A Bayesian Spatial Modeling Approach to Mortality Forecasting

Zhen Liu[1], Xiaoqian Sun[1] Yu-Bo Wang[1*]
[1]School of Mathematical and Statistical Sciences,
Clemson University, SC, USA

*Corresponding author: yubow@clemson.edu

Abstract

This paper extends Bayesian mortality projection models for multiple populations considering the stochastic structure and the effect of spatial autocorrelation among the observations. We explain high levels of overdispersion according to adjacent locations based on the conditional autoregressive model. In an empirical study, we compare different hierarchical projection models for the analysis of geographical diversity in mortality between the Japanese counties in multiple years, according to age. By a Markov chain Monte Carlo (MCMC) computation, results have demonstrated the flexibility and predictive performance of our proposed model.

Keywords: Mortality Projection; Overdispersion; MCMC; Conditional Autoregressive Model

1 Introduction

Mortality rates have been varying in most countries over half a century. Many researchers work on mortality projection modelings applied in wide social fields, such as demography and insurance. With the development of approaches, people focus on the mortality information evolving from single population into multiple regions. For some adjacent districts, the conditions of mortality would be influenced mutually by each other and potential missing data issues can be solved using the accurate estimates from neighboring areas. Hence, spatiotemporal modeling and stochastic modeling of mortality forecasting for age groups has drawn more attentions in decades.

Besag et al. (1991) suggested a model which combined an autoregressive component for adjacent spatial effects and independent error terms component for the observed variance without explanations by the spatial structure for disease data. However, the temporal effect to the spatial heterogeneity is not discussed to establish the disease risk mapping model over time.

Knorr-Held (2000) considered the interaction of spatial and temporal factors and introduced a Bayesian spatio-temporal model of Ohio lung cancer accounting for the variation mixed in the geographical diversity involved across each period. The asso-

ciation of temporal varying parts overcomes the drawback of limitation of the single spatial model for panel data.

Arató et al. (2006) worked on the research of age-dependent deaths in 150 geographic regions of Hungary by the Bayesian approach. They assumed a binomial distribution as the underlying model for respective age, region and groups in 1997. This structure focuses on the mortality estimation in a single year rather than the projection for following periods.

Greco and Scalone (2013) proposed a extended Poisson log-linear model for mortality risk of every provinces in Italy from 1992 to 2009, which sets a Bayesian hierarchical spatio-temporal modeling framework based on a decomposition for mortality rate. The authors propose separate models for each age class without borrowing strength from each age groups and the stochastic structure of forecasting is also dropped in this model. For more countries, see Congdon (2007), Rau and Schmertmann (2017), Bergeron-Boucher et al. (2018), etc.

Lee and Carter (1992) contributed a log-bilinear mortality projection model including separate age-specified parameters and time-series structure parameters which represent the trend of mortality variation over time. The Lee Carter (LC) model implies the future mortality rate will follows the same tendency as past periods, but without considering the accidental effects and improvements of society, since only the stochastic temporal component is responsible for variation in the estimation.

Considering the limitation of variation resources in the original LC model, Brouhns et al. (2002) suggested a hierarchical Poisson log-bilinear model. The extended model keep the basic form of mortality risk and capture more information of variability via including death and exposure to risk. It is commonly named Poisson Lee-Carter (PLC) model in following researches.

Wong et al. (2018) proposed a Poisson Log-normal Lee Carter (PLNLC) model, which introduced an overdispersion term with a normal distribution to the structure of mortality rate. Their method overcomes drawbacks of mean–variance equality assumption in the PLC model by adding more mortality variations across units. For other extensions of log normal structure about mortality rate, readers are referred to Cairns

et al. (2006), Delwarde et al. (2007), Hyndman and Ullah (2007), etc.

To reflect more possible uncertainties which affect predicted values and prediction intervals, Czado et al. (2005) implemented the Poisson LC model in the Bayesian framework by Markov chain Monte Carlo (MCMC) sampling. Besides these works on the mortality projections for single population, some researchers also discuss the related extension of the LC model within multiple populations (see Li and Lee (2005), Pedroza (2006), Cairns et al. (2011), Antonio et al. (2015), Liu et al. (2020), etc.).

In this paper, we intend to take advantage of age-specific factors and stochastic structures in LC model for the projections to fill the gap of the integrated estimation and prediction of age-stratified mortality rates for multiple nationwide adjacent regions under Bayesian paradigm in a long run. Most of mortality research focused on the modeling of age-time alone or separate spatial effect in the single period. An essential feature of our model is to link spatial effects to time effects for each age division in the estimation of mortality with multiple populations. We employ spatial factors in the model, which extends with univariate distribution to a multivariate normal structure of overdispersion term involving more information among areas. The variances are assumed to be independent for each population in the frame of univariate overdispersion. This assumption probably leads to the lack of variation issue. Comparably, our overdispersion terms based on dependent covariance structure overcome the previous drawback since considering more interactions of adjacent regions.

The remainder of the paper is organized as follows. In section 2, we presents a brief description of trends in mortality for every county in Japan. Section 3 introduces the proposed model along with the prior settings and detailed steps of an Markov chain Monte Carlo (MCMC) sampling. The real data is applied to illustrate the advantage of the proposed approach in Section 4. Finally, we conclude with a discussion in Section 5.

2 Geographical Mortality Data

In order to perform age-specific mortality levels at county's level, we take into account mortality rates by representative age groups and counties in three separate years between 1975 and 2005. The adopted data are collected by the Japan Mortality Database (JMD) [1] referring to 45 Japanese counties from 1975.

From figures 1 to figure 3, counties are classified by direct mortality rates level in age group of 30, 50 and 70 for three given years (1975, 1990 and 2005). The general declining trend which occurred in each age group over time can be demonstrated in these figures. Indeed, we can easily see that the first maps in four figures are dominated by the darkest highest mortality categories, since almost all the counties fell in the highest intervals. On the other hand, we can clearly see the lowest mortality categories in 2005.

Spatial correlation of mortality is supported by the fact that higher death rates are concentrated in some visible cluster of counties. In this respect, it indicates that the importance of spatial structure components when modeling mortality rates for sub-populations in terms of county.

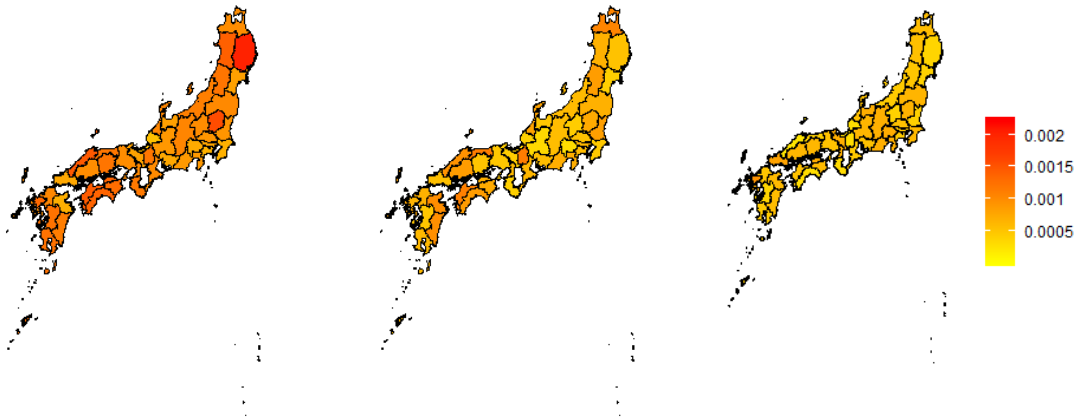


Fig. 1. Maps of observed mortality rates by county at **age 30** in Japan at 1975 (left), 1990 (middle) and 2005 (right).

