A Penalized Framework for Distributed Lag Non-Linear Models

Antonio Gasparrini,[1,2,*] Fabian Scheipl,[3] Ben Armstrong,[1] and Michael G. Kenward[2]

[1]Department of Social and Environmental Health Research, London School of Hygiene & Tropical Medicine,
15–17 Tavistock Place, London WC1H 9SH, U.K.
[2]Department of Medical Statistics, London School of Hygiene & Tropical Medicine, Keppel Street,
London WC1E 7HT, U.K.
[3]Department of Statistics, Ludwig Maximilians University, Munich, Germany
*email: antonio.gasparrini@lshtm.ac.uk

SUMMARY. Distributed lag non-linear models (DLNMs) are a modelling tool for describing potentially non-linear and delayed dependencies. Here, we illustrate an extension of the DLNM framework through the use of penalized splines within generalized additive models (GAM). This extension offers built-in model selection procedures and the possibility of accommodating assumptions on the shape of the lag structure through specific penalties. In addition, this framework includes, as special cases, simpler models previously proposed for linear relationships (DLMs). Alternative versions of penalized DLNMs are compared with each other and with the standard unpenalized version in a simulation study. Results show that this penalized extension to the DLNM class provides greater flexibility and improved inferential properties. The framework exploits recent theoretical developments of GAMs and is implemented using efficient routines within freely available software. Real-data applications are illustrated through two reproducible examples in time series and survival analysis.

KEY WORDS: Distributed lag; Exposure-lag-response; Generalized additive models; Latency; Penalized splines.

1. Introduction

Distributed lag models (DLMs), originally proposed in econometrics by Almon (1965) and more recently in epidemiology by Schwartz (2000), constitute an elegant analytical framework to describe associations characterized by a delay between an input and a response in time series data. DLMs model the response y_t observed at time t in terms of past occurrences $x_{t-\ell}$ of a predictor x. The new quantity ℓ, the *lag*, defines a new space that expresses the temporal structure of the association. In standard DLMs, a parametric function is applied to model the shape of the lag structure, usually polynomials or less often regression splines. However, the estimated shape is dependent on the chosen parametric form, for instance, the degree of the polynomial or the number and locations of the splines knots. More sophisticated smoothing techniques have been proposed to address this issue in DLMs, including penalized splines through generalized additive models (GAMs) (Zanobetti et al., 2000; Muggeo, 2008; Rushworth et al., 2013; Obermeier et al., 2015) or Bayesian approaches with the definition of prior distributions (Welty et al., 2009). While these methods offer greater flexibility and more advanced model selection procedures, they rely on the strong assumption of a linear or linear-threshold dose-response relationship, and are only applicable to time series data.

Recent work has addressed these limitations. First, Armstrong (2006) and Gasparrini et al. (2010) extended DLMs to *distributed lag non-linear models* (DLNMs), a framework to describe bi-dimensional *dose-lag-response* associations potentially varying non-linearly in the dimensions of

predictor intensity and lag. Second, Gasparrini (2014) generalized DLMs and DLNMs beyond the time series setting, extending their application to other designs and data structures. However, the current version of DLNMs still requires the user to select the parametric form of the functions expressing the dose-lag-response relationship. Model selection procedures based on information criteria have been proposed, but they lack a solid theoretical basis, and have been shown to partly affect the inferential properties of the estimators (Gasparrini, 2014; Obermeier et al., 2015).

In this contribution, we propose an extended DLNM class developed through penalized splines regression. This development provides greater flexibility for modelling potentially complex bi-dimensional dose-lag-response relationships, and offers built-in model selection and inferential procedures based on recent theoretical work on GAMs. In addition, we extend the methodology further to accommodate a priori assumptions on the shape of the lag structure through the definition of specific penalties. This general framework is applicable to model either linear or non-linear lagged relationships, in various study designs based on either time series or other data structures, and includes most of the models described above as special cases. This extension is fully implemented in freely available software routines.

The article is structured as follows: Section 2 briefly revisits the definition of DLNMs. Section 3 illustrates the extension to a penalized version of DLNMs. In Section 4, we present a simulation study for the assessment of the performance and inferential properties of both standard and extended versions. In Section 5, penalized DLNM are applied in two reproducible

illustrative examples. A final discussion is provided in Section 6. Additional information and results are provided in the Web material.

2. The DLNM Framework

In time series data, DLMs and DLNMs model a response y_t measured at time $t = 1, \ldots, m$ in terms of lagged occurrences of a predictor x_t, represented by the vector $\mathbf{q}_t = [x_{t-\ell_0}, \ldots, x_{t-L}]^\mathsf{T}$, with ℓ_0 and L as minimum and maximum lags, respectively, (Gasparrini et al., 2010). The framework can be extended beyond the time series setting by including an additional indexing structure, allowing each response $y_{i,t}$, with $i = 1, \ldots, n$, to depend on equally spaced lagged values $\mathbf{q}_{i,t} = [x_{i,t-\ell_0}, \ldots, x_{i,t-L}]^\mathsf{T}$. Here, each observation identified by the index i, for instance, a specific subject followed in time, refers to a different *exposure profile* that determines the lagged dose pattern at time t. Time series data represent a special case where $i = t$, while extensions to more complex designs such as survival or repeated-measures longitudinal data are straightforward. See Gasparrini (2014) for a more detailed explanation of the extension beyond time series data and related algebraic definitions.

The association is represented through a function s, defined as:

$$s(\mathbf{q}_{i,t}) = s(x_{i,t-\ell_0}, \ldots, x_{i,t-L}) = \sum_{\ell=\ell_0}^{L} f \cdot w(x_{i,t-\ell}, \ell). \quad (1)$$

Here, the bi-dimensional dose-lag-response function $f \cdot w(x, \ell)$ is composed of two marginal functions: the standard dose-response function $f(x)$, and the additional lag-response function $w(\ell)$ that models the lag structure in the space $\boldsymbol{\ell} = [\ell_0, \ldots, L]^\mathsf{T}$. Parameterization of f and w is obtained by applying known basis transformations to the vectors $\mathbf{q}_{i,t}$ and $\boldsymbol{\ell}$, producing marginal basis matrices $\mathbf{R}_{i,t}$ and $\mathbf{C}$ with dimensions $(L - \ell_0 + 1) \times v_x$ and $(L - \ell_0 + 1) \times v_\ell$, respectively. Identifiability constraints require a reparameterization of $\mathbf{R}$ (see Section 3.2). The function s, here termed *cross-basis* function and parameterized by coefficients $\boldsymbol{\eta}$, is constructed by:

$$s(x_{i,t-\ell_0}, \ldots, x_{i,t-L}; \boldsymbol{\eta}) = (\mathbf{1}_{L-\ell_0+1}^\mathsf{T} \mathbf{A}_{i,t}) \boldsymbol{\eta} = \mathbf{w}_{i,t}^\mathsf{T} \boldsymbol{\eta}, \quad (2)$$

with $\mathbf{w}_{i,t}$ as a set of known transformations derived from $\mathbf{A}_{i,t}$, which in turn is computed by a row-wise Kronecker product between the two basis matrices (Eilers et al., 2006), as:

$$\mathbf{A}_{i,t} = (\mathbf{R}_{i,t} \otimes \mathbf{1}_{v_\ell}^\mathsf{T}) \odot (\mathbf{1}_{v_x}^\mathsf{T} \otimes \mathbf{C}), \quad (3)$$

with $\mathbf{1}_j$ as a vector of 1's with length j. The $n \times (v_x \cdot v_\ell)$ cross-basis matrix $\mathbf{W}$, obtained by applying (1)–(3) to the full set of n observations, can be included in the design matrix of standard regression models, such as generalized linear models (GLMs) or Cox proportional hazard models, to estimate the parameters $\boldsymbol{\eta}$.

