12-14-2011

Flexible Distributed Lag Models using Random Functions with Application to Estimating Mortality Displacement from Heat-Related Deaths

Roger D. Peng
Johns Hopkins University, rpeng@jhsph.edu

Flexible Distributed Lag Models using Random Functions with Application to Estimating Mortality Displacement from Heat-Related Deaths

Matthew J. HEATON and Roger D. PENG

As climate continues to change, scientists are left to analyze the effects these changes will have on the public. In this article, a flexible class of distributed lag models are used to analyze the effects of heat on mortality in four major metropolitan areas in the U.S. (Chicago, Dallas, Los Angeles, and New York). Specifically, the proposed methodology uses Gaussian processes as a prior model for the distributed lag function. Gaussian processes are adequately flexible to capture a wide variety of distributed lag functions while ensuring smoothness properties of process realizations. Additionally, the proposed framework allows for probabilistic inference of the maximum lag. Applying the proposed methodology revealed that mortality displacement (or, harvesting) was present for most age groups and cities analyzed suggesting that heat advanced death in some individuals. Additionally, the estimated shape of the DL functions gave evidence that prolonged heat exposure and highly variable temperatures pose a threat to public health.

Key Words: Climate change; Gaussian process; Public health; Harvesting.

1. INTRODUCTION

In their fourth assessment (IPCC 2007), the Intergovernmental Panel on Climate Change (IPCC) reports that as the global climate changes, temperatures will become more variable and extreme (Tebaldi, Hayhoe, Arblaster, and Meehl 2006) and heat waves will lengthen and strengthen in intensity (Meehl and Tebaldi 2004). Such changes are concerning as many studies have already linked extreme temperatures and heat waves to excess mortality (see, e.g., O'Neill, Zanobetti, and Schwartz 2003; Kovats and Hajat 2008; Anderson and Bell 2009). Other studies argue that excess mortality is expected to increase given the projected changes in climate (Li, Sain, Mearns, Anderson, Kovats, Ebi, Bekkedal, Kanarek, and Patz 2011; Peng, Bobb, Tebaldi, McDaniel, Bell, and Dominici 2011). For these reasons, understanding the relationship between mortality (or morbidity) and heat is one of many keys to understanding how public health will be affected by global climate changes.

M. J. Heaton is Post-doctoral Scientist, Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000 (email: heaton@ucar.edu); and R. D. Peng is Associate Professor, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205 (email: rpeng@jhsph.edu).

A common statistical tool for studying the relationship between mortality and heat are distributed lag (DL) models which relate a response observed at time t, say Y_t, to covariates measured at the current and previous time periods, say $X_t, \ldots, X_{t-M}$ for some maximum lag M, via the mean such that

$$g(\mathbb{E}(Y_t \mid X_t, \ldots, X_{t-L})) = \beta_0 + \sum_{\ell=0}^{M} \theta_\ell X_{t-\ell} \tag{1.1}$$

where $g(\cdot)$ represents an appropriate link function (e.g. identity, log, etc.). The $\{\theta_\ell : \ell = 1, \ldots, M\}$ is called the distributed lag (DL) function and describes how past covariates interact with the response at time t. In terms of relating heat to mortality, DL models match *a priori* intuition in that heat-related deaths could be the result of either (i) shorter-term (i.e. same or previous day) exposure to extreme heat resulting in heat stroke or (ii) prolonged heat exposure resulting in disease or dehydration. To highlight a few results, Braga, Zanobetti, and Schwartz (2001) used DL models to analyze time series data from 12 U.S. cities and found that the public in those cities with colder climates were more susceptible to short term heat-related deaths than cities in warmer climates. In a follow-up study, Braga, Zanobetti, and Schwartz (2002) discovered a "harvesting" (i.e. mortality displacement) effect such that heat advanced death by a few days (thus prematurely depleting the cohort of susceptible persons) such that mortality on day t was negatively associated with temperatures on days $t-3$, $t-4$, and so on. Hajat, Armstrong, Gouveia, and Wilkinson (2005) compared the DL function of heat-related deaths across different demographic profiles and concluded that poorer populations were more susceptible to heat-related illnesses over both short- and long-term heat exposures.

The goal of this study is to investigate the relationship between heat and mortality in four major metropolitan areas (Chicago, Dallas, Los Angeles, and New York) in the United States from 2001 to 2005. Non-accidental related mortality counts and daily temperature for each city were obtained from the National Morbidity, Mortality and Air Pollution Study (NMMAPS) database (Samet, Zeger, Dominici, Curriero, Coursac, Dockery, Schwartz, and Zanobetti 2000). Figure 1 displays the empirical autocorrelation function $\mathbb{C}\mathrm{orr}(Y_t, X_{t-\ell})$ for $\ell = 0, \ldots, 10$ after accounting for a day of week effect, age, and annual trends where Y_t is the mortality count on day t and X_t is the average daily temperature on day t. From Figure 1, both Chicago and Los Angeles may exhibit mortality displacement behavior in that mortality counts are negatively correlated with daily temperature at lags greater than 2 or 3. In contrast, Dallas seems to exhibit resiliency to heat while New York is more susceptible to long term heat exposure. Here, DL models are used to study these empirical relationships in greater detail.

While simple in composition, effectively using DL models for data analysis can be challenging for a few reasons. First, obtaining *precise* estimates of the DL function in (1.1) is difficult because the covariates $X_t, \ldots, X_{t-M}$ are often highly collinear resulting in inflated standard errors. Solutions to reining in the standard errors include

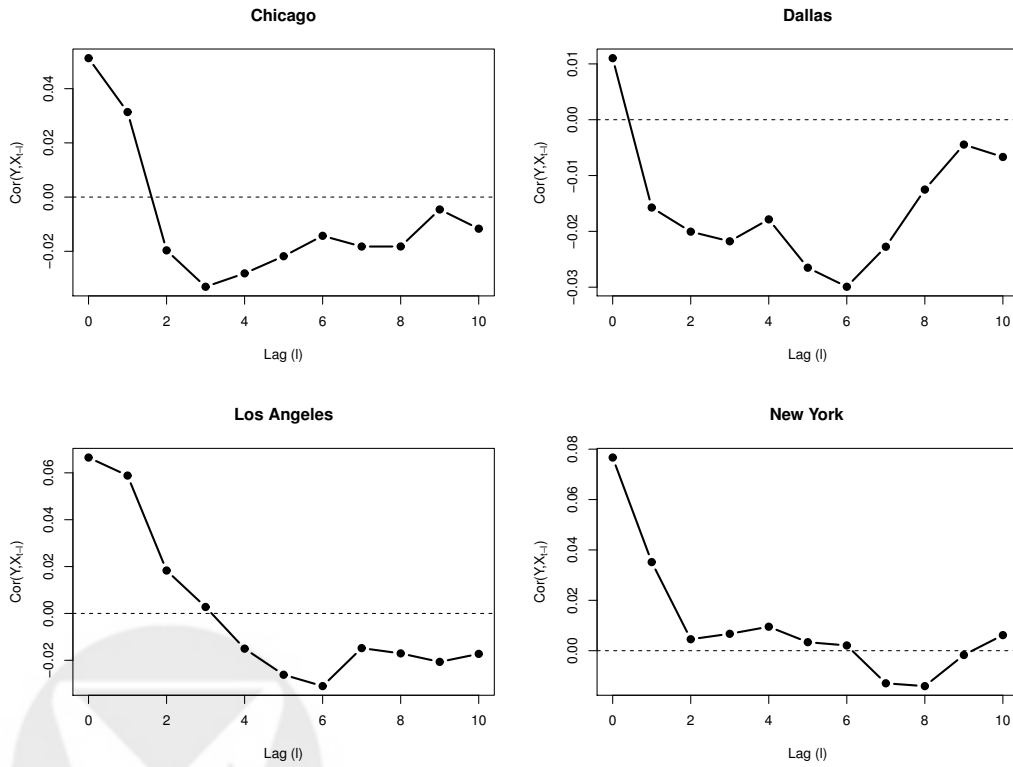*Figure 1. Empirical autocorrelation function $\mathbb{C}orr(Y_t, X_{t-\ell})$ for $\ell = 0, \ldots, 10$ after accounting for a day of week effect, age, and annual trends for each city in the current study. Note the different scales used on the y-axes.*

