

# Flexible Distributed Lag Models Using Random Functions With Application to Estimating Mortality Displacement From Heat-Related Deaths

Matthew J. HEATON and Roger D. PENG

Changes in the distribution of ambient temperature, due to climate change or otherwise, will likely have a negative effect on public health. Characterizing the relationship between temperature and mortality is a key aspect of the larger problem of understanding the health effect of climate change. 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 to construct 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. The proposed framework also 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 highly variable temperatures pose a threat to public health. This article has supplementary material online.

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

## 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 et al. 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,

---

Matthew 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, USA (E-mail: [heaton@ucar.edu](mailto:heaton@ucar.edu)). Roger D. Peng is Associate Professor, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA (E-mail: [rpeng@jhsph.edu](mailto:rpeng@jhsph.edu)).

© 2012 International Biometric Society  
*Journal of Agricultural, Biological, and Environmental Statistics*, Volume 17, Number 3, Pages 313–331  
DOI: [10.1007/s13253-012-0097-7](https://doi.org/10.1007/s13253-012-0097-7)

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 et al. 2011; Peng et al. 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.

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  $X_t, \dots, X_{t-M}$ , for some maximum lag  $M$ , via the mean such that

$$g(\mathbb{E}(Y_t | X_t, \dots, X_{t-M})) = \beta_0 + \sum_{\ell=0}^M \theta_\ell X_{t-\ell} \quad (1.1)$$

where  $g(\cdot)$  represents an appropriate link function (e.g. identity, log, etc.). The  $\{\theta_\ell : \ell = 0, \dots, M\}$  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. Health effects using distributed lag models are typically quantified using the parameter  $\beta_1 = \sum_{\ell=0}^L \theta_\ell$  which can be interpreted as the effect of a unit increase in  $X_t$  on  $g(\mathbb{E}(Y_{t^*}))$  for days  $t^* = t, t+1, t+2$ , etc. For example, if  $\beta_1 > 0$  then the left-hand side of (1.1) increases for a unit increase in  $X_t$  for days  $t, t+1, t+2$ , etc.

To highlight only a few previous 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 et al. (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 for the springs and summers (April 1–September 30) of the years 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 et al. 2000). Table 1 provides summary statistics for the data set by city. Figure 1 displays the empirical partial correlation function  $\text{corr}(Y_t, X_{t-\ell} | X_1, \dots, X_{t-\ell-1})$  for  $\ell = 0, \dots, 14$  after accounting for smooth functions of 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

Table 1. Summary statistics for the daily mortality (number of deaths per day) and temperature (in degrees Fahrenheit) data set in each city.  $Q_i$  indicates the  $i$ th quartile where  $i = 1, 2, 3$ .

Statistic	Chicago		Dallas		Los Angeles		New York	
	Mort.	Temp.	Mort.	Temp.	Mort.	Temp.	Mort.	Temp.
Min	4	40	2	56	10	58	10	50
$Q_1$	19	62	12	78	26	66	29	66
$Q_2$	26	70	18	82	35	69	40	73
Mean	33	68	20	82	46	69	50	71
$Q_3$	49	75	27	87	72	71	77	77
Max	84	90	49	96	107	85	121	91

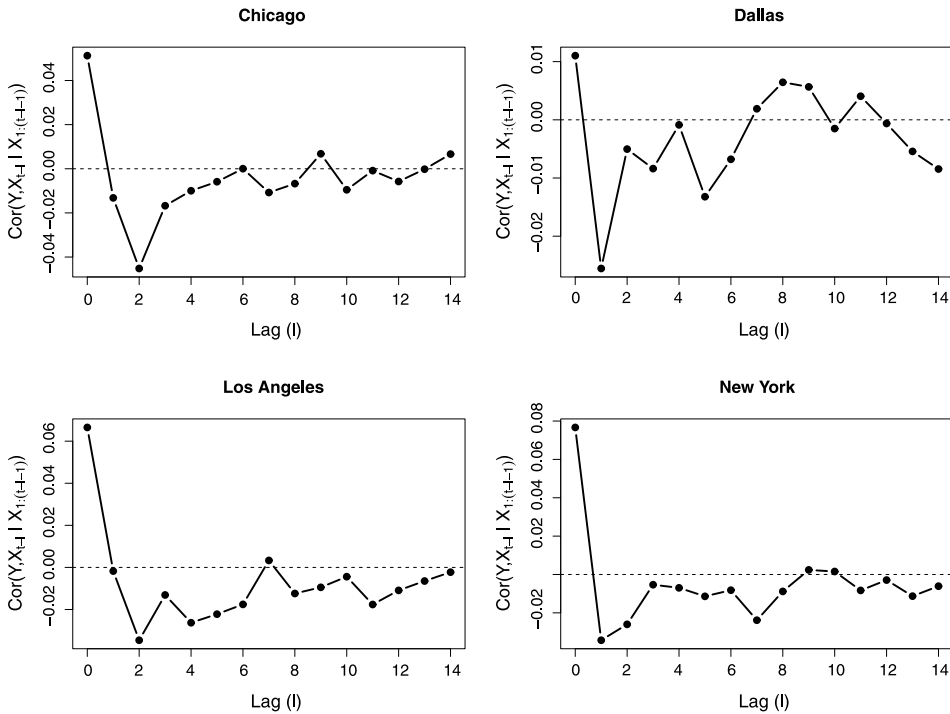


Figure 1. Empirical partial autocorrelation function  $\text{corr}(Y_t, X_{t-l} | X_1, \dots, X_{t-l-1})$  for  $l = 0, \dots, 14$  after accounting for a day of week effect, age, and annual trends for each city in the current study. Note the different scales used on the y-axes.

2 or 3. In contrast, Dallas seems to exhibit resiliency to heat (note the y-axis scale changes in Figure 1) while New York is more susceptible to current day 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  $X_t, \dots, X_{t-M}$  are often highly collinear resulting in inflated standard errors. Solutions to reining in the standard errors include averaging covariates in temporal windows (Curriero et al. 2002; Bell, Samet, and Dominici 2004;

Welty and Zeger 2005; Caffo et al. 2011), constraining  $\{\theta_\ell\}$  to follow a function such as a polynomial (Schwartz 2000) or spline (Zanobetti et al. 2000), or building in strong prior knowledge to borrow strength across the estimates (Welty et al. 2009; Peng, Dominici, and Welty 2009). Similarly to Welty et al. (2009) and Peng, Dominici, and Welty (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 to construct a 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  $\{\theta_\ell : \ell \in \mathcal{D}\}$  for some continuous region  $\mathcal{D}$  follows a Gaussian process if and only if for any finite collection of locations  $\ell_1, \dots, \ell_n$  the vector  $(\theta_{\ell_1}, \dots, \theta_{\ell_n})'$  follows a multivariate Gaussian distribution. Here, the Matérn (Matérn 1986) correlation function is used to induce strong a priori correlation between  $\theta_{\ell_i}$  and  $\theta_{\ell_j}$  to not only control standard errors of  $\theta_\ell$  but also control the smoothness properties of the DL function.

