Alguns Aspectos de Modelos Espaço-Temporais

Letícia Cavalari Pinheiro

DISSERTAÇÃO APRESENTADA AO DEPARTAMENTO DE ESTATÍSTICA
DA UNIVERSIDADE FEDERAL DE MINAS GERAIS PARA
OBTENÇÃO DO TÍTULO DE MESTRE EM ESTATÍSTICA

Programa: Mestrado em Estatística
Orientador: Prof. Dr. Renato Martins Assunção
Co-Orientador: Profa. Dra. Ilka Afonso Reis

Belo Horizonte, dezembro de 2009

Alguns Aspectos de Modelos Espaço-Temporais

Este exemplar corresponde à redação
final da dissertação devidamente corrigida
e defendida por (Letícia Cavalari Pinheiro)
e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Prof. Dr. Renato Martins Assunção (orientador) - UFMG.
- Profa. Dra. Ilka Afonso Reis (co-orientador) - UFMG.
- Profa. Dra. Rosângela Helena Loschi - UFMG.
- Prof. Dr. Valdério Anselmo Reisen - UFES.
- Profa. Dra. Patrícia Klarmann Ziegelmann - UFRGS.

Agradecimentos

Agradeço a minha família por, cada um de sua forma, serem tão especiais. Em primeiro lugar agradeço a minha mãe, que esteve ao meu lado durante toda a trajetória da minha vida até aqui, sempre amenizando o peso das dificuldades, potencialializando as alegrias e valorizando os esforços. Por ela ser o meu exemplo de mãe e de mulher, tão amorosa, companheira, prestativa, disposta... fazendo da minha caminhada muito mais amena e melhor. Ao meu pai, que mesmo estando longe fisicamente sempre participou de todas as etapas, agradeço pelo apoio e amor incondicionais, por ser um grande incentivador e um exemplo de força e persistência. Ao meu irmão Ricardo, por ser o meu melhor amigo, pelos conselhos, pelas longas conversas e pelos momentos de descontração. Ao meu irmão Alexandre, por me ensinar a importância de ter pensamento positivo e de correr atrás dos meus objetivos, por torcer sempre por mim. A minha irmã Isabela, pelo exemplo de superação e força, pela alegria que irradia e pelo carinho constante. Ao Marquinhos, que entrou em minha vida durante essa caminhada e passou a ser tão importante, agradeço pelo incentivo, pela compreensão, pela paciência, pelo sorriso constante e por ser meu melhor refúgio, me proporcionando momentos tão agradáveis.

Agradeço aos meus orientadores por todo o conhecimento que me transmitiram. Ao Renato, por ser minha referência, por ter me dado um voto de confiança e me orientado tão bem nessa jornada, pelo ombro amigo quando precisei, pelo incentivo constante e pelo exemplo profissional. A Ilka, pela disponibilidade, pela organização, pela calma transmitida. A todos os professores com os quais tive contato durante o curso, que sem dúvida foram muito importantes na minha formação. Por eles terem o dom de ensinar e desempenhar tão bem esse papel, agradeço a inspiração. A professora Mariângela e a Val, agradeço pela determinação na obtenção dos dados para o nosso trabalho e pela parceria nessa pesquisa.

Agradeço a todos os meus amigos, que são o meu porto seguro. Não vou citar nomes para não esquecer de ninguém, mas que cada um sinta o meu abraço de agradecimento individualmente. Aos amigos de colégio, que guardo com muito carinho em meu coração e tenho certeza que serão meus amigos para sempre. Aos amigos do mestrado que caminharam junto comigo nesses dois anos e se tornaram muito mais do que simples colegas. Aos amigos do LESTE, pelos conhecimentos compartilhados e por proporcionarem um ambiente de trabalho tão prazeroso. Aos outros amigos que surgiram na minha vida e fizeram dela mais alegre e completa. Saibam todos o quanto são importantes para mim.

Enfim, agradeço a todos que participaram de alguma forma dessa minha conquista. Sem vocês nada disso não seria possível.

Resumo

Este trabalho apresenta diferentes estudos envolvendo dados espaço-temporais. No Capítulo 1, estudamos as matrizes de covariância de modelos bayesianos com efeitos de interação espaço-temporal. Para isso, apresentamos os possíveis tipos de efeitos aleatórios espaciais e temporais em modelos bayesianos e atribuimos a eles distribuições *a priori* comumente utilizadas. Construímos possíveis efeitos aleatórios espaço-temporais a partir da interação entre um efeito temporal e um espacial. Calculamos as matrizes de covariância a priori para os modelos com interação espaço-tempo e as escrevemos na forma de produto de Kronecker entre as matrizes de covariância a priori dos efeitos temporal e espacial. Conseguimos visualizar mais claramente o efeito de cada tipo de interação possível, relacionando as matrizes de covariância *a priori* com as estruturas de dependência espacial e/ou temporal envolvidas nos modelos estudados. Como exemplo, apresentamos o estudo das matrizes de covariância *a priori* de dois modelos específicos existentes na literatura e fazemos suas interpretações.

No Capítulo 2, estudamos métodos bayesianos para dados de área espaço-temporais. Buscamos modelos para serem ajustados às taxas de incidência de Leishmaniose Visceral na cidade de Belo Horizonte, utilizando dados dos anos de 2000 a 2008. Fazemos o ajuste de três modelos distintos aos dados, sendo que um deles segue o procedimento clássico e os outros dois são modelos bayesianos. Comparamos os resultados obtidos e selecionamos o modelo que parece acompanhar melhor a evolução espaço-temporal das taxas de incidência de Leishmanise Visceral em Belo Horizonte. Conseguimos observar a evolução espaço-temporal da doença durante os anos estudados de forma mais clara. A partir desse modelo, há possibilidade de que sejam feitas projeções para os próximos anos. Essas projeções podem ser úteis para classificar áreas prioritárias para ações de combate e prevenção da doença.

No Capítulo 3, utilizamos dados espaço-temporais na forma de padrões pontuais, onde a localização dos eventos é aleatória. Apresentamos a função K_{12}, cujo objetivo é testar independência espacial entre dois processos pontuais estacionários observados dentro de um mesmo polígono. Por exemplo, queremos testar se há independência entre a distribuição geográfica de uma espécie de árvores A em relação à outra espécie B, tendo suas localizações dentro de uma floresta. Isto é, queremos saber se árvores da espécie A tendem a se desenvolver mais próximas de árvores da espécie B, mais distantes destas, ou ainda se as duas distribuições geográficas são independentes uma da outra. Baseados nessa função K_{12}, desenvolvemos outra função semelhante para lidar com dados espaço-temporais e a denominamos Função Kt_{12}. Apresentamos a definição matemática, o algoritmo e os testes realizados em dados gerados computacionalmente. Apresentamos ainda uma aplicação a dados reais de Leishmaniose Visceral na cidade de Belo Horizonte.

Palavras-chave: dados espaço-temporais, modelos bayesianos espaço-temporais, matriz de covariância *a priori*, teste de independência espaço-temporal, Função K.

Abstract

This theses presents different types of studies involving spatio-temporal data. In Chapter 1, we study the covariance matrices of Bayesian models with effects of space-time interaction. For that purpose, we present the possible types of spatial and temporal random effects and assign generally used *prior* distributions for them. We introduce the possible space-time random effects, constructed from the interaction between a temporal and a spatial effect. We calculate the *prior* covariance matrices for the models with space-time interaction and write them in the form of Kronecker product between the *prior* covariance matrices of the a temporal and a spatial effects. We see more clearly the effect of each type of interaction, relating the *prior* covariance matrix with the structure of spatial and/or temporal dependence involved in the models studied. As an example, we present the study of the *prior* covariance matrices of two specific models in the literature and make their interpretations.

In Chapter 2, we study Bayesian methods for space-time areal data. We look for models to be adjusted to the incidence rates of Visceral Leishmaniasis in Belo Horizonte, using data from 2000 to 2008. We fit three different models to the data, one of them follows the classic inference procedure and the other two are Bayesian models. We compare the results and select the model that seems to represent better the spatio-temporal evolution of incidence rates of Visceral Leishmaniasis in Belo Horizonte. We observe more clearly the spatio-temporal evolution of the disease during the years studied. From this model, there is possibility to make projections for the coming years. These projections may be useful to classify priority areas for action and prevention of disease.

In Chapter 3, we use spatio-temporal data in the form of point patterns, where the events location is random. We introduce the K_{12} function, which objective is to test spatial independence between two stationary point processes observed within the same polygon. For example, we want to test independence between the geographic distribution of specie A of trees in relation to other specie B, having its locations in a forest. That is, we want to know if the trees of specie A tend to grow closer to trees of specie B, more distant of these, or whether the two geographic distributions are independent of each other. Based on this function K_{12}, we have developed a similar one to deal with spatio-temporal data and called it function Kt_{12}. We present the mathematical definition, the algorithm and tests based on computationally generated data. We also present an application to real data of Visceral Leishmaniasis in Belo Horizonte.

Palavras-chave: spatio-time, spatio-time Bayesian models, *prior* covariance matrix, spatio-time independence test, K-Function.

Sumário

Lista de Figuras

Capítulo 1

Estudo das Matrizes de Covariância de Modelos Bayesianos com Interação Espaço-Temporal

1.1. Introdução

O estudo de métodos estatísticos para análise de dados espaciais é importante em diversas áreas, tais como ecologia, epidemiologia, demografia, geografia, entre outros. Os dados espaciais contém, além de valores de possíveis atributos de interesse, as localizações espaciais relativas às observações.

Os dados espaciais envolvidos em um problema podem ser divididos em quatro categorias [3]: dados em forma de uma configuração espacial de pontos (dados pontuais), dados espacialmente contínuos, dados de área e dados de interação espacial. Neste trabalho utilizamos os dados de área, que serão descritos a seguir.

Ao tratar com dados de área, as observações estão associadas a determinadas unidades de área, como setores censitários, municípios, microrregiões, etc. Cada observação é a realização de uma variável aleatória. O valor é associado à área como um todo e não a um ponto particular dentro dela. Suponha que a região de estudo seja particionada em N áreas. É feita uma observação do atributo de interesse para cada uma delas. Em diversos estudos epidemiológicos é frequente a observação dos eventos espaciais em diferentes períodos de tempo, de forma a obter dados espaço-temporais. Nesse caso, os dados estão associados não só à área, mas também ao tempo em que foram observados.

Modelos estatísticos espaço-temporais para mapeamento de doenças utilizando dados de área se tornaram muito populares em epidemiologia. Frequentemente, o atributo de interesse em estudos deste tipo é o número de casos de uma determinada doença, e o objetivo dos modelos é estimar o risco relativo da doença em cada área e período de tempo. O problema é que o número de casos e a população em risco para uma única área em um único período de tempo costumam ser muito pequenos para produzir uma estimativa do risco próxima da realidade. Portanto, pode ser interessante utilizar informações das áreas vizinhas para estimar o risco relativo. Muitos trabalhos de pesquisa foram desenvolvidos nessa área, principalmente na década passada [1, 4, 3, 6, 7], sendo que modelos bayesianos foram frequentemente adotados para suavizar a distribuição dos riscos

relativos nas áreas.

Knorr-Held [6] propõe uma estrutura unificada para a análise Bayesiana de dados no espaço e no tempo. São apresentados quatro tipos diferentes de distribuições *a priori* para a interação espaço-tempo. Cada um deles implica em um certo grau de dependência *a priori* para os parâmetros de interação e corresponde ao produto de um efeito espacial com um temporal.

Um caso particular desta estrutura proposta foi apresentado recentemente por Martínez-Beneito et al. [7], que propõem uma abordagem espaço-temporal para mapeamento de doenças, combinando idéias de séries temporais e de modelagem espacial. Dessa forma, eles buscam construir um modelo que define uma estrutura espaço-temporal na qual os riscos relativos são espacial e temporalmente dependentes, ao mesmo tempo. Os autores definem a covariância *a priori* como um produto de Kronecker [13] das matrizes de covariância temporal e espacial. A covariância *a posteriori* é de difícil cálculo e interpretação devido à complexidade das operações matriciais envolvidas.

Faremos uma extensão de um resultado para modelos com efeitos espaciais [1], obtido a partir do estudo das estruturas de covariância *a priori* e *a posteriori* de modelos bayesianos com verossimilhança seguindo uma distribuição normal multivariada, e com distribuição *a priori* Condicional Auto-Regressiva (CAR) [2] atribuída ao efeito espacialmente estruturado. Utilizando resultados de álgebra linear, os autores verificaram que a covariância entre duas áreas do mapa é proporcional à soma ponderada das chances de passar de uma área para a outra em k passos (com $k = 1, 2, 3, ...$), considerando um passeio aleatório em um grafo definido pela estrutura de vizinhança geográfica.

Neste trabalho, apresentamos uma análise das matrizes de covariância *a priori* de modelos bayesianos espaço-temporais. Com a utilização da estrutura unificada proposta por Knorr-Held [6], construímos uma estrutura para as matrizes de covariância *a priori* de cada tipo especificado de interação espaço-temporal. Faremos interpretações dos resultados obtidos, objetivando associar as estruturas de covariância *a priori* à estrutura de dependência espaço-tempo atribuida ao modelo. Como casos particulares, apresentaremos as análises das matrizes de covariância *a priori* dos modelos de Assunção et al [1] e de Martínez-Beneito at al. [7].

Apresentaremos, na Seção 2, as metodologias que foram utilizadas como base para este trabalho. Na Seção 3, descreveremos as estruturas de covariância *a priori* para modelos com os diferentes tipos de interação espaço-tempo. Na Seção 4, mostraremos o caso particular da matriz de covariância do modelo proposto por Martínez-Beneito et al.[7] e do modelo proposto por Assunção et al. [1]. E finalmente, na Seção 5, apresentaremos as conclusões sobre o trabalho realizado e as propostas de alguns trabalhos futuros.

1.2. Metodologia

1.2.1. O Modelo CAR

Seja uma região **D**, particionada em N áreas disjuntas, $A_1, ..., A_N$, ou seja, $A_i \cap A_j = \emptyset$ e $\cup_{i=1}^{N} A_i = \mathbf{D}$. Por exemplo, as 140 áreas de abrangência dos centros de saúde de Belo Horizonte formam uma partição da cidade. Seja y_i o valor observado de um determinado fenômeno na área A_i . O interesse é modelar o processo estocástico $Y(A_i), i = 1, ..., N$ ou simplesmente $\mathbf{y} = (y_1, ..., y_N)$. Utilizaremos informações provenientes das observações de áreas vizinhas para prever o comportamento de cada área. Um modelo muito utilizado para isso é o modelo autoregressivo condicional - CAR [2].

