# Distributed lag interaction models with two pollutants

Yin-Hsiu Chen, Bhramar Mukherjee and Veronica J. Berrocal

*University of Michigan, Ann Arbor, USA*

**Summary.** Distributed lag models (DLMs) have been widely used in environmental epidemiology to quantify the lagged effects of air pollution on a health outcome of interest such as mortality and morbidity. Most previous DLM approaches consider only one pollutant at a time. We propose a distributed lag interaction model to characterize the joint lagged effect of two pollutants. One natural way to model the interaction surface is by assuming that the underlying basis functions are tensor products of the basis functions that generate the main effect distributed lag functions. We extend Tukey's 1 degree-of-freedom interaction structure to the two-dimensional DLM context. We also consider shrinkage versions of the two to allow departure from the specified Tukey interaction structure and achieve bias–variance trade-off. We derive the marginal lag effects of one pollutant when the other pollutant is fixed at certain quantiles. In a simulation study, we show that the shrinkage methods have better average performance in terms of mean-squared error across various scenarios. We illustrate the methods proposed by using the 'National morbidity, mortality, and air pollution study' data to model the joint effects of particulate matter and ozone on mortality count in Chicago, Illinois, from 1987 to 2000.

*Keywords*: Shrinkage; Time series; Tukey's single degree-of-freedom test for non-additivity; Two-dimensional distributed lag interaction models

## 1. Introduction

The association between air pollution and adverse health outcomes has been an important public health concern and a topic of extensive research in environmental epidemiology (Pope and Dockery, 2006). The short-term, or acute, effects of air pollution exposure on health outcomes, such as mortality and cardiovascular events, have been widely studied (Pope *et al.*, 1995; Dominici *et al.*, 2006). However, most studies so far have considered adverse health effects of exposure to a single pollutant (Dominici *et al.*, 2010). When ambient concentration data are available for multiple pollutants, it is standard practice to analyse their effects one at a time by fitting multiple single-pollutant models. However, the health burden from simultaneous exposure to multiple pollutants may differ from the sum of individual effects and the mode of action can be synergistic or antagonistic (Mauderly, 1993). A multipollutant approach that considers the joint effects of chemical mixtures of exposures is likely to yield a more accurate assessment of health risk (Billionnet *et al.*, 2012). A variety of approaches have been proposed to estimate the health effects of multiple pollutants (Sun *et al.*, 2013), including the least absolute shrinkage and selection operator (the lasso) (Tibshirani, 1996), classification and regression trees (Hu *et al.*, 2008) and Bayesian kernel machine regression (Bobb *et al.*, 2014). However, very few methods so far have considered the problem of capturing the lagged effect of two pollutants and their

*Address for correspondence*: Yin-Hsiu Chen, Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA.
E-mail: yinhsiuc@umich.edu

potential interactions over a biologically meaningful time period. Single-day pollution measures might underestimate risk when there is a cumulative effect of air pollution over a time window preceding a health event (Roberts, 2005).

Distributed lag models (DLMs) are a class of models that are often used to include lagged measures of concentration levels of an ambient air pollutant simultaneously. The parametric DLM assumes that the lag effect coefficients are a function of the lags, such as lower degree polynomials (Almon, 1965). Generalized additive DLMs (Zanobetti *et al.*, 2000) use penalized regression splines (Marx and Eilers, 1998) to represent the distributed lag (DL) function in a more flexible manner. The Bayesian DLM (Welty *et al.*, 2009) was proposed to incorporate prior knowledge about the DL function through specification of the prior variance–covariance matrix of lag coefficients. Most of the discussion regarding DLMs has been in the context of a single pollutant and only few distributed lag interaction models (DLIMs) with two pollutants have been attempted. Extensions to higher dimensions include bivariate constrained DLIMs (CDLIMs) (Muggeo, 2007) and high degree DLMs (HDDLMs) (Heaton and Peng, 2014). Muggeo (2007) jointly modelled the temperature and air particular matter with aerodynamic diameter less than $10\,\mu$m ($PM_{10}$) main effect in the same way as a parametric DLM with two separate sets of basis functions. Tensor products of the two are employed to characterize the joint DL surface for the temperature–$PM_{10}$ interaction. Heaton and Peng (2014) extended the DLM framework to incorporate higher order interactions between lagged predictors corresponding to a single exposure, using a Gaussian process prior as a dimension reduction tool.

Tukey's 1 degree-of-freedom (DF) test for non-additivity (Tukey, 1949) is a parsimonious approach to model the interaction term as a scaled product of its corresponding main effects (Chatterjee *et al.*, 2006; Maity *et al.*, 2009). In this paper, we extend Tukey's model to DLIMs where the interaction is parameterized as a scaled product of two DLM main effects. We shall consider estimation and inference under such an extension in both frequentist and Bayesian frameworks. We also propose a Bayesian constrained DLIM (BCDLIM) approach to characterize the joint effect of two pollutants. Instead of shrinking all main effects and interaction effects toward 0, we set a prespecified parametric CDLIM as the shrinkage target in this approach. The BCDLIM can strike a desirable bias–variance trade-off in a data-adaptive way.

The rest of the paper is organized as follows. In Section 2, we first review the existing methods, including

(a) the unconstrained DLIM (UDLIM) and
(b) the CDLIM.

We then introduce the proposed new methods

(i)   the Tukey DLIM (TDLIM),
(ii)  the Bayesian TDLIM (BTDLIM) and
(iii) the BCDLIM.

In Section 3, we conduct a simulation study to evaluate the operating characteristics of the five methods. In Section 4, we illustrate the methods by analysing data from the 'National morbidity, mortality, and air pollution study' (NMMAPS) to estimate the lagged effects of particulate matter with diameter less than $10\,\mu$m ($PM_{10}$) and ozone ($O_3$) concentration on mortality in Chicago, Illinois, from 1987 to 2000. We conclude with a discussion in Section 5.

There are several novel features of this paper. First, we extend the DLM to the DLIM to handle two pollutants. We attempt to characterize the changes in a true DL function corresponding to one exposure when the other is fixed at different values. Extending the well-known

Tukey model for interaction to the DLIM is another innovation. Finally, using data-adaptive shrinkage to allow for an unconstrained interaction model to shrink towards a parametric DLIM structure is a new contribution to the literature. More broadly, the paper posits new ideas for thinking about interaction structures between a pair of time series predictors with potential lagged effects on an outcome. This approach bears relevance beyond air pollution epidemiology.

## 2.  Methods

Let $x_{1t}$ denote the first exposure measured at time $t$ (e.g. $PM_{10}$), $x_{2t}$ denote the second exposure measured at time $t$ (e.g. $O_3$), $y_t$ denote the response measured at time $t$ (e.g. daily mortality count) and $\mathbf{z}_t$ denote the vector of covariates at time $t$, such as temperature and humidity, in addition to a constant 1 corresponding to the intercept parameter. Let $T$ be the length of the time series, and $L_1$ and $L_2$ be the maximum number of lags considered for the first and second exposure respectively. In addition, we denote with $\mathbf{X}_{1t} = (x_{1t}, \ldots, x_{1,t-L_1})^{\mathrm{T}}$ and $\mathbf{X}_{2t} = (x_{2t}, \ldots, x_{2,t-L_2})^{\mathrm{T}}$ the vector of lagged exposures and with $\mathbf{X}_{It} = \mathbf{X}_{1t} \otimes \mathbf{X}_{2t}$, where '$\otimes$' is the Kronecker product, the $(L_1 + 1)(L_2 + 1)$ elements that refer to the two-way interaction terms between the two exposures. The log-linear Poisson DLIM with all pairwise interactions between lagged measurements of the two exposures is described as

$$y_t | \mathbf{z}_t, \mathbf{X}_{1t}, \mathbf{X}_{2t}, \mathbf{X}_{It} \sim \mathrm{Poisson}(\mu_t), \tag{1}$$

$$\begin{aligned}
\log(\mu_t) &= \mathbf{z}_t^{\mathrm{T}} \boldsymbol{\alpha} + \mathbf{X}_{1t}^{\mathrm{T}} \boldsymbol{\beta}_1 + \mathbf{X}_{2t}^{\mathrm{T}} \boldsymbol{\beta}_2 + \mathbf{X}_{It}^{\mathrm{T}} \boldsymbol{\gamma} \\
&= \mathbf{z}_t^{\mathrm{T}} \boldsymbol{\alpha} + \sum_{i=0}^{L_1} x_{1,t-i} \beta_{1i} + \sum_{j=0}^{L_2} x_{2,t-j} \beta_{2j} + \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \gamma_{ij} x_{1,t-i} x_{2,t-j}
\end{aligned} \tag{2}$$

where $\boldsymbol{\alpha}$ represents the effect of covariates, $\boldsymbol{\beta}_1 = (\beta_{10}, \ldots, \beta_{1L_1})^{\mathrm{T}}$ is the $(L_1 + 1)$-vector of lagged main effects of the first exposure, $\boldsymbol{\beta}_2 = (\beta_{20}, \ldots, \beta_{2L_2})^{\mathrm{T}}$ is the $(L_2 + 1)$-vector of lagged main effects of the second exposure and $\boldsymbol{\gamma} = \mathrm{vec}(\boldsymbol{\Gamma}) = (\gamma_{00}, \gamma_{01}, \ldots, \gamma_{L_1 L_2})^{\mathrm{T}}$ where $\boldsymbol{\Gamma}$ is the $(L_1 + 1) \times (L_2 + 1)$ matrix of interaction effects. Our primary goal is to estimate the main effects $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ and the interaction effects $\boldsymbol{\gamma}$. For simplicity, we leave out $\mathbf{z}_t^{\mathrm{T}} \boldsymbol{\alpha}$ in subsequent presentations.