The dose-lag-response surface can be recovered by predicting effects $\hat{\beta}_{x,\ell}$ on a grid of predictor values x and lag ℓ. For ease of interpretation, $\hat{\beta}_{x,\ell}$ are defined as specific contrasts of $f \cdot w(x, \ell)$ by centering the dose-response function $f(x)$ to

a reference value of the predictor x. These effects $\hat{\beta}_{x,\ell}$ are interpreted in the usual scale of risk ratio or difference. In particular, the analysis commonly focuses on specific summaries, such as estimated lag-response associations at a given predictor value, or the overall dose-response association obtained by cumulating the risk across the lag period. Algebraic and interpretational details are given elsewhere (Armstrong, 2006; Gasparrini et al., 2010; Gasparrini, 2014).

3. Penalized DLNMs

A penalized extension of DLNM can be described within the family of GAMs (Hastie and Tibshirani, 1990; Wood, 2006a). These models extend the strong parametric form of GLMs by allowing the linear predictor to include flexible smooth functions of the covariates. A recent development of GAMs defines smooth components through penalized regression splines, using low-rank basis terms and a simple form of penalized likelihood (Wood, 2006b). This definition provides theoretically well-grounded estimators implemented using efficient and numerically stable routines (Wood, 2008, 2011). We refer to the two versions of the method as unpenalized and penalized DLNMs, sometimes using the shortcuts GLMs and GAMs, respectively.

3.1. *Penalized Likelihood*

In unpenalized models, the dose-lag-response association defined by a DLNM can be estimated by maximizing the model likelihood $l(\boldsymbol{\eta}, \boldsymbol{\gamma})$ in terms of the model parameters $[\boldsymbol{\eta}^\mathsf{T}, \boldsymbol{\gamma}^\mathsf{T}]^\mathsf{T}$, with $\boldsymbol{\eta}$ corresponding to coefficients of the cross-basis and $\boldsymbol{\gamma}$ to coefficients of additional covariate terms in the model, respectively. The idea underlying the extension to penalized DLNM is to form a richly parameterized cross-basis, and then to apply penalties through its parameters $\boldsymbol{\eta}$ to smooth the dose-lag-response surface. Following similar developments for tensor product bi-dimensional smoothing (Currie et al., 2004), a penalized version $l_p(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ of the model likelihood is obtained by:

$$l_p(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = l(\boldsymbol{\eta}, \boldsymbol{\gamma}) - \frac{1}{2} \boldsymbol{\eta}^\mathsf{T} \Big(\lambda_x \left(\mathbf{S}_x \otimes \mathbf{1}_{v_\ell}^\mathsf{T} \right) + \lambda_\ell \left(\mathbf{1}_{v_x}^\mathsf{T} \otimes \mathbf{S}_\ell \right) \Big) \boldsymbol{\eta}. \quad (4)$$

Here, the penalization of $\boldsymbol{\eta}$ is obtained through *penalty matrices* $\mathbf{S}_x$ and $\mathbf{S}_\ell$ and *penalty* (or *smoothing*) *parameters* $\boldsymbol{\lambda} = [\lambda_x, \lambda_\ell]^\mathsf{T}$ that control the degree of smoothness of the surface. The definition in (4) offers several advantages. First, it allows different degrees of penalization along the two dimensions of the dose-lag-response function $f \cdot w(x_{t-\ell}, \ell)$, by independently calibrating the smoothness in the two marginal spaces x and ℓ through λ_x and λ_ℓ, respectively. In addition, a mix of penalized and unpenalized functions can be defined, for example, when strong parametric assumptions can be made for either $f(x)$ or $w(\ell)$ in (1)–(3), with the exclusion of the related smoothing parameter and penalty matrix from (4). The framework proposed above, therefore, includes previously proposed penalized DLM (Zanobetti et al., 2000; Obermeier et al., 2015) by specifying a linear unpenalized $f(x)$. Extension to models with multiple cross-basis terms or additional penalized terms are straightforward.

3.2. *Choice of the Smoother*

Smooth terms in a GAM can be specified by different *smoothers*, characterized by alternative basis functions and penalties. In penalized DLNMs, the choice of the smoother determines the basis transformations used to generate $\mathbf{R}_{i,t}$ and $\mathbf{C}$ in (2)–(3) and the penalties that form $\mathbf{S}_x$ and $\mathbf{S}_\ell$ in (4). Here, we describe two options, although others are available (Wood, 2006a).

The first smoother, labeled PS, is based on P-splines (Eilers and Marx, 1996), which offer good performance in multidimensional smoothing and both simplicity and flexibility in the penalty definition. The basis matrix of this smoother is composed of v B-splines of degree p, defined by $v + p + 1$ equally spaced knots. The smoothing is obtained by penalizing the difference between coefficients corresponding to adjacent splines using a difference order d. The penalty matrix is derived as:

$$\mathbf{S} = \mathbf{D}_d^\mathsf{T}\mathbf{D}_d, \tag{5}$$

with $\mathbf{D}_d$ as a $(v - d) \times v$ difference matrix of order d. Examples of the first two orders are:

$$\mathbf{D}_1 = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}; \ \mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & \cdots & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -2 & 1 \end{pmatrix}.$$
$$\tag{6}$$

A second smoother, labeled CR, is based on cubic regression splines with penalties on second derivatives. As described in Wood (2006a), for computational convenience the basis matrix of this smoother is derived using a special parameterization of v natural cubic splines, where only one spline is non-zero at each of the v knots. These knots can be placed everywhere along the range of the predictor, and by default at equally spaced quantiles. The smoothing is induced by penalizing the second derivative of the function, with more complex computation required to derive the penalty matrix $\mathbf{S}$ (Green and Silverman, 1994).