[1] http://www.ipss.go.jp/p-toukei/JMD/index-en.asp

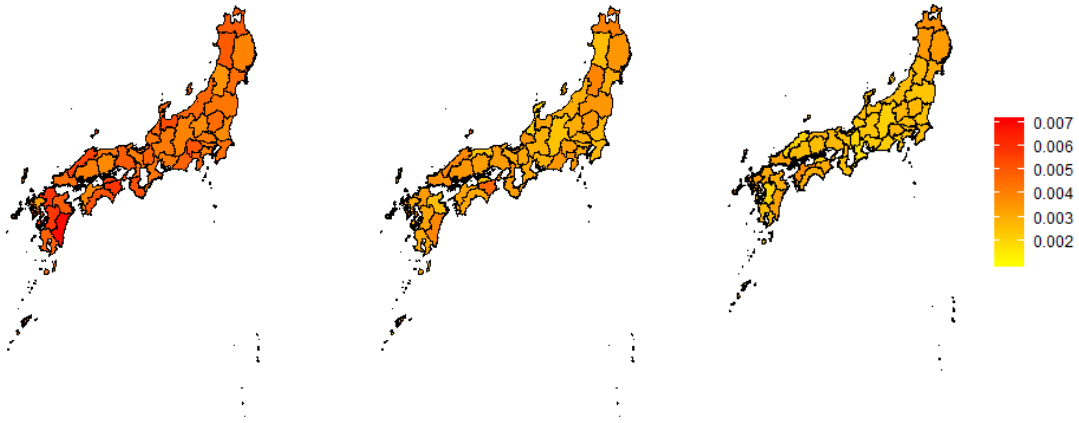


Fig. 2. Maps of observed mortality rates by county at **age 50** in Japan at 1975 (left), 1990 (middle) and 2005 (right).

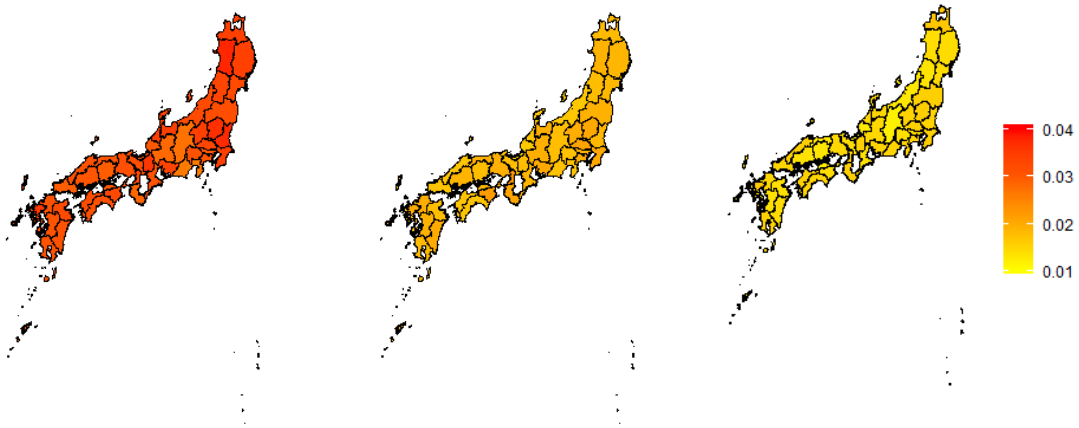


Fig. 3. Maps of observed mortality rates by county at **age 70** in Japan at 1975 (left), 1990 (middle) and 2005 (right).

3 The Proposed Model

3.1 Literature Review

Besag et al. (1991) discussed a spatial modeling method to estimate relative risks of diseases under a Bayesian framework. The idea is based on the decomposition of spatial structure component and specified regional random effects for a risk parameter η_i. It can be depicted as

$$\eta_i = \mu + \phi_i + \theta_i, \tag{1}$$

where μ is a fixed effect representing the overall risk level across all areas. ϕ_i is an conditional autoregressive spatial component and θ_i is an ordinary random effect for diversity. Knorr-Held (2000) introduced a Bayesian spatio-temporal model with a expression for log odds as a combination of temporal and spatial main effects and spatio-temporal interactions.

$$\eta_{it} = \ln\{\pi_{it}/(1-\pi_{it})\} \qquad \text{with} \qquad \eta_{it} = \alpha_t + \theta_i + \phi_i + \delta_{it}, \tag{2}$$

where α_t and is a temporal effects of year t. θ_i and ϕ_i represent features of area i with individual spatial structure and δ_{it} is a spatio-temporal parameter. The object of analysis is combined 55 to 64 years white male in this model. Further hierarchical modeling and extensions stratified by age are not described in detail.

In the Lee-Carter (LC) model, the central death rate $m_{x,t}$ for age x at year t is formulated by the following equation:

$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t} \quad , \tag{3}$$

where α_x denotes the logarithm of mortality rate averaged over time, whereas β_x denotes the age-specified sensitivity from the averaged pattern as mortality trend varies. The trend of mortality change in periods is represented by the time parameter κ_t. The component $\epsilon_{x,t}$ denotes the error term, with mean 0 and an identical variance σ_ϵ^2,

reflecting extra variations excluding age and time effects in the model.

Brouhns et al. (2002) modified the LC model into the following Poisson framework:

$$D_{x,t} \mid \mu_{x,t} \sim \text{Poisson}(E_{x,t}\mu_{x,t}) \qquad \text{with} \quad \log \mu_{x,t} = \alpha_x + \beta_x \kappa_t, \tag{4}$$

where $D_{x,t}$ is the death toll for the group aged x at time t, $E_{x,t}$ is the matched exposure at risk and $\mu_{x,t}$ is the corresponding theoretic mortality rate.

3.2 Bayesian Lee-Carter Model with Spatial Structure

Let θ_i denote the i^{th} spatial random effect for $i = 1, 2, \ldots, n$. We propose the Bayesian Lee-Carter model with spatial structure as follows,

$$\begin{aligned} & D_{x,t}^{(i)} \mid \mu_{x,t}^{(i)} \sim \text{Poisson}(E_{x,t}^{(i)}\mu_{x,t}^{(i)}), \\ & \text{with} \quad \log \mu_{x,t}^{(i)} = \alpha_x + \beta_x \kappa_t + \gamma_x \theta_i, \\ & \kappa_t = \varphi_1 + \varphi_2 t + \rho[\kappa_{t-1} - \varphi_1 - \varphi_2(t-1)] + \epsilon_t, \end{aligned} \tag{5}$$

where $\epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_\kappa^2)$. Here, the roles of α_x, β_x and κ_t are same as ones in the LC model. Moreover, γ_x denotes the overall sensitivity of the mortality aged x to all regions. Additionally, we expect to explore the interaction between age and spatial effects to explain the extra spatial heterogeneity accounting for variability in different generations.

The first two lines can be viewed as a generalization of PLC model (4) to a multi-population problem at geographical level, and the last one describes the autoregressive of order one (AR1) with drift structure of κ_t.

Let

$$\boldsymbol{U}_{N\times N}=\begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -\rho & 1 & & & \vdots \\ 0 & -\rho & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \cdots & & -\rho & 1 \end{bmatrix}, \boldsymbol{W}=\begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_N \end{bmatrix},$$

$\boldsymbol{\varphi}=(\varphi_1,\varphi_2)'$, $\boldsymbol{Q}=\boldsymbol{U'U}$, $n+1$ dependence structures of time effects can be written as $\boldsymbol{\kappa}\sim N(\boldsymbol{W\varphi},\sigma_\kappa^2\boldsymbol{Q}^{-1})$ where $\boldsymbol{\kappa}=(\kappa_1,\kappa_2,\ldots,\kappa_N)'$.

