted probabilities are sorted into bins (e.g. [0, 0.2],(0.2, 0.4], etc.) and the mean for each bin is compared to the observed relative frequency of the event for that bin in the reliability diagram.

Perfectly calibrated model predictions should fall on the 1 to 1 line. For both the daily dataset and the fourth highest values, the base model simulation line is attenuated to the no resolution reference line (i.e. the observed relative frequency of an exceedance across all sites), meaning that the base model does a poor job of separating exceedance events from non-exceedance events. Looking across all summer days, Model 3 tends to be biased high for predicted probabilities >0.2. However when the data are subset to only the fourth highest MD8 at every site, Model 3 is consistently biased low. While Model 3 tends to overpredict the mean ozone across all days, as we saw in Table 1, this model underpredicts at higher observed concentrations levels, a deficiency also seen in the base model simulation.

For the daily MD8 values, Model 6 is much better calibrated, although it also exhibits some high bias for predicted probabilities >0.5. For the fourth highest MD8 data, Model 6 performs very well when it predicts an ozone exceedance with probability falling within [0.0, 0.2] or (0.8, 1.0], meaning these very confident predictions are also very accurate. The barplot inserts show that the Model 6 predictions are also sharper than Model 3, with many more predicted probabilities falling within these two extreme bins. Model 6 does overpredict the probability of exceedance for the intermediate probability bins, suggesting some high bias for the fourth highest summer ozone value (however note the sample sizes for these bins are very small, on the order of 10-25 pairs). A companion study addresses this performance issue for high concentrations by proposing a Bayesian RFM based on a combination of flexible semiparametric quantile regression for the center of the ozone distribution and a parametric extreme value distribution for the tail (*23*). Note that for other model applications, such as evaluating the impact of emissions reductions on the secondary ozone standard, accurate predictions of lower and mid-range ozone values are emphasized more than extreme values. Thus it is important to have a modeling approach that is well calibrated across the entire ozone distribution.

Results

Estimated scaling factors

Posterior summaries for the scaling factors in the reduced form CMAQ model as well as for the bias and error variance terms in Model 6 are given in Table 2. The posterior distributions for the six scaling factors are consistently negative. This agrees with the cross validation results described earlier which showed the base model predictions tended to be too high for this study period. The estimates for the NO_x factors are consistent with the inverse modeling analysis described in Napelenok et al. which also found that urban NO_x emissions are overestimated in this domain. The largest overestimation is in anthropogenic VOCs which have the largest ozone sensitivities in Atlanta and other urban areas (see supplemental Figure S3). Note that the NO_x and VOC scaling factors represent time-invariant domain-wide bias in the emission inventory for this simulation period. The true range of uncertainty in emissions inputs likely changes over time and space. The posterior distributions of the scaling factors presented here are estimates of the aggregate uncertainty levels in the emissions from this specific emissions inventory, for this simulation domain and time period. In addition, while these empirical estimates fit the data well, it is possible that they are affected by deficiencies in other aspects of the CMAQ modeling system not included in the reduced form model.

Even after adjusting the CMAQ model based on the six parameters in the RFM, spatially varying bias remains in the $C(\mathbf{s},t|\alpha)$ predictions. The RFM model with adjusted emissions and boundary conditions tends to overpredict the lower observed ozone concentrations indicated by the estimates for $\beta_0 < 0$ at most grid cells. More importantly, the model tends to underpredict the higher ozone concentrations, indicated by $\beta_1 > 1$ at most grid locations. These statistical bias factors are used to account for uncertainty in the emissions inputs not captured by the DDM sensitivity fields, as well as uncertainty from other inputs such as the MM5 meteorological model and issues of incommensurability that arise when comparing a point measurement to a grid cell average model prediction. The error variance, σ_e^2, indicates how much variability in the observations is not explained by the bias-adjusted RFM. Without this term, the range of predicted values would be too narrow compared to observed values, as we saw with Model 3.