The tensor product form of the cross-basis in (1)–(3) requires constraints to ensure the identifiability of the regression parameters. Specifically, the identifiability constraints are absorbed into $\mathbf{R}$ and $\mathbf{S}_x$ through a reparameterization. This step coincides with the exclusion of the intercept from $f(x)$ in unpenalized DLNMs, and ensures that the $n \times (v_x \cdot v_\ell)$ cross-basis matrix $\mathbf{W}$ can be full rank. Additional details are given in Wood (2006a) and the references cited above.

3.3. *Alternative Penalties on the Lag Structure*

Specific assumptions can be made about the shape of the relationship in the lag dimension. These assumptions can be incorporated through additional penalties, which fall into two main categories. First, *varying ridge penalties* can be imposed to shrink different parts of the lag-response curve towards the null value. These type of penalties can be used with either PS

or CR smoothers, and takes two alternative forms:

$$\mathbf{S}_\ell = \mathbf{P}_{v_\ell}, \tag{7a}$$

$$\mathbf{S}_\ell = \mathbf{C}^\mathsf{T}\mathbf{P}_\ell\mathbf{C}. \tag{7b}$$

Here, $\mathbf{P}$ is a pre-specified diagonal matrix of weights p, which in (7a) are applied directly to the v_ℓ coefficients $\boldsymbol{\eta}$, while in (7b) are chosen for the $L - \ell_0 + 1$ lags and mapped into $\boldsymbol{\eta}$ through the basis matrix $\mathbf{C}$ defined in (3). These were previously discussed in Muggeo (2008) and Obermeier et al. (2015).

A second type are *varying difference penalties* that can be applied to enforce a different degree of smoothness along the lag-response curve. These penalties are naturally defined for PS smoothers, and while technically applicable with CR smoothers as well, they are less theoretically grounded for the latter. They take the forms:

$$\mathbf{S}_\ell = \mathbf{D}_d^\mathsf{T}\mathbf{P}_{v_\ell}\mathbf{D}_d, \tag{8a}$$

$$\mathbf{S}_\ell = \mathbf{C}^\mathsf{T}\mathbf{D}_d^\mathsf{T}\mathbf{P}_\ell\mathbf{D}_d\mathbf{C}, \tag{8b}$$

where $\mathbf{P}$ defines weights p for $v_\ell - d$ and $L - \ell_0 + 1 - d$ differences between coefficients in (8a) and lags in (8b), respectively, while $\mathbf{D}_d$ is a matrix defined in (6) of consistent dimensions.

Single or multiple penalties for the lag structure as in (5) or (7)–(8) can be imposed in the same model by defining for each of them the smoothing parameters λ_ℓ and penalty matrices $\mathbf{S}_\ell$ in (4). See Sections 4–5 for specific examples.

3.4. *Estimation*

After the model has been defined by the choice of basis terms and penalty matrices, maximization of the penalized log-likelihood $l_p(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ in (4) is solved through standard estimation methods for GAMs (Wood, 2006a). Briefly, a penalized iterative reweighted least square (P-IRLS) method is integrated with multiple smoothing parameter selection to estimate the degree of smoothness. Alternative methods are available for the estimation of the smoothing parameters $\boldsymbol{\lambda}$ within P-IRLS, such as generalized cross validation (GCV), unbiased risk estimator (UBRE, essentially scaled AIC) and (restricted) maximum likelihood (REML and ML), all implemented using reliable and computationally efficient routines (Wood, 2008, 2011). Simulations indicate that REML and ML are superior in terms of mean-square error performance and smoothing properties (Wood, 2011).

Approximate point-wise confidence intervals of the dose-lag-response surface and its summaries are computed from the estimated posterior (co)variance matrix of the coefficients $\hat{\boldsymbol{\eta}}$, derived using empirical Bayesian estimators (Marra and Wood, 2012). These account for the inherent bias affecting smooth terms and have been shown to provide confidence intervals with *across-the-function* frequentist coverage close to nominal. Although the estimators applied here neglect the uncertainty in the estimation of the smoothing parameters $\boldsymbol{\lambda}$, this has little effect on interval performance in real-data settings (Marra and Wood, 2012).
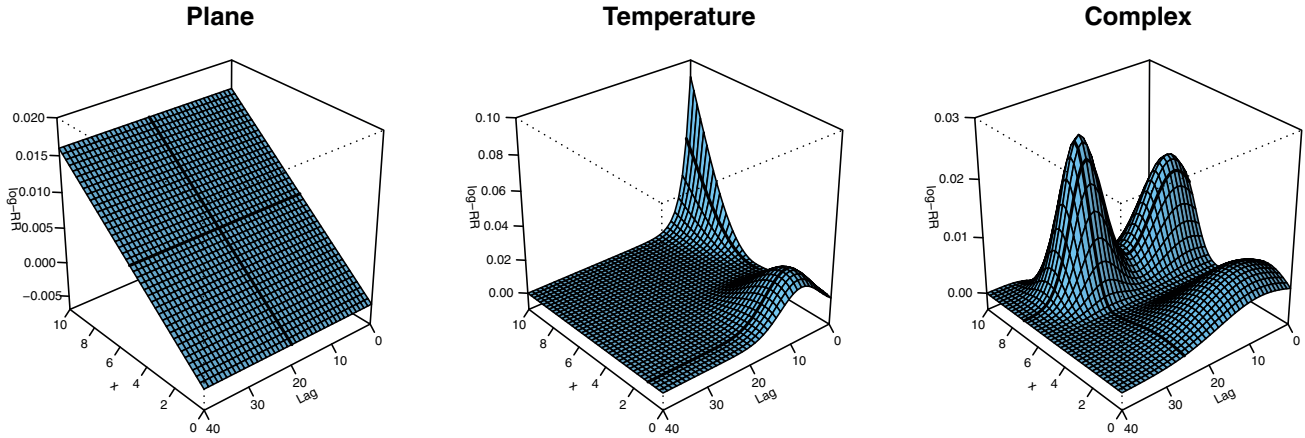
Figure 1. Simulated scenarios representing different bi-dimensional dose-lag-response associations. The bold black lines represent the dose-response and lag-response relationships used to compare the fit of different models in Figure 2.

The smoothness of the dose-lag-response surface can be quantified in terms of *effective degrees of freedom* (*edf*), with the high boundary usually represented by the product of the dimensions of the two marginal basis matrices, $v_x \times v_\ell$ (when $\lambda_x = \lambda_\ell = 0$), and the lower boundary determined by the product of their null space dimensions (when $\lambda_x \to \infty$ and $\lambda_\ell \to \infty$). The null space dimension of each marginal basis is equal to the order of the penalty, that is, the difference order d in PS or the order of derivative (usually 2) in CR smoothers, respectively, minus any constraint (Wood, 2006a).

4. Simulation Study

To assess the performance and inferential properties of different versions of penalized DLNMs and to compare them with the standard unpenalized approach, we performed a simulation study based on scenarios of dose-lag-response surfaces with varying degree of complexity.

