A Bayesian model for quantifying the change in mortality associated with future ozone exposures under climate change

Stacey E. Alexeeff, Gabriele G. Pfister, Doug Nychka

August 13, 2014

National Center for Atmospheric Research, Boulder, CO, U.S.A.

Abstract

Climate change is expected to have many impacts on public health and the environment, including changes in surface ozone concentrations. A key public health concern is the potential increase in ozone-related summertime mortality if surface ozone concentrations rise in response to climate change. A major source of uncertainty in the health impacts is the variability of the modeled ozone concentrations. Although ozone formation depends partly on summertime weather which exhibits considerable year-to-year variability, previous health impact studies have not incorporated the variability of ozone into their prediction models. A common data reduction step to handle the Big Data is to use double averaging of the ozone, which reduces dimensionality but also results in a loss of information about variability. We propose a Bayesian model and Monte Carlo estimation method for quantifying health effects of future ozone. An advantage of this approach is that we estimate the uncertainty in both the health effect association and the modeled ozone concentrations. Using this modeling approach, we quantify the expected change in ozone-related summertime mortality in the contiguous United States between 2000 and 2050 under a changing climate. This analysis yields more realistic estimates and inferences, providing clearer interpretation for decision making regarding the impacts of climate change.

1 Introduction

Climate change due to increased greenhouse gases may have an impact on risks to human health. One concern is that ozone concentrations at the surface level could increase as a consequence of climate change, in particular because of the photochemical formation of ozone [IPCC, 2007]. Because of this concern, interventions to reduce ground level air pollution emissions have been proposed to mitigate the increase in air pollution. Climate model simulations with atmospheric chemistry can help investigate these complex interactions and estimate future surface ozone concentrations.

Numerous epidemiological studies and meta-analyses have reported positive associations between short-term changes in ozone exposure and daily mortality rates [Stieb et al., 2002, Bell et al., 2004, 2005, Ito et al., 2005, Levy et al., 2005] Based on an overall synthesis of evidence from several scientific disciplines, the most recent U.S. EPA integrated science assessment has concluded that "there is likely to be a causal relationship between short-term exposures to O_3 and total mortality" [EPA, 2013]. This scientific evidence raises concern over the health impacts on daily mortality if ozone levels increase in the future.

Future expected mortality attributable to changes in ozone levels has been estimated in several recent studies under different climate change scenarios [Post et al., 2012, Bell et al., 2007, Tagaris et al., 2009]. However, there is a lack of consistent statistical methodology to estimate the uncertainty associated with these estimates. For example, Post et al. [2012] reports only point estimates with no measure of uncertainty, and Bell et al. [2007] only accounts for the uncertainty in the health effect association. Future ozone projections are inherently variable because they are influenced by the year-to-year variability of weather. Yet no statistical methodology has been proposed to account for the uncertainty in the climate model simulations.

We present a Bayesian modeling approach and a Monte Carlo simulation method for quantifying health effects of future ozone concentrations under climate change.

An innovation of this approach is that the uncertainty estimate accounts for both the uncertainty in the health effect association and the variability of the estimated ozone changes from the climate model projections. We demonstrate this approach by quantifying the expected change in ozone-related summer mortality in the contiguous United States between 2000 and 2050 under a changing climate with two different emission control scenarios.

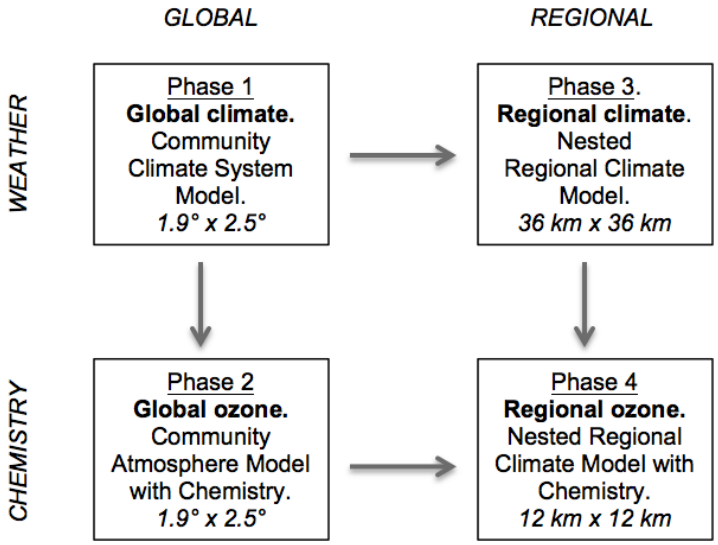
2 Data Structure of Climate Change and Health Studies

Here we describe the setup of a climate change and health study, focusing on the particular aspects that motivate the methodology of our paper. In Section 5, we apply our proposed modeling approach to this climate change data.