For model identifiability, we also want to point out that some constraints with two same ones as Lee and Carter (1992)

$$\sum_{x\in\Theta_{\text{age}}}\beta_x=1 \quad \text{and} \quad \sum_{t\in\Theta_{\text{time}}}\kappa_t=0. \tag{6}$$

Here, we suppose $\Theta_{\text{age}}=\{x_1,x_1+1,\ldots,x_1+M-1\}\equiv\{x_1,x_2,\ldots,x_M\}$, $\Theta_{\text{time}}=\{t_1,t_1+1,\ldots,t_1+N-1\}\equiv\{t_1,t_2,\ldots,t_N\}$ and $\Theta_{\text{space}}=\{\theta_1,\theta_2,\ldots,\theta_P\}$ denoted the sets of age, time and space considered in the training dataset respectively.

For all sets of spatial parameters, we assume an Conditional AutoRegressive (CAR) structure, which can be formulated as a multivariate Normal distribution in terms of mean vector and covariance matrix. This method is first introduced by Besag (1974) to analyze the similar properties of neighboring regions. We also refer extensions of CAR model about Bayesian sampling contributed by Banerjee et al. (2014) and De Oliveira (2012).

As for $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_P)'$, assuming

$$\boldsymbol{\theta}\sim N_p\left(\mathbf{0},\sigma_\theta^2(\boldsymbol{M_\theta}-\lambda\boldsymbol{W_\theta})^{-1}\right), \qquad \text{where} \qquad \lambda\in(\lambda_{min}^{-1},\lambda_{max}^{-1}) \tag{7}$$

and $\lambda_{min},\lambda_{max}$ are the smallest and largest eigenvalues in $\boldsymbol{W_\theta}$,

Denote $\boldsymbol{Q_\theta}=\boldsymbol{M_\theta}-\lambda\boldsymbol{W_\theta}$, then

$$\boldsymbol{\theta}\sim N_p\left(\mathbf{0},\sigma_\theta^2\boldsymbol{Q_\theta^{-1}}\right), \tag{8}$$

where $\boldsymbol{W_\theta}$ is the adjacency matrix whose generic ij entry $w_{ij} = 1$ if i and j are considered as neighbors, while $w_{ij} = 0$ otherwise. Matrix $\boldsymbol{M_\theta}$ is the diagonal matrix with diagonal entries equal to the number of neighbors.

Assuming the conditional distribution of θ_i as follows:

$$\theta_i \mid \theta_{-i} \sim N(\lambda \sum_{i \neq j} b_{ij}\theta_j, \sigma_i^2), \qquad \text{where} \quad i = \{1, 2, \cdots, n\}, \tag{9}$$

$$b_{ij} = \frac{w_{ij}}{w_{i+}} \quad \text{and} \quad \sigma_i^2 = \frac{\sigma_\theta^2}{w_{i+}}, \tag{10}$$

$$w_{i+} = \sum_j w_{ij} \quad \text{and} \quad w_{ij} = \begin{cases} 1 & \text{if i and j are neighbors} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

3.3 Prior Specification

3.3.1 Prior Distributions for Age Parameters

To assure the tractable full conditional distribution of α, we conduct the same variable transformation $e_x = \exp(\alpha)$ as Czado et al. (2005) and propose

$$e_x \sim \text{Gamma}(a_x, b_x), \tag{12}$$

where a_x and b_x are pre-specified constants such that

$$\pi(e_x) = (b_x)^{a_x}(e_x)^{a_x - 1}\exp(-e_x b_x)/\Gamma(a_x).$$

As for $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_M)'$, we consider the following non-informative priors

$$\boldsymbol{\beta} \sim N\left(\frac{1}{M}\boldsymbol{J}_M, \sigma_\beta^2 \boldsymbol{I}_M\right),$$
$$\sigma_\beta^2 \sim \text{InvGamma}(a_\beta, b_\beta),$$

where $\boldsymbol{J}_M$ is a $M \times 1$ vector with all elements equal to 1, $\boldsymbol{I}_M$ is an identity matrix of size M, and a_β, b_β are pre-specified constants such that

$$\pi(\sigma_\beta^2) = b_\beta^{a_\beta}(\sigma_\beta^2)^{-a_\beta-1}\exp(-b_\beta/\sigma_\beta^2)/\Gamma(a_\beta).$$

Note that the proposed priors are non-informative in the sense that they all center at $1/M$, the constraint (=1) equally shared by M age groups.
Similarly, by setting $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_M)'$, we consider the following non-informative priors

$$\boldsymbol{\gamma} \sim N\left(\frac{1}{M}\boldsymbol{J}_M, \sigma_\gamma^2 \boldsymbol{I}_M\right),$$
$$\sigma_\gamma^2 \sim \text{InvGamma}(a_\gamma, b_\gamma).$$

3.3.2 Prior Distributions for Time Parameters

Suppose $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)'$, we consider the following priors for the dependence structure of $\boldsymbol{\kappa}$

$$\boldsymbol{\varphi} \sim N_2(\boldsymbol{\varphi}_0, \boldsymbol{\Sigma}_0),$$
$$\rho \sim N(0, \sigma_\rho^2)\mathbf{1}_{\{\rho\in(-1,1)\}},$$
$$\sigma_\kappa^2 \sim \text{InvGamma}(a_\kappa, b_\kappa),$$

where $\boldsymbol{\varphi}_0$, $\boldsymbol{\Sigma}_0$, σ_ρ^2, a_κ, and b_κ are pre-specified hyperparameters, and $\mathbf{1}_{\{\rho\in(-1,1)\}}$is an indicator function equal to 1 when ρ is between -1 and 1.

3.3.3 Prior Distributions for Spatial Parameters

Finally, setting the prior of λ and σ_θ as
$\lambda \sim \text{Unif}(\lambda_{min}^{-1}, \lambda_{max}^{-1})$ and
$\sigma_\theta^2 \sim \text{InvGamma}(a_\theta, b_\theta)$.

3.4 Posterior Computation

3.4.1 Posterior distributions for Age Parameters

The full conditional distributions of age parameters are given by

$$\pi(e_x \mid \cdot) \propto \exp(-c_x e_x)(e_x)^{D_{x,\cdot}} \left|\frac{d}{de_x} g^{-1}(\alpha_x)\right| \pi(e_x)$$
$$\propto \exp\left[-(b_x + c_x)e_x\right](e_x)^{a_x + D_{x,\cdot} - 1}, \tag{13}$$

$$\pi(\beta_x \mid \cdot) \propto \prod_{i=1}^{p} \prod_{t \in \Theta_{\text{time}}} \exp\left[-E_{x,t}^{(i)} \exp(\alpha_x + \beta_x \kappa_t + \gamma_x \theta_i)\right] \times \exp(\beta_x \kappa_t D_{x,t}^{(i)})$$
$$\times \exp\left[-\frac{(\beta_x - \frac{1}{M})^2}{2\sigma_\beta^2}\right], \tag{14}$$

$$\pi(\sigma_\beta^2 \mid \cdot) \propto (\sigma_\beta^2)^{-\tilde{a}_\beta - 1} \exp(-\tilde{b}_\beta / \sigma_\beta^2), \tag{15}$$