4.1. Simulation Setting

The predictor x_t was represented by the daily temperature series in Chicago within the period 1987–2000 (Samet et al., 2000), standardized over the range 0–10. For each replicate, we simulated an outcome series y_t of daily mortality counts, with $t = 1, \ldots, 5114$, from a Poisson distribution with mean μ_t, using:

$$\log(\mu_t) = \alpha_j + \sum_{\ell=0}^{40} f_j \cdot w_j(x_{t-\ell}, \ell). \tag{9}$$

We repeated the simulations in three scenarios $j = 1, 2, 3$, with the dose-lag-response function $f_j \cdot w_j(x, \ell)$ over lag 0–40 described by:

- Scenario 1: a simple plane;
- Scenario 2: a shape resembling previously estimated temperature-mortality associations;
- Scenario 3: a complex wiggly surface.

A graphical representation of these three scenarios is offered in Figure 1, with algebraic details provided in Web Appendix A. The intercept α_j was used as a signal-to-noise parameter

to achieve a Pearson correlation coefficient between μ_t and y_t of approximately 0.5 in each scenario.

For each simulated series, we fitted alternative models where the second term in (9) is replaced by a cross-basis $s(x_t, \ldots, x_{t-40})$. The primary model, simply labeled GAM, used a penalized DLNM with PS smoothers of rank $v = 10$ (minus constraints), degree 3 (cubic B-splines) and second-order difference ($d = 2$) penalties for each marginal dimension, estimated by REML. Previous research (Wood, 2006a) indicates that basis dimension is not critical, if large enough to fit the underlying marginal shape, while the smoother and estimator were chosen for their flexibility and inferential performance, respectively. This model was compared with:

- *Alternative estimators:*
  - GLM-AIC, defined by (unpenalized) quadratic B-spline functions with the optimal number of equally spaced knots selected by minimizing AIC among combinations producing 1–10 *df* (minus constraints) in each dimension;
  - GAM-AIC, fitted by replacing REML with a UBRE-AIC estimator.
- *Alternative smoothers*:
  - GAM-CR, defined by replacing the PS with CR smoothers;
  - GAM-PS$_{2,1}$, with difference penalties of order 2 and 1 for $f(x)$ and $w(\ell)$, respectively.
- *Additional/alternative penalties for $w(\ell)$:*
  - GAM-ADD$_{LAST}$, with an additional varying ridge penalty as in (7a) with $\mathbf{p} = [0, 0, 0, 0, 0, 0, 1, 1, 1, 1]^\mathsf{T}$;
  - GAM-ALT$_{QUAD}$, that replaced entirely the penalty with a varying difference penalty as in (8b), with $p_\ell = \ell^2$;
  - GAM-ALT$_{EXP}$, that replaced entirely the penalty with a varying ridge penalty as in (7a), with $p_k = \exp(k - 1)$ and $k = 1, \ldots, v_\ell - d$.

These additional/alternative penalties follow the assumption of a lag-response that approaches the null value at the end of the lag period, or that is smoother at longer lags. See Muggeo (2008) and Obermeier et al. (2015) for details.

We assessed the performance of the eight models above using 1000 simulation replicates in each of the three scenarios depicted in Figure 1, by comparing the across-the-surface cov-

Table 1

Results of the simulation study with average time (seconds, using a 2.4 GHz PC), equivalent degrees of freedom (edf), coverage, and root mean square error (RMSE, relative to the GAM-REML model) for each scenario depicted in Figure 1 across alternative models in 1000 replicates

	Time	Scenario 1			Scenario 2			Scenario 3		
		(e)df	cov	RMSE	(e)df	cov	RMSE	(e)df	cov	RMSE
Alternative estimators										
GAM	3.70	2.81	0.97	1.00	27.87	0.91	1.00	19.42	0.92	1.00
GLM-AIC	7.94	2.93	0.83	2.67	30.42	0.85	1.41	22.87	0.81	1.83
GAM-AIC	5.73	4.54	0.96	1.77	30.49	0.91	1.10	22.93	0.95	1.10
Alternative smoothers										
GAM-CR	4.72	2.97	0.97	1.02	37.05	0.95	1.03	24.32	0.94	1.00
GAM-PS$_{2,1}$	3.83	1.72	0.98	0.63	28.26	0.91	1.03	18.32	0.92	0.91
Additional/alternative penalties for $w(\ell)$										
GAM-ADD$_{\text{LAST}}$	4.74	2.73	0.95	0.98	20.44	0.93	0.76	16.64	0.89	0.88
GAM-ALT$_{\text{QUAD}}$	3.66	2.87	0.97	1.02	25.83	0.92	0.92	19.85	0.91	1.03
GAM-ALT$_{\text{EXP}}$	4.10	7.45	0.90	2.68	25.20	0.94	0.80	23.15	0.97	0.90

erage and root mean square error (RMSE) (defined in Web Appendix B, see also Marra and Wood (2012)) using estimated effects $\hat{\beta}_{x_p,\ell_p}$ computed on a grid of predictor values $x_p = 0, 0.25, \ldots, 9.75, 10$ and lag values $\ell_p = 0, \ldots, 40$.

4.2. Results of the Simulation Study

Results are illustrated in Table 1 and Figure 2. Table 1 reports the average computing time and *edf*, the empirical coverage of 95% confidence intervals and the empirical RMSE relative to the GAM model across the surfaces. Figure 2 displays the estimated dose-response and lag-response curves corresponding to the bold black lines across the surfaces in Figure 1 for the models GAM, GLM-AIC, and GAM-ADD$_{\text{LAST}}$. The same graphical representation for the other models is provided in Web Figures S2–S3 in the online Supporting Information.

In all the scenarios, penalized DLNMs appear superior to the unpenalized counterpart. In particular, GLM-AIC shows higher RMSE (as also suggested by the wigglier curves in Figure 2), and a substantial under-coverage due to unaccounted additional variability of the model selection procedure, consistently with what previously reported (Sylvestre and Abrahamowicz, 2009; Gasparrini, 2014). The REML estimator exhibits a slightly better performance if compared to UBRE-AIC in GAM-AIC, with the latter showing higher RMSE and some evidence of undersmoothing, especially in the simplest scenario. Alternative smoothers in GAM-CR and GAM-PS$_{2,1}$ provide similar outputs, with the latter performing better in the plane scenario, which is consistent with its null space of 1 *edf*.

The model GAM-ADD$_{\text{LAST}}$ shows an improved performance in the second scenario, where the extended flat region (see Figure 2) is well fitted through the addition of a varying ridge penalty, which also helps identify the correct lag period even when the interval is extended well beyond it, as previously reported (Obermeier et al., 2015). This doubly penalized model performs well also in the other scenarios that do not match the assumption of the penalty, with only a minor bias produced in the plane scenario, as noticeable in the last part of the estimated lag-response curves in Figures 2 and S3. This

good performance is due to the possibility of virtually removing the additional penalty by estimating a very low smoothing parameter λ_ℓ. Models GAM-ALT$_{\text{QUAD}}$ and GAM-ALT$_{\text{EXP}}$, where the standard penalty was removed, perform well in the second and third scenarios, but the latter fails to fit the plane dose-lag-response surface, which is not compatible with its strong assumptions about form of the lag-response shape (see Web Figure S3).