2.1 Ozone projections for present and future

To study changes in surface ozone over the US in the next 50 years, projections of present and future ozone concentrations were generated using global climate, regional climate, and regional air quality models. The overall project to generate ozone concentrations over the contiguous U.S. proceeded in four phases of modeling, as summarized in Figure 1. Global climate projections were generated in phase 1 by the Community Climate System Model run under the IPCC SRES A2 scenario [Meehl et al., 2007]. This is a full Earth system model that includes atmosphere, ocean, cryosphere, biosphere, and land surface. In phase 2, the global Community Atmosphere Model with Chemistry generated global ozone projections at 1.9 degrees Latitude by 2.5 degrees Longitude using the RCP 8.5 emissions inventory [Lamarque et al., 2011]. In phase 3, the global climate projections of phase 1 were dynamically downscaled using the Nested Regional Climate Model to a 36 km $\times$ 36 km domain over the larger North America region [Done et al., 2013]. The results of phases 2 and 3 provided initial conditions and boundary conditions for the Nested Regional Climate Model with Chemistry, a fully coupled chemical transport model. In Phase

Figure 1: *Summary of modeling phases needed to generate 12km × 12km regional ozone concentrations over the contiguous U.S. states.*



4, this model generated hourly ozone concentrations during summertime over the contiguous U.S. for the present (ca. 2000) and future (ca. 2050) at a 12 km × 12 km resolution. The Nested Regional Climate Model with Chemistry has been extensively evaluated and compared with monitoring data, with details found in Kim et al. [2009] and Pfister et al. [2014].

Ozone concentrations were simulated for 13 years during summertime months of June, July, and August for the present and future time periods. Present time anthropogenic emissions were computed from the year 2000 emissions inventory. For the future, two anthropogenic emissions scenarios for the U.S. were considered: *Future A* where U.S. emissions continue at the year 2000 levels into the future and *Future B* where U.S. emissions are reduced in the future based on the RCP 8.5 emissions inventory. All RCPs assume a more stringent air pollution control policy in the future, and emissions of many ozone precursors such as NOx are assumed to decrease [van Vuuren et al., 2011]. Further details of the aerosol emissions inventory can be found in Lamarque et al. [2011] and details of the ozone projections can be

found in Pfister et al. [2014].

2.2 Regridding of ozone model output

One problem common to many spatial data analyses is combining data from different spatial grids. For studies of climate and health specifically, the problem is that climate-chemistry model simulations produce output on a regularly spaced grid over latitude and longitude, while population information from Census data is aggregated into irregular shaped polygons. In order to perform any analyses on these data, a method is needed to transform the ozone concentrations from the original grid to the target polygons; this process is called "regridding" in the climate change literature.

From a Big Data perspective, this regridding step is a nontrivial computational step because the transformation procedure must be applied to each daily ozone surface. For example, our study includes 92 days in summer for 13 years over 3 time periods, so the regridding step must be applied 3,588 times. In addition, the ozone data may be contained in many separate files totaling several Gbs, even after extracting only the surface-level ozone from the climate-chemistry model simulation output and averaging the hourly ozone to daily ozone.

We conjecture that one reason previous studies have averaged the ozone concentrations into a single summertime average is to avoid the computational demands of this step. By first computing a single summertime average ozone concentration for each time period and next applying the regridding procedure, the regridding procedure would only be needed 2 or 3 times instead of several thousand. Unfortunately, this shortcut approach results in a drastic loss of information about the ozone concentrations and their variability.

We propose applying the regridding step to each daily ozone surface in order to preserve the daily ozone information. To compute county-level daily ozone concentrations, we propose a two step procedure. First, fit a thin plate spline to the daily

ozone concentrations to interpolate from a 12km grid to a 2km grid, then average the ozone concentrations over the centroids of the 2km $\times$ 2km gridboxes falling within each county boundary for each of the 3,219 counties in the contiguous U.S. Supplementary Figure 1 illustrates the thin plate spline interpolation for one day of ozone. Supplementary Appendix D includes the R code for the regridding step, implemented in parallel over a computer cluster.

Previous climate and health studies have not explicitly defined their regridding procedure or its implementation. The simplest regridding approach would be to average the ozone concentrations over the centroids of the 12km $\times$ 12km gridboxes falling within each county boundary. However, this approach has the disadvantage that the irregular boundary will contain different proportions of each gridbox, which may lead to a crude approximation of the county average. To improve the accuracy of the county-level averages used for the health effect analysis, we propose the added interpolation step to a finer grid.

3 Modeling Approach

This section provides details of our proposed statistical approach for quantifying climate change-induced health effects. The first two subsections define our Bayesian framework for the distributions of ozone in the present and future and the distribution of the health effect parameter. In the following two subsections we outline how to make inferences for two quantities: the change in the total summertime mortality and the percent change in the total summertime mortality.

