Estimating the Health Impact of Climate Change with Calibrated Climate Model Output

Jingwen Zhou*[1], Howard H. Chang[2], and Montserrat Fuentes[1]

[1]Department of Statistics, North Carolina State University, Raleigh, NC, 27695

[2]Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, 30322

February 4, 2012

Abstract

Studies on the health impacts of climate change routinely use climate model output as future exposure projection. Uncertainty quantification, usually in the form of sensitivity analysis, has focused predominantly on the variability arise from different emission scenarios or multi-model ensembles. This paper describes a Bayesian spatial quantile regression approach to calibrate climate model output for examining to the risks of future temperature on adverse health outcomes. Specifically, we first estimate the spatial quantile process for climate model output using nonlinear monotonic regression during a historical period. The quantile process is then calibrated using the quantile functions estimated from the observed monitoring data. Our model also down-scales the gridded climate model output to the point-level for projecting future exposure over a specific geographical region. The quantile regression approach is motivated by the need to better characterize the tails of future temperature distribution where the greatest health impacts are likely to occur. We applied the methodology to calibrate temperature projections from a regional climate model for the period 2041 to 2050. Accounting for calibration uncertainty, we calculated the number of of excess deaths attributed to future temperature for three cities in the US state of Alabama.

KEYWORDS: Bayesian spatial quantile regression, climate change, model calibration, health impacts

*Corresponding author. Email address: jzhou3@ncsu.edu

1. INTRODUCTION

Projections from climate models provide state-of-the-art quantitative information on future climate. Recently, climate model output has been utilized to quantify the health impacts of various environmental risk factors due to climate change. However climate model simulations are deterministic, and uncertainty may be present in various modeling stages that attempt to represent the underlying physical processes with numerical models (Knutti 2008). For health impact analysis, uncertainty quantification has focused predominantly on the variability that arise from different emission scenarios or multi-model ensembles, usually in the form of sensitivity analysis (Peng et al. 2011). While the uncertainty associated with climate model output is well recognized, it is difficult to assess its magnitude for any single climate model.

The main objective of this paper is to describe a statistical approach to compare, evaluate, and resolve the differences between climate model output distribution and the observed data distribution. We describe a Bayesian spatial quantile regression model to calibrate climate model output and use daily maximum temperature as the motivating example. Temperature projections have been used extensively in health impacts analysis because of its well-established adverse health effects (Curriero et al. 2002; Anderson and Bell 2009). Moreover, temperature can act as a predictor for other environmental processes such as infectious disease transmission (Ogden et al. 2006; Remais et al. 2008), hydrological dynamics for water quality (John and Rose 2005), or ground-level ozone creation (Knowlton et al. 2004; Bell et al. 2007). By modeling climate model output to reproduce small-scale weather events in a historical period where observations are available, we wish to not only calibrate future model projections, but also incorporate projection uncertainty in the final health impact estimates.

Early approaches to evaluate model performance usually rely on linear regression analysis of observations versus model output that describes the complex and complicated spatio-temporal dependence in environmental data. One common challenge is that data sources often are not available on the same spatial scales. For example, climate model outputs are provided as average values over a grid cell, while observations from monitoring stations are taken at point locations (Eder and Yu 2006). Several approaches have been proposed to fuse the point and gridded out-

puts by either specifying a latent point-level process (Fuentes and Raftery 2005), or by directly modeling the bias between model and observed values as a spatial-temporal process (Berrocal et al. 2010; McMillan et al. 2009).

Another challenge arises from the usual Gaussian assumption in standard linear regression approach which may underestimate the tail probability for climate variables with skewed distributions (Chang et al. 2010). For example, local heat wave is often defined based on daily temperatures exceeding the 95th percentile of its local summer time climatology. Therefore improving the ability to characterize extreme temperature events is of critical importance. To this end, quantile regression is an important tool for characterizing the tail probabilities (Koenker 2005). Specifically, given the data z_1, z_2... z_T, we wish to model the cumulative distribution function (CDF) $F_Z(z) = P(Z \leq z)$ and the probability density $f_Z(z) = F_Z'(z)$. Additionally, for a given $\tau \in [0,1]$, the quantile function $q_Z(\tau)$ is defined as $q_Z(\tau) = F_Z^{-1}(\tau) = \inf\{z : F_Z(z) \geq \tau\}$. In the recent literature, efforts have been made to analyze each quantile level separately (Lavine 1995; Dunson and Taylor 2005; Kozumi and Kobayashi 2011). The calibration technique is accomplished by how quantile functions behave under transformations of random variables. For example, let the τ^{th} quantile process of the observations be $q_Y(\tau)$ and the model outputs be $q_Z(\tau)$. Similarly, denote distribution functions of Y and Z, as F_Y and F_Z respectively. Now consider the calibration function G_τ. In fact, $G_\tau(\cdot)$ is a monotonic function of the quantile functions of model output. For instance, given the quantile level τ, we have the quantile functions $q_Y(\tau)$ and $q_Z(\tau)$ to construct $G_\tau(\cdot)$ as follows:

$$q_Y(\tau) = G_\tau(q_Z(\tau)) \tag{1}$$

In other words, the transformed data $G_\tau(z_1)$, $G_\tau(z_2)$,..., $G_\tau(z_T)$ are identically distributed as Y.

In our approach, we first estimated the quantile process for climate model output using nonlinear monotonic regression during a historical period. Therefore, we do not choose a probability distribution a priori, but model the quantile function directly. One advantage is that we obtain a model for the entire quantile process instead of a fixed quantile level. By considering only the model grid cells with a monitoring location, this allows us to identify a calibration function G_τ to map the climate model quantile process to the quantile functions estimated from the observed

monitoring data. Subsequently, a spatial adjustment of the entire distribution of the model output with respect to the distribution of the monitoring data was specified to downscale the calibration function and their spatial-quantile processes (Zhou et al. In press). Finally, we evaluated the future heat wave excess mortality based on the calibrated NARCCAP data from 2041 to 2050 (see Figure 1).

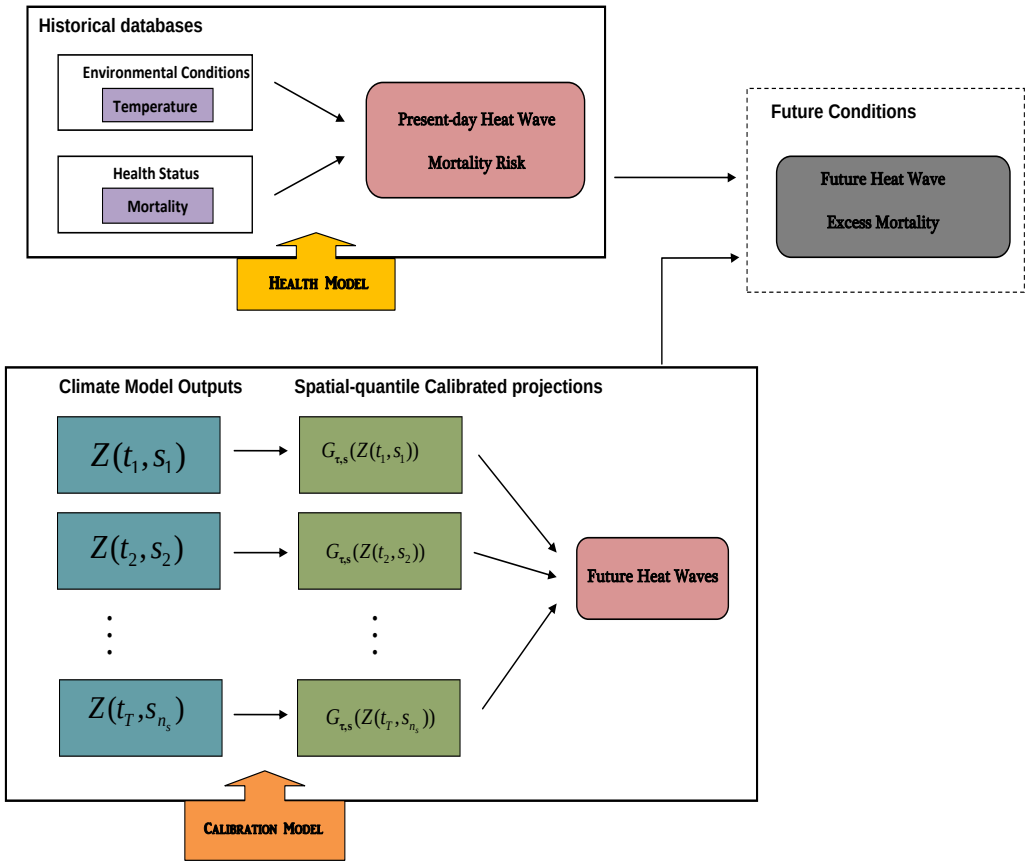

Figure 1: A process chart for our two-stage estimation. We first obtained the calibration model by comparing the original NARCCAP data with the corresponding observations through their underlying spatial-quantile processes in year 2000. Meanwhile, we fitted the health model to investigate the relationship between heat waves and mortality. Finally, we evaluated the future heat wave excess mortality based on the calibrated NARCCAP maximum temperature from 2041 to 2050.