$$\pi(\gamma_x \mid \cdot) \propto \prod_{i=1}^{p} \prod_{t \in \Theta_{\text{time}}} \exp\left[-E_{x,t}^{(i)} \exp(\alpha_x + \beta_x \kappa_t + \gamma_x \theta_i)\right] \times \exp(\gamma_x \theta_i D_{x,t}^{(i)})$$
$$\times \exp\left[-\frac{(\gamma_x - \frac{1}{M})^2}{2\sigma_\gamma^2}\right], \tag{16}$$

$$\pi(\sigma_\gamma^2 \mid \cdot) \propto (\sigma_\gamma^2)^{-\tilde{a}_\gamma - 1} \exp(-\tilde{b}_\gamma / \sigma_\gamma^2), \tag{17}$$

where the notation " $\mid \cdot$ " represents "conditional on all other parameters and data G", $c_x = \sum_{i \in \Theta_{\text{space}}} c_x^{(i)}$ and $c_x^{(i)} = \sum_{t \in \Theta_{\text{time}}} E_{x,t}^{(i)} \exp(\beta_x \kappa_t + \gamma_x \theta_i)$, $D_{x,\cdot} = \sum_{i \in \Theta_{\text{space}}} D_{x,\cdot}^{(i)}$ and $D_{x,\cdot}^{(i)} = \sum_{t \in \Theta_{\text{time}}} D_{x,t}^{(i)} - 1$, $\tilde{a}_\beta = a_\beta + \frac{M}{2}$, $\tilde{b}_\beta = b_\beta + \frac{1}{2}\left(\boldsymbol{\beta} - \frac{1}{M}\boldsymbol{J}_M\right)'\left(\boldsymbol{\beta} - \frac{1}{M}\boldsymbol{J}_M\right)$, $\tilde{a}_\gamma = a_\gamma + \frac{M}{2}$, $\tilde{b}_\gamma = b_\gamma + \frac{1}{2}\left(\boldsymbol{\gamma} - \frac{1}{M}\boldsymbol{J}_M\right)'\left(\boldsymbol{\gamma} - \frac{1}{M}\boldsymbol{J}_M\right)$.

From (13), (15), and (17) we have

$$e_x \mid \cdot \sim \text{Gamma}(a_x + D_{x,\cdot}, b_x + c_x),$$
$$\sigma_\beta^2 \mid \cdot \sim \text{InvGamma}(\tilde{a}_\beta, \tilde{b}_\beta),$$
$$\sigma_\gamma^2 \mid \cdot \sim \text{InvGamma}(\tilde{a}_\gamma, \tilde{b}_\gamma)$$

and can sample each iteration of $e^{(i)}$, σ_β^2 and σ_γ^2 from the known distributions.

3.4.2 Posterior distributions for Time Parameters

In this section, we separately discuss the sampling algorithms for common and population-specific time parameters because the dependence structure of the latter is further regularized by the dirac spike. Let $\boldsymbol{\kappa}_{-t} = \boldsymbol{\kappa} \backslash \{\kappa_t\} = (\kappa_1, \ldots, \kappa_{t-1}, \kappa_{t+1}, \ldots, \kappa_{t_N})'$ and $\eta_t = \varphi_1 + \varphi_2 t$, the full conditional distributions of $\boldsymbol{\kappa}, \boldsymbol{\varphi}, \rho$, and σ_κ^2 are proportional to

$$\pi(\kappa_t \mid .) \propto \prod_{i=1}^{n} \prod_{x \in \Theta_{\text{age}}} \exp\left[-E_{x,t}^{(i)} \exp(\alpha_x + \beta_x \kappa_t + \gamma_x \theta_i)\right] \times \exp(\beta_x \kappa_t D_{x,t}^{(i)}) \times f(\kappa_t \mid \boldsymbol{\kappa}_{-t}), \tag{18}$$

$$\pi(\boldsymbol{\varphi} \mid \cdot) \propto \exp\left[-\frac{1}{2\sigma_\kappa^2}(\boldsymbol{\varphi}'(\boldsymbol{\Sigma}^*)^{-1}\boldsymbol{\varphi} - 2(\boldsymbol{\kappa}'\boldsymbol{Q}\boldsymbol{W} + \sigma_\kappa^2 \boldsymbol{\varphi}_0' \boldsymbol{\Sigma}_0^{-1})\boldsymbol{\varphi})\right], \tag{19}$$

$$\pi(\rho \mid \cdot) \propto \exp\left[-\frac{1}{2\sigma_\kappa^2}\left(a_\rho \rho^2 + \frac{\sigma_\kappa^2}{\sigma_\rho^2}\rho^2 - 2b_\rho \rho\right)\right] 1_{\{\rho \in (-1,1)\}}, \tag{20}$$

$$\pi(\sigma_\kappa^2 \mid \cdot) \propto (\sigma_\kappa^2)^{-(a_\kappa + N/2)-1} \exp\left[-\frac{1}{\sigma_\kappa^2}\left(b_\kappa + \frac{1}{2}(\boldsymbol{\kappa} - \boldsymbol{W}\boldsymbol{\varphi})'\boldsymbol{Q}(\boldsymbol{\kappa} - \boldsymbol{W}\boldsymbol{\varphi})\right)\right], \tag{21}$$

where $f(\kappa_t \mid \boldsymbol{\kappa}_{-t})$ are the conditional distribution of κ_t based on AR(1) with a drift in (5), $\boldsymbol{\Sigma}^* = (\boldsymbol{W}'\boldsymbol{Q}\boldsymbol{W} + \sigma_\kappa^2 \boldsymbol{\Sigma}_0^{-1})^{-1}$, $a_\rho = \sum_{t=t_2}^{t_N} (\kappa_{t-1} - \eta_{t-1})^2$, and $b_\rho = \sum_{t=t_2}^{t_N} (\kappa_t - \eta_t)(\kappa_{t-1} - \eta_{t-1})$. Note that when $t = t_1$,

$$\begin{aligned} f(\kappa_t \mid \boldsymbol{\kappa}_{-t}) &\propto f(\kappa_t) f(\kappa_{t+1} \mid \kappa_t) \\ &\propto \exp\left[-\frac{1}{2\sigma_\kappa^2}[(\kappa_t - \eta_t)^2 + (\kappa_{t+1} - \eta_{t+1} - \rho(\kappa_t - \eta_t))^2]\right]; \end{aligned} \tag{22}$$

when $t_1 < t < t_N$,

$$\begin{aligned} f(\kappa_t \mid \boldsymbol{\kappa}_{-t}) &\propto f(\kappa_{t+1} \mid \kappa_t) f(\kappa_t \mid \kappa_{t-1}) \\ &\propto \exp\left[-\frac{1}{2\sigma_\kappa^2}[(\kappa_t - \eta_t - \rho(\kappa_{t-1} - \eta_{t-1}))^2 + (\kappa_{t+1} - \eta_{t+1} - \rho(\kappa_t - \eta_t))^2]\right]; \end{aligned} \tag{23}$$

when $t = t_N$,

$$f(\kappa_t \mid \boldsymbol{\kappa}_{-t}) \propto f(\kappa_t \mid \kappa_{t-1}) \propto \exp\left[-\frac{1}{2\sigma_\kappa^2}(\kappa_t - \eta_t - \rho(\kappa_{t-1} - \eta_{t-1}))^2\right]. \tag{24}$$