O modelo denominado *Conditional AutoRegressive* - CAR - é especificado por um conjunto de distribuições condicionais, no qual se propõe que a observação de uma área tem distribuição gaussiana com o parâmetro de média sendo a média ponderada das observações das áreas vizinhas e a variância sendo inversamente proporcional ao número de áreas vizinhas.

O modelo CAR é determinado por um conjunto de distribuições condicionais

$$y_i|y_{-i} \sim N(\mu_i + \sum_{j \neq i} \rho \mathbf{W}_{ij}(y_j - \mu_j), \kappa_i^2)$$

onde $\kappa_i^2 > 0, i = 1, \ldots, N$ e y_{-i} são os valores de **y** em todas as áreas do mapa, exceto a área i. Seja $\mathbf{A} = (a_{ij})$ uma matriz de vizinhança $N \times N$, tal que $a_{ij} = 1$ quando as áreas i e j são vizinhas e $a_{ij} = 0$, caso contrário. Por definição, $a_{ii} = 0$. Define-se a matriz $\mathbf{W} = (w_{ij})$ de forma que $w_{ij} = a_{ij}/d_i$, onde $d_i = \sum_j a_{ij}$ representa o número de vizinhos da área i. Observamos que cada área deve ter ao menos uma vizinha, de forma que d_i é sempre maior que 0, portanto ilhas não são consideradas. ρ é o parâmetro de correlação espacial do modelo CAR, e $\kappa_i^2 = \sigma^2/d_i$ é a variância *a priori* da área i. Se $|\rho| < 1$, o modelo CAR construído dessa maneira define uma distribuição conjunta válida para o vetor **y** dada por uma distribuição normal multivariada:

$$\mathbf{y} \sim N(\mu, (\mathbf{I}_N - \rho \mathbf{W})^{-1} \mathbf{T}^{-1})$$

em que $\mathbf{T}^{-1} = (\sigma^2)^{-1} diag\{d_1, ..., d_N\}$.

1.2.2. Matriz de covariância *a priori* da distribuição *a priori* CAR

Primeiramente, iremos utilizar um resultado importante de álgebra linear. Esse resultado diz que se **M** é uma matriz quadrada tal que cada elemento da matriz $\mathbf{M}^k$ tende a zero quando k aumenta, então a inversa $(\mathbf{I} - \mathbf{M})^{-1}$ existe [10] e é dada por

$$(\mathbf{I} - \mathbf{M})^{-1} = \mathbf{I} + \mathbf{M} + \mathbf{M}^2 + \mathbf{M}^3 + \ldots \tag{1.1}$$

Vimos na seção anterior que a matriz de covariância do modelo CAR é dada por $(\mathbf{I} - \rho \mathbf{W})^{-1} T^{-1}$. Então, seja $\mathbf{M} = \rho \mathbf{W}$ onde $|\rho| < 1$. Se $0 \leq [\mathbf{W}^k]_{ij} \leq 1$ para todo i, j e para todo inteiro k, podemos escrever

$$(\mathbf{I} - \rho\mathbf{W})^{-1} = \mathbf{I} + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \rho^3\mathbf{W}^3 + \dots \tag{1.2}$$

Para que esse resultado seja válido, precisamos garantir que $0 \leq [\mathbf{W}^k]_{ij} \leq 1$ para todo i, j e para todo inteiro k. Iremos recorrer a um resultado de teoria de grafos. Repare que podemos definir uma cadeia de Markov finita e discreta com matriz de transição dada por $\mathbf{W}$. Dessa forma, cada área A_i representa um nó de um grafo e há uma aresta conectando dois nós (ou duas áreas i e j) quando as áreas são vizinhas, ou seja, quando $w_{ij} \neq 0$. Observando a matriz $\mathbf{W}$, percebemos que as probabilidades de transição de um nó i para qualquer outro que seja seu vizinho, em um único passo, são iguais. Portanto, definimos um modelo markoviano chamado de passeio aleatório e $\mathbf{W}^k$ é a matriz de transição para os movimentos da cadeia em k passos. Esse passeio aleatório converge para uma única distribuição estacionária se a matriz $\mathbf{W}$ obedece algumas regras. Para que isso ocorra o grafo deve ser aperiódico e também conectado, ou seja, de cada nó existe um caminho de arestas conectando nós sucessivos até que qualquer outro nó escolhido arbitrariamente seja atingido.

A distribuição estacionária da cadeia de Markov definida por $\mathbf{W}$ (matriz de adjacência normalizada como definida anteriormente) é dada por $\pi = (\pi_1, ..., \pi_N)$ onde $\pi_i = d_i/D$, sendo d_i o número de áreas vizinhas à área i e $D = \sum_i d_i$ [7]. Isso implica que $[\mathbf{W}^k]_{ij} \to d_j/D$, quando $k \to \infty$.

Verificamos então que a equação (1.2) é verdadeira para o modelo definido. Dessa forma, com uma boa aproximação e para algum k, temos que

$$\begin{aligned} [(\mathbf{I} - \rho\mathbf{W})^{-1}]_{ij} &\approx [\mathbf{I}]_{ij} + \rho[\mathbf{W}]_{ij} + \dots + \rho^{k-1}[\mathbf{W}^{k-1}]_{ij} + \frac{d_j\rho^k}{D(1-\rho)} && (1.3) \\ &\approx [\mathbf{I}]_{ij} + \rho[\mathbf{W}]_{ij} + \dots + \rho^{k-1}[\mathbf{W}^{k-1}]_{ij} && (1.4) \end{aligned}$$

Sabemos que se $[\mathbf{W}^k]_{ij} > 0$, então existe ao menos uma sequência de k arestas (ou um caminho com k passos) no grafo, tal que o nó inicial é o i e o nó final é o j. Assim, o valor de $[\mathbf{W}^k]_{ij}$ é uma soma ponderada de todos os caminhos com k passos entre i e j. Por exemplo:

$$[\mathbf{W}^2]_{ij} = \sum_{k=1}^{N} W_{ik}W_{kj} = \sum_{k=1}^{N} \frac{a_{ik}}{d_i}\frac{a_{kj}}{d_k} = \frac{1}{d_i}\sum_{k=1}^{N} \frac{a_{ik}a_{kj}}{d_k} \ .$$

O produto binário $a_{ik}a_{kj}$ será igual a 1 se k é ligado a i e também a j. Assim, $[\mathbf{W}^2]_{ij}$ é proporcional a soma ponderada de todos os caminhos em 2 passos $i \to k \to j$. Observamos que, quanto maior o número de vizinhos da área k, menor será o peso dado ao caminho $i \to k \to j$.

O mesmo raciocínio se aplica para todas as potências da matriz $\mathbf{W}$. Por exemplo, se a aproximação de terceiro grau em (1.4) for suficiente, teremos que a covariância do modelo CAR entre as áreas i e j é dada aproximadamente por

$$\frac{\sigma^2}{d_j}\left([\mathbf{I}]_{ij} + \frac{\rho a_{ij}}{d_i} + \frac{\rho^2}{d_i}\sum_{k=1}^{N}\frac{a_{ik}a_{kj}}{d_k} + \frac{\rho^3}{d_i}\sum_{l=1}^{N}\sum_{k=1}^{N}\frac{a_{ik}a_{kl}a_{lj}}{d_l d_k}\right)$$

Logo, a covariância entre as duas áreas i e j é dada aproximadamente por σ^2/d_j multiplicando a soma ponderada de todos os caminhos possíveis entre a área i e a área j em 1, 2 e 3 passos.

1.2.3. Estrutura geral dos modelos com interação espaço-tempo

Seja Y_{it} o número de eventos na área i e no período de tempo t, $i = 1, \ldots, N$, $t = 1, \ldots, T$, E_{it} o número esperado de eventos para a área i no tempo t sob a hipótese de que o risco é constante no espaço e no tempo. Assume-se que Y_{it} segue uma distribuição de Poisson:

$$Y_{it} \sim Poisson(E_{it} \exp(r_{it}))$$

onde $\exp(r_{it})$ representa o risco relativo referente à área i e ao tempo t.

Queremos modelar r_{it}, que é o logaritmo natural do risco relativo. Uma primeira abordagem é supor que não existe interação espaço-temporal. Para isto, assumimos que r_{it} pode ser decomposto da seguinte forma [6]:

$$r_{it} = \mu + \alpha_t + \gamma_t + \theta_i + \phi_i \tag{1.5}$$

onde μ é a média global do risco e α_t e γ_t são efeitos temporais representando características do período de tempo t que, respectivamente, apresentam e não apresentam estrutura temporal *a priori*. Da mesma maneira, θ_i e ϕ_i representam características da área i que, respectivamente, apresentam e não apresentam estrutura espacial *a priori*.

Precisamos definir as distribuições *a priori* de cada componente de (1.5). Para μ foi escolhida uma distribuição *a priori* não informativa. A cada um dos blocos $\alpha = (\alpha_1, \ldots, \alpha_T), \gamma = (\gamma_1, \ldots, \gamma_T), \theta = (\theta_1, \ldots, \theta_N), \phi = (\phi_1, \ldots, \phi_N)$ é atribuida uma distribuição normal multivariada com média zero e matriz de precisão $\lambda\mathbf{K}$, onde λ é um escalar desconhecido a ser estimado a partir dos dados e $\mathbf{K}$ é uma matriz de estrutura conhecida. A matriz $\mathbf{K}$ será diferente para cada bloco, com intuito de descrever diferentes suposições sobre a relação *a priori* entre os parâmetros dentro do bloco. Daqui em diante chamaremos a matriz $\mathbf{K}$ de matriz de precisão, já que os valores de λ não nos interessam neste estudo, apenas as diferenças relativas entre os valores da matriz $\lambda\mathbf{K}$.

- Para α, adota-se uma distribuição *a priori* na qual os efeitos dos períodos de tempo vizinhos são levados em conta. Uma escolha simples é a de um processo auto-regressivo de ordem 1, cuja matriz de precisão $\mathbf{K}_\alpha$ é tal que [12]

$$\mathbf{K}_\alpha = \begin{bmatrix} 1 & -\rho_\alpha & & & \\ -\rho_\alpha & 1+\rho_\alpha & -\rho_\alpha & & \\ & \ddots & \ddots & \ddots & \\ & & -\rho_\alpha & 1+\rho_\alpha & -\rho_\alpha \\ & & & -\rho_\alpha & 1 \end{bmatrix}$$

- Para γ, adota-se uma distribuição *a priori* na qual o efeito dos períodos de tempo vizinhos não importam. Ou seja, o efeito γ não depende da estrutura de vizinhança temporal. Portanto,

$$\mathbf{K}_\gamma = \mathbf{I}$$

onde $\mathbf{I}$ é a matriz identidade.

- Para o efeito espacialmente estruturado θ atribui-se uma distribuição *a priori* CAR

$$(\theta|\lambda_\theta) \sim N(\mu, \lambda_\theta^{-1}(\mathbf{I}_N - \rho_\theta \mathbf{W})^{-1}\mathbf{T}^{-1})$$

com matriz de precisão $\mathbf{K}_\theta = \mathbf{T}(\mathbf{I} - \rho_\theta \mathbf{W})$, com as definições dadas nas seções anteriores. Logo, $\mathbf{K}_\theta$ tem os elementos da diagonal k_{ii} iguais ao número de vizinhos da área i, e os elementos fora da diagonal, k_{ij}, iguais a -1 quando as áreas i e j são vizinhas e iguais a 0 caso contrário.

- Para o efeito espacial não estruturado ϕ, a distribuição adotada também implica ausência de correlação entre as áreas. Dessa forma, temos

$$\mathbf{K}_\phi = \mathbf{I}$$

onde $\mathbf{I}$ é a matriz identidade.

A Figura 1.1 mostra um esquema representando os efeitos descritos.

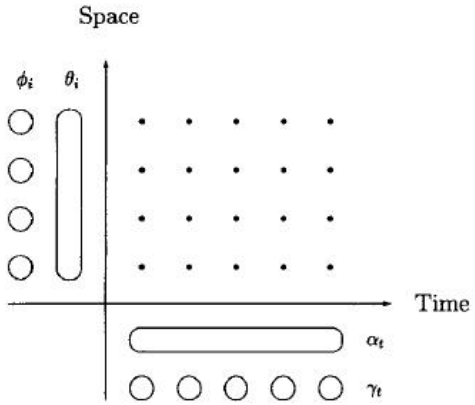


Figura 1.1. Representação simbólica dos efeitos principais. Círculos representam independência *a priori* e ovais representam dependência *a priori*. As observações no espaço-tempo são indicadas pelos pontos.

Finalmente, além dos efeitos individuais espaciais ou temporais, gostaríamos de modelar efeitos nos quais há interação entre espaço e tempo. Dessa forma, adiciona-se parâmetros δ_{it} ao modelo descrito na equação (1.5), obtendo:

$$r_{it} = \mu + \alpha_t + \gamma_t + \theta_i + \phi_i + \delta_{it} \tag{1.6}$$

onde o vetor de parâmetros $\delta = (\delta_{11}, \ldots, \delta_{NT})$ assume uma distribuição *a priori* gaussiana com matriz de precisão $\lambda_\delta \mathbf{K}_\delta$, onde λ_δ é um escalar desconhecido a ser estimado a partir dos dados e $\mathbf{K}_\delta$ é uma matriz de estrutura conhecida. Podemos perceber que se todos os $\delta_{it} = 0$, então a equação (1.6) se reduz a (1.5). Dessa forma, o efeito de interação espaço-temporal só irá capturar a variação que não pode ser explicada pelos efeitos principais separadamente.

A partir das definições dadas, temos quatro possibilidades de interações espaço-tempo, dependendo de qual dos dois efeitos temporais interage com qual dos dois efeitos espaciais. Esses quatro tipos de interações, representados na Figura 1.2 implicam em diferentes relações *a priori* entre os δ_{it}. Na próxima seção será discutido cada um dos tipos de interação, apresentadas as matrizes de precisão e calculadas as covariâncias *a priori* de cada um deles.