3.1 Prior and Posterior distributions for β

The goal of this paper is to study how best to quantify *predictions* of ozone-related mortality. For that reason, we focus on finding the distributions of posterior predictive summaries and we assume that the first step of the Bayesian analysis is already in hand. In our study, we use the previously published Bayesian meta-analysis

of Bell et al. [2005] to provide the posterior distribution of the health effect parameter, denoted by β. Specifically, we assume the posterior distribution of the health effect parameter is given by $\beta \sim N(\mu_\beta, \sigma_\beta^2)$, where μ_β and σ_β^2 are taken from the U.S.-specific meta-analysis results in Bell et al. [2005]. The authors reported that a 10-ppb increase in 24-hr ozone was associated with a 0.84% increase in total mortality (95% posterior interval 0.48% to 1.20%), based on studies of the U.S. population.

3.2 Health Effect Model for Ozone and Mortality

Suppose we have a region consisting of geographic locations $i = (1, \ldots, N)$ and consider the days in summer $t = (1, \ldots, T)$. Let Y_{it} denote the number of deaths for location i on day t. Let X_{it} denote the surface ozone concentration for location i on day t, and let $\boldsymbol{X}$ be a $N \times T$ matrix of these surface ozone concentrations for all locations in the region and all days of summer. Assume that $\boldsymbol{X}$ is a generated from a multivariate spatio-temporal distribution, $\boldsymbol{X} \sim F_{\boldsymbol{X}}$. For each geographic location, i, let the population be denoted by P_i and the mortality rate be denoted by R_i, and assume both are known from observed data. Assume $\beta \sim N(\mu_\beta, \sigma_\beta^2)$ is the posterior distribution of the health effect parameter as defined above; implicitly, β is dependent on $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$, the observed ozone and mortality data evaluated in the health studies. Let the expected mortality follow

$$E\left(Y_{it}|\beta, X_{it}; P_i, R_i\right) = P_i R_i \exp\left(\beta X_{it}\right), \tag{1}$$

which is consistent with a Poisson regression model or a generalized additive model with a log link function. Equation (1) gives an expression for the expected daily mortality for one location given β and X_{it}, and also defines a new random variable whose distribution depends on the joint distribution of β and X_{it}. Let the scalar random variable η represent the expected total summertime mortality in our region,

where

$$\eta(\beta, \boldsymbol{X}) = \sum_i \sum_t P_i R_i \exp\{\beta X_{it}\}. \quad (2)$$

We assume that β and $\boldsymbol{X}$ are independent random variables, noting the subtlety that the posterior distribution of β depends directly on the true distribution of ozone in the past, while $\boldsymbol{X}$ represents simulated ozone concentrations generated from climate-chemistry models. Conceptually, the distribution of η has similarities to computing a posterior predictive distribution, except that instead of a prediction interval given one new observation $\tilde{\boldsymbol{X}}$, we are interested in the distribution of this mean function given all possible values of $\boldsymbol{X}$ from its distribution $F_{\boldsymbol{X}}$. In the next two subsections, we define the random variables of interest using η and explain how to make valid inferences.

3.3 Inference for change in total summertime mortality.

We first consider how to make estimates and inferences for the change in expected total summertime mortality attributable to ozone in a region. Let η^F and η^P denote the expected total summertime mortality in the future and in the present respectively, as defined in equation (2). The statistical summaries used for inference are the posterior mean and the 95% credible interval. The posterior mean of $\eta^F - \eta^P$ is

$$E\left[\eta^F - \eta^P\right] = \int \eta(\beta, \boldsymbol{x}) f_\beta(\beta) f_{\boldsymbol{X}^F}(\boldsymbol{x}) d\boldsymbol{x} d\beta - \int \eta(\beta, \boldsymbol{x}) f_\beta(\beta) f_{\boldsymbol{X}^P}(\boldsymbol{x}) d\boldsymbol{x} d\beta \quad (3)$$

Consider the 95% credible interval in the central 95% posterior probability interval, such that 2.5% of the posterior probability lies above and below the interval. Denote the c.d.f. of $\eta^F - \eta^P$ by the function $\Phi(\cdot)$. Written as an integral,

$$\Phi(a) = \int_{[A:\eta(\beta,\boldsymbol{x})-\eta(\beta,\boldsymbol{u})\leq a]} f_\beta(\beta) f_{\boldsymbol{X}^F}(\boldsymbol{x}) f_{\boldsymbol{X}^P}(\boldsymbol{u}) d\boldsymbol{x} d\boldsymbol{u} d\beta \quad (4)$$

To obtain the bounds for the central 95% credible interval, we find the inverse of the c.d.f $\Phi^{-1}(p)$ for percentiles $p = 0.025, 0.975$. Note that we assume that ozone

has some dependence structure across counties and across days within summer, for distributions $f_{\boldsymbol{X}^F}$ and $f_{\boldsymbol{X}^P}$.

3.4 Inference for percent change in total summertime mortality.

Now consider the percent change in expected total summertime mortality in the future compared to the present. Let ζ denote the percent change, which is defined as $\zeta = \frac{\eta^F - \eta^P}{\eta^P}$. This function is useful for comparing changes in several regions with different population sizes.