Generally, penalized models show across-the-surface coverage close to the nominal value, although some undercoverage is evident for some models in the second scenario (see also Web Figures S4–S5 in the online Supporting Information). In addition, GAM-AIC fails to converge in 1.4% of replicates of the plane scenario, where the simulated surface represents the null space dimension of the tensor product smoother. However, the analysis of non-convergent models does not identify problems with point estimates and coverage.

5. Two Examples

As an illustration of the application of penalized DLNMs in different study designs, we replicate two published analyses. The reader can refer to the original publications for details on the analytical methods and data (Gasparrini and Leone, 2014; Gasparrini, 2014).

5.1. Outdoor Temperature and All-Cause Mortality

The first example illustrates the application of penalized DLNMs in time series data, using daily series from London in the period 1993–2006. Specifically, the relationship between counts of all-cause mortality y_t at day t and outdoor temperature $x_{t-\ell}$, accounting for up to 25 days of lag, was estimated with a quasi-Poisson GLM of form:

$$\log[\text{E}(y_t)] = \alpha + s(x_t, \ldots, x_{t-25}; \boldsymbol{\eta}) + g(t; \boldsymbol{\gamma}) + \sum_{j=1}^{6} \delta_j w_{j,t}, \quad (10)$$

with g as natural cubic splined defined by 10 *df*/year accounting for seasonal and long term trends, and w_j as an indicator of day of the week. In the original analysis (Gasparrini and
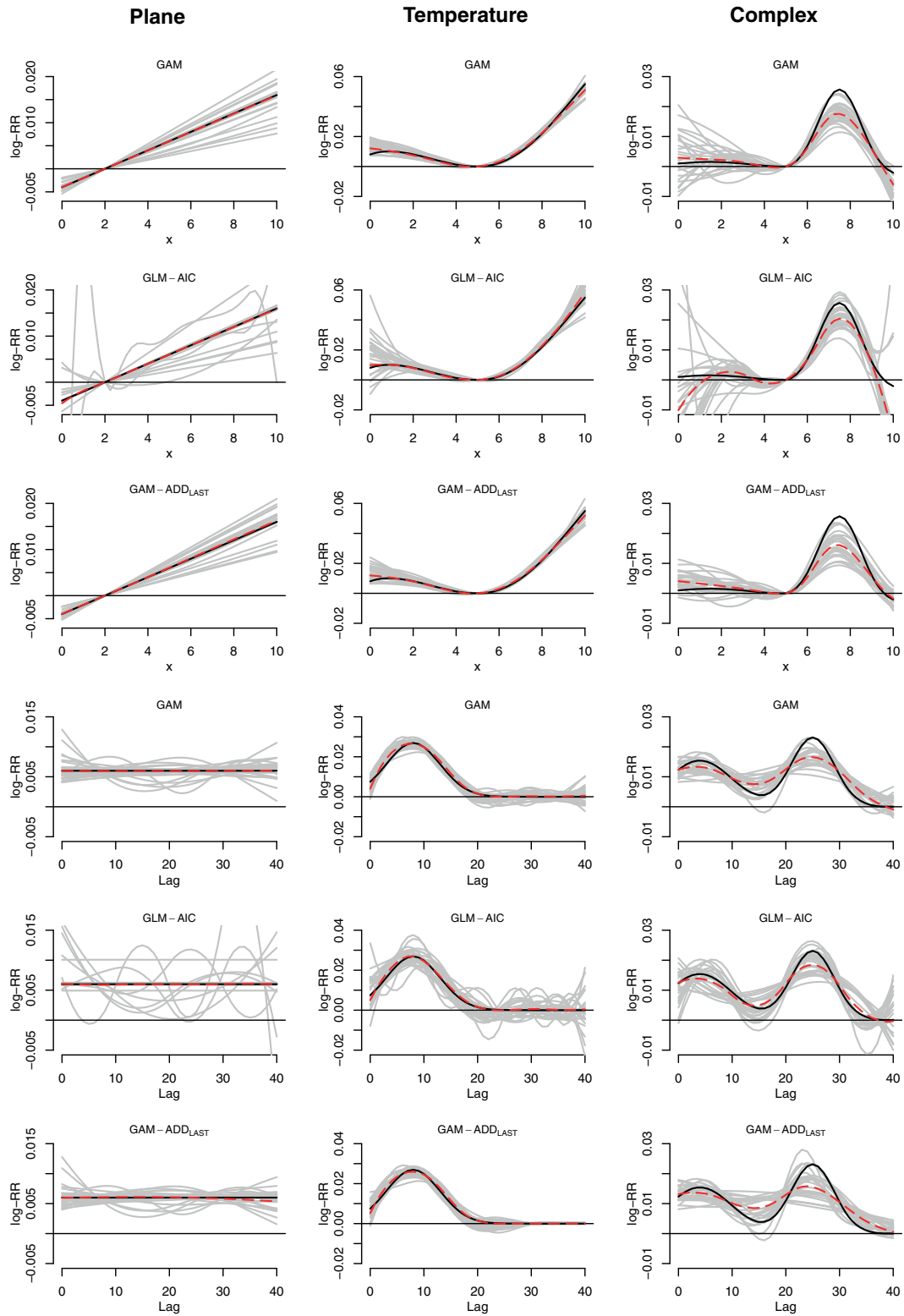
Figure 2. Results of the simulation study, illustrating the performance of three different models (see Table 1) in 1000 replicates. The panels represent the dose-response (rows 1–3) and lag-response curves (rows 4–6) corresponding to the bold black lines in the three simulated surfaces (by column) in Figure 1. Continuous gray, and dashed red and continuous black lines represent the fit from 25 random replicates, the average across all replicates, and the true simulated curves, respectively.

Leone, 2014), the dependency was modelled with an unpenalized DLNM within a GLM, using a cross-basis function s with $4 \times 5 = 20$ *df* composed of quadratic B-splines defined by 2 equally spaced internal knots for the dose-response function $f(x)$ and natural cubic splines by three equally spaced internal knots in the log scale plus intercept for the lag-response function $w(\ell)$. Boundary knots were placed by default at the ranges.

We replicated the analysis using a penalized DLNM within a GAM with a REML estimator, specifying marginal PS smoothers with dimension 10 (minus constraints) for both spaces. Penalization of $f(x)$ was enforced through a default second-order difference penalty as in (5). Extending previous models (Muggeo, 2008; Obermeier et al., 2015), we applied a double varying penalty to $w(\ell)$ using a second-order difference form (8b) with $p_k = k^2$ for $k = 0, \ldots, 23$, and a ridge penalty of form (7a) with $\mathbf{p} = [0, 0, 0, 0, 0, 0, 1, 1, 1, 1]^\mathsf{T}$. These choices are motivated by the assumption of a shape that is smoother at longer lags and approaches the null value at the end of the lag period.

