International Journal of Epidemiology, 2024, **53(3)**, dyae061
https://doi.org/10.1093/ije/dyae061
Original article

Original article

Spatial Bayesian distributed lag non-linear models (SB-DLNM) for small-area exposure-lag-response epidemiological modelling

Marcos Quijal-Zamorano [1,2,*] Miguel A Martinez-Beneito,[3] Joan Ballester[1] and Marc Marí-Dell'Olmo[4,5,6]

[1]ISGlobal, Barcelona, Spain, [2]Universitat Pompeu Fabra (UPF), Barcelona, Spain, [3]Departament d'Estadística i Investigaciò Operativa, Universitat de València, Burjassot, Valencia, Spain, [4]Agència de Salut Pública de Barcelona (ASPB), Barcelona, Spain, [5]Institut de Recerca Sant Pau (IR SANT PAU), Barcelona, Spain and [6]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

*Corresponding author. Barcelona Institute for Global Health (ISGlobal), Barcelona Biomedical Research Park (PRBB), Doctor Aiguader, 88, 08003 Barcelona, Spain. E-mail: marcos.quijal@isglobal.org

Abstract

Background: Distributed lag non-linear models (DLNMs) are the reference framework for modelling lagged non-linear associations. They are usually used in large-scale multi-location studies. Attempts to study these associations in small areas either did not include the lagged non-linear effects, did not allow for geographically-varying risks or downscaled risks from larger spatial units through socioeconomic and physical meta-predictors when the estimation of the risks was not feasible due to low statistical power.

Methods: Here we proposed spatial Bayesian DLNMs (SB-DLNMs) as a new framework for the estimation of reliable small-area lagged non-linear associations, and demonstrated the methodology for the case study of the temperature-mortality relationship in the 73 neighbourhoods of the city of Barcelona. We generalized location-independent DLNMs to the Bayesian framework (B-DLNMs), and extended them to SB-DLNMs by incorporating spatial models in a single-stage approach that accounts for the spatial dependence between risks.

Results: The results of the case study highlighted the benefits of incorporating the spatial component for small-area analysis. Estimates obtained from independent B-DLNMs were unstable and unreliable, particularly in neighbourhoods with very low numbers of deaths. SB-DLNMs addressed these instabilities by incorporating spatial dependencies, resulting in more plausible and coherent estimates and revealing hidden spatial patterns. In addition, the Bayesian framework enriches the range of estimates and tests that can be used in both large- and small-area studies.

Conclusions: SB-DLNMs account for spatial structures in the risk associations across small areas. By modelling spatial differences, SB-DLNMs facilitate the direct estimation of non-linear exposure-response lagged associations at the small-area level, even in areas with as few as 19 deaths. The manuscript includes an illustrative code to reproduce the results, and to facilitate the implementation of other case studies by other researchers.

Keywords: Small-area analysis, spatial statistics, non-linear dynamics, heat-related mortality, climate change, DLNM, Bayesian models.

Key Messages

- We adapted distributed lag non-linear models (DLNMs), the reference framework for modelling lagged non-linear associations, to Bayesian statistics for both large- and small-area studies.
- The Bayesian framework offers a wider range of estimates and tests, providing intuitive interpretations of uncertainty for derived indicators, without relying on arbitrary *P*-values.
- We extended Bayesian DLNMs (B-DLNMs) to spatial Bayesian DLNMs (SB-DLNMs) by incorporating spatial models, estimating short-term health effects of environmental exposures in small areas.
- We demonstrated the practicality of SB-DLNMs through a case study estimating temperature-mortality relationship in the 73 neighbourhoods of the city of Barcelona, and we ensured its reproducibility by detailing the mathematical formulation and the R codes used.
- SB-DLNMs provide geographical-varying risks and directly estimate non-linear lagged associations at small-area levels by modelling spatial differences and structures across small areas, even in cases with as few as 19 deaths.

Introduction

The quantification of the short-term health effects of environmental exposures, which are associated with millions of premature deaths every year,[1–5] is a major scientific question with far-reaching implications in decision and policy making. Distributed lag non-linear models (DLNMs) have rapidly

Received: 21 July 2023. **Editorial Decision:** 12 March 2024. **Accepted:** 10 April 2024

become the referent framework for the modelling of all these short-term associations.[6] DLNMs allow modelling non-linear and lagged environmental effects on health in a user-friendly implementation within an R package.[7]

For large-scale multi-location DLNM studies, it is common to apply multivariate meta-analysis to synthesize the estimates from different locations.[8–10] In this approach, independent models are individually fitted at each location, estimating the location-specific exposure-response relationship (first stage), and then they are combined through multivariate meta-analysis to derive the average exposure-response relationship across locations (second stage). The method also allows to compute the best linear unbiased prediction in each location, which is a trade-off prediction between average and location-specific exposure-response associations. Although these predictions provide a more refined understanding of the exposure-response relationship in each location, they take into account neither the spatial structure of the data nor any eventual spatial dependency between risks at different locations.

The spatial units of these multi-location DLNM studies are usually large- or medium-sized cities or regions, with research on small-area units still scarce. Nonetheless, the increasing availability of high-resolution environmental data, which can be used to assign exposures to individual postcodes or addresses,[11] has the potential to open new avenues in small-area environmental epidemiology. Moreover, even in ecological studies, where health data records are aggregated within spatial units, considering small areas reduces the ecological fallacy by providing closer estimations to individual-level risks.[11] In addition, the study of the socioeconomic and physical factors that modify the effect between environmental exposures and health becomes more reliable when estimations get closer to individual risks, given that modifying factors are generally diluted when populations are aggregated to larger-scale domains.

The study of the short-term health effects of environmental exposures in small areas is particularly susceptible to yielding unreliable estimates,[12] with only a few studies proposing methodologies to address this issue.[13,14] Some of them rely on self-matched approaches.[15–17] The case time series design uses small-area data and considers the full flexibility of DLNMs[18]; however, the method did not allow geographically varying risks, estimating a common exposure-response relationship across all small areas.[16] This method was extended to generate associations at any spatial level.[17] The approach was conceived as a two-stage design followed by a downscaling procedure, where location-specific associations were first estimated at an intermediate spatial level, ie, not the smallest-area level. Then, the first-stage coefficients were pooled through a meta-regression, which included meta-predictors at the intermediate-area level. In the final step, they used the meta-analytical model and the meta-predictors at the smallest-area level to downscale, ie, meta-predict, the risks at the smallest-area level. Consequently, the spatial variation of the risks depends on the meta-predictors, and this approach cannot directly estimate the associations at the smallest-area level if the number of cases in each unit is very small.

