

# A Bayesian Spatio-Temporal Approach to Predict Mortality Rates

Luiz Fernando Figueiredo

Supervisors: Viviana G R Lobo and João B M Pereira

## 1 Introduction

Measuring mortality across populations, namely to construct life tables, is common practice in various fields, such as demography (to inform public policy decisions), medicine (to measure and predict the spread and impact of diseases), and actuarial science (to support pensions and portfolio calculations). Modelling mortality rates provide further insights into the interested population through the inferential process, capturing mortality trends and allowing for future behaviour predictions.

Recently, several approaches based on stochastic methods have been employed to understand population behaviour better. The Lee-Carter model ([Lee and Carter, 1992](#)) was one of the pioneering stochastic mortality models and has remained widely applied in the field. This innovative work introduced a method for analyzing mortality incorporating age and time-specific parameters to capture temporal dynamics within a given population. The model relies on a factor structure with a latent factor that evolves over time (a state parameter). Over the years, several adaptations of the Lee-Carter model have been suggested: [Lee \(2000\)](#) reunited various extensions to the original Lee-Carter model, [Pedroza \(2002\)](#) implemented a state-space formulation, [Renshaw and Haberman \(2003\)](#) proposed a generalized regression model based on heteroskedastic Poisson error structures, [\(Li and Lee, 2005\)](#) extended the Lee-Carter model to multiple populations for a coherent mortality forecast. The goal is to improve predictions of future mortality by including similar patterns for the group in the model. For comparison of these models, [Cairns et al. \(2009\)](#) ranked stochastic mortality models via Bayes Information Criterion (BIC), highlighting some of the Lee-Carter extensions and their own Cairns-Blake-Dowd (CBD) stochastic model.

In a Bayesian approach, [Pedroza \(2006\)](#) proposed the Bayesian Lee-Carter, incorporating prior information into the framework and using Markov chain Monte Carlo (MCMC) methods to sample from the predictive posterior distribution. They also demonstrate how the Bayesian approach facilitates handling missing data and offers multiple extensions to the model. The Bayesian Lee-Carter model is available for implementation through `BayesMortalityPlus` package in R ([BayesMortalityPlus, 2024](#)). Similarly, [Czado et al. \(2005\)](#) considered the Bayesian Poisson log-bilinear model, an alternative to the usual Lee-Carter model, by incorporating a Poisson response for deaths. More recently, [Wong et al. \(2018\)](#) proposed two bayesian models to account for overdispersion, Poisson log-normal Lee-Carter (PLNLC) and Negative binomial Lee-Carter (NBLC), and compared them with Poisson approaches to Lee-Carter that imposes mean-variance equality.

Although significant advances have been proposed in mortality rate modelling in recent years, ranging from improved data collection and filtering to state-of-the-art stochastic

models, smaller units, such as counties or states, still suffer from missing data due to smaller population sizes and/or a limited data reporting resources.

In this context, spatial stochastic models are designed to incorporate correlation between neighbouring areas, thus allowing more data to be included in the modelling process. This approach is based on the idea that neighbouring counties, locations, or points of interest in a data grid can provide information about the areas they are close to, addressing missing data problems and capturing the dependencies between neighbouring units. This is important because data in one region can often be influenced by the data in neighbouring regions, and these spatial dependencies must be properly modelled to make more accurate and realistic predictions.

A well-known method in the literature for modelling spatial correlation in areal data is the conditional autoregressive (CAR) model ([Besag, 1974](#)), which incorporates spatial dependencies through full conditional distributions based on neighbouring locations. This approach enables the inclusion of spatial parameters in the model, helping to capture spatial correlations in the data. Several studies have successfully applied the method, such as [Gómez-Rubio et al. \(2019\)](#), which consider a spatio-temporal model for the joint analysis of multiple diseases through three different causes of death registered in Spain at the provincial level, [Gibbs et al. \(2020\)](#), which present a spatial approach to modelling county-level mortality rates in the United States using dynamic linear models, and [Liu et al. \(2021\)](#), proposing a bayesian Lee-Carter spatial extension with Poisson framework to fit data from Japan. All studies utilize CAR structures to account for spatial dependencies in the model.

In this work, we aim to extend the Bayesian Lee-Carter approach proposed by [Pedroza \(2006\)](#) by incorporating spatial information to model the spatial dependencies between neighbouring areas, thus improving the accuracy of mortality predictions and analyse changes in mortality probabilities considering a stochastic model which models log mortality dynamics at age  $x$ , region  $s$  and times  $t$ . For this, we consider two datasets in this analysis: the first is from DataSUS (available at [DATASUS, 2024](#)), containing demographic and mortality information for the state of Rio de Janeiro at microregion level, covering the years 1980 to 2021. The second one is from Japan ([Japanese Mortality Database, 2024](#)), a more robust dataset used in [Liu et al. \(2021\)](#) with information for Japan's prefectures from 1947-2022. We explore mortality rates by age groups, year, and county of residence.

We introduce a pair of spatial-related parameters to capture the correlation between neighbouring areas via CAR specification ([Banerjee et al., 2014](#)) while maintaining the Log-Normal response originally proposed in [Lee and Carter \(1992\)](#). Practically, this means that we can still take advantage of the Kalman Filter ([Kalman, 1960](#)) to specify the temporal parameter and all of the extensions proposed by [Pedroza \(2006\)](#) could be applied to the model with no hassle (such as missing data treatment).

The remainder of the document is organized as follows: Section 2 introduces forecast mortality models and builds the Bayesian spatial Lee-Carter model, Section 3 details the inference process of the proposed model, Section 4 presents some exploratory analysis and fits the model to both datasets presented before and Section 5 collects partial conclusions from the application, as well as future work to be done.

## 2 Forecasting Mortality Models

In this section, we introduce the original Lee-Carter model along with its state-space

extension, which forms the foundation for developing the proposed spatial Lee-Carter model.

## 2.1 The basic Lee-Carter model

Consider  $x = 1, 2, \dots, p$  as the age index and  $t = 1, 2, \dots, T$  as the year index of the mortality information in a chosen dataset. Let  $E_{x,t}$  and  $D_{x,t}$  denote the exposure to risk and the death counts at age  $x$  and year  $t$ , respectively. We define the central death rate as  $m_{x,t} = \frac{D_{x,t}}{E_{x,t}}$  for age  $x$  and year  $t$  (Bowers, 1986).