From (19), (20), and (21), $\boldsymbol{\varphi}, \rho$ and σ_κ^2 are updated by

$$\begin{aligned}
\boldsymbol{\varphi} \mid \cdot &\sim N(\boldsymbol{\Sigma}^*(\boldsymbol{W}'\boldsymbol{Q}\boldsymbol{\kappa} + \sigma_\kappa^2\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\varphi}_0), \sigma_\kappa^2\boldsymbol{\Sigma}^*), \\
\rho \mid \cdot &\sim N\left(\frac{b_\rho}{a_\rho + \frac{\sigma_\kappa^2}{\sigma_\rho^2}}, \frac{\sigma_\kappa^2}{a_\rho + \frac{\sigma_\kappa^2}{\sigma_\rho^2}}\right) 1\{\rho \in (-1, 1)\}, \\
\sigma_\kappa^2 \mid \cdot &\sim \text{InvGamma}\left(a_\kappa + \frac{N}{2}, b_\kappa + \frac{1}{2}(\boldsymbol{\kappa} - \boldsymbol{W}\boldsymbol{\varphi})'\boldsymbol{Q}(\boldsymbol{\kappa} - \boldsymbol{W}\boldsymbol{\varphi})\right).
\end{aligned}$$

3.4.3 Posterior distributions for Spatial Parameters

The full conditional distribution is

$$\begin{aligned}
f(\theta_i \mid \cdot) \propto \prod_x \prod_t &\exp(-E_{x,t}^{(i)} \exp(\alpha_x + \beta_x\kappa_t + \gamma_x\theta_i)) \cdot \exp(\gamma_x\theta_i D_{x,t}^{(i)}) \\
&\times \exp(-\frac{1}{2\sigma_i^2}(\theta_i - \lambda\sum_{i\neq j} b_{ij}\theta_j)^2).
\end{aligned} \tag{25}$$

The posterior distributions of λ and σ_θ are

$$f(\lambda \mid \cdot) \propto \det(\Sigma_\theta)^{-\frac{1}{2}} \exp(-\frac{1}{2}\boldsymbol{\theta}'\Sigma_\theta^{-1}\boldsymbol{\theta}) \times \mathbf{1}_\lambda(\lambda_{min}^{-1}, \lambda_{max}^{-1}), \tag{26}$$

where Σ_θ denoted as $\sigma_\theta^2 Q_\theta^{-1}$.

$$\sigma_\theta^2 \sim \text{InvGamma}(a_\theta + \frac{P}{2}, b_\theta + \frac{1}{2}\boldsymbol{\theta}\boldsymbol{Q}_{\boldsymbol{\theta}}\boldsymbol{\theta}'). \tag{27}$$

4 Numerical Analysis

4.1 Dataset description

The data to illustrate our proposed method is deaths and exposures to risk in 45 counties (excluding Hokkaido and Okinawa without adjacent county) of Japan from the Japanese Mortality Database (JMD). We consider each county as a single population, and calibrate the model on the data from 1975 to 2004 for ages 0-99 while the data from 2005 to 2016 is separated for the validation purpose.

4.2 Initial values of prior distributions

Following the prior specifications in Sections 3.3.1-3.3.3, we consider $a_x = b_x = 1$, $a_\beta = b_\beta = 0.01$, and $a_\gamma = b_\gamma = 0.001$ as age-related hyperparameters. For those hyperparameters related to the spacial and time factors, they are set as $a_\theta = b_\theta = 0.1$, $\lambda = 0.9$, $a_\kappa = b_\kappa = 0.001$, $\sigma_\rho^2 = 1$, $\boldsymbol{\varphi}_0 = (0, 0)'$, $\Sigma_0 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$, respectively. It has to be mentioned that the pre-specified values here are similar to the ones in Czado et al. (2005) and non-informative relative to the size ($30 \times 100 \times 45$) of our analyzing data set.

4.3 Model Estimation

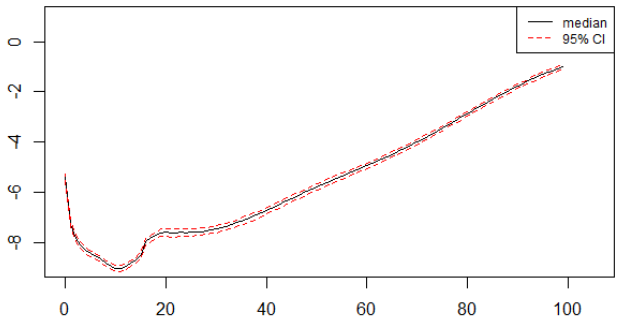
To evaluate the performance of our spatial model, we generate an MCMC sample of 20000 iterations with the first 10000 as burn-in period, and summarize the posterior medians of $\boldsymbol{\alpha}^{(i)}$, $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{(i)}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\kappa}^{(i)}$ along with the 95% highest posterior density (HPD) intervals in sections 4.3.1 and 4.3.2, respectively.

4.3.1 Estimation for Age Parameters

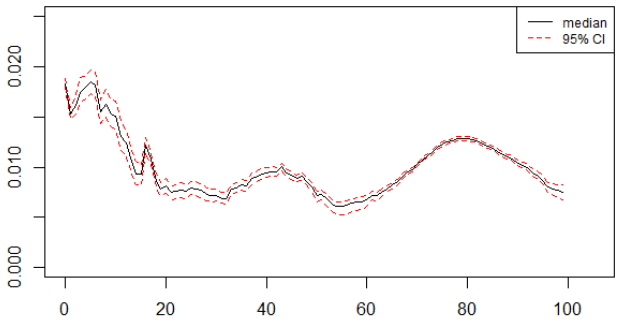
Figures 4(a) present the results of α_x under the spatial model, where the 95% HPD intervals are obtained by the method suggested in Hoff (2009). From Figure 4(a), we notice that the posterior distributions of α_x have small variances, and that there are

some features in the estimated curves: The decline from the infant stage to teenager is likely related to the immune system strengthened with growing age. Then, when ages are around 16-21, the health condition may not be the only decisive factor for the hump. It might be blamed on unnatural deaths caused by immature behaviors. For the adult-and-elder stage, the curves consistently increase since deaths happening in this stage are more related to aging.

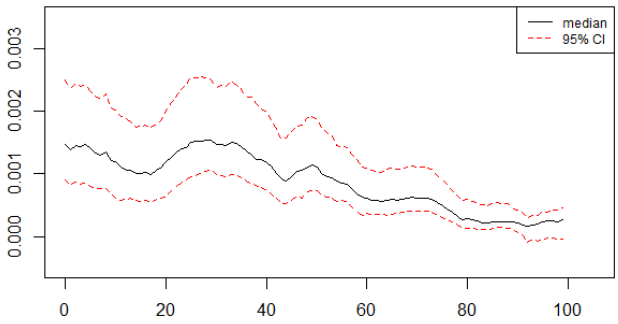
Similarly, Figure 4(b) and 4(c) shows the posterior median and 95% HPD interval of factor β_x and γ_x, respectively. It can be seen that β_x are more sensitive to all time change at children and seniors stage and γ_x is lack of sensitivity to spacial variations with increasing ages.



(a)



(b)



(c)

Fig. 4. Plots of the estimated medians of α_x, β_x and γ_x with their 95% HDP intervals.

4.3.2 Time parameters

In Figure 5, we present the posterior medians and 95% HPD intervals of κ_t. In addition to the years 1975-2004, the follow-up 12-year ahead projections of time effects are also provided via the posterior predictive distributions.To obtain a sample from the posterior predictive distribution of κ_t, we have

$$\kappa_{N+t'}^{[j]} \sim N\left(\varphi_1^{[j]} + \varphi_2^{[j]}(N+t') + \rho^{[j]}[\kappa_{N+t'-1}^{[j]} - \varphi_1^{[j]} - \varphi_2^{[j]}(N+t'-1)], (\sigma_\kappa^2)^{[j]}\right)$$

for iterations $j = 1, 2, \ldots, 10000$ and $t' = 1, 2, \ldots, 12$. From Figure 5, we observe decreasing trends in most of time windows, which might be attributed to the advances in medical improvement and social welfare.