On the other hand, and in contrast to this approach, other studies considered the spatial structure of the data with spatial Bayesian methods, directly estimating the short-term health effects of warm temperatures on mortality in small areas.[19] However, they did not account for lagged non-linear associations, but instead only considered one single linear term. We here extend this last approach by formulating, for the first time, spatial Bayesian distributed lag non-linear models (SB-DLNMs) for the estimation of reliable small-area lagged non-linear associations, and we illustrate the methodology for the particular case of the temperature-mortality relationship in the 73 neighbourhoods of the city of Barcelona. In our approach, we adapt the DLNM modelling framework to Bayesian statistics, considering both the spatial structure and the potential presence of spatial dependence in the associations. To extend single-location independent Bayesian distributed lag non-linear models (B-DLNMs) to SB-DLNMs, we employ Bayesian hierarchical models by including spatial components in the model.

The paper is organized as follows. In the Methods section, we formalize B-DLNMs and SB-DLNMs, using different case-crossover and time-series designs. In the Results section, we apply the proposed models to a small-area temperature-mortality case study in the city of Barcelona. Finally, the Discussion section discusses the advantages and limitations of the method, and explores future steps to upscale the use of the approach to large-scale epidemiological analyses, beyond small-area studies. The illustrative code to reproduce the results is available in the Supplementary data (Supplementary Text S1, available as Supplementary data at *IJE* online) and in a GitHub repository [https://github.com/marcosqz/sbdlnm_smallarea_casestudy].

Methods

Here we present four B-DLNMs to assess short-term exposure-response associations. The first two models generalize location-independent DLNMs to a Bayesian framework, using two common designs to control for seasonal and long-term trends: case-crossover (Model 1),[20] and splines of time (Model 2).[21] In addition, we extend these two models to SB-DLNMs by incorporating Leroux models in a single-stage approach to account for the spatial dependence between the location-independent parameters (Models 3 and 4).[22] Although our case study focuses on the short-term health effects of warm temperatures on mortality in the city of Barcelona, these four models can be easily adapted to other locations, exposure and response variables and model configurations.

As a common notation standard for all four models, DLNMs use a cross-basis matrix that defines the (potentially) non-linear relation between exposure and response across a range of lags.[6] Two separate basis functions are applied to the exposure (with dimension ν_x) and lag (ν_l) spaces, and combined to generate the cross-basis matrix $\boldsymbol{W}$. This combination involves tensor products and the re-arrangements of the two splines basis functions.[6] Matrix $\boldsymbol{W}$ consists of $n_c = \nu_x \times \nu_l$ columns, representing linear-term coefficients and defining the joint non-linear effect of the exposure and lag dimensions.

Independent models: B-DLNM with case-crossover (Model 1) and time-series (Model 2) designs

In Model 1, we implement the B-DLNM with a case-crossover design[20] which has been widely used to study the short-term health effects of environmental exposures.[19,23] In this model, we stratify the data by area, year, month and day of week, where daily temperatures and counts of deaths in a given area are compared with those in the same area, day of the week, month and year. Consequently, each stratum is composed of 4 or 5 case-days, depending on the specific day of the week,

month and year. We exclude stratums without deaths as they do not contribute to the likelihood in these models.[24] This design is supposed to control automatically for seasonal and long-term trends, day of week effect and other potential confounders that vary slowly over time.[19]

In Model 2, we implement splines of time to account for long-term and seasonal trends, following the standard approach in time-series studies of the short-term association between temperature and mortality.[21] To capture the seasonal confounding patterns in mortality, we use flexible spline functions. Specifically, we transform the date and day-of-year into two matrices, $\boldsymbol{T}$ (long-term) and $\boldsymbol{S}$ (seasonality), where the columns correspond to the functions derived from the respective spline bases. The number of coefficients estimated in each area to control for these trends depends on the dimension of the spline bases, here noted as n_t and n_s for long-term and seasonal trends, respectively. If the study period is continuous, the date can be directly transformed using a single spline of time with enough flexibility per year.

For the independent models, we analyse each area individually by extending frequentist location-independent DLNMs to a Bayesian framework. We use the following formulations:

$$y_{d,a} \sim \text{Poisson}(\lambda_{d,a})$$

$$\text{Model 1 (independent-case-crossover): } \log(\lambda_{d,a}) = \alpha_{s(d,a)} + \sum_{c=1}^{n_c} \beta_{a,c} \cdot W_{d,a,c}$$

$$\text{Model 2 (independent-time-series): } \log(\lambda_{d,a}) = \alpha_{a,o(d)} + \sum_{c=1}^{n_c} \beta_{a,c} \cdot W_{d,a,c} + \sum_{t=1}^{n_t} \gamma_{a,t} \cdot T_{d,a,t} + \sum_{s=1}^{n_s} \delta_{a,s} \cdot S_{d,a,s}$$

For each day (d, integer from 1 to the number of days) and area (a, integer from 1 to the number of areas), the counts of death ($y_{d,a}$) follow a Poisson distribution with mean parameter $\lambda_{d,a}$. We fit generalized linear models relating $\lambda_{d,a}$ via a logarithm link function with the cross-basis matrix ($\boldsymbol{W}$, with n_c columns representing the different exposure-lag predictors expressed in the respective bases). Moreover, the models include:

- for the case-crossover design (Model 1) an intercept $\alpha_{s(d,a)}$ for each stratum (**s**, integer from 1 to the number of stratums with cases) that is shared among days in the same area, year, month and day of week. This intercept is intended to match observations within stratums;
- for the time-series design (Model 2), intercepts $\alpha_{a,o(d)}$ controlling for the day of week (***o***, integer from 1 to 7) in each area a, the long-trend matrix ($\boldsymbol{T}, n_t$ columns) and the seasonality matrix ($\boldsymbol{S}$, n_s columns).

