UNIVERSITY OF LEEDS

Article:
Aykroyd, RG orcid.org/0000-0003-3700-0816 (2019) Bayesian modeling of temperature-related mortality with latent functional relationships. Communications in Statistics - Theory and Methods, 48 (1). pp. 3-14. ISSN 0361-0926

https://doi.org/10.1080/03610926.2017.1421223

Bayesian modeling of temperature-related mortality with latent functional relationships

Robert G Aykroyd

Department of Statistics, University of Leeds, Leeds, LS2 9JT, UK

(e-mail: r.g.aykroyd@leeds.ac.uk)

December 12, 2017

Abstract

It is common for the mortality rate to increase during periods of extreme temperature and for the minimum mortality rate to depend on factors such as the mean summer temperature. In this paper, local correlation is explicitly described using a generalized additive model with a spatial component which allows information from neighbouring locations to be combined. Random walk and random field models are proposed to describe temporal and spatial correlation structure. Further, joint spatial-temporal modelling is proposed by including a temperature-related mortality term. This will make use of existing data more efficiently and should reduce prediction variability. The methods are illustrated using simulated data based on real mortality and temperature data.

Keywords: Bayesian methods, demography, generalised additive models, maximum likelihood, spatial-temporal.

1 Introduction

There are several important implications of the observation that death rates increase under usual extremes of temperature (National Academy of Sciences, 2010; McMichael and Coauthors, 2008). Gasparrini et al. (2015) report a large-scale study using data from 384 locations and have shown that this is a worldwide issue. In particular, this should influence local government policy on current social welfare, but also it has longer term worldwide implications in the face of global warming. The realisation that this will have a critical influence on human life cannot be ignored. It has, however, also been noted that the minimum mortality rate is not equal in different regions, and in particular that this appears to be related to local average temperature. Hence, a possible conclusion is that mortality rates might not increase if an overall change in mean temperature takes place over an extended period of time, but will have a significant effect if the annual temperature profile becomes more volatile — that is the difference between summer highs and winter lows increases. To produce reliable information on which to base critical decisions is therefore of vital importance.

Although the modelling approach is described in terms of a spatial-temporal mortality data structure, it can equally be used to described a wide range of other, apparently, unrelated problems and hence has more general importance.

2 Model definitions

2.1 Data structure

Consider data collected at a set of n known locations, $\{\boldsymbol{s}_i : i = 1, \ldots, n\}$, within some bounded region, $\mathcal{F} \subset \mathbb{R}^2$ – such as in Figure 1(a)[‡]. At each location, other variables are recorded over a long period which might simply be labelled sequentially as $j = 1, \ldots, N$. The main variables of interest are mortality, m_{ij}, and temperature, t_{ij} – as in Figure 1(b)[‡]. It is assumed that there are also M demographic and socio-economic covariates which will

be denoted $x_{ijk}, k = 1, \ldots, M$. This leads to a full dataset in terms of the following variables:

$$\begin{aligned} \text{mortality:} \quad & \boldsymbol{m} = \{m_{ij} : i = 1, \ldots, n, j = 1, \ldots, N\}, \\ \text{locations:} \quad & \boldsymbol{s} = \{\boldsymbol{s}_i : i = 1, \ldots, n\}, \\ \text{temperature:} \quad & \boldsymbol{t} = \{t_{ij} : i = 1, \ldots, n, j = 1, \ldots, N\}, \\ \text{covariates:} \quad & \boldsymbol{X} = \{x_{ijk}, i = 1, \ldots, n, j = 1, \ldots, N, k = 1, \ldots, M\}. \end{aligned} \tag{1}$$

Some of these variables will vary smoothly with location or time while others will be independent of space and time. A key aspect of the proposed approach is to model each type appropriately — this will be discussed in Section 3.

Now considering key relationships between variables. In particular, it is assumed that mortality depends on temperature, through a temperature-mortality curve – as in the example in Figure 1(c)*. These will vary between locations, and even over time, but will have similarly shaped curves across time and space with key aspects dependent on selected demographic variables – this level of further modelling will not be considered here.

2.2 Generalized additive modelling

This structure is well-suited to a description using a generalized additive model (GAM) approach (Hastie et al., 2009) which is an extension of the generalized linear model (GLM) approach – see also Fahrmeir and Lang (2001); Fahrmeir et al. (2004). In the GLM, the mean of the response variable, $E[Y]$, depends on a set of explanatory variables, $X_1, \ldots, X_p$, through a link function, η, as

$$\eta(E[Y]) = \alpha + \beta_1 X_1 + \cdots + \beta_p X_p. \tag{2}$$

Then the usual linear regression models have the identity link, $\eta(E[Y]) = E[Y]$, and the random errors are assumed to be independent and identically distributed normal random variables. In the GAM the linear relationships are simply replaced by non-linear functions, hence $f_1(X_1), \ldots, f_p(X_p)$,

$$\eta(E[Y]) = \alpha + f_1(X_1) + \cdots + f_p(X_p). \tag{3}$$

‡Data from Tank and Coauthors (2002) and *based on Todd and Valleron (2015).

