Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Brooke Alhanti Date

Methods for Estimating the Effect of Air Pollution on Asthma under a Changing Climate

By

Brooke Alhanti

Doctor of Philosophy

Biostatistics

Howard H. Chang, Ph.D.
Advisor

Lance A. Waller, Ph.D.
Committee Member

Stefanie E. Sarnat, Sc.D.
Committee Member

Yang Liu, Ph.D.
Committee Member

Yi-Juan Hu, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Methods for Estimating the Effect of Air Pollution on Asthma under a Changing Climate

by

Brooke Alhanti
MPH, Emory University, 2009
BS, Brigham Young University, 2007

Advisor:

Howard H Chang, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2016

Abstract

Methods for Estimating the Effect of Air Pollution on Asthma under a Changing Climate

By
Brooke Alhanti

Climate models are complex mathematical representations of global and regional climate, which vary over space and time. Estimates of local meteorological variables provided by climate models are biased. More accurate estimates of future meteorological conditions are essential to investigating health impacts of climate change. Here we consider the association between ambient air pollution and respiratory disease under a changing climate. We use regional climate and air quality model outputs and develop calibration methods that aim to produce more accurate projections.

We first assess the association between ambient ozone (O_3) and fine particulate matter ($PM_{2.5}$) and asthma in the metro Atlanta area by age under recent meteorological conditions (1993-2009). We then propose a quantile-mapping approach where the quantile values of climate model outputs and historical data are regressed against each other with integrated piecewise splines to create calibrated projections.

We utilize copulas to develop a bivariate quantile calibration method that simultaneously calibrates the marginal distribution of each variable while capturing dependence between variables. This method estimates the bias between monotonic increasing quantile functions for climate models and monitoring data and applies this estimated bias to future climate projections. The Gumbel and Frank copulas are used to estimate the dependence between daily average temperature and solar radiation. Inference is conducted under a Bayesian framework to account for all sources of uncertainty in the projection. A cross-validation study is performed using historical data to evaluate the proposed approach. We apply our method to projections from four different climate models in Atlanta and find evidence for higher mean temperature and lower mean solar radiation in 2041-2070 compared to 1991-2000.

Finally, we consider spatial dependence in climate model outputs over a contiguous grid. We develop a spatial calibration method that calibrates climate model outputs over multiple locations and projects calibrated values on a gridded area. This spatial extension allows calibrated projections at locations that do not have a monitor. We project future levels of two air pollutants: O_3 and $PM_{2.5}$ on a spatial field covering northern Alabama and Georgia. We subsequently use these projected pollutant concentrations to project future asthma cases in Atlanta and Birmingham.

Methods for Estimating the Effect of Air Pollution on Asthma under a Changing Climate

by

Brooke Alhanti
MPH, Emory University, 2009
BS, Brigham Young University, 2007

Advisor:

Howard H Chang, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2016

Contents

List of Figures

List of Tables

Specific Aims

Over the past 50 years it has become clear that the warming of the Earth's climate system is unequivocal. However, how much warming will occur and the local effects of global climate change are uncertain. Climate models, which are mathematical representations of climate, are used to understand and predict climatic behavior. They are complex models based on atmospheric, ocean, and land variables, which vary over space and time. Climate models are built on a global grid system and then downscaled to regional and local levels. The different combinations of climate models and downscaling methodologies results in estimates that retain a large amount of uncertainty, however, because climate models are deterministic much of that uncertainty is unmeasured.

Interest in the future impacts of climate change is increasing as scientists and policymakers have realized accurate predictions of local impacts are needed to justify the social and economic costs of combating climate change. One such impact of interest is how climate change will affect human health. Climate change has the potential to impact human health in a myriad of ways, from direct impacts such as heat-related illness to more complicated impacts such as forced migrations due to environmental degradation. How climate change will affect air pollution and related health effects is a particularly complex question. Meteorological variables play a large role in the formation, transport, and removal of ambient air pollution; therefore, how these variables change in the future will directly influence air pollution concentrations.

Fine particulate matter ($PM_{2.5}$) concentrations are affected by temperature, solar radiation, and precipitation. Heat and solar radiation can influence the chemical constituents of $PM_{2.5}$, and precipitation can lower the concentration of $PM_{2.5}$ in ambient air. Ozone is particularly sensitive to temperature and solar radiation as it is a secondary pollutant formed by the photochemical oxidation of carbon monoxide and volatile organic compounds in the presence of nitrous oxides and sunlight. Many scientists believe as the Earth warms, ozone levels

will increase (assuming current emissions of greenhouse gases remain somewhat constant) [4, 5].

The particles that make up $PM_{2.5}$ are small enough to be inhaled deep into the lungs, leading to worsened respiratory health. $PM_{2.5}$ has been linked to asthma in numerous epidemiological and clinical studies. Ozone is also known to exacerbate asthma by leading to inflammation in the airways. Patients with asthma are more likely to have respiratory distress when exposed to ozone than non-asthmatic patients. Furthermore, rates of asthma are increasing globally at a pace too rapid to be explained by genetics or changes in diagnostic characterization. Scientific inquiry has looked to environmental exposures to explain this increase [6].

The relationships between meteorological variables, air pollution, and asthma naturally lead to the question: How will climate change impact asthma in the future? This dissertation contributes to a growing body of scientific literature informing the answer.

Aim 1. Determine the association between ambient air pollutants (O_3 and $PM_{2.5}$) and asthma in Atlanta. To address this aim, we examine the association between population-weighted ambient air pollution concentrations measured daily from 1993-2009 in Atlanta, GA and emergency department (ED) visits for asthma. Daily time-series analyses are conducted to estimate associations by age group (0-4, 5-18, 19-39, 40-64, and 65+ years). Sub-analyses are performed stratified by race and sex.

Aim 2. Simultaneously calibrate climate model outputs with multiple variables. Calibration methods can be used to improve the accuracy of climate change model predictions on the local level. Previous methods only used one variable at a time, resulting in the loss of dependence between variables. We develop a method that calibrates climate model output for multiple variables simultaneously (e.g., temperature and solar radiation) by using quantile functions and copulas. This method also provides a measure of uncertainty.

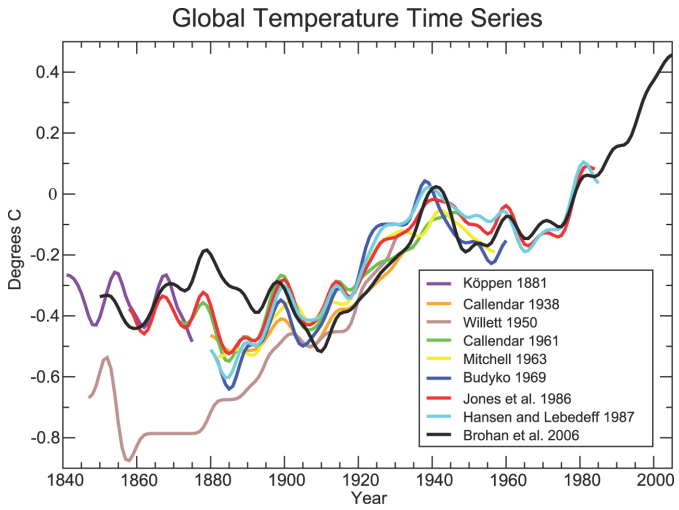
Aim 3. Extend climate model calibration to spatially to project future ambient air pollution and asthma ED visits We extend our calibration method beyond a single location to multiple 12 km x 12 km grid cells over northern Georgia and Alabama. This extension is accomplished via the inclusion of spatially-varying coefficients in our calibration approach. Calibrated future levels of O_3 and $PM_{2.5}$ are used to project how asthma morbidity will be effected by the changing climate, based on observed associations between ambient air pollutants and asthma morbidity discussed in Aim 1.

1 Introduction

The Earth's climate system is an interactive system consisting of the atmosphere, land surface, water, and living organisms. Climate, which is typically described in terms of temperature, precipitation, and wind, is often defined as "average weather" and is influenced by both internal and external dynamics. These external dynamics are referred to as "forcings" and can be natural (e.g., volcano erupting) or anthropogenic (e.g., carbon emission from power generation) [7]. Over the past 50 years, it has become clear that the "warming of the climate system is unequivocal" and that anthropogenic forcing has contributed to climate change [8]. However, how much warming will occur and where temperatures changes will be the most drastic is still uncertain. Climate scientists try to answer these questions with climate models.

Climate models are mathematical representations of climate used to understand and predict climatic behavior. They are based on a large number of variables from the atmosphere, ocean, and land, and must make many simplifying assumptions to perform complex calculations. These complicated models output a large number of climate variables and the quality of the estimation of individual output variables is dependent on the climate model and it's inputs. The figure below shows an example of the challenges in estimating observed temperature alone. Each line represents an estimate of surface temperature over large regions over time. Some models (Köppen) used land air temperatures while others (Brohan) used both land air temperature and sea surface temperatures. All time series were smoothed and adjusted to be comparable [7].

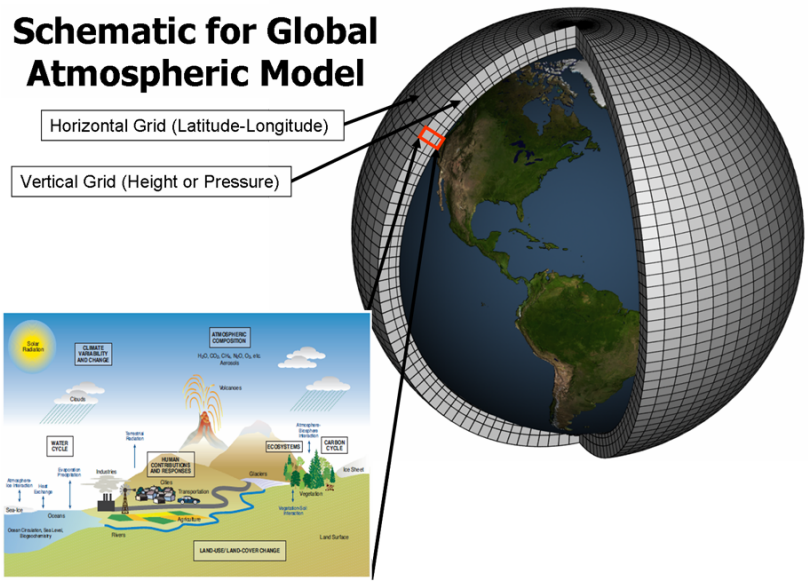
Figure 1: Global surface temperature models over time (1840-2000) [9]



This example illustrates how climate models must use multiple data sources to produce their estimates. When projecting future climate, models also must take into account the current changing climate as well as potential changes in anthropogenic forcing. In order to simplify future projections, the International Panel on Climate Change (IPCC) developed four scenarios, called Representative Concentration Pathways (RCPs) to represent a range of climate outcomes from low future greenhouse gas (GHG) emissions to very high future GHG emissions. RCPs are defined by their total radiative forcing (a cumulative measure of human emissions of GHGs in Watts/sq^2) [8]. All recent climate change model predictions take place under one of the RCP scenarios.

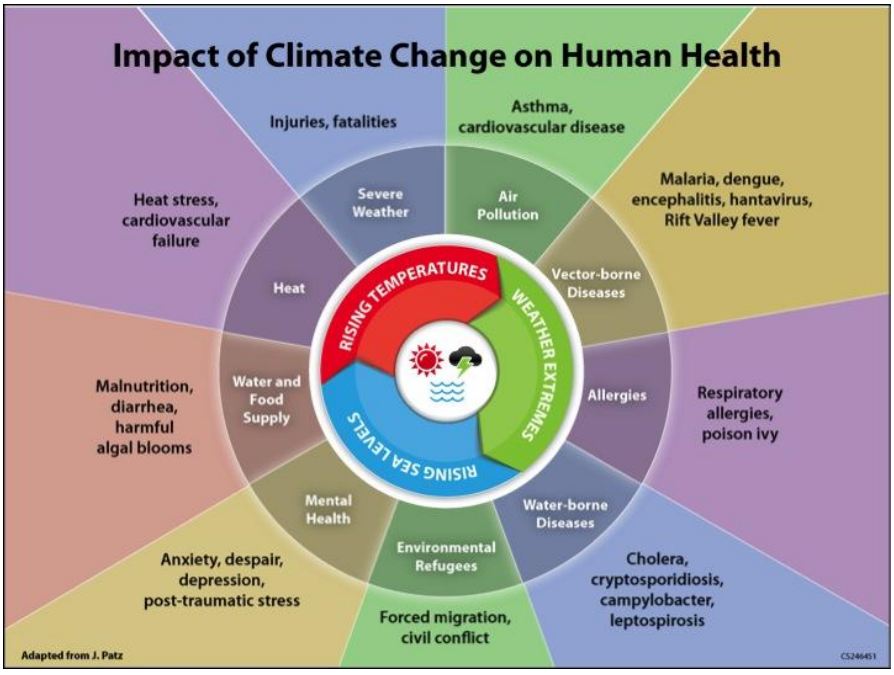
Modern climate models are typically divided into two categories: Global Circulatory Models (GCMs) and Regional Climate Models (RCMs). GCMs model the entire Earth's climate based grid cells over time. The size of the grid cell and the time-step is dependent partly on computing power. RCMs are nested models which take GCM levels for a region (e.g., Europe or North America) and use those levels as initial and boundary conditions for a more finely gridded model. RCMs are used to inform assessments of the impact of climate change on regions and nations as local information is necessary to justify social and economic policies related to climate change adaptation and mitigation [10].

Figure 2: Schematic for global atmospheric model [2]



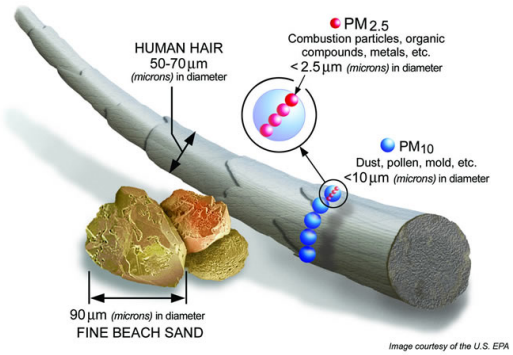
Simulating the most realistically accurate climate models possible is important for assessing the potential impact of climate change on human health. Adaptation and mitigation to the changing climate are complex and expensive endeavors with high political and economic costs. The World Meteorological Organization asserted, "only by assessing what the real impact will be on different countries will it be possible to justify difficult social and economic policies to avert a dangerous deterioration in the global climate" [10]. Climate change has the potential to endanger human health across the globe. The figure below was developed by the Centers of Disease Control and Prevention (CDC) to illustrate the complex and intertwined relationship between climate change and human health [3]. The inner circle shows how climate change factors(e.g., rising temperatures, increasing CO_2 levels, etc.) lead to outcomes such as extreme heat or air pollution, which in turn lead to adverse health events (e.g., asthma, cardiovascular disease, etc.).

Figure 3: Climate change impacts on human health [3]



Rising temperatures and unmitigated GHG emissions will affect several types of air pollution. Particulate matter (PM) is a mixture of solid particles and liquid droplets found in the air. $PM_{2.5}$ is the designation used for particles with diameters less than 2.5 μg [4]. $PM_{2.5}$ concentrations are affected by weather, such as heat and precipitation. Temperature and solar radiation can influence the chemical composition of $PM_{2.5}$'s constituents and precipitation can reduce the overall concentration in ambient air [11]. $PM_{2.5}$ is particularly dangerous to human health as the particles are small enough to be inhaled deep into the lungs, beyond the lungs natural defenses (Figure 4). $PM_{2.5}$ causes inflammation of the airways, resulting in decreased lung function, respiratory symptoms (e.g., coughing and difficulty breathing), and aggravated asthma [12, 13].

Figure 4: Fine particulate pollution [4]



Another pollutant impacted by climate change is ground-level ozone. Tropospheric ozone (O_3) is formed through photochemical reactions involving carbon monoxide (CO), volatile organic compounds (VOC), and nitrogen oxides (NO_x) in the presence of sunlight. These reactions occur more often in the presence of higher temperatures and solar radiation leading to higher O_3 concentrations on hot, sunny days [5]. The projected effect of climate change on ozone concentrations varies by location. As temperatures rise, air quality models project increased ozone production; however, higher temperatures also lead to faster destruction of ozone. As ozone is a secondary pollutant, the role of changing emissions standards will also affect the level of ozone concentrations. Within all the uncertainty about projected global ozone levels, there is a growing consensus that ozone levels will increase in urban areas [14, 15, 16, 17, 18].

Exposure to ambient ozone can lead to a number of adverse health outcomes, including increasing the likelihood of asthma-related events [19, 20, 21, 14, 22, 23]. Exposure to ozone induces airway inflammation, resulting in respiratory responses [18]. Patients with asthma are at increased risk for adverse effects of ozone exposure because they have weakened pulmonary reserves to tolerate airway inflammation and ozone interacts with the underlying pathophysiology of asthma. Ozone exposure can induce airway inflammation, which appears to worsen a patient's underlying asthma status. Studies have shown asthmatic patients exposed to ozone had a greater influx of markers of airway inflammation than

non-asthmatic controls [24, 6]. The observed epidemiological links between ambient ozone and asthma are supported by experimental chamber studies and suggest a causal relationship between ozone exposure and increased likelihood of an asthmatic response [6].

In the chapters below we also explore calibration of climate model outputs as a way to increase the accuracy of climate projection models. We propose a method to calibrate climate model outputs to historical observed data and then apply the relationship to future projections. The calibration is done using quantile functions specified using I-splines (see Section 3.3). The quantile relationship between the observed data and the projected estimates is determined and then used to predict more accurate climate variable estimates in the future. Quantile functions are useful in climate research as they can capture tail probabilities better than mean regression techniques. These tail probabilities are important in climate research as extreme values are often the most related to adverse health events, such as heat waves. Quantile functions modeled with splines also allow us to flexibly capture smooth non-linear relationships. Multiple model variables can be used to inform the calibration of a single climate variable through creating multivariate distributions with a copula, which provides a method for characterizing a joint distribution from two variables. Once bivariate calibration has been performed at a single location the calibration can be extended to a spatial field encompassing multiple climate model grid cells. This spatial calibration is achieved through the use of spatially varying calibration parameters. Spatial calibration of projected concentrations of ozone and $PM_{2.5}$ are used to project changes in future asthma morbidity resulting from changes in ambient air pollutant concentrations.

2 Aim 1: Air Pollution and Respiratory Disease

2.1 Background

Many epidemiological studies have documented relationships between ambient air pollution and asthma [20]. Adverse effects of ambient air pollution on asthma appear to disproportionately affect children [25, 26, 27, 28, 29, 30], who also have the highest prevalence of asthma [31]. The increased susceptibility of children may be due to physiological differences between adults and children (underdeveloped pulmonary and immune systems and smaller airways that become more obstructed when inflamed) and higher levels of outdoor physical activity in children [32, 33, 34, 22]. Only limited previous studies consider whether and how associations differ by age group. Analyses comparing air pollution effects across age groups can help to identify susceptible populations —an essential component for both focused prevention efforts and accurate risk assessment. Here we investigated short-term associations between ambient air pollution concentrations and asthma emergency department (ED) visits across five age groups in Atlanta, GA. Additional stratification by sex and race was also considered to examine whether the differences in asthma-air pollution associations across age groups may be explained by sex and race.

Air quality and ED visit data were collected in 20 counties around Atlanta (years 1993-2009). The air quality data included daily population-weighted average concentrations of ambient 8-hour maximum O_3 (ppb) and 24-hour average $PM_{2.5}$ ($\mu g/m^3$). The population-weighted average concentrations were calculated using data from all available monitors for each pollutant in the study regions. We included monitors from the US EPA Air Quality System, the SouthEastern Aerosol Research and Characterization (SEARCH) network, and the Assessment of the Spatial Aerosol Composition (ASACA) network (Figure 5). The monitoring data were first spatially interpolated across the studys geographic domain and then a population-weighted average concentration estimate was derived using

tract-level population estimates from the 2010 US Census (Figure 6). Daily minimum, maximum, and dew-point temperature data were obtained from the National Climate Data Center.

Figure 5: Air pollution monitoring locations in the Atlanta metro region with spatial domain of study area

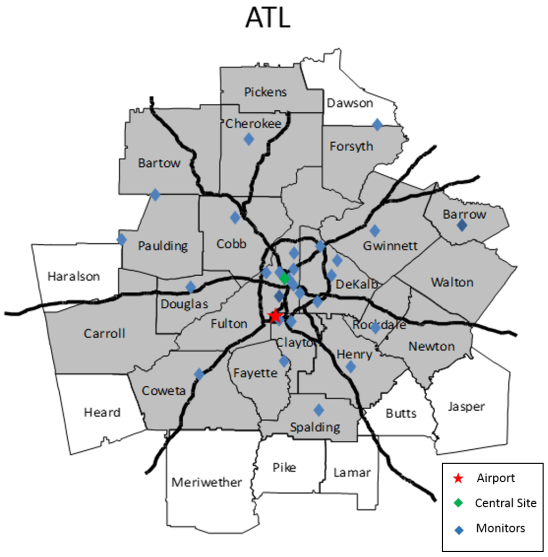
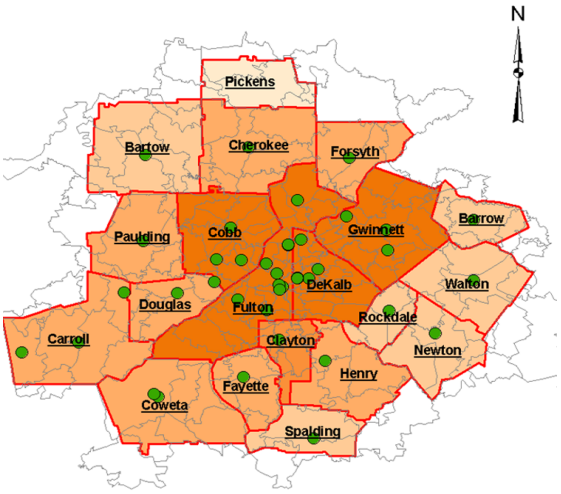


Figure 6: Locations of Atlanta hospitals providing daily emergency department (ED) data from 1993-2009



Patient-level ED visit data were obtained from individual hospitals and the Georgia Hospital Association. Asthma ED visits were identified using the primary International Classification of Diseases, 9th Revision diagnosis code for the record as either 493 (asthma) or 786.07 (wheeze). Identified visits were aggregated by

day to obtain daily counts of asthma ED visits overall, and for five age groups: 0-4, 5-18, 19-39, 40-64, and 65+ years.

2.2 Analysis

For each age group, daily time-series Poisson regression models were used to assess the relationship between asthma ED visit counts and ambient pollutant concentrations. Analyses were conducted separately for each age group.

$$log(Y_{ik}) = \beta_0 + \beta_1 x_i + \beta_2 dw_i + \sum_{a=1}^{7} \beta_a hol_{ai} + \sum_{b=1}^{9} \beta_b hospid_{bi} + \sum_{c=1}^{3} \beta_c n(temp0_{ci}) + \sum_{d=1}^{3} \beta_d n(temp12_{di}) + \sum_{e=1}^{3} \beta_e n(dp12_{ei}) + \sum_{f=1}^{8} \beta_f n(time_{fi})$$

Y_{ik}: Daily ED count (i) for age group k
x_i: Daily 3-day moving average pollutant concentration
$n()$: Natural cubic spline

In these analyses, our a priori temporal metric for all air pollutant concentrations was a 3-day moving average (of the current day, the previous day, and 2 days prior), chosen based on previous research [22, 35, 36]. The models controlled for meteorology via natural cubic splines with 3 degrees of freedom (df) of lag 0 minimum temperature, 2-day moving average (of the previous day, and 2 days prior) minimum temperature, and 2-day moving average (of the previous day, and 2 days prior) dew point temperature. Temporal controls included indicator variables for day of week, federal holidays, and days commonly celebrated as holidays (e.g., the day after Thanksgiving) and a natural cubic spline of time (df=8 per year). The models also included indicator variables for the periods when there were changes in the number of hospitals that contributed to the total ED counts over time. Standard errors of the model coefficients were scaled to account for over-dispersion in ED counts.

Rate ratios (RR) were scaled to a pollutant-specific standard unit. The standard unit for each pollutant was determined as the interquartile range (IQR) of data: 28 ppb for O_3 and 8 $\mu g/m^3$ for $PM_{2.5}$. Pairwise statistical significance for the difference in age-specific RRs was assessed based on standard errors obtained from independent stratified analyses at a type I error rate of 0.05, assuming the RRs estimated from the separate datasets were not correlated.