A second challenge in using DL models is selecting a 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). However, 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). A second solution is to assume that  $M$  is known and choose its value based on a priori intuition (Roberts 2005). However,  $M$  is rarely known and determining its value is typically a goal of the analysis. A third option is to allow  $M$  to be large enough to capture all possible effects of interest and use credible (or confidence) intervals for  $\{\theta_t\}$  to infer a smaller “true” maximum lag (e.g. the smallest lag such that all credible intervals for subsequent lagged effects contain zero). With this approach, the collinearity of  $X_t, \dots, X_{t-M}$  mentioned above creates more problems because the standard errors of  $\{\theta_t\}$  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 finite large lag  $M$  is first chosen to specify a finite DL model. Next, a second “true” maximum lag  $L$  is introduced such that  $\theta_\ell = 0$  if  $\ell \geq L$ . Given  $M$  and  $L$ , a model for  $\theta_0, \dots, \theta_{L-1}$  is constructed via the conditional properties of the multivariate normal distribution. Specifically, a valid model specification is provided for all  $L = 1, \dots, M$  enabling inference for  $L$ . By treating the maximum lag as a parameter of interest, estimates of lagged effects are averaged over different values for  $L$ . More modeling details and discussion on implementing a Markov chain Monte Carlo (MCMC) algorithm for posterior inference is provided in Section 2.

The primary contributions of this article are outlined as follows: (i) the use of random functions (specifically, Gaussian processes) as a 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  $\ell \geq L$ ; and (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 an 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 shortcomings 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,  $Y_t$  a response variable of interest observed at time  $t$ , and  $X_t$  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. Since 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

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

where, as in the Introduction section,  $g(\cdot)$  denotes an appropriate link function (e.g. identity, log, etc.) and  $\{\theta_\ell : \ell = 0, \dots, M\}$  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  $\theta_\ell$ . For DL models, the sum  $\beta_1 = \sum_{\ell=0}^M \theta_\ell$  is often of interest and quantifies the effect of a unit increase in  $X_t$  on  $g(\mathbb{E}(Y_{t^*}))$  for days  $t^* = t, t+1, t+2$ , etc. Note, however, that the value of  $\beta_1$  alone can be deceiving. For example, notice that  $\beta_1$  can be approximately 0, suggesting a null effect between covariates and response, yet  $\theta_\ell \neq 0$  for all  $\ell$  suggesting that *all* past covariates affect the response. Thus, the distributed lag function  $\{\theta_\ell : \ell = 0, \dots, M\}$  is the primary quantity of interest and  $\beta_1$  is a useful univariate summary.

In practice, the value of  $M$  in (2.1) is often set large enough such that  $\theta_\ell$  for  $\ell = 0, \dots, M$  captures all lagged effects of interest and confidence (or credible) intervals are used to determine a “true” maximum lag  $L \leq M$  at which  $\theta_\ell = 0$  for  $\ell \geq L$ . That is, if  $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2)'$  where  $\boldsymbol{\theta}_1 = (\theta_0, \dots, \theta_{L-1})$  and  $\boldsymbol{\theta}_2 = (\theta_L, \dots, \theta_M)$ , then, conditional on  $L$ ,  $\boldsymbol{\theta}_2 = \mathbf{0}$ . The modeling strategy proposed here also considers  $M$  to be fixed at some large value yet estimates the maximum lag parameter  $L \leq M$  from the data. For example, in the heat-related mortality example discussed in the Introduction section and analyzed in Section 4,  $M = 60$  because intuition states that, with near-certainty, the effect of the average daily temperature two months ago should have no effect on mortality today, yet, a maximum lag  $L \leq M$  is estimated where the effect of average daily temperature at lags greater than  $L$  is zero.

For a fixed  $L$ , prior knowledge of  $\boldsymbol{\theta}_1$  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 of at-risk 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 be trivial 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  $\theta_\ell \rightarrow 0$  smoothly as  $\ell \uparrow L$ . Thus, a model for  $\theta_1$  should be flexible at small lags yet constrained at larger lags.

To motivate the modeling strategy for  $\theta_1 | L$ , first consider a model for  $\theta = (\theta_0, \dots, \theta_M)'$  where, again,  $M$  is fixed at a large value to ensure that all desired effects are included in  $\theta$ . Let

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

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

$$\rho_\theta(\|i - j\|) = \left( \frac{1}{2^{\nu-1} \Gamma(\nu)} \right) (\phi_\theta \|i - j\|)^\nu K_\nu(\phi_\theta \|i - j\|) \quad (2.3)$$

with  $\nu$  a smoothness parameter,  $\phi_\theta$  a decay parameter and  $K_\nu$  the modified Bessel function of the second kind of order  $\nu$ . 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,  $\nu$  is used to control the smoothness properties of the distributed lag function. Following Welty et al. (2009) and Peng, Dominici, and Welty (2009), the DL function is hypothesized to be smooth over time; thus, here  $\nu$  is fixed at 3 a priori. Borrowing a term from spatial statistics, the parameter  $\phi_\theta$  is a decay parameter and controls the correlation at fixed temporal distances. In other words, as  $\phi_\theta$  increases,  $\text{corr}(\theta_\ell, \theta_{\ell+u})$  decreases for fixed  $u \in \mathbb{R} \setminus 0$ . Finally,  $\sigma_\theta^2$  is a variance term that controls the maximum (and minimum) height of the DL function. For example, if  $\sigma_\theta^2$  is small (large) then  $\max_\ell \{\theta_\ell : \ell \in \{0, \dots, M\}\}$  will, with high probability, also be small (large) in absolute value. Importantly, Zhang (2004) showed that the parameter  $\phi_\theta$  is not consistently estimable and fixing  $\phi_\theta$  a priori while estimating  $\sigma_\theta^2$  is sufficient. Thus  $\phi_\theta$  will be treated as fixed for the remainder of this article and, for convenience,  $\mathbf{R}_\theta(\phi_\theta)$  will be written as simply  $\mathbf{R}_\theta$ . In initial investigations of the methodology proposed here,  $\phi_\theta$  was treated as an unknown parameter but results were insensitive to its value. Specifically, several values for  $\phi_\theta$  were used and each value resulted in equivalent estimates of the model parameters. As a general guideline for choosing  $\phi_\theta$ , an intuitive approach is to solve  $\text{corr}(\theta_\ell, \theta_{\ell+u} | \phi_\theta) = \rho_\theta(u | \phi_\theta) = 0.05$  for  $\phi_\theta$  where  $u$  is a user-specified distance at which the correlation between lagged effects decays to 0.05. Initial investigations suggest that any  $u \geq 3$  resulted in an equivalent model fit.