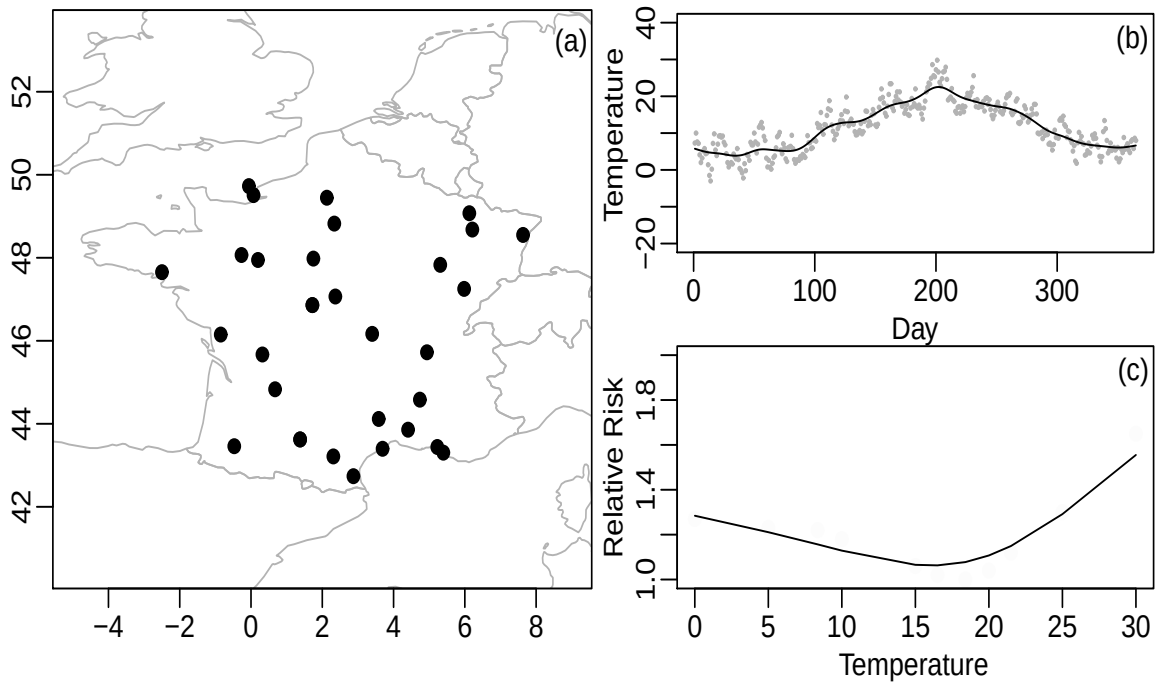

Figure 1: Data: (a) spatial locations, (b) typical daily temperature pattern through a year, and (c) typical relationship between mortality and temperature.

Note that some of the functions could be linear producing a mixture of non-linear and linear components. Unknown non-linear functions are usually estimated flexibly, perhaps using smoothing splines or based on local models such as autoregressive processes. As with GLMs, the form of the link function is determined by the nature of the random error, for example, an *identity* link for Gaussian errors, *logit* or *probit* for binomial probabilities and *log* for Poisson counts.

2.3 Spatial and temporal models

The specific version of the GAM to be used here, written in terms of mortality, temperature, space and time, will now be considered. The overall modelling approach is similar to that of Fahrmeir and Lang (2001); Fahrmeir et al. (2004). The relationship between the variables is illustrated in Figure 2, where arrows show plausible causal influences. The spatial-temporal

GAM contains a global mean, μ, with separable spatial and temporal components, f_1 and f_2,

$$y(\boldsymbol{s}, t) = \log m(\boldsymbol{s}, t) = \mu + f_1(\boldsymbol{s}) + f_2(t) + \epsilon(\boldsymbol{s}, t) \tag{4}$$

where the errors ϵ are independent and identically distribution normal random variables, $\epsilon \sim N(0, \sigma^2)$. It is also possible to include a term which depends on a set of other covariates, as represented in Figure 2.

As a novel extension, an additional component is added which explicitly described changes in mortality caused by changes in temperature

$$y(\boldsymbol{s}, t) = \log m(\boldsymbol{s}, t) = \mu + f_1(\boldsymbol{s}) + f_2(t) + f_3(\texttt{temp}(\boldsymbol{s}, t)) + \epsilon(\boldsymbol{s}, t). \tag{5}$$

It is this additional, *latent* component which contains the novel modelling in this paper. This is a function of both location and time, but instead a *hidden* mortality-temperature function is defined which can be estimated using localized data collapsed over time.

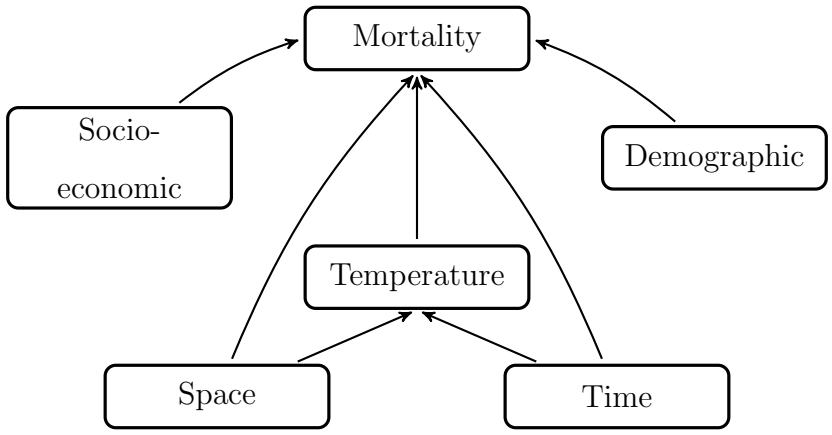

Figure 2: Diagram showing the causal relationship and hence the conditional independence structure.