The Lee-Carter model (Lee and Carter, 1992) models the central death rates as:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \quad (1)$$

where  $\alpha_x$  and  $\beta_x$  are age-specific parameters and  $\kappa_t$  is a time-varying index representing the level of mortality. The parameter  $\alpha_x$  captures the general age pattern of mortality, while  $\beta_x$ , also known as the improvement parameter, indicates how rapidly mortality rates decline for different age groups in response to changes in the time index  $\kappa_t$ . The error term  $\epsilon_{x,t}$ , with mean 0 and variance  $\sigma_\epsilon^2$ , reflects age-specific historical influences not captured by the model and are assumed to be independent. To ensure model identification, some constraints have been applied:  $\beta_x$  to sum to unity and  $\kappa_t$  to sum to zero. This implies that the  $\alpha_x$  are simply the averages over time of the log-mortality rates. The estimation process uses singular value decomposition to find a least squares solution, with  $\kappa_t$  reestimated using Box-Jenkins methodology.

As presented by Pedroza (2002), the Lee-Carter model can be written as a state-space model for  $p$  age groups (or single ages). For this, let  $\mathbf{y}_t = (\ln(m_{1,t}), \ln(m_{2,t}), \dots, \ln(m_{p,t}))'$  be a vector of log mortality rates at year  $t$ . Then,

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\alpha} + \boldsymbol{\beta} \kappa_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N_p(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}), \\ \kappa_t &= \kappa_{t-1} + \eta + \omega_t, \quad \omega_t \stackrel{iid}{\sim} N(0, \sigma_\omega^2), \end{aligned} \quad (2)$$

where  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_p)'$  and  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$  are age-specific vectors of parameters  $\alpha_x$  and  $\beta_x$ , respectively,  $\mathbf{I}$  is a  $p \times p$  identity matrix,  $\boldsymbol{\epsilon}_t$  and  $\omega_t$  are assumed to be independent,  $\mathbf{y}_t$  are independent and identically distributed (i.i.d.) with common variance  $\sigma_\epsilon^2$  and a random walk with drift  $\eta$  is assumed for the state vector. Pedroza's Bayesian approach specifies the Lee-Carter model as a Dynamic Linear Model (West and Harrison, 1997), considering Forward Filtering Backward Sampling (FFBS) (Carter and Kohn, 1994, Frühwirth-Schnatter, 1994) for state parameter estimation and Gibbs sampler to produce inference on the remaining parameters and predictive posterior distribution. We adopt the same approach, which will be discussed in Section 3.

## 2.2 Spatial Lee-Carter model

There is limited literature available on spatial modelling, particularly concerning state-space models. The approach we adopt is largely based on the work of Waller et al. (1997) and Liu et al. (2021), where a spatial parameter is introduced to the model using a CAR structure alongside an age-specific parameter to account for age-related spatial effects while maintaining the structure outlined in the equation (2). Let  $\mathbf{y}_{s,t} =$

$(y_{1,s,t}, y_{2,s,t}, \dots, y_{p,s,t})'$  denote the log mortality rate at year  $t$  and region  $s$ , where  $s = 1, 2, \dots, S$ . We propose the Spatial Bayesian Lee-Carter model (SBLC) following:

$$\begin{aligned} \mathbf{y}_{s,t} &= \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\gamma}\theta_s + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N_p(\mathbf{0}, \boldsymbol{\Sigma}), \\ \kappa_t &= \kappa_{t-1} + \eta + \omega_t, \quad \omega_t \stackrel{iid}{\sim} N(0, \sigma_\omega^2), \end{aligned} \quad (3)$$

with  $\theta_s$  measuring the spatial effects with respect to the region,  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_p)'$  accounting for the age-spatial effects and  $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$  a vector with different  $\sigma_x^2$  variance parameters for each age  $x$ . Both the spatial parameter and different variances per age are employed to grant more flexibility to the model. Notice that since the modelling structure presented by equation (2) is preserved, model extensions and missing data treatment presented in [Pedroza \(2006\)](#) can be employed with a simple adaptation.

To accommodate spatial correlation between neighbouring regions, we consider an Intrinsically Autoregressive (IAR) structure to parameter  $\boldsymbol{\theta}$ . According to [Banerjee et al. \(2014\)](#) and [Gibbs et al. \(2020\)](#), let  $\mathbf{W}$  be our proximity matrix, where  $w_{ij} = 1$  if  $i$  is neighbour to  $j$  and 0, otherwise. We can write the full conditional distributions of each  $\theta_i$  as:

$$\theta_i | \theta_{-i} \sim N \left( \sum_{j \neq i} \frac{1}{w_{i+}} w_{ij} \theta_j, \frac{\sigma_\theta^2}{w_{i+}} \right), \quad (4)$$

where  $w_{i+} = \sum_j w_{ij}$  is the total number of neighbours for location  $i$ .

These full conditional distributions are compatible, resulting in a joint distribution for the random vector  $\boldsymbol{\theta}$  in the form of:

$$p(\theta_1, \theta_2, \dots, \theta_S) \propto \exp \left\{ -\frac{w_{i+}}{2\sigma_\theta^2} \boldsymbol{\theta}' (\mathbf{M} - \mathbf{W}) \boldsymbol{\theta} \right\}, \quad (5)$$

where  $\mathbf{M} = \text{diag}(w_{1+}, w_{2+}, \dots, w_{S+})$  is a diagonal matrix that holds the total number of neighbours for each location

## 2.3 Constraints

Parameter constraints are crucial aspects of the Lee-Carter model, since they guarantee identifiability for the parameters to be estimated and also affect their interpretation and comparability. Therefore, it is important to clearly specify the constraints applied to each parameter.

To ensure model identification for the SBLC model, we apply the constraints seen in [Lee and Carter \(1992\)](#) to the spatial parameters  $\sum_x \gamma_x = 1$  and  $\sum_s \theta_s = 0$ , transforming the remaining parameters on averages throughout the spatial grid and guarantees the needed centering to  $\boldsymbol{\theta}$  for the IAR structure to ensure parameter identification.