The coefficients $\beta_{a,c}$, $\gamma_{a,t}$, and $\delta_{a,s}$ are the coefficients defining the temperature-lag-mortality relation, and the long-term and seasonal trends, respectively. We use independent vague prior distributions for all these coefficients and intercepts. We note that performing these independent models is equivalent to performing separate DLNM analyses for each area.

Spatial models: SB-DLNM with case-crossover (Model 3) and time-series (Model 4) designs

When dealing with small sample sizes, independent B-DLNMs may provide unreliable estimates. Moreover from an epidemiological perspective, it is reasonable to assume that the risk estimates will be more similar in closer areas than in distant ones. To address this, in the SB-DLNMs (Models 3 and 4) we add a spatial component to account for the potential spatial dependence in exposure-response associations. We include the spatial model within a Bayesian hierarchical framework, allowing direct estimation of the risk in a single step. Specifically, we include the spatial dependence in the coefficients of the cross-basis $\boldsymbol{W}$, ie, $\beta_{a,c}$, by using the Leroux conditional autoregressive (CAR) distribution.[22] We note however, that alternative spatial models can be easily implemented instead, eg, the Besag-York-Mollié model.[25]

In Models 3 (spatial case-crossover) and 4 (spatial time-series), we change the prior distributions of the coefficients $\beta_{a,c}$ of the cross-basis matrix $\boldsymbol{W}$ specified in Models 1 and 2. Instead of using independent vague priors, we use a set of random effects with spatial structure, with the following formulation:

$$\beta_{a,\,c} = \mu_c + \beta^*_{a,\,c}$$

$$\beta^*_{a,c} \sim N\left(\frac{\rho}{1-\rho+\rho \cdot n_a} \sum_{a' \sim a} \beta^*_{a',c}, \frac{\sigma_c^2}{1-\rho+\rho \cdot n_a}\right)$$

As a result, SB-DLNMs are generalizations of the independent B-DLNMs. In SB-DLNMs, the coefficient corresponding to a specific column of the cross-basis matrix $\boldsymbol{W}$ in a given area is estimated considering the coefficients of the same term in neighbouring areas. Given an area, we here define its neighbours as the set of adjacent areas, ie, those that share a border, although this spatial connectivity rule can be easily generalized to any other type of criterion. Specifically, for each area (a) and cross-basis element (c), $\beta_{a,c}$ is modelled as the sum of the mean value across the spatial units (μ_c), which describes the mean effect for the entire region, and a random effect ($\beta^*_{a,c}$) that introduces spatial variability between areas. The spatial variability $\beta^*_{a,c}$ follows a normal distribution, with the mean and variance depending on the parameter ρ. The parameter ρ defines the strength of the spatial component in the Leroux models, ranging from 0 (lowest strength, representing a heterogeneous process) to 1 (highest strength, representing a fully spatial intrinsic CAR distribution).[26] The mean and variance of the normal distribution depend on the information of the adjacent areas, ie, the number of adjacent areas (n_a) and the spatial variability parameters ($\beta^*_{a',c}$) for the same coefficient c of the adjacent area a'. Finally, σ_c^2 allows for different variances for the coefficients of the cross-basis c. We also use vague prior distributions for all the coefficients in the Leroux model.

Results

Small-area case study

In this case study, we modelled the short-term associations between temperature and mortality during the summer months in 73 small areas (neighbourhoods) of Barcelona (Supplementary Figure S1, Supplementary Table S1, available as Supplementary data at *IJE* online). As a health outcome, we used the daily neighbourhood counts of natural deaths from June to September during the period 2007–2016, obtained from the mortality registry of the city. We assigned deaths to the corresponding neighbourhood based on the residence address of the deceased individuals. In total, the dataset included 39 569 deaths, with 19 (1454) deaths in the neighbourhood

with the lowest (highest) number of deaths during the whole 10-year period (Supplementary Table S1, available as Supplementary data at *IJE* online). As the exposure variable, we used the daily neighbourhood time series of maximum temperatures from the UrbClim model,[27] calculated as the population-weighted average of the 100x100 meter gridded temperatures within the boundaries of each neighbourhood.

We applied both independent B-DLNMs and SB-DLNMs. In each neighbourhood, we used a natural cubic spline with 2 internal knots at the 50th and 90th percentiles of the neighbourhood-specific daily temperature distribution for modelling the exposure-response function ($v_x = 3$). The lag-response ranged from 0 to 8 days, and we used a natural cubic spline with 2 internal knots equally spaced in the logarithm scale and an intercept ($v_l = 4$, $n_c = 3 \times 4 = 12$). In the time-series models (models 2 and 4), we controlled for the seasonality with and interaction between the year and a natural cubic spline with 3 equally-spaced internal knots ($n_s = 4 \times 10 = 40$), the long-term trend with a linear term ($n_t = 1$), and the day of the week with 511 intercepts ($73 \times 7 = 511$). Instead, the case-crossover models (models 1 and 3), included 14857 stratums. These DLNM-modelling choices were based on previous studies.[3,28]

We executed the four models by using the WinBUGS software, which implements Markov chain Monte Carlo (MCMC) methods.[29] For each model, we run 12 chains in parallel,[30] initialized with different values. We set a burn-in of 10% of the total iterations, and we run as many iterations as necessary to achieve convergence (50 000, 100 000 or 150 000, depending on the model). Additionally, we subsampled the chains by thinning, retaining 1000 samples. Each of the 1000 samples represents a cycle of the model that generates a full set of parameter values from the posterior distribution.[31] To assess model convergence, we examined two statistics for all parameters: the Rhat (potential scale reduction factor), which we imposed to be lower than 1.1, and the effective sample size, which needed to be higher than 100.[26,32] The full code for the four B-DLNMs and SB-DLNMs in R and WinBUGS, used in this case study, is provided in the supplementary data (Supplementary Text S1, available as Supplementary data at *IJE* online) and GitHub repository (https://github.com/marcosqz/sbdlnm_smallarea_casestudy).