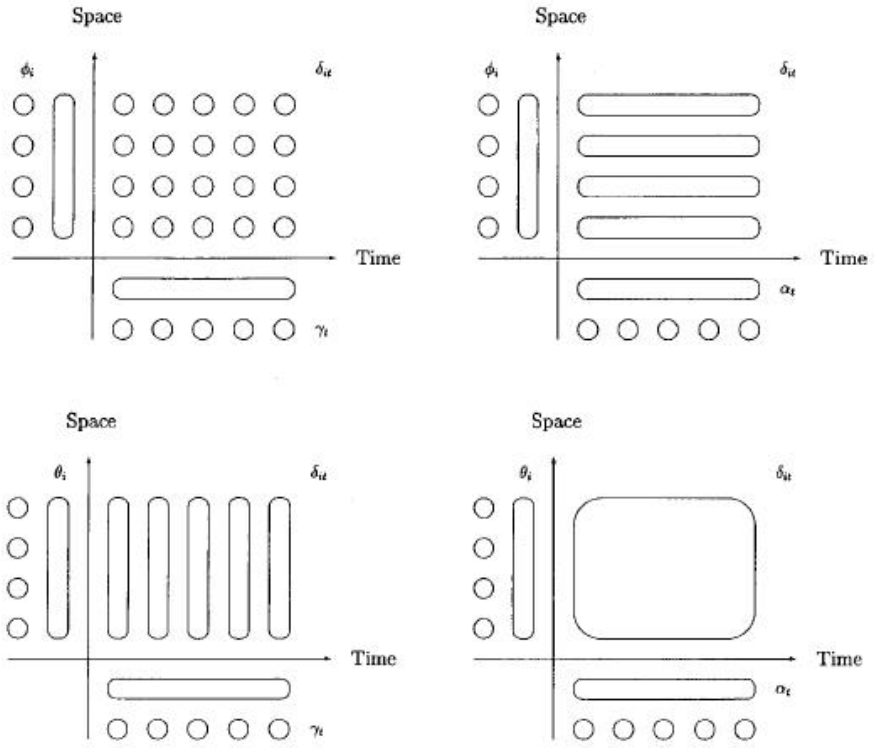


Figura 1.2. Representação simbólica dos quatro tipos de interação espaço-tempo possíveis no modelo. Círculos representam independência *a priori* e ovais representam dependêcia *a priori*.

1.3. Matriz de covariância *a priori* de modelos com interação espaço-tempo

Nesta seção prentendemos apresentar as possíveis interações espaço-tempo e fazer o estudo de suas matrizes de precisão e de covariância *a priori*. Clayton [8] sugere que a matriz de precisão $\mathbf{K}_\delta$ seja especificada pelo produto de Kronecker das matrizes de precisão dos dois efeitos principais que estão interagindo. Utilizaremos esse resultado juntamente com algumas propriedade do produto de Kronecker [13]. A partir do modelo definido na seção anterior, define-se as interações espaço-tempo [6], que podem aparecer como resultado de interações entre $\alpha, \gamma, \phi, \theta$, como descrevemos a seguir.

1.3.1. Interação entre os efeitos não estruturados espacial ϕ e temporal γ

Neste caso, temos que, *a priori*, os parâmetros de interação espaço-tempo δ_{it} são independentes no espaço e no tempo. A matriz de precisão é dada por

$$\mathbf{K}_\delta = \mathbf{K}_\phi \otimes \mathbf{K}_\gamma = \mathbf{I}_{N\times N} \otimes \mathbf{I}_{T\times T} = \mathbf{I}_N T$$

A matriz de covariância é dada pelo inverso da matriz de precisão, logo ela é

$$\mathbf{K}_\delta^{-1} = (\mathbf{K}_\phi \otimes \mathbf{K}_\gamma)^{-1} = \mathbf{K}_\phi^{-1} \otimes \mathbf{K}_\gamma^{-1} = \mathbf{I}_{N\times N} \otimes \mathbf{I}_{T\times T} = \mathbf{I}_N T$$

Portanto a covariância entre áreas distintas é sempre nula, seja no mesmo período de tempo ou em períodos de tempo distintos. A covariância de uma área em um certo período de tempo com ela mesma em outro período de tempo também é nula. Os únicos valores não nulos na matriz são os que representam a variância de cada área em cada período de tempo (os elementos da diagonal). Resumindo, para duas áreas quaisquer i e j em dois períodos de tempo t e v:

$$cov(\delta_{it}, \delta_{jv}) \propto \begin{cases} 0 & \text{se} \quad i \neq j \\ 0 & \text{se} \quad t \neq v \\ 1 & \text{se} \quad i = j \text{ e } t = v \end{cases}$$

Isto significa que, além da estrutura puramente espacial, representada por $\theta + \phi$, e da estrutura puramente temporal, representada por $\alpha + \gamma$, temos um efeito aleatório, do tipo ruído branco, em cada área i e tempo t. Esses efeitos δ_{it} são não correlacionados no espaço e no tempo.

1.3.2. Interação entre o efeito temporal estruturado α e o efeito espacial não estruturado ϕ

Seja $\delta_i = (\delta_{i1}, \ldots, \delta_{iT})$ a série temporal de δ_{it} para uma área fixa i. Neste modelo de interação, nós assumimos que, *a priori*, cada $\delta_i = (\delta_{i1}, \ldots, \delta_{iT})$ representa um passeio aleatório independente das outras áreas do mapa. A matriz de precisão é dada por

$$\mathbf{K}_\delta = \mathbf{K}_\alpha \otimes \mathbf{K}_\phi = \mathbf{K}_\alpha \otimes \mathbf{I}_N = \begin{bmatrix} \mathbf{I}_N & -\rho_\alpha \mathbf{I}_N & & & \\ -\rho_\alpha \mathbf{I}_N & (1+\rho_\alpha)\mathbf{I}_N & -\rho_\alpha \mathbf{I}_N & & \\ & \ddots & \ddots & \ddots & \\ & & -\rho_\alpha \mathbf{I}_N & (1+\rho_\alpha)\mathbf{I}_N & -\rho_\alpha \mathbf{I}_N \\ & & & -\rho_\alpha \mathbf{I}_N & \mathbf{I}_N \end{bmatrix}$$

de forma que $\mathbf{K}_\delta$ tem dimensão $NT \times NT$.

A matriz de covariância é dada pelo inverso da matriz de precisão, portanto [12]

$$\mathbf{K}_\delta^{-1} = (\mathbf{K}_\alpha \otimes \mathbf{K}_\phi)^{-1} = \mathbf{K}_\alpha^{-1} \otimes \mathbf{K}_\phi^{-1} = \frac{1}{1-\rho_\alpha^2} \begin{bmatrix} \mathbf{I}_N & \rho_\alpha \mathbf{I}_N & \rho_\alpha^2 \mathbf{I}_N & \ldots & \rho_\alpha^{T-1} \mathbf{I}_N \\ \rho_\alpha \mathbf{I}_N & \mathbf{I}_N & \rho_\alpha \mathbf{I}_N & \ldots & \rho_\alpha^{T-2} \mathbf{I}_N \\ & \ddots & \ddots & \ddots & \\ \rho_\alpha^{T-2} \mathbf{I}_N & \rho_\alpha^{T-3} \mathbf{I}_N & \ldots & \mathbf{I}_N & \rho_\alpha \mathbf{I}_N \\ \rho_\alpha^{T-1} \mathbf{I} & \rho_\alpha^{T-2} \mathbf{I} & \ldots & \rho_\alpha \mathbf{I} & \mathbf{I} \end{bmatrix}$$

Nesse caso, temos que a covariância entre áreas distintas é nula, seja no mesmo período de tempo ou em períodos de tempo distintos. Já para a mesma área em períodos

de tempo distintos, digamos t e v, a covariância é proporcional a $\rho_\alpha^{|t-v|}$. Resumindo, para duas áreas quaisquer i e j em dois períodos de tempo t e v:

$$cov(\delta_{it}, \delta_{jv}) \propto \begin{cases} 0 & \text{se} \quad i \neq j \\ \rho_\alpha^{|t-v|} & \text{se} \quad i = j \end{cases}$$

1.3.3. Interação entre o efeito temporal não estruturado γ e o efeito espacial estruturado θ

Seja $\delta_t = (\delta_{1t}, \ldots, \delta_{Nt})$ o mapa dos efeitos δ_{it} num instante de tempo t fixo. Neste outro modelo de interação espaço-tempo, nós assumimos que, *a priori*, cada $\delta_t = (\delta_{1t}, \ldots, \delta_{Nt})$ representa um modelo CAR independente do que ocorre nos outros períodos de tempo. Esse modelo é razoável quando as tendências espaciais são completamente independentes a cada período de tempo, ou seja, não há nenhuma dependência temporal no efeito de interação δ_{it}. A matriz de precisão é dada por

$$\mathbf{K}_\delta = \mathbf{K}_\gamma \otimes \mathbf{K}_\theta = \mathbf{I}_T \otimes \mathbf{T}(\mathbf{I}_N - \rho_\theta \mathbf{W}) = \begin{bmatrix} \mathbf{T}(\mathbf{I} - \rho_\theta \mathbf{W}) & 0 & \ldots & 0 \\ 0 & \mathbf{T}(\mathbf{I} - \rho_\theta \mathbf{W}) & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & 0 & \mathbf{T}(\mathbf{I} - \rho_\theta \mathbf{W}) \end{bmatrix}$$

A matriz de covariância é dada pelo inverso da matriz de precisão. Para termos uma interpretação mais clara, usaremos o resultado da seção (2.2) para inverter a matriz $(\mathbf{I} - \rho_\theta \mathbf{W})$. Então, temos que:

$$\begin{aligned} \mathbf{K}_\delta^{-1} = (\mathbf{K}_\gamma \otimes \mathbf{K}_\theta)^{-1} &= \mathbf{K}_\gamma^{-1} \otimes \mathbf{K}_\theta^{-1} = \mathbf{I}_T \otimes (\mathbf{I}_N - \rho_\theta \mathbf{W})^{-1} \mathbf{T}^{-1} \\ &= \mathbf{I}_T \otimes [\mathbf{I}_N + \rho_\theta \mathbf{W} + \rho_\theta^2 \mathbf{W}^2 + \rho_\theta^3 \mathbf{W}^3 + \ldots] \mathbf{T}^{-1} \end{aligned}$$

Representando $[\mathbf{I} + \rho_\theta \mathbf{W} + \rho_\theta^2 \mathbf{W}^2 + \rho_\theta^3 \mathbf{W}^3 + \ldots]$ por $\mathbf{X}^{-1}$, teremos:

$$\mathbf{K}_\delta^{-1} = \begin{bmatrix} \mathbf{X}^{-1}\mathbf{T}^{-1} & 0 & \ldots & 0 \\ 0 & \mathbf{X}^{-1}\mathbf{T}^{-1} & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & 0 & \mathbf{X}^{-1}\mathbf{T}^{-1} \end{bmatrix}$$

Nesse caso, temos que a covariância de uma área com outra área ou com ela mesma em períodos distintos é nula. Já em um mesmo período de tempo, a covariância entre duas áreas i e j quaisquer é proporcional à soma ponderada de todos os caminhos possíveis entre a área i e a área j em 1, 2, 3, ... passos.

Supondo que a aproximação de grau três seja suficiente, podemos resumir assim:

$$cov(\delta_{it}, \delta_{jv}) \propto \begin{cases} 0 & \text{se} \quad t \neq v \\ \frac{\sigma^2}{d_j} \left([\mathbf{I}]_{ij} + \frac{\rho_\theta a_{ij}}{d_i} + \frac{\rho_\theta^2}{d_i} \sum_{k=1}^N \frac{a_{ik} a_{kj}}{d_k} + \frac{\rho_\theta^3}{d_i} \sum_{l=1}^N \sum_{k=1}^N \frac{a_{ik} a_{kl} a_{lj}}{d_l d_k} \right) & \text{se} \quad t = v \end{cases}$$

1.3.4. Interação entre os efeitos estruturados temporal α e espacial θ

Para o estudo realizado, essa é a forma de interação espaço-tempo mais complexa e mais interessante. Neste modelo, o efeito espacial evolui no tempo de forma que o mapa $\delta_t = (\delta_{1t}, \ldots, \delta_{Nt})$ está correlacionado com o mapa δ_{t+1} no instante $t+1$. Isto é, os δ_{it} são agora dependentes no espaço e no tempo e não podem ser decompostos em blocos independentes. Agora, os δ_{it} de áreas vizinhas de uma deteminada área i são levados em conta tanto aqueles do mesmo período de tempo quanto aqueles de períodos de tempo vizinhos para prever o valor da área i.

