Pak.j.stat.oper.res. Vol.18 No.4 2022 pp 775-788 DOI: http://dx.doi.org/10.18187/pjsor.v18i4.4095

Pakistan Journal of Statistics and Operation Research

Bayesian bivariate spatial shared component model: mapping breast and cervical cancer mortality in Southern Brazil

Diego Gafuri Silva[1], Larissa Intrebartoli Resende[2],
Elisângela Aparecida da Silva Lizzi[3]
Jorge Alberto Achcar[4], Edson Zangiacomi Martinez[5*]

*Corresponding author

1. State University of Maringá, Master Program in Biostatistics, Maringá, Brazil, dgafuri@yahoo.com.br
2. State University of Maringá, Master Program in Biostatistics, Maringá, Brazil, pg403493@uem.br
3. Federal University of Technology – Paraná, Cornélio Procópio, Brazil, elisangelalizzi@utfpr.edu.br
4. Ribeirão Preto Medical School, Universidade de São Paulo, Brazil, achcar@fmrp.usp.br
5. Ribeirão Preto Medical School, Universidade de São Paulo, Brazil, edson@fmrp.usp.br

Abstract

Spatial analysis techniques are used in the data analysis of ecological studies, which consider geographical areas as observation units. In this article, we propose a Bayesian bivariate spatial shared component model to map breast and cervical cancer mortality in Southern Brazil, based on the models introduced by Knorr-Held and Best (2001) and Held et al. (2005). Markov Chain Monte Carlo (MCMC) methods were used to spatially smooth the standardized mortality ratios (SMR) for both diseases. The local Indicator of Spatial Association (LISA) was used to verify the existence of spatial clusters in specific geographical areas. This study was carried out using secondary data obtained from publicly available health information systems.

Key Words: Spatial Analysis; Ecological Studies; Epidemiology; Bayesian Methods; Gaussian Markov Random Field.

Mathematical Subject Classification: 62F15, 62P10.

1. Background

Ecological studies have been widely used to follow time trends and distributions of diseases in populations of interest. According to Tulchinsky and Varavikova (2014), ecological studies are important for population health monitoring and the generation of hypotheses for further investigation and intervention. Spatial studies are a type of ecological study that uses aggregated data rather than individual-level data to assess the distribution of a disease or event of interest in a geographic space. Bayesian models have been extensively used in the analysis of spatial data, and a popular modeling approach has been through the conditional autoregressive distribution and their generalizations. These models are relatively flexible and can accommodate different spatial correlation structures. Since these models present great complexity for the likelihood function, the estimation of the model parameters using frequentist inference methods can be a difficult task, and a Bayesian approach could offer a convenient alternative to deal with this model structure (Lee, 2011; Martinez and Achcar, 2014). Examples of application of spatial Bayesian models in epidemiology research include the articles by Araújo et al. (2013), Martinez and Roza (2020), and Stoppa et al. (2022). In all these articles, Markov

Chain Monte Carlo (MCMC) algorithms were used in the parameter estimation. Since the convergence of MCMC can be very slow in some cases, the use of integrated nested Laplace approximation (INLA) methods is a good alternative for parameter estimation of models with spatial structures (Bakka et al., 2018; Blangiardo and Cameletti, 2015). INLA is a computational method that can fit Bayesian models in a fraction of the time required by typical MCMC sampling. De Smedt et al. (2015) compared the performance of the MCMC and INLA methods using simulated data with spatial components and under different scenarios, and observed that both methods were equivalent for parameter estimation. However, they also observed that when the disease or event of interest is statistically very rare, INLA provides worse estimates and credible intervals for the parameters of interest. Furthermore, an advantage of MCMC methods is the great flexibility that computer programs like OpenBUGS and R offer for implementing models with different structures and parameter specifications (Cowles, 2013). In addition, Goudie et al. (2020) recently introduced MultiBUGS, a software built on the existing algorithms and tools in OpenBUGS that automatically parallelizes the MCMC algorithm to speed up computation dramatically. It is important to point out that the use of OpenBugs software requires a minimum of programming expertise, where it is only necessary to introduce the likelihood function and the prior distributions of the model parameters, which makes its use very attractive to epidemiologists and statisticians using more sophisticated epidemiological models in data analysis as considered in this study. In this way, the main goal of this study is the implementation of existing spatial Bayesian models in epidemiology research introduced in the literature assuming MCMC methods, using the free OpenBUGS software and considering a breast and cervical cancer mortality data set in Brazil, which could be very useful to epidemiologists and statisticians working with mortality data.

In the present article we use a Bayesian bivariate spatial shared component model introduced in the literature to jointly map breast and cervical cancer mortality in Southern Brazil. These types of cancer are important causes of death in women over the age of 15 (Forouzanfar et al., 2011), and studies of their distribution in geographic space are helpful to support health policies, resource allocation, and to generate hypotheses on possible associations between exposures and these diseases. The proposed model is based on the shared component model introduced by Knorr-Held and Best (2001) and Held et al. (2005). While many statistical approaches have been proposed to understand the spatial patterns of a single disease, the shared component model was developed to model two or more diseases that occur in a population. The central idea of the model is to separate the underlying risk surface for each disease into a shared and a disease-specific component. The article is organized as follows. The bivariate model with shared components is described in Section 2, along with a description of the breast and cervical cancer mortality data and details on the computational implementation of the proposed method. Section 3 shows the results obtained when the method is applied to the breast and cervical cancer data. Bayesian approaches using MCMC simulation methods were used to spatially smooth the standardized mortality ratios. Section 4 provides a discussion of the results. Finally, some concluding remarks are presented in Section 5.

2. Methods

2.1. Gaussian Markov Random Field models

Gaussian Markov Random Field (GMRF) models are widely used in spatial statistics. As defined by Rue and Held (2005), a GMRF "is just a (finite-dimensional) random vector following a multivariate normal (or Gaussian) distribution". According to this definition, many of the spatial models introduced in the literature are GMRF models. Let us consider a study aiming to assess the distribution of a disease or event of interest in a geographical region divided into n mutually exclusive areas. These areas are numbered sequentially from 1 to n and each one is defined as a polygon in a vector map.

Definition 2.1. *Let us consider that $\{S_j : j \in \Delta_i\}$ are random quantities related to the neighbours Δ_i to an area i, with $i \neq j$. A Gaussian random field $\mathbf{S} \sim N(\mu, \mathbf{\Sigma})$ that satisfies*

$$p(s_i|\{s_j : j \neq i\}) = p(s_i|\{s_j : j \in \Delta_i\})$$

is a GMRF, where μ is a mean vector and $\mathbf{\Sigma}$ is a variance-covariance matrix (Rue and Held, 2005).

The Intrinsic Conditional AutoRegressive (ICAR) model introduced by Besag et al. (1991) is a special case of a GMRF model.

Definition 2.2. *The ICAR model considers*

$$S_i|S_{\{-i\}}, \upsilon, W \sim N\left(\frac{\sum_{j\in\Delta_i} S_j}{m_i}, \frac{\upsilon}{m_i}\right), \tag{1}$$

where $S_{\{-i\}}$ denotes all the elements of $\mathbf{S}$ except S_i, $\upsilon > 0$ is an unknown variance parameter, W is a spatial weights matrix, Δ_i denotes the set of neighbours of area i as defined in W, and m_i is the number of areas which are adjacent to area i. Unnormalized weights are given by $W_{ij} = 1$ if areas i and j are adjacent and $W_{ij} = 0$ otherwise.