We apply these constraints as a block-step inside the estimation algorithm, transforming the parameters as follows

$$\begin{aligned} \boldsymbol{\beta}^* &\leftarrow \frac{\boldsymbol{\beta}}{\xi}, \quad \xi = \frac{\kappa_1 - \kappa_T}{T-1}, & \kappa_t^* &\leftarrow \kappa_t - \bar{\kappa}, \quad \bar{\kappa} = \frac{1}{T} \sum_t^T \kappa_t, \\ \boldsymbol{\gamma}^* &\leftarrow \frac{\boldsymbol{\gamma}}{\gamma_+}, \quad \gamma_+ = \sum_x \gamma_x, & \theta_s^* &\leftarrow \theta_s - \bar{\theta}, \quad \bar{\theta} = \frac{1}{m} \sum_s^m \theta_s. \end{aligned}$$

As seen in Czado et al. (2005), these transformations reflect on other parameters as

$$\boldsymbol{\alpha}^* \leftarrow \boldsymbol{\alpha} + \boldsymbol{\beta}\bar{\kappa} + \boldsymbol{\gamma}\bar{\theta}, \quad \kappa_t^* \leftarrow \kappa_t^*\xi, \quad \theta_s^* \leftarrow \theta_s^*\gamma_+.$$

Then, the model equation leading to the response parameter  $\mathbf{y}_{s,t}$  remains unchanged as

$$\begin{aligned} \mathbf{y}_{s,t} &= \boldsymbol{\alpha}^* + \boldsymbol{\beta}^*\kappa_t^* + \boldsymbol{\gamma}^*\theta_s^* \\ &= (\boldsymbol{\alpha} + \boldsymbol{\beta}\bar{\kappa} + \boldsymbol{\gamma}\bar{\theta}) + \frac{\boldsymbol{\beta}}{\xi}(\kappa_t - \bar{\kappa})\xi + \frac{\boldsymbol{\gamma}}{\gamma_+}(\theta_s - \bar{\theta})\gamma_+ \\ &= \boldsymbol{\alpha} + (\boldsymbol{\beta}\bar{\kappa} - \boldsymbol{\beta}\bar{\kappa}) + (\boldsymbol{\gamma}\bar{\theta} - \boldsymbol{\gamma}\bar{\theta}) + \boldsymbol{\beta}\kappa_t + \boldsymbol{\gamma}\theta_s \\ &= \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\gamma}\theta_s. \end{aligned}$$

### 3 Inference Procedure

The inference procedure for the model described by equation 3 is similar to Pedroza (2006), with a few changes to the FFBS and the Gibbs algorithm due to the addition of spatial parameters. Consider  $\Phi = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \eta, \Sigma, \sigma_\omega^2, \sigma_\theta^2)$  to be a parameter collection of age-specific parameters, variances and drift, the dynamical parameter  $\boldsymbol{\kappa}_{1:T}$ , where  $\boldsymbol{\kappa}_{1:T} = (\kappa_1, \kappa_2, \dots, \kappa_T)'$  and spatial vector parameter  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_S)'$  to be estimated. The likelihood function of the proposed model can be written as:

$$\begin{aligned} \mathcal{L}(\mathbf{y} \mid \Phi, \boldsymbol{\kappa}_{1:T}, \boldsymbol{\theta}) &= \prod_{x=1}^p \prod_{t=1}^T \prod_{s=1}^S \mathcal{N}(y_{x,s,t}; \alpha_x + \beta_x \kappa_t + \gamma_x \theta_s, \sigma_x^2) \\ &\propto \prod_{x=1}^p (1/\sigma_x^2)^{\frac{T_S}{2}} \exp \left( -\frac{1}{2\sigma_x^2} \sum_{t=1}^T \sum_{s=1}^S (y_{x,s,t} - \alpha_x - \beta_x \kappa_t - \gamma_x \theta_s)^2 \right), \end{aligned} \quad (6)$$

where  $\mathcal{N}(\cdot; A, B)$  is a Gaussian distribution with mean  $A$  and variance  $B$ .

The choice of prior distributions for all parameters in  $\Phi$  considers vague priors and conjugate distributions, ensuring that the full conditional distributions are suitable for use in a Gibbs sampler, while also maintaining comparability with the basic Lee-Carter model. We are assuming that the priors are independent and assign  $p(\boldsymbol{\alpha}) \propto 1$ ,  $p(\boldsymbol{\beta}) \propto 1$ ,  $p(\boldsymbol{\gamma}) \propto 1$ ,  $p(\eta) \propto 1$ ,  $p(\sigma_x^2) \propto 1/\sigma_x^2$ ,  $p(\sigma_\omega^2) \propto 1/\sigma_\omega$  and  $p(\sigma_\theta^2) \propto 1/\sigma_\theta$ . Prior distribution for the starting point  $\kappa_0 \sim N(m_0, C_0)$  with known  $m_0$  and  $C_0$ . (More details on Pedroza, 2006)

Banerjee et al. (2014) points out that the joint distribution (5) is improper, however, the full conditional distributions required for the Gibbs sampler are all proper. There are methods employed to the CAR structure that result in a proper joint distribution for  $p(\theta_1, \theta_2, \dots, \theta_S)$ , however, there are pros and cons with this approach (more details of this discussion on Besag and Kooperberg, 1995) and the decision is ultimately left to the analyst.

### 3.1 Posterior distribution

Following the Bayes' theorem and assuming mutual independence between parameters, the posterior distribution of  $(\Phi, \kappa_{1:T}, \theta)$  given the log mortality rates  $\mathbf{y}$  is described by

$$\begin{aligned} p(\Phi, \kappa_{1:T}, \theta | \mathbf{y}) &\propto \prod_{x=1}^p \prod_{t=1}^T \prod_{s=1}^S \mathcal{N}(y_{x,s,t}; \alpha_x + \beta_x \kappa_t + \gamma_x \theta_s, \sigma_x^2) \\ &\quad \times \mathcal{N}(\kappa_1; a_1, R_1) \times \prod_{t=2}^T \mathcal{N}(\kappa_t; \eta + \kappa_{t-1}, \sigma_\omega^2) \\ &\quad \times \exp \left\{ -\frac{w_{i+}}{2\sigma_\theta^2} \boldsymbol{\theta}' (\mathbf{M} - \mathbf{W}) \boldsymbol{\theta} \right\} \times \pi(\Phi). \end{aligned} \quad (7)$$

where  $a_1 = \eta + m_0$  and  $R_1 = C_0 + \sigma_\omega^2$  and  $\pi(\cdot)$  stands for all prior distributions for the parameters. As seen in [Pedroza \(2006\)](#), analytical results may be too complicated for (7), requiring Markov Chain Monte Carlo (MCMC) methods to sample from it ([Gamerman and Lopes, 2006](#)). Specifically, we use the Gibbs sampler algorithm, where  $\kappa$  is sampled via FFBS, and the elements of  $\theta$  are drawn from their full conditional distributions, which are easily obtained due to the structure of the IAR models. We can summarize the estimation algorithm as:

Step 1: Sample dynamic parameter  $\kappa_{1:T}$  via FFBS detailed in subsection [3.1.1](#);

Step 2: Sample  $\alpha, \beta, \gamma, \eta, \Sigma, \sigma_\omega^2, \sigma_\theta^2$  and  $\theta_s$  from each of their posterior full conditional distributions via Gibbs sampler;

#### 3.1.1 Forward Filtering Backward Sampling

Following [Pedroza \(2006\)](#), the Forward Filtering Backward Sampling (FFBS) algorithm is applied to the estimation of parameter  $\kappa_t$ .