The output of the B-DLNMs allowed us to obtain the location-independent temperature-lag-mortality associations. First, following the algebra of DLNMs,[6] we calculated the relative risk (RR) for every combination of temperature and lag in each region, based on the 1000 independent samples. These associations are visualized as 1000 exposure-lag-response surfaces in Figure 1 (first column), which shows the associations for one of the neighbourhoods. From this ensemble of surfaces, we can obtain the associations for a specific lag (Figure 1, second column), the associations for a specific temperature (Figure 1, third row), and the overall cumulative temperature-mortality associations by summing the lag-specific contributions (Figure 1, fourth column).

Figure 2 illustrates the process of obtaining summary RR estimates and credible intervals (CrI) for the whole ensemble, in continuation of the associations shown in Figure 1. To facilitate comparison between neighbourhoods and models, and without loss of generality, we set the centring temperature as the baseline for all curves, defined as the 30th percentile of the neighbourhood daily temperature distribution. However, it is possible to centre the associations using other summary statistics, such as those derived from the posterior distribution of the minimum mortality temperature (see Supplementary Figure S2, available as Supplementary data at *IJE* online). The point estimate of the RRs is here defined as the median of the ensemble of RRs at each temperature, and the 95% CrI as the 2.5th and 97.5th percentiles.

In Figure 3, we compare the overall cumulative associations in the four models between three adjacent neighbourhoods with very different population sizes and, consequently, mortality numbers. The exposure-response curves show the closest results between the independent B-DLNMs and SB-DLNMs in Sant Andreu (Figure 3a, d). This neighbourhood has a high number of deaths during the study period (1191), so its estimate is less affected by the small sample size problem. Point estimates and 95% CrIs fall into reasonable ranges in the independent B-DLNMs, although the inclusion of the spatial component in the SB-DLNMs further reduces the uncertainty of the estimates. Instead, the two adjacent neighbourhoods of Sant Andreu, ie La Prosperitat (Figure 3b, e; 587 deaths) and La Trinitat Vella (Figure 3c, f; 132 deaths), show unusual features in the associations with independent B-DLNMs. The incorporation of information from adjacent neighbourhoods in SB-DLNMs, eg Sant Andreu, substantially reduces uncertainties and corrects the point estimate curves, aligning the values between the adjacent neighbourhoods. This effect is particularly important in neighbourhoods with the smaller sample size, such as La Trinitat Vella. The associations observed in the case-crossover and time-series analyses show comparable results, with slightly smaller uncertainties observed in the time-series design. Supplementary Figures S3 and S4 (available as Supplementary data at *IJE* online) show the comparison between independent B-DLNMs and SB-DLNMs of overall cumulative association and the lag-response association at the 99th temperature percentile for all neighbourhoods, respectively.

Finally, Figure 4 shows the spatial distribution of the RR at the 99th temperature percentile in the four models (top panels) together with the corresponding probability that the RR is greater than one (ie fraction of samples with RR >1, bottom panels). On the one hand, we observe no definite spatial pattern of the RRs in the independent B-DLNMs (Figure 4a, b), with protective RR values (ie RR <1) in many neighbourhoods (Figure 4e, f). The small numbers of cases contribute to noise, resulting in adjacent neighbourhoods showing contrasting extremes. On the other hand, SB-DLNMs estimate more consistent RR values, both spatially and in terms of magnitude, revealing clusters of neighbourhoods with similar RR values (Figure 4c, d). Importantly, SB-DLNMs address the instability and incoherence observed in the independent B-DLNMs by incorporating spatial dependencies, resulting in more plausible and coherent results (cf Figure 4a–d). When comparing the results between the case-crossover and time-series design, the addition of the spatial component gives equivalent conclusions. However, the time-series approach suggests higher statistical power to detect risk excesses (cf Figure 4g, h). Last but not least, the strength of the spatial component is high in the Leroux models, with estimated ρ values skewed toward 1: 0.72 (95% CrI, 0.19–0.99) for the case-crossover SB-DLNM, and 0.82 (95% CrI, 0.32–0.99) for the time-series SB-DLNM.

Discussion

In this study, we proposed SB-DLNMs for estimating short-term health effects of environmental exposures in small areas.