Our ability to quantify the health impacts of future climate has significant implications in guiding policies towards environmental sustainability and in protecting public health. Uncertainty quantification for climate projections is an important component in the risk assessment process. To the authors's knowledge, this is the first health impact analysis to utilize a calibration approach that directly characterizes the uncertainty in the quantiles of future weather. This is particularly useful for examining heat-related impacts where extreme events such as heat waves have been associated with considerable health risks. The main contribution of this paper is to present a statistical framework aimed to not only provide more accurate projections, but to also allow uncertainty propagation in the health impact calculations. We believe the approach described here represents a crucial step towards enhancing the applicability and relevance of the results from climate change and health studies.

The paper is organized as follows. In Section 2, we present the climate model projections and health data for a case study in the US state of Alabama. The southeast region of the US was selected because it has been shown to experience the greatest increase in 20-year return values of daily maximum temperature in the U.S. (Kharin and Zwiers 2000). In Section 3, we describe a Bayesian approach to model the spatial-quantile processes of the climate model output. In Section 4, we outline the modeling steps for risk estimation and how to utilize calibrated model projections to conduct health impact analysis associated with high temperature days and heat waves. We provide the estimation algorithm in Section 5. In Section 6 we present results of the calibration and estimates of the number of attributable deaths due to future temperature extremes. We also compared our approach to calibration using linear regression method in a cross-validation study. Finally, we end with some conclusions and remarks in Section 7.

2. DATA

Future climate projection and health impact analysis were conducted in the state of Alabama, USA for the period 2041 to 2050. We restricted the analysis to the months of May to September, a total of 153 days. Future daily maximum temperatures were obtained from the North American Regional Climate Change Assessment Program (NARCCAP). NARCCAP is an international program to assess uncertainties in regional climate projections using different combinations of regional

climate models and general circulation models. We utilized results from the Canadian Regional Climate Model (CRCM) (Caya et al. 1995) using boundary conditions from the third version of the Coupled Global Climate Model (CGCM3) (Scinocca et al. 2008). NARCCAP provides gridded output with a 50×50km spatial resolution generated under the A2 emissions scenario of the Intergovernmental Panel on Climate Change. The A2 scenario projects large population increases, high carbon dioxide emissions, weak environmental concerns, and regionally oriented economic growth with slower and more heterogeneous technological changes.

We estimated the short-term effects of high temperature on daily mortality for three urban communities in Alabama (Birmingham, Mobile, and Huntsville). Time series of daily maximum temperature, dew-point temperature, and total non-accidental deaths aggregated across the county were obtained from the National Mortality, Morbidity, and Air Pollution Study (NMMAPS) (Peng and Wealty 2004) for the period 1991 to 2000. To perform output calibration, we also obtained NARCCAP data for the year 2000. Observed daily maximum temperature for 13 monitors in Alabama were obtained from the National Oceanic Atmospheric Administration's National Climatic Data Center.

3. SYSTEM CALIBRATION AND SPATIAL QUANTILE PROCESSES

In this section, we will review the spatial-quantile calibration method for obtaining the transformation (calibration) function G_τ in (1) across space. Suppose we have two data sets $Y(t, \boldsymbol{s}_i)$ and $Z(t, \boldsymbol{B}_{\boldsymbol{s}_i})$, where $\boldsymbol{s}_i = (s_{i1}, s_{i2})$ are latitude/longitude coordinates of a temperature monitoring location and $\boldsymbol{B}_{\boldsymbol{s}_i}$ is the associated 50×50 km grid cell in which $\boldsymbol{s}_i$ lies. At each location i=1,2,...,n_s, we have time $t = 1, 2, ..., T$ replications. In order to calibrate the underlying spatial-quantile processes $Q_Z(\tau|\boldsymbol{B}_{\boldsymbol{s}})$ and $Q_Y(\tau|\boldsymbol{s})$, we extend the transformation function G_τ to be space-dependent, and we denote it as $\boldsymbol{G}_{\boldsymbol{\tau},\boldsymbol{s}}$. Then we have:

$$Q_Y(\tau|\boldsymbol{s}) = \boldsymbol{G}_{\boldsymbol{\tau},\boldsymbol{s}}(Q_Z(\tau|\boldsymbol{B}_{\boldsymbol{s}})) \tag{2}$$

Without loss of generality, we rescale both the NARCCAP data (in year 2000 and 2041-2050) and observations (year 2000) such that the range of Z and Y is within $[0, 1]$.

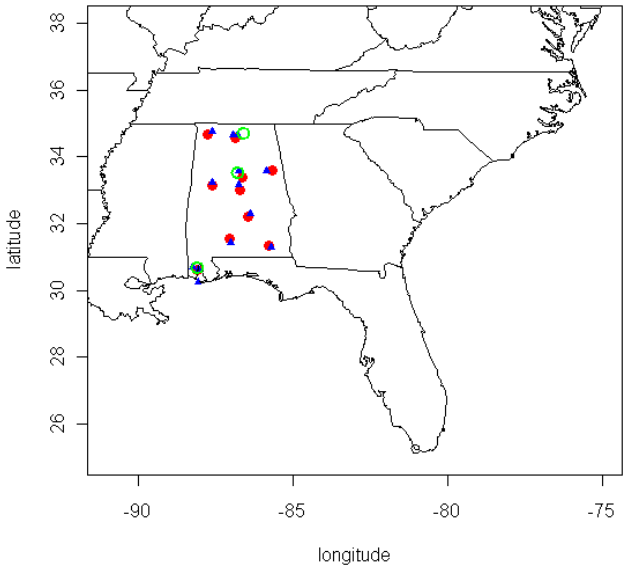

Figure 2: Locations of the monitoring sites (blue triangles), centers of the NARCCAP grid cells (red dots) and NMMAPS communities (green circles) within Alabama.

We first model the spatially-varying quantile function of the NARCCAP model outputs $Q_Z(\tau|\boldsymbol{B_s})$ for $\tau \in [0, 0.01, ..., 1]$. The monotonicity in the coordinate τ for the quantile functions is achieved through the use of the *integrated piecewise polynomial spline* (or I-splines) (Ramsay 1988; Lu and Clarkson 1999). The I-splines construct basis functions using piecewise polynomials, such that upper and lower tails of a quantile function are estimated independently from each other. For a knot sequence $\{\gamma_1, ..., \gamma_{M+h}\}$, M is the number of free parameters that determine the spline function's continuity characteristics, and h is the degree of the piecewise polynomial I_m. For all τ, there exists m such that $\gamma_m \leq \tau < \gamma_{m+1}$. For application to the important case where h=3, let: $I_1^* = \frac{(\tau - \gamma_m)}{(\gamma_{m+2} - \gamma_{m+1})}$; $I_2^* = \frac{(\tau - \gamma_{m+1})^2 - (\gamma_{m+3} - \tau)^2}{(\gamma_{m+3} - \gamma_{m+1})(\gamma_{m+2} - \gamma_{m+1})}$; $I_3^* = \frac{(\gamma_{m+3} - \tau)^3}{(\gamma_{m+3} - \gamma_{m+1})(\gamma_{m+3} - \gamma_m)(\gamma_{m+2} - \gamma_{m+1})}$ - $\frac{(\tau - \gamma_m)^3}{(\gamma_{m+3} - \gamma_m)(\gamma_{m+2} - \gamma_m)(\gamma_{m+2} - \gamma_{m+1})}$. The I-spline I_m will be piecewise cubic, zero for

$\tau < \gamma_m$ and unity for $\tau \geq \gamma_{m+3}$, with the direct expressions:

$$I_m(\tau|\gamma) = \begin{cases} 0, & \text{if } \tau < \gamma_m \\ \dfrac{(\tau-\gamma_m)^3}{(\gamma_{m+1}-\gamma_m)(\gamma_{m+2}-\gamma_m)(\gamma_{m+3}-\gamma_m)}, & \text{if } \gamma_m \leq \tau < \gamma_{m+1} \\ I_1^* + I_2^* + I_3^*, & \text{if } \gamma_{m+1} \leq \tau < \gamma_{m+2} \\ 1 - \dfrac{(\gamma_{m+3}-\tau)^3}{(\gamma_{m+3}-\gamma_{m+2})(\gamma_{m+3}-\gamma_{m+1})(\gamma_{m+3}-\gamma_m)}, & \text{if } \gamma_{m+2} \leq \tau < \gamma_{m+3} \\ 1, & \text{if } \tau \geq \gamma_{m+3}. \end{cases} \quad (3)$$

Because the I-spline is an integral of nonnegative splines, it yields monotone functions $\sum_{m=1}^{M} I_m(\tau)\beta_m(B_s)$ when combined with nonnegative values of the coefficients $\beta_m(B_s)$ (see Figure 3).