Estimation of the RFM parameters provides insight into the model inputs with the greatest bias and uncertainty. Such results can help prioritize what additional data need to be collected to produce more informed decisions. Of main interest in this application is how to apply these parameter estimates in order to predict ozone concentrations under multiple emission reduction scenarios. The following section compares the probability of exceeding the ozone standard under four different types of hypothetical NO_x reductions.

Evaluating emission reduction strategies

Traditionally, evaluation of emissions reductions is based on two modeling simulations: one representing current conditions and one representing conditions under a specific set of reductions. The impact of the emissions reductions is then estimated based on the difference (or ratio) in predicted concentrations from the two simulations. Due to the computational cost of running regional-scale air quality models, RFMs have been proposed as an efficient method for characterizing the pollutant response to many different types of emissions reductions (*24*, *25*). For example, for the current case study, the CMAQ RFM in (1) is used to estimate the ozone response to a 45% reduction in mobile NO_x emissions by setting the scaling factor for mobile NO_x sensitivities to -0.45 with the remaining scaling factors set to 0.0 (see the top row of Figure 2). However, such an approach does not account for bias and uncertainty in the deterministic model used for simulating the pollutant response.

Digar et al. (*26*) demonstrate that the DDM-based RFM can be used to simultaneously characterize the impact on model concentrations from uncertainty in model inputs and from a decrease in the true emission rates due to a specific reduction. To model ozone simulations that capture both of these changes we apply $\alpha_j^\star = (1+\alpha_j) * (1+\phi_j) - 1$ as the scaling factor for parameter j where α_j represents the correction to the baseline emissions described in the previous section and ϕ_j is the additional change from the proposed reduction. For example, for a 10% reduction in mobile NO_x emissions, where the modeled emissions inputs are estimated to be 20% too high under current conditions, α_{eMNO_X} for current conditions is -0.20 and under the reduction scenario $\alpha^*_{eMNO_X} = -0.28$. In this case, the underlying bias in the emissions input dampens the impact of the emissions reduction, i.e. the change in the ozone concentration (neglecting second order effects for simplicity) would be -0.08$S^{(1)}_{eMNO_X}$ rather than -0.1$S^{(1)}_{eMNO_X}$.

To develop a set of emissions reductions to compare for this case study, the CMAQ RFM is used to calculate the ozone response to changes in sector-specific NO_x emissions without accounting for input uncertainty (hereafter referred to as non-Bayesian RFM). Using different sets of α values, we simulate ozone data under four different emission reduction scenarios: 45% cut in mobile NO_x, 50% cut in point NO_x, 50% cut in other NO_x, and 15% cut in all NO_x sources. All four of these scenarios lower the base model predicted fourth highest MD8 by an average of approximately 3ppb across the entire domain (excluding areas with zero population), although the spatial locations with the largest decreases in ozone changes depending on the emissions sources being reduced (see supplemental Figure S4, top row). For example, the reduction in mobile NO_x results in decreases in ozone in areas downwind of high traffic, with some small increases in very urban grid cells containing sources high of NO_x emissions. In these locations, high levels of NO_x can inhibit ozone formation through titration and removal of radicals. In contrast, the largest decreases under the 50% reduction in other sources of NO_x occur along the coast due to a reduction in emissions in shipping lanes.

Using the MCMC samples from Model 6 for (α, β_0, β_1, σ_e) and different values of $\phi = (\phi_1, \ldots, \phi_p)$, we compute the CMAQ RFM corresponding to input $\alpha^\star = (\alpha^\star_1, \ldots, \alpha^\star_p)$ for the four different emission reduction scenarios. The estimated decreases in the fourth highest MD8 under the reduced emissions conditions are smaller compared to the non-Bayesian RFM results as shown in the second row of supplemental Figure S4. In addition, areas with the largest posterior mean decrease in ozone also have the largest variability in the posterior predicted value, shown in the bottom row of the same figure. These results reflect that the Bayesian predictions for the fourth highest values account for sources of uncertainty and persistent biases that are not accounted for in the base and non-Bayesian models.