Having defined a very flexible model for  $\theta$  in (2.2) above, a model for  $\theta_1 | L$  which adheres to the physical constraint that  $\theta_\ell \rightarrow 0$  smoothly as  $\ell \uparrow L$  can now be constructed via conditioning in the following manner. Partition  $\mathbf{R}_\theta$  in (2.2) as

$$\mathbf{R}_\theta = \begin{pmatrix} \mathbf{R}_{\theta,11} & \mathbf{R}_{\theta,12} \\ \mathbf{R}_{\theta,21} & \mathbf{R}_{\theta,22} \end{pmatrix}$$

such that  $\text{corr}(\theta_i, \theta_j) = \mathbf{R}_{\theta,ij}$  for  $i, j = 1, 2$ . Conditioning (2.2) on  $L$ , a distribution for  $\theta_1 | L$  follows directly from the properties of the multivariate normal distribution. That is,

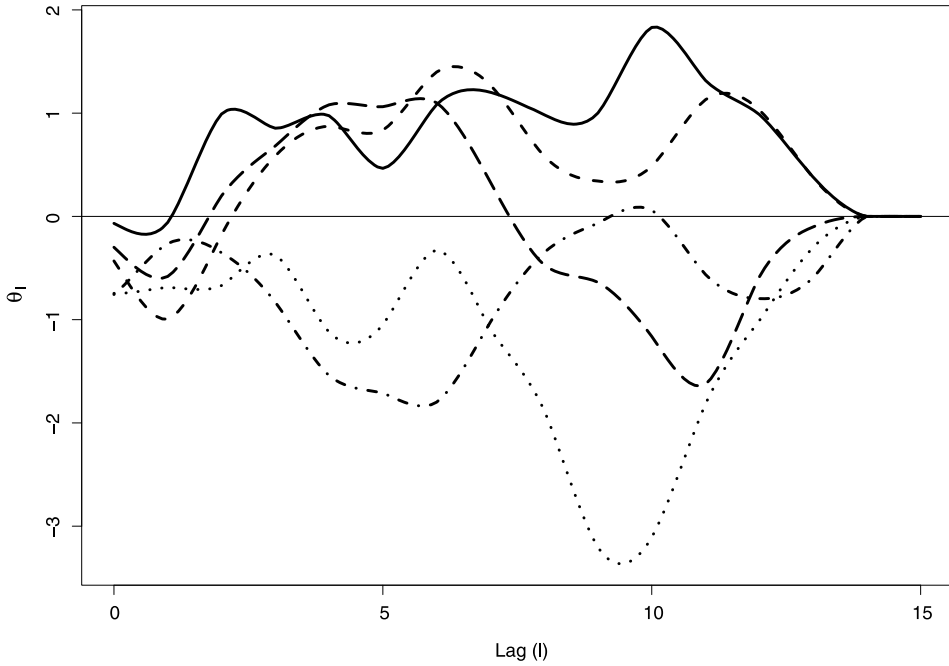


Figure 2. The non-zero portion (i.e.  $\theta_1$ ) of five DL functions randomly drawn according to the random function DL model with  $M = 21$ ,  $L = 14$ ,  $\sigma_\theta^2 = 1$ , and  $\phi_\theta = 0.18$ . Notice that the random function DL model allows for a wide variety of DL functions while constraining each of these to approach 0 as the lag increases.

let  $\theta_1|L \sim \mathcal{N}(\mathbf{0}, \sigma_\theta^2 \mathbf{R}_\theta^*(L))$  where  $\mathbf{R}_\theta^*(L) = \mathbf{R}_{\theta,11} - \mathbf{R}_{\theta,12} \mathbf{R}_{\theta,22}^{-1} \mathbf{R}_{\theta,21}$  is now  $L \times L$ . By conditioning on  $L$ , notice that the first few elements of  $\theta_1$  (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  $\theta_1$  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),  $M$  should be 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 allow  $\theta_\ell \rightarrow 0$  smoothly as  $\ell \uparrow L$ . For example, notice that if  $L = M$  then  $\theta_2 = \theta_M$  is univariate and conditioning on  $\theta_M = 0$  will not adequately capture the knowledge that  $\theta_{M+1} = \theta_{M+2} = \dots = 0$ . To avoid this,  $L$ , a priori, will be assumed to follow a discrete uniform prior where  $\max_\ell \{\ell : \Pr(L = \ell) > 0\}$  will be much less than  $M$ .

**2.2. MODEL SPECIFICATION AND PARAMETER ESTIMATION**

Let  $\mathbf{Y} = (Y_1, \dots, Y_T)'$  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  $\mathbf{Y}$  can be written as

$$\mathbf{Y} | \mathbf{X} \sim \mathcal{N}(\beta_0 \mathbf{1}_T + \mathbf{X}\boldsymbol{\theta}, \sigma_Y^2 \mathbf{I}_T), \tag{2.4}$$

where  $\mathbf{1}_T$  is the length  $T$  column vector of ones,  $\mathbf{X}$  is the  $T \times (M + 1)$  design matrix with  $t$ th row  $(X_t, X_{t-1}, \dots, X_{t-M})$ ,  $\boldsymbol{\theta} = (\theta_0, \dots, \theta_M)'$ ,  $\sigma_Y^2$  is the error variance, and  $\mathbf{I}_T$  is the rank  $T$  identity matrix. Notice that (2.4) assumes that all of the temporal correlation in  $\mathbf{Y}$  is explained through  $\mathbf{X}$ . Admittedly, this may be overly simplistic because  $\mathbf{Y}$  may have excess temporal correlation beyond that explained by  $\mathbf{X}$ ; however, this extension is not considered here and focus is placed on estimating the DL function  $\boldsymbol{\theta}$ . Partition  $\mathbf{X} = [\mathbf{X}_{1:L}, \mathbf{X}_{(L+1):(M+1)}]$  such that  $\mathbf{X}_{a:b}$  are columns  $a$  through  $b$  of  $\mathbf{X}$ . Conditioning (2.4) on  $L$  gives

$$\mathbf{Y} \mid \mathbf{X}, L \sim \mathcal{N}(\beta_0 \mathbf{1}_T + \mathbf{X}_{1:L} \boldsymbol{\theta}_1, \sigma_Y^2 \mathbf{I}_T) \quad (2.5)$$