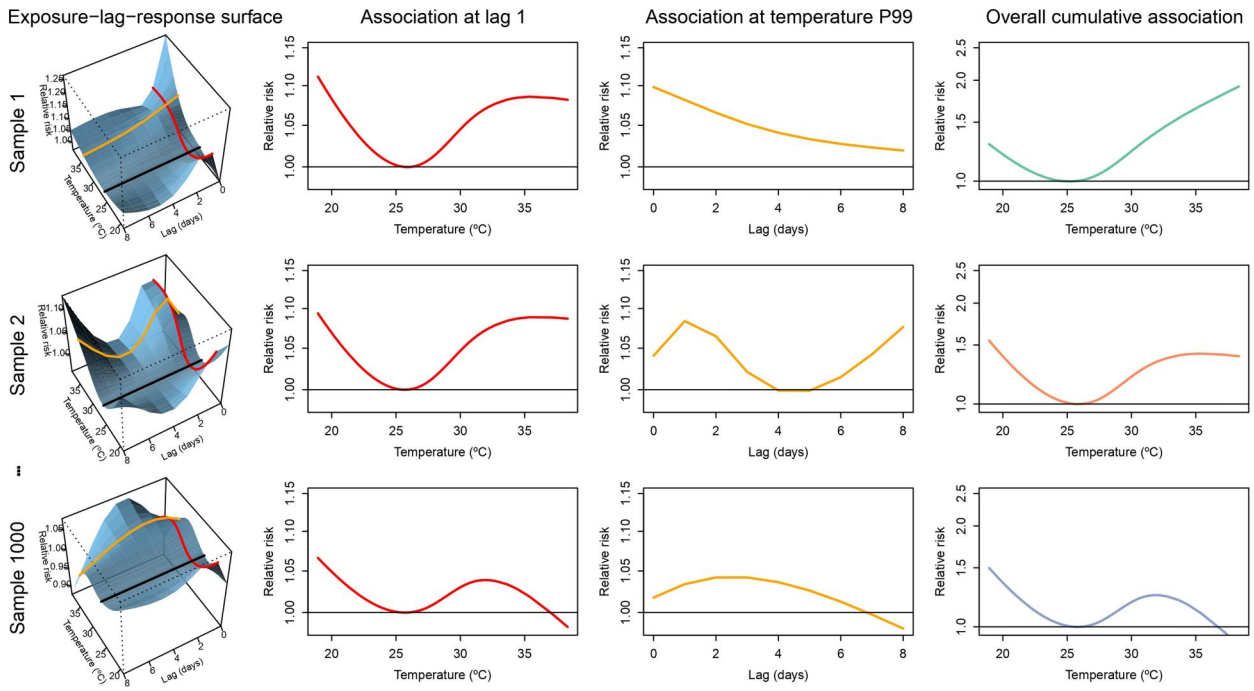

Figure 1. Ensemble of associations between temperature, lag and mortality derived from a Bayesian distributed lag non-linear model (B-DLNM). The results are based on the spatial Bayesian DLNM (SB-DLNM) with the case-crossover design (Model 3) in the el Raval neighbourhood (2007–2016). Each row corresponds to one of the (1000) independent samples generated by the model. All associations are centred in the 30th percentile of the temperature distribution (indicated by the black line in the surfaces). The first column shows exposure-lag-response surfaces. The second column shows temperature-specific associations at lag 1. The third column shows lag-specific associations at the 99th percentile (P99) of the temperature distribution. The fourth column shows overall cumulative temperature-mortality associations

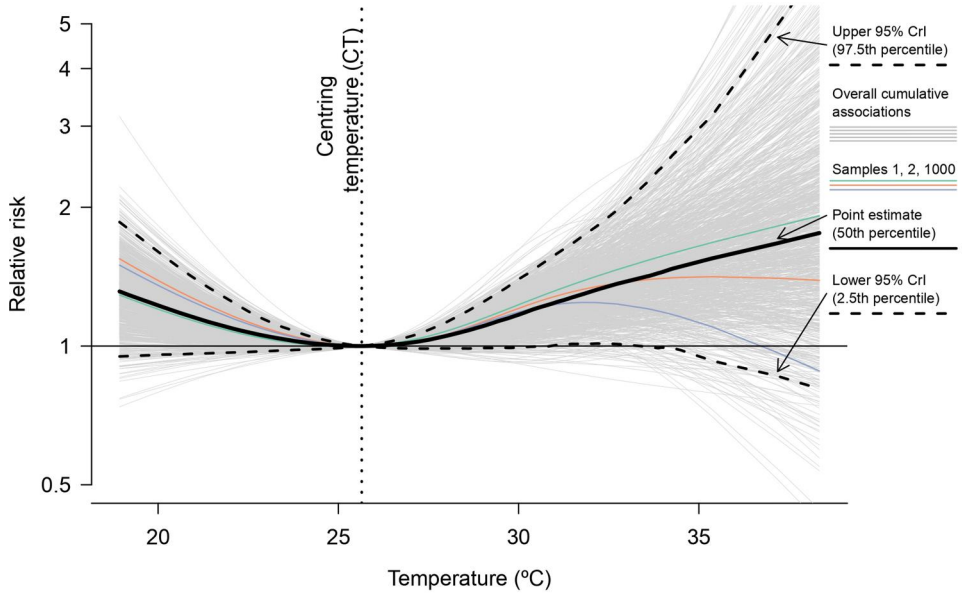

Figure 2. Summary estimates of relative risks derived from a Bayesian distributed lag non-linear model (B-DLNM). The results are based on the spatial Bayesian DLNM (SB-DLNM) with the case-crossover design (Model 3) in the el Raval neighbourhood (2007–2016). The point estimates and confidence intervals (CrI) of the relative risks are derived from the ensemble (of 1000 samples) of overall cumulative temperature-mortality associations. The overall cumulative associations are calculated by summing the lag-specific contributions and are centred on the 30th percentile of the temperature distribution

We adapted the DLNM framework to Bayesian statistics, and used spatial models to account for spatial dependences. These one-stage models were formulated for the two most common designs in this type of analysis: the case-crossover and time-series designs.

In the practical case study, we implemented independent B-DLNMs and SB-DLNMs to assess the short-term association between temperature and mortality in 73 neighbourhoods in Barcelona. We illustrated how to process the outputs of the four Bayesian models, and highlighted the benefits of incorporating the spatial component to address unstable estimates from small numbers. Indeed, estimates obtained from independent B-DLNMs are unstable and unreliable, particularly in neighbourhoods with lower numbers of deaths, with noisy spatial