The GAM used 35.45 *edf* to model the dose-lag-response surface, and suggests a strong and short-term association with heat and a more delayed association with cold temperatures, consistently with previous results (Gasparrini et al., 2015). The estimates, reported in the first row of Figure 3, are very similar to those from the original analysis, replicated in the second row. However, it is interesting to note the effect of the double varying penalty in the estimated lag-response at 29 °C, with the curve shrinking toward the null at lags higher than 15. In addition, while the cross-basis specification of the unpenalized DLNM was originally defined a priori, an AIC-based selection suggests a very complex and implausible model with $10 \times 9 = 90$ *df*, with estimates illustrated in the third row of Figure 3.

As previously mentioned (Section 3.1), the flexibility of this modelling framework allows a mix of penalized and unpenalized functions. As an example, we replaced the PS smoother for $f(x)$ with an unpenalized double-threshold function, that is, linear splines which model a straight relationship below 17 °C and above 21 °C, and a flat region in between. Results are displayed in the last row of Figure 3. This model uses only 10.64 *edf* to define the dose-lag-response surface, although this comes at the price of making strong parametric assumptions for one of the two spaces. The same approach can be used to specify simpler penalized DLMs, by selecting a linear function as $f(x)$.

5.2. *Occupational Radon and Lung Cancer Mortality*

The second example describes the extension of penalized DLNMs to individual time-to-event data, using a cohort of 3347 miners working in the Colorado Plateau mines, with follow-up at December 31, 1982. Specifically, the association between an indicator of occurrence of lung cancer death $y_{i,t}$ for subject i at age t, and yearly occupational radon exposure $x_{i,t-\ell}$, measured in working-level months (WLM), with lags of 2–40 years, was estimated with a Poisson GLM of form:

$$\log[\mathrm{E}(y_{i,t})] = \alpha + s_x(x_{i,t-2}, \ldots, x_{i,t-40}; \boldsymbol{\eta}_x)$$
$$+ s_z(z_{i,t-2}, \ldots, z_{i,t-40}; \boldsymbol{\eta}_z) + g(t; \boldsymbol{\gamma}) + \delta c_{i,t}.$$
$$(11)$$

This GLM approximates the Cox proportional hazard model applied in the original analysis (Gasparrini, 2014) by splitting the follow-up time of each individual into 1-year periods, and modelling the baseline risk with a cubic B-spline function $g(t)$ with 5 *df*. This allows the use of penalized splines implemented within GAMs with survival data. Other terms in the model are a cross-basis function s_z to control for the lagged effect of smoking z, and a linear term for calendar year c. In the original analysis, the association with radon was modelled with a cross-basis function s_x composed of quadratic B-splines functions with a single internal knot at 59.4 WLM/year and 13.3 years of lag for $f(x)$ and $w(\ell)$, respectively, and boundary knots at the respective ranges. The intercept was excluded from the latter, assuming no effect for exposures experienced within the first two years. This model, using a total of 9 *df* to define the association, was selected by minimizing AIC.

The analysis was replicated with a penalized DLNM using a GAM with a REML estimator, using marginal CR smoothers with dimension 11 (minus constraints) and 10 for exposure and lag spaces, respectively. The use of the CR smoother allows placing the knots of the dose-response function $f(x)$ at equally spaced intervals in the log scale, accounting for the highly skewed distribution of radon exposure, and allows excluding the intercept in $s(\ell)$ following previous assumptions. In addition to the default penalty on the second derivative, enforced in both spaces, we added a varying ridge penalty of form (7b) to $w(\ell)$ with $p_\ell = 1$ if $\ell > 30$ and 0 otherwise, thus assuming no additional risk 30 years after the exposure to radon.

Results are displayed in Figure 4. The penalized DLNM (first row) indicates a peak in lung cancer mortality risk approximately 11 years after the exposure to radon. The non-linear dose-response shows how the risk flattens out above 50 WLM/year. The model used a total of 8.03 *edf* to describe the association. These findings are consistent with the unpenalized DLNM fitted with a GLM (second row of Figure 4), which closely approximates the original estimates from the Cox model illustrated by Gasparrini (2014, Figure 2). However, the addition of the ridge penalty in the GAM produces more precise estimates at the end of the lag period, suggesting that the risk completely disappears 30 years after the exposure.