Now consider log-mortality data $\boldsymbol{y} = \{y_{ij} : i = 1, \ldots, n, j = 1, \ldots, N\}$, where $y_{ij} = \log m_{ij}$ for $i = 1, \ldots, n, j = 1, \ldots, N$. From this description the likelihood and, in turn, the

following log-likelihood can be defined

$$l(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2} \sum_{i=1,j=1}^{n,N} \left(y_{ij} - \mu - f_1(\boldsymbol{s}_i) - f_2(t_j) - f_3(\texttt{temp}(\boldsymbol{s}_i, t_j))\right)^2 \tag{6}$$

where vector $\boldsymbol{\theta}$ includes all parameters involved in the unknown functions. It is important to note that this is an under-determined model and hence the usual *sum to zero* constraints will be included

$$\sum_{i=1}^{n} f_1(\boldsymbol{s}_i) = 0, \quad \sum_{j=1}^{N} f_2(t_j) = 0 \text{ and } \sum_{i=1,j=1}^{n,N} f_3(\texttt{temp}(\boldsymbol{s}_i, t_j) = 0.$$

In turn each of the unknown functions will now be considered – following the general approach of Fahrmeir and Lang (2001); Fahrmeir et al. (2004). The hierarchical relationship between data, component functions and other model parameters is shown in Figure 3.

Suppose that f_1 is a smooth function of space, but that it will only be considered at points corresponding to the recorded locations. That is the values of $\boldsymbol{f}_1 = \{f_{1,i} : i = 1, \ldots, n\}$ corresponding to $\boldsymbol{s} = \{s_i : i = 1, \ldots, n\}$. To impose smoothness, it will be assumed that these form a Gaussian random field such that $(f_{1,i} - f_{1,i'}) \sim N(0, 1/\beta_1^2)$, where i' corresponds to a neighbour of i and for $i = 1, \ldots, N$. In general, two locations will be considered neighbours if they are close – this will be discuss in more detail in Section 4.3. Although Gaussian models are used to describe a wide variety of spatial processes they would not be suitable for situations where large deviations are more likely where a heavy-tailed distributions, such as a Student-t, might be more appropriate. In other situations a non-symmetric distribution might provide a better description (see for example, Garcia-Papani et al., 2017). Such generalisations can be accommodated in the proposed framework.

In a similar way, suppose that f_2 is a smooth function of time, but that it will only be considered at points corresponding to the recorded times, that is the values of $\boldsymbol{f}_2 = \{f_{,j} : j = 1, \ldots, n_T\}$. To impose smoothness, it will be assumed that these form a Gaussian random walk process such that $(f_{2,j} - f_{2,j-1}) \sim N(0, 1/\beta_2^2)$ for $j = 2, \ldots, N$. Again, the proposed modelling approach can easily be generalized to take into account heavy-tailed or symmetric distribution should that be needed.

Finally, for the choice of f_3. At a given location and time, then there is a corresponding temperature and f_3 is a smooth function of temperature. Suppose that values $\boldsymbol{f}_3 = \{f_{3,l} : l = 1, \ldots, n_T\}$ corresponding to a set of temperatures $\boldsymbol{T} = \{T_l : l = 1, \ldots, n_T\}$ — these are chosen to cover the range of the measured temperatures, but are not the measured temperatures. To impose smoothness, it will be assumed that these form a Gaussian random walk process such that $(f_{3,l} - f_{3,l-1}) \sim N(0, 1/\beta_3^2)$, for $l = 2, \ldots, n_T$. Should it be necessary, then non-Gaussian and non-symmetric models can be used within the general framework.

Taking all parameters together, including the three parameters in the prior distributions, gives $\boldsymbol{\theta} = \{\mu, \boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \beta_1, \beta_2, \beta_3\} = \{\boldsymbol{\theta}_f, \boldsymbol{\theta}_\beta\} = \{\theta_1, \ldots, \theta_p\}$ which is of length $p = n + N + n_T + 4$. The model fitting process is then one of estimating $\boldsymbol{\theta}$ from all available data – this is the theme of the next section.

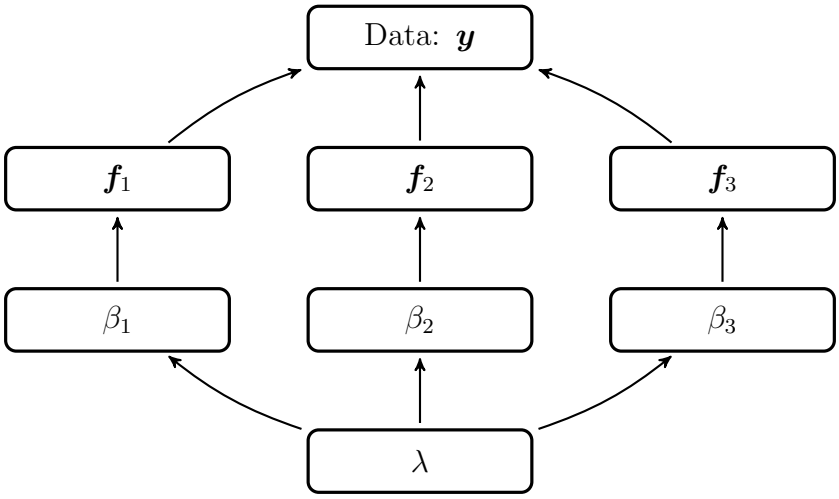

Figure 3: Diagram showing the hierarchical parameter structure.