Another way to compare the emissions reductions is to use the posterior distributions of the fourth highest MD8 to calculate the probability of a predicted exceedance of the NAAQS at any grid location. The probability of exceedance can be used to communicate the relative success of different emissions reductions in meeting the target air quality level. For example, the posterior probability of exceeding the current ozone standard of 75ppb based on the Bayesian RFM after a 45% reduction in mobile NO_x emissions is less than 0.50 throughout much of the domain, however the probability of exceedance in the largest urban areas is still very high (Figure 2). Table 3 shows the aggregate impact of each emission reduction across all monitoring locations and across the entire domain. Results are also weighted by population density to better quantify the societal benefits of the different emissions reductions in terms of human exposure. The ranking for each emission reduction is shown for the non-Bayesian and the Bayesian RFM output. Based on the non-Bayesian regional mean values (row 2), mobile NO_x reductions are favored over the other types of reductions. Population weighting of the non-Bayesian RFM results leads to point NO_x reductions being preferred over mobile NO_x reductions. This is due to the high population areas in Atlanta, DC, and Baltimore where NO_x acts as a sink of ozone, reducing the benefits from the mobile source reductions (see middle left-hand plot in supplemental Figure S4).

The Bayesian-RFM provides different rankings. For both the regional and population-weighted mean, reducing other NO_x sources produces the smallest probability of exceedance, whereas these sources were ranked last in the non-Bayesian comparison. The impact of reduction in NO_x from other sources, seen in supplemental Figure S4 are more wide-spread compared to the mobile and point-source reductions, although the magnitude of the decrease in urban areas is much less. More importantly, the uncertainty in the model estimates for the ozone response to other NO_x sources, characterized by the spread of the posterior distribution at each grid cell, is much smaller across most of the domain compared to the mobile and point reductions (see bottom row of S4). This is consistent with the results in Table 2 which shows the posterior distribution for α_{eONO_X} has much less bias compared to the other variables, which translates to less dampening of the impact of the emissions reduction on the ozone response. Note that these rankings are specific to the simplified emission reduction scenarios used in this case study. In a regulatory application, the control options would likely be more targeted to specific regions/sectors, thus leading to different conclusions. The important result here is that through this approach, the observation-based estimates for the bias and uncertainty in the input parameters have been propagated to the final model output used for comparing different scenarios.

Discussion

The statistically calibrated RFM presented here is shown to provide predicted ozone concentrations and predicted probabilities of exceedances that are both accurate and reliable. While the results of the emission reduction scenario comparisons are specific to the emissions inventory and AQM simulation used in this case study, these examples demonstrate how this approach can be used to refine and supplement the evaluation of traditional emission reduction pathways. Propagating the uncertainty in the modeling inputs to the model output metric of interest, in this fashion can help to improve confidence in air quality management decisions.

Several extensions to this methodology would be of interest. For example, a more comprehensive RFM could include explicit representation of uncertainties in the meteorological inputs. Such a model could be developed with existing technology by utilizing an ensemble of meteorological model inputs similar to the approach in (*27*). An additional modification to the RFM would be to introduce spatially varying emission adjustments by developing DDM-3D sensitivities to emissions in specific geographic regions. Furthermore, the sector definitions used in this study could be refined. For example, separating point-sources with continuous emissions monitoring data (e.g. large electrical generating units) from other point sources would allow for improved estimates of the bias and uncertainty in each sector type. In terms of the statistical model, the approach presented in (*23*) utilizing extreme value theory could be used to improve the calibration of the Bayesian RFM for high ozone concentrations.

Acknowledgement

We thank Christian Hogrefe, Robert Pinder and Norm Possiel from the U.S. Environmental Protection Agency (EPA) for their comments and suggestions throughout the development of this work. 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 Agency's administrative review and approved for publication.

Supporting Information Available

Time series and spatial plots of O3 sensitivities to the six parameters of the reduced form CMAQ model, PIT histograms of the predicted MD8 ozone concentrations, and spatial plots of the impact of four emission reduction scenarios on the fourth highest maximum eight-hour average ozone (MD8). This material is available free of charge via the Internet at `http://pubs.acs.org`.