*Remark 1.* Expressions (1) and (2) model the conditional mean response at a time point $t$ given the current and past measurements of the two exposures. The non-null interaction effect in equation (2) implies that the lagged effects of the first exposure depend on the level of the second exposure, and vice versa. It is noted that the interaction effects in equation (2) are not symmetric, namely $\gamma_{ij} \neq \gamma_{ji}$ for $i \neq j$. A natural quantity of interest is the marginal effect of one exposure at a certain lag given the other exposure fixed at a certain level such as the median or a specified quantile. Algebraically, if we fix the second exposure at $x_2^*$ across all lags, the marginal lag effects of the first exposure at lag $i$ can be written as $\beta_{1i}^* = \beta_{1i} + x_2^* \Sigma_{j=0}^{L_2} \gamma_{ij}$ for $i = 0, \ldots, L_1$. The vector representation is

$$\boldsymbol{\beta}_1^{\mathrm{m}}(x_2^*) = \boldsymbol{\beta}_1 + x_2^* \boldsymbol{\Gamma} \mathbf{1} \tag{3}$$

where $\mathbf{1}$ is a vector of 1s. Similarly, if we fix the first exposure at $x_1^*$, the marginal lag effects of the second exposure at lag $j$ can be written as $\beta_{2j}^* = \beta_{2j} + x_1^* \Sigma_{i=0}^{L_1} \gamma_{ij}$ for $j = 0, \ldots, L_2$ with vector representation $\boldsymbol{\beta}_2^{\mathrm{m}}(x_1^*) = \boldsymbol{\beta}_2 + x_1^* \boldsymbol{\Gamma}^{\mathrm{T}} \mathbf{1}$. Throughout the rest of this paper, we shall summarize the estimates of $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\gamma} = \mathrm{vec}(\boldsymbol{\Gamma})$ on the basis of the above expressions.

## 2.1.    Existing methods

### 2.1.1.    Unconstrained distributed lag interaction model

The UDLIM does not impose any constraints on coefficients $\psi = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\gamma}^T)^T$ in equation (2). The UDLIM coefficients can be simply estimated via maximum likelihood estimation:

$$\hat{\psi}_{\text{UDLIM}} = \arg\max_{\psi} \sum_{t=1}^{T} \{y_t \mathbf{X}_t^T \psi - \exp(\mathbf{X}_t^T \psi) - \log(y_t!)\},$$

where $\mathbf{X}_t = (\mathbf{X}_{1t}^T, \mathbf{X}_{2t}^T, \mathbf{X}_{It}^T)^T$. Standard frequentist inference based on large sample theory of maximum likelihood estimates can be drawn subsequently. However, because of the collinearity between serially measured exposure levels and the large number of parameters (i.e. $L_1 + L_2 + 2$ main effect terms and $(L_1 + 1)(L_2 + 1)$ interaction terms), the lagged effect estimates may be less efficient with inflated variance and the estimated DL functions could be highly variable.

### 2.1.2.    Constrained distributed lag interaction model

The parametric DLIM imposes a smooth structure on lagged effect coefficients by assuming that each lag coefficient is a linear combination of known basis functions measured at its lag index. The CDLIM extends this configuration to two-dimensional scenarios. Assume that $B_{11}(\cdot), \ldots, B_{1p_1}(\cdot)$ are the $p_1$ basis functions applied to $\boldsymbol{\beta}_1$ and $B_{21}(\cdot), \ldots, B_{2p_2}(\cdot)$ are the $p_2$ basis functions applied to $\boldsymbol{\beta}_2$. The main effects coefficients are assumed to be of the form $\beta_{1i} = \Sigma_{m=1}^{p_1} B_{1m}(i)\theta_{1m}$ for $i = 0, \ldots, L_1$ and $\beta_{2j} = \Sigma_{n=1}^{p_2} B_{2n}(j)\theta_{2n}$ for $j = 0, \ldots, L_1$ where $\{\beta_{1i}\}$ and $\{\beta_{2j}\}$ are elements of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ respectively, and $\{\theta_{1m}\}$ and $\{\theta_{2n}\}$ are free parameters to be estimated. To smooth the interaction surface, Muggeo (2007) utilized tensor products of marginal basis functions. The element corresponding to the interaction between $x_{1,t-i}$ and $x_{2,t-j}$ can be expressed as $\gamma_{ij} = \Sigma_{m=1}^{p_1} \Sigma_{n=1}^{p_2} B_{1m}(i) B_{2n}(j)\theta_{Imn}$.

Define $\mathbf{C}_1$ as an $(L_1 + 1) \times p_1$ transformation matrix (Gasparrini *et al.*, 2010) where the element $(i+1, m)$ is $B_{1m}(i)$ and, similarly, define $\mathbf{C}_2$ as an $(L_2 + 1) \times p_2$ transformation matrix where the element $(j+1, n)$ is $B_{2n}(j)$. Denote $\boldsymbol{\theta}_1 = (\theta_{11}, \ldots, \theta_{1p_1})$, $\boldsymbol{\theta}_2 = (\theta_{21}, \ldots, \theta_{2p_2})$ and $\boldsymbol{\theta}_I = (\theta_{I11}, \theta_{I12}, \ldots, \theta_{Ip_1p_2})$; the CDLIM coefficients can be written in terms of the free parameters to be estimated as

$$\left.\begin{aligned} \boldsymbol{\beta}_1 &= \mathbf{C}_1 \boldsymbol{\theta}_1, \\ \boldsymbol{\beta}_2 &= \mathbf{C}_2 \boldsymbol{\theta}_2, \\ \boldsymbol{\gamma} &= (\mathbf{C}_1 \otimes \mathbf{C}_2)\boldsymbol{\theta}_I. \end{aligned}\right\} \tag{4}$$

The free parameters $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_I$ can be obtained by maximizing the log-likelihood function

$$\sum_{t=1}^{T} \{y_t(\mathbf{W}_{1t}^T\boldsymbol{\theta}_1 + \mathbf{W}_{2t}^T\boldsymbol{\theta}_2 + \mathbf{W}_{It}^T\boldsymbol{\theta}_I)^T - \exp(\mathbf{W}_{1t}^T\boldsymbol{\theta}_1 + \mathbf{W}_{2t}^T\boldsymbol{\theta}_2 + \mathbf{W}_{It}^T\boldsymbol{\theta}_I) - \log(y_t!)\}$$

where $\mathbf{W}_{1t} = \mathbf{C}_1^T\mathbf{X}_{1t}$, $\mathbf{W}_{2t} = \mathbf{C}_2^T\mathbf{X}_{2t}$ and $\mathbf{W}_{It} = (\mathbf{C}_1 \otimes \mathbf{C}_2)^T\mathbf{X}_{It}$. Let $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \boldsymbol{\theta}_I^T)^T$, a vector of length $p_1 + p_2 + p_1p_2$, and $\mathbf{C} = \text{diag}(\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_1 \otimes \mathbf{C}_2)$. The CDLIM estimator can be written as $\hat{\psi}_{\text{CDLIM}} = \mathbf{C}\hat{\boldsymbol{\Theta}}$ and $\text{cov}(\hat{\psi}_{\text{CDLIM}}) = \mathbf{C}\,\text{cov}(\hat{\boldsymbol{\Theta}})\mathbf{C}^T$.

## 2.2.    Proposed methods

### 2.2.1.    Tukey's distributed lag interaction model

The underlying foundation of Tukey's model for interaction is a latent variable framework (Chatterjee *et al.*, 2006). Suppose that we define a surrogate variable for each exposure that

aggregates the temporal lagged effect of the exposure through a weighted sum at time $t$, namely

$$s_{1t} = \sum_{i=0}^{L_1} w_{1i} x_{1,t-i}, \quad s_{2t} = \sum_{j=0}^{L_2} w_{2j} x_{2,t-j}. \tag{5}$$

If we assume that the association between $y_t$, $\mathbf{X}_{1t}$ and $\mathbf{X}_{2t}$ is through the interaction model

$$\log(E[y_t]) = \mu_0 + \mu_1 s_{1t} + \mu_2 s_{2t} + \mu_I s_{1t} s_{2t}. \tag{6}$$

Substituting equation (5) in equation (6), we can obtain

$$\log(E[y_t]) = \mu_0 + \sum_{i=0}^{L_1} \mu_1 w_{1i} x_{1,t-i} + \sum_{j=0}^{L_2} \mu_2 w_{2j} x_{2,t-j} + \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \mu_I w_{1i} w_{2j} x_{1,t-i} x_{2,t-j}$$

$$= \mu_0 + \sum_{i=0}^{L_1} \beta_{1i} x_{1,t-i} + \sum_{j=0}^{L_2} \beta_{2j} x_{2,t-j} + \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} \gamma_{ij} x_{1,t-i} x_{2,t-j}$$

where $\beta_{1i} = \mu_1 w_{1i}$, $\beta_{2j} = \mu_2 w_{2j}$ and $\gamma_{ij} = \mu_I w_{1i} w_{2j}$. Note that we can express the interaction coefficient as

$$\gamma_{ij} = \beta_{1i} \beta_{2j} \frac{\mu_I}{\mu_1 \mu_2},$$

a scaled product of the corresponding main effect coefficients. This motivates the use of Tukey's style interaction in our context. The surrogate variables $s_{1t}$ and $s_{2t}$ represent summary exposures over all the lags of the two exposures. Coefficients $\mu_0$, $\mu_1$, $\mu_2$ and $\mu_I$ characterize the overall combined effects of the two exposures in association with outcome measurement at lag 0. The lag measurements of the two exposures interact through the two surrogate variables in the simple pairwise interaction model that is described in equation (6). Estimating the lagged effects in this model is the same as estimating the relative weights to combine the exposure lagged measurements into a summary surrogate variable. To extend the classical Tukey interaction structure to DLIMs, we now assume that the main effects are specified in the same way as in the CDLIM with constrained parameterization such that $\beta_1 = \mathbf{C}_1 \theta_1$ and $\beta_2 = \mathbf{C}_2 \theta_2$ as in equation (4). In matrix form, the interaction coefficients can be expressed under Tukey's model as

$$\boldsymbol{\gamma} = \eta(\boldsymbol{\beta}_1 \otimes \boldsymbol{\beta}_2) = (\mathbf{C}_1 \otimes \mathbf{C}_2)\{\eta(\theta_1 \otimes \theta_2)\}.$$