temporal windows (Bell, Samet, and Dominici 2004; Welty and Zeger 2005; Caffo, Peng, Dominici, Louis, and Zeger 2011), constraining $\{\theta_\ell\}$ to follow a function such as a polynomial (Schwartz 2000) or spline (Zanobetti, Wand, Schwartz, and Ryan 2000), or building in strong prior knowledge to borrow strength across the estimates (Welty, Peng, Zeger, and Dominici 2009; Peng, Dominici, and Welty 2009). Similar to Welty et al. (2009) and Peng et al. (2009), this collinearity issue is handled here by inducing strong *a priori* correlation between the parameters of the DL function while not constraining the shape to follow a particular function. Specifically, Gaussian processes (i.e. random functions) are used as a prior model for $\{\theta_\ell\}$ to this end (see the books Banerjee, Carlin, and Gelfand 2004; Cressie and Wikle 2011, for detailed discussions of Gaussian processes and their properties). A stochastic process $\{\theta_\ell : \ell \in \mathscr{D}\}$ for some continuous region $\mathscr{D}$ follows a Gaussian process if and only if for any finite collection of locations $\ell_1, \ldots, \ell_n$ the vector $(\theta_{\ell_1}, \ldots, \theta_{\ell_n})'$ follows a multivariate Gaussian distribution. Here, the Matérn (Matérn 1986) correlation function is used to induce strong *a priori* correlation between θ_{ℓ_i} and θ_{ℓ_j} to not only control standard errors of θ_ℓ but also control the smoothness properties of the DL function.

A second challenge in using DL models is choice of the maximum lag M. One possible solution is to allow M to be infinite yielding the Koyck distributed lag model (see Frances and van Oest 2004). For studies relating heat to mortality, allowing M to be infinite does not match *a priori* knowledge that M is not only finite but probably quite small (perhaps only one or two weeks). The majority of analyses employing DL models assume that M is known and choose its value based on *a priori* intuition. By choosing M to be large enough to capture all possible effects of interest, credible (or confidence) intervals for $\{\theta_l\}$ could be used to infer a smaller "true" maximum lag (e.g. the largest lag such that a credible interval does not contain zero). With this approach, the collinearity of $X_t, \ldots, X_{t-M}$ mentioned above creates more problems because the standard errors of $\{\theta_l\}$ are inflated. A major contribution of this work is to treat the maximum lag as a parameter of interest and infer its value using the data. Specifically, a large lag M is first chosen and a Gaussian process prior distribution is used for $\theta_0, \ldots, \theta_M$. Next, a second "true" maximum lag L is introduced such that $\theta_\ell = 0$ if $\ell \geq L$. Because conditioning on L is equivalent to conditioning on $\theta_L = \cdots = \theta_M = 0$, the conditional distribution $[\theta_0, \ldots, \theta_{L-1} \mid L]$ is well defined by the properties of the multivariate Gaussian distribution where $[\cdot]$ denotes a general probability distribution. Thus, a valid model specification is available for all $L = 1, \ldots, M$ enabling convenient inference for L. More modeling details and discussion on implementing a Markov chain Monte Carlo (MCMC) algorithm for posterior inference is provided in Section 2.

To conclude this Introduction, the primary contributions are outlined as follows: (i) the use of random functions (specifically, Gaussian processes) as a prior model for the DL function which alleviates collinearity issues through strong *a priori* correlation between parameters and inducing smoothness properties while not overly constraining the DL function; (ii) full posterior inference on the maximum lag L such that $\theta_\ell = 0$ if $\ell \geq L$.

(iii) an in-depth analysis of the relationship between heat and mortality in four major U.S. metropolitan areas. Section 2 details the modeling strategy used throughout this article including a discussion of how to effectively implement a MCMC algorithm for posterior inference. Section 3 validates the proposed methodology via simulation studies and Section 4 describes the results from the heat-related death analysis. Finally, Section 5 concludes, identifies short-comings of the analysis and details open research questions.

2. DISTRIBUTED LAG MODELS USING RANDOM FUNCTIONS

2.1. METHODOLOGY

For purposes of methodological development, let $t \in \mathbb{Z}$ denote time, Y_t denote a response variable of interest observed at time t and X_t denote the univariate response variable that is, potentially, associated with Y_t. Recall that the application of interest in this article is the mortality displacement of heat-related deaths such that X_t will denote average daily temperature. Because temperature data in the years 2001-2005 time frame is highly abundant and typically measured with high precision, assume that X_t is known for all t of interest.

In order to capture lagged effects of X_t on the response Y_t, consider the distributed lag model given by,

$$g(\mathbb{E}(Y_t \mid \{X_t : t \in \mathbb{Z}\})) = \beta_0 + \sum_{\ell=0}^{M} \theta_\ell X_{t-\ell}, \tag{2.1}$$

where, as in the previous section, $g(\cdot)$ denotes an appropriate link function (e.g. identity, log, etc.), and $\{\theta_\ell : \ell = 0, \ldots, M\}$ is the DL function quantifying the effect of $X_{t-\ell}$ on $\mathbb{E}(Y_t)$. Specifically, for a unit increase in $X_{t-\ell}$, holding all else constant, $g(\mathbb{E}(Y_t))$ will change by θ_ℓ. For DL models, the sum $\beta_1 = \sum_{\ell=0}^{\infty} \theta_\ell$ is often of interest quantifies the effect of a unit increase in X_t on $\mathbb{E}(Y_{t'})$ for *all* future days $t' = t+1, t+2$, etc. Note, however, that the value of β_1 alone can be deceiving. For example, notice that β_1 can be approximately 0, suggesting a null effect between covariates and response, yet $\theta_\ell \neq 0$ for all ℓ suggesting that *all* past covariates affect the response. Thus, the distributed lag function $\{\theta_\ell : \ell = 0, \ldots, M\}$ is the primary quantity of interest and β_1 is a useful univariate summary. Here, the value of M is set large enough such that $\{\theta_\ell : \ell \in \{0, \ldots, M\}\}$ captures all lagged effects of interest. For example, in the heat-related mortality example discussed in the Introduction, $M = 60$ because intuition states that, with near certainty, the effect of the average daily temperature two months ago should have no effect on mortality today. As is discussed more below, a second truncation level L (which will be estimated using the data) will be introduced which represents the "true" maximum lag at which $\theta_\ell = 0$ for $\ell \geq L$.

Prior knowledge of $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_M)'$ varies greatly from application to application. For example, when relating temperature to mortality, the heat-related mortality displacement hypothesis suggests that high temperatures in recent days may advance the death

individuals. Under such a hypothesis, the resulting DL function would be positive at small lags, negative at moderate lags then taper off to zero. As a different example, when relating pollution exposure to mortality, the DL function may have a moderate positive effect for small lags, peak at moderate lags, then decrease to zero for larger lags (Schwartz 2000). In almost all cases, however, the physical constraints of the problem suggest $\theta_\ell \to 0$ smoothly as $\ell \uparrow M$. Thus, a model for $\boldsymbol{\theta}$ should be flexible at small lags yet constrained at larger lags.