To incorporate the spatial data to the filtering recursions, we can rewrite the model equation stacking the log mortality rates to each region as such:

$$\mathbf{E}[\mathbf{Y}_t] = \mathbf{A}\alpha + \mathbf{A}\beta\kappa_t + (\boldsymbol{\gamma} \otimes \boldsymbol{\theta}), \quad (8)$$

where  $\mathbf{Y}_t$  is a stacked log-mortality matrix with dimensions  $pS \times 1$ ,  $\mathbf{A}$  a  $pS \times p$  stacking matrix for age related parameters and  $(\boldsymbol{\gamma} \otimes \boldsymbol{\theta})$  representing the  $pS \times 1$  kronecker matrix product between the age-vector  $\boldsymbol{\gamma}$  and spatial vector  $\boldsymbol{\theta}$ . The stacked quantities can be written as:

$$\mathbf{Y}_t = \begin{bmatrix} y_{1,t,1} \\ \dots \\ y_{S,t,1} \\ y_{1,t,2} \\ \dots \\ y_{S,t,p} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}, \quad (\boldsymbol{\gamma} \otimes \boldsymbol{\theta}) = \begin{bmatrix} \gamma_1 \theta_1 \\ \dots \\ \gamma_1 \theta_S \\ \gamma_2 \theta_1 \\ \dots \\ \gamma_n \theta_S \end{bmatrix}. \quad (9)$$

The latent equation describing  $\kappa_t$  remains the same.

With our new specification, assuming that given initial prior information  $D_0$  at  $t = 0$ , at any future time  $t$  the available information is  $D_t = \{\mathbf{Y}_t, D_{t-1}\}$  and  $(\kappa_{t-1} | D_{t-1}) \sim N(m_{t-1}, C_{t-1})$  for each  $t = 1, 2, \dots, T$ , we can compute the Kalman filter quantities, following [West and Harrison \(1997\)](#):

1. The 1-step ahead prior density is  $(\kappa_t | D_{t-1}) \sim N(a_t, R_t)$ , where

$$\begin{aligned} a_t &= \mathbf{E}[\kappa_t | D_{t-1}] = \mathbf{E}[\kappa_{t-1} + \eta + \omega_t | D_{t-1}] = \eta + m_{t-1}, \\ R_t &= \mathbf{V}[\kappa_t | D_{t-1}] = \mathbf{V}[\kappa_{t-1} + \eta + \omega_t | D_{t-1}] = C_{t-1} + \sigma_\omega^2. \end{aligned} \quad (10)$$

2. The 1-step ahead predictive density is  $(\mathbf{Y}_t | D_{t-1}) \sim N(\mathbf{f}_t, \mathbf{Q}_t)$ , where

$$\begin{aligned} \mathbf{f}_t &= \mathbf{E}[\mathbf{Y}_t | D_{t-1}] = \mathbf{A}\boldsymbol{\alpha} + \mathbf{A}\boldsymbol{\beta}a_t + (\boldsymbol{\gamma} \otimes \boldsymbol{\theta}), \\ \mathbf{Q}_t &= \mathbf{V}[\mathbf{Y}_t | D_{t-1}] = \mathbf{A}\boldsymbol{\beta}(R_t)\boldsymbol{\beta}'\mathbf{A}' + (\mathbf{I}_{pS \times pS} \otimes \boldsymbol{\Sigma}). \end{aligned} \quad (11)$$

3. Finally, the posterior density is  $(\kappa_t | D_t) \sim N(m_t, C_t)$ , where

$$\begin{aligned} m_t &= a_t + R_t\boldsymbol{\beta}'\mathbf{A}'\mathbf{Q}_t^{-1}(\mathbf{Y}_t - \mathbf{f}_t), \\ C_t &= R_t - R_t\boldsymbol{\beta}'\mathbf{A}'\mathbf{Q}_t^{-1}\mathbf{A}\boldsymbol{\beta}(R_t)R_t. \end{aligned} \quad (12)$$

The Kalman filter, as described above, corresponds to the Forward Filtering step. The second part of the algorithm, Backward Sampling, involves sampling  $\boldsymbol{\kappa}_{1:T}$  from the distribution  $p(\boldsymbol{\kappa}_{1:T} | D_T)$ . This is done sequentially, for  $t = T - 1, \dots, 1$  by sampling each  $\kappa_t$  from  $N(h_t, H_t)$ , where  $h_t = m_t + B_t(\kappa_{t+1} - a_{t+1})$ ,  $H_t = C_t - B_tR_{t+1}B_t'$  and  $B_t = C_tR_{t+1}^{-1}$ .

For prediction, we compute the  $k$ -steps ahead distribution. For each time  $t$  and  $k > 0$ , distributions for  $\kappa_{t+k}$  and  $\mathbf{Y}_{t+k}$  given  $D_t$  are  $(\kappa_{t+k} | D_t) \sim \mathcal{N}(a_t(k), R_t(k))$  and  $(\mathbf{Y}_{t+k} | D_t) \sim \mathcal{N}(\mathbf{f}_t(k), \mathbf{Q}_t(k))$ , where  $a_t(0) = m_t$  and  $R_t(0) = C_t$ . Then, sample from  $(\mathbf{Y}_{t+k} | D_t)$  to obtain prediction mortality estimates.

## 4 Application

In this work, we aim to capture mortality trends across the entire spatial grid and perform model diagnostics to assess whether the spatial model offers improvements over the marginal (no spatial information) Lee-Carter model. We will fit both models considering two datasets: (i) mortality data from Brazilian state of Rio de Janeiro (available via tabnet platform [DATASUS, 2024](#)), see section 4.1) and (ii) mortality data from Japan (available on [Japanese Mortality Database, 2024](#), see section 4.2).

Convergence diagnostics were performed for all posterior chains, but the detailed diagnostics plot and metrics for each parameter of the proposed model will be omitted throughout the paper.