Nesse caso, a matriz de precisão *a priori* é dada por

$$\mathbf{K}_\delta = \mathbf{K}_\alpha \otimes \mathbf{K}_\theta = \begin{bmatrix} 1 & -\rho_\alpha & & & \\ -\rho_\alpha & (1+\rho_\alpha) & -\rho_\alpha & & \\ & \ddots & \ddots & \ddots & \\ & & -\rho_\alpha & (1+\rho_\alpha) & -\rho_\alpha \\ & & & -\rho_\alpha & 1 \end{bmatrix} \otimes \mathbf{T}(\mathbf{I} - \rho_\theta \mathbf{W})$$

$$= \begin{bmatrix} \mathbf{T}(\mathbf{I}-\rho_\theta\mathbf{W}) & -\rho_\alpha\mathbf{T}(\mathbf{I}-\rho_\theta\mathbf{W}) & & & \\ -\rho_\alpha\mathbf{T}(\mathbf{I}-\rho_\theta\mathbf{W}) & (1+\rho_\alpha)\mathbf{T}(\mathbf{I}-\rho_\theta\mathbf{W}) & -\rho_\alpha\mathbf{T}(\mathbf{I}-\rho_\theta\mathbf{W}) & & \\ & \ddots & \ddots & \ddots & \\ & -\rho_\alpha\mathbf{T}(\mathbf{I}-\rho_\theta\mathbf{W}) & (1+\rho_\alpha)\mathbf{T}(\mathbf{I}-\rho_\theta\mathbf{W}) & -\rho_\alpha\mathbf{T}(\mathbf{I}-\rho_\theta\mathbf{W}) & \\ & & -\rho_\alpha\mathbf{T}(\mathbf{I}-\rho_\theta\mathbf{W}) & \mathbf{T}(\mathbf{I}-\rho_\theta\mathbf{W}) \end{bmatrix}$$

A matriz de covariância é dada pelo inverso da matriz de precisão. Para termos uma interpretação mais clara, usaremos o resultado da seção 1.2.2 para inverter a matriz $(\mathbf{I} - \rho_\theta \mathbf{W})$. Para inverter a matriz que representa a parte temporal, usaremos o resultado já visto em um caso anterior neste trabalho e mostrado em [12]. Então, temos que:

$$\mathbf{K}_\delta^{-1} = (\mathbf{K}_\alpha \otimes \mathbf{K}_\theta)^{-1} = \mathbf{K}_\alpha^{-1} \otimes \mathbf{K}_\theta^{-1} =$$

$$= \frac{1}{1-\rho_\alpha^2} \begin{bmatrix} 1 & \rho_\alpha & \rho_\alpha^2 & \ldots & \rho_\alpha^{T-1} \\ \rho_\alpha & 1 & \rho_\alpha & \ldots & \rho_\alpha^{T-2} \\ & \ddots & \ddots & \ddots & \\ \rho_\alpha^{T-2} & \rho_\alpha^{T-3} & \ldots & 1 & \rho_\alpha \\ \rho_\alpha^{T-1} & \rho_\alpha^{T-2} & \ldots & \rho_\alpha & 1 \end{bmatrix} \otimes [\mathbf{I} + \rho_\theta\mathbf{W} + \rho_\theta^2\mathbf{W}^2 + \rho_\theta^3\mathbf{W}^3 + \ldots]\mathbf{T}^{-1}$$

Representando $[\mathbf{I} + \rho_\theta\mathbf{W} + \rho_\theta^2\mathbf{W}^2 + \rho_\theta^3\mathbf{W}^3 + \ldots]$ por $\mathbf{X}^{-1}$, teremos:

$$\mathbf{K}_\delta^{-1} = \frac{1}{1-\rho_\alpha^2} \begin{bmatrix} \mathbf{X}^{-1}\mathbf{T}^{-1} & \rho_\alpha\mathbf{X}^{-1}\mathbf{T}^{-1} & \rho_\alpha^2\mathbf{X}^{-1}\mathbf{T}^{-1} & \ldots & \rho_\alpha^{T-1}\mathbf{X}^{-1}\mathbf{T}^{-1} \\ \rho_\alpha\mathbf{X}^{-1}\mathbf{T}^{-1} & \mathbf{X}^{-1}\mathbf{T}^{-1} & \rho_\alpha\mathbf{X}^{-1}\mathbf{T}^{-1} & \ldots & \rho_\alpha^{T-2}\mathbf{X}^{-1}\mathbf{T}^{-1} \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \rho_\alpha^{T-2}\mathbf{X}^{-1}\mathbf{T}^{-1} & \rho_\alpha^{T-3}\mathbf{X}^{-1}\mathbf{T}^{-1} & \ldots & \mathbf{X}^{-1}\mathbf{T}^{-1} & \rho_\alpha\mathbf{X}^{-1}\mathbf{T}^{-1} \\ \rho_\alpha^{T-1}\mathbf{X}^{-1}\mathbf{T}^{-1} & \rho_\alpha^{T-2}\mathbf{X}^{-1}\mathbf{T}^{-1} & \ldots & \rho_\alpha\mathbf{X}^{-1}\mathbf{T}^{-1} & \mathbf{X}^{-1}\mathbf{T}^{-1} \end{bmatrix}$$

Nesse caso, em um mesmo período de tempo, a covariância entre duas áreas i e j quaisquer é proporcional à soma ponderada de todos os caminhos possíveis entre a área i e a área j em 1, 2, 3, ... passos. Já em períodos de tempo distintos t e v, a covariância entre as áreas i e j é proporcional a $\rho_\alpha^{|t-v|}$ vezes a soma ponderada de todos os caminhos possíveis entre a área i e a área j em 1, 2, 3, ... passos. Dessa forma, a covariância diminui a medida que os períodos de tempo em questão são mais distantes um do outro, já que $|\rho_\alpha| < 1$, o que faz sentido.

Supondo que a aproximação de grau três seja suficiente, podemos resumir assim: para duas áreas i e j e dois períodos de tempo t e v,

$$\text{cov}(\delta_{it}, \delta_{jv}) \propto \rho_\alpha^{|t-v|} \left[\frac{\sigma^2}{d_j} \left([\mathbf{I}]_{ij} + \frac{\rho_\theta a_{ij}}{d_i} + \frac{\rho_\theta^2}{d_i} \sum_{k=1}^{N} \frac{a_{ik}a_{kj}}{d_k} + \frac{\rho\theta^3}{d_i} \sum_{l=1}^{N} \sum_{k=1}^{N} \frac{a_{ik}a_{kl}a_{lj}}{d_l d_k} \right) \right]$$

1.4. Exemplos de matrizes de covariância *a priori* de modelos específicos

Nesta seção, faremos o estudo da matriz de covariância *a priori* de dois modelos específicos [1, 7].

Para definir os modelos, considere a seguinte notação:

- Y_{it} é o número de casos na área i no período de tempo t.
- r_{it} é o logaritmo natural da taxa de incidência na área i no período t.
- $i = 1, \cdots, N$.
- $t = 1, \cdots, T$.

Assuma que o número de casos de uma determinada doença para cada área e cada período de tempo segue uma distribuição de Poisson:

$$[Y_{it} \mid P_{it}, r_{it}] \sim Poisson(P_{it} \exp(r_{it}))$$

Cada um dos modelos apresentados irá modelar o logaritmo natural da taxa de incidência de uma forma distinta, como apresentado a seguir.

1.4.1. Matriz de covariância *a priori* do modelo proposto por Assunção et al.

O procedimento proposto [1] modela o logaritmo natural das taxas de incidência da doença linearmente em relação ao tempo, da seguinte forma:

$$r_{it} = \alpha_i + \beta_i t$$
$$\alpha = (\alpha_1, \ldots, \alpha_N) \sim CAR(\sigma_\alpha^2)$$
$$\beta = (\beta_1, \ldots, \beta_N) \sim CAR(\sigma_\beta^2)$$

Definimos α como o vetor que contém os α_i's de todas as áreas, representando possíveis características que são espacialmente estruturadas e não apresentam uma tendência

temporal. E definimos β como o vetor que contém os β_i's de todas as áreas, representando possíveis características que são espacialmente estruturadas e variam linearmente em relação ao tempo.

As distribuições de probabilidade *a priori* dos parâmetros de precisão σ_α^{-2} e σ_β^{-2} são definidas de forma a serem vagas, já que não há conhecimento algum sobre eles *a priori*. Elas são dadas então por:

$$\tau_\alpha = \sigma_\alpha^{-2} \sim Gamma(a, b)$$
$$\tau_\beta = \sigma_\beta^{-2} \sim Gamma(c, d)$$

onde a, b, c e d são valores conhecidos.

Para encontrar os termos da matriz de covariâncias do modelo proposto, são realizados os cálculos a seguir:

$$\text{var}(r_t|\sigma_\alpha^2, \sigma_\beta^2) = \text{var}(\alpha + \beta t) = \text{var}(\alpha) + t^2\text{var}(\beta) = \Sigma_\alpha + t^2\Sigma_\beta$$

$$\begin{aligned} \text{cov}(r_t, r_{t+k}|\sigma_\alpha^2, \sigma_\beta^2) &= \text{cov}(\alpha + \beta t, \alpha + \beta(t+k)) \\ &= \text{var}(\alpha) + t(t+k)\text{var}(\beta) = \Sigma_\alpha + t(t+k)\Sigma_\beta \end{aligned}$$

Portanto, a covariância *a priori* entre a área i no tempo t e a área j no tempo v, $i, j = 1, \ldots, N$ e $t, v = 1, \ldots, T$, é dada por

$$\text{cov}(r_{it}, r_{jv}) = [\Sigma_\alpha]_{ij} + tv[\Sigma_\beta]_{ij}$$

Já vimos a forma geral da matriz de covariância de um efeito aleatório com distribuição *a priori* CAR. Assim podemos definir $\Sigma_\alpha = (\mathbf{I} - \rho_\alpha \mathbf{W})^{-1}\mathbf{T}_\alpha^{-1}$ e $\Sigma_\beta = (\mathbf{I} - \rho_\beta \mathbf{W})^{-1}\mathbf{T}_\beta^{-1}$, utilizando as definições de **W**, **T** e ρ dadas nas seções anteriores. Podemos encontrar os coeficientes da matriz de covariância a partir dessas definições e das igualdades a seguir:

$$\Sigma_\alpha = (\mathbf{I} - \rho_\alpha \mathbf{W})^{-1}\mathbf{T}_\alpha^{-1} = [\mathbf{I} + \rho_\alpha \mathbf{W} + \rho_\alpha^2 \mathbf{W}^2 + \rho_\alpha^3 \mathbf{W}^3 + \ldots]\mathbf{T}_\alpha^{-1}$$
$$\Sigma_\beta = (\mathbf{I} - \rho_\beta \mathbf{W})^{-1}\mathbf{T}_\beta^{-1} = [\mathbf{I} + \rho_\beta \mathbf{W} + \rho_\beta^2 \mathbf{W}^2 + \rho_\beta^3 \mathbf{W}^3 + \ldots]\mathbf{T}_\beta^{-1}$$

Finalmente, a covariância *a priori* entre as áreas i no período de tempo t e j no período de tempo v é dada por:

$$cov(r_{it}, r_{jv}) \approx \frac{\sigma_\alpha}{d_i}[[\mathbf{I}]_{ij}\frac{\rho_\alpha a_{ij}}{d_i} + \frac{\rho_\alpha^2}{d_i}\sum_{k=1}^{N}\frac{a_{ik}a_{kj}}{d_k} + \ldots] + tv\frac{\sigma_\beta}{d_i}[[\mathbf{I}]_{ij}\frac{\rho_\beta a_{ij}}{d_i} + \frac{\rho_\beta^2}{d_i}\sum_{k=1}^{N}\frac{a_{ik}a_{kj}}{d_k} + \ldots]$$

Observamos para esse modelo, que a covariância *a priori* entre duas áreas quaisquer em dois períodos de tempo quaisquer é composta pela soma da covariância do efeito aleatório espacialmente estruturado α com a covariância do efeito aleatório também espacialmente estruturado β que recebe um peso igual ao produto entre os períodos de tempo t e v.

Representando por Σ_α e Σ_β as matrizes de covariância *a priori* de α e β, respectivamente, temos que a matriz de covariância deste modelo pode ser escrita da seguinte forma:

$$I_T \otimes \Sigma_\alpha + \begin{bmatrix} 1 & 2 & \dots & T-1 & T \\ 2 & 4 & \dots & 2(T-1) & 2T \\ \vdots & \ddots & \vdots & \vdots & \\ T-1 & 2(T-1) & \dots & (T-1)^2 & (T-1)T \\ T & 2T & \dots & T(T-1) & T^2 \end{bmatrix} \otimes \Sigma_\beta$$

Portanto, o primeiro termo dessa soma de matrizes representa a matriz de covariância do efeito α espacialmente estruturado e temporalmente não estruturado. Essa matriz é exatamente da forma vista no Seção 1.3.3. Já o segundo termo da soma de matrizes acima, representa a matriz de covariância do efeito β espacial e temporalmente estruturado. Ela é composta pelo produto de Kronecker das matrizes temporal e espacial. Dessa vez, a matriz temporal é diferente da que vimos na Seção 1.3.4, que era de um modelo auto-regressivo de órdem 1. Isso se deve à diferença da modelagem temporal utilizada no modelo descrito aqui, que trata o tempo de forma linear.

1.4.2. Matriz de covariância *a priori* do modelo proposto por Matinez-Beneito et al.

Define-se que o logaritmo natural da taxa de incidência para o primeiro período de tempo observado é composto pela soma de um intercepto e dois efeitos aleatórios:

$$r_{i1} = \mu + \alpha_1 + (1-\rho^2)^{-\frac{1}{2}}(\theta_{i1} + \phi_{i1}) \tag{1.7}$$

$$\theta_{i1} \sim N(0, \sigma_\theta^2)$$

$$\Phi_1 = (\phi_{1,1}, \dots, \phi_{N,1}) \sim CAR(\sigma_\phi^2)$$

Foram incluídos no modelo o efeito aleatório estruturado espacialmente ϕ e o efeito aleatório não estruturado espacialmente θ para descrever o padrão espacial do risco, de forma a garantir flexibilidade suficiente para que possam haver estimativas de taxas bastante diferentes em áreas próximas uma da outra.

No modelo descrito, ρ corresponde à correlação temporal, μ representa a média do risco levando em conta todas as regiões e todos os períodos de tempo e α_1 representa a variação da média da taxa de incidência do primeiro período de tempo em relação aos demais períodos.

Temos agora as expressões que modelam o logaritmo natural da taxa de incidência nos períodos de tempo seguintes ($t = 2, \cdots, T$):

$$r_{it} = \mu + \alpha_t + \rho(r_{i(t-1)} - \mu - \alpha_{t-1}) + \theta_{it} + \phi_{it} \tag{1.8}$$

$$\theta_{it} \sim N(0, \sigma_\theta^2)$$

$$\Phi_t = (\phi_{1t}, \dots, \phi_{Nt}) \sim CAR(\sigma_\phi^2)$$

$$\alpha = (\alpha_1, \dots, \alpha_T) \sim CAR(\sigma_\alpha^2)$$

Podemos notar que os termos estruturados e os não estruturados espacialmente são ambos independentes no tempo e tambem mutuamente independentes em todos os períodos de tempo.

Como consequência da dependência temporal definida pela distribuição *a priori* dada aos α's, os valores esperados para as taxas de incidência em cada região e em cada período de tempo não dependerão apenas dos valores de suas áreas vizinhas no mesmo período. Eles também irão depender de seus valores em outros períodos de tempo. Dessa forma, as estimativas das taxas de incidência são temporalmente dependentes. Por outro lado, o efeito aleatório espacialmente estruturado em cada período de tempo garante a dependência espacial das estimativas. Assim, o modelo definido permite a transferência de informação entre períodos de tempo e regiões vizinhas.

As distribuições *a priori* usadas para os hiperparâmetros são:

$$\sigma_{\phi}^{-2}, \sigma_{\theta}^{-2}, \sigma_{\alpha}^{-2} \sim Gamma(a, b)$$

$$\rho \sim U(-1, 1)$$

$$\mu \sim N(0, c)$$

Percebe-se que as distribuições *a priori* para o valor médio da taxa de incidência para todos os períodos de tempo e para a precisão dos efeitos aleatórios são definidas de forma que os hiperparâmetros têm a intenção de expressar informações bem vagas.