Several sensitivity analyses on model specification were performed. Various degrees of freedom for all non-linear effects were explored as well as different meteorological controls, such as the use of maximum temperature versus minimum temperature. Sub-analyses were also performed to examine whether the pattern of age-specific associations between pollutant concentrations and asthma ED visits differed by sex and race. For these analyses, race was dichotomized into white and non-white (black, Hispanic, and other). Race-specific models were only considered for 1993 through 2006 because information on race was unavailable after 2006. Atlanta data between 1993 and 2006 were also missing race information to varying extents by age group (as high as 56% for asthma ED visits in the 0-4 year age category). All analyses were performed in R 2.15.1 (R Foundation for Statistical Computing, Vienna, Austria).

2.3 Results

Descriptive statistics for the pollutants are shown in Table 1. Asthma ED counts totaled 389,771 in Atlanta. Mean daily counts were highest in the 5-18 year age group, followed by the 0-4 year age group (Table 2). Males had approximately twice the mean daily counts of females in the younger age groups (0-4 and 5-18 years); in the older age groups this pattern reversed to females having about twice as many mean daily cases as males. Non-whites had more mean daily cases in the younger age groups than whites, but as age increased mean daily case counts became more even between groups. Time series plots show the patterns

of asthma ED visits by age group (Figure 7) and pollutant concentrations over time (Figure 8). The younger age groups (0-4, 5-18) have a seasonal pattern of asthma ED visits, while the older age groups have less temporal variance. The pollutant concentrations have a clear seasonal pattern and $PM_{2.5}$ has an overall slight decrease over time.

Table 1: Descriptive Statistics of Population Weighted Averages of Ambient Pollutant Concentrations in Atlanta (1993-2009)

Pollutant	**Daily Mean**	**SD**
O_3 (ppb)	43.7	19.1
$PM_{2.5}$ ($\mu g/m^3$)	14.1	6.8

Table 2: Mean Daily Counts of Asthma ED Visits in Atlanta (1993-2009), by Age Group: Total and by Sex and Race

Age Group	**Total**	**Male**	**Female**	**White**	**Non-white**
0-4 years	16.78	10.42	6.24	2.22	4.5
5-18 years	18.24	10.75	7.3	2.78	6.2
19-39 years	13.36	4.65	8.53	4.39	6.16
40-64 years	11.33	3.60	7.62	3.68	4.53
65+ years	1.53	0.49	1.04	0.75	0.48

Figure 7: Daily counts of asthma emergency department (ED) visits by age group for 2006-2009 in the Atlanta

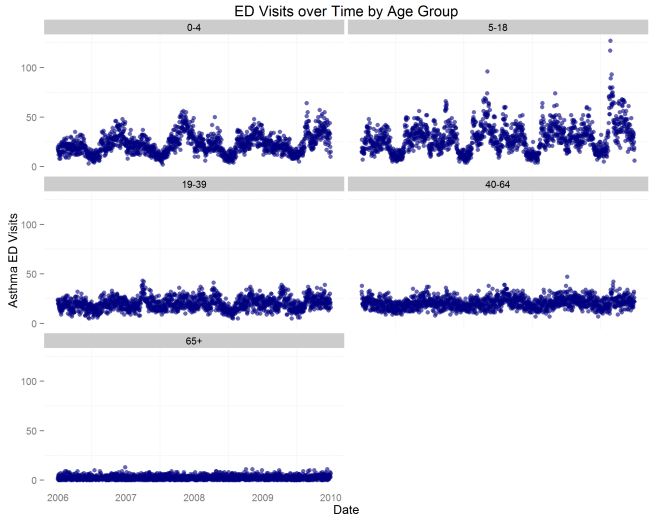
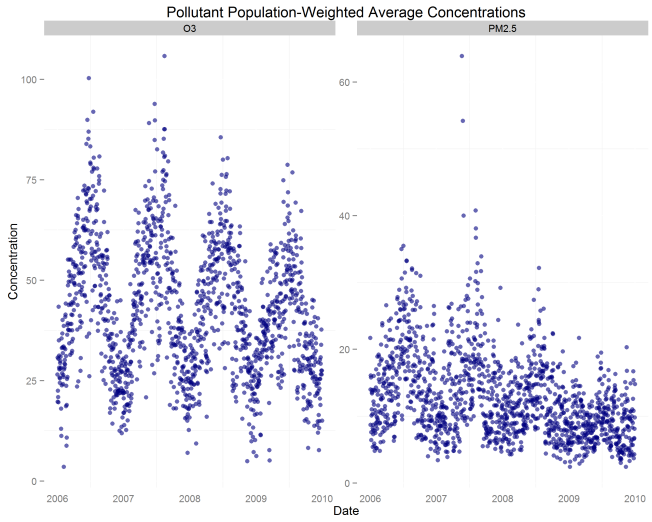


Figure 8: Daily population-weighted average concentrations of O_3 and $PM_{2.5}$ for 2006-2009 in the Atlanta



The strongest associations between pollutant concentrations and asthma ED visits were found in the 5-18 year age group (Table 3). Ambient O_3 was more strongly associated with asthma ED visits (RR per 28 ppb=1.05, 95% confidence interval (CI): 1.01, 1.08) in the 5-18 year old age group compared to the other age groups. The 19-39 age group also had a statistically significant positive association (RR per 28 ppb=1.04, CI: 1.01, 1.07). Ambient O_3 concentrations and asthma ED visits were not significantly associated in the other age groups with rate ratios per 28 ppb of 1.00 (CI: 0.98, 1.03) in 0-4 year olds, 1.01 (CI: 0.98, 1.04) in 40-64 year olds, and 1.01 (CI: 0.96, 1.07) in those 65 years and older. For ambient $PM_{2.5}$ a statistically significantly positive association with asthma ED visits was found in the 5-18 year old age group with a RR per $8\mu g/m^2$ of 1.03 (CI: 1.01, 1.05). Ambient $PM_{2.5}$ was also significantly associated with the 40-64 year olds (RR per $8\mu g/m^2$=1.02, CI: 1.00, 1.04), and 65+ year olds (RR per $8\mu g/m^2$=1.04, CI: 1.00, 1.08). Young children aged 0-4 years old (RR per $8\mu g/m^2$=1.01, CI: 0.99, 1.03) and 19-39 year olds (RR per $8\mu g/m^2$=1.01, CI: 1.00, 1.03) did not have a significant association between ambient $PM_{2.5}$ and asthma ED visits.

Table 3: Rate Ratios for the Association Between a Standard Unit Increase in Three-Day Moving Average Pollutant Concentrations and Asthma ED visits in Atlanta (1993-2009)*

Pollutant	**Age Group**	**RR**	**95% CI**	
O_3	0-4	1.004	0.977	1.031
	5-18	1.046**	1.014	1.079
	19-39	1.041**	1.011	1.071
	40-64	1.010	0.978	1.042
	65+	1.011	0.955	1.070
$PM_{2.5}$	0-4	1.009	0.993	1.025
	5-18	1.028**	1.009	1.047
	19-39	1.015	0.998	1.032
	40-64	1.020**	1.002	1.038
	65+	1.039**	1.003	1.077

* Standard units: 28 ppb for O_3 and 8 $\mu g/m^3$ for $PM_{2.5}$

** Significant at the α=.05 level

Sensitivity analyses on model specification showed similar results for a wide range of degrees of freedom for both the temperature and time splines. In general, results using minimum temperature and maximum temperature were similar. However, the use of maximum temperature attenuated the association between the O_3 concentration and asthma-related ED visits, particularly among the younger age groups (0-4, 5-18, and 19-39 years). This attenuation was more substantial as the degrees of freedom in the time spline were increased, and may be attributed to over control. Ozone formation is highly sensitive to temperature, and the Pearson correlation between 3-day moving average ozone and maximum temperature was 0.72 compared to a lower correlation of 0.60 between ozone and minimum temperature in Atlanta where both temperature variables are represented as specified in our main model lag 0 temperature plus 2-day moving average (of the previous day, and 2 days prior).

The observed trend of 5-18 year olds having higher RRs, generally persisted in the sub-analyses by sex and race. Figures 9 and 10 illustrates the pattern of RRs and corresponding 95% confidence intervals for each pollutant (axes vary by pollutant for better visualization of the results). However, the higher RRs in the

5-18 year age group appeared to be partially driven by high RRs in non-whites. Considering the results for 5-18 year olds, associations of O_3 among non-whites (RRs of 1.16 (1.10, 1.21) per 28 ppb O_3) were significantly stronger (2-sided P <0.001) than among whites (RR of 0.97 (0.91, 1.04) per 28 ppb O_3). In addition, associations were generally stronger among males than females across all age groups, although there were no statistically significant differences. No significant differences in associations between the demographic sub-groups were observed for $PM_{2.5}$, although patterns of weighted average associations by sex and race across age groups were generally similar to O_3 with weaker associations overall. The non-white 5-18 year olds had a RR of 1.16 (1.10, 1.21) compared to whites with a RR of 0.97 (0.91, 1.04).

Figure 9: Rate ratios for the association between a standard unit increase in three-day moving average O_3 concentrations and asthma ED visits in Atlanta (1993-2009)

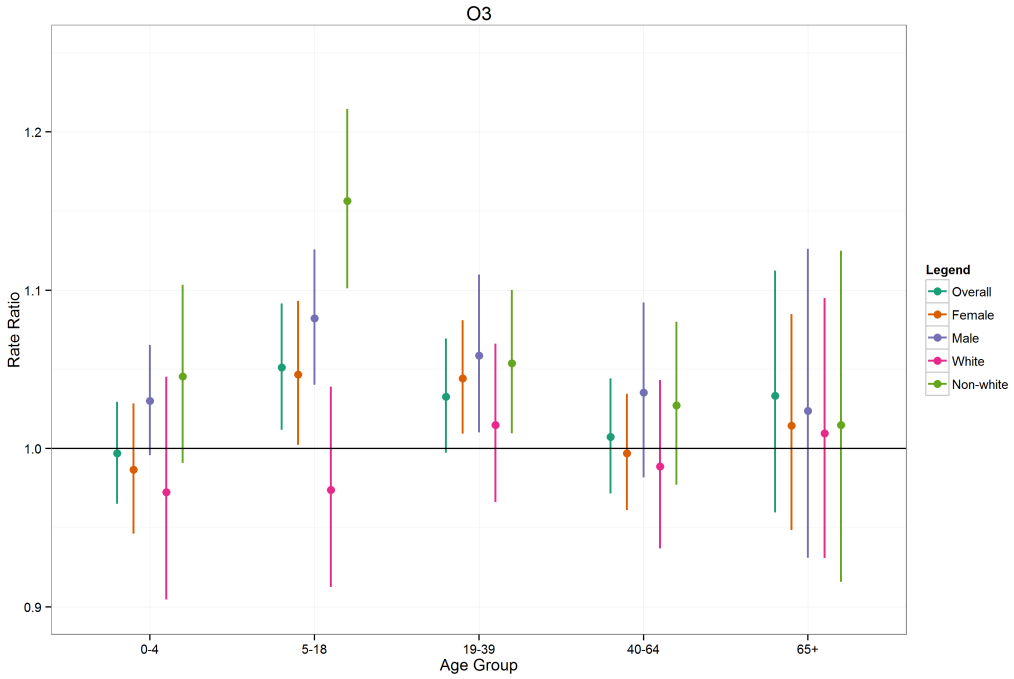
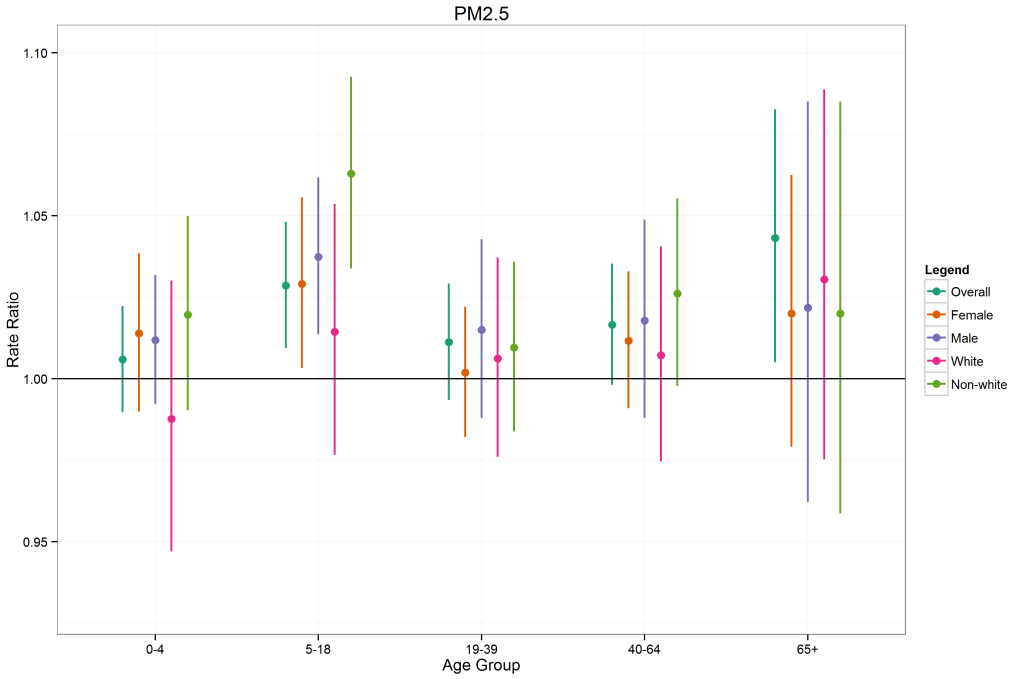


Figure 10: Rate ratios for the association between a standard unit increase in three-day moving average $PM_{2.5}$ concentrations and asthma ED visits in Atlanta (1993-2009)



2.4 Discussion

We found consistent positive associations between short-term concentrations of ambient O_3 and $PM_{2.5}$ and asthma ED visits among children 5-18 year olds compared to younger and older age groups. These results are similar to those found elsewhere. In a study of 8 cities in Korea, children under 15 years old were more likely than older age groups to be hospitalized for asthma when concentrations of O_3, NO_2, and PM_{10} were higher [27]. Likewise, short-term increases in ambient $PM_{2.5}$ and O_3 concentrations were also associated with an elevated risk of asthma-related hospitalization among 6 to 18 year olds in New York City [26]. A study on ED visits for asthma in an industrial Canadian city found stronger associations with increases in O_3, CO, NO_2, and SO_2 for children between 2 and 14 than for adults [25].

Our results of strong associations in the pediatric groups may be related to several well-documented susceptibility factors experienced by children. For

example, children may be more susceptible to harmful effects of air pollution because their less-developed respiratory system may become more obstructed when exposure to air pollution causes inflammation in their smaller airways. Children also have less developed immune systems than adults, higher ventilation rates, and higher air pollution exposures per body weight than adults, which may increase the severity of the effects of pollutant exposure [33, 34]. Children may also experience higher levels of exposure to ambient air pollution than adults because they are more physically active and spend more time outside, thus increasing the frequency and duration of their exposure [32].

We also examined sex and race as potential factors explaining observed age-related differences in air pollution-asthma associations. The pattern of strong associations among 5-18 year olds appeared to be partially driven by non-white and male patients. For example, non-whites appeared to be more vulnerable to short-term increases in ambient pollution, particularly O_3 and NO_2, than whites in the 5-18 year old age group. A study in Phoenix found both black and Hispanic children (0-14 years old) had stronger associations between NO_2 concentrations and asthma hospitalization than white children [37]. Other studies have found effect modification by race for O_3 and $PM_{2.5}$. Strickland et al. found children aged 2 to 16 with African American mothers had greater susceptibility to effects of O_3 on asthma ED visits than children with white mothers [38]. Similarly, a study of children in Texas found African American and Hispanic children had a stronger association between increases in ambient O_3 and initial asthma diagnosis than white children [39]. A national study found African American adults had a stronger association between asthma attacks and $PM_{2.5}$ concentrations than white adults [40]. Racial and ethnic minorities may be overall more susceptible to asthma exacerbations in response to asthma triggers as several studies have found African Americans and Hispanics of all ages have more poorly controlled asthma than whites and that this pattern persists even in high income groups [41, 42, 43].

Considering stratification by sex, associations between asthma ED visits and ambient air pollutant concentrations were generally stronger among males than females in the 5-18 year old age group, with significant differences observed for CO and NO_2. Pediatric males may spend more time outside (e.g., playing on sports teams) than females, which would lead to increased exposure. These results by sex are similar to those by Son et al., who found males had higher risk ratios between asthma hospitalizations and CO and NO_2 than females, however their results were not statistically significant [27]. A study in Toronto did not find any significant differences in the associations between asthma hospitalizations and $PM_{2.5}$ concentrations by sex among children 6 to 12 years old [44]. Overall, few studies have examined modification of the effects of ambient air pollution on asthma by sex, thus this may be a topic for further investigation.

There are several limitations to our study. First, our epidemiological results may be impacted by potential measurement error in our exposure data. We attempted to limit this measurement error by calculating population-weighted averages of air pollution concentration data available from multiple monitoring sites in each city on each study day. This ensures the daily pollution values used in our epidemiological models represent population exposures to ambient concentrations better than those from single fixed-site monitors. It is important to note, however, that the population-weighted average concentrations were weighted with overall population counts across each study area, not age-specific population counts, and thus our age-specific RRs may be impacted by differential measurement error. Second, asthma ED visits were identified only using primary ICD-9 code information. Not including visits with asthma listed in secondary ICD-9 codes may have reduced our sample size and thus the power of the study. However, since asthma listed in secondary codes indicates co-occurrence with other morbidities that in some cases may be the main reason for the ED visit, using secondary ICD-9 code information to identify asthma visits may result in a non-specific case definition and attenuate the apparent association between

pollutant concentrations and health care visits. Third, there was missing a large amount of race data in the demographic sub-analyses. Finally, sub-analyses were limited to race and sex; there may be other unmeasured factors that differed across age groups that might explain further observed age-related differences in air pollution-asthma associations.

In summary, our results suggest children with asthma are disproportionately affected by acute changes in ambient air pollution, compared with adults. This may be particularly explained by stronger risks faced by non-white and male populations in this age group.

3 Aim 2: Simultaneously calibrate climate model outputs with multiple variables

3.1 Background

The complexity of climate models combined with the many assumptions necessary to build a model lead to models that are inherently biased compared to observations. The aphorism attributed to George E.P. Box that, "All models are wrong, but some are useful" particularly pertains to climate models [45]. While we know no climate model is completely accurate, but they do provide an idea of how climate works now and how it will change in the future. There is some evidence that calibrating specific variables outputted from climate models to historically observed values can improve predictability in that variable [46].

Calibration methods that currently exist have several disadvantages. First, commonly used calibration methods, such as variance-scaling, include some temporal component in the method. Calibration on a daily or monthly level artificially limits the potential full distribution of the variable being calibrated to a monthly. For example, if we are interested in projecting daily average temperatures in 2050 our primary concern would be the value of the hottest temperature, not the month in which temperature occurred. By linking the calibration to temporal periods we may lose the upper and lower tails after calibration. Our proposed method uses quantile functions to calibrate climate model outputs over their entire distributions, ignoring the short-term temporal trends and avoiding the lose of information that may result from temporal calibration. The second disadvantage of currently existing calibration methods is the lack of simultaneous calibration. Current calibration methods calibrate each climate model variable one at a time, ignoring the physical complexity of the climate system. Climate variables are highly interconnected and univariate calibration methods lose information, particularly the dependence structure of the

variables. A calibration approach which considers multiple variables and retains the dependence structure may reduce uncertainty in the calibrated variables.

The proposed approach estimates quantile functions using I-splines to model the marginal distributions of meteorological variables from observed data and climate model outputs. The joint distribution is specified via a copula. Our goal is to calibrate the climate model projections to observed values while retaining the between-variable dependence structure. This method may lead to a more accurate projection of future meteorological patterns and provide a measure of uncertainty for the projections.

3.2 Quantile Functions

Quantile functions are useful tools for quantifying tail probabilities. Correctly estimating such tail probabilities is important in climate model outputs as extreme values (e.g., heat waves, higher ozone levels) are usually the most related to adverse health effects . Furthermore, the relationship between observed and modeled meteorological variables can be non-linear and their relationship cannot be summarized solely by the mean, as in linear regression [47]. Considering the relationships between quantile functions of the observed and modeled data over the study period provides more information than a simple linear regression would.

For a random variable, Y, the quantile function of Y is the inverse of the cumulative distribution function:

$$Q_Y(\tau) = F_Y^{-1}(\tau),\ \tau \in [0,1]$$

where τ represents the quantile value of Y: $P[Y < Q(\tau)] = \tau$. For example, if Y_j is the median Y value, $\tau_j = 0.5$. More recently, researchers have used techniques to estimate entire quantile functions with the purpose of identifying empirical distributions [48, 47].

Traditional linear regression models are structured as $Y = X\beta + \epsilon$ where X is a matrix of explanatory variables and Y is a vector containing the outcome of interest. In quantile regression, the Y vector is replaced with a vector containing the τ^{th} quantile of Y.

$$Q(\tau) = X^T\beta(\tau)$$

This regression models the relationship between X and the conditional quantiles of Y|X=x. Regression coefficients can be estimated with the formula: $\hat{\beta}(\tau) = argmin \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T\beta)$ or with Bayesian regression methods. Thus, $\beta(\tau)$ provides the covariate effect for each τ^{th} quantile.

Here we are not using true quantile regression as there are no covariates being regressed against (rather, we use an I-spline based on a simple sequence from 0 to 1).

3.3 I-Splines

Splines are basis functions for piecewise regressions which connect at interior knots. They consist of polynomial functions that are designed to provide a sufficient degree of smoothness at the knots. Splines enable the fitting of smooth regression lines for non-linear relationships. Splines can also be used to constrain regression models (e.g., monotonic increasing).

As discussed in Zhou, et al., when simultaneously modeling multiple quantiles the quantile curves can cross, which leads to an invalid distribution for the outcome of interest [49, 50, 48]. One way to prevent this issue is by constraining the quantile function with piecewise polynomial functions to ensure monotonicity. Integrated piecewise polynomial splines (I-splines) can construct such a monotone function. I-splines are calculated by performing piecewise integration of M-splines. M-splines, basis functions with order k, $M_i(\bullet|k,t), i = 1,...,n$, are designed so that any piecewise polynomial associated with the knot sequence, t, can be represented as a linear combination $f = \sum a_i M_i$, where $a_i \geq 0$ and $\sum a_i = 1$. The M-spline, $M_i(x|k,t)$, is only defined in the interval $[t_i, t_{i+k}]$ and is zero elsewhere. Additionally, $\int M_i(x)dx$=1 and each M_i has the properties of a probability density function over the interval $[t_i, t_{i+k}]$. First order M-splines are uniform densities over $[t_i, t_{i+1}]$ while second order M-splines are triangle densities over $[t_i, t_{i+2}]$ with the mode at t_{i+1} [51]. M-splines are calculated as follows:

$M_i(x|1,t) = \frac{1}{t_{i+1}-t_i}$, where $t_i \leq x < t_{i+1}$ and 0 otherwise

$M_i(x|k,t) = \frac{k(x-t_i)M_i(x|k-1,t)+(t_{i+k}-x)M_{i+1}(x|k-1,t)}{(k-1)(t_{i+k}-t_i)}$, when k>1

I-splines are calculated by integrating each piecewise section of the M-spline from the minimum of the appropriate knot-interval (L) to x. Hence, I-splines are monotonic increasing.

$$I_i(x|k,t) = \int_L^x M_i(u|k,t)\ du$$

3.3.1 I-Spline Examples

For example, if we are creating I-splines of order k=3 with 2 interior knots at 0.3 and 0.6 we would start by creating the M-splines. The third order M-splines are built upon the second order M-splines, which are based on the first order M-splines. Thus, we begin with the creation of the first order M-splines. First, we identify the knot sequence t, whose length is determined by the final order of the splines we are creating (here, k=3) and the number of interior knots. The exterior knots are repeated k times with the interior knots only occurring once. In this example, we create 3rd order I-splines with 2 interior knots, therefore our knot sequence will be t = {0,0,0,.3,.6,1,1,1}.