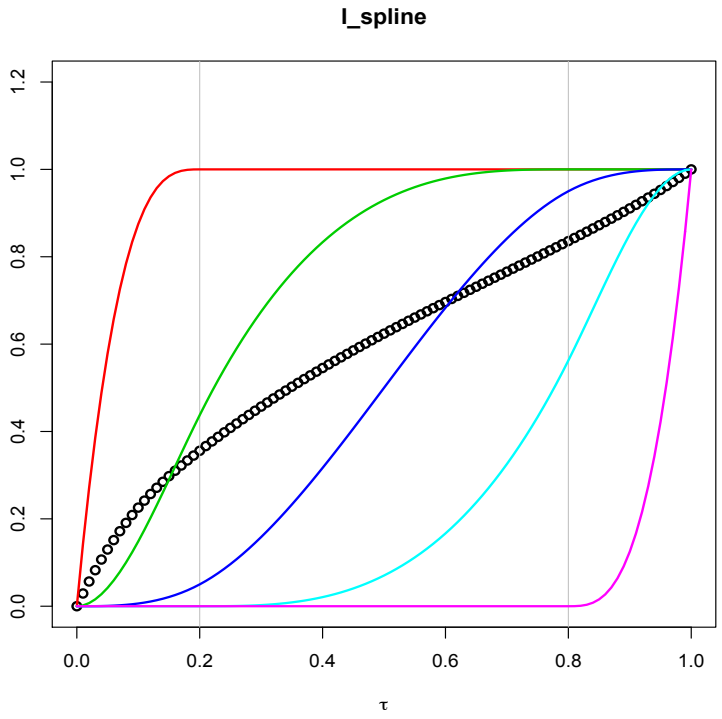

Figure 3: An example of the integrated spline (I-spline). The I-splines are defined on [0, 1] of order h=3 and associated with interior knots 0.2, 0.8. Each I-spline is piecewise cubic, and a monotonic curve is obtained by the linear combination of these M=5 splines with nonnegative coefficients (1.2,2,1.2,1.2,0.4)/6. The I-spline functions ensure non-crossing quantiles with both flexibility and constrains at different percentiles.

Subsequently, the process $Q_Z(\tau|\boldsymbol{B_s})$ is defined as:

$$Q_Z(\tau|\boldsymbol{B_s}) = \boldsymbol{I}(\tau)' \tilde{\boldsymbol{\beta}}_{\boldsymbol{B_s}} = \beta_0(\boldsymbol{B}_s) + \sum_{m=1}^{M} I_m(\tau)\beta_m(\boldsymbol{B}_s) \tag{4}$$

where $\boldsymbol{I}(\tau)'$ denotes the corresponding I-spline basis functions. Because the I-spline is an integral of nonnegative splines, it yields monotonically nondecreasing functions $\sum_{m=1}^{M} I_m(\tau)\beta_m(B_s)$ when combined with nonnegative values of the coefficients $\beta_m(B_s)$.

To ensure the quantile constraints, we introduce latent unconstrained variables $\beta_m^*(\boldsymbol{B}_s)$ and take:

$$\beta_m(\boldsymbol{B}_s) = \begin{cases} \beta_m^*(\boldsymbol{B}_s) & \text{if } \beta_m^*(\boldsymbol{B}_s) \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Therefore a model using $\tilde{\boldsymbol{\beta}}(\boldsymbol{B_s})$ induces via (4) a quantile process of $Q_Z(\tau|\boldsymbol{B}_s)$. Without loss of generality, we choose the knots series within $\gamma_1 = 0$ and $\gamma_{M+h} = 1$. The quantile process thus satisfies the boundary conditions $Q_Z(0|\boldsymbol{B}_s) = \beta_0(\boldsymbol{B}_s)$, and $Q_Z(1|\boldsymbol{B}_s) = \beta_0(\boldsymbol{B}_s) + \sum_{m=1}^{M}\beta_m(\boldsymbol{B}_s)$. In addition, assume $\beta_m^*(\boldsymbol{B}_s)$ have prior $\beta_m^*(\boldsymbol{B}_s) \sim N\left(\mu_{m_{n_s}}, \Sigma^m_{n_s \times n_s}\right)$, with $\Sigma^m_{(\boldsymbol{B_s},\boldsymbol{B_{s'}})} = \sigma^2_{m_{\boldsymbol{B}}} \exp(-||\boldsymbol{s}-\boldsymbol{s'}||/\rho_{m_{\boldsymbol{B}}})$. The full conditional distribution of $\pi(\beta_m(\boldsymbol{B}_s)|Z)$ is then given by $f(Z|\beta_m(\boldsymbol{B}_s), \beta_m^*(\boldsymbol{B}_s))\ \pi(\beta_m(\boldsymbol{B}_s)\ |\beta_m^*(\boldsymbol{B}_s))\ \pi(\beta_m^*(\boldsymbol{B}_s))$, and is obtained using the Metropolis-Hastings algorithm for further calibration.

We then calibrate the NARCCAP spatial-quantile process with the observations in (2) by modeling the observed quantiles of Y as follows:

$$Q_Y(\tau|\boldsymbol{s}) = \boldsymbol{I}\left(\eta_{\boldsymbol{s}}(\tau, \hat{\tilde{\boldsymbol{\beta}}}_{\boldsymbol{B_s}})\right)' \tilde{\boldsymbol{\alpha}}(\boldsymbol{s}) = \alpha_0(\boldsymbol{s}) + \sum_{m=1}^{M} I_m\left(\eta_{\boldsymbol{s}}(\tau, \hat{\tilde{\boldsymbol{\beta}}}_{\boldsymbol{B_s}})\right)\alpha_m(\boldsymbol{s}) \tag{6}$$

where $\eta_{\boldsymbol{s}}(\tau, \tilde{\boldsymbol{\beta}}_{\boldsymbol{B_s}})$ is a monotonic function defined as:

$$\eta_{\boldsymbol{s}}(\tau, \hat{\tilde{\boldsymbol{\beta}}}_{\boldsymbol{B_s}}) = Q_Z(\tau|\boldsymbol{B_s}; \hat{\tilde{\boldsymbol{\beta}}}_{\boldsymbol{B_s}}) \tag{7}$$

Therefore, we first model quantile functions of the model output $\eta_{\boldsymbol{s}}(\tau, \tilde{\boldsymbol{\beta}}_{\boldsymbol{B_s}})$ as a function of τ, then we model the quantile values $Q_Y(\tau|\boldsymbol{s})$ to obtain the calibration function $\boldsymbol{G}_{\boldsymbol{\tau,s}}$ between the two quantile processes (see Formula (2)). $\hat{\tilde{\boldsymbol{\beta}}}_{\boldsymbol{B_s}}$ are sampled from the posterior distribution $\pi(\beta_m(\boldsymbol{B}_s)|Z)$. Note that we mainly focus on $\eta_{\boldsymbol{s}}(\tau, \tilde{\boldsymbol{\beta}}_{\boldsymbol{B_s}}) \in [0,1]$ because Z is standardized to be between 0 and 1. Therefore, for non-negative $\alpha_m(\boldsymbol{s})$ values we obtain a monotonic mapping between

the model output and observed quantile functions. Again, we assume $\alpha_m(\boldsymbol{s})$ are spatially-varying coefficients modeled as multivariate mean-zero Gaussian spatial processes $\alpha_0(\boldsymbol{s}), \alpha_1(\boldsymbol{s}), ..., \alpha_M(\boldsymbol{s})$, such that, $\text{Cov}(\alpha_m(\boldsymbol{s}), \alpha_m(\boldsymbol{s'}))=\sigma^2_{\boldsymbol{ms}} \exp(-||\boldsymbol{s}-\boldsymbol{s'}||/\rho_{\boldsymbol{ms}})$ and $\rho_{\boldsymbol{ms}}$ is the spatial decay parameter for the Gaussian process $\alpha_m(\boldsymbol{s})$, $m = 0, 1, ..., M$.