The approach proposed here is to first assume a highly flexible model for $\boldsymbol{\theta}$ then condition on the physical constraints of the problem. First, to flexibly model $\boldsymbol{\theta}$, treat $\boldsymbol{\theta}$ as a realization of a random function. Specifically, assume $\boldsymbol{\theta}$ is a realization from a Gaussian process at the finite set of locations $\{0,\ldots,M\}$ such that,

$$\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \sigma_\theta^2 \boldsymbol{R}_\theta(\phi_\theta)) \tag{2.2}$$

where $\mathbf{0}$ is the zero-vector, σ_θ^2 is a constant variance term, $\boldsymbol{R}_\theta(\phi_\theta)$ is a $(M+1)\times(M+1)$ correlation matrix with ij^{th} element given by the Matérn correlation function

$$\rho_\theta(\|i-j\|) = \left(\frac{\sigma_\theta^2}{2^{\nu-1}\Gamma(\nu)}\right)(\phi_\theta\|i-j\|)^\nu K_\nu(\phi_\theta\|i-j\|)$$

where ν is a smoothness parameter, ϕ_θ is a decay parameter, and K_ν is the modified Bessel function of the second kind of order ν. One particularly useful property of the Matérn correlation function is that realizations can be shown to be k times mean square differentiable if and only if $\nu > k$ (see Handcock and Stein 1993; Handcock and Wallis 1994). Thus, ν is used to control the smoothness properties of the distributed lag function. Following Welty et al. (2009) and Peng et al. (2009), the DL function is hypothesized to be smooth over time; thus, here ν is fixed at 3 *a priori*. Borrowing a term from spatial statistics, the parameter ϕ_θ is a decay parameter and controls the correlation at fixed temporal distances. In other words, as ϕ_θ increases, $\mathbb{C}\mathrm{orr}(\theta_\ell, \theta_{\ell+u})$ decreases for fixed $u \in \mathbb{R}\setminus 0$. Finally, σ_θ^2 is a variance term that controls the maximum (and minimum) height of the DL function. For example, if σ_θ^2 is small (large) then $\max_\ell\{\theta_\ell : \ell \in \{0,\ldots,M\}\}$ will, with high probability, also be small (large) in absolute value. Importantly, Zhang (2004) showed that the parameter ϕ_θ is not consistently estimable and fixing ϕ_θ *a priori* while estimating σ_θ^2 is sufficient. Thus ϕ_θ will be treated as fixed for the remainder of this article. In initial investigations of the methodology proposed here, ϕ_θ was treated as an unknown parameter but results were insensitive to its value.

Having defined a very flexible model for $\boldsymbol{\theta}$ in (2.2) above, consider now incorporating the physical constraints of the problem. Under (2.2), the physical constraint that $\theta_\ell \to 0$ smoothly as $\ell \uparrow M$ can now be enforced via conditioning in the following manner. Consider a lag time $L \in \{1,\ldots,M\}$ such that $\theta_\ell = 0$ if $\ell \geq L$ and partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ where $\boldsymbol{\theta}_1 = (\theta_0,\ldots,\theta_{L-1})'$ and $\boldsymbol{\theta}_2 = (\theta_L,\ldots,\theta_M)'$. Furthermore, partition

$$\boldsymbol{R}_\theta = \begin{pmatrix} \boldsymbol{R}_{\theta,11} & \boldsymbol{R}_{\theta,12} \\ \boldsymbol{R}_{\theta,21} & \boldsymbol{R}_{\theta,22} \end{pmatrix}$$
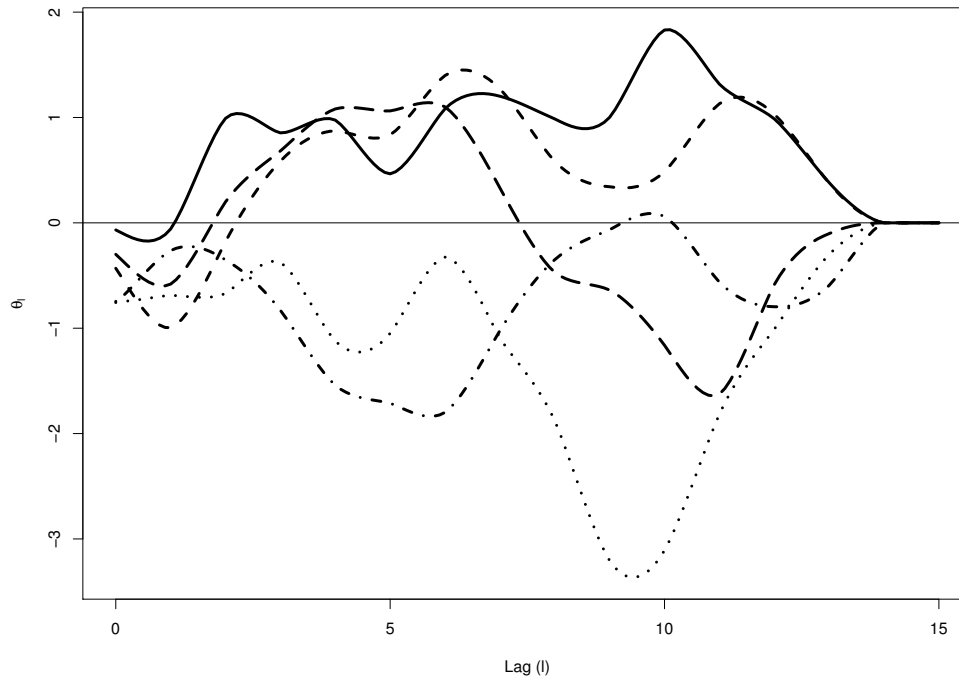
Figure 2. The non-zero portion (i.e. $\boldsymbol{\theta}_1$) of five DL functions randomly drawn according to the random function DL model with $M = 21$, $L = 14$, $\sigma_\theta^2 = 1$, and $\phi_\theta = 0.18$. Notice that the random function DL model allows for a wide variety of DL functions while constraining each of these to approach 0 as the lag increases.

such that $\mathbb{C}\mathrm{orr}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) = \boldsymbol{R}_{\theta,ij}$ for $i, j = 1, 2$. Conditioning (2.2) on L, the distribution of the non-zero portion of the DL function $\boldsymbol{\theta}_1$ follows directly from the properties of the multivariate normal distribution. That is, $\boldsymbol{\theta}_1 | L \sim \mathcal{N}(\mathbf{0}, \sigma_\theta^2 \boldsymbol{R}_\theta^\star(L))$ where $\boldsymbol{R}_\theta^\star(L) = \boldsymbol{R}_{\theta,11} - \boldsymbol{R}_{\theta,12} \boldsymbol{R}_{\theta,22}^{-1} \boldsymbol{R}_{\theta,21}$ is now $L \times L$. By conditioning on L, notice that the first few elements of $\boldsymbol{\theta}_1$ (i.e. the lagged effects at small lags) are less constrained than the lagged effects at larger lags. To illustrate this behavior, Figure 2 plots five random draws of $\boldsymbol{\theta}_1$ using $M = 21$, $L = 14$, $\phi_\theta = 0.18$ and $\sigma_\theta^2 = 1$. Notice that at early and moderate lags, the functions look quite different; yet, all taper off to zero as the lag increases.

Returning to the discussion of the truncation level M in (2.1), the role of M is now more clear given the introduction of the maximum lag L. Specifically, M is chosen large enough such that, with prior probability 1, $\theta_\ell = 0$ if $\ell \geq M$. In practice, the value of M should often be chosen to be much larger than the maximum possible value of L in order to properly capture the effect of conditioning on $\boldsymbol{\theta}_2 = \mathbf{0}$. For example, notice that if $L = M$ then $\boldsymbol{\theta}_2 = \theta_M$ is univariate and conditioning on $\theta_M = 0$ will not adequately capture the knowledge that $\theta_{M+1} = \theta_{M+2} = \cdots = 0$. To avoid this, L, *a priori*, will be assumed to follow a discrete uniform prior where $\max_\ell \{\ell : \mathrm{Pr}(L = \ell) > 0\}$ will be much less than M.