Considerando agora o parâmetro de correlação temporal, a distribuição *a priori* de ρ foi escolhida de forma a garantir a estacionariedade da série temporal, considerando que ela tem uma estrutura auto-regressiva de ordem 1. Na equação 1.7 o termo $(1-\rho^2)^{-1/2}$ é introduzido para fazer com que a matriz de covariâncias de $Y_{\cdot 1}$ fique igual a matriz de covariância estacionária da série $(Y_{\cdot j})_{j=1}^{\infty}$. A definição do logaritmo natural da taxa para os períodos de tempo seguintes é feita de forma análoga.

É interessante observar que os incrementos temporais de um período para o seguinte têm uma estrutura espacial, de forma que regiões vizinhas terão tendências temporais das taxas de incidência similares, assim como elas têm estimativas geográficas das taxas de incidência similares.

Para encontrar os termos da matriz de covariâncias do modelo proposto, são realizados os cálculos a seguir:

$$\text{var}(\mathbf{r}_1|\mu, \alpha, \sigma_{\phi}, \sigma_{\theta}) = \text{var}((1-\rho^2)^{-1/2}(\theta_1 + \phi_1)) = (1-\rho^2)^{-1}\Sigma$$

$$\text{var}(\mathbf{r}_2|\mu, \alpha, \sigma_{\phi}, \sigma_{\theta}) = \text{var}(\rho\mathbf{r}_1 + (\theta_2 + \phi_2)) = \rho^2\text{var}(\mathbf{r}_1) + \text{var}(\theta_2 + \phi_2) = (1-\rho^2)^{-1}\Sigma$$

onde Σ é a matriz de covariância de $\theta_i + \phi_i$.

Procedendo da mesma maneira, é fácil mostrar que a variância é a mesma para todos os períodos de tempo, sendo igual a $(1-\rho^2)^{-1}\Sigma$.

Para as covariâncias, temos:

$$\begin{aligned} cov(\mathbf{r}_j, \mathbf{r}_{j+k}) &= cov(\mathbf{r}_j, \rho^k\mathbf{r}_j + \sum_{t=0}^{k-1}\rho^t(\theta_{j+k-t} + \phi_{j+k-t})) \\ &= cov(\mathbf{r}_j, \rho^k\mathbf{r}_j) = \rho^k var(\mathbf{r}_j) = \rho^k(1-\rho^2)^{-1}\Sigma \end{aligned}$$

Conclui-se que a matriz de covariâncias para o logaritmo natural da taxa de incidência, segundo o modelo proposto, é dada por

$$\frac{1}{1-\rho^2}\begin{bmatrix} \Sigma & \rho\Sigma & \dots & \rho^{T-1}\Sigma \\ \rho\Sigma & \Sigma & \dots & \rho^{T-2}\Sigma \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1}\Sigma & \rho^{T-2}\Sigma & \dots & \Sigma \end{bmatrix} = \Lambda \otimes \Sigma$$

onde Λ representa a matriz de correlação de uma série temporal auto-regressiva de ordem 1 e tamanho T, ou seja:

$$\Lambda = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{bmatrix}$$

Neste ponto, chegamos a um resultado muito semelhante ao que tivemos na seção anterior para os modelos com interação entre os efeitos temporais e espaciais, ambos estruturados. Temos que a matriz de precisão corresponde ao produto de Kronecker entre as matrizes de precisão temporal e espacial, já que ela é dada por $(\Lambda \otimes \Sigma)^{-1} = \Lambda^{-1} \otimes \Sigma^{-1}$.

Vamos estudar os coeficientes da covariância *a priori* para o modelo proposto. Para isso, define-se a matriz de covariância espacial $\Sigma = (\mathbf{I} - \rho_\phi \mathbf{W})^{-1}\mathbf{T}^{-1} + \sigma_\theta \mathbf{I}$, utilizando as definições de **W** e **T** dadas nas seções anteriores. Para encontrarmos os valores dos coeficientes da matriz de covariância $\Lambda \otimes \Sigma$, realizamos alguns cálculos.

$$\begin{aligned} \Sigma &= (\mathbf{I} - \rho_\phi \mathbf{W})^{-1}\mathbf{T}^{-1} + \sigma_\theta \mathbf{I} \\ &= [\mathbf{I} + \rho_\theta \mathbf{W} + \rho_\theta^2 \mathbf{W}^2 + \rho_\theta^3 \mathbf{W}^3 + \dots]\mathbf{T}^{-1} + \sigma_\theta \mathbf{I} \\ &= (\mathbf{T}^{-1} + \sigma_\theta \mathbf{I}) + [\rho_\theta \mathbf{W} + \rho_\theta^2 \mathbf{W}^2 + \rho_\theta^3 \mathbf{W}^3 + \dots]\mathbf{T}^{-1} \end{aligned}$$

Observamos que os coeficientes de $(\mathbf{T}^{-1} + \sigma_\theta \mathbf{I})$ só são diferentes de zero na diagonal da matriz Σ, ou seja, quando estamos considerando a variância de cada área. Nos termos que representam as covariâncias entre áreas distintas, esses coeficientes são iguais a zero.

Com relação à matriz relativa à parte temporal Λ, observamos que, a menos da constante $(1-\rho^2)^{-1}$, os coeficientes da matriz têm uma forma geral, dada por $\rho^{|t-v|}$, onde t e v são a linha e a coluna da matriz, representado os dois períodos de tempo entre os quais estamos calculando a covariância. Ao considerarmos a diagonal da matriz Λ, observamos que a variância de todos os períodos é proporcional a 1, ou seja é constante em relação ao tempo.

Finalmente, a covariância entre as áreas i no período de tempo t e j no período de tempo v é dada por:

$$cov(r_{it}, r_{jv}) \approx \begin{cases} \rho^{|t-v|}\{(\frac{\sigma_\phi}{d_i} + \sigma_\theta)[\mathbf{I}]_{ij} + \frac{\sigma_\phi}{d_i}[\frac{\rho_\phi a_{ij}}{d_i} + \frac{\rho_\phi^2}{d_i}\sum_{k=1}^{N}\frac{a_{ik}a_{kj}}{d_k} + \dots]\} & \text{se} \quad i = j \\ \rho^{|t-v|}\frac{\sigma_\phi}{d_i}[\frac{\rho_\phi a_{ij}}{d_i} + \frac{\rho_\phi^2}{d_i}\sum_{k=1}^{N}\frac{a_{ik}a_{kj}}{d_k} + \dots] & \text{se} \quad i \neq j \end{cases}$$

Observamos para este modelo, que a covariância entre duas áreas distintas i e j em um mesmo período de tempo ($t = v$) é proporcional à soma ponderada de todos os caminhos possíveis entre a área i e a área j em 1, 2, 3, ... passos. Já em períodos de tempo distintos t e v, a covariância entre as áreas distintas i e j é proporcional a $\rho_\alpha^{|t-v|}$ vezes a soma ponderada de todos os caminhos possíveis entre a área i e a área j em 1, 2, 3, ... passos. Dessa forma, a covariância diminui a medida que os períodos de tempo em questão são mais distantes um do outro, já que $|\rho_\alpha| < 1$. Esses termos representam a interação entre o efeito temporalmente estruturado e o efeito espacialmente estruturado, semelhante ao que vimos na Seção 1.3.4.

Ao considerar a variância de uma determinada área ($i = j$ e $t = v$) ou a covariância de uma área com ela mesma em períodos de tempo distintos ($i = j$ e $t \neq v$), os elementos correspondentes da matriz levam em conta um termo referente à variância de θ, já que esse é o efeito não estruturado espacialmente (variância não nula e constante e covariância entre áreas distintas nula). Esse elemento é proporcional a $\rho^{|t-v|}\sigma_\theta[\mathbf{I}]_{ij}$ e representa a interação entre o efeito temporal estruturado e o efeito espacial não estruturado θ, semelhante ao que vimos na Seção 1.3.2.

Representando por Σ_θ e Σ_ϕ as matrizes de covariância *a priori* de θ e ϕ, respectivamente, e utilizando uma propriedade do produto de Kronecker [13], temos:

$$\begin{aligned}
\Lambda \otimes \Sigma &= \Lambda \otimes (\Sigma_\theta + \Sigma_\phi) \\
&= \Lambda \otimes (\sigma_\theta \mathbf{I} + (\mathbf{I} - \rho_\phi \mathbf{W})^{-1}\mathbf{T}) \\
&= (\Lambda \otimes \sigma_\theta \mathbf{I}) + (\Lambda \otimes (\mathbf{I} - \rho_\phi \mathbf{W})^{-1}\mathbf{T})
\end{aligned}$$

Portanto, o padrão geral dos valores encontrados para os coeficientes da matriz de covariância *a priori* desse modelo pode ser descrito como a interação entre o efeito temporalmente estruturado e o efeito espacialmente estruturado mais a interação entre o efeito temporalmente estruturado e o efeito espacialmente não estruturado.

1.5. Conclusões

Neste trabalho fizemos uma revisão de tipos possíveis de efeitos aleatórios espaciais e temporais em modelos bayesianos. Classificamos tais efeitos de acordo com as estruturas de dependência espacial ou temporal atribuídas, *a priori*, a cada um deles. Para isso, atribuimos distribuições *a priori* utilizadas com frequência para cada tipo de efeito aleatório em questão. Construimos as possíveis interações espaço-tempo resultantes das combinações entre um dos efeitos espaciais e um dos temporais. Para os modelos com estas interações, escrevemos as matrizes de covariância *a priori* utilizando o produto de Kronecker entre as matrizes de covariância *a priori* de um efeito puramente espacial e outro puramente temporal.

Dessa maneira, fizemos algumas interpretações interessantes dos coeficientes das matrizes de covariâcia *a priori* para os modelos com interação espaço-tempo. Vizualizamos mais claramente o efeito de cada tipo de interação possível entre os efeitos espaciais e temporais, relacionando as matrizes de covariância *a priori* com as estruturas de dependência espacial e/ou temporal envolvidas nos modelos estudados.

1.6. Referências Bibliográficas

[1] Assunção RM, Krainski E. Neighborhood dependence in Bayesian spatial models. *Biometrical Journal.* 2009; **51**:851–869.

[2] Assunção RM, Reis IA, Oliveira CL. Diffusion and prediction of leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. *Statistics in Medicine.* 2001; **20**:2391–2335.

[3] Bailey TC, Gatrell AC. *Interactive Spatial Data Analysis.* 1995; Longman Scientific & Technical, England.

[4] Bernardinelli L, Montomoli C. Empirical Bayes Versus Fully Bayesian Analisys of Geographical Variation in Disease Risk. *Statistics in Medicine.* 1992; **11**:983–1007.

[5] Besag J. Spatial interaction and the statistical analysis of lattice systems (with discussions). *Journal of the Royal Statistical Society, Series B* 1974; **36**:192–236.

[6] Besag J, Kooperberg C. On conditional and intrinsic autoregressions. *Biometrika.* 1995; **82**:733–746.

[7] Brémaud P. *Markov Chains.* Springer-Verlag, New York.

[8] Clayton DG. *Generalized linear mixed models. In Markov Chain Monte Carlo in Practice.* (eds Gilks WR, Richardson S, Spiegelhalter DJ.) Chapman & Hall: London. 275–301.

[9] Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine.* 2000; **19**:2555–2567.

[10] Iosifescu M. *Finite Markov Processes and Their Applications.* John Wiley and Sons, New York.

[11] Martinez-Beneito MA, Lopez-Quilez A, Botella-Rocamora P. An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine.* 2008; **27**:2874–2889.

[12] Rue H, Held L. *Gaussian Markov Random Fields; Theory and Applications.* 2005; Chapman & Hall/CRC Press, Boca Raton.

[13] Steeb WH. *Matrix Calculus and Kronecker Product with Applications and C++ Programs.* 1997; World Scientific Publishing, Singapore.

Capítulo 2

Ajuste de Modelos Bayesianos Espaço-Temporais para Leishmaniose Visceral na Cidade de Belo Horizonte

2.1. Introdução

O estudo de métodos estatísticos para análise de dados espaciais é importante em diversas áreas, tais como ecologia, epidemiologia, demografia, geografia, entre outros. Os dados espaciais contém, além de valores de possíveis atributos de interesse, as localizações espaciais relativas às observações.

Uma região de estudo pode ser particionada em uma quantidade n de áreas de formas distintas, dependendo do problema a ser analisado. Por exemplo, uma cidade pode ser particionada em regiões, bairros, setores censitários, áreas de abrangência dos centros de saúde, etc. Deve ser escolhida a partição mais conveniente. Ao fazer uma partição da região de estudo em n áreas, deve-se garantir que toda a região está contida na união dessas áreas e que nenhuma área tem interseção com outra.

A cada uma das n áreas observadas é associada uma variável aleatória. É comum em estudos epidemiológicos que essas variáveis aleatórias representem a quantidade de casos de determinada doença ou a mortalidade devido a uma causa específica, ocorridos em cada área durante um certo período de tempo. Caso haja interesse em acompanhar as tendências da variável de interesse ao longo do tempo, é comum que sejam associados a cada área valores medidos em períodos de tempo distintos. Dessa forma, são obtidas informações associadas a cada área e a cada período de tempo, constituindo os dados espaço-temporais. Frequentemente, o objetivo desses estudos epidemiológicos é estimar as taxas de incidência de doenças, acompanhando sua distribuição espacial e suas tendências temporais.

Ao analisar doenças relativamente raras e áreas com populações pequenas, as estimativas das taxas de incidência podem não refletir bem a realidade. Isso dificulta o acompanhamento das tendências de interesse. Uma alternativa é utilizar informações das áreas vizinhas e/ou de períodos de tempo subsequentes para estimar a taxa de incidência em uma área em um certo período de tempo, suavizando a distribuição das estimativas. Estudos desse tipo têm se tornado muito populares em epidemiologia [1, 4, 3, 6, 7] e os

resultados alcançados costumam ser satisfatórios.

Neste trabalho, faremos um estudo de modelos bayesianos espaço-temporais aplicados a dados de Leishmaniose Visceral (LV) na cidade de Belo Horizonte. A LV [9] é uma doença infecciosa, caracterizada por febre, aumento do fígado e do baço, anemia, entre outros sintomas, levando o doente à morte se não for devidamente tratada. A transmissão se dá através da picada de flebotomíneo contaminado pelo parasita causador da doença. Atualmente, sabe-se que o cão doméstico, entre outros animais, tem importante papel na manutenção da doença, pois age como hospedeiro do parasita. O tempo de incubação (prazo que se passa entre a picada do mosquito transmissor da doença e a manifestação dos sintomas nas pessoas) da LV varia de 10 dias a 24 meses, com média de 2 a 6 meses.