The sample quantiles are bounded by the sample minimum and maximum. In order to model the tail behaviors of the distribution better, we describe a Bayesian mixture approach to recover the density function. It is based on (1) the quantile functions estimated from the data for characterizing the central tendency; and (2) a parametric density function φ_Z for characterizing the tails. In our application, we treat φ_Z as the density function of the generalized Pareto distribution (GPD) (Hosking and Wallis 1987):

$$\varphi_Z(z;\omega) = \begin{cases} \omega(1-\kappa\omega z)^{1/\kappa-1} & \text{if } \kappa \neq 0 \\ \omega \exp(-\omega z) & \text{if } \kappa = 0. \end{cases} \tag{8}$$

We extend the work of Tokdar and Kadane (2011) for a bounded response variable to deal with an unbounded variable. Given a quantile function $q_Z(\tau)$ we define its density as

$$f_Z(z) = \begin{cases} C_L \varphi_Z\left(q_Z(\tau_{z_{(l)}}) - z; \omega_L\right) & \text{if } z < q_Z(\tau_{z_{(l)}}) \\ \dfrac{1}{\dfrac{\partial q_Z(\tau)}{\partial \tau}\Big|_{\tau=\tau_{(z)}}} & \text{if } q_Z(\tau_{z_{(l)}}) \leq z \leq q_Z(\tau_{z_{(u)}}) \\ C_U \varphi_Z\left(z - q_Z(\tau_{z_{(u)}}); \omega_U\right) & \text{if } z > q_Z(\tau_{z_{(u)}}). \end{cases} \tag{9}$$

where $z_{(l)}$ and $z_{(u)}$ are lower and upper cut points that satisfy $q_Z(\tau_{z_{(l)}}) = z_{(l)}$ and $q_Z(\tau_{z_{(u)}}) = z_{(u)}$, respectively. The Newton's Recursion method (Tokdar and Kadane 2011) is used to approximate τ_z. Therefore, equation (9) allows us to evaluate the likelihood given a threshold δ. For example, if $\tau_{z_{(l)}} = \delta$, $\tau_{z_{(u)}} = 1-\delta$ and $\delta = 0.1$, then we evaluate the central 80% density using the equation (4), and the 10% upper/lower tail's density using the generalized Pareto distribution (8).

To ensure a proper likelihood, the constants in equation (9) are given by:

$$C_L = \tau_{z_{(l)}} \quad , \quad \omega_L = \frac{1}{C_L}\left[\frac{\partial q_Z(\tau)}{\partial \tau}\Big|_{\tau=\tau_{z_{(l)}}}\right]^{-1}$$

$$C_U = 1 - \tau_{z_{(u)}} \quad , \quad \omega_U = \frac{1}{C_U}\left[\frac{\partial q_Z(\tau)}{\partial \tau}\Big|_{\tau=\tau_{z_{(u)}}}\right]^{-1}$$

Because $C_L \times \omega_L$ and $C_U \times \omega_U$ are the density evaluated at the point where the generalized Pareto distribution joins the quantile derived distribution (See Figure 4). As a result, ω_L and ω_U ensure that equation (9) is continuous. Finally, technical details of estimating the parameter κ are provided in the Appendix.

4. HEALTH EFFECT ESTIMATION AND IMPACT ANALYSIS

Relative change in the rate of mortality associated with variation in daily maximum temperature was estimated via Poisson regression

$$\log E(y_t^c) = ns(\text{temp}_t^c, 3) + \text{confounders},$$

where $temp_t^c$ and y_t^c are the observed daily maximum temperature and the number of deaths on day t in city c, respectively. We modeled the effects of temperature as a non-linear function using natural cubic splines with 3 degrees of freedom denoted by $ns(\text{temp}_t^c, 3)$. Following Anderson and Bell (2011), we also included the following confounders in the model: (1) indicators for day of the week; (2) indicators for city; (3) indicators of age group (under 65, 65 to 75, above 75); (4) city and age-group interactions; (5) smooth function of calendar date $ns(t, 10 \times 3)$; and (6) smooth functions of dew-point temperature for the current day and previous three-day averages using natural cubic splines with 2 degrees of freedom. In our analysis, we did not find significant interactions between the effects of temperature and age-groups. We also assumed the effects of weather are identical across the three cities.

Because the effect of temperature is non-linear, we used the metric by Peng et al. (2011) to summarize the overall relative risk associated with high temperature and heat waves. The relative risks (RR) were calculated separately for each city as follows

$$\hat{RR}^c = \frac{\frac{1}{n_1^c}\sum_t \exp\{ns(\text{temp}_t^c, 3)\} \times I^c_{\text{risk},t}}{\frac{1}{n_0^c}\sum_t \exp\{ns(\text{temp}_t^c, 3)\} \times I^c_{\text{ref},t}}$$

where $I^c_{\text{risk},t}$ and $I^c_{\text{ref},t}$ are indicators for whether the daily maximum temperature on day t represents an *at-risk* day or a *reference* day. Similarly n^c_1 and n^c_0 denote the total number of at-risk and reference days in city c. Therefore the above metric describes the ratio of average attributable mortality between the at-risk and reference days. For example, at-risk days are defined as heat wave days and reference days are non-heat wave days.

To estimate the overall effect of high temperature at different thresholds, we set the reference temperature to be less than 30°C, the median value between 1991 to 2000. We then calculated the RR associated with daily maximum temperatures exceeding values 32.5°C, 34°C, 35.5°C, or 37°C (approximately the 90^{th}, 95^{th}, 97.5^{th} and 99^{th} percentile, respectively). There exists no universal definition of heat waves and various metrics have been proposed (Karl and Knight 1997; Huth et al. 2000). We used a similar definition from a recent national population study (Anderson and Bell 2011) to estimate the additional adverse effect of heat waves by creating indicators for heat wave days. Specifically, we identified heat waves as a period of ≥ 2 consecutive days with daily maximum temperature higher than the city-specific quantile thresholds. We examined heat waves defined based on the 90^{th}, 95^{th}, 97.5^{th}, and 99^{th} quantiles. For the heat wave RR, reference days were defined as all non-heat wave days.

Finally, we quantified current and future health impacts of temperature by calculating the expected number of excess deaths attributable to high temperature days as

$$\sum_c N_c \times (RR^c - 1) \times M_c,$$

where N_c is the expected daily mortality in city c, and M_c is the number of at-risk days over the 10-year period of 1991 to 2000, or 2041 to 2050. We estimated N_c by the mean daily mortality across all days under 30°C for each city.

5. ESTIMATION

Estimation was carried out in a Bayesian framework via Markov chain Monte Carlo (MCMC) in two stages (see Figure 1). In the first stage, we estimated the calibration parameters using the

Metropolis-Hastings algorithm with data in year 2000. To sample from $\pi(\tilde{\boldsymbol{\beta}}, \rho_{m_{\boldsymbol{B}}}, \sigma^2_{m_{\boldsymbol{B}}}|Z)$ and $\pi(\tilde{\boldsymbol{\alpha}}, \rho_{ms}, \sigma^2_{m_{\boldsymbol{s}}}|Y, Z)$ sequentially, the associated density function was approximated from quantile function $Q_Y(\tau|Z, \boldsymbol{s})$ using a grid of 101 equally-spaced quantile values $\tau_k \in [0, 1]$ and I-splines with interior knots at (0.3, 0.7). We then obtained posterior predictive samples of the calibrated future NARCCAP data based on the nonlinear transformation function $\boldsymbol{G}_{\boldsymbol{\tau},\boldsymbol{s}}$,

$$\boldsymbol{G}_{\boldsymbol{\tau},\boldsymbol{s}}(z(t, \boldsymbol{B}_{\boldsymbol{s}}), \tilde{\boldsymbol{\alpha}}) = \alpha_0 + \sum_{m}^{M} I_m(z(t, \boldsymbol{B}_{\boldsymbol{s}}))\alpha_m, \tag{10}$$

where $\tilde{\boldsymbol{\alpha}}$ are posterior samples from the MCMC. Finally, $\boldsymbol{G}_{\boldsymbol{\tau},\boldsymbol{s}}(z(t, \boldsymbol{B}_{\boldsymbol{s}}), \tilde{\boldsymbol{\alpha}})$ was re-scaled to its original range.

In the second stage, we fitted the health model in a separate MCMC run. Posterior distributions for the relative risks were obtained from the posterior samples of the temperature-mortality dose-response function $ns(\text{temp}^c_t, 3)$. We combined the uncertainty in both climate projection and risk estimation in calculating the number of excess deaths (ED) as follows. For the future period, the number of at-risk days M_c was calculated for each posterior time series of the calibrated NARCCAP daily maximum temperature. We then took an *exposure simulation* approach (Gryparis et al. 2009) where for each posterior sample of the relative risk, a realization of M_c was randomly drawn from its posterior samples. We then pooled the posterior samples of ED and calculated its posterior median and 95% credible intervals.