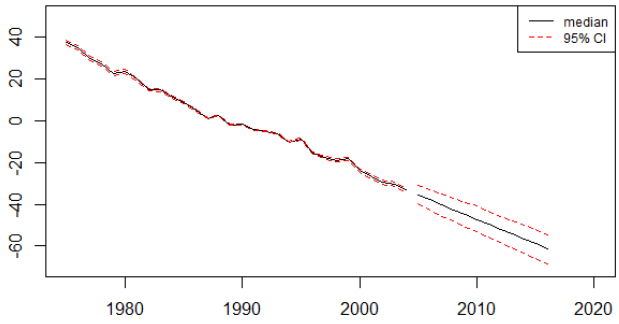


Fig. 5. Plots of the estimated medians of κ_t with their 95% HDP intervals and the 20-years ahead projections of these parameters.

4.4 Model Evaluation

Through this part, we are working on performances of the spatial model, the Poisson Lee-Carter (PLC) model and Poisson log-normal Lee–Carter (PLNLC) model. With a common measurement, we list the square of Pearson residuals under the PLC model for each population given as:

$$r_{x,t}^{(i)2} = \frac{(D_{x,t}^{(i)} - E_{x,t}^{(i)}\hat{\mu}_{x,t}^{(i)})^2}{E_{x,t}^{(i)}\hat{\mu}_{x,t}^{(i)}} \text{ where } \hat{\mu}_{x,t}^{(i)} = \exp(\hat{\alpha}_x^{(i)} + \hat{\beta}_x^{(i)}\hat{\kappa}_t^{(i)})$$

and estiamtes $\hat{\alpha}_x^{(i)}$, $\hat{\beta}_x^{(i)}$ and $\hat{\kappa}_t^{(i)}$ are posterior medians via Bayesian sampling. Here, heat maps of $r_{x,t}^{(i)2}$ are opted to visualize the lack of fit of the PLC model for mortality

data in three Japanese counties, as depicted in Fig 6 (a) to Fig 8 (a).

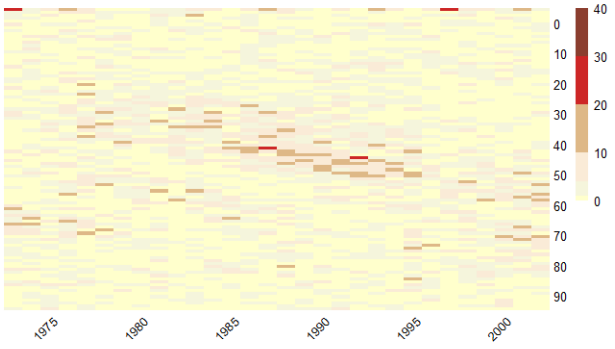
By a considerable amount of dark color cells dispersing around various regions in the heat maps, there is one evidence showing the lack of fit in the PLC model with ignoring the spatial effect.

Similarly, heat maps of the squared Pearson residuals for our models can be constructed. The expression of squared residuals for our spatial model follows

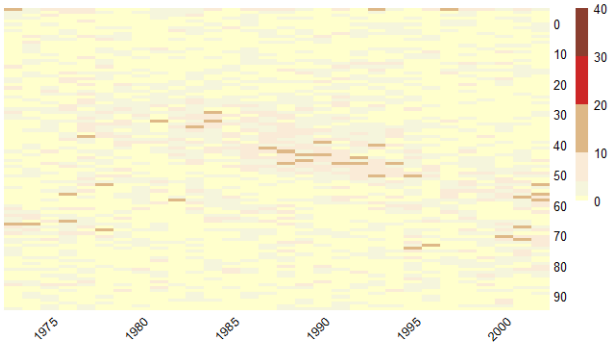
$$r_{x,t}^{(i)2} = \frac{(D_{x,t}^{(i)} - E_{x,t}^{(i)}\hat{\mu}_{x,t}^{(i)})^2}{E_{x,t}^{(i)}\hat{\mu}_{x,t}^{(i)}} \qquad \text{where } \hat{\mu}_{x,t}^{(i)} = \exp(\hat{\alpha}_x + \hat{\beta}_x\hat{\kappa}_t + \hat{\gamma}_x\hat{\theta}^{(i)})$$

and here the posterior medians of the parameters α_x, β_x, γ_x, κ_t and θ_i are plugged into the expression for an estimate.

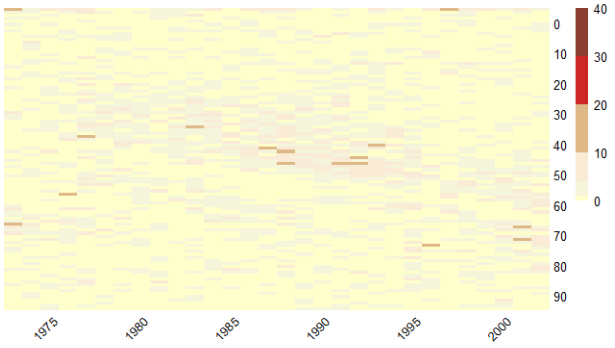
Fig 6 (c) to 8 (c) shows that there exist more lighter color cells in the heat maps of the spatial model than those in the other two models, which implies an overall improvement in goodness of fit totally.



(a)

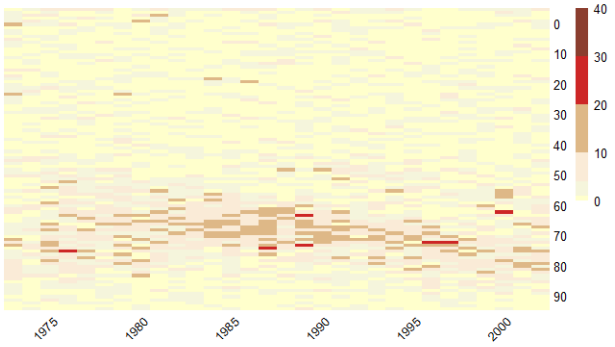


(b)

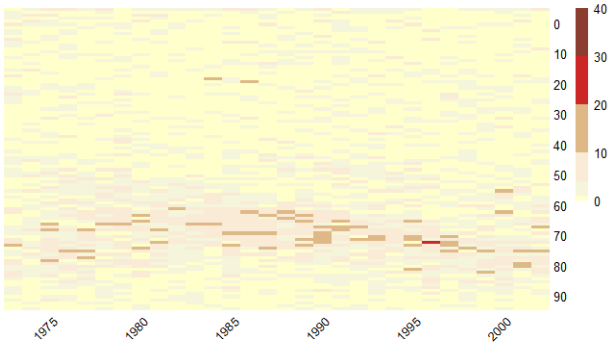


(c)

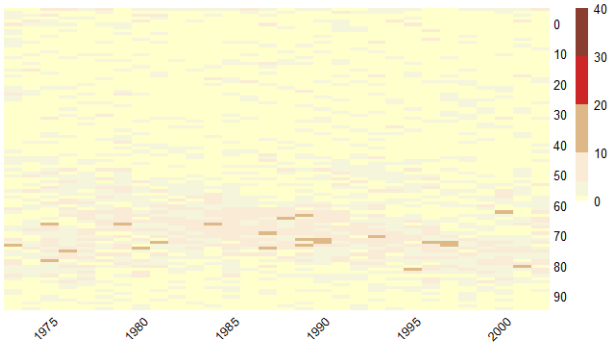
Fig. 6. Heat map of squared Pearson residuals $r_{x,t}^{(i)^2}$ under the PLC model (top), the PLNLC model (middle) and the spatial model (bottom) for Miyagi.