2.2. MODEL SPECIFICATION AND PARAMETER ESTIMATION

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_T)$ denote the vector of observed response variables. Also consider, for the moment, that the support of each Y_t is $\mathbb{R}$ such that a Gaussian assumption is appropriate (perhaps following a transformation). Using (2.1), a model for $\boldsymbol{Y}$ can be written as,

$$\boldsymbol{Y} \mid \boldsymbol{X} \sim \mathcal{N} \left(\beta_0 \boldsymbol{1}_T + \boldsymbol{X} \boldsymbol{\theta}, \sigma_Y^2 \boldsymbol{I}_T \right), \tag{2.3}$$

where $\boldsymbol{1}_T$ is the length T column vector of ones, $\boldsymbol{X}$ is the $T \times M + 1$ design matrix with t^{th} row $(X_t, X_{t-1}, \ldots, X_{t-M})$, $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_M)'$, σ_Y^2 is the error variance and $\boldsymbol{I}_T$ is the rank T identity matrix. Notice that (2.3) assumes that all of the temporal correlation in $\boldsymbol{Y}$ is explained through $\boldsymbol{X}$. Admittedly, this may be overly simplistic because $\boldsymbol{Y}$ may have excess temporal correlation beyond that explained by $\boldsymbol{X}$; however, this extension is not considered here and focus is placed on estimating the DL function $\boldsymbol{\theta}$. Partition $\boldsymbol{X} = [\boldsymbol{X}_{1:L}, \boldsymbol{X}_{(L+1):(M+1)}]$ such that $\boldsymbol{X}_{a:b}$ are columns a through b of $\boldsymbol{X}$. Conditioning (2.3) on L, gives,

$$\boldsymbol{Y} \mid \boldsymbol{X}, L \sim \mathcal{N} \left(\beta_0 \boldsymbol{1}_T + \boldsymbol{X}_{1:L} \boldsymbol{\theta}_1, \sigma_Y^2 \boldsymbol{I}_T \right) \tag{2.4}$$

where, again, $\boldsymbol{\theta}_1 \mid L \sim \mathcal{N}(\boldsymbol{0}, \sigma_\theta^2 \boldsymbol{R}_\theta^\star(L))$ as in the previous section.

The unknown parameters in model (2.4) are $L, \beta_0, \sigma_Y^2, \sigma_\theta^2$, and the DL function $\boldsymbol{\theta}_1$. To emphasize, a strong contribution of the proposed modeling strategy here is to treat the maximum lag L as an additional unknown parameter and estimate it from the data. The approach here is Bayesian and parameter estimation is achieved via MCMC simulation. That is, each parameter is drawn successively from its' complete conditional distribution (i.e. the distribution of the parameter given all other parameters and the data). With the exception of the maximum lag L, implementing a Gibbs sampler to obtain realizations from the posterior distribution is straight forward. For example, conditional on the maximum lag L, the complete conditional distribution for $\boldsymbol{\theta}_1$ is Gaussian (i.e. $\boldsymbol{\theta}_1$ is a conjugate parameter). Likewise, if conjugate prior distributions are appropriate for the intended application, then, again conditional on L, β_0, σ_θ^2, and σ_Y^2 can all be drawn directly from their complete conditional distribution without the use of Metropolis-Hastings (MH) type algorithms. However, sampling L is more difficult.

The natural instinct is to draw L from its complete conditional distribution using a MH algorithm but complications immediately arise. For example, consider a proposal value $L^\star$ such that $L^\star > L$. Evaluating the likelihood in this case would require values for $\theta_L, \ldots, \theta_{L^\star - 1}$ which are not available in the current draw of $\boldsymbol{\theta}_1$. Likewise, if $L^\star < L$ it may not be reasonable to simply "zero-out" the elements of $\boldsymbol{\theta}_1$ to evaluate (2.4) given $L^\star$. Thus rejection of $L^\star$ may result *not* because the value of $L^\star$ is poor but because the value of $\boldsymbol{\theta}$ used to evaluate the likelihood is poor. The proposed solution to this is to use a collapsed Gibbs sampler (see van Dyk and Park 2008) and sample $(\boldsymbol{\theta}_1, L)$ from its joint complete conditional distribution $[\boldsymbol{\theta}_1, L \mid \boldsymbol{Y}, \boldsymbol{X}, \sigma_Y^2, \sigma_\theta^2, \phi_\theta] = [\boldsymbol{\theta}_1 \mid L, \boldsymbol{Y}, \boldsymbol{X}, \sigma_Y^2, \sigma_\theta^2, \phi_\theta][L \mid \boldsymbol{Y}, \boldsymbol{X}, \sigma_Y^2, \sigma_\theta^2, \phi_\theta]$. That is, first sample L conditional on everything *except* $\boldsymbol{\theta}_1$ then sample $\boldsymbol{\theta}_1$ from its conjugate

complete conditional distribution. Obviously, this results in a valid (i.e. converges to the correct stationary distribution) Markov chain because any joint distribution can be factored $[A, B] = [A \mid B][B]$ and sampled from accordingly.

In order to facilitate sampling L, a convenient prior distribution is the discrete uniform distribution on $\{1, \dots, L_{\max}\}$ where $L_{\max} < M$ and, often, $L_{\max} \ll M$ (see the discussion above). Using a discrete uniform prior for L yields,

$$[L = \ell \mid \boldsymbol{Y}, \boldsymbol{X}, \sigma_Y^2, \beta_0, \sigma_\theta^2] = \frac{\mathscr{N}\left(\boldsymbol{Y}; \beta_0 \mathbf{1}_T, \sigma_\theta^2 \boldsymbol{X}_{1:\ell} \boldsymbol{R}_\theta^\star(\ell) \boldsymbol{X}_{1:\ell}' + \sigma_Y^2 \boldsymbol{I}_T\right)}{\sum_{i=1}^{L_{\max}} \mathscr{N}\left(\boldsymbol{Y}; \beta_0 \mathbf{1}_T, \sigma_\theta^2 \boldsymbol{X}_{1:i} \boldsymbol{R}_\theta^\star(i) \boldsymbol{X}_{1:i}' + \sigma_Y^2 \boldsymbol{I}_T\right)} \tag{2.5}$$

where $\mathscr{N}(x; \mu, \sigma^2)$ denotes the Gaussian distribution function with mean μ and variance σ^2 evaluated at x. Using (2.5), L can be drawn directly following which $\boldsymbol{\theta}_1$ can be drawn directly from its complete conditional distribution as discussed above.

Because (2.1) can be used with more than just the Gaussian distribution, turn now to the case where Y_t is distributed as a Poisson random variable. Using (2.1), a model for Poisson data is given by,

$$Y_t \mid L, \boldsymbol{x}_t(L), \beta_1, \boldsymbol{\theta}_1, \sigma_\theta^2 \overset{ind}{\sim} \mathscr{P}\left(\exp\left\{\beta_0 + \boldsymbol{x}_t'(L)\boldsymbol{\theta}_1\right\}\right) \tag{2.6}$$

where $\mathscr{P}(\mu)$ represents the Poisson distribution with mean μ and $\boldsymbol{x}_t'(L) = (X_t, X_{t-1}, \dots, X_{t-L+1})'$. The apparent difficulty with using the Poisson distribution, rather than the Gaussian distribution, is loss of conjugacy for the parameters β_0 and $\boldsymbol{\theta}_1$ (notice that σ_θ^2 may still be conjugate even in the Poisson setting depending on the choice of prior distribution). Additionally, the difficulty of sampling L persists. For sampling β_0 and $\boldsymbol{\theta}_1$, the adaptive Metropolis algorithm of Haario, Saksman, and Tamminen (2001) works well here and its use is highly recommended to achieve efficient posterior simulation.

To sample L in the Poisson setting, the collapsed Gibbs sampler method is again used. However, the Poisson setting presents a new computational challenge. Specifically, the posterior probability $[L = \ell \mid \boldsymbol{Y}, \boldsymbol{X}, \sigma_\theta^2]$ given by,

$$\frac{\int_{\mathbb{R}^\ell} \prod_{t=1}^T \mathscr{P}\left(Y_t; \exp\left\{\beta_0 + \boldsymbol{x}_t'(\ell)\boldsymbol{\theta}_{1:v}\right\}\right) \mathscr{N}\left(\boldsymbol{\theta}_{1:\ell}; \boldsymbol{0}, \sigma_\theta^2 \boldsymbol{R}_\theta^\star(\ell)\right) d\boldsymbol{\theta}_{1:\ell}}{\sum_i \int_{\mathbb{R}^i} \prod_{t=1}^T \mathscr{P}\left(Y_t; \exp\left\{\beta_0 + \boldsymbol{x}_t'(i)\boldsymbol{\theta}_{1:i}\right\}\right) \mathscr{N}\left(\boldsymbol{\theta}_{1:i}; \boldsymbol{0}, \sigma_\theta^2 \boldsymbol{R}_\theta^\star(i)\right) d\boldsymbol{\theta}_{1:i}} \tag{2.7}$$