By modeling the calibration and health models separately, we broke the feedback between the mortality data and the estimated exposures. This *directional* Bayesian approach (Gelman 2004) not only reduces computational burden, but it also avoids the potential unintuitive assumption that health data could provide information on future exposure, for example in a causal pathway. Another benefit of fitting the exposure and health model separately is that often different metrics of exposure or sets of confounders are examined in a sensitivity analysis. For example, in our analysis different definitions of heat wave and extreme temperature were defined, and fitting a joint model repeatedly is computationally expensive.

6. APPLICATION: CALIBRATION OF ALABAMA MAXIMUM TEMPERATURE

We first compared the uncalibrated NARCCAP output and the observations in terms of their entire distribution for year 2000. At τ= 0.05, 0.5, and 0.95, we calculated the empirical root mean integrated squared error as:

$$RMISE = \left[n_s^{-1} \sum_{i=1}^{n_s} (\hat{Q}_Z(\tau|\boldsymbol{B}_{\boldsymbol{s}_i}) - \hat{Q}_Y(\tau|\boldsymbol{s}_i))^2 \right]^{1/2}$$

For the uncalibrated model output, the $RMISE$ at the 5^{th} quantile was equal to 3.93 (1.35 for the 50^{th} percentile, and 1.72 for the 95^{th} percentile). After calibration, the corresponding $RMISE$ at τ= 0.05, 0.5, and 0.95 were 0.66, 0.37, and 0.32, respectively..

The improvement in RMISE can be visually depicted in Figure 5. We plot the estimated quantile functions for the observed temperature at the three cities, as well as the quantile functions estimated with the uncalibrated NARCCAP model outputs and the calibrated outputs using our Bayesian algorithm. We also considered a naive calibration approach using linear regression (LM). This demonstrates the large discrepancies between the distributions of the NARCCAP output and the observed data at their lower tails. Moreover, Figure 5 highlights the ineffectiveness of linear model to characterize the tail behaviors. Across monitors, we also found that the NARCCAP data present a different spatial pattern from the observed spatial structure, indicating that the process-based numerical models are biased toward "lower tails" and may not capture the spatial correlations existed in the monitoring data.

Because our objective is to perform calibration of future climate variables, we conducted a 5-fold cross-validation study to examine the performance of out-of-sample prediction. At each site, we first randomly split the original daily maximum temperature data (T = 153) into 5 subsamples. Then we retained a single subsample as the validation set for testing our spatial-quantile calibration, and the remaining 4 subsamples were used as training data. This process was repeated 5 times, with each of the 5 subsamples used once as the validation data. We calculated $\hat{Q}_Y(\tau|\boldsymbol{s}_i)$ (the predicted quantiles of the observations at location $\boldsymbol{s}_i$), $\tilde{Q}_{S_Z}(\tau|\boldsymbol{s}_i)$ (the averaged 5-fold quantiles of the Bayesian calibrated data), and $\tilde{Q}_{L_Z}(\tau|\boldsymbol{s}_i)$ (the empirical quantile values calculated using the fitted values obtained from a linear regression of $Y_{\boldsymbol{s}_i}$ on $Z_{\boldsymbol{B}_{\boldsymbol{s}_i}}$ and a

Normality residual assumption), at $\tau \in [0.01, 0.97]$ and location $\boldsymbol{s}_i$. The root mean squared error $RMSE(\hat{Q}, \tilde{q}|\boldsymbol{s}_i) = [K^{-1}\sum_{k=1}^{K}(\hat{Q}(\tau_k, \boldsymbol{s}_i) - \tilde{q}(\tau_k, \boldsymbol{s}_i))^2]^{1/2}$ is calculated for both the linear regression method and our Bayesian approach at each location $\boldsymbol{s}_i$. To compare the predictive performance of different methods, we used the difference root mean squared error defined as

$$DRMSE = n_s^{-1}\sum_{i=1}^{n_s}\frac{RMSE(\hat{Q}_Y(\tau|\boldsymbol{s}_i), \tilde{Q}_{S_Z}(\tau|\boldsymbol{s}_i)) - RMSE(\hat{Q}_Y(\tau|\boldsymbol{s}_i), \tilde{Q}_{L_Z}(\tau|\boldsymbol{s}_i))}{RMSE(\hat{Q}_Y(\tau|\boldsymbol{s}_i), \tilde{Q}_{L_Z}(\tau|\boldsymbol{s}_i))}.$$

The DRMSE between the linear regression method and the quantile calibration method range from -81.6% to -55.3% across monitor. The average DRMSE over monitors was -68.7%, indicating a 68.7% decrease in RMSE when quantile regression was used compared to linear regression. In addition, the quantile regression achieves a 74.9% reduction in RMSE compared to the uncalibrated NARCCAP data. As a result, successfully calibrating the entire distribution of model outputs would be necessary for the model-based projections in the future.

Future projections of model outputs during the period 2041 to 2050 were calibrated similarly as in year 2000. We first scaled the NARCCAP data between 0 and 1, which allows an implementation of the transformation function $\boldsymbol{G}_{\boldsymbol{\tau},\boldsymbol{s}}$ in (10), as well as the associated posterior samples to be developed in the future. Figure 6 provides the integrated quantile curves for the uncalibrated and calibrated NARCCAP data over the decade 2041 to 2050. For comparison purposes, we added the historical values to the same graph at the three targeted cities. In the US state of Alabama, raw climate model outputs suggest that future warming at the 95^{th} percentile was equal to $5.47°C$ ($3.04°C$ for the 50^{th} percentile and a decreasing $1.34°C$ for the 5^{th} quantile). However, after calibration, the increased maximum temperature at τ= 0.05, 0.5, and 0.95 were $1.73°C$, $2.90°C$ and $4.27°C$, respectively. Note that the largest discrepancies mostly occurred at the lower tail of the distribution.

Table 1 gives the estimated number of excess deaths due to high temperature and heat waves for the historical period (1991-2000) and the future period (2041 - 2050). The estimates are presented as annual average across the three study cities in Alabama. The health impact of future temperature were based on calibrated climate model output where the 95% credible intervals reflect both uncertainty in the temperature exposure-response relationship, and the uncertainty

in model output. The average number of deaths on days with maximum temperature under 30 °C was approximately 29.8 per day (16.2 for Birmingham, 8.6 for Mobile, and 4.9 for Huntsville). Similarly average mortality rates were observed on non-heat wave days defined with the 95% quantile threshold. Assuming similar future baseline mortality rates, we found a considerable increase mortality attributed to high temperature. This is dominated by increases in the number of days with extreme temperature. For example, Table 2 gives the observed and projected average number of high temperature days and heat wave days per year in Birmingham Alabama. Similar patterns were observed in the other two cities.

7. DISCUSSION

Uncertainty in climate projections can arise from various sources. We described a statistical calibration approach that models the distributional discrepancy between model outputs and historical observations. This allows us to calibrate future projections, as well as propagating its uncertainty in a health impact analysis. This differs from previous studies where future exposures from climate model outputs are assumed to be deterministic. We chose daily maximum temperature as the motivating example; however the proposed approach can be applied to other weather variables such as precipitation, solar radiation, or cloud cover that may be associated with adverse health outcomes through various pathways. In our study region of Alabama, we found the disagreement between observed and modeled temperatures to be the greatest at the lower tail of the distribution. Consequently, the estimated health impacts due to future extreme temperatures were similar between the calibrated and the original model projections. This result may to vary across locations, especially if other weather variables or climate models are examined, and should be systematically explored in future analysis.

Our health impact calculations also suffer from many common limitations in estimating future disease burden due to climate change. For example, we did not account for the expected changes in population structure, behaviors, size, or other factors that may influence the underlying health status of the population. Moreover, studies have shown that the temperature-mortality relationships exhibit heterogeneity across different NMMAPS regions (Anderson and Bell 2011; Curriero

et al. 2002). While the attributable deaths calculation was not based on future population size, the IPCC scenario used for climate modeling includes population growth. We also recognize that our findings may be dependent on the chosen climate model and emission scenario. It is straightforward to conduct similar analysis with calibrated projections from different climate models following a recent exemplary effort by Peng et al. (2011). Nonetheless, we report a significant increase in mortality attributable to future temperature extremes even in a region that is characterized by hot and humid summer seasons.

NARCCAP conducted 12 sets of climate projections with different combination of RCM-GCM pairs. RCMs are downscaled version of GCM for studying climate evolution at a finer spatial resolution driven by initial values and boundary conditions from the GCM outputs. Recent studies have shown that the choice of both GCM and RCM can influence projections across space and time (Kaufman and Sain 2010). Various methods have been proposed for assessing intermodal variability (Jun et al. 2008; Smith et al. 2009) and for combining model outputs (Sain et al. 2011). We only utilized one set of the NARCCAP experiment and applying calibration across multiple models raises several interesting research questions. For example, do calibrated projections exhibit less intermodal variation? Also, can the degree of discrepancy between modeled and observed values serve as a guide for determining averaging weights for different projections?