Note that the interaction structure corresponding to the TDLIM is a special case of the CDLIM with $\theta_I = \eta(\theta_1 \otimes \theta_2)$. The number of parameters that are used for modelling the interaction effect reduces from $p_1 p_2$ to 1. The model without interaction is nested within the Tukey structure with the scalar parameter set to 0, assuming non-null main effects. The free parameters $\theta_1$, $\theta_2$ and $\eta$ can be estimated by maximizing the log-likelihood function

$$\sum_{t=1}^{T} [y_t \{\mathbf{W}_{1t}^{\mathrm{T}} \theta_1 + \mathbf{W}_{2t}^{\mathrm{T}} \theta_2 + \eta \mathbf{W}_{It}^{\mathrm{T}}(\theta_1 \otimes \theta_2)\} - \exp\{\mathbf{W}_{1t}^{\mathrm{T}} \theta_1 + \mathbf{W}_{2t}^{\mathrm{T}} \theta_2 + \eta \mathbf{W}_{It}^{\mathrm{T}}(\theta_1 \otimes \theta_2)\} - \log(y_t!)]. \tag{7}$$

The TDLIM is a non-linear regression model where the objective function (7) involves products of the parameters. Linear approximation using a first-order Taylor series expansion can be applied for parameter estimation and statistical inference. However, empirically, we found that the approximation accuracy using a first-order approximation is poor and the asymptotic variance is far from the empirical variance. We therefore consider an iterative approach for estimation (details are provided in the on-line supplementary appendix A.1). The value of the objective function decreases at each step and the solution is guaranteed to converge.

We recognize that the likelihood function (7) is non-convex in terms of the parameters so the convergence to a global maximum is not guaranteed by the iterative procedure. However, in our numerical studies, when the main effects are bounded away from zero, the choice of various initial values did not affect the final parameter estimates. When at least one of the main effects are close to the null value, the parameter $\eta$ is not identifiable and estimation instability occurs in these cases. For statistical inference, we consider a standard bootstrap by resampling observations with replacement to obtain standard errors and confidence intervals.

### 2.2.2. *Bayesian Tukey distributed lag interaction model*

In the proposed BTDLIM, the main effects are parametrically specified in the same way as in equation (4) and the interaction effects are modelled in the spirit of the TDLIM. The distinction from the presentation in the previous section is that the BTDLIM allows a departure from Tukey's interaction structure in a data-adaptive way. The BTDLIM assumes that the scalar parameter can vary across different interaction terms through the prior specification

$$\gamma = \eta \odot (\beta_1 \otimes \beta_2), \qquad \eta \sim N\{\mathbf{0}, \sigma^2 \Sigma(\omega)\}$$

where $\eta = (\eta_{00}, \eta_{01}, \ldots, \eta_{L_1 L_2})^{\mathrm{T}}$ is the vector of scalars, '$\odot$' is the operator denoting elementwise multiplication, $\sigma^2$ is the common variance and $\Sigma$ is the correlation matrix parameterized by a single parameter $\omega > 0$. The correlation between $\eta_{ij}$ and $\eta_{i^* j^*}$ is given by $\omega^{\sqrt{\{(i-i^*)^2 + (j-j^*)^2\}}}$ assuming an exponential structure. The prior on $\eta$ relaxes the strict specification of Tukey's interaction structure. The amount of departure from Tukey's model is controlled by the parameter $\omega$. At one extreme, when $\omega = 0$, no structure is imposed on the interaction effects. The interaction coefficients are simply a reparametrization of the UDLIM coefficients in equation (2). At the other extreme when $\omega = 1$, the model degenerates to the TDLIM and enforces the interaction coefficients to follow the Tukey structure exactly. When $\omega$ approaches 1, the correlation between neighbouring coefficients is larger, resulting in a smoother interaction surface.

To complete the model specifications, we assign $\theta_1 \sim N(\mathbf{0}, 100^2 \mathbf{I})$ and $\theta_2 \sim N(\mathbf{0}, 100^2 \mathbf{I})$ as vague priors for the main effects coefficients. We assume a non-informative prior (Gelman, 2006) on the variance parameter $\sigma^2 \sim \mathrm{IG}(a = 0.001, b = 0.001)$ where $a$ and $b$ are the shape and scale parameters of the inverse gamma distribution. To alleviate the computational burden and to keep the prior uninformative, we let $\omega$ have a discrete uniform prior on $\{0.1, 0.2, \ldots, 1\}$. The marginal posterior density of $\beta_1$, $\beta_2$ and $\gamma$ is not available in closed form. We use a Metropolis–Hastings algorithm within a Gibbs sampler to approximate the posterior distribution and obtain the BTDLIM estimator as the posterior mean with the corresponding highest posterior density interval as the corresponding credible interval. The full conditional distributions are presented in the on-line supplementary appendix A.2.

### 2.2.3. *Bayesian constrained distributed lag interaction model*

The CDLIM is a fully parametric model. The dimension reduction from $L_1 + 1 + L_2 + 1 + (L_1 + 1)(L_2 + 1)$ parameters to $p_1 + p_2 + p_1 p_2$ parameters results in a gain of efficiency in estimation. However, the benefit can be counterbalanced by potential bias when the underlying structure for the DL functions or surface is misspecified. We propose a BCDLIM to shrink UDLIM estimates in a smooth manner towards a prespecified CDLIM.

Let $B_{11}^+(\cdot), \ldots, B_{1,L_1+1}^+(\cdot)$ be $L_1 + 1$ basis functions for the first exposure, e.g. *B*-spline basis functions of degree 3 (cubic) with intercept and $L_1 - 3$ equispaced internal knots positioned

between 0 and $L_1$. Note that the basis functions describe the non-linearity in the DL function, but the exposure effect at each lag is still assumed to be linear. Let $\mathbf{T}_1$ be the corresponding $(L_1+1) \times (L_1+1)$ transformation matrix. Let $\mathbf{T}_2$ denote the square transformation matrix with dimension $(L_2+1) \times (L_2+1)$, constructed in a similar manner for the second exposure, and let the transformation matrix for the interaction parameter be $\mathbf{T}_I = (\mathbf{T}_1 \otimes \mathbf{T}_2)$ with dimension $(L_1+1)(L_2+1) \times (L_1+1)(L_2+1)$. If we apply the transformation operators $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_I$ to the CDLIM, the resulting estimator would be identical to the UDLIM estimator since a full rank transformation on the coefficients does not change the model fit. However, if we imposed shrinkage on the coefficients by using an $L_2$-penalty, the CDLIM and UDLIM estimators would be different since the shrinkage is employed in different parameter spaces. The UDLIM estimator can be viewed as choosing $B_{1m}^+(i) = I(m = i+1)$ for $m = 1, \ldots, L_1+1$ and $B_{2n}^+(j) = I(n = j+1)$ for $n = 1, \ldots, L_2+1$, where $I(\cdot)$ is an indicator function, corresponding to $\mathbf{T}_1 = \mathbf{I}$ and $\mathbf{T}_2 = \mathbf{I}$. Although the two sets of estimates share the same shrinkage target (i.e. the zero line), the solution paths are different. If the basis functions that are selected for $\mathbf{T}_1$ and $\mathbf{T}_2$ are smooth, the CDLIM with shrinkage leads to smooth estimates.