is no longer available in closed form where $\mathscr{P}(x; \mu)$ is the Poisson distribution with mean μ evaluated at x and $\boldsymbol{\theta}_{1:i} = (\theta_0, \dots, \theta_{i-1})'$. Notice that the numerator in (2.7) is simply $\mathbb{E}_{\boldsymbol{\theta}_{1:\ell}}(\prod_{t=1}^T \mathscr{P}(Y_t; \exp\{\beta_0 + \boldsymbol{x}_{t|\ell}'\boldsymbol{\theta}_{1:\ell}\})$ where $\boldsymbol{\theta}_{1:\ell} \sim \mathscr{N}(\boldsymbol{0}, \sigma_\theta^2 \boldsymbol{R}_\theta^\star(\ell))$ and could be calculated using Monte Carlo integration. Monte Carlo integration, however, could be imprecise if the product $\prod_{t=1}^T \mathscr{P}(Y_t; \exp\{\beta_0 + \boldsymbol{x}_t'(\ell)\boldsymbol{\theta}_{1:v}\})$ is highly concentrated relative to $\mathscr{N}(\boldsymbol{0}, \sigma_\theta^2 \boldsymbol{R}_\theta^\star(\ell))$. Rather, the approach used here is to calculate the marginal posterior probability $[L = v \mid \boldsymbol{Y}, \boldsymbol{X}]$ by calculating the marginal likelihoods for each value of L using the method of Chib and Jeliazkov (2001). The resulting collapsed Gibbs sampler follows by first sampling L from $[L \mid \boldsymbol{Y}, \boldsymbol{X}]$ followed by sampling $\boldsymbol{\theta}_1$ and σ_θ^2 from their respective complete conditional distributions via Metropolis steps.

Table 1. Simulation study model scenarios where $\phi(x \mid \mu, \sigma)$ is the Gaussian probability density function with mean μ and standard deviation σ evaluated at x and $\phi^{(1)}(\cdot)$ is its first derivative. For each scenario, $\{X_t\}$ were taken to be the average daily temperatures in Chicago between April 1 and September 30, 2005. The distributed lag functions for each scenario are given by the solid black lines in Figure 3.

Scenario	Model
(a)	$Y_t \overset{ind}{\sim} \mathscr{N}\left(25 + 0.17 \sum_{\ell=0}^{13} \sin(2\pi\ell/14) X_{t-\ell}, 25\right)$
(b)	$Y_t \overset{ind}{\sim} \mathscr{P}\left(\exp\left\{1.75 + 0.03 \sum_{\ell=0}^{13} \sin(2\pi\ell/14) X_{t-\ell}\right\}\right)$
(c)	$Y_t \overset{ind}{\sim} \mathscr{N}\left(-55 + 1.75 \sum_{\ell=0}^{13} \phi(\ell \mid 4, 14/3) X_{t-\ell}, 25\right)$
(d)	$Y_t \overset{ind}{\sim} \mathscr{P}\left(\exp\left\{-15 + 0.3 \sum_{\ell=0}^{13} \phi(\ell \mid 4, 14/3) X_{t-\ell}\right\}\right)$
(e)	$Y_t \overset{ind}{\sim} \mathscr{N}\left(125 + 1.75 \sum_{\ell=0}^{13} \phi^{(1)}(\ell \mid -3, 7) X_{t-\ell}, 25\right)$
(f)	$Y_t \overset{ind}{\sim} \mathscr{P}\left(\exp\left\{18 + 0.3 \sum_{\ell=0}^{13} \phi^{(1)}(\ell \mid -3, 7) X_{t-\ell}\right\}\right)$

3. SIMULATION STUDIES

3.1. SIMULATION STUDY OUTLINE AND PRIOR DISTRIBUTIONS

The goal of this section is to study how the proposed modeling strategy of Section 2 recovers various forms of DL functions based on different distributional assumptions for Y_t. For this simulation study, 100 data sets, consisting of $T = 153$ values of Y_t (5 months of daily data), were simulated for each model scenario given in Table 1. Graphically, the distributed lag functions for each scenario are given by the solid black lines in Figure 3. The general strategy in selecting these scenarios was to choose three different functional forms (a sine function, Gaussian density function, and derivative of the Gaussian density function) for the "true" DL function and then study the ability of the model to recover this function assuming a Gaussian and Poisson distribution for Y_t. Thus, the six scenarios listed in Table 1 are a 3×2 full factorial design with three levels of DL functions and two distributional assumptions. Note that for each scenario, the DL function is zero if the lag $\ell \geq 14$. The additive and multiplicative constants were chosen to scale the simulated data to approximately match the scale of the mortality dataset which is of primary interest in this article.

The model given by Equation (2.4) was fit to each dataset in scenarios (a), (c), and (e) while (2.6) was fit to each simulated data set in scenarios (b), (d), and (f). For both the Gaussian and Poisson models, M was assumed to be 30 and a discrete prior distribution was assumed for L with equal probability assumed for the values of $\{1, 4, 7, 10, 14, 17, 21\}$. This is a moderately coarse grid for L yet is still informative as to the maximum non-zero lag. A finer grid for L could be used but computation, especially for the Poisson model which relies on the methods of Chib and Jeliazkov (2001) to calculate marginal likelihoods, could quickly become daunting. For the Gaussian models, the prior distribution for β_0 was assumed to be $\mathscr{N}(0, 100^2)$ while for the Poisson models $\beta_0 \sim \mathscr{N}(0, 10^2)$. Note that for the
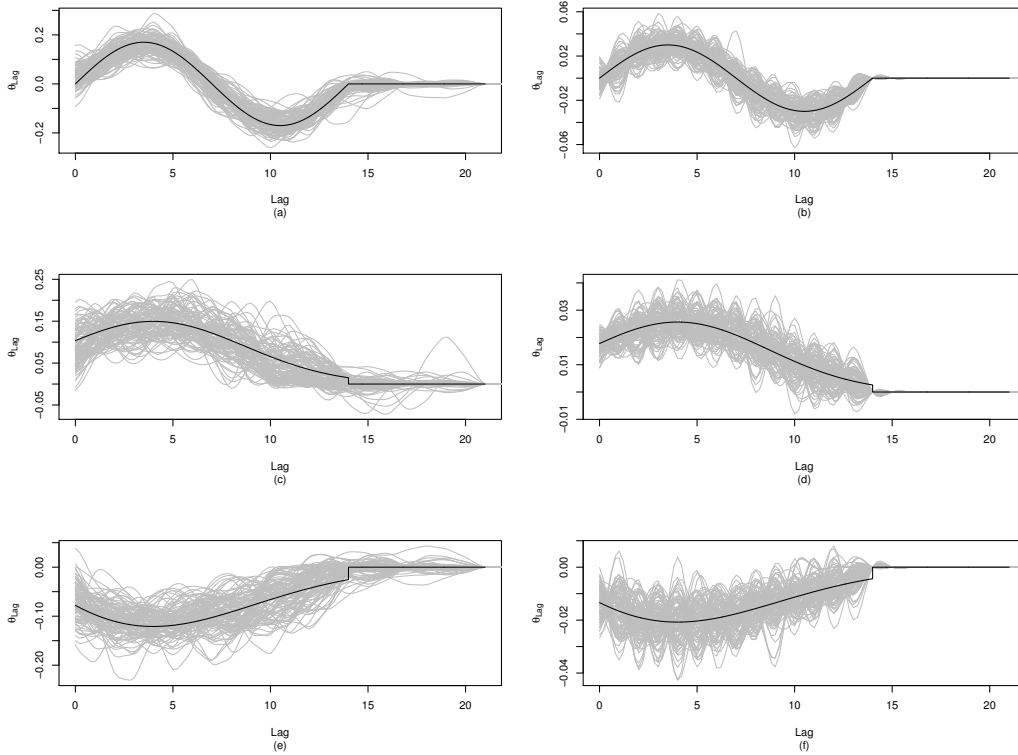
Figure 3. The true distributed lag functions (solid black lines) for each simulation scenario along with the 100 posterior mean DL functions for each simulated data set (gray lines).

Poisson case, the value of β_0 is associated with $\log(\mathbb{E}(Y_t))$ and a $\mathcal{N}(0,10^2)$ prior is quite non-informative on the log-scale. Finally, proper inverse gamma (denoted by $\mathscr{IG}(\cdot)$) prior distributions were assumed for both σ_θ^2 and σ_Y^2; specifically, σ_θ^2 and σ_Y^2 were assumed to be *a priori* independent $\mathscr{IG}(3, \text{rate} = 3)$ random variables.