### 4.1 Dataset 1: Mortality in Rio de Janeiro

In this study, we consider an analysis of the mortality dataset for the state of Rio de Janeiro, represented by the IBGE microregions grid, which consists of 18 microregions across the entire state (see Figure 1). We have exposure count data from 1980 to 2021, which will be used for analysis from 1980 to 2014, with 2015 to 2019 left out for prediction validation. The age groups range from 20-24 to 80+ totaling in 13 age groups. The COVID-19 pandemic years 2020-2021 will be excluded from the current analysis but will be examined in future studies. Note that DATASUS classifies mortality information within two categories: "deaths per occurrence", which allocates death notifications based on where it happened within the spatial grid, and "deaths per residency", which allocates death notifications based on where the victim was born in the spatial grid. To carry out

this study, it is chosen "deaths per residency" category to mitigate the effect of mortality inflation due to persons seeking out treatment in more developed microregions.

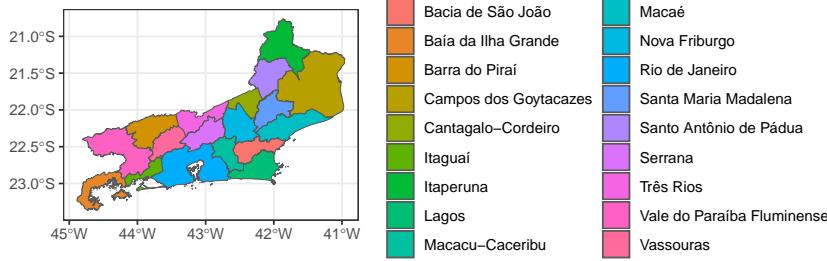


Figure 1: Dataset 1: Map of the IBGE Microregions for the state of Rio de Janeiro.

Figure 2 shows the mortality trends for Rio de Janeiro in the years 1980, 2000, and 2019 for three different age groups. We can observe an overall improvement pattern over the years (except for the Três Rios microregion in 2019 for the 25-29 age group, where it appears that external causes caused a spike in the mortality rate). Additionally, spatial correlation between microregions is evident, with neighboring regions exhibiting similar mortality patterns.

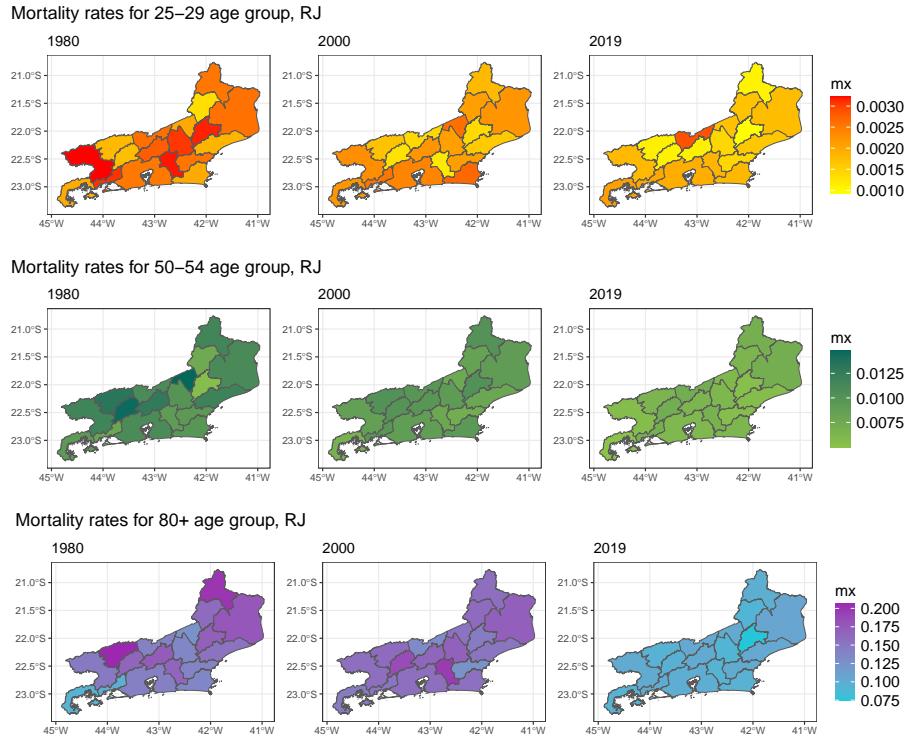


Figure 2: Data: Heat map of mortality rates throughout spatial grid. Rio de Janeiro, years 1980-2000-2019 and age groups 25-29, 50-54 and 80+.

The parameters' estimation for the SBLC model after fitting data for Rio de Janeiro can be found in Figure 3. Remember that, parameters  $\alpha_x$ ,  $\beta_x$ , and  $\gamma_x$  define mortality behaviors for the age groups, with  $\alpha_x$  offering a generalized mortality curve and  $\beta_x$  and

$\gamma_x$  measuring a general improvement and strength of spatial effects, respectively. Here, we can see the improvement captured by the model for each age group, aligned with the result from Figure 2. The local-age parameter  $\gamma_x$  also indicates that the spatial effects are less relevant for older age groups, except for the 50-54 to 65-69 age groups, where spatial factors appear to have a significant impact on their mortality rates. The spatial parameter  $\theta_s$  shows the impact of each microregion in Rio de Janeiro on mortality, with Santa Maria Madalena presenting the lowest mortality rates and Serrana, Vassouras and Itaguaí having the highest mortality rates of all microregions. The parameter  $\kappa_t$  follows its usual pattern of decreasing mortality rates over the years.

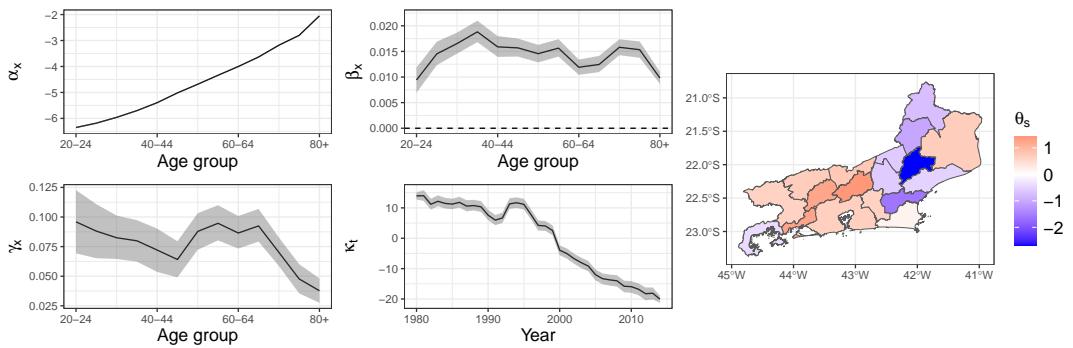


Figure 3: Posterior summaries: Medians with 95% credible interval for parameters  $(\alpha_x, \beta_x, \gamma_x, \kappa_t)$  and median for spatial parameter  $(\theta_s)$  of the fitted SBLC model for Rio de Janeiro.