Instead of shrinking the model coefficients towards 0, we consider shrinking them to a non-null target, determined by the transformation matrices $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_I = (\mathbf{C}_1 \otimes \mathbf{C}_2)$ for the CDLIM that is defined in equation (4). Without loss of generality, we describe only how to construct the non-null shrinkage target for the first exposure. We first separate $\mathbf{T}_1$ into two parts—$\mathbf{C}_1$ and $\mathbf{C}_1^c$ where $\mathbf{C}_1^T \mathbf{C}_1^c = \mathbf{0}$. We make use of this orthogonal decomposition to obtain $\mathbf{C}_1^c$ whose columns span the complementary column space of $\mathbf{C}_1$. $\mathbf{C}_1$ and $\mathbf{C}_1^c$ define the decomposition of the transformations corresponding to shrinkage towards a prespecified target and shrinkage towards 0 respectively. The orthogonal projection of $\mathbf{T}_1$ onto the complementary column space of $\mathbf{C}_1$ is given by $\mathbf{P}_1 = (\mathbf{I} - \mathbf{C}_1(\mathbf{C}_1^T \mathbf{C}_1)^{-1} \mathbf{C}_1^T)\mathbf{T}_1$. Using singular value decomposition, we can write $\mathbf{P}_1 = \mathbf{U}_1 \mathbf{D}_1 \mathbf{V}_1^T$ where $\mathbf{U}_1$ contains the columns of left singular vectors, $\mathbf{D}_1$ is a diagonal matrix with eigenvalues of $\mathbf{P}_1$, and $\mathbf{V}_1$ contains the columns of right singular vectors. Since the rank of $\mathbf{P}_1$ is $L_1+1-p_1$, we can write $\mathbf{U}_1 = (\mathbf{U}_{11} \; \mathbf{U}_{12})$ where $\mathbf{U}_{11}$ is an $(L_1+1) \times (L_1+1-p_1)$ matrix with columns of singular vectors corresponding to non-zero eigenvalues in $\mathbf{D}_1$, whereas $\mathbf{U}_{12}$ is an $(L_1+1) \times p_1$ matrix with columns of singular vectors corresponding to the eigenvalues of 0. We consider $\mathbf{C}_1^c = \mathbf{U}_{11}$. It is easy to show that $\mathbf{C}_1^T \mathbf{C}_1^c = \mathbf{0}$ and the $p_1$ columns of $\mathbf{C}_1$ and the $L_1+1-p_1$ columns of $\mathbf{C}_1^c$ span the entire $\mathbb{R}^{L_1+1}$. In other words, shrinkage through the columns of $\mathbf{C}_1^c$ defines the CDLIM estimate as the shrinkage target. The complementary matrices $\mathbf{C}_2^c$ and $\mathbf{C}_I^c$ for the second exposure and interaction can be constructed using $\mathbf{C}_2$ and $\mathbf{T}_2$, and $\mathbf{C}_I$ and $\mathbf{T}_I$ respectively in a similar way.

The likelihood corresponding to the above specification is given by

$$\mathbf{Y}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma} \sim \text{Poisson}\{\exp(\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{X}_I\boldsymbol{\gamma})\}$$

where $\mathbf{Y} = (y_1, \ldots, y_T)^T$, $\mathbf{X}_1 = (\mathbf{X}_{11}, \ldots, \mathbf{X}_{1T})^T$, $\mathbf{X}_2 = (\mathbf{X}_{21}, \ldots, \mathbf{X}_{2T})^T$ and $\mathbf{X}_I = (\mathbf{X}_{I1}, \ldots, \mathbf{X}_{IT})^T$. The prior specifications corresponding to the BCDLIM parameters are

$$\boldsymbol{\beta}_1 = \mathbf{C}_1\boldsymbol{\theta}_1 + \mathbf{C}_1^c\boldsymbol{\theta}_1^c,$$
$$\boldsymbol{\beta}_2 = \mathbf{C}_2\boldsymbol{\theta}_2 + \mathbf{C}_2^c\boldsymbol{\theta}_2^c,$$
$$\boldsymbol{\gamma} = \mathbf{C}_I\boldsymbol{\theta}_I + \mathbf{C}_I^c\boldsymbol{\theta}_I^c,$$
$$\boldsymbol{\theta}_1 \sim N(\mathbf{0}, 100^2\mathbf{I}),$$
$$\boldsymbol{\theta}_2 \sim N(\mathbf{0}, 100^2\mathbf{I}),$$
$$\boldsymbol{\theta}_I \sim N(\mathbf{0}, 100^2\mathbf{I}),$$

$$\theta_1^c \sim N(\mathbf{0}, \sigma_1^2 \mathbf{I}),$$
$$\theta_2^c \sim N(\mathbf{0}, \sigma_2^2 \mathbf{I}),$$
$$\theta_I^c \sim N(\mathbf{0}, \sigma_I^2 \mathbf{I})$$

where $\theta_1$, $\theta_2$ and $\theta_I$ are the coefficients without shrinkage and $\theta_1^c$, $\theta_2^c$ and $\theta_I^c$ are the coefficients to be shrunk towards 0. In other words, $\beta_1$, $\beta_2$ and $\gamma$ are shrunk towards $\mathbf{C}_1\theta_1$, $\mathbf{C}_2\theta_2$ and $\mathbf{C}_I\theta_I$ respectively. To complete the model specification, we assign hyperpriors on the variance parameters as

$$\sigma_1^2 \sim \mathrm{IG}(a_0, b_0),$$
$$\sigma_2^2 \sim \mathrm{IG}(a_0, b_0),$$
$$\sigma_I^2 \sim \mathrm{IG}(a_0, b_0).$$

We fix $a_0 = b_0 = 0.001$ to assume a non-informative hyperprior (Gelman, 2006). A Metropolis–Hastings algorithm within a Gibbs sampler can be used to approximate the posterior distribution of the model parameters. The full conditional distributions are provided in the on-line supplementary appendix A.3. The hyperpriors of the the BCDLIM can alternatively be viewed as penalty terms in penalized likelihood. The dual representation is presented in supplementary appendix A.4.

## 3. Simulation study

We conducted a simulation study to compare the estimation performance of the five methods that were introduced in Section 2 under different settings. We implemented the three frequentist methods by using the built-in R function `glm` and the two Bayesian methods by calling the software 'just another Gibbs sampler' using R package `rjags` (Lunn *et al.*, 2009). The average computation times for 1000 data sets under each method are provided in the on-line supplementary appendix A.5 Table 1. All simulations were performed in R version 3.3.1 (R Core Team, 2017).

### 3.1. Simulation settings
We generated two separate exposure time series ($i = 1, 2$) of length 1000 days with mean 3 and first-order auto-correlation equal to 0.5 from the model $x_{it} = 0.5x_{it-1} + \epsilon_{it}$ where $\epsilon_{it} \sim^{\mathrm{IID}} N(0, 0.75)$ for $i = 1, 2$ and $t = 1, \ldots, 1000$. We set $L_1 = L_2 = 9$ for both data generation and model fitting. The outcome $y_t$ is generated from a Poisson distribution with mean $\exp(\beta_0 + \mathbf{X}_{1t}^{\mathrm{T}}\beta_1 + \mathbf{X}_{2t}^{\mathrm{T}}\beta_2 + \mathbf{X}_{It}^{\mathrm{T}}\gamma)$ for $t = 1, \ldots, 1000$ where $\mathbf{X}_{1t}$, $\mathbf{X}_{2t}$ and $\mathbf{X}_{It}$ are defined as in Section 2. Let $\beta_0 = 3$ and consider two DL functions for the main effect coefficients $\beta_1$ and $\beta_2$—

- (a) a cubic and
- (b) a function with departure from cubic.

We consider five different underlying true interaction structures for $\gamma$—

- (i) no interaction,
- (ii) Tukey's style interaction,
- (iii) Kronecker product interaction,
- (iv) sparse interaction and
- (v) unstructured interaction.

The exact specifications are available in the on-line supplementary appendix A.6. In total, nine simulation scenarios, including all combinations of the two main effect coefficients (a) and

(b) and five interaction effect coefficients (i)–(v), except the combination of (b) and (iii), are considered. Exclusion of the combination of (b) and (iii) is because the Kronecker product interaction cannot be constructed when the corresponding main effects are not fully parametric as their underlying basis functions are undefined. In all simulations, we assume that the lag structure of the CDLIM, TDLIM, BTDLIM and BCDLIM is a cubic polynomial in the lags for all model fitting purposes.

### 3.2. Evaluation metrics

The marginal lagged effects of the first exposure defined in equation (3) depend on the level at which the second exposure is fixed. One way to eliminate the effect of the second exposure is to integrate it out. We consider the use of a finite Riemann sum to approximate numerically the integral $\beta_1^* = \int \beta_1^*(x_2)dx_2 \approx (1/S)\Sigma_{s=1}^{S}\beta_1^*(x_2^{[q(s-0.5)/S]})$ where $x_2^{[q(s-0.5)/S]}$ is the $(s-0.5)/S$th quantile of $x_2$. The empirical bias and empirical relative efficiency of the above quantity with $S = 20$ are used to summarize the simulation results across various scenarios. The squared bias is computed as $(\hat{\bar{\beta}}_1^* - \beta_1^*)^T(\hat{\bar{\beta}}_1^* - \beta_1^*)$ where $\hat{\bar{\beta}}_1^*$ is the average of the estimates that are obtained from the 1000 simulated data sets. The empirical mean-squared error (MSE) is computed as $(1/1000)\Sigma_{j=1}^{1000}||\hat{\beta}_{1j}^* - \beta_1^*||_2^2$. The relative efficiency is expressed with respect to the MSE of the UDLIM estimate, namely the MSE of the UDLIM divided by the MSE of a certain method. We emphasize that the efficiency is defined through the MSE rather than the variance in this paper. Because of the symmetry between $x_1$ and $x_2$, we present results for only the marginal lagged effects of $x_1$.

### 3.3. Simulation results

Results for the setting with main effects generated from a cubic DL function are summarized in Table 1. As we can observe in scenario (i), e.g. no interaction, all methods are more efficient than the UDLIM with relative efficiency ranging from 6.27 to 19.24. The empirical squared bias is minimal for the UDLIM (0.02), CDLIM (0.00) and BCDLIM (0.00) and is moderately small for the TDLIM (0.19) and the BTDLIM (0.13). Null interaction is a special case of Tukey's model

**Table 1.** Empirical squared bias and empirical relative efficiency (measured with respect to the MSE of the UDLIM estimate) of marginal lagged effects across five two-dimensional DL interaction models based on 1000 simulation data sets†

| Interaction structure | Metric | Results for the following models: | | | | |
|---|---|---|---|---|---|---|
| | | UDLIM | CDLIM | TDLIM | BTDLIM | BCDLIM |
| (i), no interaction | Squared bias | 0.02 | 0.00 | 0.19 | 0.13 | 0.00 |
| | Relative efficiency | 1.00 | 6.82 | 19.24 | 8.09 | 6.27 |
| (ii), Tukey's structure | Squared bias | 0.01 | 0.00 | 0.01 | 0.01 | 0.00 |
| | Relative efficiency | 1.00 | 6.14 | 18.66 | 6.71 | 5.76 |
| (iii), Kronecker product | Squared bias | 0.02 | 0.00 | 1.05 | 0.90 | 0.00 |
| | Relative efficiency | 1.00 | 6.68 | 3.45 | 2.77 | 6.17 |
| (iv), sparse | Squared bias | 0.00 | 66.22 | 67.14 | 1.43 | 0.08 |
| | Relative efficiency | 1.00 | 0.07 | 0.07 | 1.71 | 2.80 |
| (v), unstructured | Squared bias | 0.00 | 93.08 | 93.98 | 1.08 | 0.09 |
| | Relative efficiency | 1.00 | 0.05 | 0.05 | 1.88 | 2.70 |

†The lagged effects of both exposures are generated from the same cubic DL function.