2.2. Standardized mortality ratios

Standardized mortality ratio (SMR) is an indirect method of adjustment of rates by age, which describes numerically how great is the mortality due to a disease in a specified area in a given year compared with the rate found considering the geographic space of interest as a whole (Ulm, 1990). In a general way, to obtain the SMR values for gender s ($s = 1$ for males, $s = 2$ for females) and period of time t ($t = 1, \ldots, T$), we first consider the corresponding number of deaths due to the disease, observed for the whole geographic space, in A different age groups. These numbers, divided by the population for the corresponding gender, age group and period, are the rates of deaths for the whole geographic space, which are denoted by g_{sta}. The expected number z_{ist} of deaths for each area i in the period t according to the gender s is given by

$$z_{ist} = \sum_{a=1}^{A} g_{sta} \times m_{ista}, \tag{2}$$

where m_{ista} is the number of inhabitants in the area i with gender s at the period t and age group a ($a = 1, \ldots, A$). The SMR is thus given by

$$SMR_{ist} = \frac{y_{ist}}{z_{ist}},$$

where y_{ist} is the number of deaths notified in the area i at the period t, considering gender s (Silva-Lizzi et al., 2016). An SMR equals to one suggests that the observed mortality exceeds expectations in the population under study. A SMR of 1 suggests that the observed number of deaths is not different from what would be predicted for the population of the whole geographic space. In the present article, we will not consider the s and p indexes in our notation, since the database covers only one period (the year 2019) and is composed only of women.

2.3. The bivariate model with shared components

Let us consider an ecological study whose objective is to simultaneously assess the spatial distribution of deaths due to two different diseases or events of interest. Let y_{ik} and z_{ik} denote, respectively, the observed and expected number of deaths reported in area i ($i = 1, \ldots, n$) due to a specific disease k, $k = 1, 2$. Under a Bayesian framework, we assume the model

$$y_{ik} \sim Poisson(z_{ik}e^{\eta_{ik}}),\ k = 1, 2,$$

where $e^{\eta_{ik}}$ corresponds to the SMR for the area i and disease k,

$$\eta_{i1} \sim N(\alpha_1 + u_{1i}\delta + u_{2i}, \tau_1)$$

and

$$\eta_{i2} \sim N\left(\alpha_2 + \frac{u_{1i}}{\delta}, \tau_2\right).$$

The parameters α_1 and α_2 are disease specific intercepts, $\mathbf{u}_1$ is the latent shared component common to both diseases and $\mathbf{u}_2$ is the component only relevant to disease 1. The unknown parameter $\delta > 0$ is described by Held et al. (2005) as a parameter that "allow for a different risk gradient of the shared component for the two diseases". Let us assume a normal prior distribution for $\log\delta$, that is,

$$\log\delta \sim N(0, 0.17).$$

As a motivation for the choice of the value 0.17 assigned to the prior variance, we refer to Knorr-Held and Best (2001). Held et al. (2005) suggested the use of "weakly"informative prior inverse-gamma distributions for τ_1 and τ_2. Thus,

we assume

$$\tau_k \sim IG(1, 0.01),\ k = 1, 2,$$

where $IG(a, b)$ denotes an inverse-gamma distribution with probability density function given by

$$\pi(\tau_k) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{\tau_k}\right)^{a+1} \exp\left(-\frac{b}{\tau_k}\right),$$

with mean $b/(a-1)$ for $a > 1$, variance $b^2/[(a-1)^2(a-2)]$, and mode $b/(a+1)$. This implies in a prior mode value given by $0.01/2 = 0.005$ for τ_k, with infinite expectation and variance. The shared components $\mathbf{u}_1$ and $\mathbf{u}_2$ are assumed to follow a ICAR Normal prior (1) with sum-to-zero constraints on the random effect terms, unnormalized weights, and unknown variance parameters υ_1 and υ_2, respectively. For these variance parameters we also use inverse-gamma prior distributions given by

$$\upsilon_k \sim IG(1, 0.01),\ k = 1, 2.$$

For a Bayesian analysis of this model, we considered the use of Markov Chain Monte Carlo (MCMC) methods to simulate samples of the joint posterior distribution for the parameters of the proposed model (Gilks et al., 1995). A known advantage of using Bayesian methods is that this methodology incorporates the prior expert information about the model parameters given under a prior distribution, in addition to the data information given by the likelihood function, leading to more accurate inference results. In spatial data analysis, Bayesian models have been widely used for their efficiency in quantifying uncertainty and handling complex hierarchical problems associated with modern simulation techniques.

A satisfactory model should take into account all the spatial trends in the data and expect no spatial correlation for the model's residuals (Anderson and Ryan, 2017). Under a Bayesian framework, Carlin and Louis (2008) describe a residual as

$$r_{ik} = y_{ik} - \frac{1}{G}\sum_{g=1}^{G} E\left(y_{ik}|\theta_{ik}^{(g)}\right)$$

where G is the number of MCMC iterations, $E\left(y_{ik}|\theta_{ik}^{(g)}\right)$ is the expected value for the posterior predictive distribution, and $\{\theta_{ik}^{(g)}\}$ is a set of parameters sampled from the posterior distribution (Lawson, 2018). Considering a Poisson likelihood and expectation $z_{ik}e^{\widehat{\eta}_{ik}}$, the residuals can be approximated by

$$r_{ik} = y_{ik} - z_{ik}e^{\widehat{\eta}_{ik}},$$

where $\widehat{\eta}_{ik}$ are the average values of the parameter η_{ik} obtained from the simulated MCMC posterior sample (Lawson, 2018). Standardized residuals are given by

$$r_{ik}^{*} = \frac{y_{ik} - z_{ik}e^{\widehat{\eta}_{ik}}}{\sqrt{z_{ik}e^{\widehat{\eta}_{ik}}}}. \tag{3}$$

Anderson and Ryan (2017) proposed the use of a spatial Moran's index (I^2) (Moran, 1950) as a method to measure the extent of spatial autocorrelation in the residual values from the model fit. The closer I^2 is to zero, better the model accounts for the spatial autocorrelation.

2.4. The breast and cervical cancer mortality data

To simplify the administration of the Brazilian Unified Health System (SUS), the Ministry of Health office organizes the division of the Brazilian territory into Health Regions. The Health Regions are geographical areas composed of groupings of municipalities aiming for better planning of health services among them (Bousquat et al., 2019). The Brazilian municipalities are grouped into 450 Health Regions. In the present study, we considered all the 68 Health Regions in the Southern region of Brazil. This region is composed of three federative units (Paraná, Santa Catarina, and Rio Grande do Sul), as shown in the map presented in Figure 1. The Southern region is bordered to the east by the Atlantic Ocean.