Within each order there are multiple M-splines, indexed by i. The number of M-splines in each order is equal to the length of the knot sequence minus k. Thus, the first order M-splines consist of 7 splines. Using the above equation for first order M-splines, we calculate them as follows (these set $\frac{1}{0}$ to 0):

$M_1(x|1,t) = \frac{1}{0-0} = 0$

$M_2(x|1,t) = \frac{1}{0-0} = 0$

$M_3(x|1,t) = \frac{1}{0.3-0} = 0.3$

$M_4(x|1,t) = \frac{1}{0.6-0.3} = 0.3$

$M_5(x|1,t) = \frac{1}{1-0.6} = 0.4$

$M_6(x|1,t) = \frac{1}{1-1} = 0$

$M_7(x|1,t) = \frac{1}{1-1} = 0$

The second order M-splines use these results and are constructed as follows:

$$M_1(x|2,t) = \frac{2(x-0)M_1(x|1,t)+(0-x)M_2(x|1,t)}{1(0-0)} = 0$$

$$M_2(x|2,t) = \frac{2(x-0)M_2(x|1,t)+(0.3-x)M_3(x|1,t)}{1(0.3-0)} = \frac{2(0.3-x)}{.3^2}$$

$$M_3(x|2,t) = \frac{2(x-0)M_3(x|1,t)+(0.6-x)M_4(x|1,t)}{1(0.6-0)} = \frac{2}{.6}\left[\frac{x}{0.3} + \frac{0.6-x}{0.3}\right]$$

$$M_4(x|2,t) = \frac{2(x-0.3)M_4(x|1,t)+(1-x)M_5(x|1,t)}{1(1-0.3)} = \frac{2}{.7}\left[\frac{x-0.3}{0.3} + \frac{1-x}{0.4}\right]$$

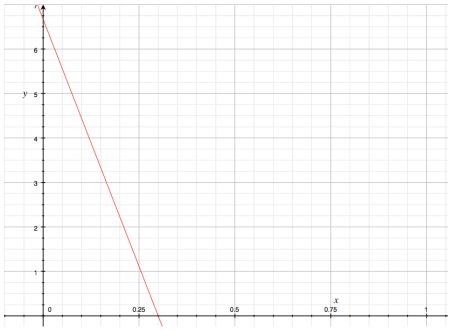
$$M_5(x|2,t) = \frac{2(x-0.6)M_5(x|1,t)+(1-x)M_6(x|1,t)}{1(1-0.6)} = \frac{2}{.4}\left[\frac{x-0.6}{0.4}\right]$$

$$M_6(x|2,t) = \frac{2(x-1)M_6(x|1,t)+(1-x)M_7(x|1,t)}{1(1-1)} = 0$$

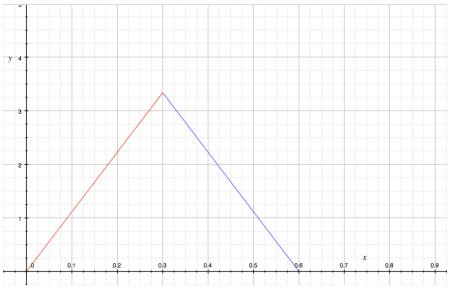
While the first order splines are uniform linear functions, the second order splines are piecewise triangles. Therefore, $M_3(x|2,t)$ and $M_4(x|2,t)$ are separated into 2 piecewise sections with the breakpoint at the interior knots (0.3 and 0.6, respectively). The area under each of the basis functions integrates to 1. The figure below graphically depicts the second order M-splines (Figure 11).

Figure 11: Second-order M-splines example. Different colors represents different sections of the piecewise functions.

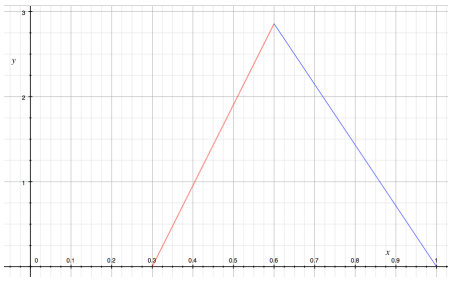
(a) $M_2(x|2,t)$



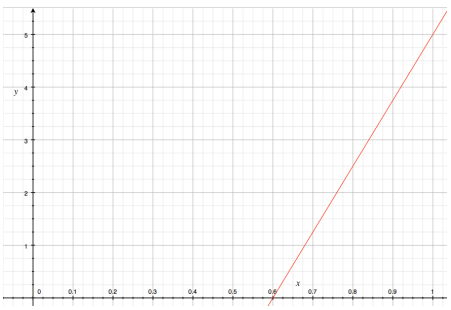
(b) $M_3(x|2,t)$



(c) $M_4(x|2,t)$



(d) $M_5(x|2,t)$



The third order M-splines are piecewise polynomial functions and are constructed from the second order M-splines. As with the second order M-splines, each piecewise polynomial function exists only within the boundaries set by the knot sequence. The area under each of lines integrates to 1. The figure below graphically depicts the third order M-splines with knots at (0, 0.3, 0.6, 1).

$$M_1(x|3,t) = \frac{3(x-0)M_1(x|2,t)+(0.3-x)M_2(x|2,t)}{2(0.3-0)} = \frac{3(0.3-x)^2}{0.3^2}$$

$$M_2(x|3,t) = \frac{3(x-0)M_2(x|2,t)+(0.6-x)M_3(x|3,t)}{2(0.6-0)} = \frac{100x}{3} - \frac{250x^2}{3} \text{ for (0, 0.3)}$$

and $= \frac{3(0.6-x)^2}{0.3(0.6^2)}$ for (0.3, 0.6)

$$M_3(x|3,t) = \frac{3(x-0)M_3(x|2,t)+(1-x)M_3(x|4,t)}{2(1-0)} = \frac{50x^2}{3} \text{ for (0, 0.3)}$$

$= \frac{-650x^2}{21} + \frac{200x}{7} - \frac{30}{7}$ for (0.3, 0.6)

and $= \frac{75}{7}(1-x)^2$ for (0.6, 1)

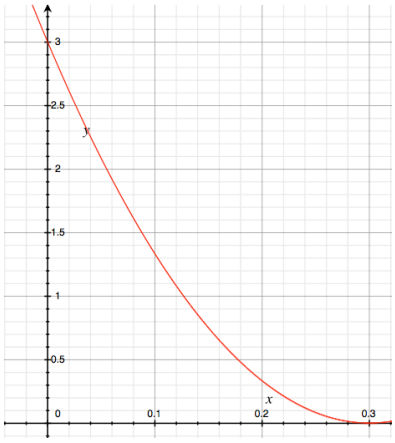
$$M_4(x|3,t) = \frac{3(x-0.3)M_4(x|2,t)+(1-x)M_3(x|5,t)}{2(1-0.3)} = \frac{1000}{49}(x-0.3)^2 \text{ for (0.3, 0.6)}$$

and $= \frac{3}{0.28^2}(-1.1x^2 + 1.64x - 0.54)$ for (0.6, 1)

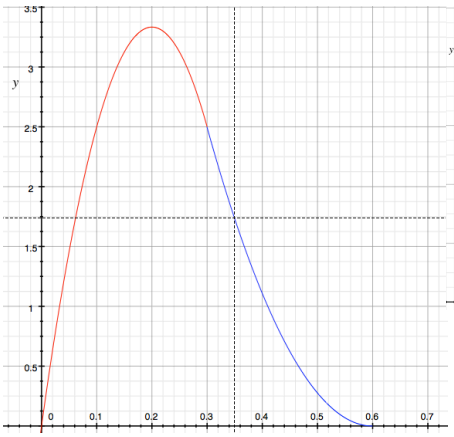
$$M_5(x|3,t) = \frac{3(x-0.6)M_5(x|2,t)+(1-x)M_3(x|6,t)}{2(1-0.6)} = \frac{3(x-0.6)^2}{0.4^3}$$

Figure 12: Third-order M-splines example. Different colors represents different sections of the piecewise functions.

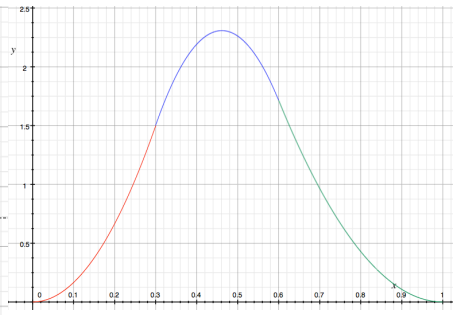
(a) $M_1(x|3,t)$



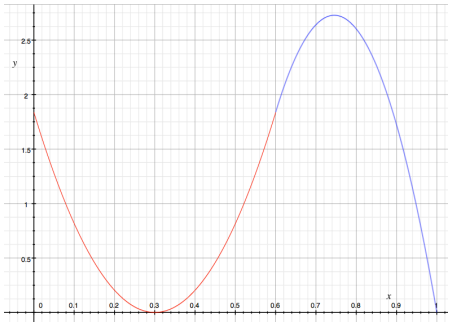
(b) $M_2(x|3,t)$



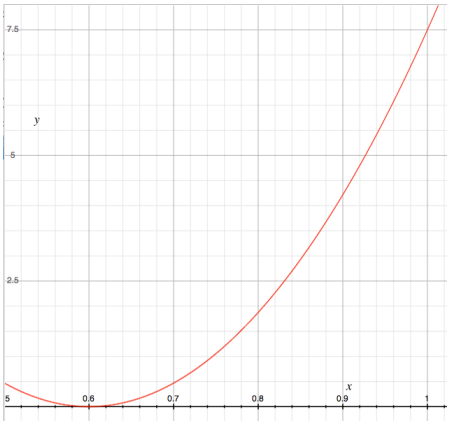
(c) $M_3(x|3,t)$



(d) $M_4(x|3,t)$



(e) $M_5(x|3,t)$



Once the third order M-splines have been calculated, the I-splines are created by integrating each piecewise section of the M-splines over their respective intervals. For example, the I-spline corresponding to $M_2(x|3,t)$ would be created in two sections - one ranging across (0, 0.3) and the other from (0.3, 0.6) as shown below (Figure 13).

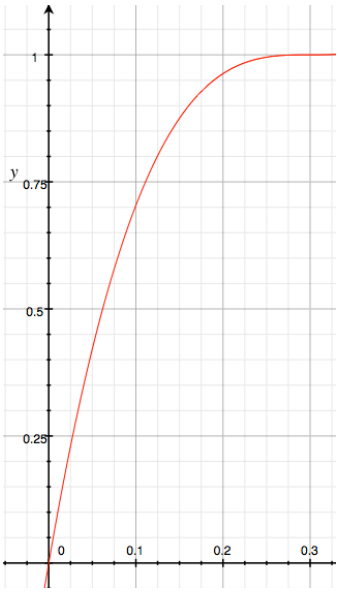
$I_2(x|3,t)$:

for (0, 0.3), $\int_0^x \frac{100}{3}u - \frac{250}{3}u^2 du = \frac{50}{3}x^2 - \frac{250}{9}x^3$

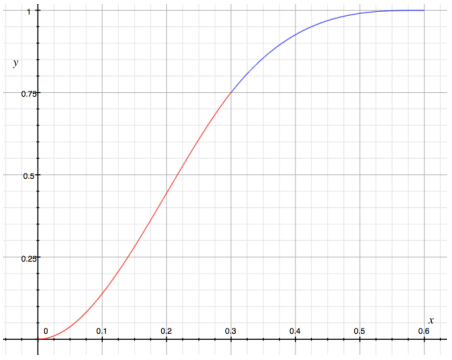
for (0.3, 0.6), $\int_{0.3}^x \frac{250}{9}(0.6-u)^2 du = \frac{250}{27}[(x-0.6)^3 + 0.027]$

Figure 13: Third-order I-splines example. Different colors represents different sections of the piecewise functions.

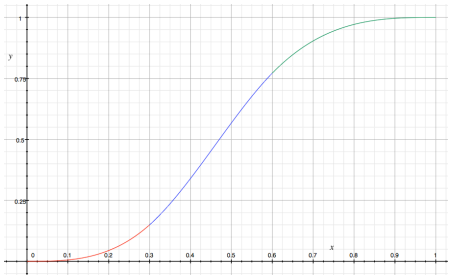
(a) $I_1(x|3,t)$



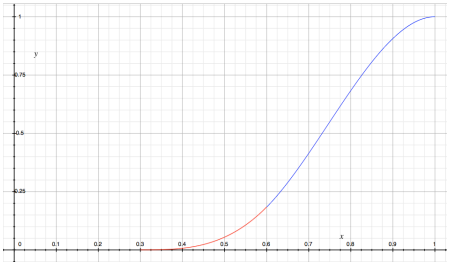
(b) $I_2(x|3,t)$



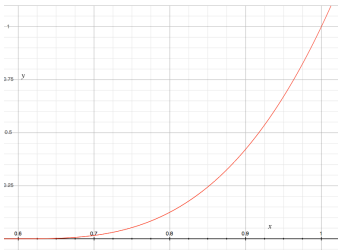
(c) $I_3(x|3,t)$



(d) $I_4(x|3,t)$



(e) $I_5(x|3,t)$



For the univariate calibration, the initial I-spline was created from a sequence, τ_i, increasing every .005 from [0,1] with 10 interior knots set at t=(0.05, 0.10, 0.15, 0.20, 0.40, 0.60, 0.80, 0.85, 0.90, 0.95).

3.4 Bayesian Inference

Bayesian inference methods are used to estimate the regression coefficients, here $Q(\tau;\beta) = I(\tau)\beta$ where $I(\tau)$ is the basis function design matrix. The Metropolis-Hastings algorithm (a Markov Chain Monte Carlo (MCMC) method) is used here as the posterior distributions of the unknown parameters are unknown and not in closed form. The general form of the Metropolis-Hastings algorithm is based on comparing the current parameter values to proposed values and then choosing whether to accept or reject the proposed parameter value. The algorithm begins by specifying a vector of initial estimates, β_c. For this analysis, the initial beta estimates were selected by minimizing the least-squares regression model between the $Q(\tau_i)$ values and the I-spline design matrix, under the constraint that $\sum \beta_i = 1$. This constraint forces the quantile function to have a minimum of 0 and a maximum of 1, which is necessary to ensure the estimated quantile function goes through (0,0) and (1,1).

Once the initial parameters are estimated, the algorithm proposes a new set of parameters, β_p, based on a specified distribution. Here we use a Log-normal distribution ($P(\beta_p|\beta_c) = \frac{1}{\beta_p\sigma\sqrt{2\pi}}e^{-\frac{(ln\beta_p-\mu)^2}{2\sigma^2}}$) to randomly generate proposed β estimates, with the parameters specified as $\mu = \beta_c$ and $\sigma^2 = \hat{V} * T$ where $\hat{V}$ is the estimated variance of the initial $\hat{\boldsymbol{\beta}}$ from the least-squares regression and T is a tuning parameter to adjust the acceptance rate.

The proposed parameters (β_p) are evaluated against the current value (β_c) by comparing the likelihood for the proposed parameter value to that of the initial estimates based on the likelihood ratio:

$$R = \frac{L(y|\beta_p)L(\beta_p)P(\beta_c|\beta_p)}{L(y|\beta_c)L(\beta_c)P(\beta_p|\beta_c)}$$

$L(y|\beta)$ is the data likelihood based on the given β, $L(\beta)$ is the prior and is often assumed to be the same for both β_c and β_p, and $P(\beta_p|\beta_c)$ (and its reciprocal

distribution) is the proposal distribution. For this application, the data likelihood is calculated empirically via the inverse derivative of the quantile function.

$$L(y|\beta_p) = \sum_{i=1}^{n} log(f_Y(y_i))$$

$$f_Y(y_i) = \frac{1}{dQ|_{\tau_i=\{\tau:Q(\tau_i;\beta)=y_i\}}} = M(\tau_i) * \beta$$

where $M(\tau_i)$ is the M-spline design matrix corresponding to the I-spline basis function.

The proposed and initial likelihoods are compared with a likelihood ratio test. If β_p produces a likelihood ratio that meets the acceptance probability (decided as min{1,R}), then β_p becomes the new β_c. If the proposed likelihood fails to meet the acceptance probability, β_p is discarded and β_c remains. Then, the algorithm begins again by proposing another set of beta estimates. This process is carried out for a predetermined number of iterations (here, n=10,000).

The initial proposed samples can be outside of the posterior distribution, therefore, the first 2,000 estimates were discarded (the burn-in). The resulting iterations were also thinned by every 100 samples to minimize autocorrelation in the samples, resulting in m=800 simulations. Convergence of the MCMC regression model was assessed using trace plots. These plots track the accepted parameter values over each iteration and show a general trend of convergence. A deviance plot was also created and showed a general decreasing trend (indicating better model fit as the model converged).

3.5 Univariate Calibration Method

In order to calibrate the climate projection model to observed data, we first convert the raw data (both the observed and the projection model, separately) to a sorted, standardized range from [0,1]. We then use the relationship between the data's distributions and the quantile functions to estimate τ values representative of the data's quantile value.

For i=1,...n, n=total number of time points, i= day in historical period, Y_i^{obs} = daily measurements at monitor, and Y_i^{RCM} = daily 50 km x 50 km grid RCM output linked to the monitor.

$$Y_i^{obs} \sim f_{obs}(y_i^{obs}|\beta_{obs})$$

$$Y_i^{RCM} \sim f_{RCM}(y_i^{RCM}|\beta_{RCM})$$

$$f_{obs}(y_i^{obs}|\beta_{obs}) = \frac{1}{\frac{d}{d\tau}Q^{obs}(\tau_i^{obs}|\beta_{obs})|_{\tau_i^{obs}=\{\tau: Q^{obs}(\tau_i;\beta_{obs})=y_i^{obs}\}}}, \ \tau \in [0,1]$$

$$f_{RCM}(y_i^{RCM}|\beta_{RCM})=\frac{1}{\frac{d}{d\tau}Q^{RCM}(\tau_i^{RCM}|\beta_{RCM})|_{\tau_i^{RCM}=\{\tau: Q^{RCM}(\tau_i;\beta_{RCM})=y_i^{RCM}\}}}, \tau\in[0,1]$$

The derivative of the quantile function can be simply calculated using the M-splines, which is one attractive feature of using I-splines. The distribution of Y is needed to calculate the data likelihood in the MCMC estimation of the beta coefficients, as explained in Section 3.4. The quantile function is given below:

Let x_j be an interior sequence of quantile values, $j = 1, ..., 1000$.

$$Q^{obs}(x_j;\beta_{obs}) = I(x_j)\beta_{obs}$$

$$Q^{RCM}(x_j;\beta_{RCM}) = I(x_j)\beta_{RCM}$$

where $I(x_j)$= J x (h-4) design matrix of third-order I-splines for a knot sequence of $\{t_1, t_2, ..., t_h\}$ and the β estimates are constrained with $\sum \beta_{i,obs} = 1, \sum \beta_{i,RCM} = 1, \beta_{obs} > 0$, and $\beta_{RCM} > 0$. For identifiability purposes, we set $\beta_1 = 1$. An MCMC algorithm is used to estimate the $\hat{\beta}_{obs}$ and $\hat{\beta}_{RCM}$ coefficients in the quantile functions, which are then used to estimate values for $\hat{\tau}_i^{obs}$ and $\hat{\tau}_i^{RCM}$.

Once the $\hat{\beta}$ estimates have been estimated, the calibration step can begin. Let Z_k^{RCM} be future RCM outputs on day k, where k = day in future period.

$$\hat{\tau}_k^{RCM} = \{\tau : Q^{RCM}(\tau_k; \beta_{RCM}) = z_k^{RCM}\}$$

Using the future $\hat{\tau}_k^{RCM}$ we use the relationship between τ_i^{obs} and τ_i^{RCM} to predict the future $\hat{\tau_k^{obs}}$.

$$\hat{\tau}_k^{obs} = \hat{g}(\hat{\tau}_k^{RCM})$$

where

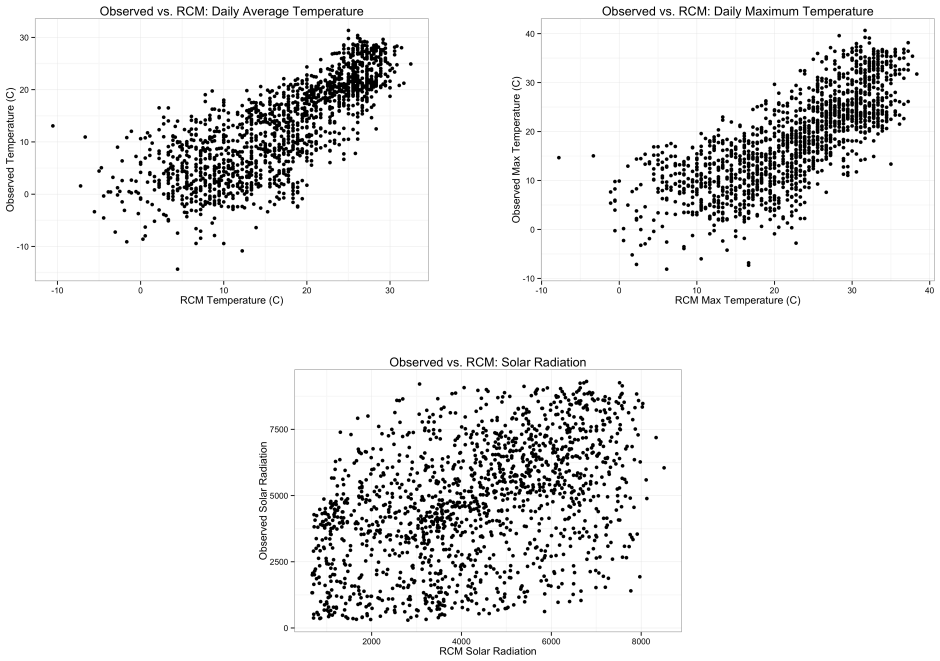
$$\tau_i^{obs} = \hat{g}(\tau_i^{RCM}) = I(\tau_i^{RCM})\gamma^{cal}$$

where γ^{cal} is subject to the $\sum \gamma = 1$ constraint. The calibrated $\hat{\tau_k^{obs}}$ values are converted to standardized climate variables through the quantile function, $Q^{RCM}(\tau_k; \beta_{RCM}) = z_k^{RCM}$. Finally, the standardized variable is transformed to the scale of the original climate variables and re-sorted into daily values.

3.6 Application of Univariate Method

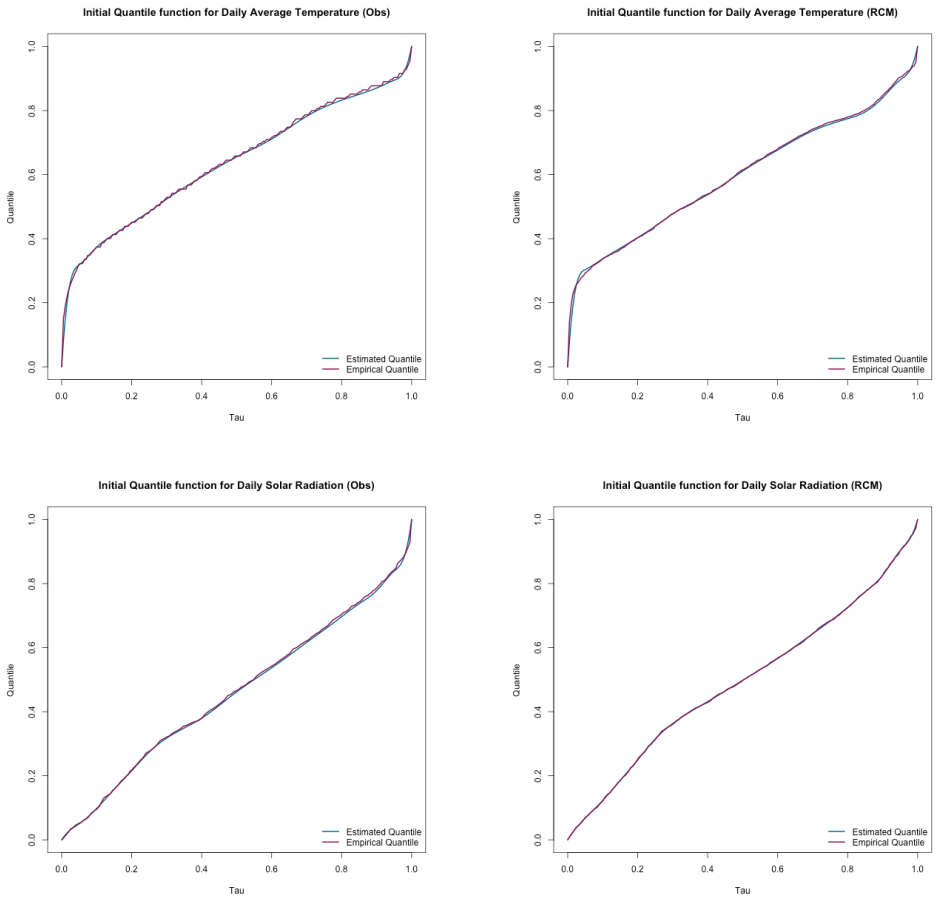
We applied the method described in Section 3.5 to data ranging from January 1, 1996 to December 30, 2000. Observed meteorologic variables were obtained from daily measurements taken at the Hartsfield-Jackson Atlanta International Airport. The climate model data was produced by the North American Regional Climate Change Assessment Program (NARCCAP) using the Canadian Regional Climate Model as the RCM and the Community Climate System Model to provide the boundary conditions [52]. The daily values for the 50 km x 50 km grid cell over the city of Atlanta were used as the climate model (RCM) data. Average daily temperature, maximum daily temperature, and daily solar radiation were assessed.