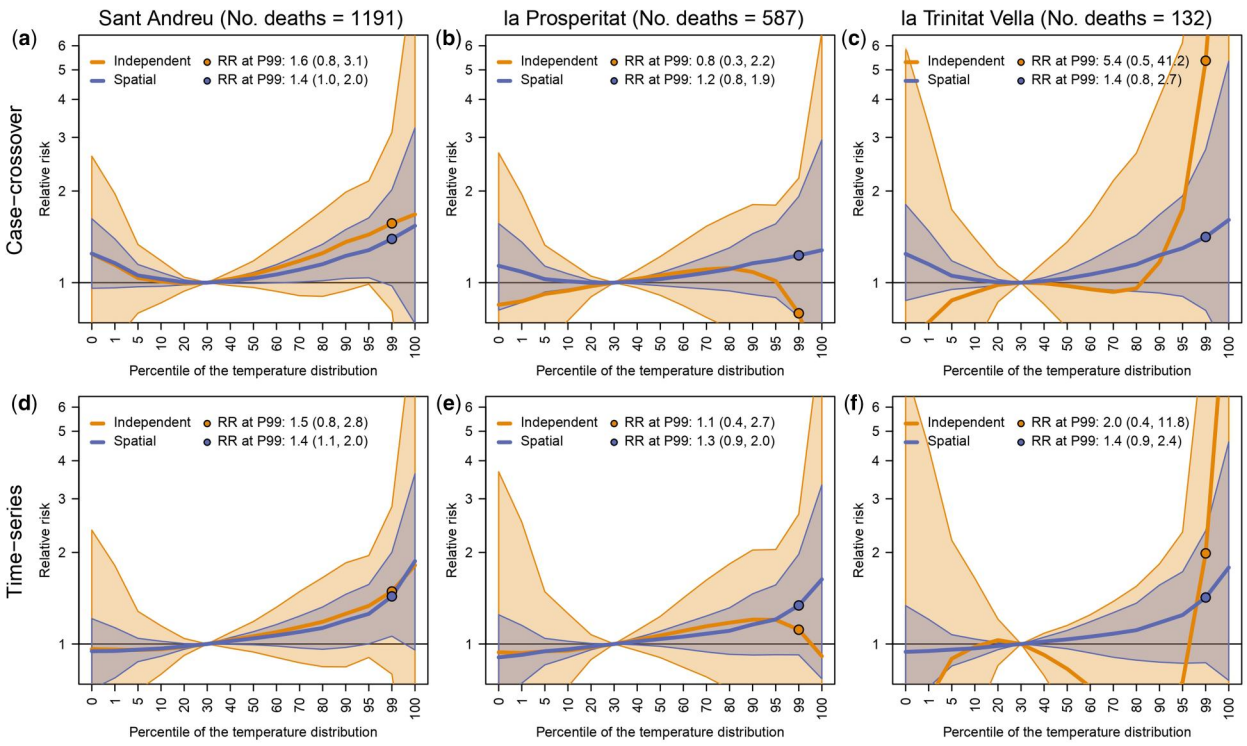

Figure 3. Overall cumulative temperature-mortality associations in three adjacent neighbourhoods of Barcelona (2007–2016) for independent Bayesian distributed lag non-linear model (B-DLNM) and spatial Bayesian distributed lag non-linear model (SB-DLNM). We selected three adjacent neighbourhoods with high (a, d), medium (b, e) and low (c, f) numbers of deaths. The top row (a, b, c) shows the associations for the case-crossover design, and the bottom row (d, e, f) shows the association for the time-series design. In each neighbourhood, the associations for the location-independent Bayesian DLNM (orange; Models 1 and 2) and spatial Bayesian DLNM (blue; Models 3 and 4) are represented. The thick curves represent the points estimates (ie median values) of the association between the percentile of the neighbourhood temperature distribution and the relative risk of mortality. To facilitate the comparison, the baseline risk is centred at the 30th percentile of the neighbourhood temperature distribution. The coloured areas represent the 95% credible interval of the estimates. The relative risks (RRs) at the 99th percentile (P99) are provided at the top of the figures, including the point estimates (also marked with dots in the curves) and their corresponding 95% credible intervals

patterns. In contrast, SB-DLNMs provide more reasonable estimates and reduce uncertainties by considering the spatial structure among adjacent neighbourhoods. Differences in RR between adjacent neighbourhoods are spatially smoothed, revealing clusters of neighbourhoods with elevated risks.

So far, most of the approaches for studying temperature-mortality associations in small areas have used the case-crossover design or similar self-matched approaches.[15,16,18,19,33] With SB-DLNMs, we model the complex relation between temperature and mortality by incorporating flexible splines accounting for the lagged non-linear effects. In comparison, Bennet and colleagues only used a single coefficient (ie a linear term) for this association in a case-crossover design.[19] Just as in SB-DLNMs, the case time series design uses the full flexibility of DLNMs. Whereas the case time series is a computationally efficient and flexible tool for analysing small-area data,[16,18] SB-DLNMs address some of its limitations. SB-DLNMs can directly estimate temperature-mortality associations at the small-area level by modelling spatial differences and structures across small areas, even in cases with as few as 19 deaths. In contrast, the case time series design assumes a common association across all small areas, lacking the geographical variation estimated in SB-DLNMs. While the case time series can be extended in a multilevel approach, predicting risk at the small-scale through a second-stage meta-analysis using other environmental meta-predictors,[17] SB-DLNMs can directly estimate risk in a one-stage model, relying only on the associations from the temperature and mortality data. As an additional innovative feature, we also formulate SB-DLNMs with the time-series design, widely used in large-scale epidemiological studies, providing somewhat more precise estimates in our case study compared with the case-crossover design.

The proposed SB-DLNMs have some limitations that should be acknowledged. The first and most significant limitation is the long execution times and high computational demands, which are common in Bayesian methods, particularly in spatial statistics involving spatial correlations. There is significant room for improvement in the implementation of the SB-DLNMs presented here (such as using an alternative MCMC engine faster than WinBUGS) which, in its current state, took several days to converge in our non-optimized setting. Nevertheless, this study aims to present the model specification for SB-DLMNs, so that future studies may develop more efficient implementations of SB-DLNMs for studies with a larger number of areas.[16,17] Due to these computational challenges, we did not conduct a simulation study that further validates the robustness of the models under different controlled conditions. By providing a reproducible R code, we encourage the community of environmental epidemiology to test SB-DLNMs for their own data and studies. Additionally, some of our assumptions involve not incorporating spatial models to control for seasonality,[13] not accounting for potential overdispersion[34] and assuming the same spatial parameter (ρ) across the coefficients of the cross-basis. The flexibility of the Bayesian framework allows the easy adaptation of the model to address these issues.

Extending beyond small-area analysis, SB-DLNMs could be directly applied in multi-city or large-area studies to