O controle da LV nos grandes centros urbanos tem sido um problema para as Secretárias de Saúde locais e para o Ministério da Saúde. As ações têm alto custo e demandam grande quantidade de mão de obra. Por isso, a identificação detalhada dos locais de risco permite a adoção de estratégias direcionadas e específicas, e portanto mais efetivas no controle da LV. Nesse contexto, o estudo realizado neste trabalho da distribuição e da evolução espaço-temporal da doença serve de ferramenta para direcionar estratégias de controle e priorizar as áreas de maior risco de incidência de LV humana.

No trabalho desenvolvido por Assunção et al. [1] para os anos de 1994 a 1996, pode-se acompanhar o espalhamento da LV humana em Belo Horizonte, a partir de sua entrada no município pelas regiões Leste e Nordeste. Também foram identificadas áreas em outras regiões do município com um crescimento mais rápido nas taxas da doença, o que poderia caracterizar novos focos. O modelo utilizado por Assunção et al. [1] para ajustar os dados de LV considera que as taxas de incidência são espacialmente estruturadas, ou seja, áreas vizinhas tendem a ter estimativas do risco relativo semelhantes. Quanto à parte temporal, o modelo considera uma tendência de crescimento (ou decrescimento) linear da taxa de incidência em relação ao tempo. Esse será um dos modelos que ajustaremos aos dados mais recentes de LV em Belo Horizonte.

Em um trabalho mais recente, Martínez-Beneito et al. [7] propõem uma abordagem espaço-temporal para mapeamento de doenças, combinando idéias de séries temporais e de modelagem espacial. Eles constroem um modelo que define uma estrutura na qual as estimativas das taxas de incidência são estruturadas espacial e temporalmente, ao mesmo tempo. Dessa forma, áreas vizinhas e períodos de tempo subsequentes tendem a ter estimativas das taxa de incidência semelhantes. Também ajustaremos esse modelo aos dados atuais de LV em Belo Horizonte.

Na Seção 2.2 deste trabalho, descreveremos os dados disponíveis para o estudo e faremos uma análise exploratória. Na Seção 2.3, faremos uma definição detalhada dos modelos que serão ajustados aos dados de LV em Belo Horizonte. Na Seção 2.4, apresentaremos os resultados dos ajustes de cada modelo proposto aos dados e faremos algumas comparações entre eles. E finalmente, na Seção 2.5, serão exibidas as conclusões sobre o trabalho realizado.

2.2. Análise Preliminar dos Dados

A área de estudo será a cidade de Belo Horizonte (BH), capital do estado de Minas Gerais. Segundo o último censo realizado pelo IBGE (2000), BH possui uma população urbana de aproximadamente 2.2 milhões de habitantes numa área de $330.90km^2$. Na cidade residem 51% da população da Região Metropolitana de Belo Horizonte (RMBH), o que resulta numa densidade demográfica de 6.8 mil habitantes por km^2, bem maior que os 460 habitantes por km^2 da RMBH.

BH é dividida administrativamente em nove regiões: Norte, Nordeste, Noroeste, Sul, Sudeste, Pampulha, Barreiro, Centro-sul e Oeste, como podemos observar na Figura 2.1. Cada uma delas corresponde a um Distrito Sanitário (DS), que gerencia as atividades de saúde em seu território. Há tambem uma delimitação geográfica que corresponde à responsabilidade territorial das unidades básicas de saúde. Estas regiões são chamadas de Áreas de Abrangência dos Centros de Saúde (AA), que totalizam 140.

Figura 2.1. Mapa da cidade de Belo Horizonte divida nas regiões referentes aos nove distritos sanitários

Os dados disponíveis para a realização desse estudo são os seguintes:

- As populações humanas, por área de abrangência, referentes aos anos de 2000 a 2008.
- A data de início de sintomas, data de notificação, evolução (cura, óbito por LV, óbito por outras causas) e endereço de residência dos casos confirmados de LV humana de residentes em BH, notificados entre 2000 e 2008.

Objetivando analisar previamente esses dados, tentamos observar se havia alguma tendência geral dos casos humanos durante os anos. Para isso, fizemos uma análise dos casos em cada ano, dividindo a área de estudos (Belo Horizonte) em distritos. Para cada distrito foi calculado o logaritmo natural da taxa de incidência de LV, que é dada pela razão

entre o número de casos e a população, para cada ano entre 2000 e 2008. Os resultados obtidos estão exibidos nos gráficos da Figura 2.2.

A série temporal dos casos para toda a cidade de Belo Horizonte pode ser vista na Figura 2.3.

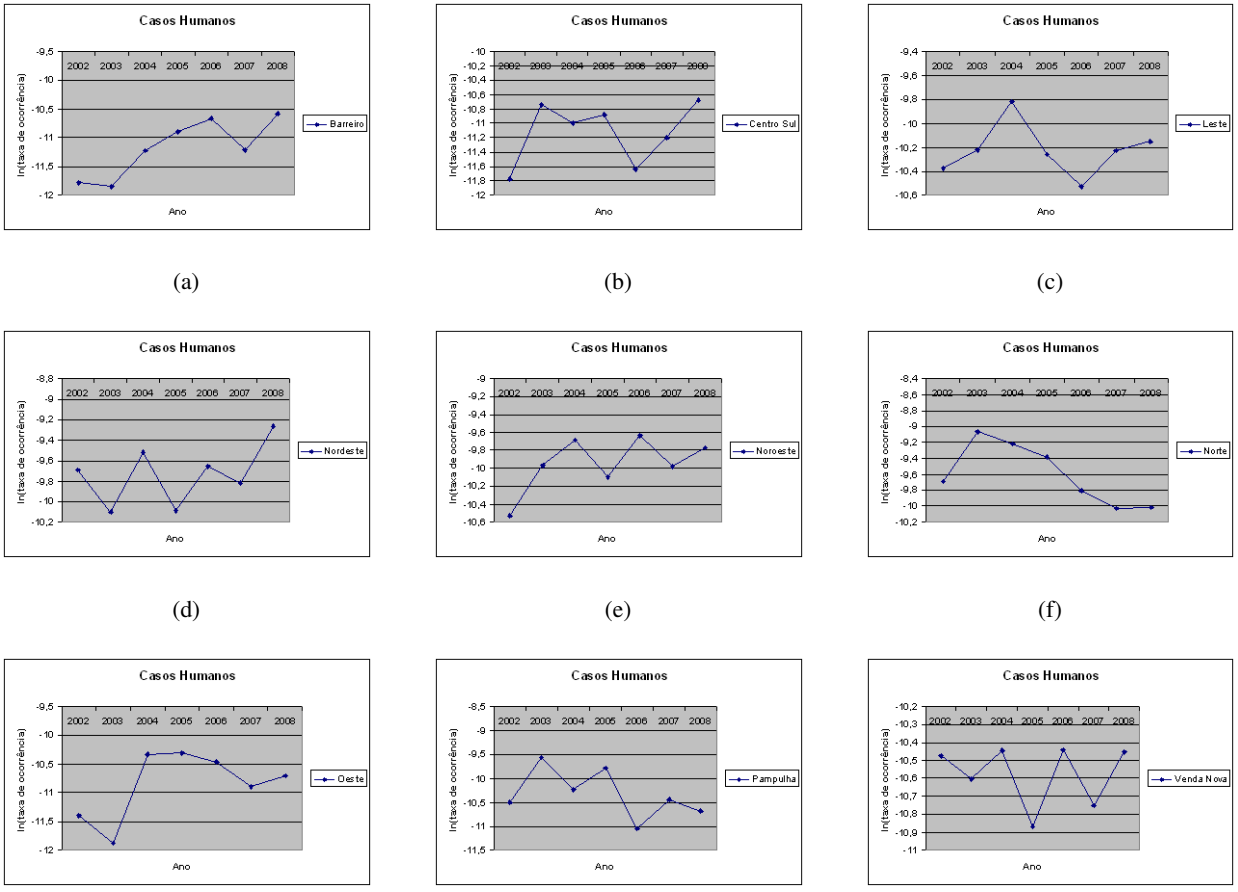


Figura 2.2. Séries Temporais referentes ao log das taxas de LV por Distrito de Belo Horizonte para os anos de 2002 **a** 2008

Observando as séries temporais por distritos, podemos ver que elas variam consideravelmante de um distrito para outro. É dificil identificar algum padrão geral que nos permita concluir que existe alguma tendência temporal nos casos de LV humana para a cidade de Belo Horizonte como um todo, tanto analisando a Figura 2.2 quanto a Figura 2.3. Portanto, não foi possível observar alguma tendência temporal a partir dessa forma de análise preliminar dos dados. Observe que o número de observações no tempo é muito pequeno para considerar modelos sofisticados de evolução temporal.

Ao analisar casos de doenças, também é interessante verificar se esses casos tendem a ocorrer mais em determinados meses ou estações do ano. A dengue, por exemplo, tem incidência elevada em Belo Horizonte no verão com relação às demais estações do ano, devido à época de chuvas ser propícia para a proliferação do mosquito transmissor. Com a intenção de observar se há alguma sazonalidade na quantidade de casos humanos de LV em Belo Horizonte, fizemos o gráfico da série temporal desses casos para cada

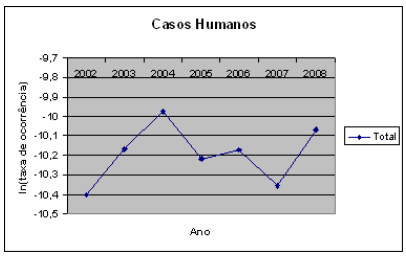


Figura 2.3. Série Temporal referente ao log da taxa de LV em Belo Horizonte para os anos de 2002 **a** 2008

mês, nos anos de 2006 a 2008 (anos em que temos os dados sobre os casos de LV humana por mês). Os resultados obtidos estão exibidos na Figura 2.4.

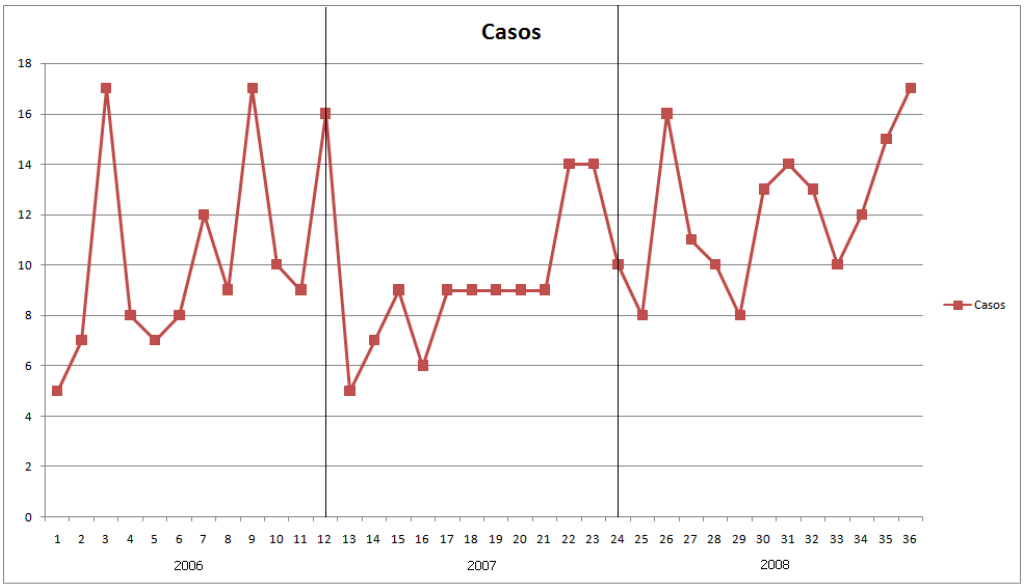


Figura 2.4. Série Temporal dos casos humanos de LV em Belo Horizonte para os anos de 2006 **a** 2008

Ao observar a série gerada não conseguimos identificar uma sazonalidade muito evidente nos dados. Obtivemos os dados de temperatura e de precipitação acumulada para os períodos em questão. Ao comparar as séries temporais dos casos humanos com as outras duas, ainda não identificamos um padrão em comum. Dessa forma, trabalharemos daqui em diante com a hipótese de que o risco de contrair a doença não varia de acordo com os meses ou as estações do ano.

2.3. Descrição dos Modelos

A partir da análise preliminar dos dados envolvidos e do problema em questão, escolhemos trabalhar com três modelos para ajustar dos dados de LV humana em Belo Horizonte.

Inicialmente, fizemos a estimação das taxas de incidência utilizando o procedimento clássico, que utiliza o conceito de taxas brutas. No entanto, quando o número de casos e a população em risco para uma única área em um único período de tempo são muito pequenos para produzir uma estimativa próxima da realidade, esse modelo gera resultados indesejáveis. Esse não é um modelo bayesiano e não utiliza informações de áreas vizinhas para estimar a taxa de incidência de uma área. Escolhemos ajustá-lo aos dados de LV para comparar seus resultados com os ajustes de modelos bayesianos que serão realizados.

Um dos modelos bayesianos a ser ajustado, o proposto por Assunção et al. [1], assume que as taxas de incidência variam no tempo de acordo com uma tendência linear específica para cada área da região de estudo. Os parâmentros relativos a essas tendências recebem uma distribuição *a priori* CAR, de forma a serem espacialmente estruturados (estimativas de uma área baseadas nas observações das áreas vizinhas). Portanto, é feita a interação entre o efeito espacial estruturado e o efeito temporal não estruturado. Esse modelo já foi aplicado para dados de LV pelos autores, que consideraram os resultados obtidos satisfatórios. Porém, o estudo foi realizado com dados de apenas três anos (1994 a 1996).

O último modelo bayesiano que utilizamos, proposto por Martínez-Beneito et al. [7], combina idéias de séries temporais e de modelagem espacial. Nesse modelo, é introduzida uma interação entre os efeitos temporal e espacialmente estruturados. Dessa forma, é construído um modelo incluindo uma estrutura na qual as taxas de incidência são espacial e temporalmente dependentes dos vizinhos (no espaço e no tempo), ao mesmo tempo. Acreditamos que este modelo, publicado recentemente, pode ter uma boa aplicação no nosso problema.