Analysis was carried out by obtaining $10,000$ draws from the posterior distribution of model parameters via a Gibbs sampling algorithm. If lacking conjugacy, draws were obtained using the adaptive Metropolis algorithm of Haario et al. (2001). An analysis of the posterior draws using trace plots suggested that the Gaussian and Poisson models required approximately $5,000$ and $10,000$ draws, respectively, as burn-in.

3.2. STUDY RESULTS

Figure 3 displays the 100 marginal posterior mean DL functions (that is, $\mathbb{E}(\theta_\ell \mid \boldsymbol{Y}, \boldsymbol{X})$ for $\ell = 0, \ldots, 21$) in each simulation scenario with the true DL function overlaid. According to Figure 3, on average, the model is finding the correct DL function. Figure 4 displays the empirical coverage (EC) rates of 95% central credible intervals for θ_ℓ. In general, the EC rates are close to the nominal rate of 0.95 yet the uncertainty is slightly understated in scenarios (a) and (e).

Table 2 displays the empirical bias, root mean square error (RMSE), coverage rate. The Berkeley Electronic Press
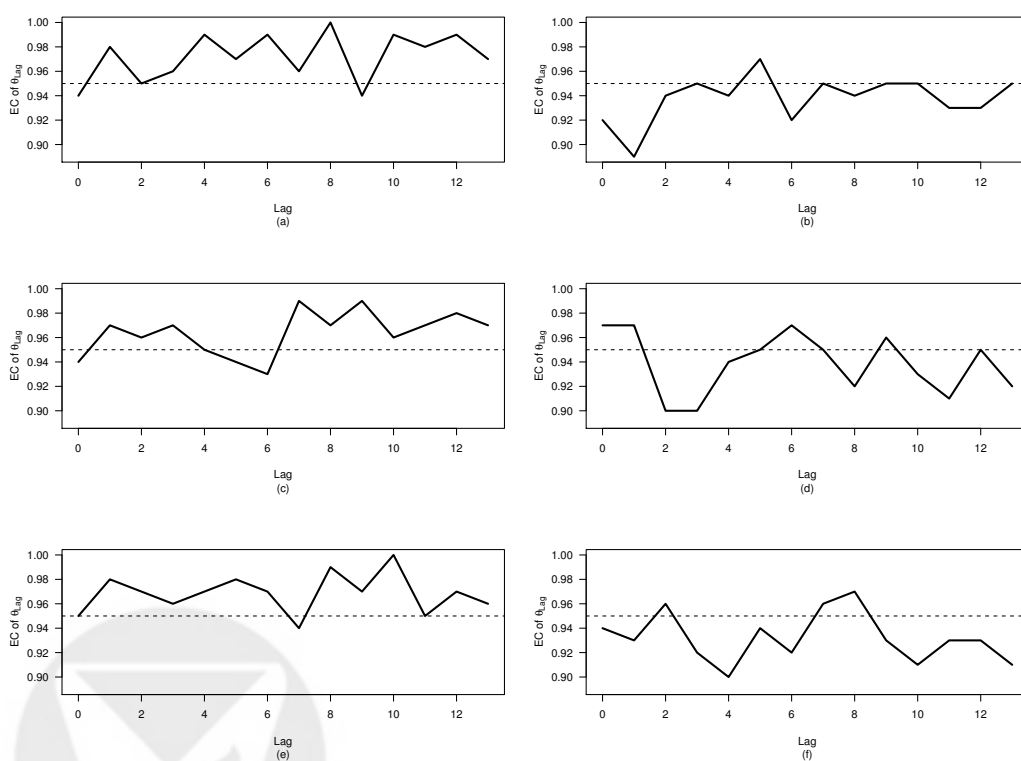
Figure 4. Empirical coverage (EC) rates of 95% central credible intervals for θ_ℓ (solid lines). The dashed horizontal line indicates the nominal rate of 95%. The empirical rates are near the nominal rates of 0.95 but the uncertainty is slightly understated in scenarios (a) and (e).

Table 2. Empirical bias, root mean square error (RMSE) and 95% central credible interval coverage and width for the cumulative effect parameter $\beta_1 = \sum_{\ell=0}^{L} \theta_\ell$. The true value of β_1 was $0.00, 0.00, 1.42, 0.24, -1.22$, and -0.20 for scenarios (a)-(f), respectively. The model fits seem to be recovering the correct parameter with near nominal coverage.

Scenario	Bias	RMSE	Coverage	Width
(a)	0.00	0.11	0.96	0.46
(b)	0.00	0.01	0.98	0.03
(c)	-0.07	0.14	0.91	0.47
(d)	-0.00	0.01	0.94	0.03
(e)	0.06	0.14	0.91	0.48
(f)	-0.00	0.00	0.91	0.02

Table 3. Average, across simulated data sets, posterior probability distribution $[L \mid \boldsymbol{Y}, \boldsymbol{X}]$. The model is able to correctly identify the true maximum lag of $L = 14$.

Scenario	$[L \mid \boldsymbol{Y}, \boldsymbol{X}]$						
	1	4	7	10	14	17	21
(a)	0.000	0.000	0.000	0.000	0.629	0.272	0.099
(b)	0.000	0.000	0.000	0.000	0.999	0.001	0.000
(c)	0.000	0.000	0.005	0.304	0.431	0.173	0.087
(d)	0.000	0.000	0.000	0.067	0.933	0.000	0.000
(e)	0.000	0.000	0.002	0.179	0.489	0.221	0.110
(f)	0.000	0.000	0.000	0.038	0.962	0.000	0.000

width of a 95% central credible interval for the cumulative effect of *all* past covariates given by $\beta_1 = \sum_{\ell=0}^{L} \theta_\ell$. Again, the model fits seem to recover to correct value (indicated by low bias) and have near nominal coverage rates of 95%.

Table 3 displays the average, across simulated data sets, posterior probability distribution $[L \mid \boldsymbol{Y}, \boldsymbol{X}]$ for each simulation scenario. Notice that the posterior mode of $[L \mid \boldsymbol{Y}, \boldsymbol{X}]$ is located on the correct value of $L = 14$ suggesting that both the Gaussian and Poisson models are able to correctly recover the maximum lag. For the Gaussian case, notice that scenarios (c) and (e) place a substantial amount of posterior mass at $L = 10$ and $L = 17$. For scenario (c), notice that $\sum_{\ell=10}^{13} \theta_\ell / \sum_{\ell=0}^{13} \theta_l = 0.12$ so that the final four lags only account for 12% of the cumulative effect of the covariates on $\mathbb{E}(Y_t)$. Because of the use of a log-link function, this 12% is amplified creating a larger discrepancy between $\mathbb{E}(Y_t \mid L = 10)$ and $\mathbb{E}(Y_t \mid L = 14)$ in the Poisson setting. Hence, the Poisson setting is more sensitive to changes in the DL function due to the use of the log-link.

4. APPLICATION TO HEAT-RELATED DEATHS

Let Y_{cayt} denote the number of non-accidental mortalities for city c (Chicago, Dallas, Los Angles, New York), age category $a \in \{< 65, 65\text{-}74, \geq 75\}$, year $y \in \{2001,\ldots,2005\}$ and time $t = 1,\ldots,183$ where $t = 1$ corresponds to April 1 and $t = 183$ corresponds to September 30. In similar notation let x_{cyt} denote the average daily temperature for city c, year y and time t. The underlying model was assumed to be,

$$Y_{cayt} \overset{ind}{\sim} \mathscr{P}\left(\exp\left\{\beta_{ca0} + \boldsymbol{x}'_{cyt}(L)\boldsymbol{\theta}_{ca1} + \boldsymbol{z}'_{cayt}\boldsymbol{\gamma}_{cay}\right\}\right)$$

where β_{ca0} is the intercept for (city, age category) $= (c,a)$, $\boldsymbol{x}'_{cyt}(L)$ is the vector of lagged temperatures with $M = 60$, $\boldsymbol{\theta}_{ca1} = (\theta_{ca10},\ldots,\theta_{ca1(L-1)})'$ is the vector of DL coefficients, $\boldsymbol{z}_{cayt}$ is a vector of confounding covariates including day of week and a natural cubic spline basis in time with three degrees of freedom per year (15 degrees of freedom total), and $\boldsymbol{\gamma}_{cay}$ is the associated vector of coefficients. Vague Gaussian prior distributions were used for each coefficient in $\boldsymbol{\gamma}_{cay}$. Exploratory analysis of each city suggested that the maximum lag was no more than 14 days. Thus, the prior distribution for L was discrete with equal probability on the values $\{1,3,5,7,9,11,14\}$. Inference for model parameters was done via MCMC sampling using 25,000 draws from the posterior after an initial burn in of 50,000 draws.