Figure 14: Scatter plots of regional climate model outputs versus observed values for daily average temperature, maximum daily temperature, and daily solar radiation in Atlanta (1996-2000)



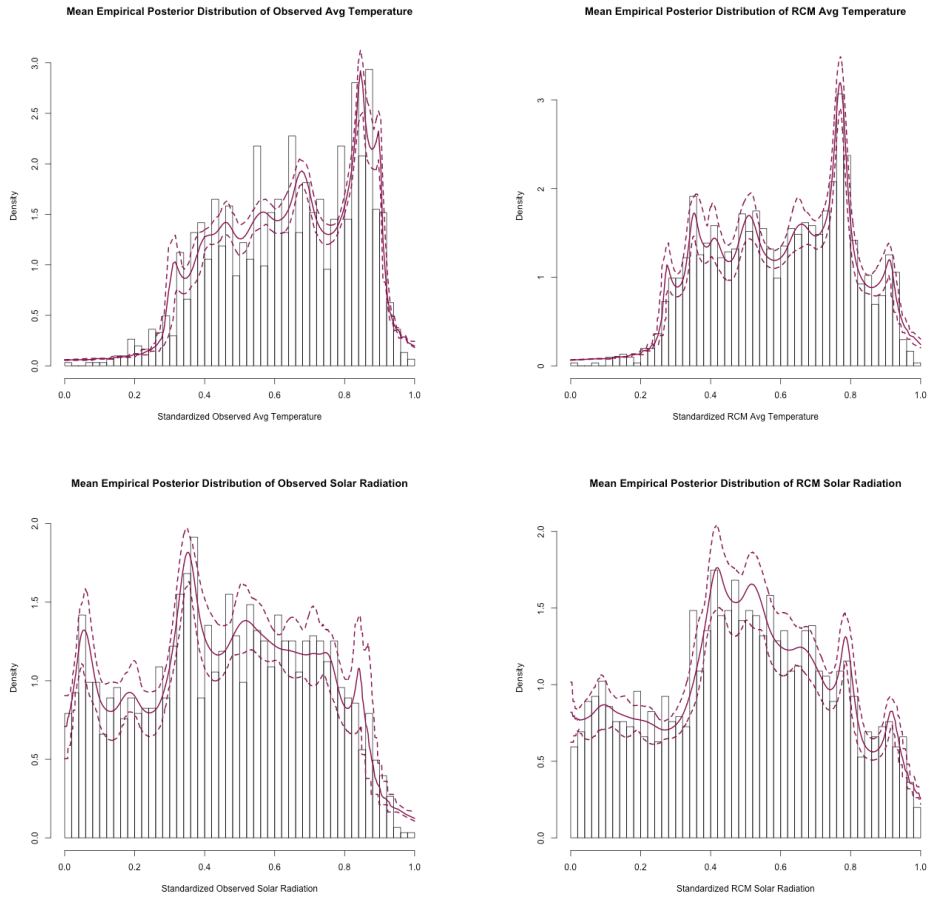
As seen in Figure 14, there was a fair amount of variability in the prediction capability of the climate model outputs. For each variable, we first identified the empirical quantile functions of both the observed and RCM datasets (Figure 15).

Figure 15: Estimated and empirical quantile functions for daily average temperature and daily solar radiation in Atlanta (1996-2000)



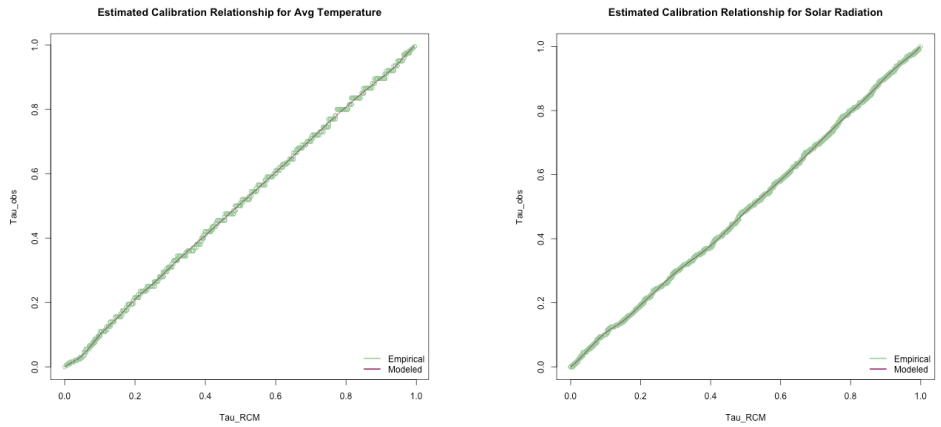
The quantile functions for average temperature and solar radiation were estimated in the MCMC using I-splines. The corresponding densities are given in Figure 16.

Figure 16: Histograms of observed and projected daily mean temperature and solar radiation with accompanying empirical data density functions in Atlanta (1996-2000)



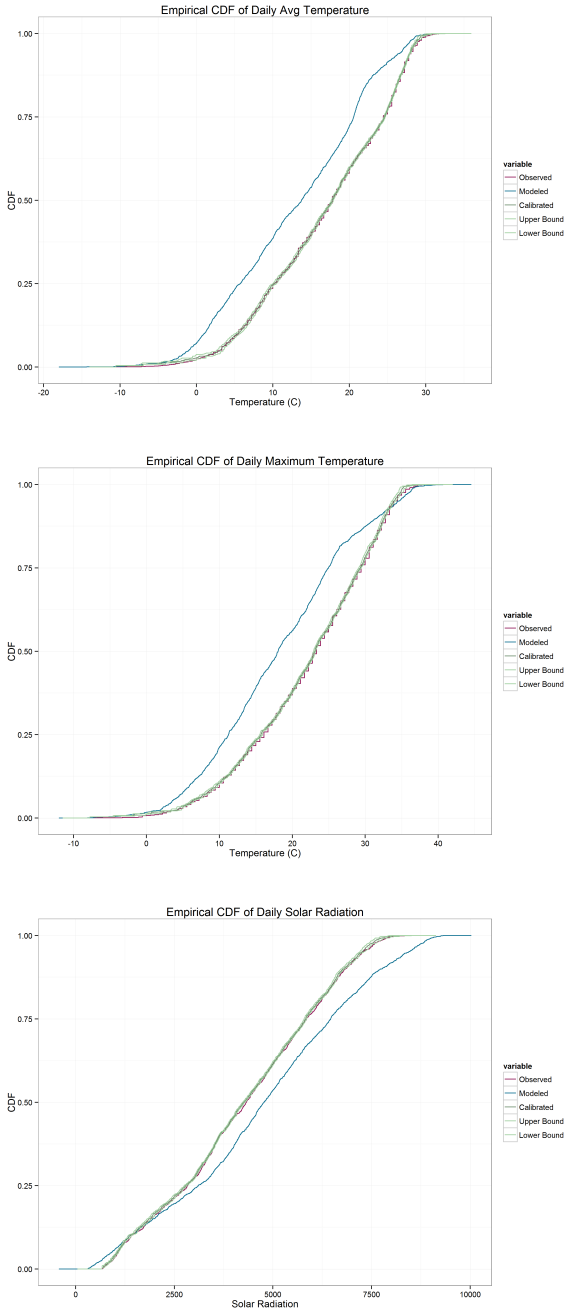
The resulting posterior quantile estimates were then regressed against each other to obtain the $\hat{\beta}$'s associated with the calibrated τ's. Figure 17 below shows how well the predicted line fit the observed relationship between $\hat{\tau}_{obs}$ and $I(\hat{\tau}_{RCM})$.

Figure 17: Predicted relationship between $\hat{\tau}_{obs}$ and $I(\hat{\tau}_{RCM})$ for average temperature and solar radiation in Atlanta (1996-2000)



The resulting calibrated quantile values, τ_c were then converted back into their respective climate variables. The empirical CDF plots below show how the calibrated values were more accurate predictors of climate variable quantile levels than the RCM values (Figure 18).

Figure 18: Empirical CDFs of observed, uncalibrated projections, and calibrated projections for daily average temperature, maximum temperature, and solar radiation in Atlanta (1996-2000). The calibrated projections include 95% confidence bands around the empirical CDF



The average root mean squared error (RMSE) of the calibrated model outputs was consistently less than the RMSE of the RCM model (Table 4). The RMSE of the RCM model was calculated directly, $RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{RCM})^2}{n}}$. The RMSE

of the calibrated model was the mean of the RMSE of each MCMC iteration, $\frac{\sum_{j=1}^{m} r_j}{m}$, where $r_j = \sqrt{\frac{\sum_{i=1}^{n}(y_i^{obs} - y_{ij}^{cal})^2}{n}}$.

Table 4: Average Root Mean Squared Error (RMSE) of RCM and Calibrated Models

Variable	RCM	Calibrated
Average Temperature	4.05	0.50
Maximum Temperature	4.29	0.58
Solar Radiation	563.57	53.32

The univariate calibrated model had the same correlations between the variables as the RCM model did (Table 5). The dependence structure seen in the observed data cannot be preserved through the univariate calibration; therefore, a multivariate calibration method is needed and developed later (see Section 3.7.2).

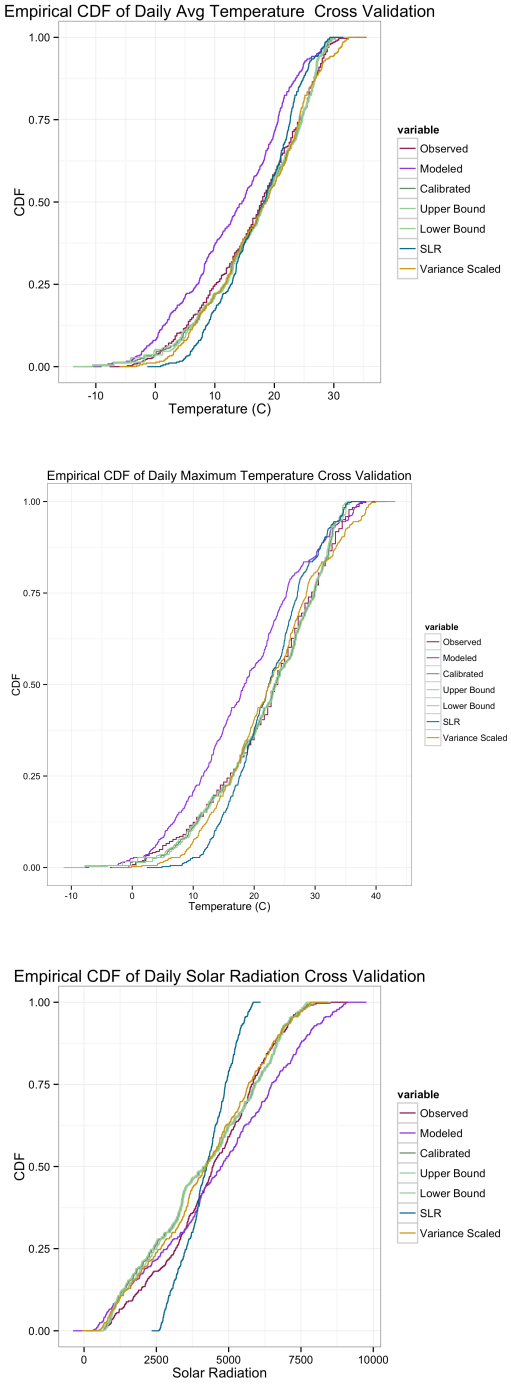
Table 5: Correlations between Climate Variables

Variable	Observed	RCM	Calibrated
Average Temperature & Maximum Temperature	0.981	0.968	0.968
Average Temperature & Solar Radiation	0.607	0.478	0.478
Maximum Temperature & Solar Radiation	0.685	0.593	0.593

We performed a cross validation analysis of the univariate method to assess its performance using four years of RCM and observed data. The first three years were used as a training dataset and the last year was used as a testing dataset. The training dataset was used to estimate the $\boldsymbol{\beta}$ coefficients and those $\hat{\beta}$'s were used to calculated the calibrated $\hat{\tau}_k^{cal}$ values, which were transformed to climate variables using the observed values of the training dataset. We compared our method to a simple linear regression (SLR) calibration between the RCM and observed data and to variance-scaling (VS), a standard calibration method often used in climate research [46]. The variance-scaling method involved adjusting the predicted values from the RCM so they will have the same mean and variance and the observed data from the training period.

Our proposed quantile calibration method was better at capturing the tail distributions of the climate variables than the SLR or VS methods. Figure 19 shows the empirical CDFs resulting from the testing dataset.

Figure 19: Cross-validation: empirical CDF plots of observed, uncalibrated projections, calibrated projections (with 95% confidence bands), simple linear regression projections, and variance-scaling projections for daily average temperature, maximum temperature, and solar radiation in Atlanta (1996-2000)



Overall, the quantile method performed well in the cross validation. The calibrated predicted quantile levels fit the observed quantile levels better than the RCM model. Temperature variables performed better than solar radiation.

This may be because the climate model is designed to prioritize the accuracy of temperature projections over the accuracy of other variables' projections. The RMSE of the calibrated variables was less than the RMSE of the uncalibrated RCM variables, but not less than the SLR or VS methods (Table 6). However, while these methods performed well in the center of the data distribution, neither of them capture the tail distributions well. The correlations between variables were similar to the overall analysis (Table 7).

Table 6: Cross Validation: Average Root Mean Squared Error (RMSE) of Different Methods

Variable	RCM	Calibrated	SLR	Variance-scaled
Average Temperature	3.46	0.89	2.27	1.09
Maximum Temperature	4.04	0.87	2.89	1.65
Solar Radiation	580.53	410.35	990.52	307.87

Table 7: Cross Validation: Correlations between Climate Variables

Variable	Observed	RCM	Calibrated	SLR	Variance-scaled
Avg Temp & Max Temp	0.983	0.967	0.967	0.967	0.967
Avg Temp & Solar Radiation	0.636	0.456	0.456	0.456	0.456
Max Temp & Solar Radiation	0.716	0.584	0.584	0.584	0.584

3.7 Bivariate Calibration Method

3.7.1 Background

The methods described to this point are designed for a univariate calibration (e.g., calibrating a RCM-generated maximum daily temperature to observed maximum daily temperature). These univariate techniques, as well as other current methods and applications of climate model calibration, only consider single variables one at a time, ignoring the physical complexity of the climate system where climate variables interact and influence each other. For example, days with higher temperatures are associated with higher solar radiation, while temperature and precipitation tend to have an inverse relationship. Therefore, an approach that considers multiple variables jointly may result in reduced uncertainty in the calibrated outputs.

Additionally, a multivariate approach can ensure that the dependence structure between those variables is retained after calibration. Retaining the dependence structure may result in more accurate projections of adverse impacts, which are often driven simultaneously by multiple conditions. Multiple climate model variables are routinely used simultaneously as inputs for projecting future infection disease transmission and air quality. For example, warmer temperatures and higher precipitation levels can synergistically lead to larger mosquito populations. Similarly, ozone and particulate matter (air pollution) are affected by multiple meteorological conditions.

The main objective of this section is to describe a statistical approach to perform bivariate joint calibration on the distribution-level. Specifically, we extend the quantile calibration of Zhou et al. [48] by incorporating the observed dependence between variables. This is accomplished via a copula to link the marginal distributions of two climate variables to allow joint calibration. We apply our calibration approach to examine projections of future temperature and solar radiation in the city of Atlanta, which is currently in non-attainment for ozone pollution under the National Ambient Air Quality Standards. Ozone is a

secondary pollutant and its formation is highly affected by temperature and solar radiation. We consider projections from four regional climate models (RCMs). RCMs are used to inform the impact of climate change on regions and nations, as accurate local information is necessary to justify social and economic policies related to climate change adaptation and mitigation [10]. Our goal is to obtain improved climate projections that can be used to evaluate health impacts in subsequent analyses.

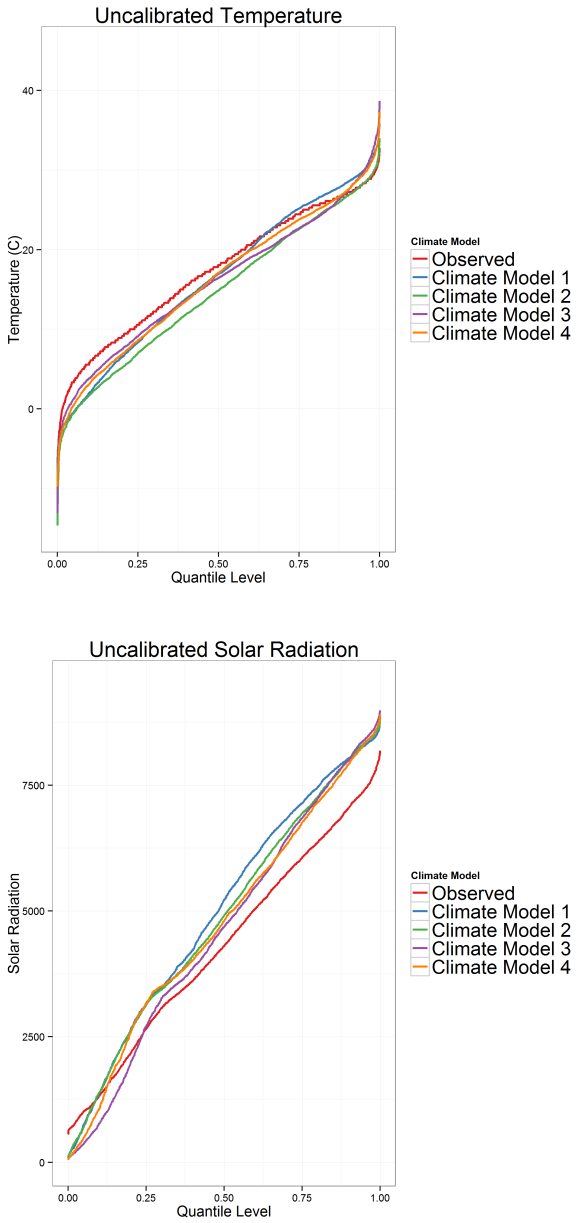
3.7.2 Motivating Example

We perform the bivariate calibration method using average daily temperature and total daily solar radiation as our two variables. Four climate models are calibrated for comparison purposes using observed monitoring data. Daily meteorological conditions for 1991-2000, including average temperature and total solar radiation (global horizontal irradiance), were obtained from the National Climatic Data Center and the National Solar Radiation Data Base for monitors located at the Hartsfield-Jackson Atlanta International Airport.

Climate model outputs were obtained from the North American Regional Climate Change Assessment Program (NARCCAP) and included daily average surface temperature and daily solar radiation values from the 50 km by 50 km grid cell including the Atlanta monitors [53]. These outputs were obtained for the historical period of 1991-2000 and for the future period of 2041-2070. NARCCAP, which is a public database of RCM simulations, provided simulations which were conducted under the IPCC Special Report on Emissions Scenarios (SRES) A2 emissions scenario [54]. The A2 scenario represents the higher end of IPCC emission scenarios and entails large population increases, high carbon dioxide emissions, and weak environmental concerns. To assess variabilities in RCM projections under the same emission scenario, we examined NARCCAP simulations from four combinations of two different RCMs driven by boundary conditions from four different global climate models (GCMs). The RCMs include

the Canadian Regional Climate Model (CRCM) and the Handley Regional Model 3 (HRM3). The GCMs include the Community Climate model version 3 (CCSM3), the Canadian Global Climate Model version 3 (CGCM3), the Geophysical Fluid Dynamics Laboratory (GFDL) Climate Model version 2.1, and the United Kingdom Hadley Centre Climate Model version 3 (HadCM3). Detailed descriptions on the RCM and GCM characteristics are summarized by NARCCAP online
(http://www.narccap.ucar.edu/). The following 4 RCM-GCM combinations were conducted by NARCCAP and examined in this study: CRCM-CCSM, CRCM-CGCM3, HRM3-GFDL, and HRM3-HadCM3. As Figure 20 shows, the uncalibrated climate models project higher temperature and solar radiation values than were observed.

Figure 20: Observed values and historic climate simulations for average temperature and solar radiation in Atlanta (1991-2000)



3.7.3 Copulas

Copulas are one method of determining a joint distribution provided the marginal distributions are known. Copulas are multivariate distribution functions where the marginal distribution for each variable is uniform. The copula completely characterizes the dependence relationship between the variables. In order to use

copulas, random variables are transformed to uniformly distributed marginals via the cumulative distribution function:

$$(U_1, U_2) = (F_1(Y_1), F_2(Y_2))$$

The copula, $C_{\mathbf{Y}}$, is defined as the joint cumulative distribution of (U_1, U_2) such that:

$$C(u_1, u_2) = P(U_1 \leq u_1, U_2 \leq u_2)$$

This relationship allows the random variables, (Y_1, Y_2) to be rewritten as $(Y_1, Y_2) = (F_1^{-1}(U_1), F_2^{-1}(U_2))$ leading to:

$$C(u_1, u_2) = P(Y_1 \leq F_1^{-1}(u_1), Y_2 \leq F_2^{-1}(u_2))$$

By Skalr's theorem, the joint probability density can then be written as the product of the derivative of the copula and the marginal densities [55].

$$f_{\mathbf{Y}}(\mathbf{x}) = f_{y1}(y_1) \cdot f_{y2}(y_2) \cdot c_{\mathbf{Y}}(u_1, u_2) \tag{1}$$

There are many different types of copula families, grouped by parametric form. The two most common copula families are elliptical and Archimedean copulas. A systematic method for selecting a copula has not been determined although the science of copula selection is a popular area of research. For this analysis, we will investigate the fit of the Gumbel and the Frank copulas, members of the Archimedean family.

The Gumbel copula form is:

$$C_\alpha(u_1, u_2) = exp\{-[(-log\, u_1)^\alpha + (-log\, u_2)^\alpha]^{1/\alpha}\},\ \alpha \in [1, \infty)$$

where $\hat{\alpha} = \frac{1}{1-k}$ and k is the Kendall's τ coefficient of the observed data.

The bivariate Frank copula form is:

$$C(u,v;\delta) = -\delta^{-1} log\left(\frac{1-e^{-\delta}-(1-e^{-\delta u})(1-e^{-\delta v})}{1-e^{-\delta}}\right), -\infty < \delta < \infty$$

where $\hat{\delta}$ is estimated by setting Kendall's $\tau = 1 + 4\delta^{-1}[\mathrm{D}_1(\delta)$ - $1]$, where $\mathrm{D}_1(\mathrm{x})=x^{-1}\int_0^x t^k(e^t - 1)^{-1}dt$, and solving for δ. The estimation of δ was accomplished with the CDVine R package [56].

The choice of the Gumbel copula, which entails greater dependence in the right tail, is based on previous work in climate change research [55]. The bivariate Frank copula was chosen as a comparison to the Gumbel copula because it models a similar, but slightly different dependence structure between the two variables. The copula determines the dependence between the two variables and contributes information on that dependence to the joint distribution of the variables. Because parametric form of copula determines the parametric form of the joint distribution, the selection of copula affects how well the joint distribution describes the true dependence between the variables. Copula selection is a current area of interest in statistics. Here, parametric forms were selected based on the copula properties, specifically maintaining tail dependence (compared to the Gaussian copula, for example, which does not retain tail dependence).

3.7.4 General Modeling Framework

The goal of our calibration method is to first estimate the discrepancy in distributions between the climate model simulation and the observed monitoring data. We characterize the entire distribution using quantile functions. For a continuous random variable, Y, the quantile function of Y is the inverse of the strictly monotone cumulative distribution function:

$$Q_Y(\tau) = F_Y^{-1}(\tau),\ \tau \in [0,1]$$

where τ represents the quantile level of Y: $P[Y < Q_Y(\tau)] = \tau$. For example, $Q_Y(0.5)$ is the median of Y. The quantile function uniquely defines the distribution with density given by:

$$f_Y(y) = \frac{1}{\frac{dQ_Y(\tau^*)}{d\tau}\big|_{\tau^*=\{\tau:Q_Y(\tau)=y\}}} \tag{2}$$

In the univariate quantile calibration described above, a non-decreasing function is used to calibrate the quantile function of the numerical model output towards the quantile function of the observations. However, univariate calibration methods do not account for the fact that meteorological variables do not exist independently from each other and affect each other's values. Therefore, calibration ideally should accommodate multiple variables in a unified statistical model. We use copulas to define the target joint distribution and retain the observed dependence structure of the climate variables in the calibrated outputs. The above copula decomposition of the bivariate likelihood allows us to perform joint quantile-based calibration as follows:

- Step 1: for the historical period, we assume the quantile function associated with each marginal density $f_k(y_k)$, for variable k=$\{1,2\}$ is given by $g_k(Q_k^{RCM}(\tau))$, where $Q_k^{RCM}(\tau)$ is the corresponding empirical quantile function of the climate model output and $g_k(\cdot)$ is a non-decreasing calibration function. The copula density $C_{\boldsymbol{Y}}(u_1, u_2)$ is also estimated from the observed monitoring data to capture the desired dependence structure.