Para as definições dos modelos a seguir, considere as seguintes notações:

- Y_{it} é a contagem dos casos da doença na área i no período de tempo t.
- P_{it} é a população na área i no período de tempo t.
- ξ_{it} é a taxa de incidência da doença na área i no período de tempo t.
- $i = 1, \cdots, N$.
- $t = 1, \cdots, T$.

Vamos assumir que o número de casos de uma determinada doença para cada área e cada período de tempo segue uma distribuição de Poisson:

$$[Y_{it} \mid P_{it}, \xi_{it}] \sim Poisson(P_{it}\xi_{it})$$

A seguir, apresentamos as definições dos modelos utilizados:

2.3.1. Modelo de Taxas Brutas

A estimativa de taxa bruta é dada pelo estimador de máxima verossimilhança da taxa de incidência da doença, ou seja, $\hat{\xi_{it}} = Y_{it}/P_{it}$.

Ao lidar com as taxas brutas de doenças relativamente raras (em regiões pequenas encontramos muitas áreas com zero ou pouquíssimos casos), pode haver uma grande variabilidade nas estimativas $\hat{\xi_{it}}$ obtidas para as áreas. Nesse caso, esta modelagem pode não ser uma boa opção se pretendemos acompanhar tendências de variação espacial e evolução temporal.

2.3.2. Modelo proposto por Assunção et al.

Definimos $\xi_{it} = \exp(r_{it})$, onde $\exp(r_{it})$ representa a taxa de incidência da doença na área i no período de tempo t.

O procedimento proposto modela o logaritmo natural das taxas de incidência da doença linearmente em relação ao tempo, da seguinte forma:

$$r_{it} = ln(\xi_{it}) = \alpha_i + \beta_i t$$

Definimos α como o vetor que contém os α_i's de todas as áreas, representando possíveis características que são espacialmente mas não são temporalmente estruturadas de acordo com o considerado no modelo. E definimos β como o vetor que contém os β_i's de todas as áreas, representando possíveis características que são espacialmente estruturadas e variam linearmente em relação ao tempo.

Para que haja dependência espacial entre os valores das taxas de incidência, de forma que áreas vizinhas tenham valores semelhantes, a distribuição *a priori* utilizada para os parâmetros α e β é a Condicional Autoregressiva (CAR) [2], que é dada por:

$$P(\alpha \mid \tau_\alpha) \propto exp(\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(\alpha_i - \alpha_j)^2)$$

$$P(\beta \mid \tau_\beta) \propto exp(\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(\beta_i - \beta_j)^2)$$

onde w_{ij} é um elemento da matriz de vizinhança, cujo valor é igual a 1 se as áreas i e j são vizinhas, e igual a 0, caso contrário.

Essa mesma distrubuição CAR pode ser definida a partir das distribuições condicionais seguintes:

$$[\alpha_i \mid \alpha_{-i}] \sim Normal(\overline{\alpha}_{-i}, n_i\tau_\alpha)$$

$$[\beta_i \mid \beta_{-i}] \sim Normal(\overline{\beta}_{-i}, n_i\tau_\beta)$$

onde α_{-i} e β_{-i} representam, respectivamente, os valores de α e β de todas as áreas, exceto a área i, e $\overline{\alpha}_{-i}$ e $\overline{\beta}_{-i}$ representam, respectivamente, as médias de α e β nas áreas vizinhas

à área i. Os valores $n_i\tau_\alpha$ e $n_i\tau_\beta$ representam as precisões de α_i e β_i, respectivamente, sendo τ_α e τ_β os parâmetros de precisão e n_i o número de vizinhos da área i.

As distribuições de probabilidade *a priori* dos parâmetros τ_α e τ_β são definidas de forma a serem bem vagas, já que não há conhecimento algum sobre esses parâmetros de precisão *a priori*. Elas são dadas então por:

$$\tau_\alpha \sim Gamma(a, b)$$

$$\tau_\beta \sim Gamma(c, d)$$

onde a, b, c e d são valores conhecidos. Tipicamente a, b, c e d recebem valores pequenos, tais como 10^{-5}, de modo que o valor esperado *a priori* é igual a 1 e a variância é muito grande (10^5).

Definidas as distribuições *a priori* para os parâmetros, podemos agora definir a distribuição de probabilidade *a posteriori* conjunta, dada por:

$$P_{post}[\alpha, \beta, \tau_\alpha, \tau_\beta \mid Y, P] \propto L(Y \mid \alpha, \beta, \tau_\alpha, \tau_\beta, P) \times P_{priori}(\alpha, \beta, \tau_\alpha, \tau_\beta)$$

onde $L(Y \mid \alpha, \beta, \tau_\alpha, \tau_\beta, P)$ representa a verossimilhança do modelo de Poisson descrito nessa seção e $P_{priori}(\alpha, \beta, \tau_\alpha, \tau_\beta)$ representa a distribuição *a priori* conjunta resultante da multiplicação das distribuições *a priori* de α, β, τ_α e τ_β.

2.3.3. Modelo Proposto por Martinez-Beneito et al.

Definimos $\xi_{it} = \exp(r_{it})$, onde $\exp(r_{it})$ representa a taxa de incidência da doença na área i no período t.

O logaritmo natural da taxa de incidência para o primeiro período de tempo observado é composto pela soma de um intercepto e dois efeitos aleatórios:

$$r_{i1} = \mu + \alpha_1 + (1 - \rho^2)^{-\frac{1}{2}}(\theta_{i1} + \phi_{i1})$$
$$\theta_{i1} \sim N(0, \sigma_\theta^2)$$
$$\Phi_1 = (\phi_{1,1}, \ldots, \phi_{N,1}) \sim CAR.Normal(\sigma_\phi^2)$$

Foram incluídos no modelo o efeito aleatório estruturado espacialmente (ϕ) e o efeito aleatório não estruturado espacialmente (θ) para descrever o padrão espacial da taxa de incidência, de forma a garantir flexibilidade suficiente para que possam haver estimativas bastante diferentes em áreas próximas uma da outra.

No modelo descrito, ρ corresponde à correlação temporal, μ representa a média da taxa de incidência levando em conta todas as regiões e todos os perodos de tempo e α_1 representa a variação da média do risco do primeiro período de tempo em relação aos demais períodos.

Á seguir, estão as expressões que modelam o risco relativo nos períodos de tempo seguintes ($t = 2, \cdots, T$):

$$r_{it} = \mu + \alpha_t + \rho(r_{i(t-1)} - \mu - \alpha_{t-1}) + \theta_{it} + \phi_{it}$$

$$\theta_{it} \sim N(0, \sigma_\theta^2)$$
$$\Phi_t = (\phi_{1t}, \ldots, \phi_{Nt}) \sim CAR.Normal(\sigma_\phi^2)$$
$$\alpha = (\alpha_1, \ldots, \alpha_T) \sim CAR.Normal(\sigma_\alpha^2)$$

Podemos notar que os termos espaciais estruturados e não estruturados são ambos independentes no tempo e também mutuamente independentes em todos os períodos de tempo.

Como consequência da dependência temporal definida pela distribuição *a priori* dada aos α's, os valores esperados para as taxas de incidência em cada região e em cada período de tempo não dependerão apenas das estimativas de suas áreas vizinhas no mesmo período. Eles também irão depender de suas estimativas em outros períodos de tempo. Dessa forma, as estimativas das taxas de incidência são temporalmente dependentes. Por outro lado, o efeito aleatório espacialmente estruturado em cada período de tempo garante a dependência espacial dessas estimativas. Assim, o modelo definido permite a tranferência de informação entre períodos de tempo e regiões vizinhas.

As distribuições *a priori* usadas para os hiperparâmetros são:

$$\sigma_\phi^{-2}, \sigma_\theta^{-2}, \sigma_\alpha^{-2} \sim Gamma(a, b)$$
$$\rho \sim U(-1, 1)$$
$$\mu \sim N(0, c)$$

Tipicamente a e b recebem valores pequenos, tais como 10^{-5}, de modo que o valor esperado para a precisão *a priori* é igual a 1 e a variância é muito grande (10^5).

Percebe-se que as distribuições *a priori* para o valor médio da taxa de incidência para todos os períodos de tempo e para a precisão dos efeitos aleatórios são definidas de forma que os hiperparâmetros têm a intenção de expressar informações bem vagas.

Considerando agora o parâmetro de correlação temporal, a distribuição *a priori* de ρ foi escolhida de forma a garantir a estacionariedade da série temporal, considerando que ela tem uma estrutura auto-regressiva de ordem 1.

Dadas as definições, temos então uma modelagem flexível para descrever a evolução temporal das taxas de incidência de uma maneira similar à utilizada comumente (e também neste modelo) para a evolução espacial.

2.4. Resultados do Ajuste dos Modelos aos Dados de Leishmaniose Visceral em Belo Horizonte

Não é possível obter as expressões analíticas para as distribuições *a posteriori* dos dois modelos bayesianos utilizados neste trabalho devido à grande dimensão do problema. Algumas medidas de comparação, como o Fator de Bayes, são difíceis de serem calculadas. Sendo assim, utilizamos o DIC (*Deviance Information Criterion*) [5] para fazer essa comparação entre os modelos descritos na seçaõ anterior. O DIC é bastante utilizado em seleção de modelos bayesianos nos quais as distribuições *a posteriori* são obtidas por simulações MCMC (*Markov chain Monte Carlo*).

O desvio é definido por $D(\theta) = -2\log(p(y|\theta)) + C$, onde y representa os dados, θ representa os parâmetros desconhecidos e $p(y|\theta)$ é a função de verossimilhança do modelo. C é uma constante que sempre se cancela ao comparar diferentes modelos e, portanto, ela não precisa ser conhecida. A esperança $\bar{D} = \mathbf{E}_{\theta|y}[D(\theta)]$ é uma medida de quão bem o modelo se ajusta aos dados, de forma que quanto maior é seu valor, pior é o ajuste. O número efetivo de parâmetros do modelo é calculado por $p_D = \bar{D} - D(\bar{\theta})$, onde $\bar{\theta}$ é a esperança *a posteriori* de θ. Quanto maior é seu valor, mais fácil para o modelo se ajustar aos dados.

O DIC é dado por:

$$DIC = p_D + \bar{D}$$

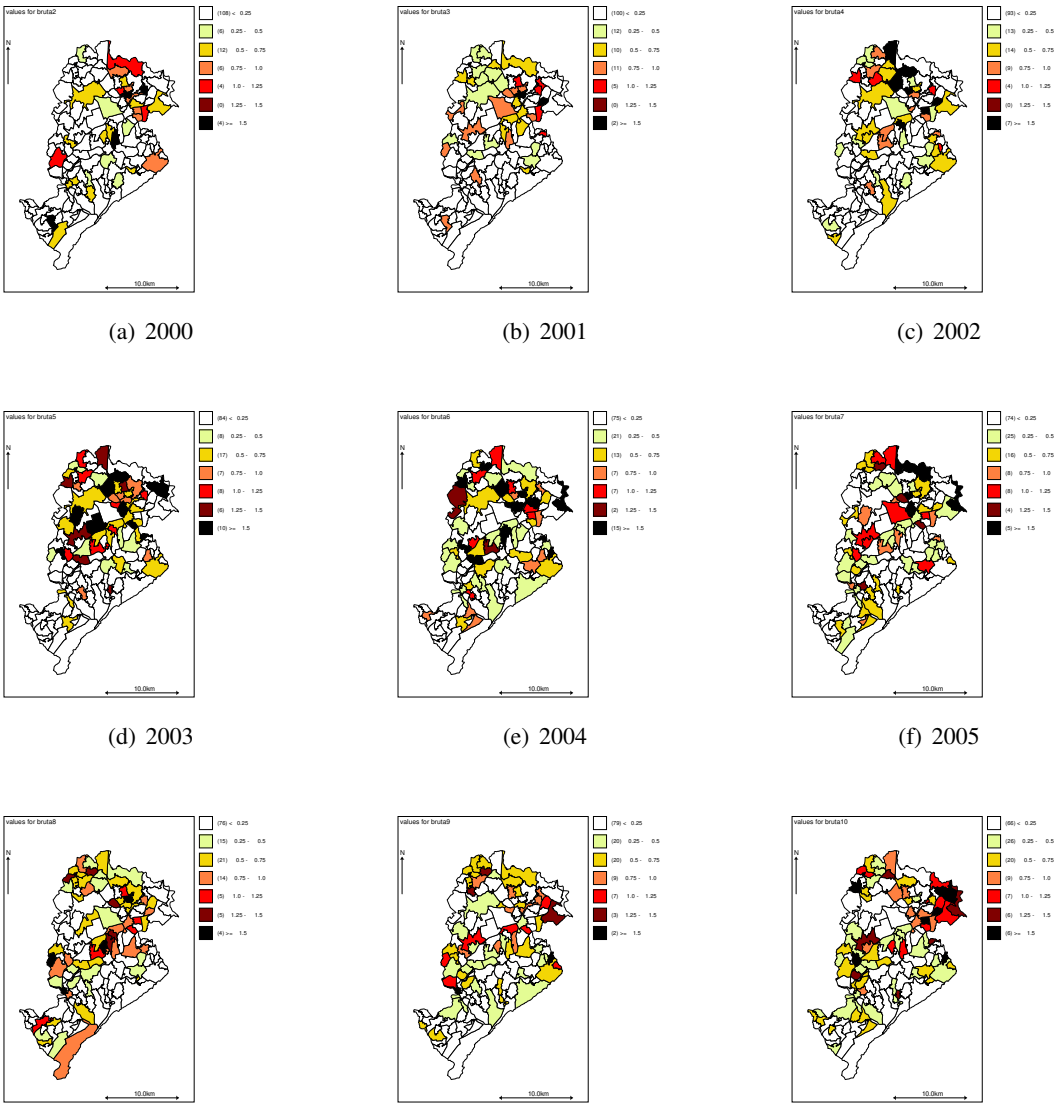
A idéia é que modelos com valores do DIC menores devem ser preferidos do que modelos com valores maiores. Os modelos são penalizados tanto pelo $\bar{D}$, que favorece o melhor ajuste, quanto pelo número efetivo de parâmetros p_D. Enquanto o valor de $\bar{D}$ vai diminuindo com o aumento do número de parâmetros no modelo, o p_D compensa esse efeito, favorecendo modelos com menor número de parâmetros.

Os modelos descritos na Seção 2.3 foram implementados no software WinBUGS [9]. Os códigos fontes desenvolvidos e utilizados estão no apêndice. Com um *burn in* de 5000 simulações, rodamos outras 20000 simulações de cada um deles. A partir dessas simulações, foram calculados os DICs (para os modelos bayesianos) e produzidos os mapas exibidos a seguir para cada modelo implementado, reproduzindo as estimativas da taxa de incidência de LV em Belo Horizonte para cada área de abrangência em cada ano, entre 2000 e 2008.