Information on the number of deaths due to breast and cervical cancer reported in 2019 was obtained from the Information System on Mortality (SIM/DATASUS) (Morais and Costa, 2017). The anatomical location of tumors was

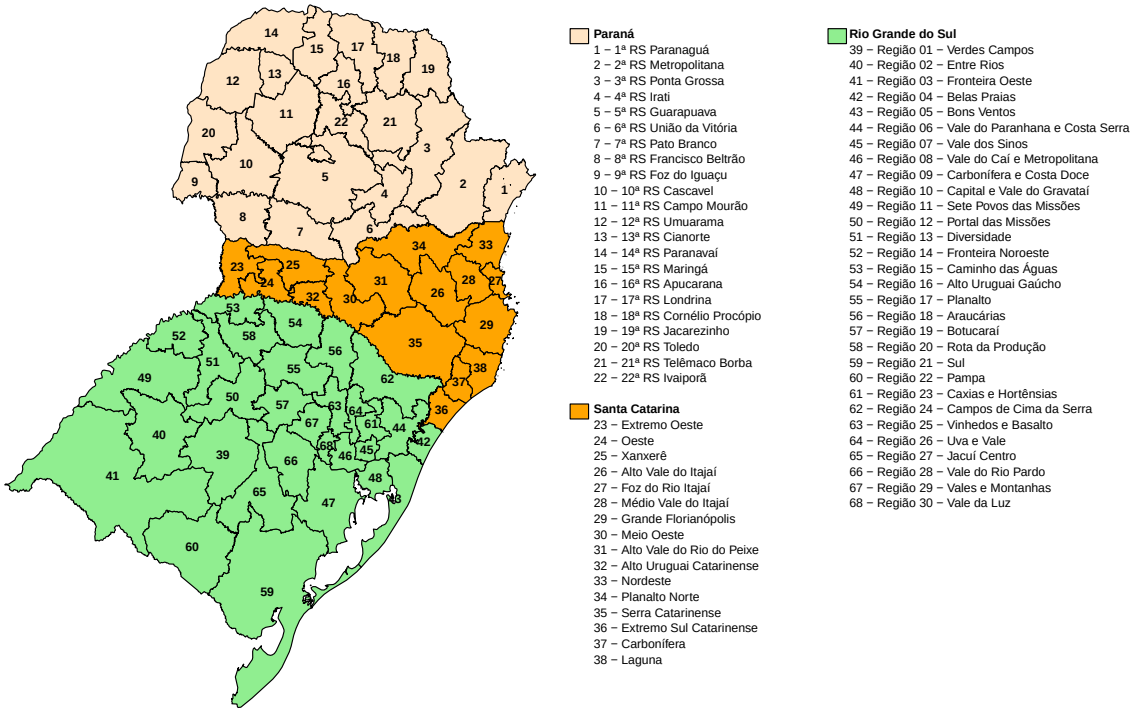

Figure 1: Health Regions in Southern Brazil. The Southern region of Brazil is divided into three federative units (Paraná, Santa Catarina, and Rio Grande do Sul) and 68 Health Regions.

coded according to the tenth revision of the International Classification of Diseases (ICD-10). Cancer disease types were selected according to these codes: breast (C50) and cervix (C53). The number of inhabitants in each one of the Health Regions by age groups ($m_{ia}, i = 1, \ldots, 68, a = 1, \ldots, 7$) was obtained from the Brazilian health statistics site TabNet/DATASUS, where the information for 2019 was calculated by projections based on demographic census data. To obtain the expected number z_{ik} of deaths due to breast ($k = 1$) and cervical ($k = 2$) cancer for each area i in 2019 (equation (2)), we considered the following age groups: 20-29 years ($a = 1$), 30-39 years ($a = 2$), 40-49 years ($a = 3$), 50-59 years ($a = 4$), 60-69 years ($a = 5$), 70-79 years ($a = 6$), and 80 and older ($a = 7$).

Summary statistics for the observed number of deaths due to breast and cervical cancer and the number of women aged 20 years and older living in each of the 68 Health Regions in Southern Brazil in 2019 are shown in Table 1.

Table 1: Summary statistics for the observed number of deaths due to breast and cervical cancer and the population size (women aged 20 years and older) of the 68 Health Regions in Southern Brazil included in this study. Data for 2019.

	n	Mean	Median	1^{st} Quartile	3^{rd} Quartile	Minimum	Maximum
Deaths by breast cancer	68	45.6	26.5	16.75	44.75	3	375
Deaths by cervical cancer	68	14.2	10.0	6.0	16.0	0	96
Population size	68	222,470	145,280	96,652	223,221	49,771	1,881,102

2.5. Computational implementation of the model

In the OpenBUGS software, we can use the `car.normal` function to fit the spatial ICAR model. Its syntax is given by

```
S[1:N] ~ car.normal(adj[], weights[], num[], tau)
```

where N is the number of areas, adj[] is the adjacency vector defined for the neighbourhoods, weights[] is a vector of length N giving the weights for each pair of neighbours, num[] is a vector of the number of neighbours for each area, and tau is the variance parameter (Lawson, 2013). Vectors adj[] and weights[] must be of the same length, and unnormalized weights are obtained setting all the elements in weights[] equal to 1. Figure 1 shows that the Health Region 1 has two neighbors (labeled as 2 and 33), the Health Region 2 has four neighbors (1, 3, 6, and 34), the Health Region 3 has seven neighbors (2, 4, 5, 6, 19, 21, and 22), the Health Region 4 has three neighbors (3, 5, and 6), and so on. Therefore, in the OpenBUGS syntax the vector num[] is given by

```
num = c(2, 4, 7, 3, 7, 8, 6, 5, 3, 7, 7, 5, 4, 3, 5, 5, 4, 3, 3, 3, 6, 5, 4,
5, 6, 5, 3, 5, 5, 8, 5, 5, 4, 6, 7, 4, 3, 3, 5, 4, 4, 3, 4, 7, 5, 6,
6, 5, 5, 8, 5, 2, 5, 6, 6, 7, 6, 5, 4, 4, 4, 7, 7, 6, 7, 5, 4, 5)
```

and the 16 first elements of the vector adj[] are given by

```
adj = c(2, 33, 1, 3, 6, 34, 2, 4, 5, 6, 19, 21, 22, 3, 5, 6, ...)
```

The OpenBUGS code used to implement the bivariate model with shared components is presented in an Appendix at the end of the manuscript. We first simulated 5,000 initial samples as a burn-in sample to eliminate the effect of the initial values for the parameters of the model in the algorithm. Next, we simulated another $G = 1,000,000$ samples, taking every 100th sample to obtain approximately uncorrelated samples of the joint posterior distribution of all parameters of the model. The Bayesian estimates for the parameters were obtained as the mean of the simulated samples drawn from the joint posterior distribution for all parameters of the model, that is, we are assuming a quadratic loss function to get the Bayesian estimators of interest. The convergence of the simulated sequences was monitored by using traceplots and the Geweke diagnostic criterion (Cowles and Carlin, 1994). This convergence criterion is based on a z score that compares the difference in the two means of non-overlapping sections of a simulated Markov chain, divided by the asymptotic standard error of the difference. This z score asymptotically follows a standardized normal distribution, so we obtain convergence for a chain if its corresponding absolute z score is less than 1.96. Geweke z-scores were obtained using the geweke.diag function of the MCMCpack package of the R software (Martin et al., 2011). Highest probability density (HPD) intervals for the parameters of interest were obtained using the Monte Carlo method described by Chen and Shao (1999). Box and Tiao (2013) state that an HPD interval has two main properties. First, the density for every point inside the interval is greater than that for every point outside the interval. Second, for a given probability content, the HPD interval is the one of shortest length (Chen and Shao, 1999). The Monte Carlo standard error (MC error) was used as a measure of how accurately the mean of the Monte Carlo samples for a given parameter estimates its true posterior expectation. As a rule of thumb, the number G of simulated MCMC samples is considered sufficient when the MC error is less than 1% to 5% of the posterior standard deviation (SD) for a parameter of interest (Lunn et al., 2013).