- Step 2: the calibration functions $g_k(\cdot)$ estimated from Step 1 are applied to the quantile functions of the climate model outputs for the future period to obtain a calibrated future marginal distribution for each variable.

- Step 3: to obtain calibrated outputs with the correct dependence at the daily level, we simulate realizations of (u_1, u_2) from the historical copula $C_{\boldsymbol{Y}}$ and the calibrated future quantile function from Step 2 for the same number of future days.

3.7.5 Model Specification

We model quantile functions for the observed data as smooth functions using I-splines as described in Section 3.3 to ensure monotonicity. Here we selected a knot sequence of {0.1, 0.25, 0.5, 0.75, 0.9, 0.95} based on a sensitivity analysis. We model $Q^{obs}(\tau) = \sum_i \beta_i I_i(Q^{RCM}(\tau))$ where $Q^{obs}(\tau)$ is the quantile function of the historical observed data and $Q^RCM(\tau)$ is the empirical quantile function for the climate model data. The marginal quantiles are modeled separately for each climate variable. $I_i(\cdot)$ denotes the i-th I-spline basis function, and β_i is the corresponding regression coefficient. $Q^{obs}(\tau)$ is monotone increasing if $\beta_i > 0$ for all i. Without loss of generality, we standardize each variable to be between 0 and 1. We then assume that $\sum_i \beta_i = 1$. This constraint forces the quantile functions to go through (0,0) and (1,1) such that the corresponding density has support over the [0,1] x [0,1] square. The dimension of the I-splines is dependent on the number of knots selected for the splines. Preliminary cross-validation study of various knot selection options determined placing knots at the 10th, 25th, 50th, 75th, 90th, and 95th quantile levels of the data was sufficiently dense. Joint calibration is accomplished via the copula function.

In addition to the bivariate copula-based calibration, a two-stage quantile calibration method was considered. This method uses the univariate quantile calibration method laid out by Zhou, et al. to calculate the marginal calibration functions independently [48]. The resulting calibrated quantile function, $g_k(Q_k^{RCM}(\tau))$, is used to identify the quantile levels, τ_i, matching the historic observed daily measured values for each variable. The resulting τ_i from both variables are linked by day. These pairs are randomly re-sampled with replacement for the same number of future days, analogous to the simulation step in the general modeling framework. Therefore, the two-stage approach can be viewed as a non-parametric bootstrap method to the copula. This non-parametric method is similar to the Schaake shuffle frequently used in climate research [57, 58, 59, 60].

3.7.6 Parameter Estimation

Parameter estimation and inference are carried out under a Bayesian framework. Bayesian inference is particularly useful for propagating different sources of uncertainties into the final climate projections. The random-walk Metropolis–Hastings algorithm, a Markov Chain Monte Carlo (MCMC) method, is used because the full conditional distributions of all model parameters do not correspond to standard distributions.

The positivity and sum-to-one constraints on the β_i's are handled using a latent variable approach. Specifically, we define $\beta_i = \frac{e^{\beta_i^*}}{\sum_i e^{\beta_i^*}}$ where β_i^* is the unconstrained latent variable. For identifiability purposes, we set β_1^*=0. Here we use a multivariate normal distribution to propose β_i^* values for each climate variable. A truncated normal distribution is used to propose α values for the Gumbel copula with a lower bound of 1 and an upper bound of 200. A normal distribution is used to propose δ values for the Frank copula. The proposal distributions are tuned to achieve acceptance rates of approximately 35%.
Non-informative flat priors are used for β_i^*, α, and δ. The data likelihood, Equation (1), is given via the inverse derivative of the quantile function for the marginal density using Equation (2) and the selected copula density. The Gumbel copula density is given by: $c(u, v, \alpha) = exp\{-[x^\alpha + y^\alpha]^{1/\alpha}\}[(x^\alpha + y^\alpha)^{1/\alpha} + \alpha - 1][x^\alpha + y^\alpha]^{1/(\alpha-2)}(xy)^{\alpha-1}(uv)^{-1}$ (where x=-log u and y=-log v). The bivariate Frank copula density is given by $c(u, v, \delta) = \frac{\delta(1-e^{-\delta})e^{-\delta(u+v)}}{[1-e^{-\delta}-(1-e^{-\delta u})(1-e^{-\delta v})]^2}$. With the use of I-splines, the marginal density has a closed-form expression given by the corresponding M-splines. The density is evaluated over a grid of 200 evenly spaced quantile values within [0,1].

A total of 500,000 iterations were run and thinned every 100 iterations. After burn-in the final number of posterior samples was 3000. Convergence of the MCMC algorithm was assessed using trace plots of key model parameters and deviance.

3.7.7 Cross-Validation Study

We compared our bivariate calibration method to other calibration methods currently used in climate science using the 10 years of historical data (1991-2000) of average daily temperature and daily total solar radiation in a five-fold cross-validation analysis.

The bivariate quantile calibration methods (Gumbel copula, Frank copula, and two-stage) were compared to linear regression and the variance-scaling method. The linear regression method estimated the linear association between daily climate model outputs and the corresponding observed monitor values. The regression coefficients are then applied to the climate model output on each day in the testing dataset. The variance-scaling method forces the mean and variance of a climate model output to equal the mean and variance of historic observed values [46]. First, the mean of the climate model output is linearly scaled by the difference between the historical monthly mean (μ_m) of the observed data and the monthly mean of the climate model data: let $T_1 = \bar{Y}^{RCM}_{future} + \mu_m(\bar{Y}^{obs}_{hist}) - \mu_m(\bar{Y}^{RCM}_{hist})$. The mean-corrected values are then centered to have zero means: $T_2 = T_1 - \mu_m(T_1)$. Next, the standard deviations are used to scale the observed to the climate model: $T_3 = T_2 * [\frac{\sigma_m(Y^{obs}_{hist})}{\sigma_m(T_2)}]$. Finally, calibrated outputs are obtained by shifting T_3 back using the corrected mean: $T_3 + \mu_m(T_1)$.

Calibration methods were compared by calculating the continuous ranked probability score (CRPS) of the quantile values over a sequence of 101 quantile levels between the calibrated climate model outputs for the testing dataset and the corresponding observed values. As seen in Table 8, the bivariate quantile calibration methods generally had a lower CRPS than the other calibration methods for both average temperature and solar radiation. The Gumbel copula method had the lowest CRPS for average temperature in two of the climate models while the Frank copula method had the lowest CRPS for solar radiation across three of the climate models.

The variance scaling method resulted in CRPSs higher than those of the bivariate calibration method in all cases except the daily average temperature in the HRM3-hadCM3 model and the solar radiation in the CRCM-CGCM3 model. The linear regression calibration generally performed the worst of the calibration methods. Here the calibrated output represents the conditional mean of the observation given the model output (regression towards the mean), resulting in poorly characterized distribution tails.

Table 8: Continuous ranked probability scores (CRPS) from a cross-validation analysis of methods used in calibrating climate models. The bivariate quantile methods are compared to univariate scaling methods on the grid cell corresponding to the monitor in Atlanta, GA. The lowest CRPS for each variable-model combination is bold.

Climate Model:	Model 1 [a]	Model 2 [b]	Model 3 [c]	Model 4 [d]
Temperature (C)				
No Calibration	1.639	2.689	1.663	1.478
Linear Regression	0.426	0.598	0.776	0.462
Variance Scaling	0.511	0.642	0.771	0.456
Two-Stage	0.563	0.566	0.627	0.507
Gumbel	0.313	0.303	**0.346**	**0.271**
Frank	**0.272**	**0.260**	0.410	0.366
Solar Radiation (W/m^2)				
No Calibration	747	605	535	540
Linear Regression	174	141	126	135
Variance Scaling	175	107	**106**	142
Two-Stage	147	138	110	120
Gumbel	**103**	107	154	115
Frank	108	**92**	107	**103**

[a] CRCM-CCSM
[b] CRCM-CGCM3
[c] HRM3-GFDL
[d] HRM3-HadCM3

Table 9 gives the Pearson's correlation and Kendall's tau of the calibrated model output in the testing dataset compared to those observed. The correlation of the calibrated climate model outputs was best retained by the two-stage quantile calibration method. This was expected as the linkage of the calibrated outputs

is based on the observed daily quantile values of the two variables. The copula calibration method led to correlations that underestimated the observed historic correlation. The linear regression and variance scaling correlations were the same as the uncalibrated correlations as both methods are monotone transformations. The variance scaling and linear regression methods resulted in correlations that were slightly higher than the observed correlation in two models and slightly lower than the observed correlations in the other two models.

Table 9: Correlation measures from a cross-validation analysis of methods used in calibrating climate models. The bivariate quantile methods are compared to univariate scaling methods on the grid cell corresponding to the monitor in Atlanta, GA.

Climate Model:	Model 1 [a]	Model 2 [b]	Model 3 [c]	Model 4 [d]
Pearsons Correlation				
Observed	0.564	0.564	0.564	0.564
No Calibration	0.685	0.613	0.536	0.461
Linear Regression	0.685	0.613	0.536	0.461
Variance Scaling	0.685	0.613	0.536	0.461
Two-Stage	0.575	0.566	0.557	0.561
Gumbel	0.343	0.370	0.392	0.382
Frank	0.490	0.482	0.493	0.484
Kendall's Tau				
Observed	0.403	0.403	0.403	0.403
No Calibration	0.492	0.433	0.375	0.330
Linear Regression	0.492	0.433	0.375	0.330
Variance Scaling	0.492	0.433	0.375	0.330
Two-Stage	0.391	0.409	0.411	0.411
Gumbel	0.277	0.268	0.286	0.275
Frank	0.344	0.337	0.346	0.339

[a]CRCM-CCSM
[b]CRCM-CGCM3
[c]HRM3-GFDL
[d]HRM3-HadCM3

3.7.8 Application to Climate Projections

We assess future climate by estimating the mean and 95th percentile of the climate variables over the period 2041-2070. These statistics are compared to those during

the historical period (1991-2000). The quantile functions of the calibrated climate models had less variability between them than the uncalibrated climate models, particularly in the upper tails of the distributions for both temperature and solar radiation (Figures 21 and 22). Figure 21 shows that the quantile functions of the calibrated historic climate models (1991-2000) fit the quantile function of the observed monitoring data better than the quantile functions of the uncalibrated climate models. The raw future period (2041-2070) climate model outputs have considerably higher upper tails than the observations. The calibrated distributions are closer to the historic distribution than the uncalibrated distributions thus demonstrating our ability to resolve distributional discrepancies. Figures 21 and 22 also show that the two copula forms gave similar projections. We note the posterior intervals for the two-stage calibration are much smaller. This is likely because the estimated τ_i values being resampled are assumed known.

Figure 21: Quantiles functions of average daily temperature and total daily solar radiation from raw climate model projections and calibrated climate model simulation during the historical period (1991-2000). Calibrated models are calibrated with the best method based on the CRPS from the cross-validation.

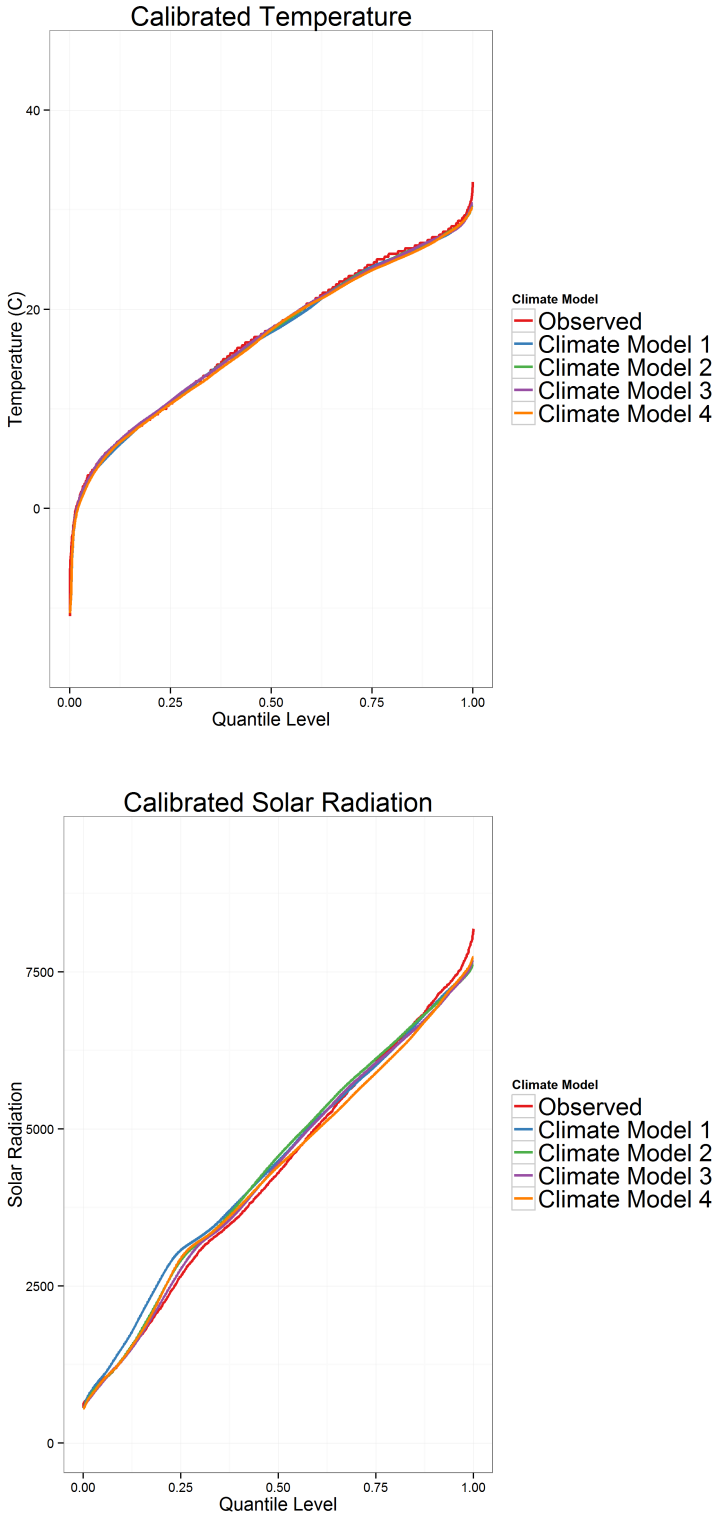
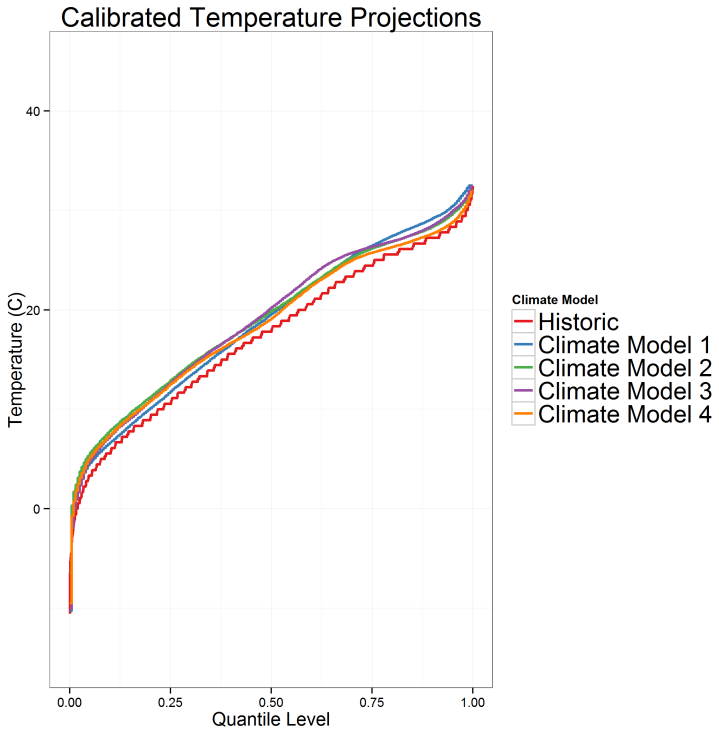
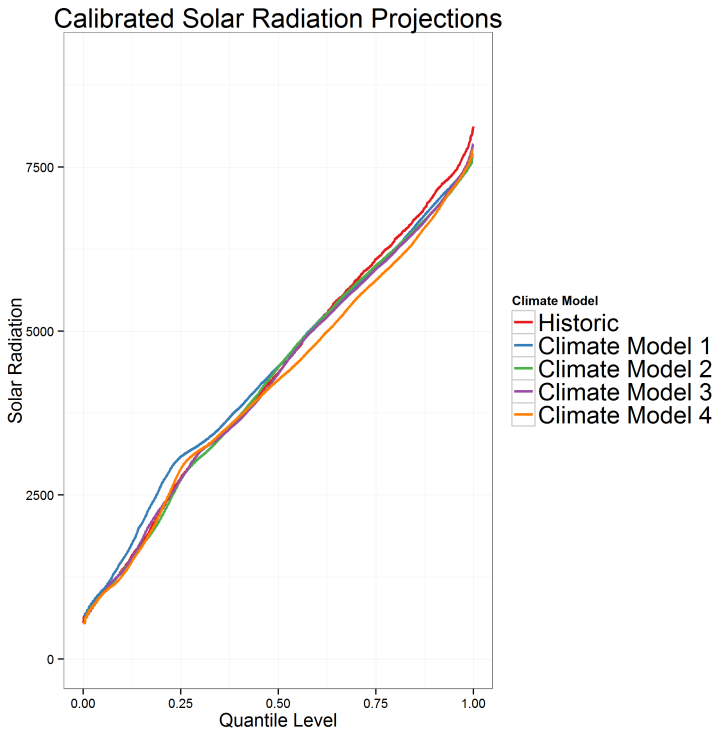


Figure 22: Quantiles functions of average daily temperature and total daily solar radiation from raw climate model projections and calibrated climate model projections (2041-2070). Climate model projections are compared to the quantile functions from the historic monitoring data (1991-2000). Calibrated models are calibrated with the best method based on the CRPS from the cross-validation.





Calibrated climate models had higher projected mean values for temperature and lower projected mean values for solar radiation in comparison to uncalibrated climate models (Figure 23). At the 95th percentile level, calibrated climate models had lower projected values for both temperature and solar radiation compared to uncalibrated climate model outputs (Figure 24). More importantly, calibration reduces the between-model variability in projections considerably. The differences between the calibrated values and the historic values were less than the differences between the historic values and the uncalibrated climate model values at the 95th percentile in all climate models. The calibrated climate models projected an increase in temperature of 1.68 to 1.81 degrees in the mean and an increase of 0.73 to 2.04 degrees in the 95th percentile. For solar radiation, the calibrated model outputs indicate no change in mean solar radiation and a decrease in solar radiation at the 95th percentile. This is the opposite conclusion to the conclusion obtained from using uncalibrated model outputs, which projects an increase in solar radiation at both the mean and the 95th percentile.

Our estimates of the Gumbel copula parameter, α, were similar across the four climate models with $\hat{\alpha} = \{1.32, 1.29, 1.32, 1.31\}$. Our estimates of the Frank copula parameter, δ, were also similar across the four climate models with $\hat{\delta} = \{3.31, 3.22, 3.33, 3.25\}$.

Figure 23: Projected mean percentile values for average daily temperature and total daily solar radiation values from raw climate model projections and calibrated climate model projections (2041-2070). Climate model projections are compared to observed values from historic monitoring data (1991-2000). Posterior mean and 95% posterior intervals are shown when available.

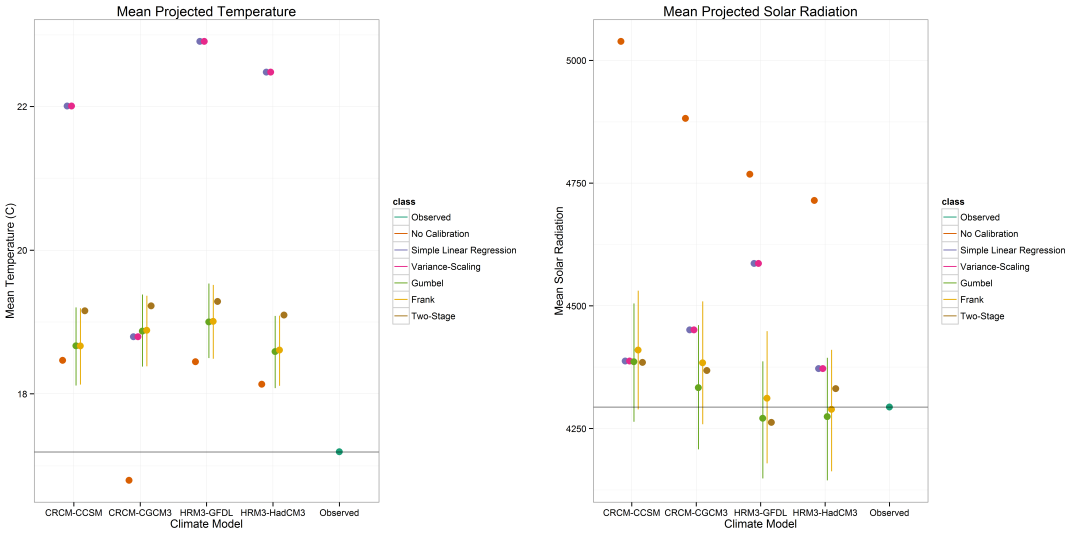
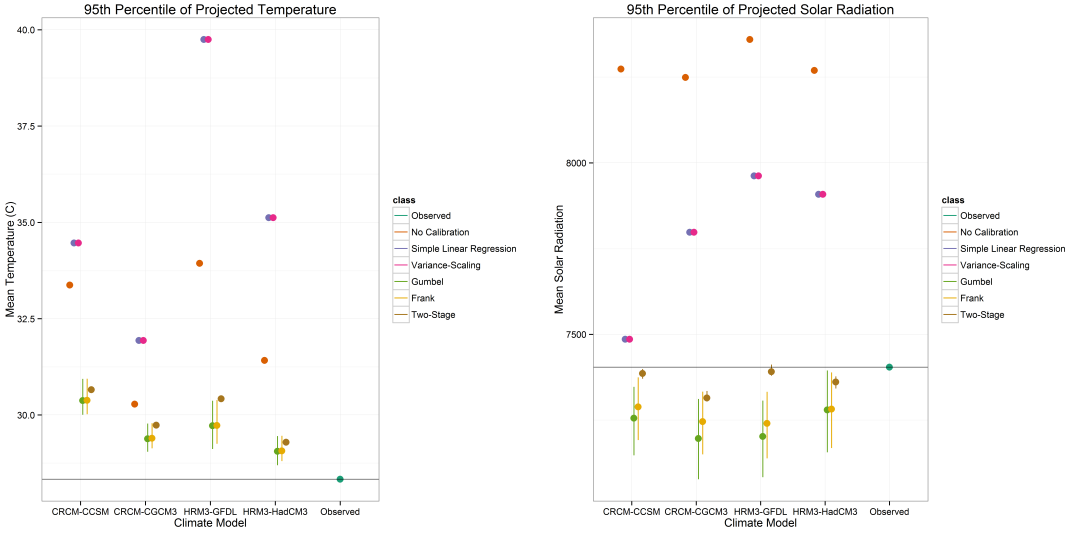


Figure 24: Projected 95th percentile values for average daily temperature and total daily solar radiation values from raw climate model projections and calibrated climate model projections (2041-2070). Climate model projections are compared to observed values from historic monitoring data (1991-2000). Posterior mean and 95% posterior intervals are shown when available.



3.7.9 Discussion

Improving the quality of climate model projections is a priority in both climate science and public health science. Accurate climate projections are important

for policy makers when deciding on how to respond to the threat of climate change. Adaptation (adjusting to unpreventable climate change) and mitigation (preventing further climate change as much as possible) to the changing climate are complex and expensive endeavors with high political and economic costs. The World Meteorological Organization asserted that "only by assessing what the real impact will be on different countries will it be possible to justify difficult social and economic policies to avert a dangerous deterioration in the global climate" [10]. Accurate and consistent climate projections, especially those with uncertainty estimates, are essential to providing decision-makers information about the impacts of climate change.

One such impact of climate change is human health. Climate change has the potential to endanger human health across the globe through several interconnected channels, including higher temperatures, more extreme weather, and sea level rise. These shifting climate factors can lead to heat waves with increased frequency and intensity, higher levels of air pollution, changes in vector ecology, increasing allergens, decreased water quality, food supply interruptions, and environmental degradation [3]. Direct heat-related mortality and morbidity (e.g., heat stroke) are expected to increase as the global temperature rises [61, 62, 63]. Increased temperatures can also indirectly affect human health through mechanisms such as increases in ambient air pollution concentrations [4, 14]. Asthma aggravation has been tied to increases in ambient air pollution levels as air pollution can induce airway inflammation [22, 12].