(a)

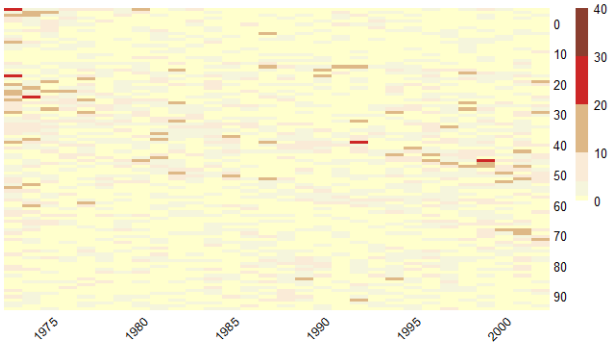


(b)

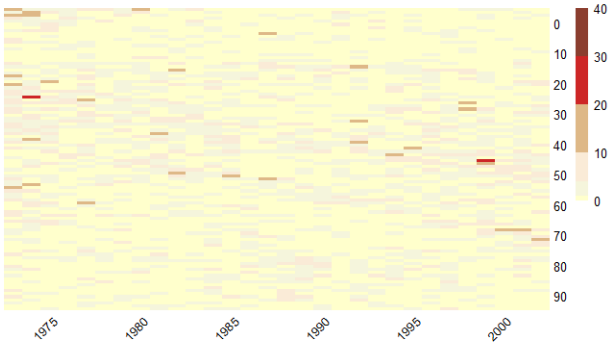


(c)

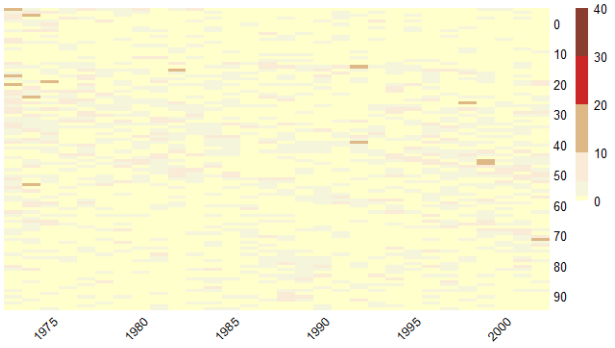
Fig. 7. Heat map of squared Pearson residuals $r_{x,t}^{(i)^2}$ under under the PLC model (top), the PLNLC model (middle) and the spatial model (bottom) for Tokyo.



(a)



(b)



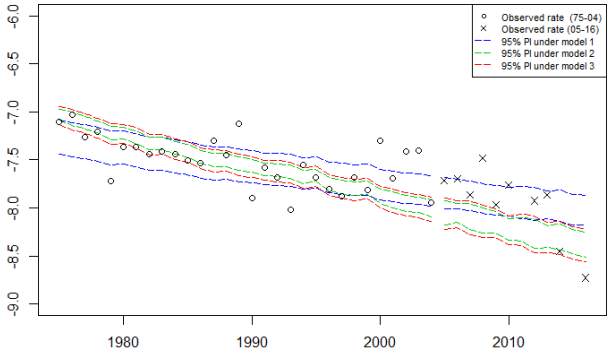
(c)

Fig. 8. Heat map of squared Pearson residuals $r_{x,t}^{(i)2}$ under under the PLC model (top), the PLNLC model (middle) and the spatial model (bottom) for Ehime.

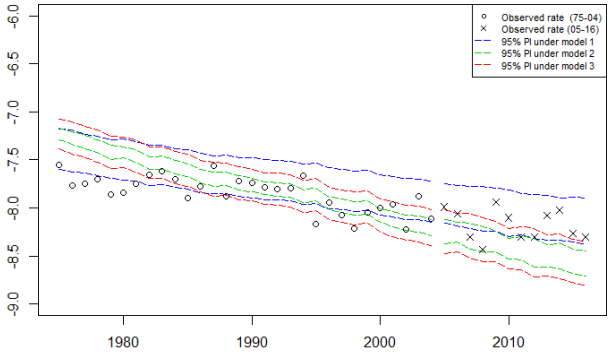
Further, we evaluate our spatial model by the comparison of variability for previous three models. Figures 9 to 11 present the 95% HPD intervals of simulated log mortality rates at three selected ages 20, 50, and 70 for three models, respectively. In addition to the training time window (years 1975-2004), 12-year ahead projections are provided

to assess the prediction ability of our spatial model. For convenience, model 1, model 2 and model 3 denote the spatial model, the PLC model and the PLNLC model in the following discussion.

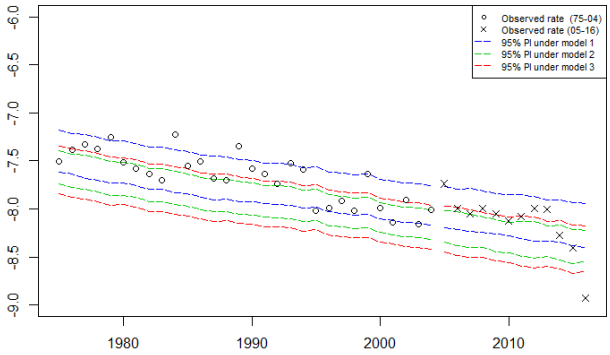
From Figure 9 to 11, we see that almost every observed points of 1975-2004 fall in the 95% prediction intervals (blue curves). Moreover, the spatial model provides solid 12-years ahead projections because validated log rates are down the middle of those blue curves. We also notice that the prediction intervals may properly present the variability of data by introducing additional spatial terms.



(a)

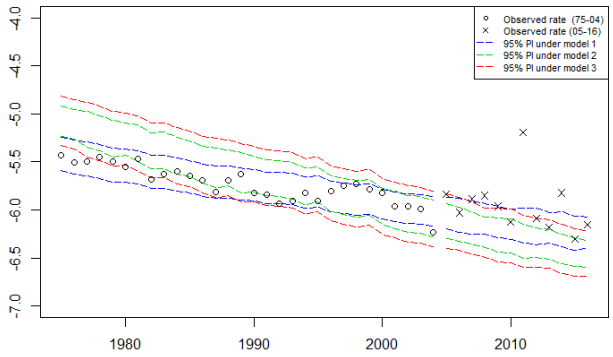


(b)

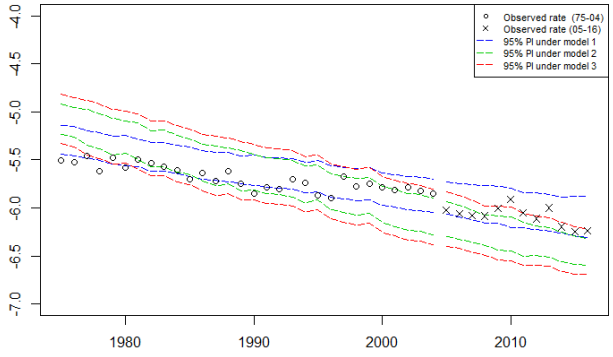


(c)

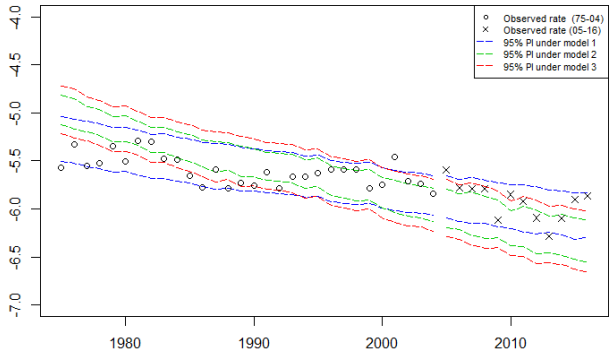
Fig. 9. Plots of the observed log death rates, fitted log death rates, the associated 12-years ahead projection of the log death rates and 95% HDP intervals aged 20 under three models for Miyagi (upper), Tokyo (middle) and Ehime (lower).