The Moran's index (I^2) was used to estimate the degree of spatial autocorrelation (Moran, 1950). This index is a generalization of Pearson's correlation coefficient, and an I^2 value close to zero indicates that a measure of interest tends to have a random spatial distribution in the geographic region of interest. The Moran's index was calculated using the moran.test function of the spdep package of the R software. Local Indicators of Spatial Association (LISA) were used to estimate the spatial autocorrelation between the Health Regions and thus to identify the local clusters (Anselin, 1995). Spatial autocorrelation of LISA is classified into four different classes, including high-high, low-low, high-low, and low-high, which respectively mean "high surrounded by high", "low surrounded by low", "high surrounded by low", and "low surrounded by high". A fifth class is commonly called "not significant" and encompasses all areas in which there are no important associations. LISA maps showing these classifications in each Health Region were used to identify spatial clusters and examine the spatial patterns of mortality due to breast and cervical cancer in Southern Brazil.

The code to reproduce this work is also available in GitHub (https://github.com/edsonzmartinez/sharedcomponents).

3. Results

The computer code provided in the Appendix was used in OpenBUGS and MultiBUGS software, both installed on a personal computer with an 11th Gen Intel®Core™i7 processor (2.80 GHz). Using the OpenBUGS software, the processing time for 1,000,000 iterations was 648 seconds, and using the MultiBUGS software, this processing time was 638.5 seconds. Since the model involves a relatively small number of parameters and observations, both software showed very close running times. The obtained results from both software were similar.

Table 2: Bayesian estimates (posterior means) of the parameters.

Parameter	Estimate	SD	MC error	95% HPD interval	Geweke z-score
α_1	-0.08949	0.02763	0.0003	(-0.1434, -0.0354)	-0.048
α_2	-0.02340	0.04240	0.0004	(-0.1038, 0.0620)	0.972
δ	1.119	0.335	0.0045	(0.4951, 1.7630)	0.835
τ_1	0.01211	0.00955	0.0001	$(3.383 \times 10^{-6}, 0.0307)$	1.369
τ_2	0.02491	0.01833	0.0002	$(1.798 \times 10^{-5}, 0.0597)$	-1.251
υ_1	0.03111	0.02306	0.0003	$(2.273 \times 10^{-5}, 0.0754)$	-0.422
υ_2	0.04237	0.03407	0.0005	$(9.383 \times 10^{-6}, 0.1090)$	-1.825

Table 2 shows the Bayesian estimates of the parameters α_1, α_2, δ, τ_1, τ_2, υ_1, and υ_2, obtained from the MultiBUGS software, together with their sample standard deviation (SD), MC error, 95% HPD intervals, and Geweke z-scores. The MC errors for each parameter are less than 5% of their corresponding SD, indicating that the number of MCMC simulated samples is sufficient to have a correct estimation. In addition, all absolute Geweke z-scores are smaller than 1.96 suggesting the convergence of the simulated Markov chains to a stable distribution. The posterior samples for these parameters are graphically described in Figure 2. The traceplots presented in the first column of the Figure 2 show the evolution of the samples generated by the MCMC method over the iterations, also indicating good convergence of the simulation algorithm. The second column of Figure 2 shows unimodal posterior distributions for the parameters, and the autocorrelation function (ACF) plots presented in the third column show that the autocorrelation between successive samples is reasonably low. We also verified the convergence of the simulated samples for the shared components u_{1i} and for u_{2i}, $i = 1, \ldots, 68$, but we do not show the results here due to space limitation.

Bayesian estimates and 95% HPD intervals for the SMR for breast and cervical cancer are visually represented in Figure 3. The Health Regions 10 to 14, 17, 18, 20, 22, and 25 are highlighted in green in the panel (a) of Figure 3. These regions have relatively low SMR estimates for breast cancer and corresponding HPD intervals which do not contain the value 1. Except for region 25, all these Health Regions are located in the state of Paraná (see Figure 1). The Health Regions 33, 41, 48, and 59 have relatively high SMR estimates for breast cancer and are highlighted in red in the panel (a) of Figure 3. The Health Region 33 is located in the northeast of Santa Catarina state, and the regions 41, 48, and 49 are located in the Rio Grande do Sul state. Panel (b) of Figure 3 visually suggests that the variability of the SMR estimates for cervical cancer is lower than that observed for breast cancer, and all the corresponding HPD intervals contain the value 1. The maps in the Figure 4 describe the distribution of the Bayesian estimates for SMR for breast and cervical cancer across the geographic space.

Figure 5 shows LISA maps of the smoothed estimates for SMR for breast and cervical cancer. The obtained Moran's index with value 0.537 suggests a non-homogeneous distribution pattern of the SMR for breast cancer in the southern region of Brazil. Panel (a) of Figure 5 identifies spatial High-High clusters in the southwest region of Rio Grande do Sul state and in the coastal regions of Paraná and Santa Catarina states. Low-Low clusters are identified in western Paraná state. The obtained Moran's index with value 0.298 indicates that the spatial distribution of the SMR for cervical cancer has a lower heterogeneity degree when compared to the SMR distribution for breast cancer in the southern region of Brazil. Nevertheless, panel (b) of Figure 5 identifies High-High spatial clusters in the southwest region of Rio Grande do Sul state and Low-Low clusters in the northwest of Paraná state. High-Low and Low-High classes were absent in our data set.

Standardized residuals for the Bayesian model, given by the expression (3), are shown in Figure 6, together with their corresponding spatial Moran's indexes. From these plots we can see that the majority of the standardized residuals are in the range of $(-2, 2)$, suggesting the absence of extremely low or extremely high values. The Moran's indexes for the spatial distribution of the standardized residuals are given by -0.083 (breast cancer) and -0.039 (cervical cancer). As these values are close to zero, we conclude that the proposed model is able to explain the spatial distribution of the SMR for the two diseases in the Southern Brazilian region.

For comparison, we also fitted a similar Bayesian model with both components $\mathbf{u}_1$ and $\mathbf{u}_2$ assuming a double exponential (Laplace) prior distribution in place of a Gaussian prior distribution. This distribution is denoted by `car.l1` in the OpenBUGS syntax. Comparing the fit of the models with shared components following a double exponential distribution and a Gaussian distribution to the data, the overall goodness-of-fit measured by the deviance information