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

The unknown parameters in model (2.5) 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 straightforward. 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  $L, \beta_0, \sigma_\theta^2$ , and  $\sigma_Y^2$  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 an MH algorithm but complications immediately arise. For example, consider a proposal value  $L^*$  such that  $L^* > L$ . Evaluating the likelihood in this case would require values for  $\theta_L, \dots, \theta_{L^*-1}$  which are not available in the current draw of  $\boldsymbol{\theta}_1$ . Likewise, if  $L^* < L$ , it then may not be reasonable to simply “zero-out” the elements of  $\boldsymbol{\theta}_1$  to evaluate (2.5) given  $L^*$ . Thus rejection of  $L^*$  may result *not* because the value of  $L^*$  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  $[\boldsymbol{\theta}_1, L \mid \mathbf{Y}, \mathbf{X}, \sigma_Y^2, \sigma_\theta^2, \beta_0] = [\boldsymbol{\theta}_1 \mid L, \mathbf{Y}, \mathbf{X}, \sigma_Y^2, \sigma_\theta^2, \beta_0][L \mid \mathbf{Y}, \mathbf{X}, \sigma_Y^2, \sigma_\theta^2, \beta_0]$  where  $[\cdot]$  denotes a general distribution. 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  $\{1, 2, \dots, L_{\max}\}$  where  $L_{\max} < M$  and, often,  $L_{\max} \ll M$  (see the discussion above). Using a discrete uniform prior for  $L$  yields

$$[L = \ell \mid \mathbf{Y}, \mathbf{X}, \sigma_Y^2, \beta_0, \sigma_\theta^2] = \frac{\mathcal{N}(\mathbf{Y}; \beta_0 \mathbf{1}_T, \sigma_\theta^2 \mathbf{X}_{1:\ell} \mathbf{R}_\theta^*(\ell) \mathbf{X}'_{1:\ell} + \sigma_Y^2 \mathbf{I}_T)}{\sum_{i=1}^{L_{\max}} \mathcal{N}(\mathbf{Y}; \beta_0 \mathbf{1}_T, \sigma_\theta^2 \mathbf{X}_{1:i} \mathbf{R}_\theta^*(i) \mathbf{X}'_{1:i} + \sigma_Y^2 \mathbf{I}_T)} \quad (2.6)$$



where  $\mathcal{N}(x; \mu, \sigma^2)$  denotes the Gaussian distribution function with mean  $\mu$  and variance  $\sigma^2$  evaluated at  $x$ . Using (2.6),  $L$  can be drawn directly, following which  $\theta_1$  can be drawn directly from its complete conditional distribution as discussed above.

Since (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

$$Y_t | L, \mathbf{x}_t(L), \beta_1, \theta_1, \sigma_\theta^2 \stackrel{ind}{\sim} \mathcal{P}(\exp\{\beta_0 + \mathbf{x}'_t(L)\theta_1\}) \tag{2.7}$$

where  $\mathcal{P}(\mu)$  represents the Poisson distribution with mean  $\mu$  and  $\mathbf{x}'_t(L) = (X_t, X_{t-1}, \dots, X_{t-L+1})'$ . The apparent difficulty with using the Poisson, rather than the Gaussian, distribution is in loss of conjugacy for the parameters  $\beta_0$  and  $\theta_1$  (notice that  $\sigma_\theta^2$  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  $\beta_0$  and  $\theta_1$ , 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  $[L = \ell | \mathbf{Y}, \mathbf{X}, \beta_0, \sigma_\theta^2]$  given by

$$\frac{\int_{\mathbb{R}^\ell} \prod_{t=1}^T \mathcal{P}(Y_t; \exp\{\beta_0 + \mathbf{x}'_t(\ell)\theta_{1:\ell}\}) \mathcal{N}(\theta_{1:\ell}; \mathbf{0}, \sigma_\theta^2 \mathbf{R}_\theta^*(\ell)) d\theta_{1:\ell}}{\sum_i \int_{\mathbb{R}^i} \prod_{t=1}^T \mathcal{P}(Y_t; \exp\{\beta_0 + \mathbf{x}'_t(i)\theta_{1:i}\}) \mathcal{N}(\theta_{1:i}; \mathbf{0}, \sigma_\theta^2 \mathbf{R}_\theta^*(i)) d\theta_{1:i}} \tag{2.8}$$

is no longer available in closed form where  $\mathcal{P}(x; \mu)$  is the Poisson distribution with mean  $\mu$  evaluated at  $x$  and  $\theta_{1:i} = (\theta_0, \dots, \theta_{i-1})'$ . Notice that the numerator in (2.8) is simply  $\mathbb{E}_{\theta_{1:\ell}}(\prod_{t=1}^T \mathcal{P}(Y_t; \exp\{\beta_0 + \mathbf{x}'_{t\ell}\theta_{1:\ell}\}))$  where  $\theta_{1:\ell} \sim \mathcal{N}(\mathbf{0}, \sigma_\theta^2 \mathbf{R}_\theta^*(\ell))$  and could be calculated using Monte Carlo integration. Monte Carlo integration, however, could be imprecise if the product  $\prod_{t=1}^T \mathcal{P}(Y_t; \exp\{\beta_0 + \mathbf{x}'_t(\ell)\theta_{1:v}\})$  is highly concentrated relative to  $\mathcal{N}(\mathbf{0}, \sigma_\theta^2 \mathbf{R}_\theta^*(\ell))$ . Rather, the approach used here is to calculate the marginal posterior probability  $[L = \ell | \mathbf{Y}, \mathbf{X}]$  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  $[L | \mathbf{Y}, \mathbf{X}]$  followed by sampling  $\beta_0, \theta_1$  and  $\sigma_\theta^2$  from their respective complete conditional distributions via Metropolis steps.