The posterior probability $[L = 3 \mid \boldsymbol{Y},\boldsymbol{X}]$ was at least 0.95 but typically upwards of 0.99. The older than 75 age category in Chicago placed 5% posterior mass at $L = 5$. These results seem to coincide with the findings of Braga et al. (2001) and Braga et al. (2002) but contradict those of Hajat et al. (2005) would found significant lags of up to 4 weeks. However, Hajat et al. (2005) used various demographic information that was not available here. Thus, direct comparison of the two findings is not advisable.

Figure 5 displays the posterior mean distributed lag function by city and age group expressed in terms of percent increase in mortality due to unit increase in temperature; i.e. $100 \times (\exp\{\theta_l\} - 1)$. In the majority of cases, evidence of mortality displacement or harvesting is present with the largest effects occurring on the most recent days and negative effects thereafter. Thus, in most cases heat advances the death of at-risk individuals. Comparing across age groups, the younger age groups have the smallest effect for all cities except New York. Thus, heat-related effects are amplified for the older age groups.

Figure 6 displays the posterior distribution of the percent increase in mortality due to a unit increase in temperature for the past L time periods (i.e. $100 \times (\exp\{\beta_1\} - 1)$) by city and age group. Dallas seems to be the least susceptible to heat effects. This is not surprising as Dallas is a very warm climate where citizens most already but used to warm temperatures. In contrast, Los Angeles, the elderly in Chicago, and New York seem to be more susceptible to heat. Comparing the age groups, in all cities except Dallas the elderly had a larger cumulative effect of heat on mortality. For Dallas, the posterior densities for all age groups are centered near zero. This information combined with the distributed lag

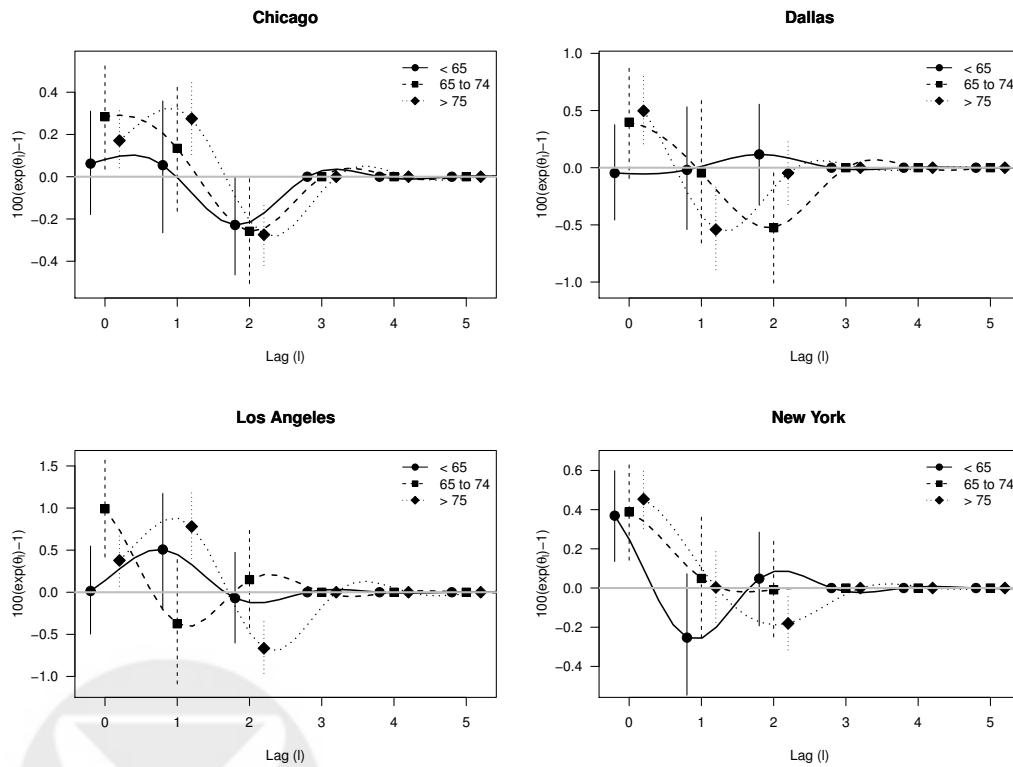*Figure 5. Estimated (i.e. posterior mean) DL function by city and age group expressed in terms of percent increase in mortality due to unit increase in temperature $(100 \times (\exp\{\theta_l\} - 1))$. Most cases suggest a mortality displacement or harvesting effect where deaths are advanced by a day or two. Furthermore, the heat effects for the older than 75 age group are exaggerated compared to the younger age groups.*
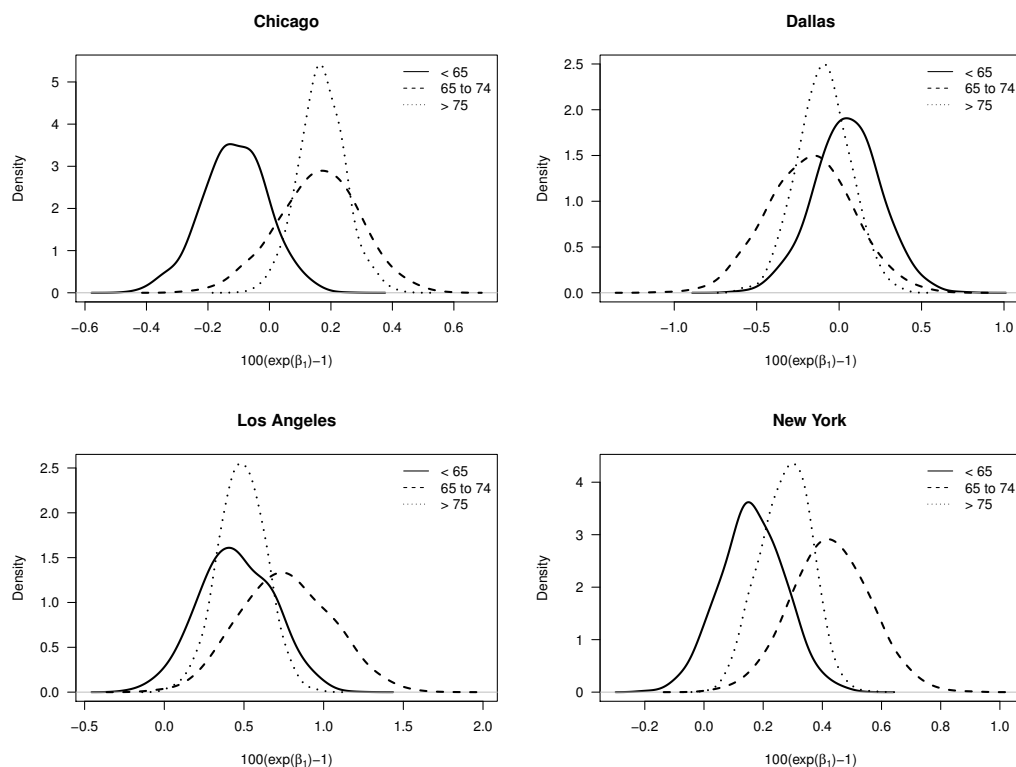
Figure 6. Posterior density of $100 \times (\exp\{\beta_1\} - 1)$; i.e. percent increase in mortality due to a unit increase in temperature for the past L time periods by city and age group. In most instances, the majority of the density curve is greater than zero suggesting that a uniform increase in temperature would result in more heat-related deaths. Dallas (Los Angeles) seems to be the least (most) susceptible to heat effects.

curves in Figure 5 suggest that heat only advances death of those highly susceptible persons in Dallas which mortality displacement creates a net-zero effect of heat on mortality.