3 A Bayesian spatial-temporal approach

3.1 Bayesian framework

In the following the data will be denoted $\boldsymbol{y} = \{y_i : i = 1, \ldots, n\}$ and the unknown model parameters as $\boldsymbol{\theta} = \{\mu, \boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \beta_1, \beta_2, \beta_3\} = \{\boldsymbol{\theta_f}, \boldsymbol{\theta_\beta}\} = \{\theta_1, \ldots, \theta_p\}$. The key ingredients in the Bayesian approach are the likelihood function, denoted as $f(\boldsymbol{y}|\boldsymbol{\theta}) \equiv l(\boldsymbol{\theta})$, and the prior distribution, denoted $p(\boldsymbol{\theta})$, which combine to give the posterior distribution

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \Big/ f(\boldsymbol{y}). \tag{7}$$

The hierarchical structure described in Figure 2 in terms of the measured variables can be converted into one in terms of the model parameters as in Figure 3. The conditional independence structure can then be exploited to re-write the likelihood and prior distribution in Equation (7) as

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = f(\boldsymbol{y}|\boldsymbol{\theta_f})p(\boldsymbol{\theta_f}|\boldsymbol{\theta_\beta})p(\boldsymbol{\theta_\beta}) \Big/ f(\boldsymbol{y}). \tag{8}$$

This recognises that the likelihood depends on $\boldsymbol{\theta_f}$ only, and that the prior distribution can be re-written $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta_f}|\boldsymbol{\theta_\beta})p(\boldsymbol{\theta_\beta})$. The hyper-prior distribution $p(\boldsymbol{\theta_\beta})$ may also depend on parameters, but these will not be treated as random variables and hence are not included in these expressions, but will be discussed in Section 3.3,

An estimate of $\boldsymbol{\theta}$ can be found as the value using the maximum a posteriori (MAP) estimate or the posterior expectation, also know as the posterior mean (PM), which are defined as

$$\widehat{\boldsymbol{\theta}}_{MAP} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\boldsymbol{y}) \quad \text{and} \quad \widehat{\boldsymbol{\theta}}_{PM} = \int_{\boldsymbol{\theta}} \boldsymbol{\theta}\, p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}. \tag{9}$$

For more robust estimation the posterior median can be used, which is the Bayes' estimator under a linear loss function. It is also important to examine other summaries of the posterior distribution, such as posterior credible intervals, or to perform hypothesis tests. The estimation of all the unknown parameters, and other posterior inference, will be through Markov chain Monte Carlo methods which will be briefly discussed in Section 3.3.

3.2 Prior and hyper-prior models

A key aspect of the Bayesian approach to statistical modelling is the explicit inclusion of prior information. In some cases, this will be subjective information elicited from experts, or it could be explicit information derived from related experiments leading to an empirical Bayes approach. Here prior belief about the smoothness of the unknown functions and parameters will be quantified – see also Fahrmeir and Lang (2001); Fahrmeir et al. (2004). Further discussion of prior distribution choice can be found in Berger (2010)

The definitions given in Section 2 already define prior distributions with

$$p(\boldsymbol{f}_i|\beta_i) = \frac{\beta_i^m}{(2\pi)^{m/2}} \exp\left\{-\frac{\beta_i^2}{2}\sum_{j=1}^{m}(f_{i,j} - f_{i,j'})^2\right\}, \quad i = 1, 2, 3. \tag{10}$$

Note that for $\boldsymbol{f}_1$ the subscript j' denotes the spatial neighbours of j, whereas for $\boldsymbol{f}_2$ and $\boldsymbol{f}_3$ it can, more simply, be replaced by $j - 1$. Note that the earlier mentioned sum-to-zero constraints make these proper distributions with finite integrals.

This modelling has introduced additional hyper-parameters, $\beta_i : i = 1, 2, 3$, which will be taken as unknown and estimated after the inclusion of the hyper-prior distributions

$$p(\beta_i) = \lambda \exp\{-\lambda/\beta_i\}, \qquad \beta_i > 0;\ \lambda > 0,\ i = 1, 2, 3. \tag{11}$$

It is plausible that individual λ parameters might be needed, but in this preliminary work a single value, chosen after initial experiments, has been used.

3.3 Model fitting using the MCMC algortihm

A standard random–walk Metropolis–Hastings algorithm is used to produce approximate samples from the posterior distribution by simulating a Markov chain. The use of such methods for parameter estimation, and more general density exploration is widespread – a review can be found in Robert and Casella (2011), then for theoretical details see Gamerman and Lopes (2006); Brooks et al. (2011), for general practical examples see the collection by Gilks et al. (1995) and for an early example applied to spatial modelling see Besag et al. (1991).

	Min.	1st Qu.	Median	3rd Qu.	Max
μ	-	-	0.5	-	-
$\boldsymbol{f}_1$	-0.1778	-0.1317	-0.0051	0.0639	0.2704
$\boldsymbol{f}_2$	-0.0045	-0.0007	0.0000	0.0007	0.0045
$\boldsymbol{f}_3$	-0.1163	-0.0913	-0.0260	0.0668	0.2037

Table 1: Summary of the true components of log-mortality.