We would like to evaluate the performance of our proposed model in terms of goodness-of-fit and predictive capabilities for the predicted mortality rates. For this, we consider two models for comparison: the usual Lee-Carter model (BLC), which considers a single population, and the Spatial Lee-Carter model (SBLC). The evaluation will be done visually, comparing the fitted and predicted log mortality rates from both models against the observed and left out data.

Figure 4 shows the fitted and predicted log mortality rates for Rio de Janeiro, Itaperuna and Santa Maria Madalena microregions for some age groups. The black cross represents the holdout data. The population of each microregion is 12,489,824 for Rio de Janeiro, 209,926 for Itaperuna, and 30,449 for Santa Maria Madalena, according to the 2021 estimates from DataSUS.

We observe that the SBLC model lacks the flexibility of the BLC (marginal) model, failing to capture the mortality level across age groups (controlled by parameter  $\alpha$ ) and producing poor predictions over time ( $\kappa_t$ ). It is worth noting that the increasing mortality rate observed in the holdout data for the 25-29 age group in the Rio de Janeiro microregion could not be predicted by either model. This discrepancy becomes more pronounced with the volume of data: the Rio de Janeiro microregion shows the poorest fit of the three regions, followed by Itaperuna and Santa Maria Madalena. Predictive performance is similarly affected by the poor fit, with inflated credible intervals for the Rio de Janeiro microregion and narrower intervals for Santa Maria Madalena.

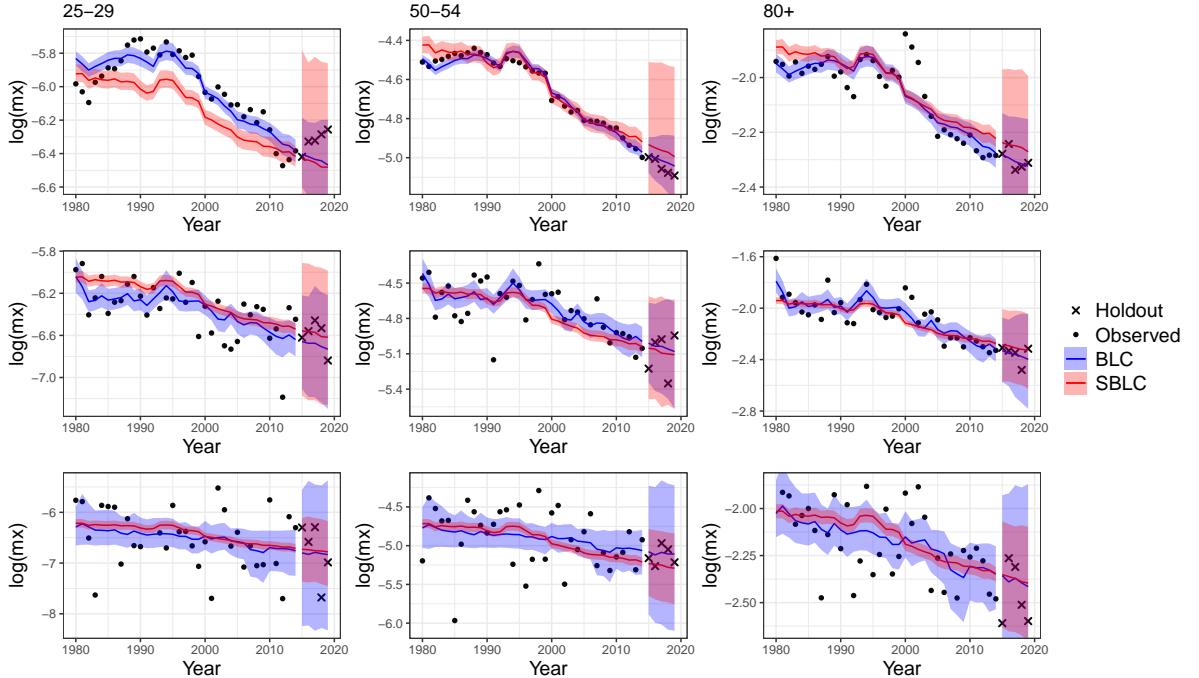


Figure 4: Posterior summaries: fitted log mortality rates of Rio de Janeiro microregions Rio de Janeiro (top row), Itaperuna (middle row) and Santa Maria Madalena (bottom row) for BLC (blue) and SBLC (red) models. Age groups 25-29, 50-54 and 80+ (columns) and years 1980-2014 with prediction for 2015-2019.

This shows us that the SBLC's poor fit could be due to the model parameters static nature and heterogeneity in the data collected, with  $\alpha$  and  $\kappa_t$  being forced to average the spatial particularities of each microregion. The odd behaviour for the predictive intervals also indicates that denoting the variances  $\Sigma$  as age-specific was not enough flexibility for the data. Overall, the SBLC model performance for this particular dataset was below standards.

## 4.2 Dataset 2: Mortality in Japan

The dataset for Japan is presented by its 45 prefectures with 43 available for fitting, as two prefectures are islands with no neighbours to incorporate into the model. The data spans from 1947 to 2022 and covers ages 0 to 110. To compare the performance of the fit for the Japan data with that of Rio de Janeiro, we gathered a smaller sample with the same dimensions as observed in the study for dataset 1. We considered a temporal window of 35 years (1980–2014, with 2015–2019 as the holdout period) and the 13 age groups. For Japan's dataset, the category '80-84' was treated as equivalent to the category '80+' for the sake of comparison. This comparison will help us understand the impact that data quality may have on the SBLC model.

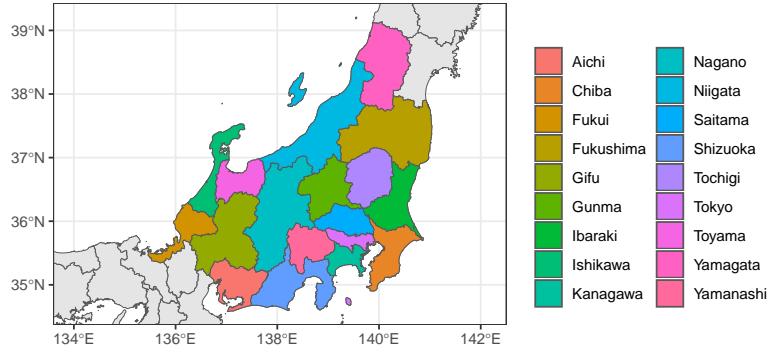


Figure 5: Data: Japan prefectures, with included regions colored.