**Table 2.**  Empirical squared bias and empirical relative efficiency (measured with respect to the MSE of the UDLIM estimate) of marginal lagged effects across five two-dimensional DL interaction models based on 1000 simulation data sets†

| Interaction structure | Metric | Results for the following models: | | | | |
|---|---|---|---|---|---|---|
| | | *UDLIM* | *CDLIM* | *TDLIM* | *BTDLIM* | *BCDLIM* |
| (i), no interaction | Squared bias | 0.02 | 69.51 | 70.03 | 7.39 | 0.10 |
| | Relative efficiency | 1.00 | 0.24 | 0.25 | 1.59 | 3.25 |
| (ii), Tukey's structure | Squared bias | 0.01 | 990.83 | 1023.84 | 35.50 | 0.09 |
| | Relative efficiency | 1.00 | 0.00 | 0.00 | 0.05 | 1.35 |
| (iv), sparse | Squared bias | 0.01 | 210.32 | 215.94 | 10.80 | 0.52 |
| | Relative efficiency | 1.00 | 0.02 | 0.02 | 0.35 | 1.78 |
| (v), unstructured | Squared bias | 0.01 | 989.93 | 1019.06 | 31.83 | 0.10 |
| | Relative efficiency | 1.00 | 0.00 | 0.00 | 0.04 | 1.34 |

†The lagged effects of both exposures are generated from the same cubic-like DL function (moderate departure from a cubic function).

with $\eta = 0$. Because TDLIMs correctly specify the main effects and interaction effects with a smaller number of parameters, it achieves the highest efficiency (19.24). In scenario (ii) where the non-null interaction effects are of Tukey form, all methods have similar, though slightly smaller, relative efficiency in comparison with scenario (i), ranging from 5.76 to 18.66. Again, the TDLIM has the highest relative efficiency as expected. Scenario (iii) represents the situation where the true interaction structure departs from Tukey's form. We can see now that the TDLIM (3.45) is less efficient than the CDLIM (6.68) because of the bias that is introduced in estimating the interaction surface. However, the TDLIM is still more efficient than the UDLIM (1.00) and the BTDLIM (2.77). The CDLIM correctly specifies both main effects and interaction effects in this scenario and attains the highest efficiency.

Across scenarios (i) and (ii), we note that the squared bias and relative efficiency of the BTDLIM always fall between those of the CDLIM and TDLIM, suggesting that the BTDLIM successfully performs shrinkage and achieves a better average performance. In addition, we can observe that the BCDLIM (relative efficiency 6.27, 5.76, 6.17) is slightly less efficient than the CDLIM (relative efficiency 6.82, 6.14, 6.68) across the three scenarios. The difference is due to the flexibility of the BCDLIM that accounts for a possible departure from Kronecker product type of interaction structure. Scenarios (iv) and (v) are situations where the UDLIM is the only method that can unbiasedly estimate the interaction surface. As expected, both the CDLIM and the TDLIM suffer from serious bias and the gains of efficiency from dimension reduction diminish substantially. The class of interaction surfaces that the CDLIM and TDLIM can describe is restricted. Note that all the methods jointly estimate the main effects and interaction effects and thus misspecifying the interaction effects could possibly distort the estimation of the main effects as they are not orthogonal. The BCDLIM is less biased and more efficient than the BTDLIM across the two scenarios. Across all scenarios when the main effects are correctly specified, the BCDLIM has the best average performance in terms of estimation efficiency.

We summarize the results where the main effects deviate from a cubic DL function in Table 2. Both the CDLIM and the TDLIM are seriously biased, largely because of the misspecification of the main effect terms. These two methods are the least efficient. If we contrast scenarios (i) and (ii), we can observe that misspecification of the main effects not only influences the accuracy of estimation of the main effect DL function, but also the interaction DL surface. The BTDLIM

is biased across the board as well, with squared bias ranging from 7.39 to 35.50 respectively. It is more efficient than the UDLIM only in situations where there is no interaction. The BCDLIM is slightly biased across different scenarios with the squared bias ranging from 0.09 to 0.52. The BCDLIM leads to gains in efficiency with reduced bias. The relative efficiencies are 3.25, 1.35, 1.78 and 1.34 across the four scenarios. Summarizing the results in Tables 1 and 2, it is clear that the BCDLIM approach has desirable MSE properties across the scenarios, offering a robust and efficient solution to this problem.

## 4. Application

### 4.1. Data overview and modelling

We apply the five methods compared in Section 3 to the NMMAPS data. We jointly model daily time series of

(a) $PM_{10}$ and
(b) $O_3$

in association with all-cause non-accidental mortality counts in Chicago, Illinois, for the period between 1987 and 2000. Details with respect to data assembly are available from `http://www.ihapss.jhsph.edu/data/NMMAPS/`. Zanobetti *et al.* (2000) indicated that it is unlikely that lags beyond 2 weeks would have a substantial effect. We therefore set $L_1 = L_2 = 14$ for $PM_{10}$ and $O_3$ levels respectively.

Previous studies showed that it is crucial to account for meteorologic variables as potential confounders in the analysis of air pollution effects (Welty and Zeger, 2005). Dominici *et al.* (2005, 2007) highlighted the need to adjust carefully for a broad set of confounders and to explore their functional forms. We specify the adjustment covariates in the same way as Dominici *et al.* (2005) and focus on the choice of the lag structure in our application. We acknowledge that there may be more optimal adjustment models when we introduce interaction effects. Let $x_{1tk}$, $x_{2tk}$, $y_{tk}$ and $\mathbf{z}_{tk}$ denote the $PM_{10}$ level, $O_3$ level, mortality count and vector of time varying covariates, measured on day $t$ for age group $k$ for $t = 1, \ldots, 5114$ and $k = 1, 2, 3$ respectively. The three age categories are 'greater than or equal to 75 years old', 'between 65 and 74 years old' and 'less than 65 years old'. $PM_{10}$ and $O_3$ were shared exposures across the three age groups so we have $x_{ltk} \equiv x_{lt}$ for $l = 1, 2$. For each group $k$, we assume that, given $PM_{10}$, $O_3$ and other time varying confounders, the mortality count in Chicago on day $t$ is a Poisson random variable $Y_{tk}$ with mean $\mu_{tk}$ such that

$$\log(\mu_{tk}) = \mathbf{X}_{1t}^{\mathrm{T}} \boldsymbol{\beta}_1 + \mathbf{X}_{2t}^{\mathrm{T}} \boldsymbol{\beta}_2 + \mathbf{X}_{It}^{\mathrm{T}} \boldsymbol{\gamma} + \mathbf{z}_{tk}^{\mathrm{T}} \boldsymbol{\alpha}$$

$$= \mathbf{X}_{1t}^{\mathrm{T}} \boldsymbol{\beta}_1 + \mathbf{X}_{2t}^{\mathrm{T}} \boldsymbol{\beta}_2 + \mathbf{X}_{It}^{\mathrm{T}} \boldsymbol{\gamma} + \alpha_0 + \sum_{j=1}^{2} \alpha_{1j} I(k=j) + \sum_{j=1}^{6} \alpha_{2j} I(\mathrm{dow}_t = j)$$

$$+ \mathrm{ns}(\mathrm{temp}_t; 6 \text{ DFs}, \boldsymbol{\alpha}_3) + \mathrm{ns}(\overline{\mathrm{temp}}_t^{(3)}; 6 \text{ DFs}, \boldsymbol{\alpha}_4)$$

$$+ \mathrm{ns}(\mathrm{dptp}_t; 3 \text{ DFs}, \boldsymbol{\alpha}_5) + \mathrm{ns}(\overline{\mathrm{dptp}}_t^{(3)}; 3 \text{ DFs}, \boldsymbol{\alpha}_6)$$

$$+ \mathrm{ns}(t; 98 \text{ DFs}, \boldsymbol{\alpha}_7) + \mathrm{ns}(t; 14 \text{ DFs}, \boldsymbol{\alpha}_8) I(k=1) + \mathrm{ns}(t; 14 \text{ DFs}, \boldsymbol{\alpha}_9) I(k=2)$$

where $\mathbf{X}_{1t} = (x_{1t}, \ldots, x_{1,t-14})^{\mathrm{T}}$, $\mathbf{X}_{2t} = (x_{2t}, \ldots, x_{2,t-14})^{\mathrm{T}}$, $\mathbf{X}_{It} = \mathbf{X}_{1t} \otimes \mathbf{X}_{2t}$ and $\underline{\mathrm{ns}}(\cdot)$ denotes the natural spline with specified DFs. Predictors $\mathrm{dow}_t$, $\mathrm{temp}_t$, $\overline{\mathrm{temp}}_t$, $\mathrm{dptp}_t$ and $\overline{\mathrm{dptp}}_t$ represent the day of the week, the current day's temperature, adjusted average lag 1–3 temperature, the current day's dewpoint temperature and the adjusted average lag 1–3 dewpoint temperatures for day $t$. The indicator variables allow different baseline mortality rates within each age group and within

each day of the week. The smooth term for time $t$ is to adjust for long-term trends and seasonality and the choice of 98 DFs corresponds to 7 DFs per year over the 14-year time horizon. The last two product terms separate smooth functions of time with 2 DFs per year for each age group contrast. The primary goal is to estimate the coefficients $\beta_1$, $\beta_2$ and $\gamma$, whereas $\alpha$ is the set of covariate parameters. A four-degree polynomial DL function is applied to both $\beta_1$ and $\beta_2$ for the CDLIM, TDLIM, BTDLIM and BCDLIM. The analysis is performed in R version 3.3.1 and the source code is available from `https://github.com/yinhsiuc/NMMAPS_DLIM`. The computational times are provided in the on-line supplementary appendix A.4 Table 2 and the summary statistics corresponding to $PM_{10}$ and $O_3$ levels are provided in supplementary appendix A.7.