Calibration is one strategy to improve the accuracy of climate model projections. The proposed bivariate quantile calibration method is aimed at improving the accuracy of the upper-tail climate projections in particular. Calibrating climate model outputs against historical observations via a statistical model also provides uncertainty measures (here, a posterior predictive interval) that can employed in the decision-making process. Under our Bayesian framework, parameters for the marginal densities and the copula are estimated jointly. This

approach better propagates uncertainties in projections compared to methods that employ a stepwise estimation approach [46, 64, 65].

Several elements of the proposed calibration method warrant further examination. First, how well the dependence structure between the variables is retained in the calibrated models outputs is influenced by the choice of copula. In our application, the Gumbel and Frank copulas appear to underestimate the dependence between temperature and solar radiation. Copula selection is an active area of research including work in the use of goodness-of-fit statistics [66], copula structures [67], and information criteria [68, 69]. Copula selection can be also driven by the dependence structure, such as symmetry and tail dependence, that one wishes to capture based on knowledge of the physical system. The use of a parametric copula facilitates our approach to be extended to more than two variables.

Second, any calibration method is dependent on the quality and assumptions of the climate model used, where minimal discrepancies during the historical period will lead to smaller uncertainty in the calibrated projections. This analysis considered only four climate models and assessment of more climate models of different spatial resolution and global regions are needed to fully assess the benefits and limitations of statistical calibration in climate science research.

Third, output calibration makes the necessary but naive assumption that bias observed during the historical period in the climate model will be identical to the future periods. Finally, we only consider calibration at a single location. Climate model bias is likely to exhibit spatial trends and one natural extension is to consider spatial quantile processes [47] when performing calibration at multiple locations. Incorporating spatial dependence will also allow interpolation of the calibration parameters such that locations with no monitoring data can be calibrated.

4 Aim 3: Projection of Future Ambient Air Pollution and Asthma Emergency Department Visits

4.1 Background

Climate models have been used to estimate future health impacts due to air pollution, often with a focus on ozone and respiratory disease [70, 71, 72, 73, 74, 75]. These studies used climate model outputs, including O_3 concentrations, to project changes in future mortality under climate change scenarios. Both O_3 concentrations and air pollution mortality are expected to increase in the future based on the findings of these studies. Research on health impacts resulting from changes in future $PM_{2.5}$ concentrations has had mixed findings [76, 77]. $PM_{2.5}$ concentrations are heavily dependent on emissions and any projection must make assumptions on future environmental policy. Climate models that assume a reduction in the emissions contributing to $PM_{2.5}$ project a subsequent reduction in $PM_{2.5}$ concentrations and a corresponding reduction in morbidity and mortality due to $PM_{2.5}$.

The reliability of all health impact projections is dependent on the accuracy of the climate models on which they are based. Here we will extend the bivariate calibration method detailed above in Section 3.7 by considering the spatial structure of climate model outputs on a contiguous areal grid. We develop a spatial calibration method that calibrates climate model outputs over a gridded area. This spatial calibration allows for local calibration (at each grid cell) to potentially improve the accuracy of the climate model outputs. The bivariate calibration also allows for the retention of the observed correlation structure between climate model variables.

This spatial extension allows the projection of calibrated climate model variables at locations that do not have a monitoring site to provide observed

data. We project calibrated values for two air pollutants: O_3 and $PM_{2.5}$. We subsequently use these projected pollutant concentrations to project future asthma emergency department (ED) visits in two metropolitan areas: Atlanta, GA and Birmingham, AL.

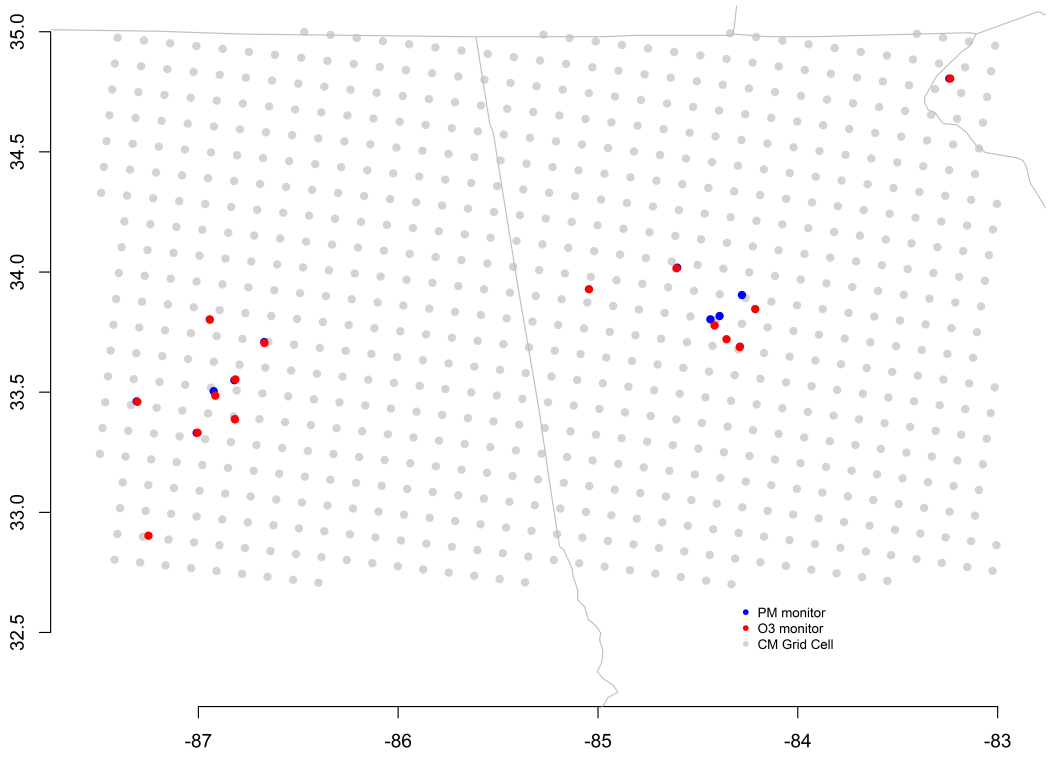
4.2 Data

The climate model used in this analysis is a coupled global and region climate modeling system that links a global climate-chemistry model with regional weather (WRF) and air quality (CMAQ) models [78, 79, 70]. The WRF-CMAQ climate model has a spatial resolution of 12 km x 12 km grid cells. The spatial domain of the climate model was limited to latitudes between 32.7° and 35° and longitudes between -83° and -87.5° (northern Georgia and Alabama). At each grid cell the climate model simulates values of 8-hour maximum O_3 (ppb) and 24-hour average $PM_{2.5}$ ($\mu g/m^3$) for 2001-2004 and 2055-2059. For this analysis the 8.5 RCP scenario is used, meaning the WRF model assumes high CO_2 emissions and large population increases. The CMAQ model also makes assumptions about air pollutant emissions. Most importantly for this work, the CMAQ model assumes a large reduction in the emissions of the constituents of $PM_{2.5}$, resulting in a model that greatly under-estimates $PM_{2.5}$ concentrations in the historic period.

The observed data consists of 8-hour maximum O_3 (ppb) and 24-hour average $PM_{2.5}$ ($\mu g/m^3$) measured from 2001-2004 at air quality monitors located within the study area. The O_3 and $PM_{2.5}$ monitors that were not co-located were linked to the closest monitor by determining the shortest great circle distance between the pollutants' monitors. Monitoring pairs ($PM_{2.5}$ monitors each matched to the closest O_3 monitor) that had observed data during the summer months (May - September) 2001-2004 were selected for use. Distances between O_3 and $PM_{2.5}$ monitors ranged from 0 km to 8.9 km. In the end, 17 monitors were used (Figure 25). Climate model grid cells were linked to the observed monitors by finding the

closest grid cell centroid for each monitor pair. Both climate model and observed data were limited to warm months (May - September) in this analysis.

Figure 25: Study area for spatial calibration analysis with air quality monitors



4.3 Data Descriptives

The mean observed correlation between the pollutants at each monitoring location was 0.588 while the mean correlation between the pollutants at each linked grid cell from the climate model was -0.058. The correlations at each monitoring location is displayed in Table 10. The climate model consistently underestimates the temporal correlation between O_3 and $PM_{2.5}$. As Figures 26 and 27 show, the temporal patterns of the pollutants are not well simulated by the uncalibrated climate model during the historical period. Specifically, the model simulation shows less day-to-day variation than the observations.

Table 10: Pearson's Correlation of Observed and Uncalibrated Data

Monitor Pair	Observed	Uncalibrated (2001-2004)
1	0.745	-0.117
2	0.688	-0.084
3	0.718	-0.11
4	0.657	-0.101
5	0.683	-0.021
6	0.493	-0.049
7	0.672	0.092
8	0.661	-0.106
9	0.58	-0.031
10	0.496	-0.031
11	0.499	-0.07
12	0.433	-0.043
13	0.581	-0.016
14	0.471	-0.058
15	0.663	-0.099
16	0.55	-0.031
17	0.413	-0.112
Mean	**0.588**	**-0.058**

Figure 26: Time series of daily observed O_3 and $PM_{2.5}$ at a monitor pair in northeast Birmingham, AL during summer 2002

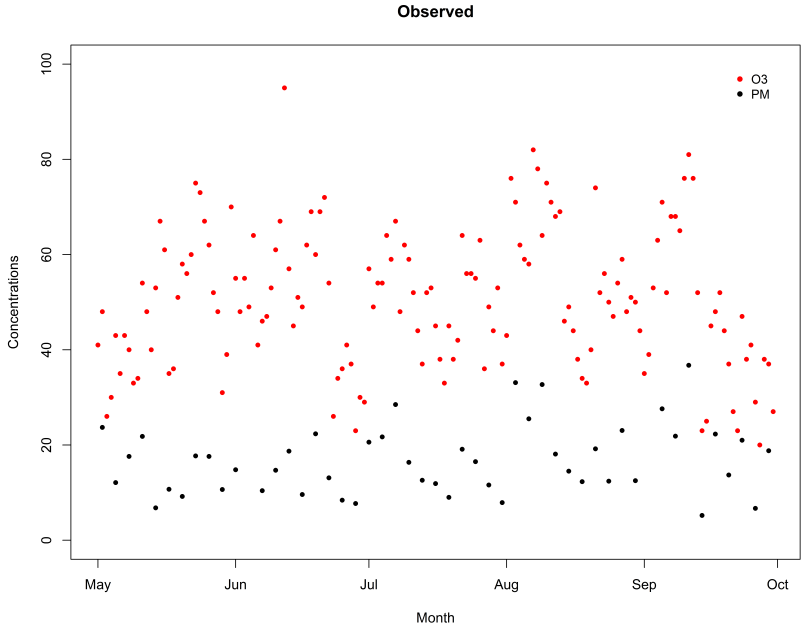
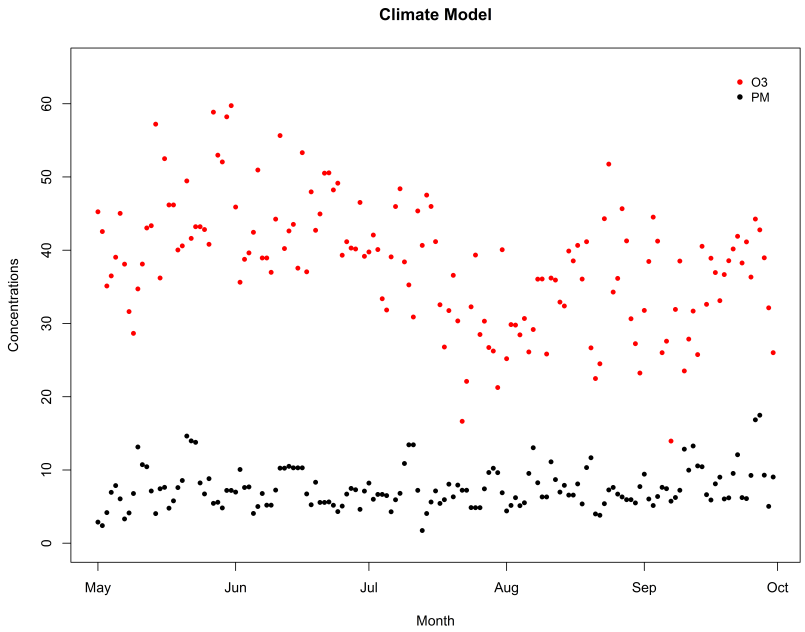


Figure 27: Time series of daily simulated O_3 and $PM_{2.5}$ at a monitor pair in northeast Birmingham, AL during summer 2002



As Figure 28 shows, the uncalibrated climate model is biased upwards in the hindcast for O_3 resulting in higher concentrations of O_3 than were actually

observed. As Figure 29 shows, uncalibrated climate model outputs of $PM_{2.5}$ over-predicts days with lower $PM_{2.5}$ concentrations and under-predicts days with higher $PM_{2.5}$ concentrations.

Figure 28: Quantiles of projected and observed O_3 at a monitor pair in northeast Birmingham, AL, 2001-2004

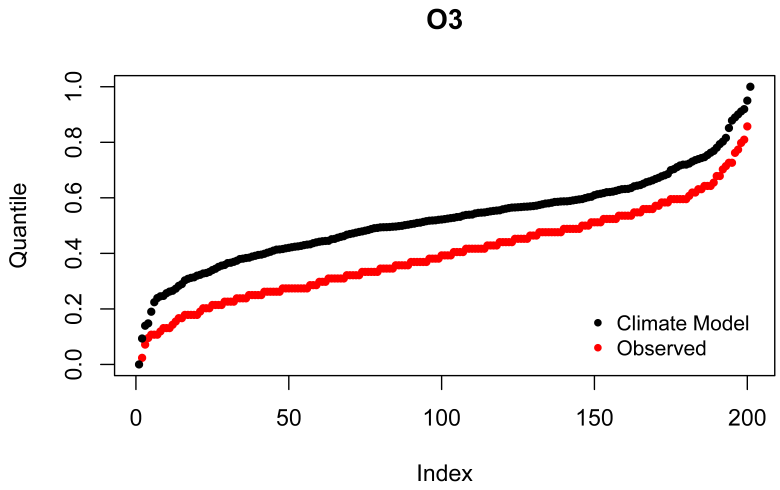
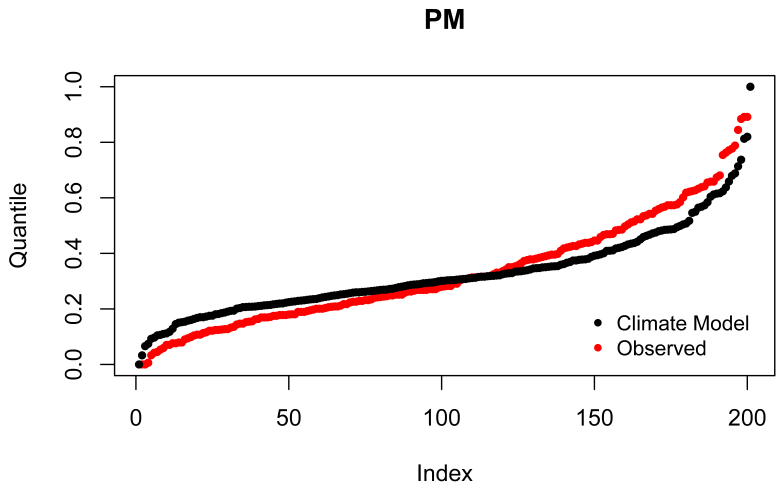


Figure 29: Quantiles of projected and observed $PM_{2.5}$ at a monitor pair in northeast Birmingham, AL, 2001-2004



4.4 Parameter Estimation

Estimation was done in two-stages as the model was too large to estimate at once. The first stage comprises of fitting the calibration model separately at each

monitor location. The second stage uses the parameters from all monitor locations and pools them spatially to obtain a spatially-varying calibration function. Here the knot sequence of the I-splines used to model the empirical quantile functions was {0.1, 0.25, 0.5, 0.75, 0.95}.

4.4.1 Statistical Model

There are S monitoring locations matched to S climate model grid cells. At each location, denoted by s, a bivariate calibration modeled is fitted following the method proposed in Aim 2.

At each location s, $Q^{Obs}_{s,PM}(\tau)$ is the quantile function for the observed $PM_{2.5}$ values, $Y^{Obs}_{s,PM}$, and $Q^{Obs}_{s,O_3}(\tau)$ is the quantile function for the observed O_3 values, Y^{Obs}_{s,O_3}. Similarly, $Q^{CM}_{s,PM}(\tau)$ is the quantile function for the linked modeled $PM_{2.5}$ values, $Y^{CM}_{s,PM}$, and $Q^{CM}_{s,O_3}(\tau)$ is the quantile function for the matched-location modeled O_3 values, Y^{CM}_{s,O_3}.

The model is set up as follows:

$$Q^{Obs}_{s,PM}(\tau) = g_{s,PM}[Q^{CM}_{s,PM}(\tau); \beta_{s,PM}]$$

$$Q^{Obs}_{s,O_3}(\tau) = g_{s,O_3}[Q^{CM}_{s,O_3}(\tau); \beta_{s,O_3}]$$

$$f(Y^{Obs}_{s,PM}, Y^{Obs}_{s,O_3}) = f(Y^{Obs}_{s,PM}) f(Y^{Obs}_{s,O_3}) c(\tau^{Obs}_{1,s}, \tau^{Obs}_{2,s}; \theta)$$

where $c(\tau^{Obs}_{1,s}, \tau^{Obs}_{2,s}; \theta)$ is the copula density, τ_1 are the quantile levels for $PM_{2.5}$, and τ_2 are the quantile levels for O_3. The density of $f(Y)$ where $Y = \{Y^{Obs}_{s,PM}, Y^{Obs}_{s,O_3}, Y^{CM}_{s,PM}, Y^{CM}_{s,O_3}\}$ is given by:

$$f(Y) = \frac{1}{\frac{dQ_Y(\tau^*)}{d\tau}\big|_{\tau^* = \{\tau : Q_Y(\tau) = y\}}} \quad ;$$

$$\tau_{1,s}^{Obs} = \{\tau : Q_{s,PM}^{Obs}(\tau) = Y_\tau\} \quad ;$$

$$\tau_{2,s}^{Obs} = \{\tau : Q_{s,O_3}^{Obs}(\tau) = Y_\tau\}$$

The priors associated with the above model's parameters are given below. We consider 2 copula functions: Gumbel and Frank:

$$\theta_{Gubmel} \sim TN(\theta, 1, 1, \infty)$$

$$\theta_{Frank} \sim N(\theta, 0.1)$$

The priors for $\beta_{s,PM}$ and $\beta_{s,O3}$ are N(0, δ), where δ equals the variance of the initial β estimates multiplied by a tuning parameter, with constraints:

$$\sum_i \beta_i = 1$$
$$\beta_i = \frac{e^{\beta_i^*}}{\sum_i e^{\beta_i^*}}$$
$$\beta_1^*=0$$

The one-stage estimation results in a variable-specific $\hat{\boldsymbol{\beta}}$ vector and corresponding $\hat{\boldsymbol{\Sigma}}$ covariance matrix for every monitoring location. Each of the $\hat{\boldsymbol{\beta}}$ vectors consist of the regression coefficients describing the calibration relationship at that location and have the dimensions (K x 1) where K is determined by the dimensions of the calibration design matrix and is dependent on the number of knots in the I-spline. Following the calibration at each location, the calibration parameters, $\hat{\boldsymbol{\beta}}_s$, are pooled across all sites for each variable.

For s = 1, 2, . . . , S,

$$\hat{\boldsymbol{\beta}}_{s,PM} \sim N(\boldsymbol{\beta}_{s,PM}, \hat{\boldsymbol{\Sigma}}_{s,PM})$$

$$\hat{\boldsymbol{\beta}}_{s,O_3} \sim N(\boldsymbol{\beta}_{s,O_3}, \hat{\boldsymbol{\Sigma}}_{s,O_3})$$

$$\boldsymbol{\beta}_{s,PM} \sim N(\boldsymbol{\mu}_{PM}, \boldsymbol{V}_{PM})$$

$$\boldsymbol{\beta}_{s,O_3} \sim N(\boldsymbol{\mu}_{O_3}, \boldsymbol{V}_{O_3})$$

We combine the location-specific parameters into a single vector. For each pollutant separately, the (K x 1) $\hat{\boldsymbol{\beta}}$ vectors are stacked into one large $\boldsymbol{\beta\cdot}$ vector with the dimensions (SK x 1).

$$\hat{\boldsymbol{\beta}}\boldsymbol{\cdot} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{\mathbf{1}} \\ \hat{\boldsymbol{\beta}}_{\mathbf{2}} \\ \vdots \\ \hat{\boldsymbol{\beta}}_{\boldsymbol{S}} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_{11} \\ \vdots \\ \hat{\beta}_{1K} \\ \hat{\beta}_{21} \\ \vdots \\ \hat{\beta}_{2K} \\ \vdots \\ \hat{\beta}_{SK} \end{bmatrix}$$

$$\boldsymbol{\beta}\boldsymbol{\cdot} = \begin{bmatrix} \boldsymbol{\beta}_{\mathbf{1}} \\ \boldsymbol{\beta}_{\mathbf{2}} \\ \vdots \\ \boldsymbol{\beta}_{\boldsymbol{S}} \end{bmatrix} = \begin{bmatrix} \beta_{11} \\ \vdots \\ \beta_{1K} \\ \beta_{21} \\ \vdots \\ \beta_{2K} \\ \vdots \\ \beta_{SK} \end{bmatrix}$$

A vector of the mean of the $\hat{\boldsymbol{\beta}}_k$ values across all locations, $\boldsymbol{\mu}$ is created. This vector is repeated S times to create one long vector, $\boldsymbol{\mu\cdot}$ with the dimensions (SK x 1).

$$\boldsymbol{\mu}\boldsymbol{\cdot} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix}$$

The covariance matrix is similarly set up where $\hat{\boldsymbol{\Sigma}}_s$ represents the (K x K) covariance matrix at each monitor location. These $\boldsymbol{\Sigma}_s$ covariance matrices are set up in a large block diagonal covariance matrix:

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\cdot}} = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_1 & & & \\ & \hat{\boldsymbol{\Sigma}}_2 & & 0 \\ 0 & & \ddots & \\ & & & \hat{\boldsymbol{\Sigma}}_S \end{bmatrix}$$

For each location s, $\boldsymbol{V}$ is a (K x K) diagonal vector of variances of each β_k coefficient across the s locations. Covariances between β_s coefficients at the same location were assumed to be 0. Similar to the $\boldsymbol{\mu}_{\boldsymbol{\cdot}}$ vector that summarized information across all locations, a summary covariance matrix, $\boldsymbol{V}_{\boldsymbol{\cdot}}$, was created by summarizing the $\boldsymbol{V}$ for all s and has dimensions (SK x SK).

$$\boldsymbol{V}_{\boldsymbol{\cdot}} = \begin{bmatrix} \boldsymbol{V} & \cdots & \cdots & \\ \cdots & \boldsymbol{V} & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \boldsymbol{V} \end{bmatrix}$$

Our final model for $PM_{2.5}$ is:

$$\hat{\boldsymbol{\beta}}_{\cdot PM} \sim N(\boldsymbol{\beta}_{\cdot PM}, \hat{\boldsymbol{\Sigma}}_{\cdot PM})$$

$$\boldsymbol{\beta}_{\cdot PM} \sim N(\boldsymbol{\mu}_{\cdot PM}, \boldsymbol{V}_{\cdot PM})$$

Similarly for O_3, we have:

$$\hat{\boldsymbol{\beta}}_{\cdot O3} \sim N(\boldsymbol{\beta}_{\cdot O3}, \hat{\boldsymbol{\Sigma}}_{\cdot O3})$$

$$\boldsymbol{\beta}_{\cdot O3M} \sim N(\boldsymbol{\mu}_{\cdot O3}, \boldsymbol{V}_{\cdot O3})$$

Our goal is to estimate a set of $\boldsymbol{\beta}$ coefficients that can be used to simulate calibrated values of O_3 and $PM_{2.5}$ based on all of the monitor locations. We accomplish this using a Gibbs sampler where the posterior distributions are known. Here $\mathbf{X}$ represents a (SK x K) design matrix that consists of (K x K) identity matrices repeated S times.

$$\boldsymbol{\beta}_{\cdot} \sim N((\hat{\boldsymbol{\Sigma}}_{\cdot}^{-1} + \boldsymbol{V}_{\cdot}^{-1})^{-1}(\hat{\boldsymbol{\Sigma}}_{\cdot}^{-1}\hat{\boldsymbol{\beta}}_{\cdot} + \boldsymbol{V}_{\cdot}^{-1}\boldsymbol{\mu}_{\cdot}), (\hat{\boldsymbol{\Sigma}}_{\cdot}^{-1} + \boldsymbol{V}_{\cdot}^{-1})^{-1})$$

$$\boldsymbol{\mu}_{\cdot} \sim N((\boldsymbol{X}^T\boldsymbol{V}_{\cdot}^{-1}\boldsymbol{X}^{-1})^{-1}\boldsymbol{X}^T\boldsymbol{V}_{\cdot}^{-1}\boldsymbol{\beta}_{\cdot}, (\boldsymbol{X}^T\boldsymbol{V}_{\cdot}^{-1}\boldsymbol{X})^{-1})$$

$$\boldsymbol{V}_{k,k} \sim InvGamma\Big(\frac{S}{2} + \alpha_0, \frac{\sum_{s=1}^{S}(\beta_{s,k} - \mu_k)^2}{2} + \alpha_0\Big)$$

The off-diagonal portions of the covariance matrix, $\boldsymbol{V}_{\cdot}$, represent the spatial covariance between the monitor locations. If there was no spatial relationship, the off-diagonals would be 0 and $\boldsymbol{V}_{\cdot}$ could be estimated by taking the Kronecker product of an (S x S) identity matrix and the $\boldsymbol{V}$ matrix, $\boldsymbol{V}_{\cdot} = \mathbf{I} \otimes \boldsymbol{V}$. Spatial

covariances between the locations are incorporated with an exponential spatial covariance function, $\mathbf{S} = \exp(\frac{-1}{\rho} \text{ x } \mathbf{D})$ where $\mathbf{D}$ is a matrix of the pairwise distances between the monitor locations. Thus, the covariance matrix for spatial dependence $\boldsymbol{V\cdot} = \mathbf{S} \otimes \boldsymbol{V}$.