APPENDIX A. FITTING THE GENERALIZED PARETO DISTRIBUTION TO THE TAILS

When $\kappa = 0$, the GPD distribution defined as (8) reduces to an exponential distribution with mean $1/\omega$. Now we concentrate on the more difficult case where $\kappa \neq 0$. The idea is to make full use of the information contained in both central tendency and tails. Note that all the discussions below are aimed at analyzing upper tails, and it can be similarly applied on the lower tails.

If the likelihood is given by (9), we first calculate the derivative given μ_U as

$$\frac{\partial f_Z(z)}{\partial z} \mid_{z=\mu_U} = C_U\omega_U(\frac{1}{\kappa} - 1)(-\kappa\omega_U)(1 - \kappa\omega_U(z - \mu_U)) \mid_{z=\mu_U} = C_U\omega_U^2(\kappa - 1) \tag{A.1}$$

Because $\frac{\partial f_Z(z)}{\partial z} \mid_{z=\mu_U} \leqslant 0$, thus $\kappa \leqslant 1$.

Based on the likelihood, we transform the overall order statistics $z_{(u+1)}$, $z_{(u+2)}$,...,$z_{(T)}$ to the partial order statistics on the upper tail as: $z^*_{(1:q)}$, $z^*_{(2:q)}$,..., $z^*_{(q:q)}$, where $z^*_{(i:q)} = z_{(u+i)} - \mu_U$,

$i = 1, 2, ..., q$, and $q = T - u$. Such order statistics are evaluated by first equating the CDF at the observed order statistics to their corresponding percentile values:

$$F(z^*_{(i:q)}; \kappa) = p_{i:q} \tag{A.2}$$

where $p_{i:q} = \frac{1}{q}$ if $q \neq 1$. Substituting the corresponding CDF of (8) to (A.2) we have:

$$\frac{1}{\kappa} \ln(1 - \kappa \omega_U z^*_{(i:q)}) = -\Omega_i \tag{A.3}$$

where $\Omega_i = -\ln(1 - p_{i:q}) > 0$. Consider the following function of κ:

$$h(\kappa) = \ln(1 - \kappa \omega_U z^*_{(i:q)}) + \kappa \Omega_i \tag{A.4}$$

which is defined in $(-\infty, \min(1, \frac{1}{\omega_U z^*_{(i:q)}})]$. Then (A.3) is equal to:

$$h(\kappa) = 0 \tag{A.5}$$

Subsequently, we use the resulting equation (A.5) as a basis for obtaining initial parameters for κ. By taking the derivative of $h(\kappa)$, we have:

$$\frac{\partial h(\kappa)}{\kappa} = -\frac{\omega_U z^*_{(i:q)}}{1 - \kappa \omega_U z^*_{(i:q)}} + \Omega_i \tag{A.6}$$

Let $\frac{\partial h(\kappa)}{\kappa} = 0$, we have $\kappa^* = \frac{1}{\omega_U z^*_{(i:q)}} - \frac{1}{\Omega_i}$. In general, the function $h(\kappa)$ is increasing when $\kappa \in (-\infty, \kappa^*)$ and decreasing when $\kappa \in (\kappa^*, \frac{1}{\omega_U z^*_{(i:q)}})$. Considering κ's constraints, $h(\kappa)$ has the following properties:

1. $\kappa^* = 0$ if $\Omega_i = \omega_U z^*_{(i:q)}$. Therefore the function $h(\kappa)$ is increasing when $\kappa \in (-\infty, 0)$ and decreasing when $\kappa \in (0, \min(1, \frac{1}{\omega_U z^*_{(i:q)}}))$. Additionally, the solution for (A.5) is $\hat{\kappa} = 0$.

2. $\kappa^* > 0$ if $d\Omega_i = \omega_U z^*_{(i:q)}$, where $0 < d < 1$. Then $\kappa^* = (\frac{1}{d} - 1)\frac{1}{\Omega_i} > 0$ and $h(\kappa^*) > 0$. The solutions of (A.5) are: (1) $\kappa^* \geq 1 \Rightarrow \hat{\kappa} = 0$; (2) for $\kappa^* < 1$ and $\frac{1}{d\Omega_i} > 1$, if $h(1) < 0$, we use the bisection method to determine the solution in $(\kappa^*, 1)$, otherwise $\kappa = 0$; (3) if $\frac{1}{d\Omega_i} < 1$, we use the bisection method to determine the solution in $(\kappa^*, \frac{1}{d\Omega_i})$.

3. $\kappa^* < 0$ if $d\Omega_i = \omega_U z^*_{(i:q)}$, where $d > 1$. Given a negative number of large magnitude Δ, the solutions of (A.5) are: (1) $\kappa^* < \Delta \Rightarrow \hat{\kappa} = 0$; (2) $\kappa^* > \Delta$ and $h(\Delta) < 0$, we use the bisection method to obtain the solutions in (Δ, κ^*).

These estimates of $\hat{\kappa}$ are computed for $z^*_{(i:q)}$, then combined in a suitable way to obtain final estimates (i.e., use the median of each of the foregoing set of estimators to obtain a corresponding overall estimators of κ). Subsequently, if $\tau > \tau_{z_{(u)}}$, the τ^{th} quantile for the model output Z is obtained by the $\frac{\tau-\tau_{z_{(u)}}}{1-\tau_{z_{(u)}}}$ th quantile of the GPD distribution with $\hat{\kappa}$.

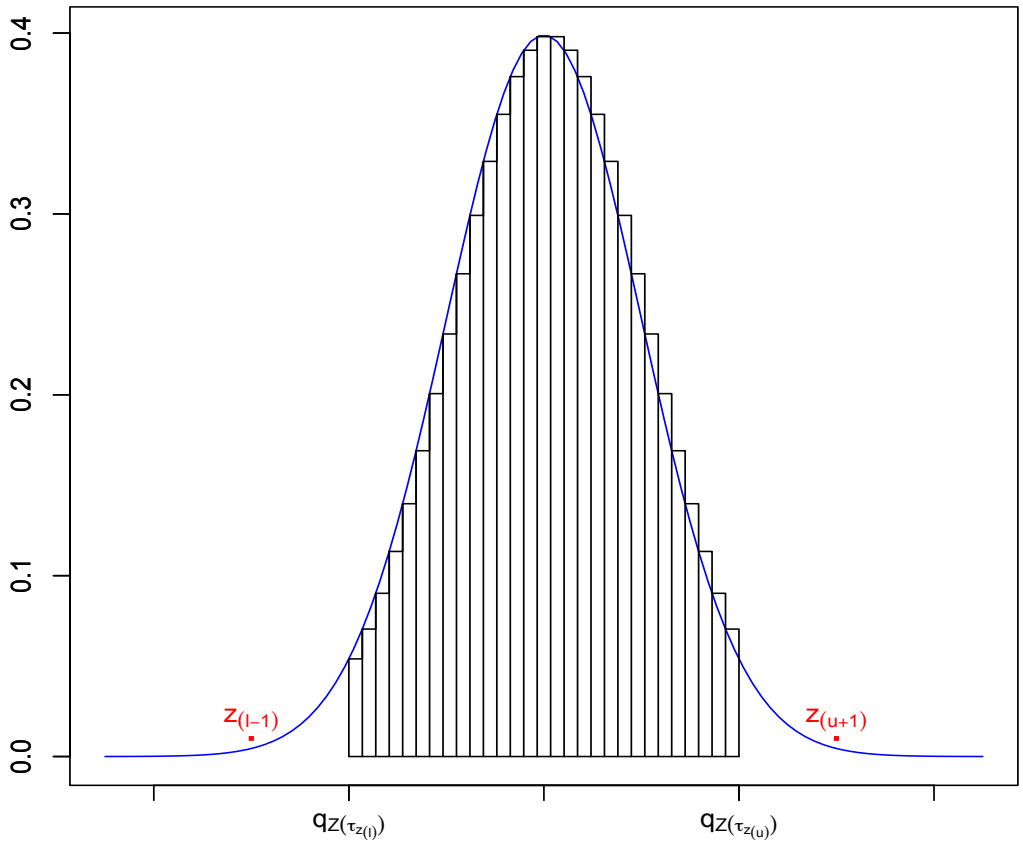

Figure 4: Likelihood estimation without specifying a density function a priori. The blue curve is a density plot of simulated data. The likelihood function expressed as a quantile-based central tendency and a generalized Pareto tail is able to characterize unbounded response variables.