As a final point for this section, note that in both the Gaussian and Poisson models (Equations (2.5) and (2.7), respectively) sampling  $L$  is computationally challenging. In the case of (2.5), note that calculating (2.6) requires inverting (and calculating the determinant of) the  $T \times T$  dense covariance matrix  $\sigma_\theta^2 \mathbf{X}_{1:\ell} \mathbf{R}_\theta^*(\ell) \mathbf{X}'_{1:\ell} + \sigma_Y^2 \mathbf{I}_T$  for all  $\ell \in \{\ell : \Pr(L = \ell) > 0\}$  at each iteration of the MCMC algorithm. However, in the Gaussian setting, this computational burden is largely alleviated through the use of the Sherman–Morrison–Woodbury inversion formula. In the case of (2.7), calculating  $[L | \mathbf{Y}, \mathbf{X}]$  using the methods of Chib and Jeliazkov (2001) requires initially running a separate MCMC chain for each  $\ell \in \{\ell : \Pr(L = \ell) > 0\}$ . Hence, for large  $T$  or large  $\#\{\ell : \Pr(L = \ell) > 0\}$  (or both), obtaining the complete conditional distribution for  $L$  can be a time-consuming task. One possible approach to partially alleviate this computational task is to select only a few “practical values” for the prior distribution  $\{\ell : \Pr(L = \ell) > 0\}$  of most interest in

Table 2. Simulation study model scenarios where  $\phi(x | \mu, \sigma)$  is the Gaussian probability density function with mean  $\mu$  and standard deviation  $\sigma$  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)	$Y_t \stackrel{ind}{\sim} \mathcal{N}(25 + 0.17 \sum_{\ell=0}^{13} \sin(2\pi \ell/14) X_{t-\ell}, 25)$
(b)	$Y_t \stackrel{ind}{\sim} \mathcal{P}(\exp\{1.75 + 0.03 \sum_{\ell=0}^{13} \sin(2\pi \ell/14) X_{t-\ell}\})$
(c)	$Y_t \stackrel{ind}{\sim} \mathcal{N}(-55 + 1.75 \sum_{\ell=0}^{13} \phi(\ell   4, 14/3) X_{t-\ell}, 25)$
(d)	$Y_t \stackrel{ind}{\sim} \mathcal{P}(\exp\{-15 + 0.3 \sum_{\ell=0}^{13} \phi(\ell   4, 14/3) X_{t-\ell}\})$
(e)	$Y_t \stackrel{ind}{\sim} \mathcal{N}(125 + 1.75 \sum_{\ell=0}^{13} \phi^{(1)}(\ell   -3, 7) X_{t-\ell}, 25)$
(f)	$Y_t \stackrel{ind}{\sim} \mathcal{P}(\exp\{18 + 0.3 \sum_{\ell=0}^{13} \phi^{(1)}(\ell   -3, 7) X_{t-\ell}\})$

the study under consideration. For example, in the heat-related study considered here, distinguishing between a 2- or 3-day maximum lag time may be of high interest. Yet, distinguishing between two longer-term maximum lag times may be less important (e.g., distinguishing between a maximum lag time of 13 or 14 days has no real practical significance). For the heat-related mortality application in Section 4, the discrete uniform prior on  $\{1, 3, 5, 7, 9, 11, 14, 21\}$  is considered. However, for the simulation study presented in Section 3, an interesting question is to determine how well the value of  $L$  can be estimated using the data. Hence, in Section 3 the prior for  $L$  is assumed to be a discrete uniform distribution on  $\{1, 2, \dots, 21\}$ .

Details of the MCMC algorithms used in this article to fit (2.5) and (2.7) are provided in the supplementary material.

### 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 2. 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 2 are a  $3 \times 2$  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 data set which is of primary interest in this article.

The model given by Equation (2.5) was fit to each data set in scenarios (a), (c), and (e) while (2.7) 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  $\{1, 2, \dots, 21\}$ . The value of  $\phi_\theta$  was fixed such that  $\rho_\theta(5) = 0.05$ . For the Gaussian models, the prior distribution for  $\beta_0$  was assumed to be  $\mathcal{N}(0, 100^2)$  while for the Poisson models,  $\beta_0 \sim \mathcal{N}(0, 10^2)$ . Note that for the Poisson case, the value of  $\beta_0$  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  $\mathcal{IG}(\cdot)$ ) prior distributions were assumed for both  $\sigma_\theta^2$  and  $\sigma_Y^2$ ; specifically,  $\sigma_\theta^2$  and  $\sigma_Y^2$  were assumed to be a priori independent  $\mathcal{IG}(3, \text{rate} = 3)$  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, Saksman, and Tamminen (2001). An analysis of the posterior draws using trace plots suggested that the Gaussian and Poisson models required approximately 5000 and 10,000 draws, respectively, as burn-in. A single data set from scenarios (a), (c), and (e) took approximately 45 minutes of computation time. A single data set from scenarios (b), (d), and (f), took approximately 3.5 hours of computation time due to the large number of values for  $L$  being considered by the prior distribution  $[L]$ .

### 3.2. STUDY RESULTS

Figure 3 displays the 100 marginal posterior mean DL functions (that is,  $\mathbb{E}(\theta_\ell | \mathbf{Y}, \mathbf{X})$  for  $\ell = 0, \dots, 21$ ) in each simulation scenario with the true DL function overlaid. According to Figure 3, on average, the model is finding the correct DL function with the possible exception of scenarios (d) and (f) in which the estimated DL functions are tapering off to zero prematurely. For scenarios (d) and (f), however, note that the final few lagged effects are near zero (the largest being 0.009 in absolute value) and are difficult to distinguish from zero. Figure 4 displays the empirical coverage (EC) rates of 95 % central credible intervals for  $\theta_\ell$  (i.e., the proportion of 95 % credible intervals containing the true value for  $\theta_\ell$ ). In general, the EC rates are close to the nominal rate of 0.95 yet the certainty is slightly understated in scenarios (a) and (e).

Table 3 displays the empirical bias (i.e.  $(100)^{-1} \sum_{i=1}^{100} (\hat{\beta}_{1i} - \beta_1)$  where  $\hat{\beta}_{1i}$  is the posterior mean of  $\beta_1$  for the  $i$ th simulated data set), root mean-square error (i.e.  $\sqrt{0.01 \sum_{i=1}^{100} (\hat{\beta}_{1i} - \beta_1)^2}$ ), coverage rate (i.e., the proportion of the 100 credible intervals for  $\beta_1$  which contain the true value of  $\beta_1$ ) and width of a 95 % central credible interval for the cumulative effect of *all* past covariates given by  $\beta_1 = \sum_{\ell=0}^L \theta_\ell$ . The values in parentheses are the empirical standard deviations associated with the values. According to Table 3, the model fits seem to recover to correct value (indicated by low bias) and have near-nominal coverage rates of 95 %.