The posterior mean of ζ is

$$E(\zeta) = \int \left[\frac{\eta(\beta, \boldsymbol{x})}{\eta(\beta, \boldsymbol{u})} - 1\right] f_\beta(\beta) f_{\boldsymbol{X}^F}(\boldsymbol{x}) f_{\boldsymbol{X}^P}(\boldsymbol{u}) d\boldsymbol{x} d\boldsymbol{u} d\beta \tag{5}$$

Deonte the c.d.f. of ζ by the function $\Pi(\cdot)$. Written as an integral,

$$\Pi(a) = \int_{\left[A: \frac{\eta(\beta,\boldsymbol{x})}{\eta(\beta,\boldsymbol{u})} - 1 \leq a\right]} f_\beta(\beta) f_{\boldsymbol{X}^F}(\boldsymbol{x}) f_{\boldsymbol{X}^P}(\boldsymbol{u}) d\boldsymbol{x} d\boldsymbol{u} d\beta \tag{6}$$

To obtain the central 95% credible interval for ζ, we find the inverse of the c.d.f $\Pi^{-1}(p)$ for percentiles $p = 0.025, 0.975$. Note that this fraction does not simplify further because of the matrix multiplication, but we can easily use Monte Carlo sampling to compute this quantity, which we explain next in Section 5.

3.5 The double averaging method and loss of variability

Several recent studies have estimated the expected future ozone-related mortality [Post et al., 2012, Bell et al., 2007, Tagaris et al., 2009]. Similar to our study, those studies have also used existing meta-analyses, including the Bell et al. [2005] meta-analysis, to quantify the health association and its uncertainty. However, the key difference between this study and previous studies is the treatment of the ozone concentrations.

Researchers have estimated future climate change and ozone-related mortality using an approach which we will call the *double averaging method*. We define the method as follows. Suppose that we have simulated daily ozone concentrations X_{tm} for $t = 1, \ldots, T$ days of summer and $m = 1, \ldots, M$ simulation years. The

double averaging method first computes average ozone over both days of summer and years of simulation, specifically, $\bar{X} = \frac{1}{T}\frac{1}{M}\sum_{t=1}^{T}\sum_{m=1}^{M} X_{tm}$, and then treats $\bar{X}$ as observed ozone data by plugging into health association equations as described in the following examples below. Our two main concerns for use of double averaging are (1) introducing bias via interchanging expectation and averaging in log-linear functions, and (2) loss of variability in the ozone projections by reducing potentially thousands of daily concentrations to a single average.

First, consider using the double averaging method to estimate the expected total summertime mortality. Using the double averaged ozone $\bar{X}$ and the posterior mean of β from a meta-analysis, the point estimate for expected mortality is computed as $P_i R_i \exp\{\beta \bar{X}\}$. Post et al. [2012] argue that average ozone can be used "because the health impact functions are nearly linear," referring to log-linear functions. However, the double averaging method will underestimate the expected total summertime mortality because $\frac{1}{T}\frac{1}{M}\sum_{t=1}^{T}\sum_{m=1}^{M}\exp\{\beta X_{tm}\} > \exp\{\beta \bar{X}\}$ by Jensen's inequality. When estimating the difference in expected total summertime mortality between the present and future, the bias will depend on the relative magnitude of the bias in the present and future estimates. Another problem is that this plug-in method yields no statistically valid method to construct a 95% credible interval around the mortality difference; plugging in the upper and lower CI limits for β would only yield individual confidence intervals around the present and future respectively with no way to estimate confidence around the difference. Indeed, Post et al. [2012] chooses to report no intervals around the point estimates.

Next, consider estimating the expected percent changes in summertime mortality using the double averaging method. Let and $\bar{X}^F$ and $\bar{X}^P$ denote the double averaged ozone over the future simulation years and present simulation years respectively. The double averaging method estimates the percent change in summertime mortality by the ratio $\exp(\beta \bar{X}^F)/\exp(\beta \bar{X}^P) - 1$. However, as outlined above, both the numerator

and denominator are biased estimates of the average over the exponential function, yielding a biased point estimate of the percent change in mortality. To obtain a 95% credible interval, the double averaging method takes the upper and lower bounds of the credible interval for β, and plugs those bounds into the mortality function ratio along with the double averaged ozone. This type of plug-in interval only reflects the uncertainty in the health parameter β, and ignores the variability of ozone. The final problem is that the double averaging method does not allow the estimates and confidence intervals to be aggregated regionally. Instead, for example, [Bell et al., 2007] reports the *average percent change across cities* by averaging the percent change over each city.

The problems with double averaging described here are also illustrated using our particular data example in Section 5 and graphics in the Supplement. Our Bayesian framework addresses the problems of past frameworks by clearly defining the statistical model and the quantities of interest, and accounting for both the variability of ozone as well as the variability of the health parameter.

4 Monte Carlo Procedure and Simplifications

A key innovation of our study is to use the empirical distribution for ozone by treating the different years of projections within each time period as independent realizations of $\boldsymbol{X}$. This utilizes the natural uncertainty inherent in the climate-chemistry simulations to characterize the uncertainty of $\boldsymbol{X}$ without making distributional assumptions about the underlying space-time process.