References

(1) US EPA, Air Quality Criteria for Ozone and Related Photochemical Oxidants (Final). Washington, DC: U.S. EPA, available online at: `http://cfpub.epa.gov/ncea/CFM/recordisplay.cfm?deid=149923`, 2006.

(2) Fishman, J.; Creilson, J. K.; Parker, P. A.; Ainsworth, E. A.; Vining, G. G.; Szarka, J.; Booker, F. L.; Xu, X. An investigation of widespread ozone damage to the soybean crop in the upper Midwest determined from ground-based and satellite measurements. *Atmos. Environ.* **2010**, *44*, 2248–2256.

(3) Avnery, S.; Mauzerall, D. L.; Liu, J.; Horowitz, L. W. Global crop yield reductions due to surface ozone exposure: 1. Year 2000 crop production losses and economic damage. *Atmos. Envir.* **2011**, *45*, 2284–2296.

(4) Digar, A.; Cohan, D. S.; Cox, D. D.; Byeong-Uk, K.; Boylan, J. W. Likelihood of Achieving Air Quality Targets under Model Uncertainties. *Environ. Sci. Technol.* **2011**, *45*, 189–196.

(5) Hogrefe, C.; Rao, S. T. Deomonstrating attainment of the air quality standards: integration of observations and model predicitons inot the probabilistic framework. *J. Air & Waste Manage. Assoc.* **2001**, *51*, 1060–1072.

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

(7) 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. *Environ. Sci. Technol.* **2006**, *40*, 3855–3864.

(8) Martien, P. T.; Harley, R. A. Adjoint sensitivity analysis for a three-dimensional photochemical model: Application to Southern California. *Environ. Sci. Technol.* **2006**, *40*, 4200–4210.

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

(10) Napelenok, S. L.; Foley, K. M.; Kang, D.; Mathur, R.; Pierce, T.; Rao, S. T. Dynamic evaluation of regional air quality model's response to emission reductions in the presence of uncertain emission inventories. *Atmos. Environ.* **2011**, *45*, 4091–4098.

(11) Byun, D.; 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.

(12) Foley, K. M.; Roselle, S. J.; Appel, K. W.; Bhave, P. V.; Pleim, J. E.; Otte, T. L.; Mathur, R.; Sarwar, G.; Young, J. O.; Gilliam, R. C.; Nolte, C. G.; Kelly, J. T.; Gilliland, A. B.; Bash, J. O. Incremental testing of the Community Multiscale Air Quality (CMAQ) modeling system version 4.7. *Geoscientific Model Development* **2010**, *3*, 204–226.

(13) Grell, G. A.; Dudhia, A. J.; Stauffer, D. R. A description of the Fifth-Generation PennState/NCAR Mesoscale Model (MM5). NCAR Technical Note NCAR/TN-398+STR, available online at: `http://www.mmm.ucar.edu/mm5/doc1.html`, 1994.

(14) Schwede, D.; Pouliot, G. A.; Pierce, T. Changes to the Biogenic Emissions Inventory System Version 3 (BEIS3). In Proceedings of the 4th CMAS Models-3 Users Conference, Chapel Hill, NC, September, 2005.

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

(16) Carter, W. P. L. Implementation of the SAPRC-99 Chemical Mechanism into the Models-3 Framework. Report to the United States Environmental Protection Agency, available online at: `http://www.cert.ucr.edu/\~carter/absts.htm\#s99mod3`, 2000.

(17) Napelenok, S. L.; 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. *Atmos. Chem. Phys.* **2008**, *8*, 5603–5614.

(18) Gelman, A.; Carlin, J. B.; Stern, H. S.; Rubin, D. B. *Bayesian data anlaysis*; Chapman & Hall: London, 1995.

(19) Casella, G.; George, E. I. Explaining the Gibbs Sampler. *Am. Stat.* **1992**, *46*, 167–174.