Given the DL functions in Figure 5 and the cumulative effects in Figure 6, one interesting conclusion is that highly variable temperatures are a strong threat to public health due to mortality displacement. While heat-waves and successively high temperatures on neighboring days is a problem (as evidenced by positive cumulative effects), notice that the shape of the DL functions suggest that the expected number of deaths from three successive high temperature days is typically less than a sequence of a cool day followed by one or two warm days. Hence, the projected variability in temperatures resulting from possible changes in climate are a threat to public health.

5. DISCUSSION AND MODEL EXTENSIONS

In this study, a flexible class of DL models were used to analyze the relationship between heat and mortality in four major metropolitan cities in the U.S. The modeling approach for this analysis used random functions (Gaussian processes) which allow flexibility for the DL func-

tion to take a wide variety of shapes while still constraining the function to decrease to zero smoothly as the lag increased. Additionally, a strong appeal of the proposed methodology was the ability to perform probabilistic inference for the maximum lag L; a contribution previously unseen in the literature. The analysis found some evidence of mortality displacement, particularly in the older populations, suggesting heat may advance death of at-risk individuals. Furthermore, the shape of the estimated DL functions suggest that not only are longer heat-waves a threat to public health but highly variable temperatures are also. Finally, the effect of heat on mortality for the cities considered in this analysis was seen to have a maximum lag of 3 or 4 days.

The analysis here was limited to all non-accidental deaths. However, heat affects various areas of the body differently suggesting different distributed lag functions for different causes of death (Braga et al. 2002). For example, excessive heat exposure may lead to dehydration and renal failure. Thus, a DL function for deaths due to renal failure would look different that the DL function for deaths due to respiratory failure. Furthermore, heat may affect various populations differently; e.g. lower income areas may be more susceptible to heat-related illness due to lack of acceptable housing. Here, all deaths were aggregated and, as such, the ability to distinguish different DL functions for different causes of death and different populations was lost. Applying the proposed methodology to different diseases and populations may give further insight into how heat affects public well-being.

Average daily temperature was used as a covariate to indicate "heat". However, average daily temperature has a number of drawbacks. First, average daily temperature includes temperatures at night and may not adequately represent how hot it is outside during daylight hours when people are most active. Thus, perhaps daily maximum temperature would be a preferred covariate to use in this case. Additionally, humidity (either absolute or relative) may play a major role in explaining morbidity or mortality as high humidity influences how fast water is lost from the body. Along these same lines, a combination of temperature and humidity may be more closely correlated with heat-related illnesses. For example, days with high temperature and high humidity may pose a more serious threat to public health than just high temperature or high humidity alone. These and other questions are a current area of research.

ACKNOWLEDGEMENTS

Protection Agency.

REFERENCES

Anderson, B. G. and Bell, M. L. (2009), "Weather-related mortality: How heat, cold, and heat waves affect mortality in the united states," *Epidemiology*, 20, 205–213.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman and Hall/CRC.

Bell, M. L., Samet, J. M., and Dominici, F. (2004), "Time-series studies of particulate matter," *Annual Review of Public Health*, 25, 247–280.

Braga, A., Zanobetti, A., and Schwartz, J. (2002), "The effect of weather on respiratory and cardiovascular deaths in 12 u.s. cities," *Environmental Health Perspectives*, 110, 859–863.

Braga, A. L., Zanobetti, A., and Schwartz, J. (2001), "The time course of weather-related deaths," *Epidemiology*, 12, 662–667.

Caffo, B. S., Peng, R. D., Dominici, F., Louis, T. A., and Zeger, S. L. (2011), "Parallel mcmc imputation for multiple distributed lag models: a case study in environmental epidemiology," in *The Handbook of Markov chain Monte Carlo*, Chapman and Hall/CRC Press.

Chib, S. and Jeliazkov, I. (2001), "Marginal likelihood from the Metropolis-Hastings output," *Journal of the American Statistical Association*, 96, 270–281.

Cressie, N. and Wikle, C. K. (2011), *Statistics for Spatio-Temporal Data*, Wiley.

Frances, P. H. and van Oest, R. (2004), "On the econometrics of the koyck model," Technical report, Economic Institute, Erasmus University Rotterdam.

Haario, H., Saksman, E., and Tamminen, J. (2001), "An adaptive Metropolis algorithm," *Bernoulli*, 2, 223–242.

Hajat, S., Armstrong, B. G., Gouveia, N., and Wilkinson, P. (2005), "Mortality displacement of heat-related deaths," *Epidemiology*, 16, 613–620.

Handcock, M. S. and Stein, M. L. (1993), "A Bayesian analysis of kriging," *Technometrics*, 35, 403–410.

Handcock, M. S. and Wallis, J. (1994), "An approach to statistical spatial-temporal modeling of meteorological fields (with discussion)," *Journal of the American Statistical Association*, 89, 368–390.

IPCC (2007), *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the fourth assessment report of the Intergovernmental Panel on Climate Change*, Cambridge University Press.

Kovats, R. S. and Hajat, S. (2008), "Heat stree and public health: a critical review," *Annual Review of Public Health*, 29, 41–55.

Li, B., Sain, S., Mearns, L. O., Anderson, H. A., Kovats, S., Ebi, K. L., Bekkedal, M., Kanarek, M. S., and Patz, J. A. (2011), "The impact of extreme heat on morbidity in milwaukee, wisconsin," *Climatic Change*. DOI: 10.1007/s10584-011-0120-y.

Matérn, B. (1986), *Spatial Variation*, 2nd ed., Berlin: Springer-Verlag.

Meehl, G. A. and Tebaldi, C. (2004), "More intense, more frequence, and longer lasting heat waves in the 21st century," *Science*, 305, 994–997.

O'Neill, M. S., Zanobetti, A., and Schwartz, J. (2003), "Modifiers of the temperature and mortality association in seven us cities," *American Journal of Epidemiology*, 157, 1074–1082.

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," *Environmental Health Perspectives*, 119, 701–706.

Peng, R. D., Dominici, F., and Welty, L. J. (2009), "A bayesian hierarchical distributed lag model for estimating the time course of risk hospitalization associated with particulate matter air pollution," *Journal of the Royal Statistical Society Series C*, 58, 3–24.

Samet, J. M., Zeger, S. L., Dominici, F., Curriero, F., Coursac, I., Dockery, D. W., Schwartz, J., and Zanobetti, A. (2000), "The National morbidity, Mortality, and Air Pollution Study Part II: morbidity and mortality from air pollution in the united states," *Research Report Health Effects Institute*, 94, 5–79.

Schwartz, J. (2000), "The distributed lag between air pollution and daily deaths," *Epidemiology*, 11, 320–326.

Tebaldi, C., Hayhoe, K., Arblaster, J. M., and Meehl, G. A. (2006), "Going to the extremes: An intercomparison of model-simulated historical and future changes in extreme events," *Climatic Change*, 79, 185–211.

van Dyk, D. A. and Park, T. (2008), "Partially collapsed Gibbs samplers: Theory and methods," *Journal of the American Statistical Association*, 103, 790–796. DOI: 10.1198/016214508000000409.

Welty, L. J., Peng, R. D., Zeger, S. L., and Dominici, F. (2009), "Bayesian distributed lag models: Estimating the effects of particulate matter air pollution on daily mortality," *Biometrics*, 65, 282–291.

Welty, L. J. and Zeger, S. L. (2005), "Are the acute effects of particulate matter on mortality in the national morbidity, mortality, and air pollution study the result of inadequate control for weather and season? a sensitivity analysis using flexible distributed lag models," *Americal Journal of Epidemiology*, 162, 80–88.

Zanobetti, A., Wand, M. P., Schwartz, J., and Ryan, L. M. (2000), "Generalized additive distributed lag models: Quantifying mortality displacement," *Biostatistics*, 1, 279–292.

Zhang, H. (2004), "Inconsistent estimation and asymptotically equal interpolations in

model-based geostatistics," *Journal of the American Statistical Association*, 99, 250–261.