As in Section 3.1, the parameter vector will simply be referred to as $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$. The Markov chain can start at any feasible point in the parameter space, then a discrete time Markov chain is created with values $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^K$. At each step a subset of the parameter vector is considered, for example a single value: μ, β_1, β_2 or β_3, or a complete function: $\boldsymbol{f}_1$, $\boldsymbol{f}_2$ or $\boldsymbol{f}_3$. Suppose that a new value for θ_i is being proposed, then $\theta_i^k = \theta_i^{k-1} + \epsilon_i$ and an obvious choice is $\epsilon_i \sim N(0, \tau_i^2)$. This proposal is accepted with probability $\min\left\{1, p(\boldsymbol{\theta}^k|\boldsymbol{y})/p(\boldsymbol{\theta}^{k-1}|\boldsymbol{y})\right\}$, otherwise the value is reset with $\theta_i^k = \theta_i^{k-1}$. After each parameter is considered the chain moves from state $\boldsymbol{\theta}^{k-1}$ at time $k-1$ to state $\boldsymbol{\theta}^k$ at time k. Reasonable proposal variances can be chosen adaptively during the burn-in period to achieve a reasonable acceptance rate. After a suitable burn-in period, to allow transient behaviour to dissipate, a sample can be collected with parameters estimated based on properties of the sample, for example the sample median can used as a point estimate and sample order statistics to estimate credible intervals.

4 Empirical results

4.1 General and data simulation

In this section a simple simulated data problem will be considered whose structure follows that previously defined. It is important to note, however, that there is only a limited attempt to make the modelling realistic, instead it is chosen to allow the general approach to be illustrated. Figures 4 shows the true functions used in the simulation.

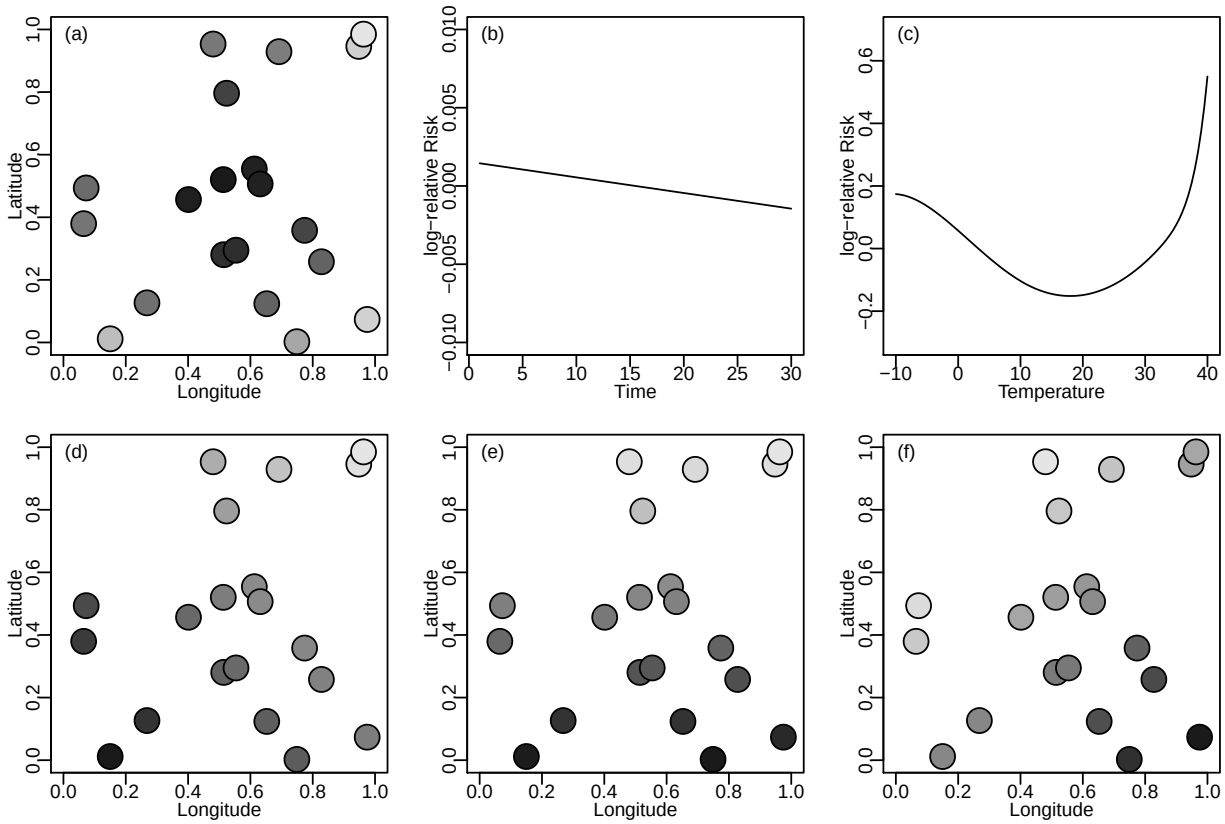

Figure 4: True log-mortality components: (a) spatial, $\boldsymbol{f}_1(\boldsymbol{s})$, (b) temporal, $\boldsymbol{f}_2(t)$, and (c) relationship between mortality and temperature $\boldsymbol{f}_3(\texttt{Temp})$, then temperature-related mortality, $\boldsymbol{f}_3(\boldsymbol{s}, t)$, at (d) $t = 6$, (e) $t = 15$ and $t = 24$.