(20) Chib, S.; Greenberg, E. Understanding the Metropolis-Hastings algorithm. *The American Statistician* **1995**, *49*, 327–335.

(21) Dennis, R. L.; Bhave, P. V.; Pinder, R. W. Observable indicators of the sensitivity of PM2.5 nitrate to emission reductions – Part II: Sensitivity to errors in total ammonia and total nitrate of the CMAQ-predicted non-linear effect of SO2 emission reductions. *Atmos. Environ.* **2008**, *42*, 1287–1300.

(22) Wilks, D. S. *Atmospheric Modeling, Data Assimilation and Predictability*; Academic Press: New York, 2006.

(23) Reich, B. J.; Cooley, D.; Foley, K. M.; Napelenok, S. L.; Shaby, B. A. Extreme value analysis for evaluating ozone control strategies. **in review**,

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

(25) Xing, J.; Wang, S. X.; Jang, C.; Zhu, Y.; Hao, J. M. Nonlinear response of ozone to precursor emission changes in China: a modeling study using response surface methodology. *Atmos. Chem. Phys.* **2011**, *11*, 5027–5044.

(26) Digar, A.; Cohan, D. S.; Cox, D. D.; Byeong-Uk, K.; Boylan, J. W. Efficient Characterization

of Pollutant-Emission Response under Parametric Uncertainty. *Environ. Sci. Technol.* **2010**, *44*, 6724–6730.

(27) Du, J. Hybrid Ensemble Prediction System: a New Ensembling Approach. Symposium on the 50th Anniversary of Operational Numerical Weather Prediction, College Park, Maryland, available online at: `http://www.emc.ncep.noaa.gov/mmb/SREF/hybrid\_50thNWP.pdf`, 2004.

Table 1: Model description for various statistical models and cross-validation results on the entire testing dataset. α Prior equal to "None" refers to α_j=0, $j = 1, \ldots, 6$. Similarly, Bias equal to "None" refers to $\beta_0 = 0; \beta_1 = 1$. The summaries are root mean squared error (RMSE), mean bias (Bias), mean predictive standard deviation (SD), coverage probability of 90% intervals (COV90), and the Brier score for exceedances of 75ppb (Brier75).

	Model				Results				
	CMAQ	α Prior	Bias	Residuals	RMSE	Bias	SD	COV90	Brier75
1	Base	None	None	None	10.14	2.92	0.00	0.00	0.055
2	Base	None	None	Indep	10.14	2.92	10.16	**0.90**	0.047
3	RFM	Informative	None	None	10.10	3.05	4.80	0.56	**0.043**
4	RFM	Uninform	None	Indep	9.66	0.08	9.63	**0.90**	0.051
5	RFM	Uninform	Constant	Indep	9.51	0.04	9.47	**0.90**	0.049
6	**RFM**	**Uninform**	**Spatial**	**Indep**	**8.70**	**0.03**	8.62	**0.90**	0.048

Table 2: Posterior summaries of the scaling factors for the six parameters of the reduced form CMAQ model as well as for $\bar{\beta}_0$, $\bar{\beta}_1$ (average bias across space) and σ_e. Results are based on the parametrization in Model 6.

	Posterior	
Variable	Mean	95% Interval
$eMNO_x$	-0.242	(-0.303, -0.186)
$ePNO_x$	-0.203	(-0.260, -0.143)
$eONO_x$	-0.087	(-0.180, 0.005)
eAVOC	-0.736	(-0.819, -0.648)
eBVOC	-0.265	(-0.302, -0.229)
bcO3	-0.113	(-0.128, -0.100)
$\bar{\beta}_0$	-7.927	(-8.765, -7.051)
$\bar{\beta}_1$	1.204	(1.187, 1.222)
σ_e	8.547	(8.473, 8.621)

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

Figure 1: Reliability diagrams for the predicted probability of (a) the daily maximum eight-hour (MD8) ozone and (b) the fourth highest MD8 ozone value exceeding a threshold of 75ppb. Results are shown for the base simulation (Model 1), the RFM based on informative priors with no Bayesian updating (Model 3), and the Bayesian RFM (Model 6). The observed relative frequency of exceedance across all monitoring sites is shown by the dotted grey line (no resolution). The barplot inserts indicate how many model/observation pairs fall within each of the 5 predicted probability bins indicated on the x-axis (e.g. [0.0, 0.2], (0.2, 0.4], etc.).