For the Gaussian likelihood scenarios (a), (c), and (e), the bias of the posterior mean for  $\sigma_Y^2$  was  $-0.03$ ,  $-0.25$ , and  $0.12$ , respectively. The root mean-square error of the posterior mean for  $\sigma_Y^2$  was 3.59, 3.54, and 3.34. The 95 % credible interval coverage for  $\sigma_Y^2$  was

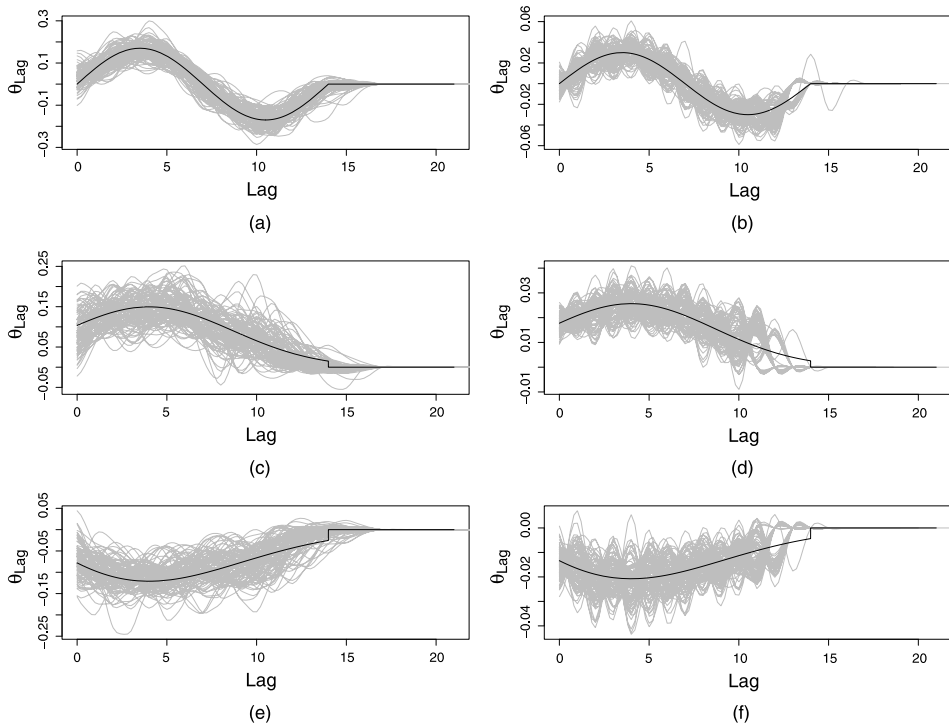


Figure 3. The true distributed lag functions (solid black lines) for each simulation scenario along with the 100 posterior mean DL functions for each simulated data set (gray lines).

0.93, 0.91, and 0.96. Thus, the model is correctly recovering the variance parameter  $\sigma_Y^2$  across the three Gaussian likelihood scenarios.

Table 4 displays the average, across simulated data sets, posterior probability distribution  $[L | Y, X]$  for each simulation scenario. For the Gaussian likelihood scenarios (a), (c), and (e), the posterior distribution spreads the posterior mass for  $L$  over values between  $L = 11$  and  $L = 15$ . For each of these scenarios, however, the posterior mean for  $L$  was 14.59, 13.22, and 13.58, respectively. Hence, on average, the value of  $L$  was chosen correctly. As mentioned above, for scenarios (d) and (f) the value of  $L$  was chosen to be too small with posterior means of 12.09 and 12.46, respectively. For these scenarios, the final lagged effects were near zero (the largest, in absolute value, being 0.009) such that distinguishing these effects from zero is very difficult given the amount of data and noise variability for these scenarios.

For comparison, two other DL models were fit to each simulated data set. First, an unconstrained DL model (i.e. non-informative independent prior distributions were used for each lagged effect  $\theta_i$ ) was fit for each of  $L = 1, \dots, 21$  and the model with the smallest Bayesian information criterion (BIC) was used for the analysis. Second, a “constrained” DL model (i.e. the model presented in Section 2.1 with  $L = M$ ) was fit and the value of  $L$  was chosen to be the smallest lag such that 95 % credible intervals for all subsequent lagged effects contained zero. The exact simulation results are provided in the supplementary materials and only a brief summary of these results is provided here to highlight the

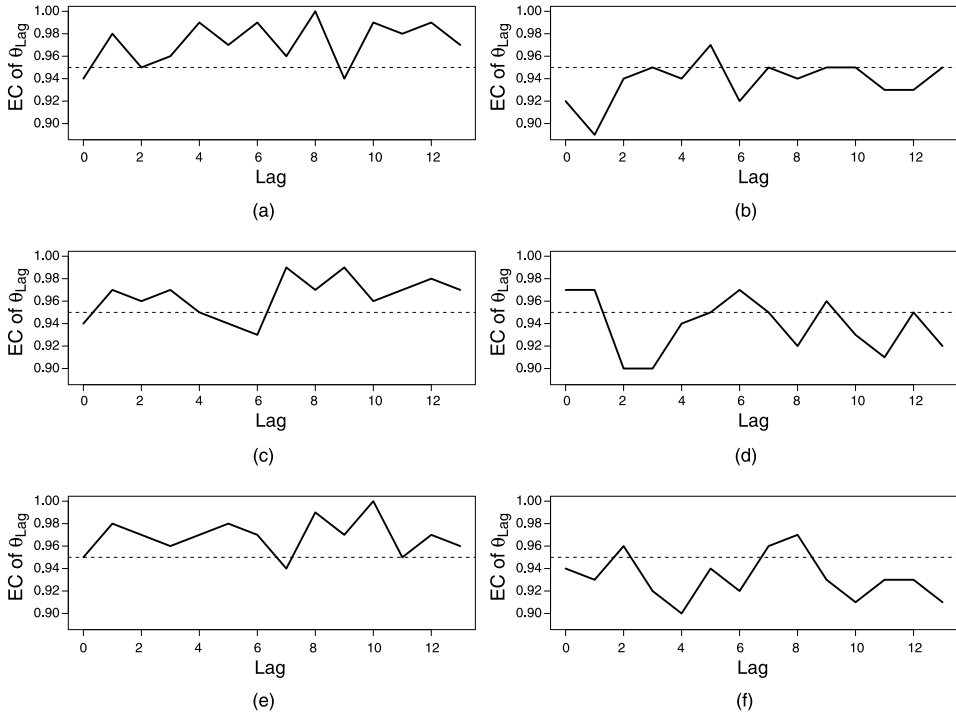


Figure 4. Empirical coverage (EC) rates of 95 % central credible intervals for  $\theta_\ell$  (solid lines). The dashed horizontal line indicates the nominal rate of 95 %. The empirical rates are near the nominal rates of 0.95 but the certainty is slightly understated in scenarios (a) and (e).