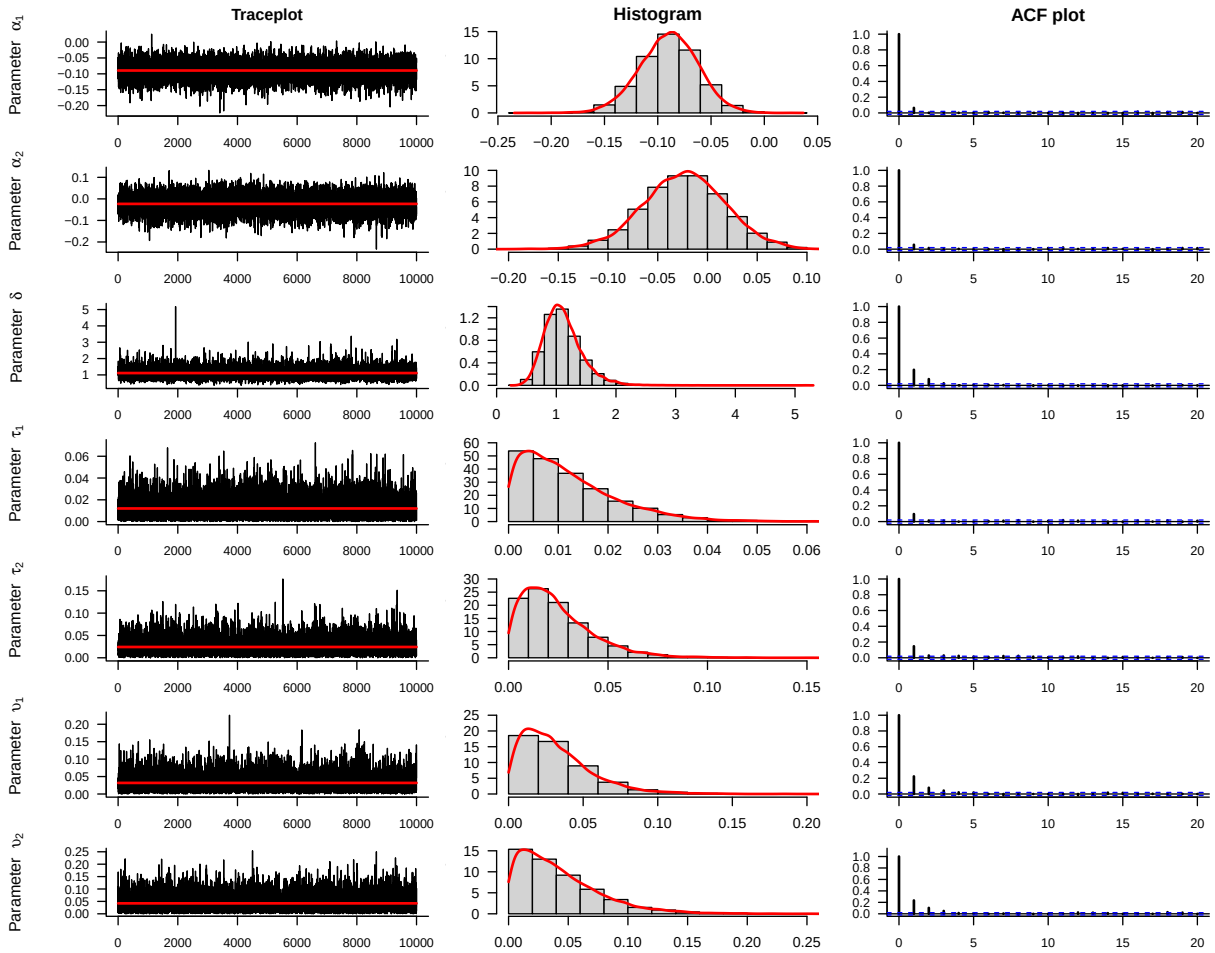

Figure 2: Posterior samples for the parameters α_1, α_2, δ, τ_1, τ_2, υ_1, and υ_2 of the bivariate model with shared components applied to the breast and cervical cancer mortality data. (a) Traceplot of posterior samples (the horizontal red lines correspond to the posterior means), (b) histogram and kernel density estimate for the posterior density, and (c) autocorrelation function (ACF) plot for the posterior samples of the model parameter.

criterion (DIC) were, respectively, 829.0 and 828.4. Since the DIC values are very close for both models, we do not have indication that one model has better fit in comparison to the other model.

4. Discussion

In the same direction as the present study, Bermudi et al. (2020) carried out a study with the objective of verifying the spatial pattern of mortality from breast and cervical cancer in the areas covered by basic health units in the metropolitan area of the São Paulo city, Brazil, showing that their obtained results are useful to help to direct resources to prevent and promote health in this geographical area. In another study, Diniz et al. (2017) aimed to identify the factors associated with the age-standardized breast cancer mortality rate in the municipalities of São Paulo state, Brazil, in the period from 2006 to 2012, showed a positive association between mammography and breast cancer mortality, and Rocha-Brischiliari et al. (2018) showed that the breast cancer mortality rate in Paraná state, Brazil, was negatively associated with illiteracy rate and positively associated with access to treatment and health services. Although these ecological studies have been important in elucidating invisible effects on the individual level, their data analysis was not formally based on a statistical model such as the one described in the present article. Considering the breast and cervical cancer mortality data as examples, we believe that the proposed model and possible extensions would be a suitable alternative for the analysis of spatial data, where the interest is to simultaneously describe the spatial pattern of the incidence, prevalence, or mortality due to two or more diseases in a region of interest.

A potential limitation of the data used in this study is due to the possible regional differences in the coverage, completeness, and quality of the information, which can lead to biased results (Felix et al., 2012). We did not use a