(a) Daily MD8 Values, TH=75

(b) 4th Highest MD8 Values, TH=75

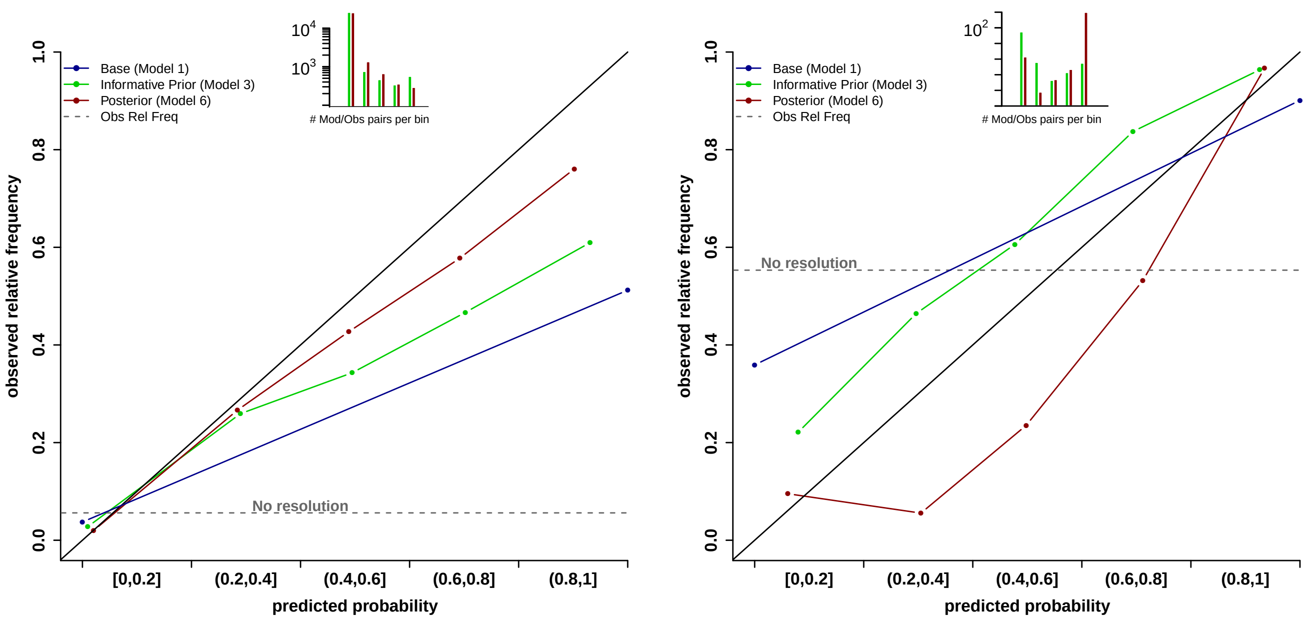

Table 3: Average probability of exceeding a standard of 75ppb for each of the four emission reduction scenarios based on the fourth highest MD8 ozone value at every grid cell in the domain (excluding areas with zero population). Population weighting is based on 2000 Census Tract Population Totals. Emission reduction scenario ranks for each row are shown in parentheses.