4.1 Monte Carlo for change in total summertime mortality

The integrals derived in Section 3.2 can be approximated by Monte Carlo as follows. First, simulate M realizations of summertime ozone in the present, $\boldsymbol{X}^{P,1}, \ldots, \boldsymbol{X}^{P,M}$ and M' realizations of summertime ozone in the future, $\boldsymbol{X}^{F,1}, \ldots, \boldsymbol{X}^{F,M'}$. Also simulate K realizations of β from the posterior distribution for the health effect,

$\beta^1, \ldots, \beta^K$. Then, Equation (2) can be approximated as

$$\frac{1}{M'}\frac{1}{K}\sum_{m'=1}^{M'}\sum_{k=1}^{K}\eta(\beta^k \boldsymbol{X}^{F,m'}) - \frac{1}{M}\frac{1}{K}\sum_{m=1}^{M}\sum_{k=1}^{K}\boldsymbol{b}^{\mathsf{T}}\exp(\beta^k \boldsymbol{X}^{P,m})\mathbf{1}. \tag{7}$$

And Equation (3) can be approximated as

$$\frac{1}{M}\frac{1}{M'}\frac{1}{K}\sum_{m=1}^{M}\sum_{m'=1}^{M'}\sum_{k=1}^{K} I\left[\eta(\beta^k \boldsymbol{X}^{F,m'}) - \eta(\beta^k \boldsymbol{X}^{P,m}) \leq a\right]. \tag{8}$$

Then we use order statistics of the differences to select the appropriate a and b to obtain the inverse of this distribution.

4.2 Monte Carlo simplification using exact integral solution

A simplification of the Monte Carlo simulation is the exact solution to the integral with respect to β. Under the assumption of Section 3 that β has a normal distribution, we obtain an exact solution to the integrals for the mean and variance of η over the distribution of β. The integral over $\boldsymbol{X}$ can then be approximated by Monte Carlo.

By Lemma 1 in Supplement Appendix A, the integrals for $E[\eta]$ and $Var[\eta]$ can be simplified to

$$E[\eta] = \int \sum_{i=1}^{N} P_i R_i \left[\sum_{t=1}^{T}\exp\left\{x_{it}\mu_\beta + \sigma_\beta^2 x_{it}^2/2\right\}\right] f_{\boldsymbol{X}}(\mathbf{x})d\mathbf{x}.$$

and

$$\begin{aligned} Var(\eta) = &\int \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} P_i R_i P_j R_j \exp\left[\mu_\beta(x_{it}+x_{js}) + \sigma_\beta^2(x_{it}+x_{js})^2/2\right] f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \\ &- \left[\int \sum_{i=1}^{N}\sum_{t=1}^{T} P_i R_i \exp\left\{x_{it}\mu_\beta + \sigma_\beta^2 x_{it}^2/2\right\} f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}\right]^2. \end{aligned}$$

Then Corollary 1 in Supplement Appendix A extends this Lemma to the difference $\eta^F - \eta^P$ where the two random variables are correlated through their dependence on β. Proofs of Lemma 1 and Corollary 1 are given in Supplement Appendix A.

The exact integral solution eliminates the need for Monte Carlo simulation over realizations of β, giving a more precise estimate and reducing the computational time. However, to use the mean and variance to construct credible intervals requires assumptions about the posterior distribution of η. By Central Limit Theorem arguments, it is reasonable to assume a Normal distribution.

4.3 Monte Carlo for percent change in total summertime mortality

The integrals derived in Section 3.3 can be approximated by Monte Carlo as follows. First, simulate M realizations of summertime ozone in the present, $\boldsymbol{X}^{P,1}, \ldots, \boldsymbol{X}^{P,M}$ and M' realizations of summertime ozone in the future, $\boldsymbol{X}^{F,1}, \ldots, \boldsymbol{X}^{F,M'}$. Also simulate K realizations of β from the posterior distribution for the health effect, $\beta^1, \ldots, \beta^K$. Then, Equation (4) can be approximated as

$$\frac{1}{M'}\frac{1}{M}\frac{1}{K}\sum_{m'=1}^{M'}\sum_{k=1}^{K}\sum_{m=1}^{M}\left\{\frac{\eta(\beta^k \boldsymbol{X}^{F,m'}) - \eta(\beta^k \boldsymbol{X}^{P,m})}{\eta(\beta^k \boldsymbol{X}^{P,m})}\right\}. \tag{9}$$

And Equation (5) can be approximated as

$$\frac{1}{M}\frac{1}{M'}\frac{1}{K}\sum_{m=1}^{M}\sum_{m'=1}^{M'}\sum_{k=1}^{K} I\left[\left\{\frac{\eta(\beta^k \boldsymbol{X}^{F,m'}) - \eta(\beta^k \boldsymbol{X}^{P,m})}{\eta(\beta^k \boldsymbol{X}^{P,m})}\right\} \leq a\right]. \tag{10}$$

The R code for the Monte Carlo sampling is given in Supplementary Appendix C.