(a)

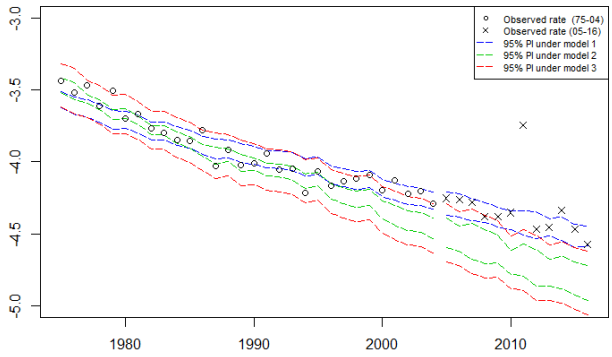


(b)

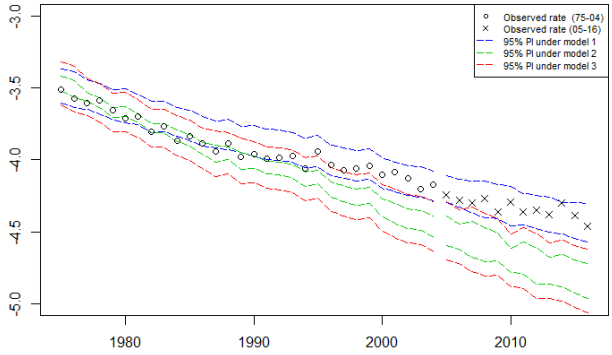


(c)

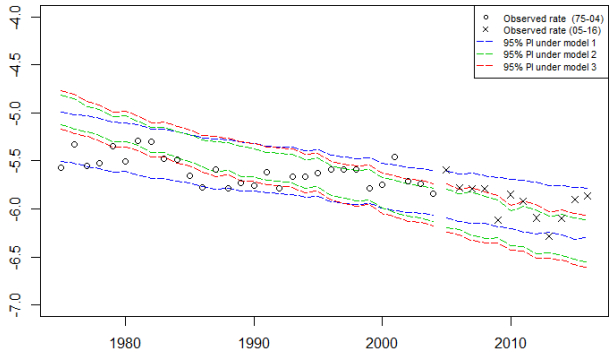
Fig. 10. Plots of the observed log death rates, fitted log death rates, the associated 12-years ahead projection of the log death rates and 95% HDP intervals aged 50 under three models for Miyagi (upper), Tokyo (middle) and Ehime (lower).



(a)



(b)



(c)

Fig. 11. Plots of the observed log death rates, fitted log death rates, the associated 12-years ahead projection of the log death rates and 95% HDP intervals aged 70 under three models for Miyagi (upper), Tokyo (middle) and Ehime (lower).

5 Conclusion

This paper presents a Bayesian approach to estimate and predict mortality for multiple neighbor regions under Poisson log-normal assumption. By illustrations of nationwide cases, we can observe the evidences of spatiotemporal effect at county level in Japan. Incorporating a CAR structure in the PLC model, the new model can simultaneously conduct the observed spatial association and temporal correlation trends, which proposed a model where regional mortality rates estimates can borrow information from each other. As a result, the issue of large uncertainty of mortality rates via direct estimation can be mitigated at specific age groups in some areas.

In our extension of projection modeling, the number of death in each population is treated as conditional independence. Some researchers consider cohort effect models to solve the dependence structure of death tolls among populations (Renshaw and Haberman (2006), Booth and Tickle (2008) , etc). However, even if the dependence between mortality rates by age is not involved in the model, estimated log-mortality rates are mostly in accord with the observed results. Hence, the assumption of conditional independence is reasonable to our proposed approach.

References

Katrien Antonio, Anastasios Bardoutsos, and Wilbert Ouburg. Bayesian Poisson log-bilinear models for mortality projections with multiple populations. *European Actuarial Journal*, 5(2):245–281, 2015.

N Miklós Arató, Ian L Dryden, and Charles C Taylor. Hierarchical Bayesian modelling of spatial age-dependent mortality. *Computational Statistics and & Data Analysis*, 51(2):1347–1363, 2006.

Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data*. CRC press, 2014.

Marie-Pier Bergeron-Boucher, Violetta Simonacci, Jim Oeppen, and Michele Gallo. Coherent modeling and forecasting of mortality patterns for subpopulations using multiway analysis of compositions: an application to canadian provinces and territories. *North American Actuarial Journal*, 22(1):92–118, 2018.

Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974.

Julian Besag, Jeremy York, and Annie Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1):1–20, 1991.

Heather Booth and Leonie Tickle. Mortality modelling and forecasting: A review of methods. *Annals of actuarial science*, 3(1-2):3–43, 2008.

Natacha Brouhns, Michel Denuit, and Jeroen K Vermunt. A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and economics*, 31(3):373–393, 2002.

Andrew JG Cairns, David Blake, and Kevin Dowd. A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718, 2006.

Andrew JG Cairns, David Blake, Kevin Dowd, Guy D Coughlan, and Marwa Khalaf-Allah. Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin: The Journal of the IAA*, 41(1):29–59, 2011.

Peter Congdon. A model for spatial variations in life expectancy; mortality in chinese regions in 2000. *International journal of health geographics*, 6(1):16, 2007.

Claudia Czado, Antoine Delwarde, and Michel Denuit. Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, 36(3):260–284, 2005.

Victor De Oliveira. Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics*, 64(1):107–133, 2012.

Antoine Delwarde, Michel Denuit, and Christian Partrat. Negative binomial version of the Lee-Carter model for mortality forecasting. *Applied Stochastic Models in Business and Industry*, 23(5):385–401, 2007.

Fedele Greco and Francesco Scalone. A space-time extension of the Lee-Carter model in a hierarchical bayesian frame-work: Modelling provincial mortality in ITALY. *Measuring Uncertainty in Population Forecasts: A New Approach.*, page 412, 2013.

Peter D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer Publishing Company, Incorporated, 1st edition, 2009. ISBN 0387922997, 9780387922997.

Rob J Hyndman and Md Shahid Ullah. Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51(10): 4942–4956, 2007.

Leonhard Knorr-Held. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in medicine*, 19(17-18):2555–2567, 2000.

Ronald D Lee and Lawrence R Carter. Modeling and forecasting US mortality. *Journal of the American statistical association*, 87(419):659–671, 1992.

Nan Li and Ronald Lee. Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42(3):575–594, 2005.

Zhen Liu, Xiaoqian Sun, Leping Liu, and Yu-Bo Wang. Bayesian poisson log-normal model with regularized time structure for mortality projection of multi-population. *arXiv preprint arXiv:2010.04775*, 2020.

Claudia Pedroza. A Bayesian forecasting model: predicting us male mortality. *Biostatistics*, 7(4):530–550, 2006.

Roland Rau and Carl Schmertmann. Bayesian modeling of small-area mortality with relational model schedules and spatially varying parameters. In *2017 International Population Conference*. IUSSP, 2017.

Arthur E Renshaw and Steven Haberman. A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and economics*, 38 (3):556–570, 2006.

Jackie ST Wong, Jonathan J Forster, and Peter WF Smith. Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics and Economics*, 83:206–221, 2018.