6. Discussion

In this contribution, we describe a penalized framework for DLNMs that provides significant developments to this modelling class, through built-in smoothness selection of potentially complex marginal functions and the flexible definition of penalties to accommodate assumptions on the lag structure. This method includes previous smoothing approaches for simpler DLMs (Zanobetti et al., 2000; Obermeier et al., 2015) as special cases, and fully extends the penalized approaches to bi-dimensional dose-lag-response surfaces. The DLNM framework unifies methods proposed to investigate lagged associations in different research fields, beyond time series analysis in environmental research. For instance, these include case-control studies in cancer epidemiology (Thomas, 1988; Hauptmann et al., 2000; Berhane et al., 2008; Richardson, 2009) and survival analysis in
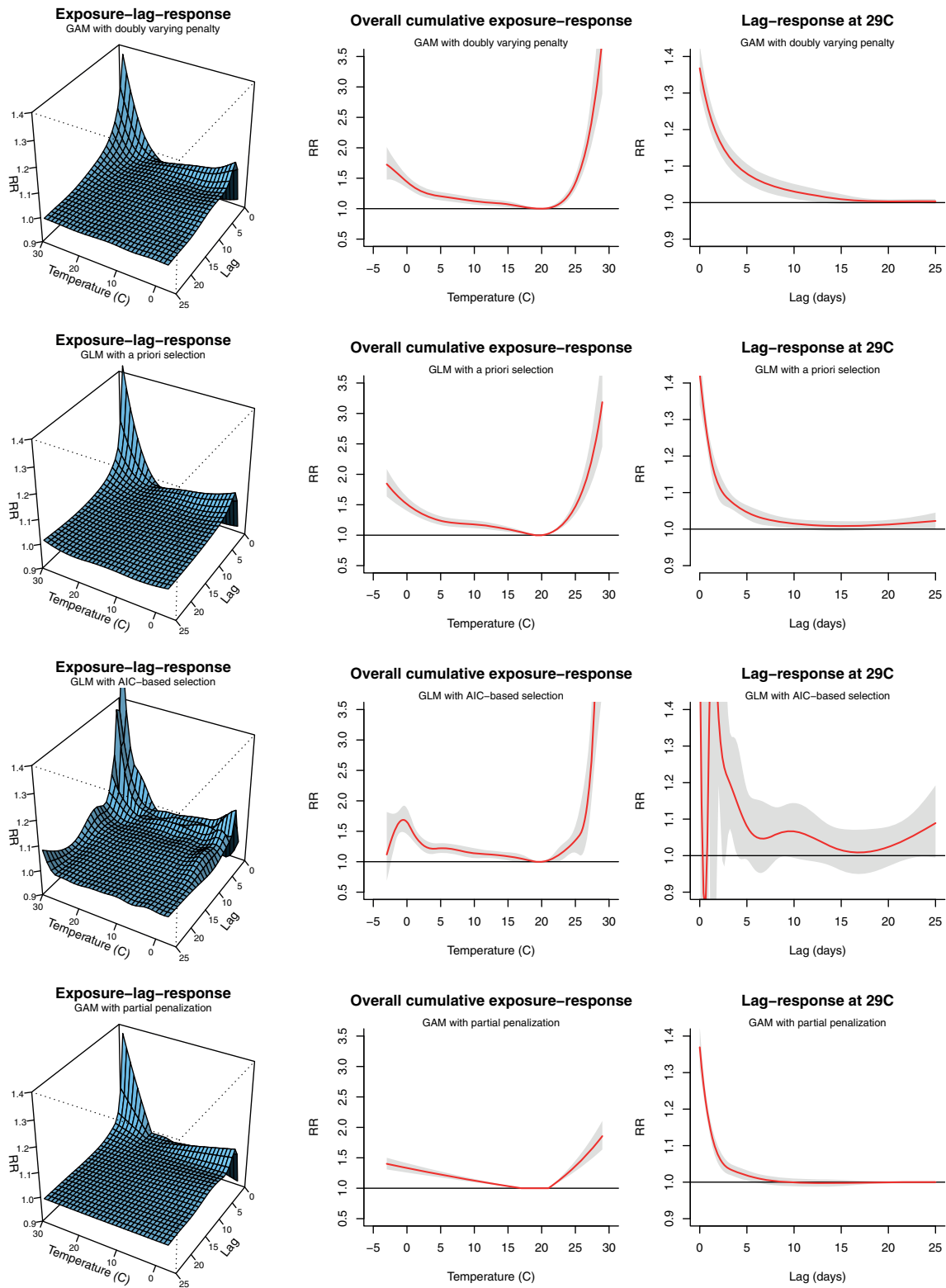
Figure 3. First example: dose-lag-response, overall cumulative dose-response, and lag-response at 29 °C (by column) summarizing the association between temperature and all-cause mortality, estimated by a GAM with double varying penalty in the lag space, GLM with a priori selection (as in Gasparrini and Leone (2014)), GLM with AIC-based selection, GAM with partial penalization (by row). London 1993–2006.
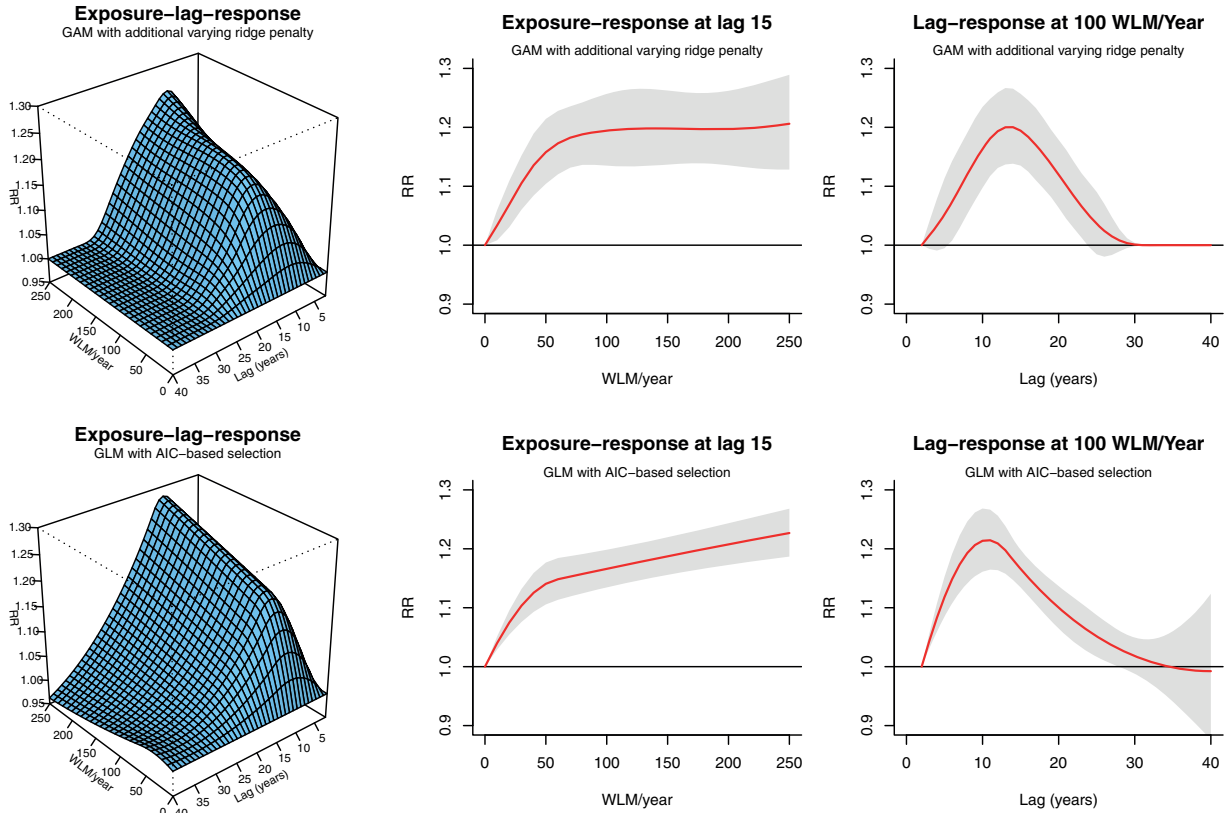
Figure 4. Second example: dose-lag-response, dose-response at lag 15, and lag-response at 100 WLM/year (by column) summarizing the association between occupational radon exposure and lung cancer mortality, estimated by a GAM with an additional varying ridge penalty in the lag space and a GLM with AIC-based selection (as in Gasparrini (2014)) (by row). Colorado Plateau Uranium miners cohort, follow-up at December 31, 1982.

pharmaco-epidemiology (Sylvestre and Abrahamowicz, 2009; Abrahamowicz et al., 2012)

This penalized version addresses the problem of choosing the appropriate degree of complexity of the DLNM. This is a critical limitation of traditional unpenalized DLNMs, for which current selection methods are not effective (as demonstrated with the first example in Section 5.1), and produce less efficient estimators (as illustrated in the simulation study in Section 4). This penalized extension is based on well-grounded theoretical results and estimation methods, recently discussed (Wood, 2006a, 2008, 2011), it can be performed with stable and efficient routines implemented in freely available software (Wood, 2006a), and it shows improved inferential properties if compared to the standard unpenalized version.