5 Application Results: Quantification of future mortality in 2050 under climate change for two emissions scenarios.

We apply our methodology to the climate change study described in Section 2. County-level population data was procured from the U.S. 2000 Census [U.S. Census Bureau]. We computed average daily mortality rates during the summertime months for each state using the population data and mortality data for year 2000 obtained from the Center for Disease Control [CDC].

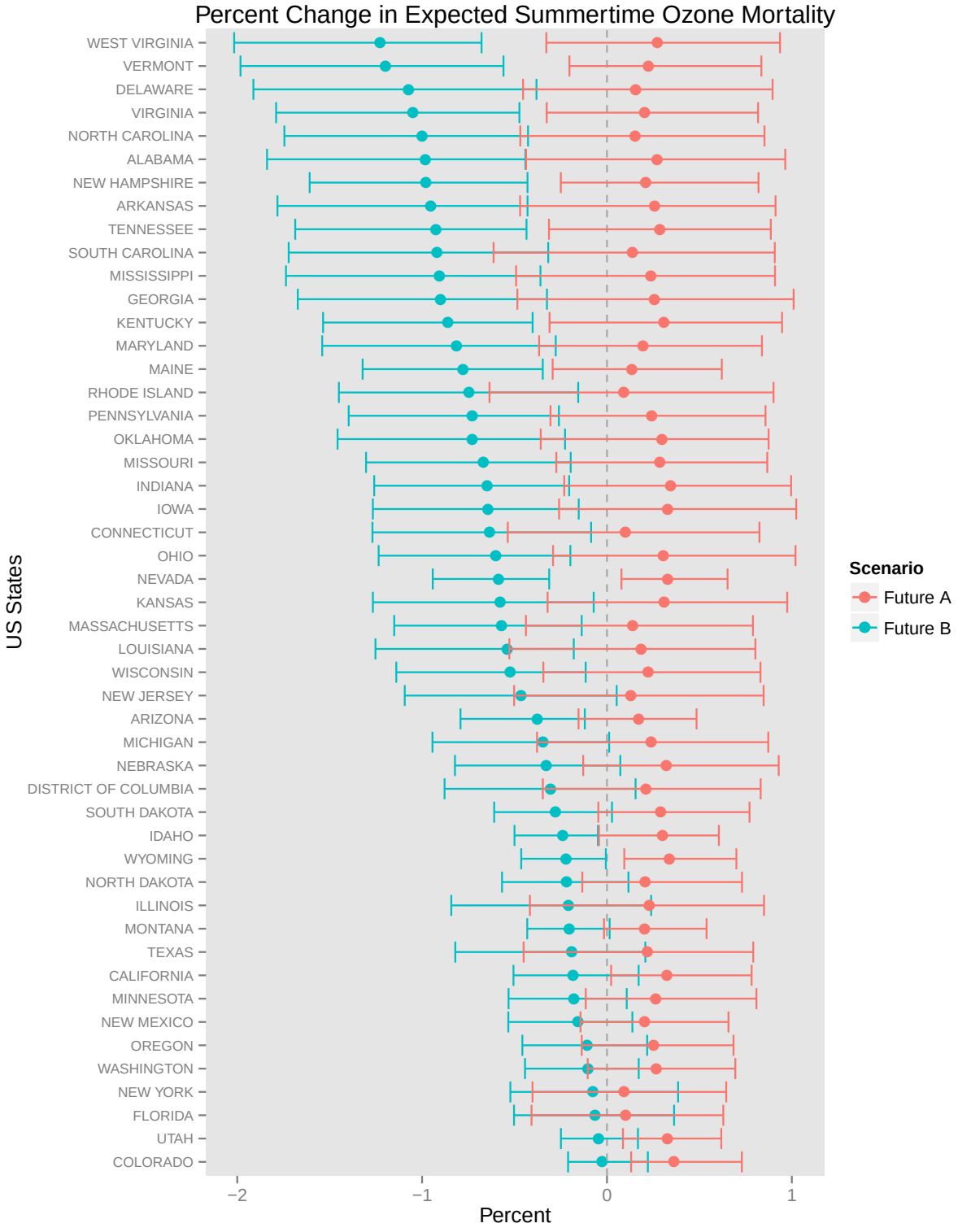
We estimate an increase of $1,212$ ozone-related deaths in summertime (95%CI:

-816 to $3,247$) in the contiguous U.S. under Future A compared to the present. In contrast, we estimate a change of $-2,543$ ozone-related deaths in summertime (95%CI: $-4,473$ to -794) under Future B in the contiguous U.S. compared to the present. Supplementary Table 1 of Supplementary Appendix B gives the results for the estimated number and 95% CI of ozone-related deaths in summertime in each state.