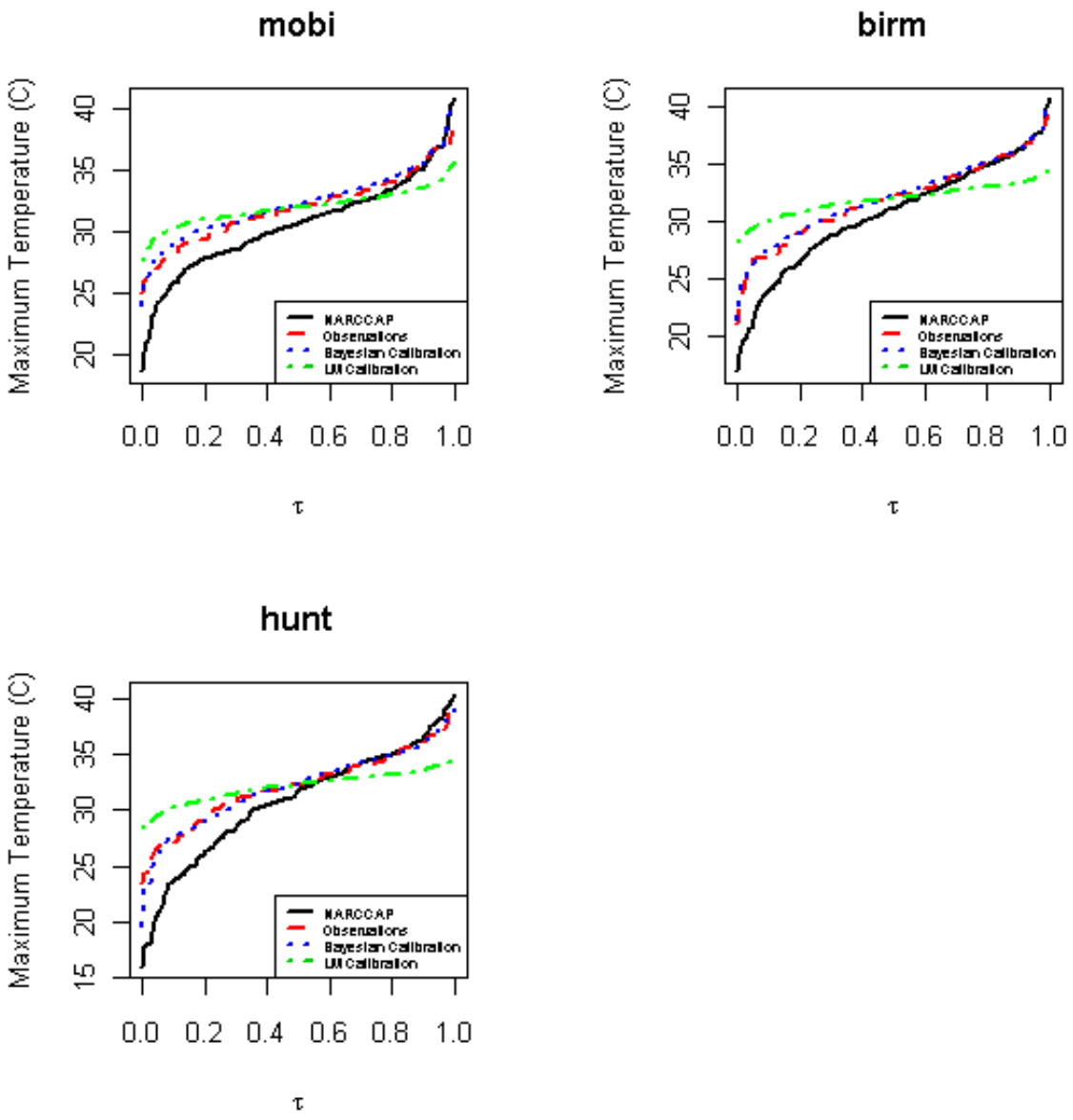

Figure 5: At three urban communities Birmingham (birm), Mobile (Mobi), and Huntsville (hunt), we plot τ^{th} quantile curves of NARCCAP outputs, observations, our Bayeisan calibrations and the simple linear regression (LM) of daily maximum temperature in year 2000, unit in Celsius (C).

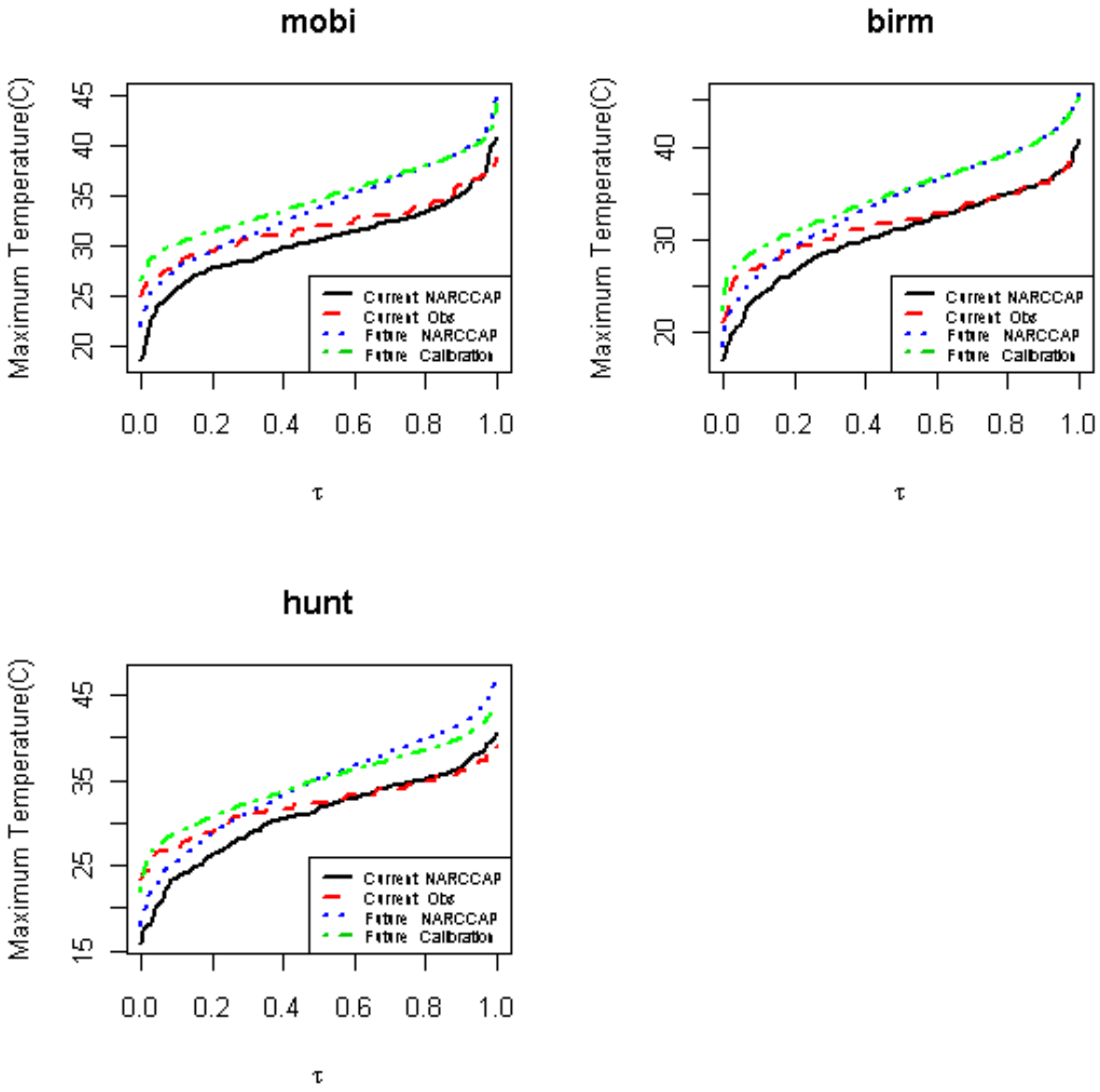

Figure 6: τ^{th} quantile curves of the historical NARCCAP data, the corresponding observations, the uncalibrated future model output and the calibrated daily maximum temperature, unit in Celsius. Note that the discrepancies mostly occurred at the lower tail of the distribution.

Table 1: Posterior medians (95% credible intervals) of annual excess deaths attributable to high temperature and heat waves across three Alabama cities. For high temperature, the baseline group is defined as days with maximum temperature under 30°C and the at-risk group is defined as days with maximum temperature above the threshold. For heat wave, the at-risk group is defined as days with maximum temperature above the quantile threshold.