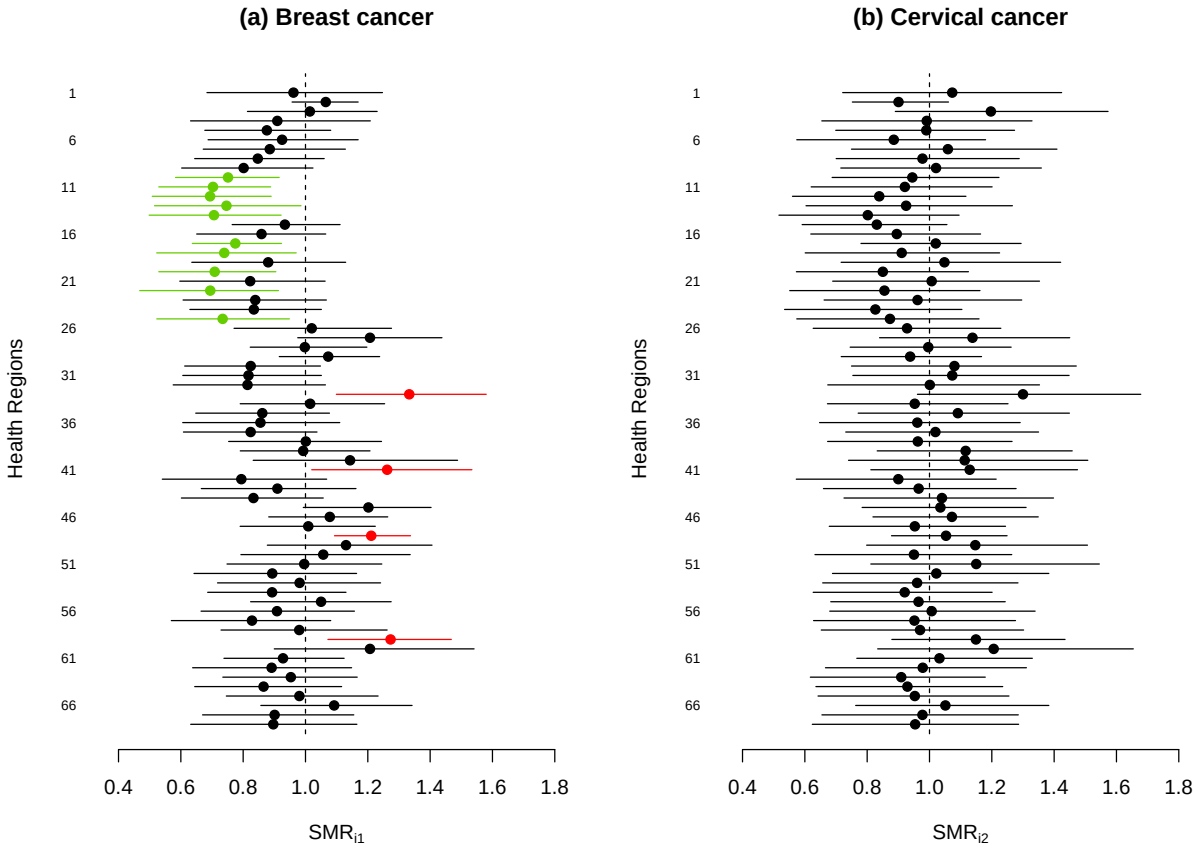

Figure 3: Bayesian estimates and 95% HPD intervals for the standardized mortality ratios (SMR) for (a) breast cancer and (b) cervical cancer. Health regions with relatively low SMR estimates and HPD intervals that do not contain the value 1 are highlighted in green, and health regions with relatively high SMR estimates are highlighted in red.

correction approach for underreporting and relocation of ill-defined causes of mortality. We also acknowledge that our results may be subject to edge effects, defined by Lawson (2013) as "any effect upon the analysis of the observed data brought about by the proximity of the window edges". The Southern Brazil region is bordered to the north by São Paulo state, to the northwest by Mato Grosso do Sul state, to the west by Argentina and Paraguay, and to the south by Uruguay. As information from the outer region is not included in the statistical model, estimates for SMR in Health Regions near boundaries can be affected by the edge position. A method proposed by Lawson (2013) to deal with edge effects is based on the inclusion of "guard areas" in the model, that is, areas outside the region of interest having a neighborhood with its boundary. However, this was not possible in the present study since our database does not have mortality information from Paraguay, Uruguay, and Argentina. Another approach was introduced by Rodeiro and Lawson (2005), which explores how the estimation of the measures of interest near boundaries can be affected by the edge position and how the error due to boundary effects propagates as we go further into the region. Their method considers successive models based on subsets of the dataset, where the first set consists of the most external areas (boundary areas), the second set consists of areas in a second ring away from the edge, the third set consists of the areas in a third ring, and so on. The map must, however, have a sufficient number of small areas to allow for "boundary hull stripping up" to five depth. Unfortunately, the map in Figure 1 shows that the number of areas in the southern region of Brazil is not large enough to use this method.

Raei et al. (2018) introduced an extension of the model with shared components, including both spatial and temporal effects. The authors used this model to map the incidence of breast and cervix uteri cancer among Iranian women over a 6-year period searching for trend changes and risk factors. Ahmadipanahmehrabadi et al. (2019) used a similar model to determine the spatial pattern and temporal trend of death risk due to colorectal and stomach cancers among provinces of Iran. Downing et al. (2008) proposed an extension of the model with shared components to more than two diseases. The authors used this model to estimate jointly the incidence rates of six smoking-related cancers in the

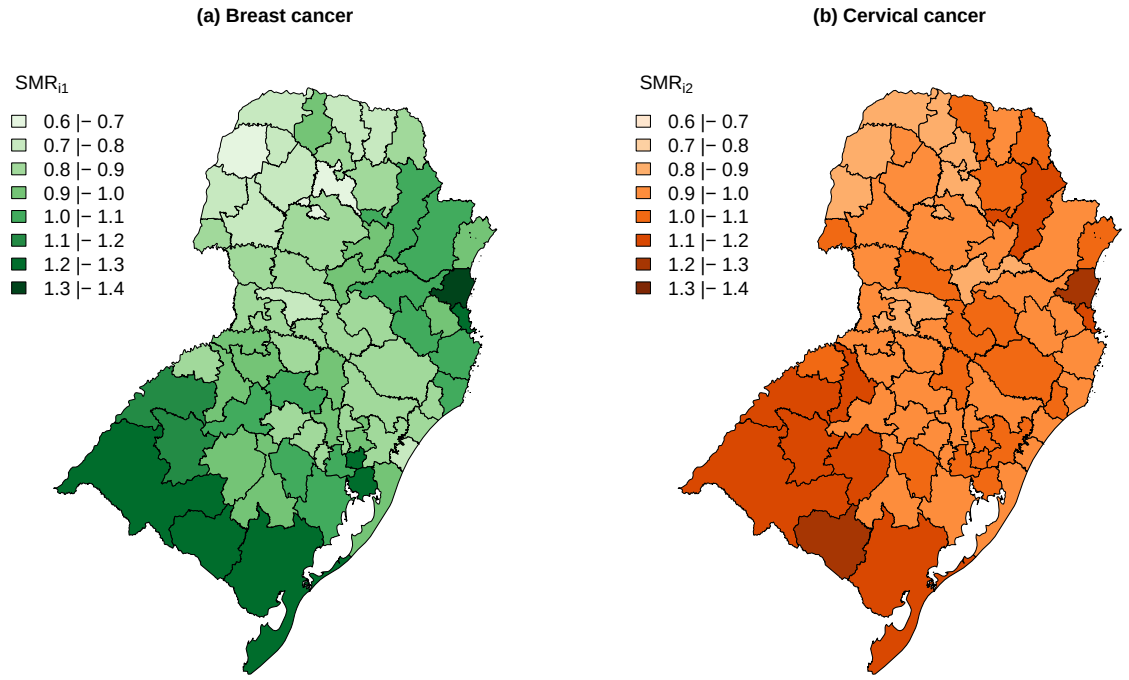

Figure 4: Bayesian estimates for standardized mortality ratios (SMR) for (a) breast cancer and (b) cervical cancer.