Figure 6 shows the mortality behaviour for the selected prefectures of Japan in the years 1980, 2000, and 2019, across three different age groups. The mortality improvement appears to be more pronounced than in the Rio de Janeiro data, with mortality rates for the 80+ age group changing drastically and in unison. Spatial correlation can also be observed, with some neighbouring prefectures displaying similar mortality patterns over the years.

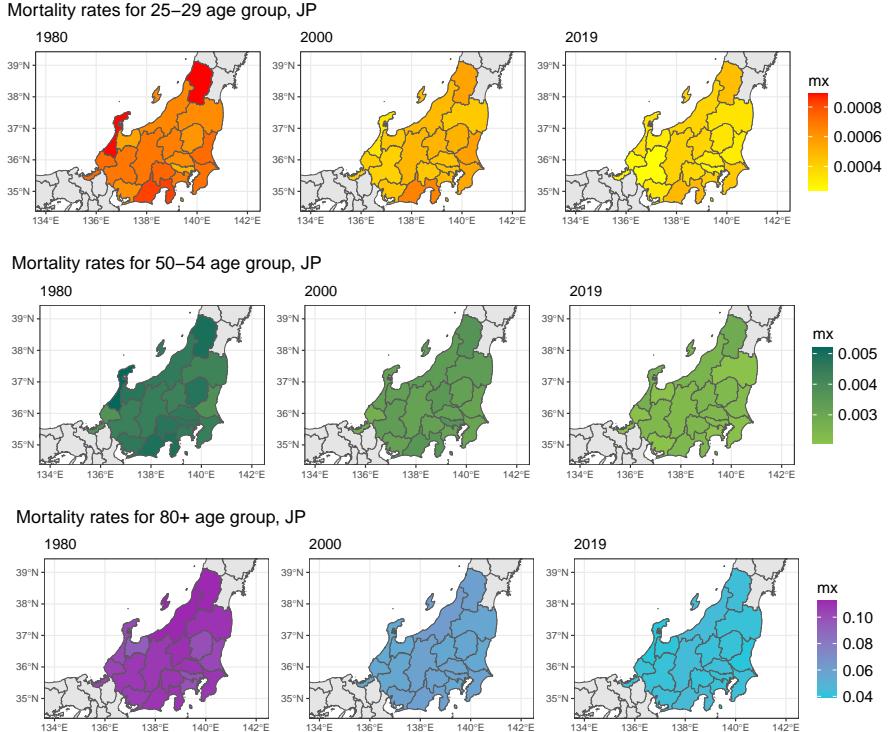


Figure 6: Data: Heat map of mortality rates throughout spatial grid. Japan, years 1980-2000-2019 and age groups 25-29, 50-54 and 80+.

Figure 7 presents the parameter estimates for the SBLC model after fitting data for Japan. Parameter  $\beta_x$  shows a somewhat increasing pattern in mortality rate improvement, with the highest being the two final age groups. Parameter  $\gamma_x$  indicates that spatial differences in mortality tend to decrease with age. This pattern is aligned with Figure 6. Parameter  $\kappa_t$  shows a decline in mortality over the years and parameter  $\theta_s$  establishes

Ishikawa prefecture as lower than average mortality rate, while Tochigi, Yamagata and Tokyo have higher than average mortality rate of the 18 prefectures studied.

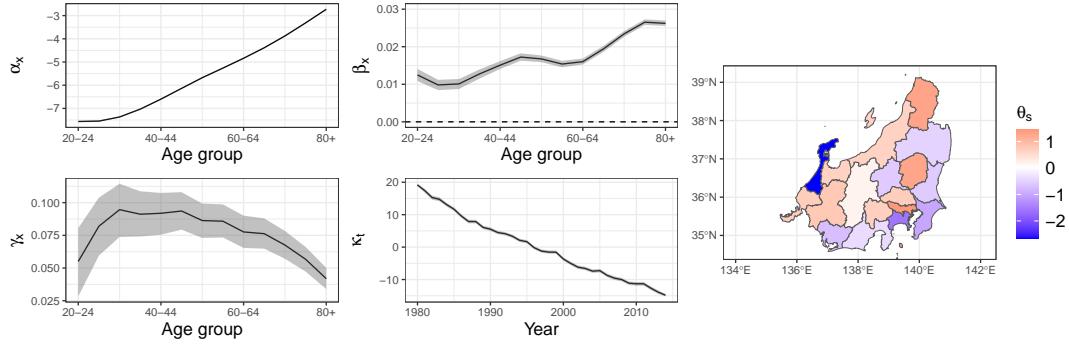


Figure 7: Posterior summaries via SBLC: Medians with 95% credible interval for parameters  $(\alpha_x, \beta_x, \gamma_x, \kappa_t)$  and median for spatial parameter  $(\theta_s)$  of the fitted SBLC model for Japan.

Figure 8 shows the fitted and predicted log mortality rates for Tokyo, Shizuoka and Fukui prefectures for some age groups, with holdout data. Notice that the number of residents in each prefecture is 13,420,622 for Tokyo (similar to Rio de Janeiro), 3,500,765 for Shizuoka and 743,651 for Fukui (both significantly bigger than Itaperuna and Santa Maria Madalena), as of Japanese Mortality Database year 2022 estimatives.