### 4.2. Estimating marginal distributed lag function

The quantity $100\exp\{10(\beta_{1i}+x_2^*\Sigma_{n=0}^{L_2}\gamma_{in})\}$ represents the percentage change in daily mortality that is associated with an increase of $10\,\mu g\,m^{-3}$ in $PM_{10}$ level at lag $i$ when the $O_3$ level is at $x_2^*$ parts per billion. Similarly, the quantity $100\exp\{10(\beta_{2j}+x_1^*\Sigma_{m=0}^{L_1}\gamma_{mj})\}$ represents the percentage change in daily mortality that is associated with an increase in $O_3$ level of 10 parts per billion at lag $j$ when $PM_{10}$ is set at $x_1^*\,\mu g\,m^{-3}$. We present the marginal lagged effects of $PM_{10}$ and $O_3$ levels in Figs 1 and 2. If we look across the panels in Fig. 1, we can observe that the fits of the UDLIM are undersmoothed and the fits of the CDLIM and TDLIM are oversmoothed, whereas those of the BTDLIM and BCDLIM are in between. When $O_3$ is at the summer average level, the oversmoothing of the CDLIM and TDLIM results in underestimation of the $PM_{10}$ effect at lag 3. For instance, the estimated percentage increases in mortality are associated with an increase of $10\,\mu g\,m^{-3}$ in $PM_{10}$ level at lag 3 when the $O_3$ level is at average summer level are 0.53%, 0.14%, 0.03%, 0.23% and 0.36% for the UDLIM, CDLIM, TDLIM, BTDLIM and BCDLIM respectively. The lower bounds of 95% confidence or credible intervals for the methods except the TDLIM are appreciably above zero. In this situation, shrinkage methods are more desirable since the CDLIM and TDLIM misspecify the DL function and potentially underestimate the relative lag effects. Similarly, we observe slight oversmoothness of the CDLIM and TDLIM on the $O_3$ effect in Fig. 2. However, the degree of underestimation of the $O_3$ effect at early lags is smaller. More similar DL functions across all methods except the UDLIM indicate that the potential misspecification of the DL function by using the CDLIM and TDLIM is minimal.

We present the marginal DL functions of $PM_{10}$ and $O_3$ levels by integrating out the other pollutant in Fig. 3. Similarly to earlier findings, shrinkage is more needed for $PM_{10}$ as the CDLIM and TDLIM tend to oversmooth the DL function in this situation. In addition, we observe that the DL function for $PM_{10}$ level starts from negative, grows to 0 and peaks at lag 3, whereas the DL function for $O_3$ level is greater than 0 at lag 0 and peaks at lag 2. The earlier peak for $O_3$ compared with $PM_{10}$ suggests a more acute effect of $O_3$ than $PM_{10}$ with an earlier window of susceptibility. We also observe that the UDLIM fits of $O_3$ fluctuate more drastically than the UDLIM fits of $PM_{10}$. This is explained by the stronger auto-correlation of the $O_3$ time series and smoothing the DL function is certainly needed and preferred in this case. We can observe that some of the estimated lagged effects are negative at larger lags for $PM_{10}$. This phenomenon is denoted as mortality displacement (Zanobetti *et al.*, 2000) and has been discovered in previous studies. Mortality displacement, which is also referred to as the harvesting effect (Zanobetti *et al.*, 2000), is the temporal shift of mortality. Usually a higher mortality rate due to the deaths of frail individuals a couple of days after a high air pollution episode is followed by a compensatory reduction in mortality rate due to the death of the more frail individuals.
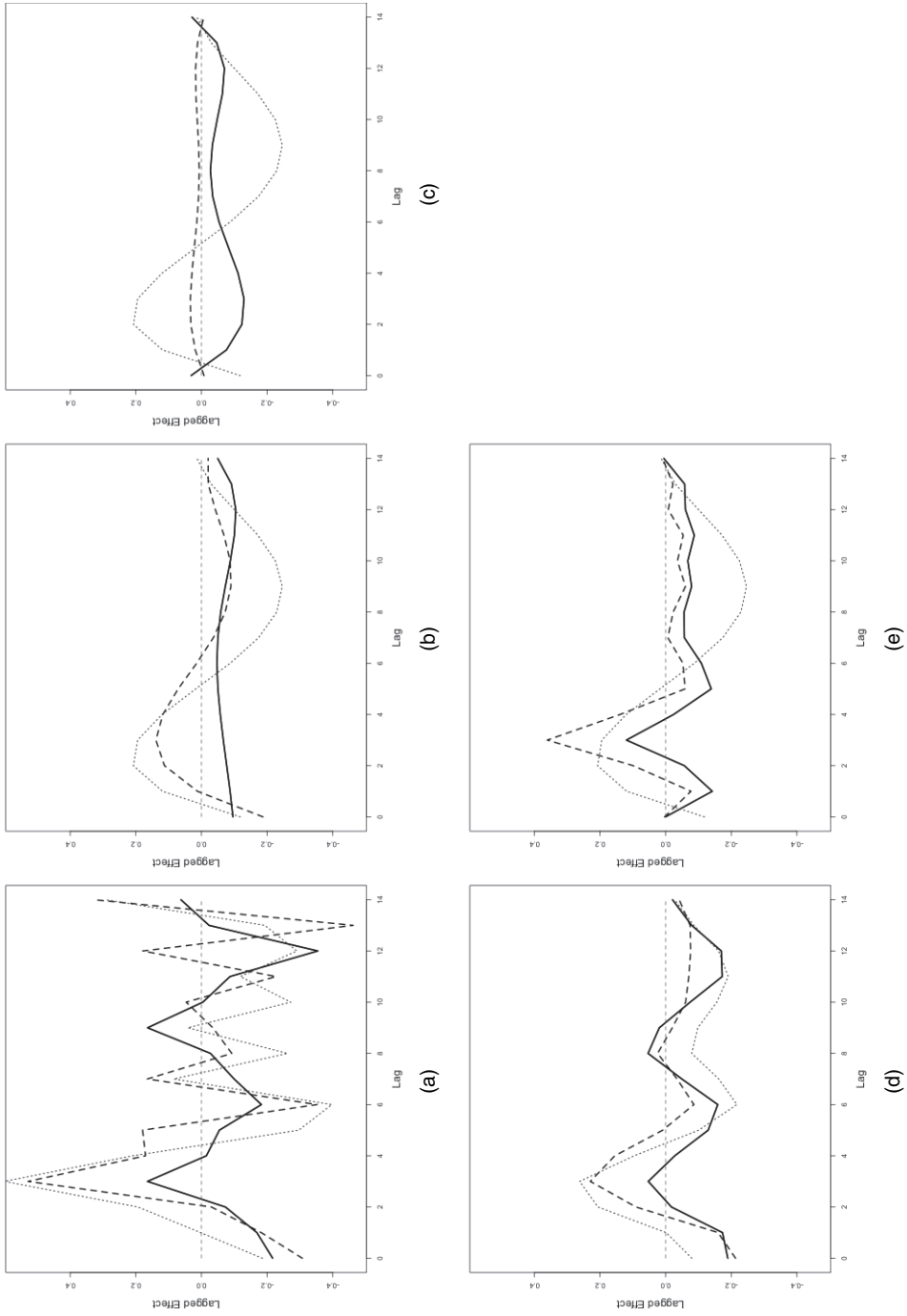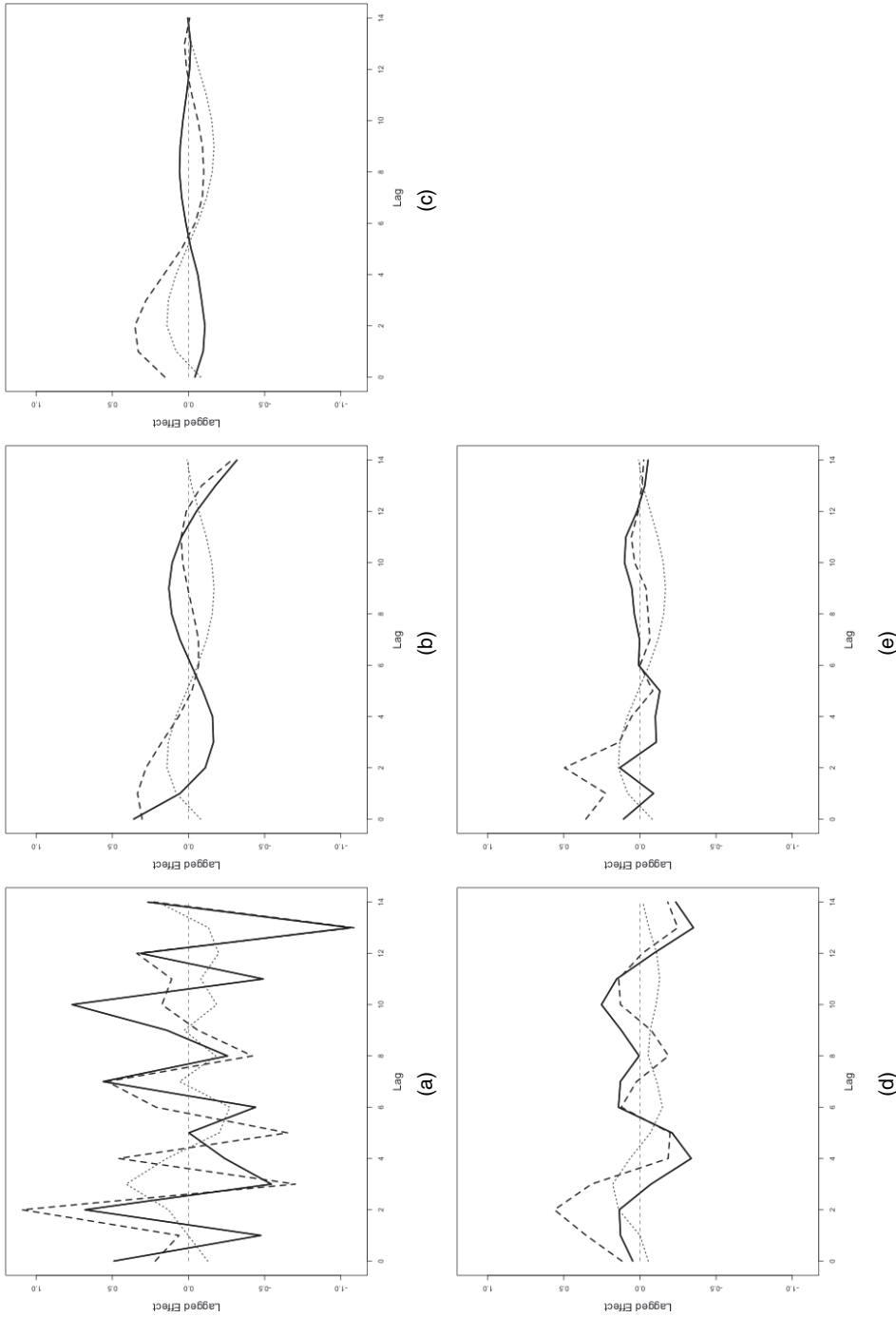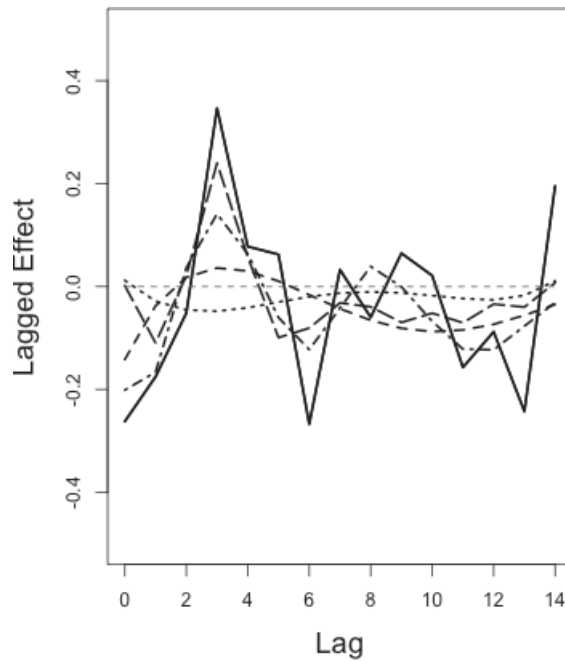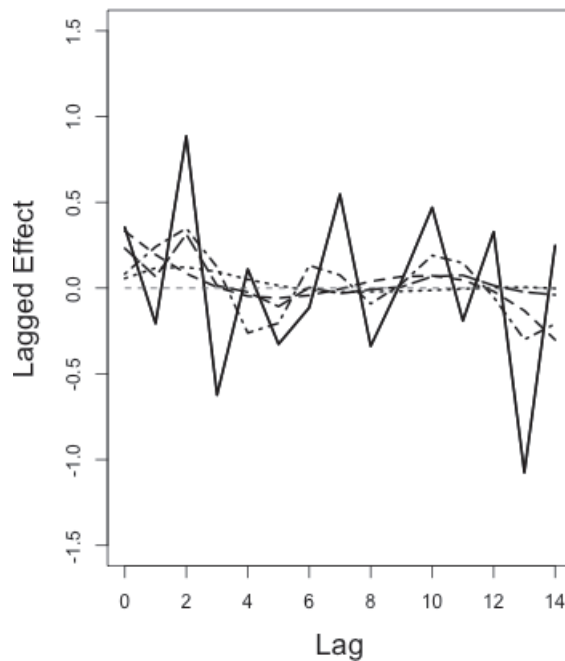
**Fig. 1.** Estimated DL functions up to 14 days for the effects of $PM_{10}$ on mortality in Chicago, Illinois, from 1987 to 2000 based on data from the NMMAPS under five estimation methods (in all panels, $O_3$ levels are fixed at the average series levels in winter (———) and the average series levels in summer (— — —); · · · · · ·, estimated DL function relative to $PM_{10}$ when $O_3$ is disregarded in a single-pollutant model for $PM_{10}$; lag effects are presented as the percentage change in mortality that is associated with a 10 $\mu$g m$^{-3}$ increase in $PM_{10}$ level): (a) UDLIM; (b) CDLIM; (c) TDLIM; (d) BTDLIM; (e) BCDLIM