The results confirm the good inferential properties of REML and UBRE-AIC estimators, with the former appearing slightly superior (Wood, 2011), and the similar performance of alternative types of smoothers (Wood, 2006a). The latter can be selected due to convenient characteristics, such as the possibility of including varying difference penalties with the PS smoother (see Section 5.1) or the flexibility in the knots placement and exclusion of intercept with the CR smoother (see Section 5.2). In particular, the inclusion of additional penalties on the lag dimension provides a way to accommodate realistic assumptions on the underlying shape. These addi-

tional penalties can be selected based on prior knowledge, and do not represent strong constraints on the lag-response shape, as their influence can be calibrated through the estimate of smoothing parameters. As previously suggested (Muggeo, 2008; Obermeier et al., 2015) and shown in the second scenario of the simulation study, additional penalties can improve the model fit and make the model less sensitive to the choice of the lag period.

Some limitations must be acknowledged. The issue of penalizing complex bi-dimensional functions has been investigated in a limited set of simulated scenarios and two real-data examples. Also, simulations show some issue with nonconvergence in the simplest scenario, where the selected *edf* tend to be close the null space of the cross-basis function, although this problem does not seem to critically affect inference. The penalized approach substantially improves the coverage properties of the confidence intervals, even though in some scenarios and for some models the empirical coverage falls short of the nominal value. In addition, the method presented here shares a known limitation of GAMs, which tend to select simpler (i.e., smoother) models when the statistical power decreases. Finally, smoothing methods for dose-lag-response relationships are difficult to validate, as the lag dimension is not directly observable in the data, thus preventing the use of standard techniques such as residual analysis. These issues

will be hopefully addressed in future research.

Penalized DLNMs can be further extended to varying-coefficients models, describing dose-lag-response relationships that change in time or along the space of other predictors, as previously described for simpler penalized DLMs (Rushworth et al., 2013) or unpenalized DLNMs (Gasparrini et al., 2015). In addition, the DLNM class has interesting links with penalized functional regression, where a functional outcome (say the shape of the dose-response) is allowed to vary depending on a functional predictor (say the lag dimension) (McLean et al., 2014; Scheipl et al., 2015), providing the possibility of further extensions through this established modelling framework.

Lagged associations occur almost universally in biomedical research, and well beyond. Penalized DLNMs offer a flexible modelling class to describe these phenomena, avoiding biases due to incorrect assumptions about the lag structure, sometimes made when using simpler approaches, and potentially extending the knowledge of the association under study. The recent extension of DLNMs beyond time series data (Gasparrini, 2014) unifies and extends methods proposed in different study designs and paves the way for original and promising applications of this modelling framework.

7. Supplementary Materials

Web Appendices, Web Figures, and R code are available at the *Biometrics* website on Wiley OnlineLibrary. In addition to Web Appendices A–B, referenced in Section 4, Web Appendix C briefly describes the software implementation in the R package dlnm. The R code fully reproduces the simulation studies and the two examples, with an updated version available from GitHub and the personal website of the first author (see Web Appendix C).

Conflict of Interest
None.

References

Abrahamowicz, M., Beauchamp, M. E., and Sylvestre, M. P. (2012). Comparison of alternative models for linking drug exposure with adverse effects. *Statistics in Medicine* **31**, 1014–1030.

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica* **33**, 178–196.

Armstrong, B. (2006). Models for the relationship between ambient temperature and daily mortality. *Epidemiology* **17**, 624–631.

Berhane, K., Hauptmann, M., and Langholz, B. (2008). Using tensor product splines in modeling exposure-time-response relationships: Application to the Colorado Plateau Uranium Miners cohort. *Statistics in Medicine* **27**, 5484–5496.

Currie, I. D., Durban, M., and Eilers, P. H. C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* **4**, 279–298.

Eilers, P. H. C., Currie, I. D., and Durban, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis* **50**, 61–76.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–101.

Gasparrini, A. (2014). Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in Medicine* **33**, 881–899.

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

Gasparrini, A., Guo, Y., Hashizume, M., Kinney, P. L., Petkova, E. P., Lavigne, E., et al. (2015). Temporal variation in heat-mortality associations: A multi-country study. *Environmental Health Perspectives* **123**, 1200–1207.

Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., et al. (2015). Mortality risk attributable to high and low ambient temperature: A multicountry observational study. *The Lancet* **386**, 369–375.

Gasparrini, A. and Leone, M. (2014). Attributable risk from distributed lag models. *BMC Medical Research Methodology* **14**, 55.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Monographs on statics and applied probability 58. London: Chapman & Hall.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, 2nd edition, London: Chapman & Hall/CRC,.

Hauptmann, M., Wellmann, J., Lubin, J. H., Rosenberg, P. S., and Kreienbrock, L. (2000). Analysis of exposure-time-response relationships using a spline weight function. *Biometrics* **56**, 1105–1108.

Marra, G. and Wood, S. N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics* **39**, 53–74.

McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F., and Ruppert, D. (2014). Functional generalized additive models. *Journal of Computational and Graphical Statistics* **23**, 249–269.

Muggeo, V. M. (2008). Modeling temperature effects on mortality: Multiple segmented relationships with common break points. *Biostatistics* **9**, 613–620.

Obermeier, V., Scheipl, F., Heumann, C., Wassermann, J., and Kuhchenhoff, H. (2015). Flexible distributed lags for modelling earthquake data. *Journal of the Royal Statistical Society: Series C* **64**, 395–412.

Richardson, D. B. (2009). Latency models for analyses of protracted exposures. *Epidemiology* **20**, 395–399.

Rushworth, A. M., Bowman, A. W., Brewer, M. J., and Langan, S. J. (2013). Distributed lag models for hydrological data. *Biometrics* **69**, 537–544.

Samet, J. M., Zeger, S. L., Dominici, F., Dockery, D., and Schwartz, J. (2000). The National Morbidity, Mortality, and Air Pollution Study (NMMAPS). Part 1. Methods and methodological issues. Technical report, Health Effects Institute.

Scheipl, F., Staicu, A.-M., and Greven, S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics* **24**, 477–501.

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

Sylvestre, M. P. and Abrahamowicz, M. (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine* **28**, 3437−3453.

Thomas, D. C. (1988). Models for exposure-time-response relationships with applications to cancer epidemiology. *Annual Review of Public Health* **9**, 451−482.

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

Wood, S. N. (2006a). *Generalized Additive Models: An Introduction with R*. Boca Raton, FL: Chapman & Hall/CRC.

Wood, S. N. (2006b). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics* **62**, 1025−1036.

Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society, Series B* **70**, 495−518.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semi-parametric generalized linear models. *Journal of the Royal Statistical Society, Series B* **73**, 3−36.

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