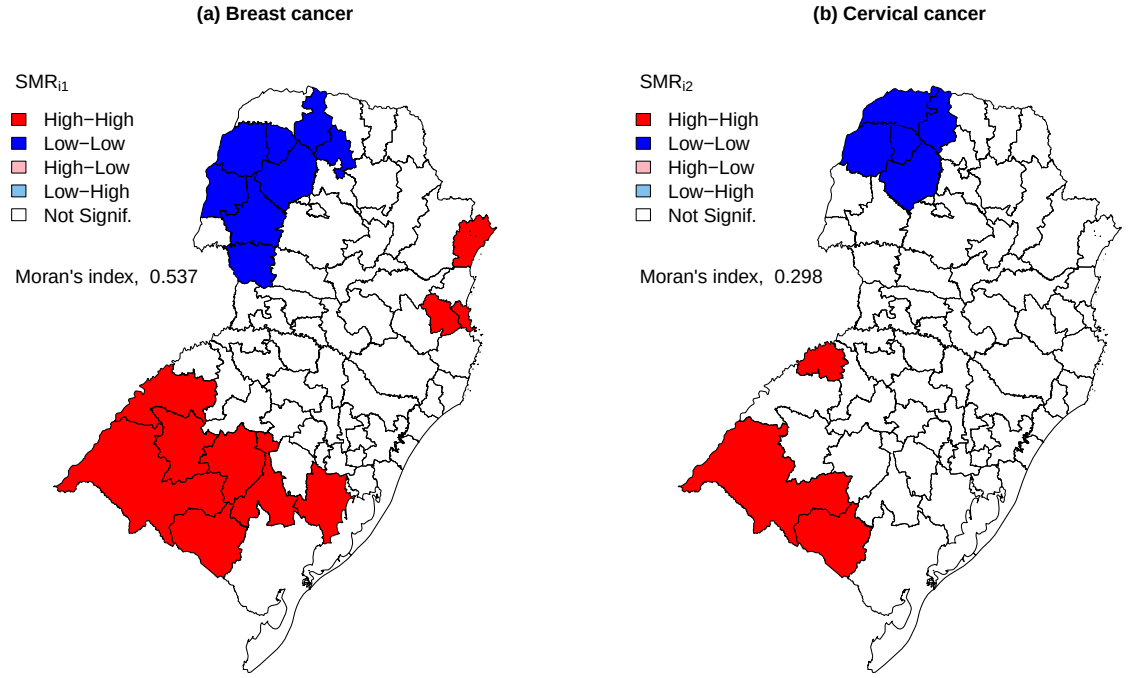

Figure 5: Global Moran's Index (I^2) and Local Indicators of Spatial Association (LISA) maps of the smoothed estimates for standardized mortality ratios (SMR) for (a) breast cancer and (b) cervical cancer.

Yorkshire region of England. These extensions of the model will be considered in our future research.

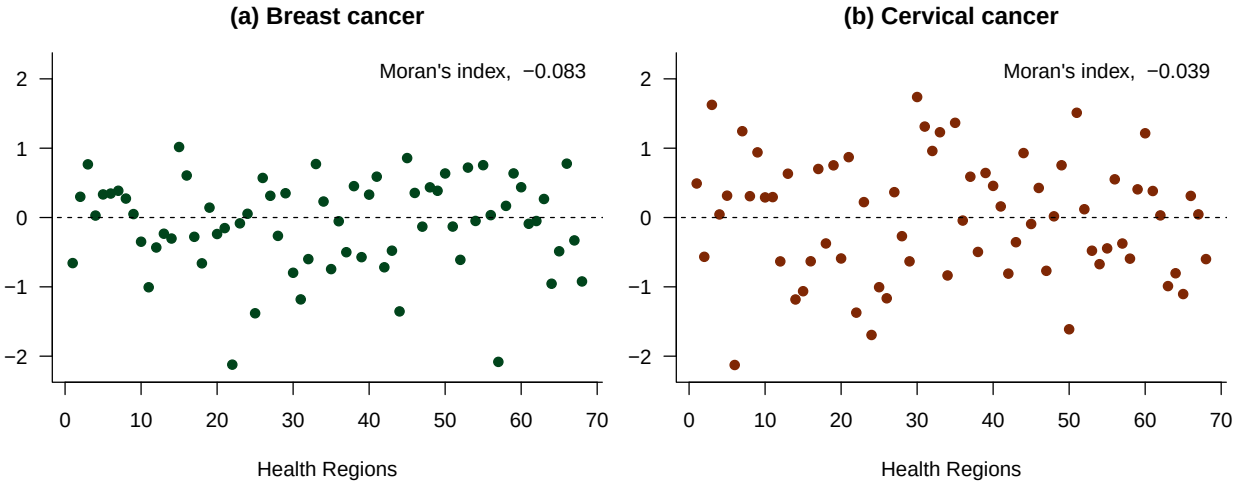

Figure 6: Standardized residuals for the Bayesian model, considering the (a) breast cancer data and the (b) cervical cancer data.

5. Concluding Remarks

The method proposed in this article is based on the models introduced by Knorr-Held and Best (2001) and Held et al. (2005). It can be easily implemented in OpenBUGS software, which requires little expertise in computer programming. As observed from the data analysis assumed in the study considering breast and cervical cancer mortality in an important region of Brazil, the computational MCMC algorithm showed good convergence in the generation of samples from the joint posterior distribution of interest and short computational time. In the application considered in this study, the model highlighted the health regions with higher or lower mortality due to the diseases under study. These results may be of great importance to health authorities in planning medical and financial resource distribution to areas with higher mortality rates.

Acknowledgements

The authors thank Landir Saviniec and Alexsandra Bezerra da Rocha for making the shapefile of the Brazilian Health Regions available on GitHub (https://github.com/lansaviniec/shapefile_das_regionais_de_saude_sus).

References

1. Ahmadipanahmehrabadi, V., Hassanzadeh, A., and Mahaki, B. (2019). Bivariate spatio-temporal shared component modeling: Mapping of relative death risk due to colorectal and stomach cancers in Iran provinces. *International Journal of Preventive Medicine*, 10:39, doi:10.4103/ijpvm.IJPVM_31_17.
2. Anderson, C. and Ryan, L. M. (2017). A comparison of spatio-temporal disease mapping approaches including an application to ischaemic heart disease in New South Wales, Australia. *International Journal of Environmental Research and Public Health*, 14:146, doi:10.3390/ijerph14020146.
3. Anselin, L. (1995). Local indicators of spatial association - LISA. *Geographical Analysis*, 27:93–115, doi:10.1111/j.1538–4632.1995.tb00338.x.
4. Araújo, V. E. M., Pinheiro, L. C., Almeida, M. C. M., Menezes, F. C., Morais, M. H. F., Reis, I. A., Assunção, R. M., and Carneiro, M. (2013). Relative risk of visceral leishmaniasis in Brazil: a spatial analysis in urban area. *PLoS Neglected Tropical Disease*, 7:e2540, doi:10.1371/journal.pntd.0002540.
5. Bakka, H., Rue, H., Fuglstad, G. A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10:61–82, doi:10.1002/wics.1443.
6. Bermudi, P. M. M., Pellini, A. C. G., Rebolledo, E. A. S., Diniz, C. S. G., Aguiar, B. S. D., Ribeiro, A. G., Failla, M. A., Baquero, O. S., and Chiaravalloti-Neto, F. (2020). Spatial pattern of mortality from breast and cervical cancer in the city of São Paulo. *Revista de Saúde Pública*, 54:142, doi:10.11606/s1518–8787.2020054002447.

7. Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–59, doi:10.1007/BF00116466.
8. Blangiardo, M. and Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons, New York, USA.
9. Bousquat, A., Giovanella, L., Fausto, M. C. R., Medina, M. G., Martins, C. L., Almeida, P. F., Campos, E. M. S., and Mota, P. H. S. (2019). Primary care in health regions: policy, structure, and organization. *Cadernos de Saude Publica*, 35:e00099118, doi:10.1590/0102–311X00099118.
10. Box, G. E. and Tiao, G. C. (2013). *Bayesian inference in statistical analysis, 3rd edition*. Chapman and Hall/CRC, Boca Raton, USA.
11. Carlin, B. P. and Louis, T. A. (2008). *Bayesian methods for data analysis*. Chapman and Hall/CRC, Boca Raton, USA.
12. Chen, M. H. and Shao, Q. M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8:69–92, doi:10.1080/10618600.1999.10474802.
13. Cowles, M. K. (2013). *Applied Bayesian statistics: with R and OpenBUGS examples*. Springer Science & Business Media, New York, USA.
14. Cowles, M. K. and Carlin, B. P. (1994). Markov Chain Monte Carlo convergence diagnostics: A comparative review. *Journal of American Statistical Association*, 91:883–904, doi:10.1080/01621459.1996.10476956.
15. De Smedt, T., Simons, K., Van Nieuwenhuyse, A., and Molenberghs, G. (2015). Comparing MCMC and INLA for disease mapping with Bayesian hierarchical model. *Archives of Public Health*, 73:O2, doi:10.1186/2049–3258–73–S1–O2.
16. Diniz, C. S. G., Pellini, A. C. G., Ribeiro, A. G., Tedardi, M. V., Miranda, M. J., Touso, M. M., Baquero, O. S., C., S. P., and Chiaravalloti-Neto, F. (2017). Breast cancer mortality and associated factors in São Paulo State, Brazil: an ecological analysis. *BMJ open*, 7:e016395, doi:10.1136/bmjopen–2017–016395.
17. Downing, A., Forman, D., Gilthorpe, M. S., Edwards, K. L., and Manda, S. O. (2008). Joint disease mapping using six cancers in the Yorkshire region of England. *International Journal of Health Geographics*, 7:1–14, doi:10.1186/1476–072X–7–41.
18. Felix, J. D., Zandonade, E., Amorim, M. H. C., and Castro, D. S. D. (2012). Evaluation of the plenitude of epidemiological variables of the Information System on Mortality of women with deaths from breast cancer in the Southeast Region - Brazil (1998 - 2007). *Ciência & Saúde Coletiva*, 17:945–953, doi:10.1590/S1413–81232012000400016.
19. Forouzanfar, M. H., Foreman, K. J., Delossantos, A. M., Lozano, R., Lopez, A. D., Murray, C. J., and Naghavi, M. (2011). Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. *The Lancet*, 378:1461–1484, doi:10.1016/S0140–6736(11)61351–2.
20. Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in Practice*. CRC Press, Boca Raton, USA.
21. Goudie, R. J. B., Turner, R. M., De Angelis, D., and Thomas, A. (2020). MultiBUGS: A parallel implementation of the BUGS modelling framework for faster Bayesian inference. *Journal of Statistical Software*, 95:7, doi:10.18637/jss.v095.i07.
22. Held, L., Natário, I., Fenton, S. E., Rue, H., and Becker, N. (2005). Towards joint disease mappings. *Statistical Methods in Medical Research*, 14:61–82, doi:10.1191/0962280205sm389oa.
23. Knorr-Held, L. and Best, N. G. (2001). A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164:73–85, doi:10.1111/1467–985X.00187.
24. Lawson, A. B. (2013). *Statistical Methods in Spatial Epidemiology*. John Wiley & Sons, New York, USA.
25. Lawson, A. B. (2018). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology. Third Edition*. Chapman and Hall/CRC, Boca Raton, USA.
26. Lee, D. (2011). A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial and Spatio-Temporal Epidemiology*, 2:79–89, doi:10.1016/j.sste.2011.03.001.
27. Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2013). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Chapman and Hall/CRC, London, UK.
28. Martin, A. D., Quinn, K. M., and Park, J. H. (2011). MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, 42:1–21.
29. Martinez, E. Z. and Achcar, J. A. (2014). Trends in epidemiology in the 21st century: time to adopt Bayesian methods. *Cadernos de Saúde Pública*, 30:703–714, doi:10.1590/0102–311X00144013.

30. Martinez, E. Z. and Roza, D. L. (2020). Ecological analysis of adolescent birth rates in Brazil: Association with Human Development Index. *Women and Birth*, 33:e191–e198, doi:10.1016/j.wombi.2019.04.002.
31. Morais, R. M. and Costa, A. L. (2017). An evaluation of the Brazilian Mortality Information System. *Saúde em Debate*, 41:101–117, doi:10.1590/0103–11042017S09.
32. Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37:17–23, doi:10.2307/2332142.
33. Raei, M., Schmid, V. J., and Mahaki, B. (2018). Bivariate spatiotemporal disease mapping of cancer of the breast and cervix uteri among Iranian women. *Geospatial Health*, 13:645, doi:10.4081/gh.2018.645.
34. Rocha-Brischiliari, S. C., Andrade, L., Nihei, O. K., Brischiliari, A., Hortelan, M. S., Carvalho, M. D. B., and Pelloso, S. M. (2018). Spatial distribution of breast cancer mortality: Socioeconomic disparities and access to treatment in the state of Paraná, Brazil. *PLoS One*, 13:e0205253, doi:10.1371/journal.pone.0205253.
35. Rodeiro, C. L. V. and Lawson, A. B. (2005). An evaluation of the edge effects in disease map modelling. *Computational Statistics & Data Analysis*, 49:45–62, doi:10.1016/j.csda.2004.05.012.
36. Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC, Boca Raton, USA.
37. Silva-Lizzi, E. A., Nunes, A. A., and Martinez, E. Z. (2016). Spatiotemporal analysis of AIDS incidence among adults in Brazil. *Current HIV Research*, 14:466–475, doi:10.2174/1570162x14666160802153025.
38. Stoppa, G., Mensi, C., Fazzo, L., Minelli, G., Manno, V., Consonni, D., Biggeri, A., and Catelan, D. (2022). Spatial analysis of shared risk factors between pleural and ovarian cancer mortality in Lombardy (Italy). *International Journal of Environmental Research and Public Health*, 13:645, doi:10.3390/ijerph19063467.
39. Tulchinsky, T. H. and Varavikova, E. A. (2014). *The New Public Health, Third Edition*. Academic Press, London, UK.
40. Ulm, K. (1990). Simple method to calculate the confidence interval of a standardized mortality ratio (SMR). *American Journal of Epidemiology*, 131:373–37, doi:10.1093/oxfordjournals.aje.a115507.

Appendix

This is the OpenBUGS code for the Bayesian bivariate model with shared components, as presented in subsection 2.3:

```
model {
for (i in 1:n) {
y1[i] ~ dpois(lambda1[i])
y2[i] ~ dpois(lambda2[i])
SMR1[i] <- exp(eta1[i])
SMR2[i] <- exp(eta2[i])
lambda1[i] <- z1[i]*SMR1[i]
lambda2[i] <- z2[i]*SMR2[i]
eta1[i] ~ dnorm(mu1[i],prec.tau1)
eta2[i] ~ dnorm(mu2[i],prec.tau2)
mu1[i] <- alpha1 + u1[i] * delta + u2[i]
mu2[i] <- alpha2 + u1[i]/delta
# Standardized Bayesian residuals:
res1[i] <- (y1[i] - lambda1[i])/sqrt(lambda1[i])
res2[i] <- (y2[i] - lambda2[i])/sqrt(lambda2[i])
}
# Prior distributions:
alpha1 ~ dflat()
alpha2 ~ dflat()
u1[1:n] ~ car.normal(adj[], weights[], num[], prec.v1)
u2[1:n] ~ car.normal(adj[], weights[], num[], prec.v2)
log(delta) <- ldelta
ldelta ~ dnorm(0,prec.delta)
tau1 ~ dgamma(1,0.01)
tau2 ~ dgamma(1,0.01)
```

```
v1 ~ dgamma(1,0.01)
v2 ~ dgamma(1,0.01)
prec.tau1 <- 1/tau1
prec.tau2 <- 1/tau2
prec.v1 <- 1/v1
prec.v2 <- 1/v2
prec.delta <- 1/0.17
for (j in 1:sumNumNeigh) {weights[j] <- 1}
}
```