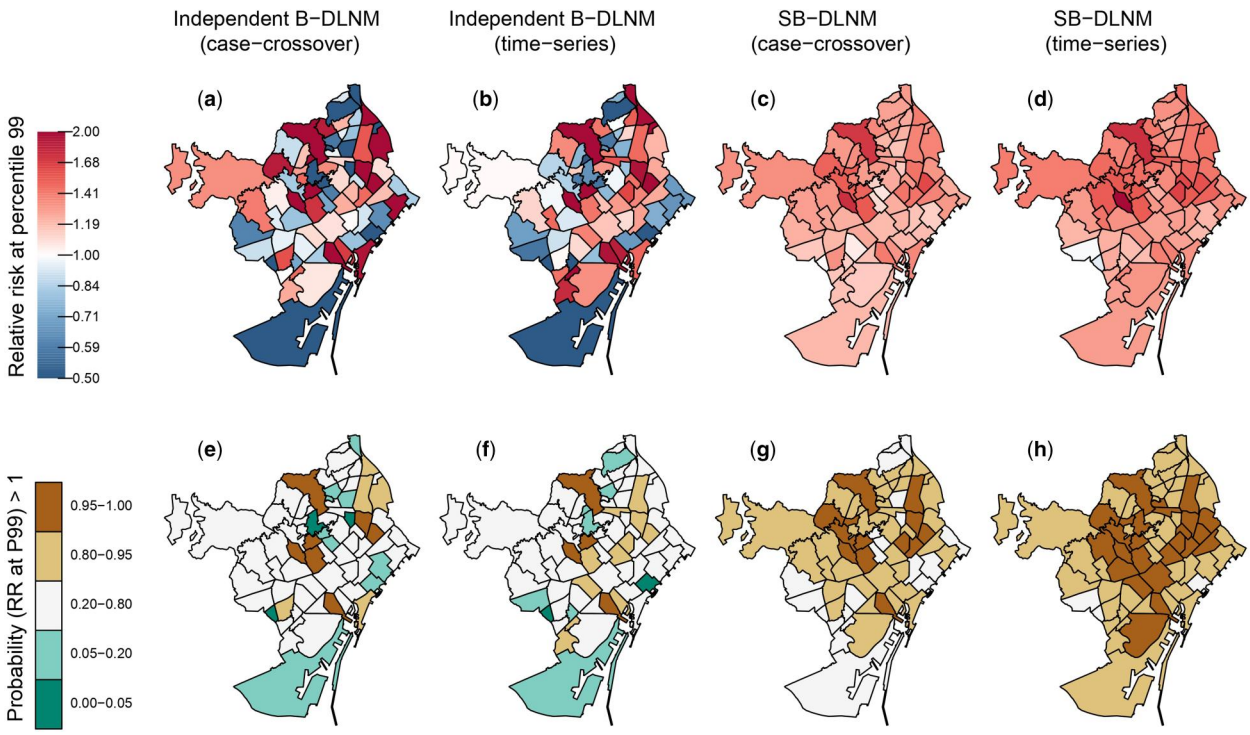

Figure 4. Spatial distribution of heat-related mortality risks in Barcelona (2007–2016) for independent Bayesian distributed lag non-linear model (B-DLNM) and spatial Bayesian distributed lag non-linear model (SB-DLNM). The top row (a–d) shows the point estimates of the relative risk (RR) of death at the 99th percentile (P99) of the neighbourhood temperature distribution, compared with the risk at the centring temperature in the 30th percentile. The bottom row (e–h) shows the probability of excess risk at the 99th percentile, which measures the probability of the relative risk being greater than 1. Panels (a), (b), (e) and (f) represent the results from the independent Bayesian DLNM, and panels (c), (d), (g), (h) represent the results from the spatial Bayesian DLNM

account for spatial correlations, which are not considered in the standard two-stage design. In both small- and large-area studies, the Bayesian framework enriches the range of estimates and tests that can be derived from temperature-mortality studies. Among others, the uncertainty of derived variables, such as the minimum mortality temperature, has in our framework an intuitive interpretation, given that it is directly calculated from the ensemble of overall cumulative exposure-response functions (see Supplementary Figure S2, available as Supplementary data at *IJE* online). In contrast, the calculation of uncertainties in a frequentist approach is more complex and requires bootstrapping techniques.[35] Furthermore, the posterior distributions in Bayesian models allow for richer estimates and tests of the risk, such as the probability of excess risk, based on the fraction of samples with RR greater than 1, rather than relying on significance tests based on arbitrary *P*-values or the width of confidence intervals. SB-DLNMs, as presented in this study, emphasize modelling spatial structures. Future SB-DLNMs studies could focus on identifying vulnerability drivers explaining these structures.[17,36,37] This could be explored by directly incorporating socioeconomic or physical factors into the risk estimation, and testing their impact on defining the spatial structures. However, the feasibility of incorporating and interpreting these modifying factors will require the development and optimization of SB-DLNMs.

Conclusion

Spatial Bayesian distributed lag non-linear models (SB-DLNMs) are a novel framework for estimating the short-term lagged health effects of environmental exposures in small areas. We extended the state-of-the-art approach for exposure-lag-response epidemiological modelling, DLNMs, by incorporating spatial models within a single-stage Bayesian framework to account for spatial dependences in risks. SB-DLNMs provide geographically varying estimation of the risks at small areas while maintaining the full flexibility of DLNMs. We provide the formulation of SB-DLNMs and guidance on managing their outputs in an available R code so other researchers can use SB-DLNMs for their own data and studies.

Ethics approval

This article is based on a secondary analysis of administrative data, and does not contain any studies with human participants performed by any of the authors. Informed consent of individuals, or approval by a medical ethics board, was not required under national regulations.

Data availability

The mortality data underlying this article were gathered from the mortality registry of Barcelona, which draws on both the mortality registry of the Department of Health of Catalonia and the municipal population registry. Data will be shared upon request to the corresponding author with permission from both the Department of Health of Catalonia and the municipal population registry. The temperature data underlying this article were provided by VITO [https://vito.be/en] by permission. Temperature data will be shared on request to the corresponding author with permission of VITO.

Supplementary data

Supplementary data are available at *IJE* online.

Author contributions

M.Q-Z. and M.M-D. conceptualized the study. M.Q-Z. and M.M-D. performed the statistical analysis. M.Q-Z. wrote the original draft including visualization. All authors contributed to the study design, the interpretation of the results and the drafting of the manuscript.

Funding

M.Q-Z. and J.B. gratefully acknowledge funding from the European Union's Horizon 2020 and Horizon Europe research and innovation programmes under grant agreement no. 865564 (European Research Council Consolidator Grant EARLY-ADAPT) [https://www.early-adapt.eu/], 101069213 (European Research Council Proof-of-Concept HHS-EWS) and 101123382 (European Research Council Proof-of-Concept FORECAST-AIR). ISGlobal authors acknowledge support from the grant CEX2018-000806-S funded by MCIN/AEI/10.13039/501100011033, and support from the Generalitat de Catalunya through the CERCA Program. M.A.M-B. acknowledges support from Project PID2022-136455NB-I00 funded by MCIN/AEI/10.13039/501100011033/FEDER, UE.

Conflict of interest

None declared.