The parameter ρ is a spatial smoothing parameter that determines how much the spatial points are influenced by the surrounding points. A rough estimate of $\rho = \frac{-log(r)}{d}$ where d is the distance between two points and r is their spatial correlation. The value of ρ was determined by selecting two points and calculating the distance between them and then conducting a sensitivity analysis over various potential spatial correlations.

4.4.2 Prediction of New Location

The posterior $\boldsymbol{\beta\cdot}$, $\boldsymbol{\mu\cdot}$, and $\boldsymbol{V\cdot}$ estimates, which incorporate the spatial relationship between the monitoring locations, are used to predict a set of β (K x 1) for a new location. For any new location, the distance matrix, $\mathbf{D}$, is expanded to include the pairwise distances between the observed locations and the new, unobserved location. This $\mathbf{D}_{new}$ (S+1) x (S+1) is used to calculate a new spatial correlation matrix, $\mathbf{S}_{new} = \exp(\frac{-1}{\rho} \text{ x } \mathbf{D}_{new})$. Subsequently, the joint covariance matrix is found, $\boldsymbol{V\cdot}_{new} = \mathbf{S}_{new} \otimes \boldsymbol{V}$.

$$\mathbf{V}_{new} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

where:

$\boldsymbol{\Sigma}_{11}$ is $(sk \text{ x } sk)$

$\boldsymbol{\Sigma}_{12}$ is $(sk \text{ x } k)$

$\boldsymbol{\Sigma}_{21}$ is $(k \text{ x } sk)$

$\boldsymbol{\Sigma}_{22}$ is $(k \text{ x } k)$

The posterior predictive distribution is the conditional multivariate Gaussian distribution.

$$\beta_{new} \mid \beta_k = N\Big(\mu_k + \Sigma_{21}\Sigma_{11}^{-1}(\beta_k - \mu_k), v_k - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Big)$$

Finally, the estimated β_{new} are used to calculate a predicted quantile function for the new location, $Q(\tau) = \beta_{new}\, I$.

4.5 Health Impact Projections

The projected pollutant values obtained using the above method were used to project how asthma ED visits will change in the future under the changing climate. Projections of the change in asthma morbidity were calculated for the 20-county metropolitan area of Atlanta, GA and the 9-county metropolitan area of Birmingham, AL.

4.5.1 Data

Air quality and ED visit data were collected in 20 counties around Atlanta and in 9 counties around Birmingham (years 2002-2008). The air quality data included daily population-weighted average concentrations of ambient 8-hour maximum O_3 (ppb) and 24-hour average $PM_{2.5}$ ($\mu g/m^3$). The population-weighted average concentrations were calculated using data from all available monitors for each pollutant in the study regions. We included monitors from the US EPA Air Quality System, the SouthEastern Aerosol Research and Characterization (SEARCH) network, and the Assessment of the Spatial Aerosol Composition (ASACA) network. The monitoring data were first spatially interpolated across the study's geographic domain and then a population-weighted average concentration estimate was derived using tract-level population estimates from the 2010 US Census. Daily minimum, maximum, and dew-point temperature data were obtained from the National Climate Data Center.

Patient-level ED visit data were obtained from individual hospitals and the Georgia Hospital Association. Asthma ED visits were identified using the primary

International Classification of Diseases, 9th Revision diagnosis code for the record as either 493 (asthma) or 786.07 (wheeze).

4.5.2 Analysis

Change in asthma morbidity is dependent on two factors: how the concentration of air pollution changes and the effect the air pollution level has on asthma. In order to estimate the change in ambient air pollutant concentration we used the calibrated future O_3 and $PM_{2.5}$ concentrations that were projected at each grid cell using the above method. Grid cells that fell within the 20- and 9-county metropolitan regions were categorized as Atlanta or Birmingham grid cells, respectively. The average concentrations of O_3 and $PM_{2.5}$ were calculated by taking the mean of the concentrations for all of the grid cells within each group. The mean observed concentrations of O_3 and $PM_{2.5}$ were calculated by taking the mean of all of the observed concentrations from the monitors that fell within the two metropolitan regions separately.

Daily time-series Poisson regression models were used to assess the relationship between asthma ED visit counts and ambient pollutant concentrations. For these analyses the temporal data were limited to the warm months (May - October) for each year. The outcome of interest was daily counts of emergency department (ED) visits for asthma. The main exposure variables were daily concentrations of O_3 (ppb) and $PM_{2.5}$ ($\mu g/m^3$). Other covariates in the health effects model were cubic splines for lag 0 maximum daily temperature, cubic splines for lag 1-2 minimum temperature, cubic splines for lag 0-2 dew point temperature, time splines for each year, indicator variables for each year, weekday, holidays, and hospital, and an interaction variable for each year and the first time spline.

The full models were:

$$log(Y_{ATL}) = \beta_0 + \beta_1 O_3 + \beta_2 PM_{2.5} + \sum_{a=1}^{3} \beta_a MaxTemp0_a + \sum_{b=1}^{3} \beta_b MinTemp12_b + \sum_{c=1}^{3} \beta_c DP02_c + \sum_{d=1}^{7} \beta_d Year_d + \sum_{e=1}^{8} \beta_e s(Time_e) + \sum_{f=1}^{7} \beta_f Year * Time1_f + \sum_{g=1}^{7} \beta_g Weekday_g + \sum_{h=1}^{6} \beta_h Holiday_h + \sum_{i=^48}^{48} \beta_i Hopsital_i + \epsilon$$

$$log(Y_{BHAM}) = \beta_0 + \beta_1 O_3 + \beta_2 PM_{2.5} + \sum_{a=1}^{3} \beta_a MaxTemp0_a + \sum_{b=1}^{3} \beta_b MinTemp12_b + \sum_{c=1}^{3} \beta_c DP02_c + \sum_{d=1}^{7} \beta_d Year_d + \sum_{e=1}^{8} \beta_e s(Time_e) + \sum_{f=1}^{7} \beta_f Year * Time1_f + \sum_{g=1}^{7} \beta_g Weekday_g + \sum_{h=1}^{6} \beta_h Holiday_h + \sum_{i=^8}^{8} \beta_i Hopsital_i + \epsilon$$

The β estimates for O_3 and $PM_{2.5}$ (and their corresponding standard errors) were used in the prediction of future asthma cases. For each city individually, the mean changes in pollutant concentrations were Δ_{PM} = mean($PM_{2.5}^{CM}$) - mean($PM_{2.5}^{Obs}$) and Δ_{O_3} = mean(O_3^{CM}) - mean(O_3^{Obs}) where the observed time period is 2001-2004 and the climate model time period is 2055-2059. Calculation of the change in daily asthma ED visits was done using the following formula:

$$\text{EED} = \text{M}^*(\exp\{\beta_{O_3}\Delta_{O_3} + \beta_{PM}\Delta_{PM}\} - 1)$$

where M = Mean annual ED visits, Δ_{O_3} = Change in mean O_3 between historic and future period, Δ_{PM} = Change in mean $PM_{2.5}$ between historic and future period, and EED = Estimated change in number of ED visits between historic and future period. Credible intervals were calculated by taking the 5^{th} and 95^{th} percentiles of iterations of the EED estimates. The EED estimate iterations were calculated by using all 3,000 iterations of Δ and simulating 3,000 β_{O3} and β_{PM} vectors from a multivariate Gaussian distribution:

$$\begin{pmatrix} \beta_{O_3} \\ \beta_{PM} \end{pmatrix} \sim N \left[\begin{pmatrix} \hat{\beta}_{O_3} \\ \hat{\beta}_{PM} \end{pmatrix}, \begin{pmatrix} \widehat{var(\beta_{O_3})} & \widehat{cov(\beta_{O_3}, \beta_{PM})} \\ \widehat{cov(\beta_{O_3}, \beta_{PM})} & \widehat{var(\beta_{PM})} \end{pmatrix} \right]$$

4.5.3 Results

The mean observed Pearson's correlation between O_3 and $PM_{2.5}$ at each linked monitoring site was 0.588. After spatial bivariate calibration the mean Pearson's correlation between projected O_3 and $PM_{2.5}$ was 0.580 for the Gumbel copula and 0.572 for the Frank copula. In contrast the mean Pearson's correlation between uncalibrated projected O_3 and $PM_{2.5}$ was -0.058. Table 11 shows the correlation values at each monitor location.

Table 11: Pearson's Correlation of Observed and Calibrated Data

Grid Cell	Observed	Gumbel	Frank	Uncalibrated
1	0.745	0.740	0.677	-0.117
2	0.688	0.675	0.664	-0.084
3	0.718	0.707	0.704	-0.110
4	0.657	0.660	0.604	-0.101
5	0.683	0.636	0.648	-0.021
6	0.493	0.494	0.505	-0.049
7	0.672	0.657	0.669	0.092
8	0.661	0.632	0.646	-0.106
9	0.580	0.583	0.598	-0.031
10	0.496	0.483	0.482	-0.031
11	0.499	0.497	0.504	-0.070
12	0.433	0.463	0.467	-0.043
13	0.581	0.588	0.559	-0.016
14	0.471	0.440	0.441	-0.058
15	0.663	0.633	0.635	-0.099
16	0.550	0.540	0.526	-0.031
17	0.413	0.427	0.396	-0.112
Mean	**0.588**	**0.580**	**0.572**	**-0.058**

For this analysis the spatial smoothing parameter ρ=60. The distance between the two monitoring locations (here, one in central Atlanta and one east of Atlanta) was 74.8 km. Spatial correlations of r=$\{0.05, 0.25, 0.5, 0.75\}$ corresponded to $\rho = \{60, 125, 250, 600\}$. Each monitor pair was left out one at a time and predicted from the other monitor pairs. The resulting estimated O_3 and $PM_{2.5}$ levels were compared to the observed levels using continuous ranked probability score (CRPS). Table 12 shows the results from the sensitivity analysis.

Table 12: Sensitivity analysis of spatial smoothing parameter values (ρ) using continuous ranked probability scores (CRPS) with the lowest CRPS in bold.

	ρ			
	60	125	250	600
Gumbel O_3	**0.098**	0.108	0.105	0.112
Gumbel $PM_{2.5}$	**0.109**	0.111	0.122	0.116
Frank O_3	**0.110**	0.115	0.121	0.113
Frank $PM_{2.5}$	**0.100**	0.104	0.108	0.123

The calibrated spatial climate models generally projected lower levels of mean O_3 than the uncalibrated climate model, although the results varied by location (Figures 30, 31 and 32). The calibrated spatial climate models projected higher levels of mean $PM_{2.5}$ than the uncalibrated climate model across most of the spatial domain (Figures 33, 34 and 35). Results were similar for the two parametric copula forms (Gumbel and Frank) for both pollutants. In Atlanta and Birmingham uncalibrated estimates for both O_3 and $PM_{2.5}$ were less than the historically observed mean concentrations for the cities' metro regions. Calibrated projected future concentrations of O_3 were less than the historic observed concentrations, while calibrated projected future concentrations of $PM_{2.5}$ were higher than historic observed concentrations (Table 13).

Table 13: Projected and Observed Concentrations of Ambient O_3 and $PM_{2.5}$ in Atlanta and Birmingham

	O_3 (ppb)	$PM_{2.5}$ ($\mu g/m^3$)
Atlanta		
Observed Historic	52.12	17.34
Uncalibrated Historic	33.95	8.93
Uncalibrated Future	50.84	4.14
Calibrated Future - Gumbel	47.1	22.07
Calibrated Future - Frank	44.12	22.69
Birmingham		
Observed Historic	46.48	18.43
Uncalibrated Historic	35.91	6.93
Uncalibrated Future	48.75	3.18
Calibrated Future - Gumbel	43.25	18.69
Calibrated Future - Frank	38.81	18.65

Figure 30: Uncalibrated projected mean O_3 (ppb) concentrations, 2055-2059

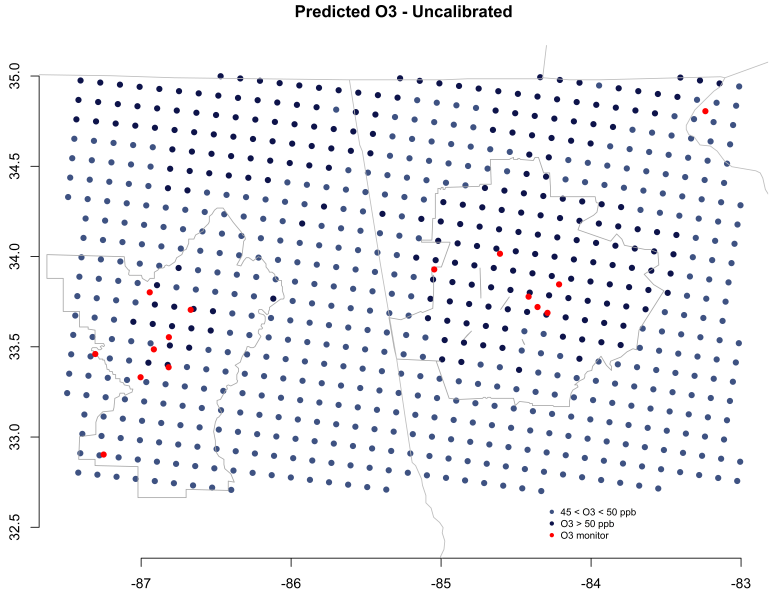


Figure 31: Gumbel calibrated projected mean O_3 (ppb) concentrations, 2055-2059

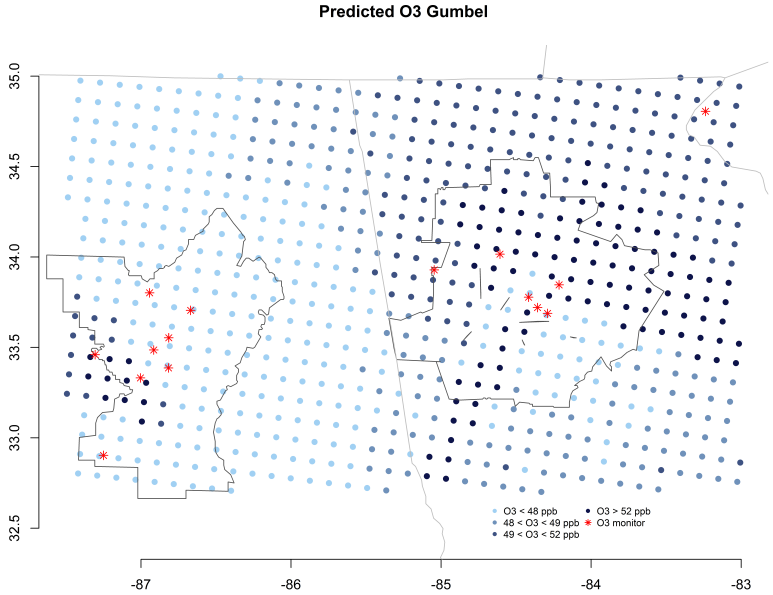


Figure 32: Frank calibrated projected mean O_3 (ppb) concentrations, 2055-2059

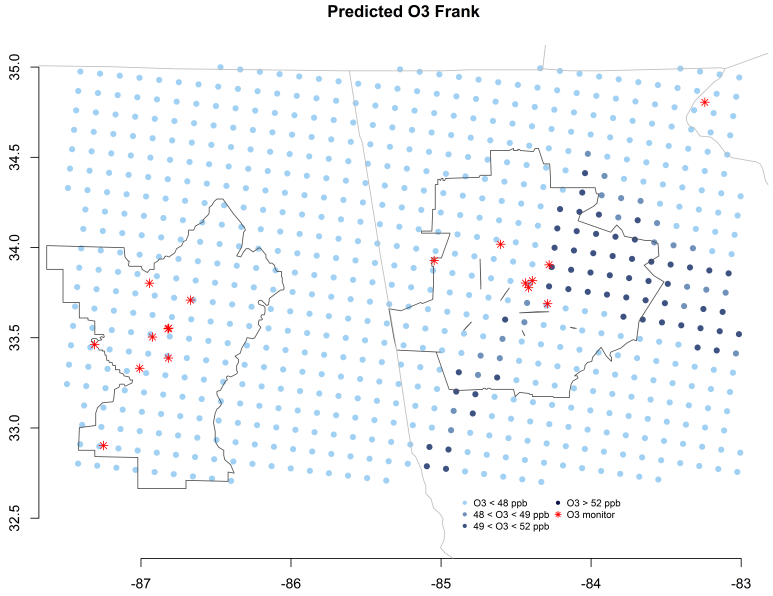


Figure 33: Uncalibrated projected mean $PM_{2.5}$ ($\mu g/m^3$) concentrations, 2055-2059

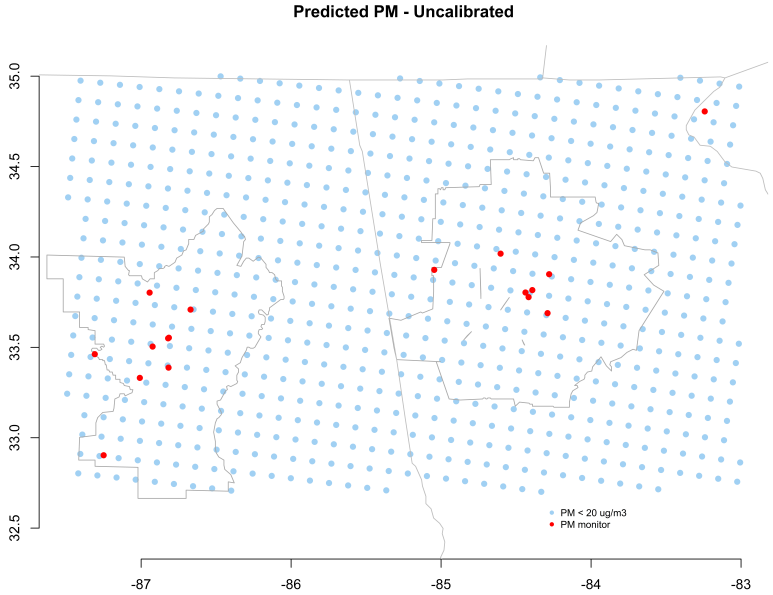


Figure 34: Gumbel calibrated projected mean $PM_{2.5}$ ($\mu g/m^3$) concentrations, 2055-2059

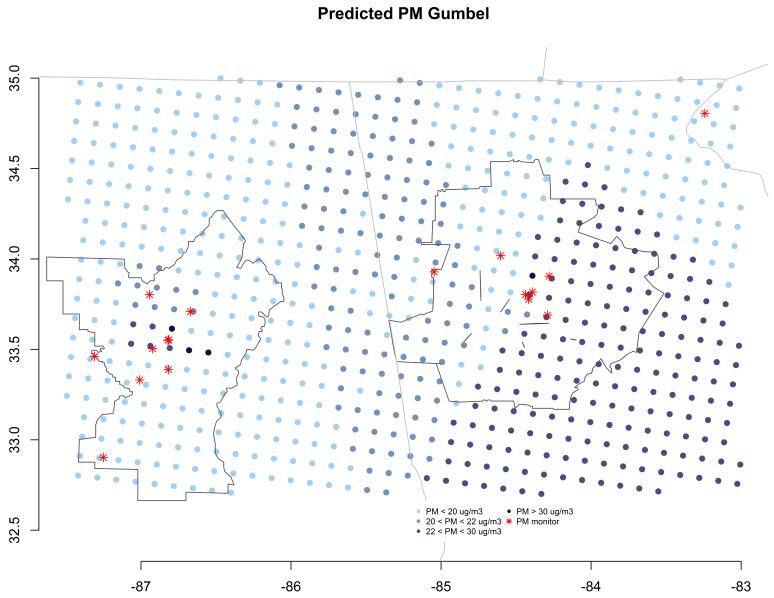
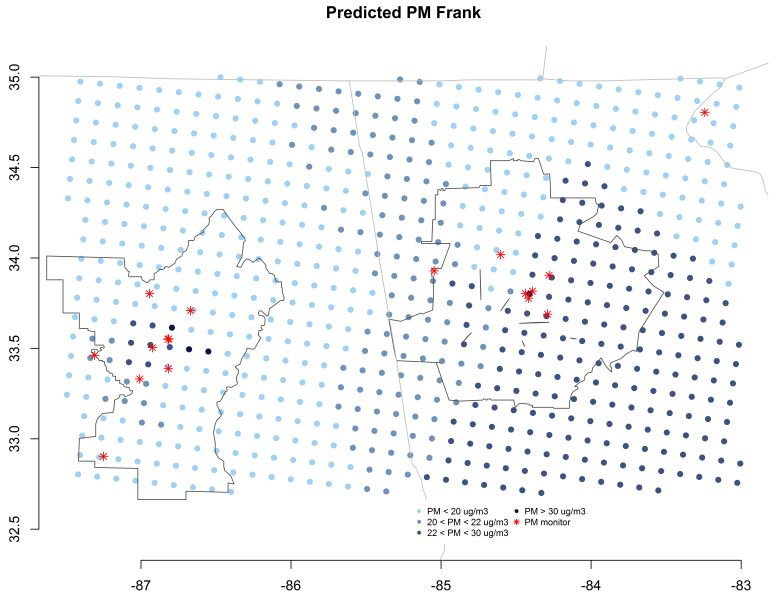


Figure 35: Frank calibrated projected mean $PM_{2.5}$ ($\mu g/m^3$) concentrations, 2055-2059



Baseline annual counts of asthma ED visits were 33,719 in Atlanta and 4,623 in Birmingham (average of total annual counts from 2006-2007). In Atlanta, the regression coefficient representing the association between asthma ED visits and O_3 was 0.0009 (SE: 0.0005) and the regression coefficient representing the association between asthma ED visits and $PM_{2.5}$ was 0.0018 (SE: 0.0008). In Birmingham, there was a null association between O_3 and asthma ED visits with a regression coefficient of 0.0000 (SE: 0.0012) and a positive association between $PM_{2.5}$ and asthma ED visits with a regression coefficient of 0.0040 (SE: 0.0019). The association between $PM_{2.5}$ and asthma ED visits was statistically significant in both Atlanta (p=.03) and Birmingham (p=.04). The covariance between O_3 and $PM_{2.5}$ was -2.4 x 10^{-7} in Atlanta and -1.5 x 10^{-6} in Birmingham.

Using the uncalibrated climate model, the projected average annual change in asthma ED visits for Atlanta was a decrease of 833 ED visits with a 95% credible interval of (-1392, -249). A reduction of 833 asthma ED visits would be a 2.47% reduction in visits compared to baseline counts. The projected average annual change in asthma ED visits for Atlanta using the calibrated climate models was an increase of 141 (-1373, 1531) visits using the Gumbel copula and 85 (-1014, 1119) visits using the Frank copula. These projections represent a 0.42% and a 0.25% annual increase in asthma ED visits, respectively.