		No Reduction	Scenario 1 -45% MNO$_x$	Scenario 2 -50% PNO$_x$	Scenario 3 -50% ONO$_x$	Scenario 4 -15% NO$_x$
Non-Bayesian RFM	Mean at AQS Sites	.358	.230 (3)	.209 (1)	.238 (4)	.227 (2)
	Regional Mean	.122	.067 (1)	.068 (2)	.073 (4)	.070 (3)
	Pop-weighted Mean	.306	.215 (2)	.212 (1)	.229 (4)	.219 (3)
Bayesian RFM	Mean at AQS Sites	.610	.486 (3)	.483 (2)	.479 (1)	.492 (4)
	Regional Mean	.368	.271 (3)	.269 (2)	.261 (1)	.273 (4)
	Pop-weighted Mean	.529	.426 (2)	.430 (3)	.422 (1)	.434 (4)

Figure 2: Spatial plots of fourth highest summer MD8 ozone concentration (ppb) under current conditions (top left) and after a 45% reduction in mobile NO_x emissions (top right) based on the base model simulation and the non-Bayesian RFM, respectively. Bottom row shows the posterior probability of the fourth highest MD8 ozone value exceeding a standard of 75ppb before (bottom left) and after (bottom right) the reduction in mobile NO_x emissions based on the Bayesian RFM output (Model 6). Results for the other emission reduction scenarios showed only minor spatial differences and are not plotted. Ozone monitoring locations are shown in black.

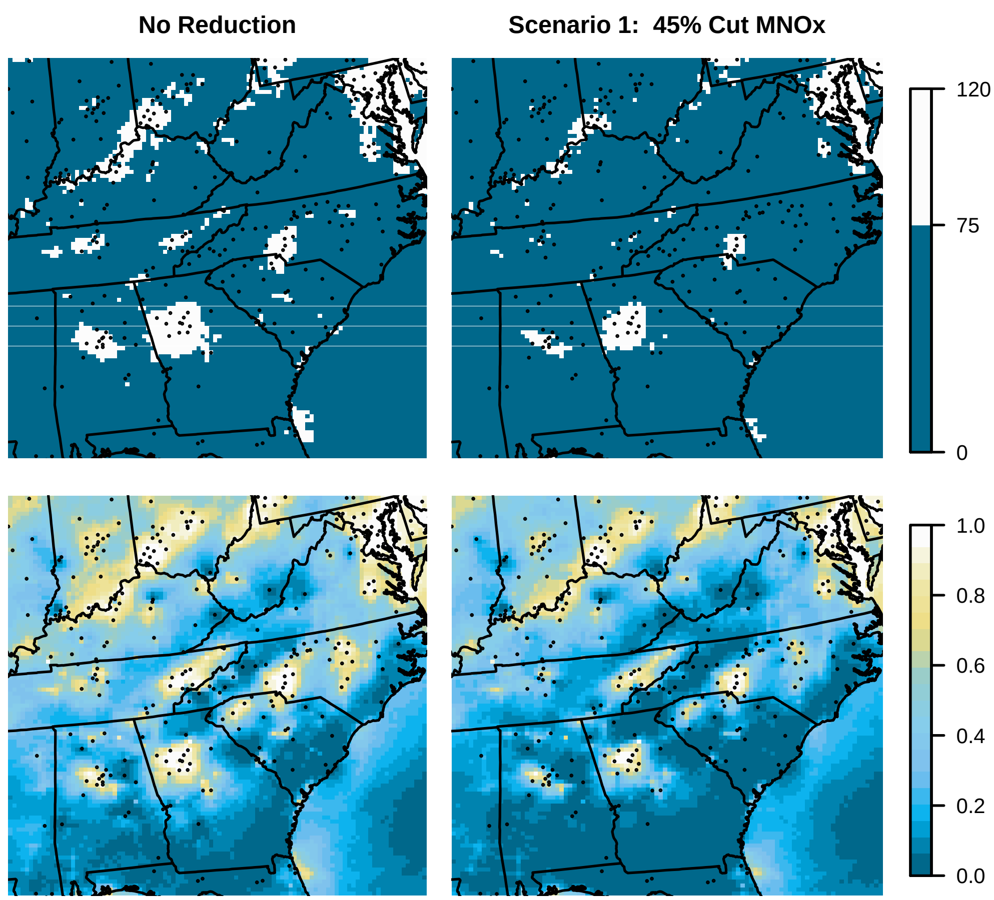

Figure 3: TOC ART