The mortality predictions show regional patterns in the expected degree of impact. Figure 2 shows the percent changes in expected total summertime mortality attributable to ozone for U.S. states. Supplementary Table 2 of Supplementary Appendix B gives the numeric version of the results shown here. Under Future A compared to the present, states most likely to see an increase total summertime mortality attributable to ozone include California and Nevada as well as the Rocky Mountain states of Colorado, Utah, and Wyoming. The northern Rocky Mountain states of Idaho and Montana show a similar pattern for Future A with an estimated 90% probability of an increase total summertime mortality attributable to ozone. States in the Mississippi Valley (Mississippi, Louisiana, Arkansas, Alabama, Tennessee, Kentucky, and Missouri) have a slight increase in mortality under Future A that may not be different than the present-day ozone mortality, while under Under Future B they have a predicted substantial decrease in mortality. In the pacific Northwestern states of Oregon and Washington the expected range of effects is small under Future A and Future B, with no clear differences compared to the present in either scenario. In Florida and New York the is estimated change in mortality in each Future shows no difference with the present.

Our estimation procedures differ from those used in previous studies, as described in Section 3.5. To illustrate the differences between these methods, we also estimate the results for each county using the double averaging method outlined in Section 3.5. Figures 2 and 3 in the Supplement show the results for each county in the state

Figure 2: *Percent change in expected total non-accidental summertime mortality attributable to changing ozone concentrations in U.S. states under two future emissions scenarios.*



of Georgia using our Monte Carlo approach versus the double averaging approach. The bias in the mean point estimate and the loss of variability are both evident. As expected, the double averaging intervals are much narrower and misleading because they do not reflect any variation in ozone concentrations. In contrast, our credible intervals reflect the variability in the daily ozone projections. The bias in the point estimate is most pronounced when the difference ozone concentrations is farther away from zero, and the narrow width of the confidence intervals is most pronounced when the difference ozone concentrations is closest to zero. This behavior is consistent with the specifc nonlinear nature of the exponential function.

We also consider estimating the difference in summertime mortality using the Monte Carlo simplification based on the exact integral solution with respect to β, as described in Section 5.2. To create a credible interval using these exact solutions for the mean and variance, we need to assume a distribution for the difference in deaths. A Normal distribution would be a reasonable choice. As an example, we computed the mean and variance for one state and consider the Normal distribution. Figure 4 in Supplement Appendix B shows a histogram of the Monte Carlo samples versus the Normal distribution using the mean and variance.

6 Discussion

We developed a Bayesian framework and a Monte Carlo simulation method to estimate expected summertime mortality attributable to change in ozone between 2000 and 2050. The key features of our methodology are (i) the propagation of uncertainty in both the health effect and the ozone projections and (ii) use of the empirical distribution of the daily ozone projections to account for their variation. In contrast to previous studies which use a double averaging method to reduce the size of the ozone data and ignore its uncertainty, we wrote efficient code and used database structures to regrid the data and maintain its daily structure, gaining more information about the uncertainty of the ozone projections. Overall, this approach

yields valid statistical inferences that give us new scientific insights into a key public health concern.

We gained an understanding of two future scenarios, showing that U.S. emission reduction strategies may benefit human health by reducing ozone mortality even in the face of a warming climate, with clear benefits in most states. In contrast, we showed that if U.S. emissions stay at their current levels under a warming climate, ozone mortality will stay the same or increase, depending on location.

The computational problems solved by this paper are the regridding step and the sampling of daily ozone concentrations instead of double-averaging. Double-averaging is a way to handle the high-dimensional nature of this Big Data, by requiring the regridding step only a few times, but this results in a substantial loss of information. We illustrated the problems of double averaging, including bias in point estimates, and confidence intervals that ignore the variability of ozone and thus are too narrow. In addition, we showed that double averaging makes inferences and interpretations about future mortality more difficult.

We also derived an exact solution to one of the integrals to reduce the Monte Carlo computational burden. However, using the mean and variance requires the additional assumption of normality, and further investigation is needed to understand the validity of this assumption and the benefits relative to the Monte Carlo sampling.

Tagaris et al. [2009] estimated approximately 300 more ozone-related premature deaths nationally based on the emissions scenario of IPCC-A1B, in contrast to our estimate of about 1,200 more deaths under IPCC-A2 with the continuation of present-day emissions. The global warming effects are more gradual in IPCC-A1B scenario compared to the IPCC-A2 scenario, so the difference in these estimates seems reasonable. In addition, Tagaris et al. [2009] used only one year of ozone data, which may not be entirely representative of the ozone distribution for the future under IPCC-A1B.

In our application, one assumption is that current population and current mortality rates do not change in the future. Of course, population is expected to increase and mortality rate is expected to decrease [Ortman and Guarneri, 2009]. To complicate matters, these changes are expected to vary demographically (e.g. more elderly people) and spatially, and those estimates are not readily available to incorporate into this analysis. The methodology proposed in this paper can be applied to changing future mortality rate estimates and future population estimates once available, and those results can be contrasted with these results to separately estimate the effect of ozone alone versus the effect of changing ozone and changing population.