**Fig. 2.** Estimated DL functions up to 14 days for the effects of $O_3$ on mortality in Chicago, Illinois, from 1987 to 2000 based on data from the NMMAPS under five estimation methods (in all panels, $PM_{10}$ levels are fixed at the average series levels in winter (————) and the average series levels in summer (– – –); · · · · ·, estimated DL function relative to $O_3$ when $PM_{10}$ is disregarded in a single-pollutant model for $O_3$; lag effects are presented as the percentage change in mortality that is associated with a 10 parts per billion increase in $O_3$ level): (a) UDLIM; (b) CDLIM; (c) TDLIM; (d) BTDLIM; (e) BCDLIM

(a)

(b)

**Fig. 3.**   Estimated DL functions up to 14 days for the effects of (a) $PM_{10}$ and (b) $O_3$ on mortality in Chicago, Illinois, from 1987 to 2000 based on data from the NMMAPS under five estimation methods (the DL functions presented here are estimated by integrating out the other pollutant; lag effects are presented as the percentage change in mortality that is associated with a 10 $\mu$g m$^{-3}$ increase in $PM_{10}$ and a 10 parts per billion increase in $O_3$): ———, UDLIM; − − −, CDLIM; ------, TDLIM; - – - –, BTDLIM; —  —, BCDLIM

### 4.3.  Assessing interaction effects

Within each panel of Figs 1 and 2, we note that the estimated DL functions of one pollutant vary with the level of the other pollutant, indicating that $PM_{10}$ might moderate the effect of $O_3$ and vice versa. For the UDLIM, CDLIM and TDLIM, we conducted a likelihood ratio test to test for $PM_{10}$–$O_3$-interactions and the $p$-values are $1.65 \times 10^{-11}$ (225 DFs), $5.33 \times 10^{-9}$ (25 DFs) and less than $10^{-4}$ (1 DF) respectively. The precision of the $p$-value of the TDLIM is only up to $10^{-4}$ because of finite bootstrap samples. For the two shrinkage methods the BTDLIM and BCDLIM, we computed the difference in deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002) between the models with and without interaction. The DIC differences are 25.56 and 68.35 respectively. It is difficult to determine a clear threshold of DIC difference for model selection (Plummer, 2008). However, models with smaller DIC are generally preferred when DIC differences are greater than 10. Coupled with the $p$-values that are obtained from the frequentist approaches, we conclude that the interaction between $PM_{10}$ and $O_3$ is evident.

From Figs 1 and 2, we can see that the summer curves are above the winter curves, suggesting that $PM_{10}$ and $O_3$ have synergistic effects on each other. Furthermore, we observe that the gaps between the curves of the three quartiles decrease beyond lag 6 and that happens across the board. The interaction between $PM_{10}$ and $O_3$ occurs at early lags. We added a dotted curve in each panel for the estimated DL function from a single-pollutant analysis (i.e. models with $PM_{10}$ alone or $O_3$ alone), representing the 'average' DL effects if we disregard the interaction effect between the two pollutants. The evidence in favour of looking at $PM_{10}$ and $O_3$ jointly is compelling.

## 5.  Discussion

In analysing NMMAPS data, we demonstrated the importance of accounting for interaction between the $PM_{10}$ and $O_3$ time series when modelling the joint pollution effect on mortality. Two major pieces of evidence support the existence of pollutant–pollutant interaction—

(a)  the marginal DL function of one pollutant varies when the level of the other changes, and
(b)  the small $p$-values from frequentist approaches and the large DIC values from the Bayesian approaches suggest evidence in favour of a $PM_{10} \times O_3$ interaction.

This adds to the finding of previous studies that supported the idea of a plausible synergism involving $PM_{10}$ and $O_3$ (Mauderly and Samet, 2009).