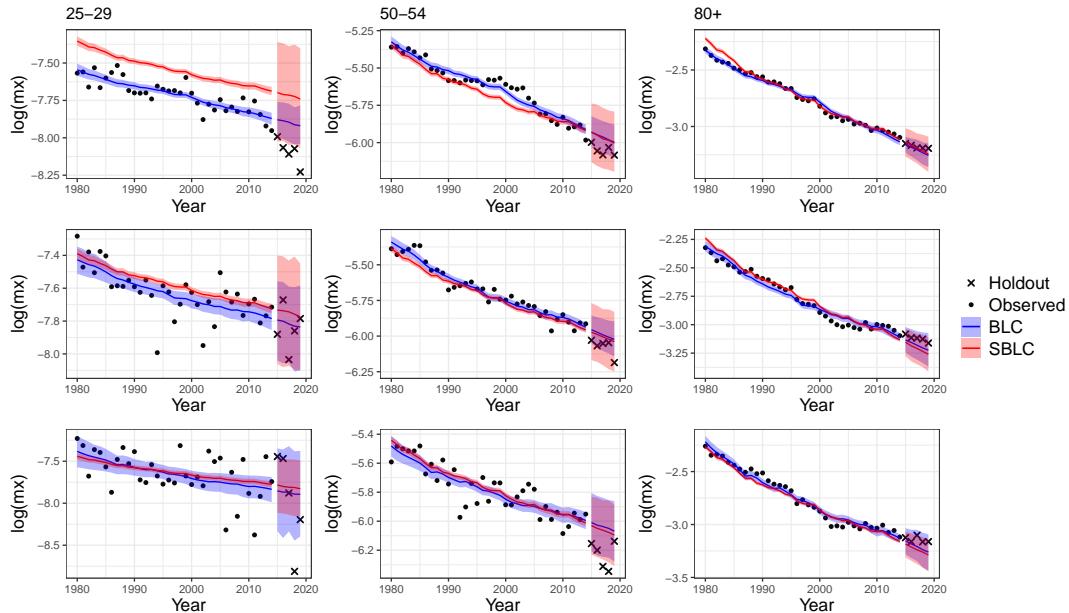


Figure 8: Posterior summaries: fitted log mortality rates of Japan prefectures Tokyo (top row), Shizuoka (middle row) and Fukui (bottom row) for BLC (blue) and SBLC (red) models. Age groups 25-29, 50-54 and 80+ (columns) and years 1980-2014 with prediction for 2015-2019.

Analogously, the same analyses presented in Section 4.1 can be done for this dataset. Parameter  $\alpha$  sometimes struggles to fit the level of mortality to the data (as shown to extreme by Tokyo prefecture in age group 25-29) and the credible intervals for the

predictions have the same behavior. The bigger data volume from Japan translates as more homogeneous mortality graduation over the spatial grid, rather diminishing the differences between the the mortality pattern captured by both models ( $\kappa_t$ ). Nonetheless, SBLC model also shows flexibility issues and lack of fit for more robust datasets.

It is worth noting that, while Japan has more spatial data and single age intervals to feed the model, the computational time to run a  $(42 \times 61) \times 30$  matrix inversion is challenging. Some solutions can be found in [Gibbs et al. \(2020\)](#), where the response matrix  $\mathbf{Y}_t$  is reduced by applying the model separately to each age group and then joining the results. Alternatively, [Li et al. \(2020\)](#) uses a Sequential Kalman filter routine ([Koopman and Durbin, 2000](#)) to increase computational efficiency by filtering only one element of  $\mathbf{Y}_t$  at a time. However, we believe that the inflexibility problem of the model would persist (and possibly worsen).

## 5 Partial Conclusions

The initial SBLC model construction proposed by this paper offered insights on spatial inference due to the interpretability of the spatial parameters added to the equation. However, the static nature to some parameters was proven to be too inflexible to fit mortality data from Rio de Janeiro's microregions and Japan's prefectures. The behavior of the mortality rates was shown to diverge greatly between regions in some age groups, causing the marginal BLC model to better capture these particularities.

Computational efficiency on the FFBS algorithm, particularly in the filtering process, for spatial models decreases greatly due to high dimension matrix inversions. Solutions presented by authors [Gibbs et al. \(2020\)](#) and [Li et al. \(2020\)](#) are being studied to mitigate this effect on the SBLC model.

Future work to be done includes building variations to the SBLC model to allow more parameter flexibility, based on the inferences obtained by this exercise, and implementing a more efficient filtering routine.

## References

- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 2nd edition.
- BayesMortalityPlus (2024). BayesMortalityPlus: Bayesian mortality models. Laboratório de Matemática Aplicada (LabMA/UFRJ), R package, GPL-3 license.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.
- Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746.
- Bowers, N. L. (1986). *Actuarial Mathematics*. Society of Actuaries.
- Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal*, 13(1):1–35.

- Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81(3):541–553.
- Czado, C., Delwarde, A., and Denuit, M. (2005). Bayesian poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, 36(3):260–284.
- DATASUS (2024). Brasil. Ministério da Saúde. Departamento de Informática do Sistema Único de Saúde. <https://datasus.saude.gov.br>. Accessed: 2024-04-04.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202.
- Gamerman, D. and Lopes, H. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Texts in Statistical Science. Taylor & Francis.
- Gibbs, Z., Groendyke, C., Hartman, B., and Richardson, R. (2020). Modeling county-level spatio-temporal mortality rates using dynamic linear models. *Risks*, 8(4):117.
- Gómez-Rubio, V., Palmí-Perales, F., López-Abente, G., Ramis-Prieto, R., and Fernández-Navarro, P. (2019). Bayesian joint spatio-temporal analysis of multiple diseases. *SORT-Statistics and Operations Research Transactions*, 43(1):51–74.
- Japanese Mortality Database (2024). Japanese Mortality Database. National Institute of Population and Social Security Research. <https://www.ipss.go.jp/p-toukei/JMD/index-en.asp>. Accessed: 2024-11-22.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, 82(Series D):35–45.
- Koopman, S. J. and Durbin, J. (2000). Fast filtering and smoothing for multivariate state space models. *Journal of time series analysis*, 21(3):281–296.
- Lee, R. (2000). The lee-carter method for forecasting mortality, with various extensions and applications. *North American actuarial journal*, 4(1):80–91.
- Lee, R. and Carter, L. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419):659–671.
- Li, J. S.-H., Zhou, K. Q., Zhu, X., Chan, W.-S., and Chan, F. W.-H. (2020). A Bayesian approach to developing a stochastic mortality model for China. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(4):1523–1560.
- Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42(3):575–594.
- Liu, Z., Sun, X., and Wang, Y.-B. (2021). A bayesian spatial modeling approach to mortality forecasting. *arXiv preprint arXiv:2102.11501*.
- Pedroza, C. (2002). *Bayesian hierarchical time series modeling of mortality rates*. Harvard University.
- Pedroza, C. (2006). A Bayesian forecasting model: predicting U.S. male mortality. *Bio-statistics*, 7(4):530–550.

- Renshaw, A. E. and Haberman, S. (2003). On the forecasting of mortality reduction factors. *Insurance: Mathematics and Economics*, 32(3):379–401.
- Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical association*, 92(438):607–617.
- West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*. Springer., 2nd edition.
- Wong, J. S., Forster, J. J., and Smith, P. W. (2018). Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics and Economics*, 83:206–221.