High Temperature			Heat Wave		
Threshold	1991-2000	2041-2050	Threshold	1991-2000	2041-2050
32.5 °C	48 (12, 88)	111 (29, 206)	90^{th} percentile	0.24 (-0.17, 0.74)	1.8 (-1.2, 5.5)
34.0 °C	31 (11, 50)	135 (48, 216)	95^{th} percentile	0.29 (0.04, 0.60)	3.9 (0.8, 7.9)
35.5 °C	20 (6, 29)	134 (44, 203)	97.5^{th} percentile	0.17 (0.10, 0.24)	4.8 (2.8, 7.1)
37.0 °C	7 (2, 11)	130 (38, 199)	99^{th} percentile	0.09 (0.06, 0.12)	5.2 (3.6, 6.9)

Table 2: Observed and projected average number of high temperature days and heat wave days per year in Birmingham, Alabama. High temperature days occur when maximum daily temperature exceeds the threshold temperature. Heat waves are defined as consecutive days with maximum daily temperature exceeding the quantile threshold. Posterior medians (95% credible intervals) are also given for the projections.

High Temperature			Heat Wave		
Threshold	1991-2000	2041-2050	Threshold	1991-2000	2041-2050
32.5 °C	45.9	106 (98, 113)	90^{th} percentile	11.1	75 (67, 86)
34.0 °C	22.7	89 (82, 97)	95^{th} percentile	5.2	61 (53, 71)
35.5 °C	12.5	72 (64, 82)	97.5^{th} percentile	2.0	49 (42, 59)
37.0 °C	4.1	54 (46, 64)	99^{th} percentile	0.9	44 (36, 52)

REFERENCES

Anderson, G. B., and Bell, M. L. (2009), "Weather-related mortality: a study of how heat, cold, and heat waves affect mortality in the United States," *Epidemiology*, 20(2), 205–213.

Anderson, G. B., and Bell, M. L. (2011), "Heat Waves in the United States: Mortality Risk during Heat Waves and Effect Modification by Heat Wave Characteristics in 43 U.S. Communities," *Environ Health Perspect*, 119(2), 210–218.

Bell, M. L., Goldberg, R., Hogrefe, C., Kinney, P. L., Knowlton, K., Lynn, B., Rosenthal, J., Rosenzweig, C., and Patz, J. A. (2007), "Climate change, ambient ozone, and health in 59 US cities," *Climatic Change*, 82, 61–76.

Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010), "A Spatio-Temporal Downscaler for Output From Numerical Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 176–197.

Caya, D., Laprise, R., andG. Bergeron, M. G., Blanchet, J. P., Stocks, B. J., and Boer, G. J. (1995), "Description of the Canadian Regional Climate Model," *Water Air Soil Poll.*, 1, 477–482.

Chang, H. H., Zhou, J., and Fuentes, M. (2010), "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.

Curriero, F. C., Heiner, K. S., Samet, J. M., Zeger, S. L., Strug, L., and Patz, J. A. (2002), "Temperature and mortality in 11 cities of the eastern United States," *American Journal of Epidemiology*, 155(1), 81–87.

Dunson, D. B., and Taylor, J. A. (2005), "Approximate Bayesian Inference for Quantiles," *Journal of Nonparametric Statistics*, 17(3), 385 – 400.

Eder, B., and Yu, S. (2006), "A performance evaluation of the 2004 release of Models-3 CMAQ. Atmospheric Environment,".

Fuentes, M., and Raftery, A. E. (2005), "Model Evaluation and Spatial Interpolation by Bayesian Combination of Observations with Outputs from Numerical Models," *Biometrics*, 61(1), 36–45.

Gelman, A. (2004), "Parameterization and Bayesian modelling," *Journal of the American Statistical Association*, 99, 537–545.

Gryparis, A., Paciorek, C. J., Zeka, A., Schwartz, J., and Coull, B. A. (2009), "Measurement error caused by spatial misalignment in environmental epidemiology," *Biostatistics*, 10, 258–274.

Hosking, J. R. M., and Wallis, J. R. (1987), "Parameter and Quantile Estimation for the Generalized Pareto Distribution," *Technometrics*, 29(3), 339–349.

Huth, R., Kysely, J., and Pokorna, L. (2000), "A GCM simulation of heat waves, dry spells, and their relationships to circulation," *Climatic Change*, 46, 29–60.

John, D. E., and Rose, J. B. (2005), "Review of factors affecting microbial survival in groundwater," *Environ. Sci. Technol.*, 39(19), 7345–56.

Jun, M., Knutti, R., and Nychka, D. W. (2008), "Spatial Analysis to Quantify Numerical Model Bias and Dependence: How Many Climate Models Are There?," *Journal of the American Statistical Association*, 103, 934–947.

Karl, T. R., and Knight, R. W. (1997), "The 1995 Chicago heat wave: how likely is a recurrence?," *Bulletin of the American Meteorological Society*, 78, 1107–1119.

Kaufman, C., and Sain, S. (2010), "Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions," *Bayesian Analysis*, 5, 123–150.

Kharin, V., and Zwiers, F. W. (2000), "Changes in Extremes in an Ensemble of Transient Climate Simulations with a Coupled Atmosphere-Ocean GCM," *J. Clim.*, 13, 3760–3788.

Knowlton, K., Rosenthal, J. E., Hogrefe, C., Lynn, B., Gaffin, S., Goldberg, R., Rosenzweig, C., Ku, J., and Kinney, P. L. (2004), "Assessing ozone-related health impacts under a change climate," *Environ Health Perspec.*, 112, 1557–1563.

Knutti, R. (2008), "Should we believe model predictions of future climate change?," *Philosophical Transactions of the Royal Soceity A*, 366, 4647–4664.

Koenker, R. (2005), *Quantile Regression* Econometric Society Monograph Series, Cambridge University Press.

Kozumi, H., and Kobayashi, G. (2011), "Gibbs sampling methods for Bayesian quantile regression," *Journal of Statistical Computation and Simulation*, pp. 1 – 14.

Lavine, M. (1995), "On an approximate likelihood for quantiles," *Biometrika*, 82, 220–222.

Lu, Z. Q., and Clarkson, D. B. (1999), "Monotone Spline and Multidimensional Scaling," *Proceedings of American Statistical Association, Section of Statistical Computing*, pp. 185–190.

McMillan, N. J., Holland, D. M., Morara, M., and Feng, J. (2009), "Combining numerical model ouput and particulate data using Bayesian space-time modeling," *Environmetrics*, 21, 48–65.

Ogden, N. H., Maarouf, A., Barker, I. K., Bigas-Poulin, M., Lindsay, M. G., Callaghan, C. J., Ramay, F., Waltner-Toews, D., and Charrn, D. F. (2006), "Climage change and the potential for range expansion of the Lyme disease vector Ixodes scapularis in Canada," *International Journal of Parasitology*, 36(1), 63–70.

Peng, R. D., Bobb, J. F., Tebaldi, C., McDaniel, L., Bell, M. L., and Dominici, F. (2011), "Toward a Quantitative Estimate of Future Heat Wave Mortality under Global Climate Change," *Environ Health Perspect*, 119(5), 701–706.

Peng, R. D., and Wealty, L. J. (2004), "The NMMAPSdata Package," *R News*, 4, 10–14.

Ramsay, J. O. (1988), "Monotone Regression Splines in Action," *Statistical Science*, 3(4), 425–441.

Remais, J. S., Liang, S., and Spear, R. C. (2008), "Coupling hydrologic and infectious disease models to explain regional differences in schistosomiasis transmission in southwestern China," *Environ. Sci. Technol.*, 42(7), 2643–2649.

Sain, S. R., Furrer, R., and Cressie, N. (2011), "A spatial analysis of multivariate output from regional climate models," *Annals of Applied Statistics*, 5(1), 150–175.

Scinocca, J., McFarlane, N., Lazare, M., Li, J., and Plummer, D. (2008), "The CCCma third generation AGCM and its extension into the middle atmosphere," *Atmos. Chem. Phys.*, 8, 7055–7074.

Smith, R. L., Tebaldi, C., Nychka, D., and Mearns, L. O. (2009), "Bayesian modeling of uncertainty in ensembles of climate models," *Journal of the American Statistical Association*, 104(485), 97–116.

Tokdar, S., and Kadane, J. (2011), "Simultaneous Linear Quantile Regression: A Semiparametric Bayesian Approach," *Bayesian Analysis*, 6(4), 1–22.

Zhou, J., Montes, F., and Jerry, D. (In press), "Calibration of numerical model output using nonparametric spatial density functions," *Journal of Agricultural, Biological, and Environmental Statistics, Accepted for publication*, .