Table 3. 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  $\beta_1$  was 0.00, 0.00, 1.42, 0.24,  $-1.22$ , and  $-0.20$  for scenarios (a)–(f), respectively. The values in parentheses are the standard deviations of the summaries. The model fits seem to be recovering the correct parameter with near-nominal coverage.

Scenario	Bias	RMSE	Coverage	Width
(a)	0.01 (0.11)	0.11 (0.02)	0.96 (0.02)	0.47 (0.03)
(b)	0.00 (0.01)	0.01 ( $1e-3$ )	0.95 (0.02)	0.03 ( $2e-3$ )
(c)	$-0.06$ (0.11)	0.13 (0.02)	0.92 (0.03)	0.46 (0.03)
(d)	$-0.01$ ( $7e-3$ )	0.01 ( $1e-3$ )	0.94 (0.04)	0.03 ( $1e-3$ )
(e)	0.06 (0.12)	0.13 (0.02)	0.91 (0.02)	0.47 (0.03)
(f)	0.00 ( $3e-3$ )	0.00 ( $2e-4$ )	0.91 (0.02)	0.02 ( $1e-3$ )

advantages of the modeling strategy in Section 2.1. First, in terms of estimating  $\beta_1$ , the unconstrained DL model often overstated the certainty associated with this parameter in that 95 % credible intervals had less than nominal coverage. In contrast, the constrained DL model attained near-nominal coverage for  $\beta_1$ . When estimating  $L$ , the unconstrained DL model would often choose values that were too small. Alternatively, the constrained DL model would often choose values for  $L$  which were too large (particularly for the

Table 4. Average, across simulated data sets, posterior probability distribution  $[L | Y, X]$ . With the exception of scenarios (d) and (f), the model was able to correctly recover  $L = 14$ . For scenarios (d) and (f), the majority of posterior mass was on values of  $L$  that were smaller than 14.

Scenario	$[L   Y, X]$								
	<11	11	12	13	14	15	16	17	>17
(a)	0.000	0.001	0.038	0.200	0.274	0.211	0.161	0.105	0.010
(b)	0.000	0.000	0.008	0.279	0.690	0.015	0.008	0.000	0.000
(c)	0.051	0.151	0.215	0.189	0.147	0.107	0.080	0.061	0.000
(d)	0.036	0.173	0.548	0.212	0.031	0.002	0.000	0.000	0.000
(e)	0.024	0.115	0.193	0.191	0.166	0.129	0.102	0.080	0.000
(f)	0.000	0.054	0.470	0.427	0.045	0.003	0.000	0.000	0.000

Poisson likelihood scenarios). Hence, compared to these alternative models, the DL model presented in Section 2.1 is superior in terms of flexibility and ability to recover parameters.

#### 4. APPLICATION TO HEAT-RELATED DEATHS

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

$$Y_{cayt} \stackrel{ind}{\sim} \mathcal{P}(\exp\{\beta_{ca0} + \mathbf{x}'_{cyr}(L_{ca})\boldsymbol{\theta}_{ca1} + \mathbf{z}'_{cayt}\boldsymbol{\gamma}_{cay}\})$$

where  $\beta_{ca0}$  is the intercept for (city, age category) =  $(c, a)$ ,  $\mathbf{x}'_{cyr}(L_{ca})$  is the vector of lagged temperatures with  $M = 60$ ,  $L_{ca}$  is the (unknown) maximum lag for city  $c$  and age group  $a$ ,  $\boldsymbol{\theta}_{ca1} = (\theta_{ca10}, \dots, \theta_{ca1(L_{ca}-1)})'$  is the vector of DL coefficients,  $\mathbf{z}_{cayt}$  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  $\boldsymbol{\gamma}_{cay}$ . The partial correlation plots in Figure 1 suggest that the maximum lag is most likely no greater than 14 days. Thus, the prior distribution for the maximum lag for city  $c$  and age group  $a$  was assumed to be discrete with equal probability on the values  $L_{ca} \in \{1, 3, 5, 7, 9, 11, 14, 21\}$ . The value of  $\phi_\theta$  was chosen such  $\text{corr}(\theta_\ell, \theta_{\ell+5}) \approx 0.05$  where the correlation function was given in Equation (2.3). 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 (see the supplementary material for details of the algorithm). The MCMC algorithm with these settings required approximately 100 minutes to run.

The posterior probability  $[L_{ca} = 3 | Y, X]$  was at least 0.95 but typically upwards of 0.99. The older than 75 age category in Chicago placed 5 % posterior mass at  $L_{ca} = 5$ . These results seem to coincide with the findings of Braga, Zanobetti, and Schwartz (2001) and Braga, Zanobetti, and Schwartz (2002) but contradict those of Hajat et al. (2005) who

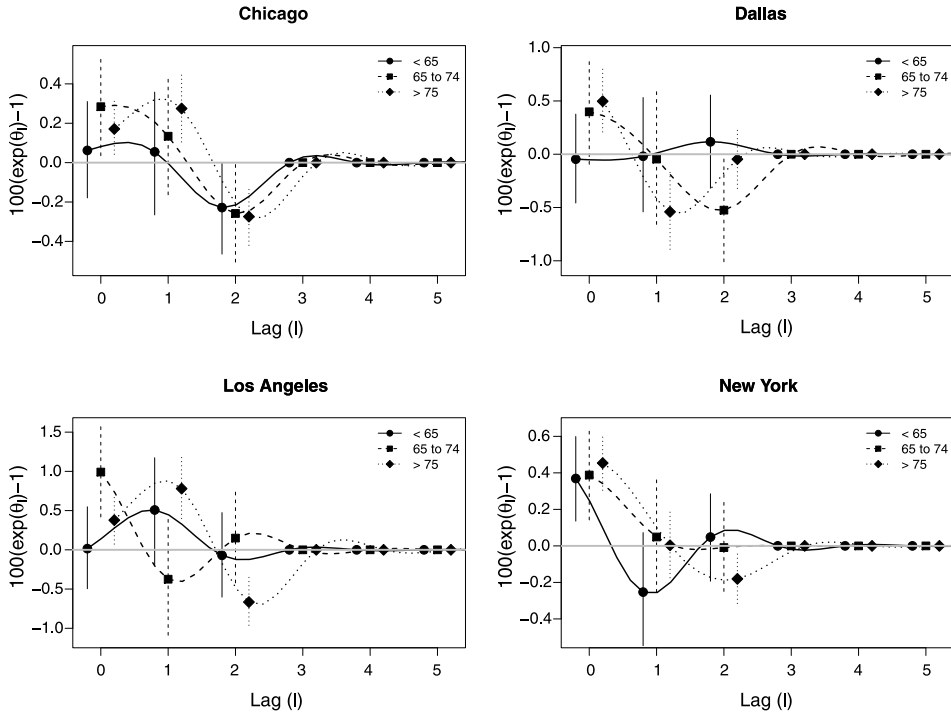


Figure 5. Estimated (i.e. posterior mean) DL function by city and age group expressed in terms of percent increase in mortality due to unit increase in temperature ( $100 \times (\exp\{\theta_l\} - 1)$ ) and 95 % credible intervals. Most cases suggest a mortality displacement or harvesting effect where 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.