In Birmingham, the projected average annual change in asthma ED visits using the uncalibrated model was an decrease of 276 (-443, -106) visits, which would be a 5.97% decrease in ED visits compared to the baseline. The calibrated climate model projected average annual change in asthma ED visits for Birmingham was a increase of 6 (-160, 169) visits using the Gumbel copula and a decrease of 4 (-153, 141) ED visits using the Frank copula (Table 13). These projections represent a 0.13% annual increase in asthma ED visits using the Gumbel copula and a 0.09% annual increase in asthma ED visits using the Frank copula.

Table 14: Projected average changes in annual asthma ED visits between 2001-2004 and 2055-2059 due to changes in concentrations of O_3 and $PM_{2.5}$

	Atlanta		
	EED	95% Interval	% Change
Uncalibrated	-833	(-1392, -249)	-2.47
Gumbel			
Combined Uncertainty	141	(-1373, 1531)	0.42
Health Association Uncertainty	135	(-173, 454)	0.40
Projection Uncertainty	146	(-1293, 1482)	0.43
Frank			
Combined Uncertainty	85	(-1014, 1119)	1.84
Health Association Uncertainty	85	(-311, 491)	1.84
Projection Uncertainty	85	(-944, 937)	1.84
	Birmingham		
	EED	95% Interval	% Change
Uncalibrated	-276	(-443, -106)	-0.82
Gumbel			
Combined Uncertainty	6	(-160, 169)	0.02
Health Association Uncertainty	5	(-25, 37)	0.01
Projection Uncertainty	6	(-146, 137)	0.02
Frank			
Combined Uncertainty	4	(-153, 141)	0.086
Health Association Uncertainty	4	(-65, 75)	0.086
Projection Uncertainty	4	(-122, 101)	0.086

Table 14 shows the difference between the projected levels and the historically observed mean levels by deciles for both O_3 and $PM_{2.5}$. In Atlanta, higher deciles show larger reductions in O_3 for the uncalibrated projections and both copula forms. For $PM_{2.5}$ in Atlanta, the uncalibrated projection has large reductions in the higher deciles, while the two copula forms project an increase in the concentration of $PM_{2.5}$ in all the deciles except the 90th. In Birmingham, the Gumbel copula projects large decreases in O_3 levels at all deciles, and the Frank copula also projects decreases in O_3 levels at deciles above the 30th percentile. The uncalibrated projection has increases in O_3 levels in the lower deciles and increases in O_3 levels starting at the 60th percentile. Similar to Atlanta, the uncalibrated

projection in Birmingham projects a reduction in $PM_{2.5}$ levels at all decile levels. The two copula forms project an increase in $PM_{2.5}$ levels in the lower deciles and a decrease in $PM_{2.5}$ levels in the higher deciles.

Table 15: Change in pollutant concentrations by quantile levels

Atlanta						
		Δ O3			Δ PM	
Percentile	Gumbel	Frank	Uncalibrated	Gumbel	Frank	Uncalibrated
10%	14	12	21	11	10	-3
20%	8	6	15	8	8	-7
30%	4	1	9	7	8	-9
40%	-1	-3	5	6	7	-11
50%	-5	-8	-1	6	6	-12
60%	-8	-12	-6	5	5	-14
70%	-11	-13	-11	4	4	-17
80%	-16	-19	-16	1	2	-20
90%	-21	-25	-23	-3	-1	-24
Mean	-5	-8	-1	5	5	-13

Birmingham						
		Δ O3			Δ PM	
Percentile	Gumbel	Frank	Uncalibrated	Gumbel	Frank	Uncalibrated
10%	-11	8	20	7	7	-6
20%	-16	4	15	5	5	-8
30%	-20	-1	11	4	4	-10
40%	-24	-5	6	3	2	-11
50%	-26	-7	4	2	1	-13
60%	-31	-10	-1	1	-1	-15
70%	-35	-13	-4	-2	-3	-18
80%	-40	-19	-10	-6	-6	-22
90%	-46	-25	-17	-10	-9	-27
Mean	-28	-8	2	1	1	-15

4.5.4 Discussion

The uncalibrated climate model projected higher concentrations of O_3 in the future period (2055-2059) compared to the calibrated climate model projections. The climate model in the historic period (2001-2004) projected lower concentrations of O_3 in the hindcast than were observed at the monitors. Therefore, we would assume the mean projected calibrated concentrations of O_3 would be higher than the uncalibrated projections. This unexpected result may be explained by the fact that the calibration function may not be accurately calibrating the upper tails of the distribution. With this somewhat small observed dataset (only four years of data and 17 monitor pairs), fewer knots could be used on the I-spline, which may have resulted in modeled quantile functions that did not accurately capture the upper tails. There was specifically a lack of data in the upper levels of the quantile functions of the observed data for O_3, making them particularly difficult to accurately model. As a result, the projected calibrated concentrations of O_3 may be attenuated toward the mean. Overall, the calibrated projections of O_3 were higher than the historic raw O_3 concentrations. However, in the urban centers of Atlanta and Birmingham, the mean projected O_3 concentrations were less than the observed O_3 concentrations.

During the hindcast, the uncalibrated climate model simulated more days with low levels of $PM_{2.5}$ and fewer days with high levels of $PM_{2.5}$, resulting projected $PM_{2.5}$ concentrations that were less than observed concentrations. This bias is corrected in the calibration step, which leads to calibrated climate models that project higher levels of $PM_{2.5}$ than observed concentrations.

The assumptions built into the uncalibrated WRF-CMAQ climate model are important in the interpretation of the $PM_{2.5}$ results. The WRF-CMAQ model assumes by 2055 there will be a 60% reduction in the emissions that contribute to $PM_{2.5}$. This assumption explains the low $PM_{2.5}$ values in the uncalibrated future climate model (2055-2059). When the downward bias from the hindcast is corrected in the future climate model, the calibrated climate model projections are

higher than both the uncalibrated hindcast and the observed values. This result highlights the major assumption of the calibration method, which is that the bias between the hindcast and the observed values will apply to the climate model in the future period.

The projected average changes in annual asthma ED visits between 2001-2004 and 2055-2059 present projected changes due only to changes in concentrations of O_3 and $PM_{2.5}$ under the assumptions of the WRF-CMAQ model. Changes in population growth, asthma incidence, and other non-pollutant exacerbators of asthma are not considered. Thus, the projected changes in annual asthma ED visits assume all conditions remain in stasis with the historic time period except the concentrations of O_3 and $PM_{2.5}$.

The uncertainty around the calibrated projected change in annual asthma ED visits encompasses the uncertainty around the estimated effect the air pollutant has on asthma ED visits as well as the uncertainty around the change in pollutant concentrations between the historic and future time periods. The uncalibrated climate model is deterministic, therefore we cannot obtain a measure of uncertainty around the uncalibrated change in pollutant concentrations between the historic and future time periods.

This analysis has several limitations. First, any work with climate models is dependent on the assumptions and accuracy of that climate model. Interpretations of results must take these assumptions into account. Second, the calibration of the upper tails of the distributions was limited by a lack of observed data. Not only were we limited to 17 monitoring sites, the number of days with very high ozone at those sites was small, making it difficult to model. Third, this analysis assumes the observed effect air pollution had on asthma in the historic period (2002-2008) in Atlanta and Birmingham will be the same effect in the future period (2055-2059). Fourth, the spatial calibration and projection method assumes the bias of climate model grid cells projection is unlinked to observed data (i.e., no monitor within 12

km by 12 km grid) can be spatially kriged using the nearby observed monitoring sites data and linked climate model projections.

5 Concluding Remarks

Climate change has the potential to impact human health in a myriad of ways, from direct impacts such as heat-related illness to more complex impacts such as forced migrations due to environmental degradation. As interest in the impacts of climate change increase, both scientists and policymakers know accurate predictions of local impacts are needed to justify the social and economic costs of combating climate change. This dissertation illustrates one method for improving the accuracy of climate model projections and estimates the impact a changing climate may have on asthma. We first assessed the association between ambient ozone (O_3) and fine particulate matter ($PM_{2.5}$) and asthma in the metro Atlanta area. We found both O_3 and $PM_{2.5}$ were associated with an increase in asthma ED visits. We next developed a bivariate calibration method to improve the accuracy of climate model projections. More accurate climate model projections should lead to more accurate estimations of future health impacts. We applied our bivariate calibration method to projections of air quality on a spatial field, which allowed us to obtain calibrated projections for two cities in the southeastern United States: Atlanta, GA and Birmingham, AL. We subsequently use these calibrated air quality levels to project future asthma ED visits. These results provide information on one potential impact of climate change on human health.

In future work we may explore expanding the bivariate calibration method to accommodate additional variables. For example, ozone and $PM_{2.5}$ concentrations are correlated with temperature and the bias correction may benefit from incorporating all three into a joint calibration. This extension would utilize copulas that allow for more than two variables (e.g., the multivariate Frank). In this research we used the Gumbel and Frank copulas as they have been used in previous climate research. Other parametric forms of copulas could also be examined.

References

[1] IPCC. IPCC Fourth Assessment Report: Climate Change 2007, 2007.

[2] National Oceanic and Atmospheric Administration. NOAA Celebrates 200 Years of Science, Service, and Stewardship, 2007.

[3] Centers for Disease Control and Prevention. Climate Effects on Health, 2015.

[4] US Environmental Protection Agency. Basic Information — Particulate Matter. 2015.

[5] US Environmental Protection Agency. What is Ozone? 2015.

[6] US Environmental Protection Agency. Health Effects of Ozone in Patients with Asthma, 2015.

[7] IPCC. Climate Change 2007: Historical Overview of Climate Change Science. Technical report, 2007.

[8] IPCC. Climate Change 2014 Synthesis Report: Summary for Policymakers. Technical report, 2014.

[9] IPCC. IPCC Fourth Assessment Report, WG1, Figure 1.3. Technical report, 2007.

[10] World Meteorological Organization. Climate Models, 2015.

[11] John P. Dawson, Pavan N. Racherla, Barry H. Lynn, Peter J. Adams, and Spyros N. Pandis. Impacts of climate change on regional and urban air quality in the eastern United States: Role of meteorology, 2009.

[12] US Environmental Protection Agency. Health — Particulate Matter. 2015.

[13] Talat Islam, W James Gauderman, Kiros Berhane, Rob McConnell, Ed Avol, John M Peters, and Frank D Gilliland. Relationship between air pollution, lung function and asthma in adolescents. *Thorax*, 62(11):957–963, 2007.

[14] Howard H Chang, Jingwen Zhou, and Montserrat Fuentes. Impact of climate change on ambient ozone level and mortality in southeastern United States. *International journal of environmental research and public health*, 7(7):2866–80, July 2010.

[15] Thomas W Hesterberg, William B Bunn, Roger O McClellan, Ali K Hamade, Christopher M Long, and Peter A Valberg. Critical review of the human data on short-term nitrogen dioxide (NO2) exposures: evidence for NO2 no-effect levels. *Critical reviews in toxicology*, 39(9):743–781, 2009.

[16] K R Smith, A Woodward, D Campbell-Lendrum, D D Chadee, Y Honda, Q Liu, J M Olwoch, B Revich, and R Sauerborn. Human health: impacts, adaptation, and co-benefits. In C B Field, V R Barros, D J Dokken, K J Mach, M D Mastrandrea, T E Bilir, M Chatterjee, K L Ebi, Y O Estrada, R C Genova, B Girma, E S Kissel, A N Levy, S MacCracken, P R Mastrandrea, and L L White, editors, *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change*, pages 709–754. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014.

[17] K. Murazaki and P. Hess. How does climate change contribute to surface ozone change over the United States? *Journal of Geophysical Research: Atmospheres*, 111(5), 2006.

[18] National Institute of Environmental Health Sciences. A Human Health Perspective on Climate Change. Technical report, 2010.

[19] J. R. Balmes. The role of ozone exposure in the epidemiology of asthma, 1993.

[20] Meng Ji, Daniel S Cohan, and Michelle L Bell. Meta-analysis of the association between short-term exposure to ambient ozone and respiratory hospital admissions, 2011.

[21] Kazuhiko Ito, Samantha F De Leon, and Morton Lippmann. Associations between ozone and daily mortality: analysis and meta-analysis. *Epidemiology (Cambridge, Mass.)*, 16(4):446–457, 2005.

[22] Matthew J. Strickland, Lyndsey A. Darrow, Mitchel Klein, W. Dana Flanders, Jeremy A. Sarnat, Lance A. Waller, Stefanie E. Sarnat, James A. Mulholland, and Paige E. Tolbert. Short-term associations between ambient air pollutants and pediatric asthma emergency department visits. *American Journal of Respiratory and Critical Care Medicine*, 182(3):307–316, 2010.

[23] Susan C Anenberg, Larry W Horowitz, Daniel Q Tong, and J Jason West. An estimate of the global burden of anthropogenic ozone and fine particulate matter on premature human mortality using atmospheric modeling. *Environmental health perspectives*, 118(9):1189–95, September 2010.

[24] M A Basha, K B Gross, C J Gwizdala, A H Haidar, and J Popovich. Bronchoalveolar lavage neutrophilia in asthmatic and healthy volunteers after controlled exposure to ozone and filtered purified air. *Chest*, 106(6):1757–65, December 1994.

[25] E Lavigne, P J Villeneuve, and S Cakmak. Air pollution and emergency department visits for asthma in Windsor, Canada. *Can J Public Health*, 103(1):4–8, 2012.

[26] R A Silverman and K Ito. Age-related association of fine particles and ozone with severe acute asthma in New York City. *J Allergy Clin Immunol*, 125(2):367–373 e5, 2010.

[27] J Y Son, J T Lee, Y H Park, and M L Bell. Short-term effects of air pollution on hospital admissions in Korea. *Epidemiology*, 24(4):545–554, 2013.

[28] P J Villeneuve, L Chen, B H Rowe, and F Coates. Outdoor air pollution and emergency department visits for asthma among children and adults: a case-crossover study in northern Alberta, Canada. *Environ Health*, 6:40, 2007.

[29] K Y Wang and T T Chau. An association between air pollution and daily outpatient visits for respiratory disease in a heavy industry area. *PLoS One*, 8(10):e75220, 2013.

[30] L C Vinikoor-Imler, E O Owens, J L Nichols, M Ross, J S Brown, and J D Sacks. Evaluating Potential Response-Modifying Factors for Associations between Ozone and Health Outcomes: A Weight-of-Evidence Approach. *Environ Health Perspect*, 122(11):1166–1176, 2014.

[31] R A Rudd and J E Moorman. Asthma incidence: data from the National Health Interview Survey, 1980-1996. *J Asthma*, 44(1):65–70, 2007.

[32] S Salvi. Health effects of ambient air pollution in children. *Paediatr Respir Rev*, 8(4):275–280, 2007.

[33] L Trasande and G D Thurston. The role of air pollution in asthma and other pediatric morbidities. *J Allergy Clin Immunol*, 115(4):689–699, 2005.

[34] M Mathieu-Nolf. Poisons in the air: a cause of chronic disease in children. *J Toxicol Clin Toxicol*, 40(4):483–491, 2002.

[35] A Winquist, J J Schauer, J R Turner, M Klein, and S E Sarnat. Impact of ambient fine particulate matter carbon measurement methods on observed associations with acute cardiorespiratory morbidity. *J Expo Sci Environ Epidemiol*, 2014.

[36] S E Sarnat, A Winquist, J J Schauer, J R Turner, and J A Sarnat. Fine Particulate Matter Components and Emergency Department Visits for

Cardiovascular and Respiratory Diseases in the St. Louis, Missouri-Illinois, Metropolitan Area. *Environ Health Perspect*, 2015.

[37] Sara E. Grineski, Joan G. Staniswalis, Yanlei Peng, and Carol Atkinson-Palombo. Children's asthma hospitalizations and relative risk due to nitrogen dioxide (NO2): Effect modification by race, ethnicity, and insurance status. *Environmental Research*, 110(2):178–188, 2010.

[38] M J Strickland, M Klein, W D Flanders, H H Chang, J A Mulholland, P E Tolbert, and L A Darrow. Modification of the effect of ambient air pollution on pediatric asthma emergency visits: susceptible subpopulations. *Epidemiology*, 25(6):843–850, 2014.

[39] J K Wendt, E Symanski, T H Stock, W Chan, and X L Du. Association of short-term increases in ambient air pollution and timing of initial asthma diagnosis among Medicaid-enrolled children in a metropolitan area. *Environ Res*, 131:50–58, 2014.

[40] K E Nachman and J D Parker. Exposures to fine particulate air pollution and respiratory outcomes in adults using two national datasets: a cross-sectional study. *Environ Health*, 11:25, 2012.

[41] Deidre Crocker, Clive Brown, Ronald Moolenaar, Jeanne Moorman, Cathy Bailey, David Mannino, and Fernando Holguin. Racial and ethnic disparities in asthma medication usage and health-care utilization: Data from the National Asthma Survey. *Chest*, 136(4):1063–1071, 2009.

[42] H Z Law, E Oraka, and D M Mannino. The role of income in reducing racial and ethnic disparities in emergency room and urgent care center visits for asthma-United States, 2001-2009. *J Asthma*, 48(4):405–413, 2011.

[43] A Roy and J P Wisnivesky. Racial and ethnic differences in the use of environmental control practices among children with asthma. *J Asthma*, 47(5):507–512, 2010.

[44] M Lin, Y Chen, R T Burnett, P J Villeneuve, and D Krewski. The influence of ambient coarse particulate matter on asthma hospitalization in children: case-crossover and time-series analyses. *Environ Health Perspect*, 110(6):575–581, 2002.

[45] G.E.P Box. Science and Statistics, 1976.

[46] Claudia Teutschbein and Jan Seibert. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, 456-457:12–29, 2012.

[47] B J Reich, M Fuentes, and D B Dunson. Bayesian Spatial Quantile Regression. *J Am Stat Assoc*, 106(493):6–20, 2011.

[48] J W Zhou, M Fuentes, and J Davis. Calibration of Numerical Model Output Using Nonparametric Spatial Density Functions. *Journal of Agricultural Biological and Environmental Statistics*, 16(4):531–553, 2011.

[49] H D Bondell, B J Reich, and H Wang. Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838, 2010.

[50] Y C Wu and Y F Liu. Stepwise multiple quantile regression estimation using non-crossing constraints. *Statistics and Its Interface*, 2(3):299–310, 2009.

[51] J. O. Ramsay. [Monotone Regression Splines in Action]: Rejoinder. *Statistical Science*, 3(4):459–461, 1988.

[52] L.O.; et al.; Mearns. The North American Regional Climate Change Assessment Program dataset. *National Center for Atmospheric Research Earth System Grid data portal, Boulder, CO.*, page Data downloaded 4/22/15.

[53] Linda O. Mearns, William Gutowski, Richard Jones, Ruby Leung, Seth McGinnis, Ana Nunes, and Yun Qian. A Regional Climate Change Assessment Program for North America, 2009.

[54] Nebojsa Nakicenovic and Robert Swart. *IPCC Special Report on Emissions Scenarios: A special report of Working Group III of the Intergovernmental Panel on Climate Change*. 2000.

[55] C. Schölzel and P. Friederichs. Multivariate non-normally distributed random variables in climate research – introduction to the copula approach, 2008.

[56] Eike Christian Brechmann and Ulf Schepsmeier. Modeling Dependence with C- and D-Vine Copulas: The R Packatge CDVine. *Journal of Statistical Software*, 52(3):1–27, 2013.

[57] Martyn Clark, Subhrendu Gangopadhyay, Lauren Hay, Balaji Rajagopalan, and Robert Wilby. The Schaake Shuffle: A Method for Reconstructing SpaceTime Variability in Forecasted Precipitation and Temperature Fields. *Journal of Hydrometeorology*, 5(1):243–262, 2004.

[58] Roman Schefzik, Thordis L. Thorarinsdottir, and Tilmann Gneiting. Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling. *Statistical Science*, 28(4):616–640, 2013.

[59] Mathieu Vrac and Petra Friederichs. Multivariate-intervariable, spatial, and temporal-bias correction. *Journal of Climate*, 28(1):218–237, 2015.

[60] Annette Möller, Alex Lenkoski, and Thordis L. Thorarinsdottir. Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society*, 139(673):982–991, 2013.

[61] Elisaveta P Petkova, Daniel A Bader, G Brooke Anderson, Radley M Horton, Kim Knowlton, and Patrick L Kinney. Heat-related mortality in a warming

climate: projections for 12 U.S. cities. *International journal of environmental research and public health*, 11(11):11371–83, 2014.

[62] Shakoor Hajat, Sotiris Vardoulakis, Clare Heaviside, and Bernd Eggen. Climate change effects on human health: projections of temperature-related mortality for the UK during the 2020s, 2050s and 2080s. *Journal of epidemiology and community health*, pages 1–8, 2014.

[63] Francesco Nordio, Antonella Zanobetti, Elena Colicino, Itai Kloog, and Joel Schwartz. Changing patterns of the temperature-mortality association by time and location in the US, and implications for climate change. *Environment international*, 81:80–6, 2015.

[64] Jie Chen, François P. Brissette, and Robert Leconte. Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of Hydrology*, 401(3-4):190–202, 2011.

[65] Fiona Johnson and Ashish Sharma. Accounting for interannual variability: A comparison of options for water resources climate change impact assessments. *Water Resources Research*, 47(4), 2011.

[66] Christian Genest, Bruno Rémillard, and David Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–213, 2009.

[67] Bo Li and Marc G. Genton. Nonparametric Identification of Copula Structures. *Journal of the American Statistical Association*, 108(502):666–675, 2013.

[68] Xiaohong Chen and Yanqin Fan. Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335, 2006.

[69] David Huard, Guillaume Évin, and Anne Catherine Favre. Bayesian copula selection. *Computational Statistics and Data Analysis*, 51:809–822, 2006.

[70] Young Min Kim, Ying Zhou, Yang Gao, Joshua S. Fu, Brent A. Johnson, Cheng Huang, and Yang Liu. Spatially resolved estimation of ozone-related mortality in the United States under two representative concentration pathways (RCPs) and their uncertainty. *Climatic Change*, 128(1-2):71–84, 2014.

[71] Michelle L. Bell, Richard Goldberg, Christian Hogrefe, Patrick L. Kinney, Kim Knowlton, Barry Lynn, Joyce Rosenthal, Cynthia Rosenzweig, and Jonathan A. Patz. Climate change, ambient ozone, and health in 50 US cities. *Climatic Change*, 82(1-2):61–76, 2007.

[72] Howard H. Chang, Jingwen Zhou, and Montserrat Fuentes. Impact of climate change on ambient ozone level and mortality in Southeastern United States. *International Journal of Environmental Research and Public Health*, 7(7):2866–2880, 2010.

[73] J. Elizabeth Jackson, Michael G. Yost, Catherine Karr, Cole Fitzpatrick, Brian K. Lamb, Serena H. Chung, Jack Chen, Jeremy Avise, Roger A. Rosenblatt, and Richard A. Fenske. Public health impacts of climate change in Washington State: Projected mortality risks due to heat events and air pollution. *Climatic Change*, 102(1-2):159–186, 2010.

[74] Kim Knowlton, Joyce E. Rosenthal, Christian Hogrefe, Barry Lynn, Stuart Gaffin, Richard Goldberg, Cynthia Rosenzweig, Kevin Civerolo, Jia Yeong Ku, and Patrick L. Kinney. Assessing ozone-related health impacts under a changing climate. *Environmental Health Perspectives*, 112(15):1557–1563, 2004.

[75] Ellen S. Post, Anne Grambsch, Chris Weaver, Philip Morefield, Jin Huang, Lai Yung Leung, Christopher G. Nolte, Peter Adams, Xin Zhong Liang, Jin Hong Zhu, and Hardee Mahoney. Variation in estimated ozone-related health impacts of climate change due to modeling choices and assumptions. *Environmental Health Perspectives*, 120(11):1559–1564, 2012.

[76] Efthimios Tagaris, Kuo-Jen Liao, Anthony J. DeLucia, Leland Deck, Praveen Amar, and Armistead G. Russell. Potential Impact of Climate Change on Air Pollution-Related Human Health Effects. *Environmental Science & Technology*, 43(13):4979–4988, 2009.

[77] Jian Sun, Joshua S Fu, Kan Huang, and Yang Gao. Estimation of future PM2.5- and ozone-related mortality over the continental United States in a changing climate: An application of high-resolution dynamical downscaling technique. *Journal of the Air & Waste Management Association*, 65(5):611–623, 2015.

[78] Y. Gao, J. S. Fu, J. B. Drake, J. F. Lamarque, and Y. Liu. The impact of emission and climate change on ozone in the United States under representative concentration pathways (RCPs). *Atmospheric Chemistry and Physics*, 13(18):9607–9621, 2013.

[79] D. C. Wong, J. Pleim, R. Mathur, F. Binkowski, T. Otte, R. Gilliam, G. Pouliot, A. Xiu, J. O. Young, and D. Kang. WRF-CMAQ two-way coupled system with aerosol feedback: Software development and preliminary results. *Geoscientific Model Development*, 5(2):299–312, 2012.