Suppose that $n = 20$ locations are monitored daily for thirty days giving $N = 30$, including weekends and without missing data. These locations have been generated at random within a unit square, and the times are simply labelled $j = 1, \ldots, N = 30$. Then the log-mortality has a global mean of $\mu = 0.5$ which corresponds to about 1.6 deaths per day. It is assumed that the mortality rate is greatest near the centre of the region and this is modelled as a quadratic surface giving a relative risk between about 0.8 to 1.3 — see Figure 4(a). Over the study period it is assumed that there is a reduction in the relative risk of about 4% — this has been exaggerated otherwise the effect would be lost in the noise — see 4(b). The localised temperature follows a simple sinusoidal pattern with variability dependent on the west-east coordinate and around a mean which depends on the north-south coordinate. The mortality

then depends on temperature in a quadratic fashion which is common across all locations and times — see 4(c) – which gives a relative risk between 0.9 and 1.6. Figure 4(d)-(f) show the temperate-related mortality components corresponding to three time points. Finally, independent Gaussian errors with standard deviation $\sigma = 0.01$ are added to the summation of these various deterministic components to give the log-mortality data values. This has been repeated $M = 100$ times to be able to assess reproducibility of the estimation process. Table 1 shows some summary statistics of the generated model components.

4.2 Algorithm implementation

The unknowns in the spatio-temporal GAM, defined in (5), with likelihood in (6) and prior distributions in (10), are estimated using the MCMC algorithm with a burn-in of 5000 iterations and a main run of 500 iterations after applying 1-in-100 thinning to remove the effects of autocorrelation. In the spatial prior, locations are considered neighbours if they are closer than 0.2 units apart, whereas in the temporal prior, adjacent times are considered neighbours. For the mortality-temperature relationship, the log relative risk at 22 equally-spaced intervals is considered with adjacent values considered neighbours. Proposal variances were automatically selected to achieve reasonable acceptance rates, and the hyper-prior parameter was fixed after a few pilot runs at $\lambda = 10^4$. It is worth noting that the algorithm was made more efficient, by balancing autocorrelation, by considering each parameter in $\boldsymbol{f}_2$ 10 times compared to only once for the other parameters. During the burn-in period, the proposal variances were updated every 100 iterations with the final values shown in Table 2.

The algorithm performed well without excessively long initial transient behaviour and autocorrelation. To illustrate typical monitoring output, consider Figure 5. The Markov path for the global mean is shown in (a) and a single element of one of the functions in

	τ_μ	τ_{f_1}	τ_{f_2}	τ_{f_3}	τ_{β_1}	τ_{β_2}	τ_{β_3}
Value	0.002	0.001	0.0001	0.001	9.70	6634	22.3

Table 2: Final values of proposal standard deviations.

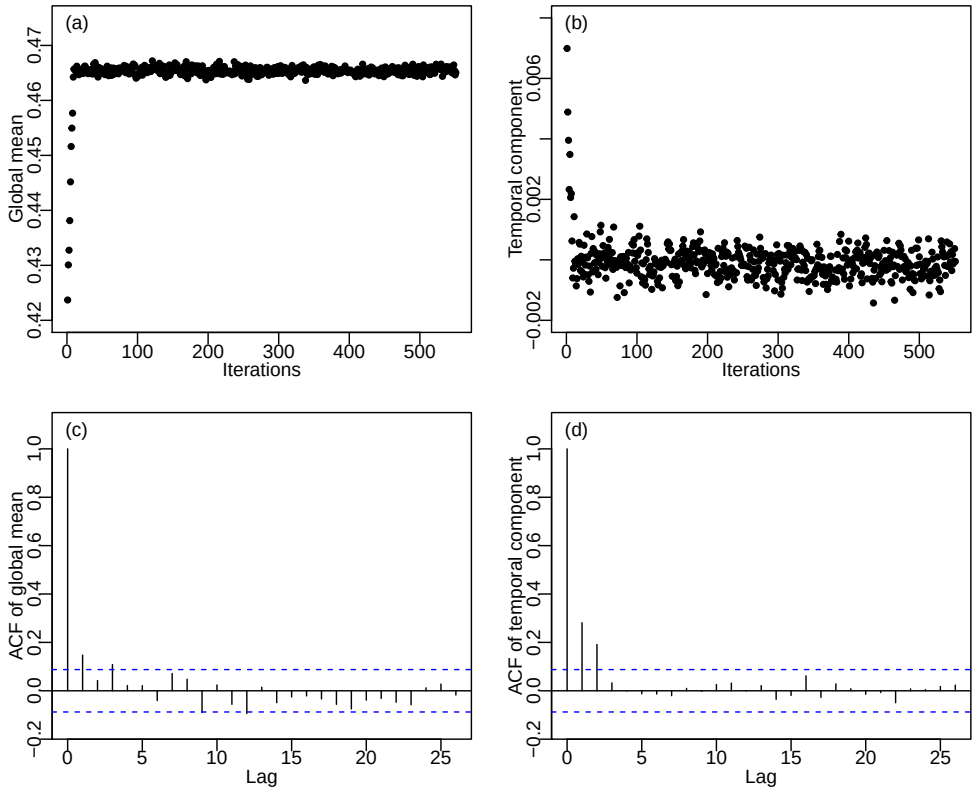

Figure 5: Typical convergence monitoring output: Markov paths of (a) $\hat{\mu}$ and (b) $\boldsymbol{f}_2(15)$, and corresponding autocorrelation functions in (c) and (d).