There are additional sources of uncertainty not accounted for in this study, such as changing population, other climate change scenarios, and uncertainty of the climate model within a scenario. However, our approach offers a consistent statistical framework conditional on the validity of the models for health and ozone, where point estimates and credible intervals can be reported that reflect the statistical uncertainty given the modeling assumptions. Overall, this statistical methodology offers many advantages and can be used to gain better quantitative insight into many pressing health questions involving climate change.

References

M. Bell, A. McDermott, S. Zeger, J. Samet, and F. Dominici. Ozone and short-term mortality in 95 us urban communities, 1987-2000. *JAMA*, 292:2372–2378, 2004.

M. Bell, F. Dominici, and J. Samet. A meta-analysis of time-series studies of ozone and mortality with comparison to the national morbidity, mortality, and air pollution study. *Epidemiology*, 16:436–445, 2005.

M. L. Bell, R. Goldberg, C. Hogrefe, P. L. Kinney, K. Knowlton, B. Lynn, J. Rosenthal, C. Rosenzweig, and J. A. Patz. Climate change, ambient ozone, and health in 50 us cities. *Climatic Change*, 82:61–76, 2007.

CDC. Mortality Multiple Cause-of-Death Public Use Record 2000.

http://www.cdc.gov/nchs/nvss/$mortality_public_use_data.htm$. Accessed: January 27, 2014.

J. Done, G. Holland, C. Bruye, L. Leung, and A. Suzuki-Parker. Modeling high-impact weather and climate: lessons from a tropical cyclone perspective. *Climatic Change*, pages 1–15, 2013. ISSN 0165-0009. doi: 10.1007/s10584-013-0954-6. URL http://dx.doi.org/10.1007/s10584-013-0954-6.

U. S. EPA. Integrated science assessment of ozone and related photochemical oxidants (final report). *U.S. Environmental Protection Agency, Washington, DC*, EPA/600/R-10/076F:1–7, 2013.

IPCC. *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007*. 2007.

K. Ito, S. F. D. Leon, and M. Lippmann. Associations between ozone and daily mortality: Analysis and meta-analysis. *Epidemiology*, 16:446–457, 2005.

S. Kim, A. Heckel, G. Frost, A. Richter, J. Gleason, J. Burrows, S. McKeen, E. Hsie, C. Granier, and M. Trainer. No2 columns in the western united states observed from space and simulated by a regional chemistry model and their implications for nox emissions. *J. Geophys. Res.*, 114(D11301):1–29, 2009. doi: 10.1029/2008JD011343.

J.-F. Lamarque, G. Kyle, M. Meinshausen, K. Riahi, S. Smith, D. Vuuren, A. Conley, and F. Vitt. Global and regional evolution of short-lived radiatively-active gases and aerosols in the representative concentration pathways. *Climatic Change*, 109 (1-2):191–212, 2011. ISSN 0165-0009. doi: 10.1007/s10584-011-0155-0. URL http://dx.doi.org/10.1007/s10584-011-0155-0.

J. I. Levy, S. M. Chemerynski, and J. A. Sarnat. Ozone exposure and mortality: an empiric bayes metaregression analysis. *Epidemiology*, 16:458–468, 2005.

G. Meehl, C. Covey, T. Delworth, M. Latif, B. McAvaney, J. Mitchell, R. Stouffer, and K. Taylor. The wcrp cmip3 multimodel dataset - a new era in climate change research. *Bull. Am. Meteorol. Soc.*, 88:13831394, 2007.

J. M. Ortman and C. E. Guarneri. United states population projections: 2000 to 2050. *United States Census Bureau*, 2009.

G. G. Pfister, S. Walters, J.-F. Lamarque, M. C. B. J. Fast, J. Wong, J. Done, G. Holland, and C. L. Bruyre. Projections of future summertime ozone over the u.s. *J. Geophys. Res. Atmos.*, 119, 2014. doi: doi:10.1002/2013JD020932.

E. S. Post, A. Grambsch, C. Weaver, P. Morefield, J. Huang, L.-Y. Leung, and C. G. Nolte. Variation in estimated ozone-related health impacts of climate change due to modeling choices and assumptions. *Environmental health perspectives*, 120:1559, 2012.

D. Stieb, S. Judek, and R. Burnett. Meta-analysis of time-series studies of air pollution and mortality: effects of gases and particles and the influence of cause of death, age, and season. *J Air Waste Manag Assoc*, pages 470–484, 2002.

E. Tagaris, K.-J. Liao, A. J. DeLucia, L. Deck, P. Amar, and A. G. Russell. Potential impact of climate change on air pollution-related human health effects. *Environmental science and technology*, 43:4979–4988, 2009.

U.S. Census Bureau. Census 2000 U.S. Gazetteer Files and TIGER/Line Shapefiles. `http://www.census.gov/geo/maps-data`. Accessed: January 27, 2014.

D. P. van Vuuren, J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G. C. Hurtt, T. Kram, V. Krey, J.-F. Lamarque, T. Masui, M. Meinshausen, N. Nakicenovic, S. J. Smith, and S. K. Rose. The representative concentration pathways: an overview. *Climatic Change*, 109:5–31, 2011.