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.  $100 \times (\exp\{\theta_l\} - 1)$ ) along with 95 % credible intervals. 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, most cases gave evidence that heat advances the death of at-risk individuals by a few days. Comparing across age groups, the younger age groups seem to have the smallest effects for all cities except New York. However, the credible intervals often overlap suggesting that different age groups may have the same DL function.

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 may already be adapted 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

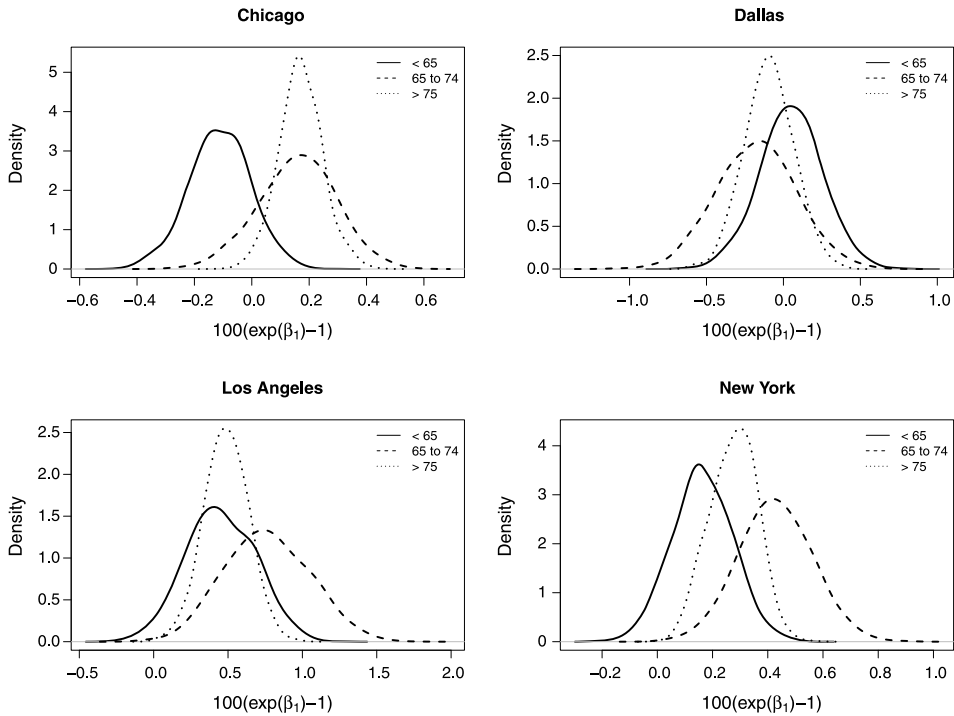


Figure 6. Posterior density of  $100 \times (\exp\{\beta_1\} - 1)$ ; i.e. percent increase in mortality due to a unit increase in temperature for the past  $L$  time periods by city and age group. In most instances, the majority of the density curve is greater than zero suggesting that a uniform increase in temperature would result in more heat-related deaths. Dallas (Los Angeles) seems to be the least (most) susceptible to heat effects.

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 curves in Figure 5 suggest that heat only advances death of highly susceptible persons in Dallas; yet, a mortality displacement effect 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, while successively high temperatures on neighboring days is a problem (as evidenced by positive cumulative effects), highly variable temperatures may also be a threat to public health. To see this, notice that the shape of many of the DL functions (particularly those corresponding to the  $>75$  age group) begin with a positive peak for 1–2 days followed by a negative valley and, finally, a small positive effect. This shape of the DL function suggests that the expected number of deaths from three successive high-temperature days (say, three consecutive days of 90-degree temperatures) is less than a more variable sequence of temperatures (say, a three-day temperature sequence of 90, 90, 70) because the negative effect is smaller in the more variable case.

## 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 allowed for the DL function 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, Zanobetti, and Schwartz 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 than 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.

## SUPPLEMENTAL MATERIALS

**Algorithms:** Details of the MCMC algorithm used to fit the flexible DL models described in this paper. (PDF)

**Additional Simulation Results:** Results from fitting an unconstrained and a constrained DL model to the simulated data sets. (PDF)

## ACKNOWLEDGEMENTS

The authors wish to thank two anonymous reviewers whose comments strengthened the article. NCAR is managed by the University Corporation for Atmospheric Research under the sponsorship of the National Science

Foundation. The work of M. J. Heaton was partially done while a graduate student at Duke University. R. D. Peng was supported by NIH Awards R21ES020152, R01ES019560, P30ES03819, and R01ES012054, and EPA grant RD-83241701. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the U.S. Environmental Protection Agency.

[Published Online August 2012.]

## 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*, London: 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. L., Zanobetti, A., and Schwartz, J. (2001), "The Time Course of Weather-Related Deaths," *Epidemiology*, 12, 662–667.
- (2002), "The Effect of Weather on Respiratory and Cardiovascular Deaths in 12 U.S. Cities," *Environmental Health Perspectives*, 110, 859–863.
- 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*, London: Chapman and Hall/CRC Press, pp. 493–511.
- Chib, S., and Jeliakov, 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*, New York: Wiley.
- 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, 80–87.
- 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: Cambridge University Press.
- Kovats, R. S., and Hajat, S. (2008), "Heat Stress 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.
- Meehl, G. A., and Tebaldi, C. (2004), "More Intense, More Frequent, 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. Applied Statistics*, 58, 3–24.
- Roberts, S. (2005), "An Investigation of Distributed Lag Models in the Context of Air Pollution," *Journal of the Air & Waste Management Association*, 55, 273–282.
- 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](https://doi.org/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," *American 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.