(b) – these take into account the 1-in-100 thinning. There is clear transient behaviour in the first few iterations. After this point there is no obvious systematic change suggesting equilibrium behaviour and hence equilibrium was declared at 50 iterations. Finally, (c) and (d) show corresponding autocorrelation functions of the thinned process after throwing away the first 50 iterations and they clearly indicate near independent observations. Sample size calculations were performed for each parameter (see for example, Aykroyd and Green, 1991) taking into account any remaining correlation with values from less than 100 to about 300 being obtained. These indicated that the main run of size 500 is more than adequate.

4.3 Fitted models

Selected results are shown in Figure 6 for the fitted global mean, the spatial and temporal components, along with the mortality-temperature relationship. A histogram of the global mean is shown in Figure 6(a) with the posterior median shown as a black vertical line and the true value as the red vertical line. The shaded area corresponds to 95% of the total area and hence the corresponding lower and upper values of μ form a 95% credible interval. Here the median is close to the true value and credible interval includes the true value.

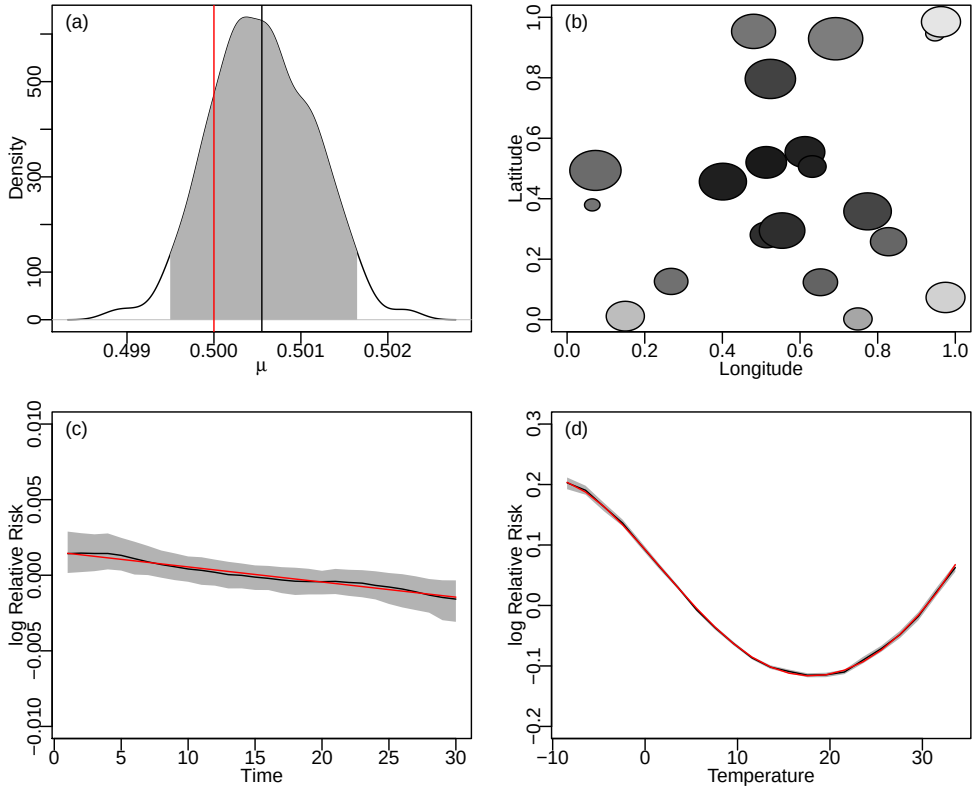

Figure 6: Posterior estimates with 95% credible intervals and true values (in red): (a) global mean, $\hat{\mu}$, (b) spatial component, $\hat{\boldsymbol{f}}_1(\boldsymbol{s})$, (c) temporal component, $\hat{\boldsymbol{f}}_2(t)$, and (d) relationship between mortality and temperature $\hat{\boldsymbol{f}}_3(\boldsymbol{s}, t)$.

In Figure 6(b) the shade of the circle indicate the posterior median values of the spatial component, $\boldsymbol{f}_1$, from grey for low values, to black for high values. The diameter of the circle indicates the posterior inter-quartile range, with small radii for low variability and large for

higher variability. This display is reminiscent of those in Besag et al. (1991). In (c) and (d) the posterior median estimates for the temporal function, $\boldsymbol{f}_2$, and temperature related mortality function, $\boldsymbol{f}_3$, are shown as a black curves within a grey area indicating the 95% posterior credible interval, with a red curve indicating the true function.

Figure 6(a) shows that the global mean can be well estimated. From (b) it is clear that the true function is well recovered with higher values in the centre, also that the estimates have small variability near the centre compared to the edges. In (c) the dependence on time is also well recovered, but with moderate uncertainty. The latent mortality-temperature relationship in (d) is also estimated well, but this time surprisingly precisely. The credible intervals will be assessed for correct coverage later.

Table 3 shows some summary results with accuracy measures averaged over the 100 replicates. In particular the mean error show that all are approximately unbiased, and the mean absolute error and root-mean-square error (RMSE) indicate very good estimation. Finally, the maximum absolute error has also been calculated and shows that even the largest errors are relatively small.