References

1. Zhao Q, Guo YLY, Ye T *et al.* Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019: a three-stage modelling study. *Lancet Planet Health* 2021;**5**:e415–25.
2. Ballester J, Quijal-Zamorano M, Méndez Turrubiates RF *et al.* Heat-related mortality in Europe during the summer of 2022. *Nat Med* 2023;**29**:1857–66.
3. Vicedo-Cabrera AM, Scovronick N, Sera F *et al.* The burden of heat-related mortality attributable to recent human-induced climate change. *Nat Clim Chang* 2021;**11**:492–500.
4. Martínez-Solanas È, Quijal-Zamorano M, Achebak H *et al.* Projections of temperature-attributable mortality in Europe: a time series analysis of 147 contiguous regions in 16 countries. *Lancet Planet Health* 2021;**5**:e446–54.
5. Quijal-Zamorano M, Martínez-Solanas È, Achebak H *et al.* Seasonality reversal of temperature attributable mortality projections due to previously unobserved extreme heat in Europe. *Lancet Planetary Health*. 2021;**5**:e573–75.
6. Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models. *Stat Med* 2010;**29**:2224–34.
7. Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. *J Stat Soft* 2011;**43**:2–20.
8. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat Med* 2012;**31**:3821–39.
9. Sera F, Armstrong B, Blangiardo M, Gasparrini A. An extended mixed-effects framework for meta-analysis. *Stat Med* 2019;**38**:5429–44.
10. Sera F, Gasparrini A. Extended two-stage designs for environmental research. *Environ Heal A Glob Access Sci Source* 2022;**21**:41.
11. Piel FB, Fecht D, Hodgson S *et al.* Small-area methods for investigation of environment and health. *Int J Epidemiol* 2021;**49**:686–99.
12. Rao JNK, Molina I. Small Area Models. In: Rao JNK, Molina I (eds). *Small Area Estimation*. Hoboken, New Jersey: John Wiley & Sons, 2015, pp. 75–96.
13. Lowe R, Lee SA, O'Reilly KM *et al.* Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *Lancet Planet Health* 2021;**5**:e209–19.
14. Konstantinoudis G, Minelli C, Vicedo-Cabrera AM, Ballester J, Gasparrini A, Blangiardo M. Ambient heat exposure and COPD hospitalisations in England: a nationwide case-crossover study during 2007-2018. *Thorax* 2022;**77**:1098–104.
15. Xu Y, Dadvand P, Barrera-Gómez J *et al.* Differences on the effect of heat waves on mortality by sociodemographic and urban landscape characteristics. *J Epidemiol Community Health* 2013;**67**:519–25.
16. Gasparrini A. A tutorial on the case time series design for small-area analysis. *BMC Med Res Methodol* 2022;**22**:129.
17. Gasparrini A, Masselot P, Schneider R *et al.* Small-area assessment of temperature-related mortality risks in England and Wales: a case time series analysis. *Lancet Planet Health* 2022;**6**:e557–64.
18. Gasparrini A. The case time series design. *Epidemiology* 2021;**32**:829–37.
19. Bennett JE, Blangiardo M, Fecht D, Elliott P, Ezzati M. Vulnerability to the mortality effects of warm temperature in the districts of England and Wales. *Nature Clim Change* 2014;**4**:269–73.
20. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 1991;**133**:144–53.
21. Madaniyazi L, Tobías A, Vicedo-Cabrera AM *et al.* Should we adjust for season in time-series studies of the short-term association between temperature and mortality? *Epidemiology* 2023;**34**:313–18.
22. Leroux BG, Lei X, Breslow N, Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence. In: Halloran M. Elizabeth, and D. Berry, eds. *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. New York, NY: Springer New York; 2000, pp. 179–91.
23. Alahmad B, Khraishah H, Royé D *et al.* Associations between extreme temperatures and cardiovascular cause-specific mortality: results from 27 countries. *Circulation* 2023;**147**:35–46.
24. Armstrong BG, Gasparrini A, Tobias A. Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis. *BMC Med Res Methodol* 2014;**14**:122.
25. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 1991;**43**:1–20.
26. Martinez-Beneito MA, Botella-Rocamora P. *Disease Mapping*. Boca Raton: Taylor & Francis, 2019.
27. De Ridder K, Lauwaet D, Maiheu B. UrbClim—a fast urban boundary layer climate model. *Urban Clim* 2015;**12**:21–48.
28. Gasparrini A, Guo Y, Hashizume M *et al.* Temporal variation in heat–mortality associations: a multicountry study. *Environ Health Perspect* 2015;**123**:1200–207.
29. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: Concepts, structure, and extensibility. *Stat Comput* 2000;**10**:325–37.
30. Vergara-Hernández C, Martinez-Beneito MA. *pbugs*. https://github.com/fisabio/pbugs (22 December 2023, date last accessed).
31. Ntzoufras I. *Bayesian Modeling Using WinBUGS*. Vol. **698**. Hoboken, New Jersey: John Wiley & Sons, 2008, pp. 1–492.
32. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. New York: Chapman and Hall/CRC, 2013.
33. Milojevic A, Armstrong BG, Gasparrini A, Bohnenstengel SI, Barratt B, Wilkinson P. Methods to estimate acclimatization to urban heat island effects on heat-and cold-related mortality. *Environ Health Perspect* 2016;**124**:1016–22.

34. Barrera-Gómez J, Puig X, Ginebra J, Basagaña X. Conditional poisson regression with random effects for the analysis of multi-site time series studies. *Epidemiology* 2023; **34**:873–78.
35. Tobías A, Armstrong B, Gasparrini A. Brief report: investigating uncertainty in the minimum mortality temperature. *Epidemiology* 2017;**28**:72–76.
36. Marí-Dell'Olmo M, Tobías A, Gómez-Gutiérrez A *et al.* Social inequalities in the association between temperature and mortality in a South European context. *Int J Public Health* 2019; **64**:27–37.
37. Sera F, Armstrong B, Tobias A *et al.* How urban characteristics affect vulnerability to heat and cold: a multi-country analysis. *Int J Epidemiol* 2019;**48**:1101–12.

International Journal of Epidemiology, 2024, 53, 1–9
https://doi.org/10.1093/ije/dyae061
Original article