In this paper, we presented five strategies to model lagged effects of two pollutants in a joint model. We reviewed two existing frequentist methods the UDLIM and CDLIM, and we proposed a frequentist TDLIM using Tukey's interaction structure, its Bayesian version and a Bayesian approach to perform shrinkage between the UDLIM and CDLIM. There are two major novelties. We adopted Tukey's 1 DF interaction structure to model two-way interactions parsimoniously. The estimation is efficient and the interaction testing is powerful. We also introduced the Bayesian version of the TDLIM (i.e. the BTDLIM) and the Bayesian version of the CDLIM (i.e. the BCDLIM). These Bayesian models allow for a departure from a prespecified structure of a DL function or surface and have been shown to be robust to misspecification. They are data adaptive and can achieve bias–variance trade-off.

Each of the five approaches has some limitations that we discuss below. The UDLIM is unbiased but potentially less efficient, especially when the auto-correlation between serial pollution measurement is large. The CDLIM imposes some structure to constrain the lag coefficients and can potentially achieve greater estimation precision. In practice, we recommend a DL structure that is no more complex than a cubic polynomial as the default choice since it is usually sufficient to capture the observed non-linear patterns as a function of the lags. Nonetheless, when the

DL structure is misspecified, the model-dependent CDLIM estimator can be seriously biased. Tukey's type of interaction has mostly been used for hypothesis testing rather than estimation in previous research. Expressing interaction effects as a scaled product of the corresponding main effects implies that the interaction effects can be non-zero only when the main effects are non-zero. This hierarchical feature results in a lack of identifiability for the scaled parameter in Tukey's model when the main effects are not present. In addition, Tukey's model is not invariant to location shifts. Different centring schemes lead to different estimates of the scale parameter $\eta$ and no universal remedy exists.

The hierarchical BCDLIM is robust to misspecification of the DL structure. The data-adaptive shrinkage can be regarded as an automatic procedure to attain a balance between the more general UDLIM and the more constrained CDLIM. The full rank transformation on the UDLIM imposes smoothness on the shrinkage path and any *a priori* knowledge about the DL structure can be incorporated. It is important to note that the BCDLIM can be extended to explore higher order interaction and multiple-pollutant scenarios. We also tried to adapt the HDDLM to two-pollutant scenarios. However, the unmodified predictive process interpolator (Banerjee *et al.*, 2008), which is the major technique that is used in HDDLMs for dimension reduction (Finley *et al.*, 2009), leads to overly smooth DL functions or surfaces which result in seriously biased estimates. We therefore decided not to include the HDDLM in this paper.

The two-pollutant DLIMs can be directly combined with DL non-linear models (Gasparrini *et al.*, 2010) to capture non-linear exposure–outcome associations flexibly by replacing the linear terms in DLM specifications with some basis functions (e.g. *B*-splines). As indicated by He *et al.* (2015), failing to account for non-linear main effects may lead to spurious detection of linear interaction terms. However, when the covariates are correlated as in our application, the signals from non-linear main effects and linear interaction effects may be indistinguishable. In addition, some regularization may be needed in this high dimensional situation to avoid overfitting. We consider this line of extension for future research.

The two-pollutant DLIM approaches that are introduced in this paper can also be extended to multipollutant situations where up to two-way interactions are considered. If one would like to consider higher order interactions and/or non-linear interactions, extension of tree-based approaches such as classification and regression trees and the Bayesian kernel machine regression can be promising. In some situations, choosing the most important pollutants among multiple candidates that are associated with a health outcome is the primary goal.

In real world settings, it is usually difficult to validate the underlying assumptions of a model-based estimator. The notion of data-adaptive shrinkage is attractive when no single estimator is universally optimal. When facing uncertainty, robust models such as the BCDLIM that have better average performance are more desirable. The BCDLIM can potentially be extended to areas outside environmental epidemiology. We hope that our work will lead to more attempts in developing two-dimensional and multi-dimensional DLIMs in the future.

## Acknowledgements

## References

Almon, S. (1965) The distributed lag between capital appropriations and expenditures. *Econometrica*, **33**, 178–196.
Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008) Gaussian predictive process models for large spatial data sets. *J. R. Statist. Soc.* B, **70**, 825–848.

Billionnet, C., Sherrill, D. and Annesi-Maesano, I. (2012) Estimating the health effects of exposure to multi-pollutant mixture. *Ann. Epidem.*, **22**, 126–141.

Bobb, J. F., Valeri, L., Henn, B. C., Christiani, D. C., Wright, R. O., Mazumdar, M., Godleski, J. J. and Coull, B. A. (2014) Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics*, **16**, 493–508.

Chatterjee, N., Kalaylioglu, Z., Moslehi, R., Peters, U. and Wacholder, S. (2006) Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. *Am. J. Hum. Genet.*, **79**, 1002–1016.

Dominici, F., McDermott, A., Daniels, M., Zeger, S. L. and Samet, J. M. (2005) Revised analyses of the National Morbidity, Mortality, and Air Pollution Study: mortality among residents of 90 cities. *J. Toxicol. Environ. Hlth* A, **68**, 1071–1092.

Dominici, F., Peng, R. D., Barr, C. D. and Bell, M. L. (2010) Protecting human health from air pollution: shifting from a single-pollutant to a multi-pollutant approach. *Epidemiology*, **21**, 187–194.

Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L. and Samet, J. M. (2006) Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *J. Am. Med. Ass.*, **295**, 1127–1134.

Dominici, F., Peng, R. D., Ebisu, K., Zeger, S. L., Samet, J. M. and Bell, M. L. (2007) Does the effect of PM10 on mortality depend on PM nickel and vanadium content?: A reanalysis of the NMMAPS data. *Environ. Hlth Perspect.*, **115**, 1701–1703.

Finley, A. O., Sang, H., Banerjee, S. and Gelfand, A. E. (2009) Improving the performance of predictive process modeling for large datasets. *Computnl Statist. Data Anal.*, **53**, 2873–2884.

Gasparrini, A., Armstrong, B., and Kenward, M. G. (2010) Distributed lag non-linear models. *Statist. Med.*, **29**, 2224–2234.

Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Baysn Anal.*, **1**, 515–534.

He, Z., Zhang, H., Lee, S., Smith, J. A., Guo, X., Palmas, W., Kardia, S. L., Roux, A. V. D. and Mukherjee, B. (2015) Set-based tests for genetic association in longitudinal studies. *Biometrics*, **71**, 606–615.

Heaton, M. J. and Peng, R. D. (2014) Extending distributed lag models to higher degrees. *Biostatistics*, **15**, 398–412.

Hu, W., Mengersen, K., McMichael, A. and Tong, S. (2008) Temperature, air pollution and total mortality during summers in Sydney, 1994–2004. *Int. J. Biometeorol.*, **52**, 689–696.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009) The BUGS project: evolution, critique and future directions. *Statist. Med.*, **28**, 3049–3067.

Maity, A., Carroll, R. J., Mammen, E. and Chatterjee, N. (2009) Testing in semiparametric models with interaction, with applications to gene–environment interactions. *J. R. Statist. Soc.* B, **71**, 75–96.

Marx, B. D. and Eilers, P. H. (1998) Direct generalized additive modeling with penalized likelihood. *Computnl Statist. Data Anal.*, **28**, 193–209.

Mauderly, J. L. (1993) Toxicological approaches to complex mixtures. *Environ. Hlth Perspect.*, **101**, 155–165.

Mauderly, J. L. and Samet, J. M. (2009) Is there evidence for synergy among air pollutants in causing health effects? *Environ. Hlth Perspect.*, **117**, 1–6.

Muggeo, V. M. (2007) Bivariate distributed lag models for the analysis of temperature-by-pollutant interaction effect on mortality. *Environmetrics*, **18**, 231–243.

Plummer, M. (2008) Penalized loss functions for Bayesian model comparison. *Biostatistics*, **9**, 523–539.

Pope, C. A. and Dockery, D. W. (2006) Health effects of fine particulate air pollution: lines that connect. *J. Air Waste Mangmnt Ass.*, **56**, 709–742.

Pope, C. A., Dockery, D. W. and Schwartz, J. (1995) Review of epidemiological evidence of health effects of particulate air pollution. *Inhaln Toxicol.*, **7**, 1–18.

R Core Team (2017) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Roberts, S. (2005) An investigation of distributed lag models in the context of air pollution and mortality time series analysis. *J. Air Waste Mangmnt Ass.*, **55**, 273–282.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.

Sun, Z., Tao, Y., Li, S., Ferguson, K. K., Meeker, J. D., Park, S. K., Batterman, S. A. and Mukherjee, B. (2013) Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. *Environ. Hlth*, **12**, 85–103.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *J. R. Statist. Soc.* B, **58**, 267–288.

Tukey, J. W. (1949) One degree of freedom for non-additivity. *Biometrics*, **5**, 232–242.

Welty, L. J., Peng, R., Zeger, S. and Dominici, F. (2009) Bayesian distributed lag models: estimating 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. *Am. J. Epidem.*, **162**, 80–88.

Zanobetti, A., Schwartz, J., Samoli, E., Gryparis, A., Touloumi, G., Atkinson, R., Le Tertre, A., Bobros, J., Celko,

M., Goren, A., Forsberg, B., Michelozzi, P., Rabczenko, D., Aranguez Ruiz, E. and Katsouyanni, K. (2002) The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement. *Epidemiology*, **13**, 87–93.

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

*Supporting information*

Additional 'supporting information' may be found in the on-line version of this article:

   'Supplementary appendix for "Distributed lag interaction models with two pollutants"'.