	$\hat{\mu}$	$\boldsymbol{f}_1$	$\boldsymbol{f}_2$	$\boldsymbol{f}_3$
Mean error	-0.0004	0.0000	0.0000	-0.0000
Mean absolute error	0.0006	0.0015	0.0009	0.0021
Max. absolute error	0.0017	0.0041	0.0022	0.0075
RMSE	0.0007	0.0018	0.0010	0.0028

Table 3: Summary of estimation accuracy averaged over 100 replicates.

Now moving on to the credible intervals, Table 4 shows the average coverage rates for the nominal 95% credible intervals. The figures for the three functions are very close to nominal value and for the global mean, recalling that the figure is calculated from only 100 cases, is also acceptable.

Before completing the analysis of the results, consider Table 5 which shows a summary of the estimated prior smoothing parameters in the function estimation. These indicate

	μ	$\boldsymbol{f}_1$	$\boldsymbol{f}_2$	$\boldsymbol{f}_3$
Coverage	0.85	0.93	0.99	0.94

Table 4: Empirical coverage rates of nominal 95% credible intervals based on 100 replicates.

very narrow bands for the estimated values which also indicates that estimation of these parameters is reliable making the overall procedures automatic and robust.

	Min.	1st Qu.	Median	3rd Qu.	Max
$\hat{\beta}_1$	48.50	49.01	49.22	49.44	49.79
$\hat{\beta}_2$	3203	3448	3546	3613	3717
$\hat{\beta}_3$	85.69	87.18	87.56	88.02	89.94

Table 5: Summary of estimated hyper-parameter values based on 100 replicates.

5 Discussion and further work

This work has presented a preliminary analysis to illustrate a novel modelling approach. It has successfully estimated separate spatial and temporal components, and importantly a latent mortality-temperature relationship. In addition, all prior parameters have also been estimated with the inclusion of hyper-prior distributions making estimation automatic. To fully demonstrate the potential for the approach it needs to me applied to a wider region containing a realistic number of spatial locations and time points. Further it is important to perform further sensitivity analysis to determine the reliability of results, and in particular to check for robustness to details in the modelling and particularly the choices of prior distributions. Once the approach has been verified in a realist situation, then possible extensions can be considered. For example, allowing local temperature-related mortality functions but constraining them to vary smoothly through space. The very promising results indicate that the methods and the general framework can become a useful tool for those modelling complex spatial-temporal problems, such as temperature-related mortality.

Acknowledgements:

I am grateful to Christos Skiadas for the initial suggestion, during ASMDA2015, of extending demographic analysis to include spatial modelling, and to Jean-Marie Robine for providing helpful suggestions which led to this work. I am also grateful to valuable comments from the anonymous referees and the Editor which have produced substantial improvements.
This paper is dedicated to the memory of Julian Besag. Although more than 25 years ago, what I learnt while in the same university department has surely influenced this work.

References

Aykroyd, R. G. and Green, P. J. (1991). Global and local priors, and the location of lesions using gamma-camera imagery, *Phil. Trans. R. Soc. Lond. A* **332**: 323–342.

Berger, J. O. (2010). *Statistical decision theory and Bayesian Analysis*, Springer Series in Statistics, Spinger-Verlag, New York.

Besag, J., York, J. C. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics, *Ann. Inst. Statist. Math.* **43**: 1–59.

Brooks, S., Gelman, A., Jones, G. and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*, Chapman & Hall/CRC.

Fahrmeir, L., Kneib, T. and Lang, S. (2004). Penalized structured additive regression for space-time data: A Bayesian perspective, *Statistica Sinica* **14**: 731–761.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors, *Journal of the Royal Statistical Society (Applied Statistics)* **50**: 201–220.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd edn, Chapman & Hall/CRC Texts in Statistical Science.

Garcia-Papani, F., Uribe-Opazo, M. A., Leiva, V. and Aykroyd, R. G. (2017). Birnbaum-Saunders spatial modelling and diagnostics applied to agricultural engineering data, *Stochastic Environmental Research and Risk Assessment* **31**(1): 105–124.

Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., Tobias, A., Tong, S.; Rocklöv, J., Forsberg, B., Leone, M., Sario, M. D., Bell, M. L., Guo, Y.-L. L., fu Wu, C., Kan, H., Yi, S.-M., de Sousa Zanotti Stagliorio Coelho, M., Saldiva, P. H. N., Honda, Y., Kim, H. and Armstrong, B. (2015). Mortality risk attributable to high and low ambient temperature: a multicountry observational study, *The Lancet* **386**: 369 – 375.

Gilks, W., Richardson, S. and Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, second edn, Springer.

McMichael, A. and Coauthors (2008). International study of temperature, heat and urban mortality: the ISOTHURM project, *International Journal of Epidemiology* **37**: 1121–1131.

National Academy of Sciences (2010). *Advancing the Science of Climate Change*, The National Academies Press, chapter Public Health, pp. 309–322.

Robert, C. and Casella, G. (2011). A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data, *Statistical Science* **26**: 102–115.

Tank, A. M. G. K. and Coauthors (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, *Int. J. of Climatol.* **22**: 1441–1453.

Todd, N. and Valleron, A. (2015). Space-time covariation of mortality with temperature: A systematic study of deaths in France, 1968-2009, *Environmental Health Perspectives* **123**: 659–664.