Apresentamos aqui os resultados obtidos para cada um dos modelos ajustados e implementados no WinBUGS. Posteriormente, faremos uma comparação entre os ajustes.

2.4.1. Modelo de Taxas Brutas

Ao gerar os gráficos com as estimativas das taxas de incidência da LV, calculadas a partir do modelo de taxas brutas, para a cidade de BH dividida em áreas de abrangência, nos anos de 2000 a 2008, obtivemos os mapas da Figura 2.5.



(a) 2000 (b) 2001 (c) 2002

(d) 2003 (e) 2004 (f) 2005

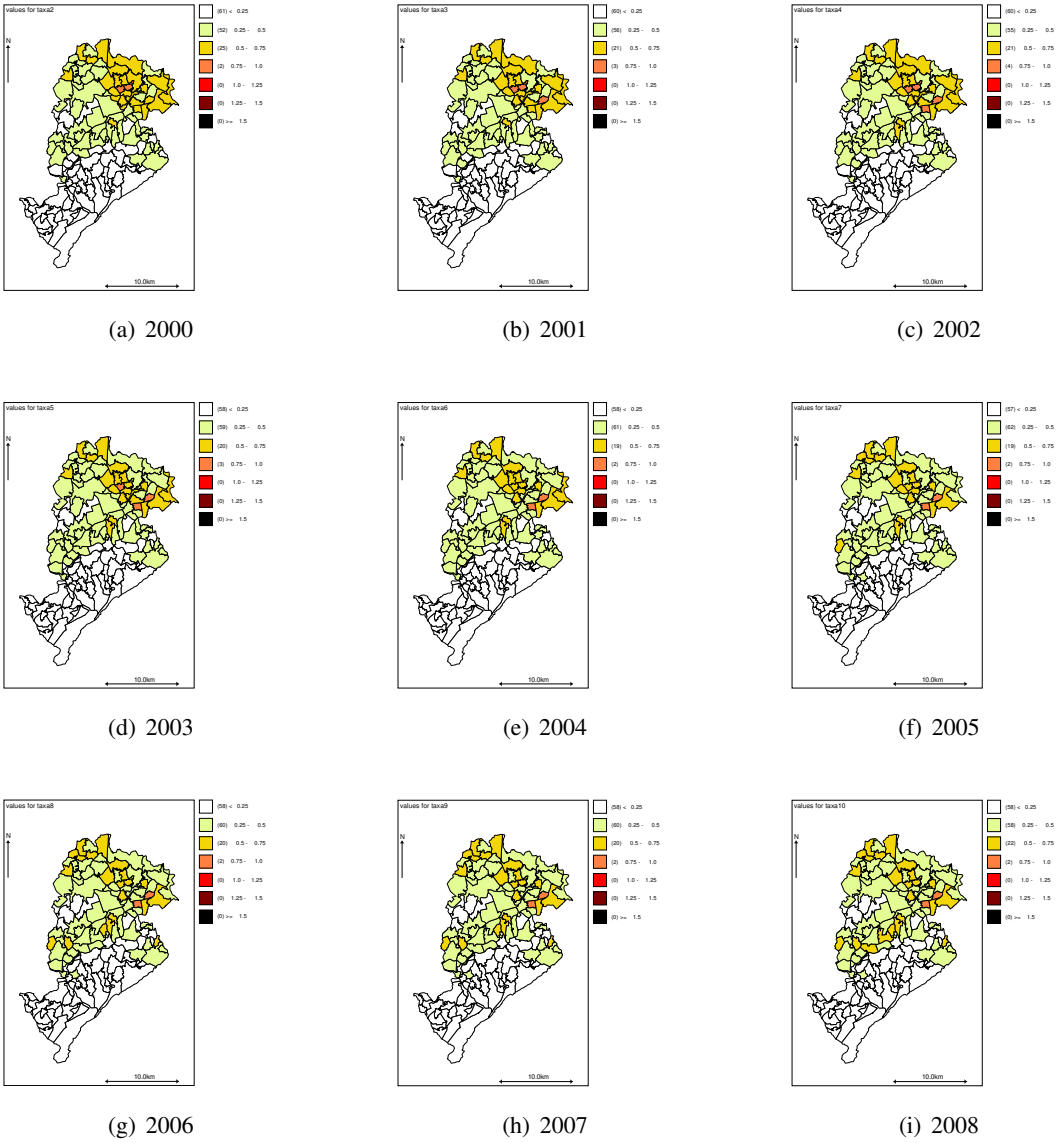
(g) 2006 (h) 2007 (i) 2008

Figura 2.5. Mapas das estimativas das taxas de incidência de LV em Humanos (a cada 10.000 **habitantes), calculadas pelo modelo de taxas brutas, na cidade de Belo Horizonte dividida em áreas de abrangência, nos anos de 2000 a 2008**

Ao observar os mapas, percebe-se a dificuldade em encontrar tendências temporais ou espaciais. Não conseguimos identificar áreas onde a LV tem aumentado ou diminuido continuamente com o passar dos anos, nem mesmo identificar trajetórias pelos quais a doença tem se espalhado, etc. Como já citamos anteriormente, a suavização das taxas através do ajuste de modelos bayesianos com efeitos espacial e/ou temporalmente estruturados pode ser uma alternativa.

2.4.2. Modelos proposto por Assunção et al.

Ao gerar os gráficos com as estimativas das taxas de incidência da LV, calculadas a partir do modelo proposto por Assunção et al., para a cidade de BH dividida em áreas de abrangência, nos anos de 2000 a 2008, obtivemos os mapas da Figura 2.6.



(a) 2000 (b) 2001 (c) 2002

(d) 2003 (e) 2004 (f) 2005

(g) 2006 (h) 2007 (i) 2008

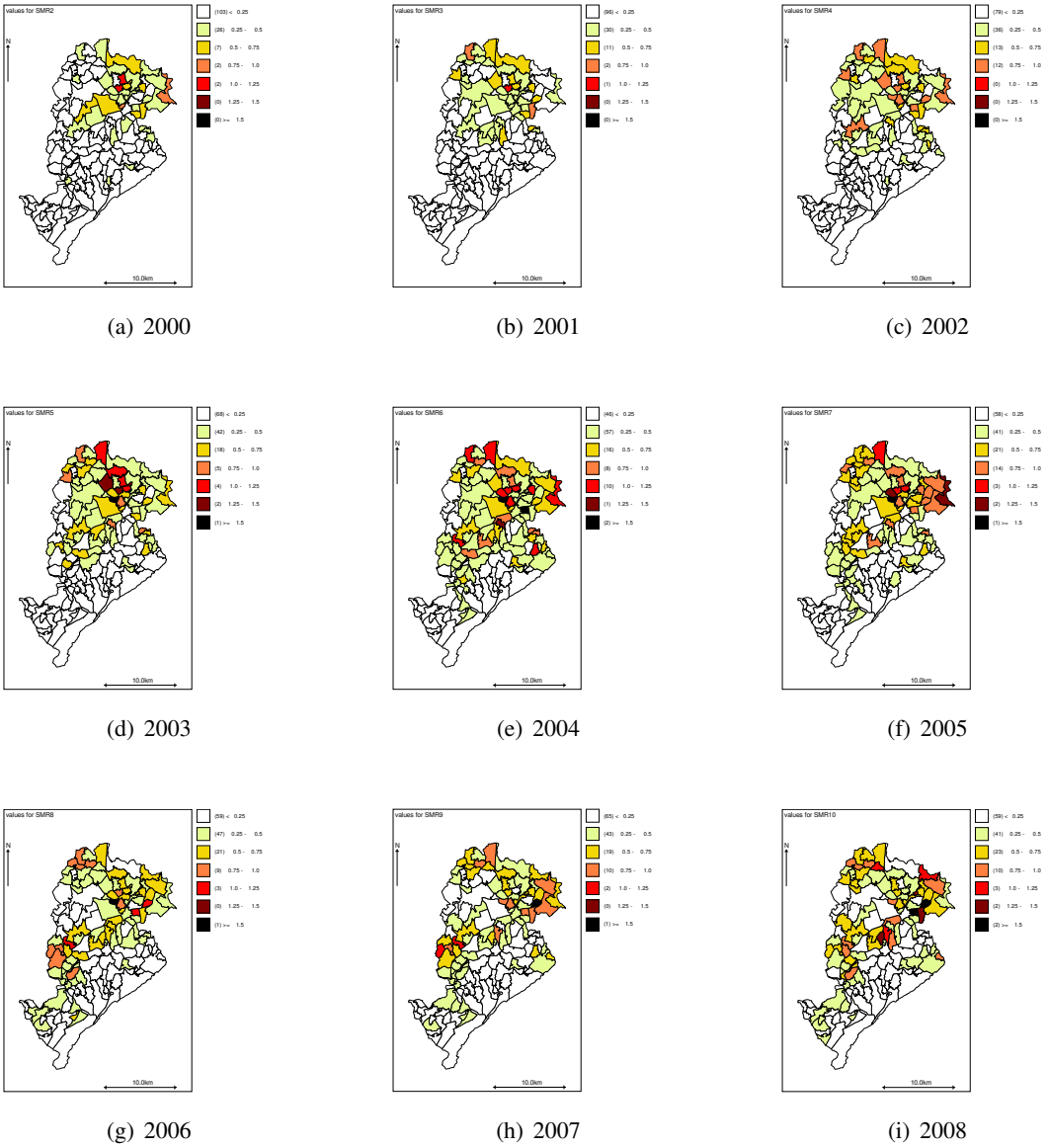
Figura 2.6. Mapas das estimativas das taxas de incidência de LV em Humanos (a cada 10.000 **habitantes), calculadas pelo modelo proposto por Assunção et al., na cidade de Belo Horizonte dividida em áreas de abrangência, nos anos de 2000 a 2008**

Observando os mapas da Figura 2.6, percebemos que há uma grande suavização das taxas em todos os anos. Os resultados obtidos por esse modelo diferem fortemente das taxas brutas. Eles não evidenciam tendências nem realçam a maioria das áreas similares às que aparecem com valores altos nos mapas feitos a partir do modelo de taxas brutas. Comparando estes mapas com os da Figura 2.5, podemos ver que há uma super suavização. Tudo isso nos indica que este modelo pode não se ajustar bem as dados de LV humana em Belo Horizonte, ou que boa parte da variação espaço-temporal observada na Figura 2.5 puro ruído não associado com o risco subjacente.

O valor do DIC calculado para este modelo foi: $19297,9$.

2.4.3. Modelo Proposto por Martínez-Beneito et al.

Ao gerar os gráficos com as estimativas das taxas de incidência da LV, calculadas a partir do modelo proposto por Martínez-Beneito et al., para a cidade de BH dividida em áreas de abrangência, nos anos de 2000 a 2008, obtivemos os mapas da Figura 2.7.



(a) 2000 (b) 2001 (c) 2002

(d) 2003 (e) 2004 (f) 2005

(g) 2006 (h) 2007 (i) 2008

Figura 2.7. Mapas das estimativas das taxas de incidência de LV em Humanos (a cada 10.000 habitantes), calculadas pelo modelo proposto por Martínez-Beneito et al., na cidade de Belo Horizonte dividida em áreas de abrangência, nos anos de 2000 a 2008

Observando os mapas da Figura 2.7, percebemos que há suavização das taxas em todos os anos, ao mesmo tempo em que há semelhança com as estimativas das taxas resultantes do modelos de taxas brutas. Nesses mapas, podemos observar algumas tendências de evolução temporal/espacial nas estimativas das taxas de incidência de LV humana em Belo Horizonte. No ano 2000, as estimativas mais altas se concentram nas regiões

Norte e Nordeste. Com o passar do tempo, pode-se observar um espalhamento da doença começando, principalmente, nas regiões da Pampulha e de Venda Nova. A região Noroeste também passa a ter estimativas elevadas das taxas de incidência. Nos anos finais já observa-se que a doença espalhou-se pela cidade como um todo, mas ainda há uma concentração de áreas com estimativas mais elevadas das taxas de incidência nas regiões Nordeste, Noroeste e Venda Nova.

O valor do DIC calculado para este modelo foi: 2729,4.

2.4.4. Comparação entre os Ajustes dos Modelos

Observando os mapas com as estimativas das taxas de incidência da LV nas áreas de abrangência de Belo Horizonte do 2000 a 2008, percebemos que:

- Nos mapas construídos com as estimativas obtidas pelo modelo de taxas brutas, não é possível acompanhar as tendências espaciais e/ou temporais, devido às grandes variações de uma área para outra vizinha ou de um ano para um ano seguinte, por exemplo. Isso se deve ao pequeno número de casos e à pouca população em cada área de abrangência, que faz com que uma mudança pequena no número de casos cause grande impacto na estimativa da taxa de incidência da área.
- Comparando os mapas construídos com as estimativas obtidas pelos modelos bayesianos ajustados com intenção de realizar uma suavização, observamos que o modelo proposto por Martínez-Beneito et al. consegue suavizar os mapas e ao mesmo tempo acompanhar as tendências de espalhamento da doença no espaço-tempo. Já o modelo proposto por Assunção et al. suaviza demais os mapas e não mostra nenhuma evolução espacial espacial e/ou tendência temporal.

Analisando os DICs obtidos para os dois modelos bayesianos, vemos que seu valor para o modelo proposto por Martínez-Beneito et al. é menor do que para o modelo proposto por Assunção et al., indicando que aquele modelo deve ajustar melhor os dados de LV em Belo Horizonte do que este.

2.5. Conclusões

Estudamos três modelos distintos, sendo que um deles segue o procedimento clássico de inferência e os outros dois, o procedimento bayesiano. Implementamos todos eles e ajustamos a dados de Leishmaniose Visceral Humana na cidade de Belo Horizonte entre os anos de 2000 e 2008. As unidades de área foram definidas como as Áreas de Abrangência dos Centros de Saúde e os períodos de tempo foram os anos.

Dos modelos bayesianos ajustados, o proposto por Assunção et al. modela a taxa de incidência da LV humana considerando um efeito aleatório espacialmente estruturado multiplicando o tempo, que entra no modelo de forma linear. Já o proposto por Martínez-Beneito et al. considera um efeito aleatório no qual há interação entre os efeitos espacialmente estruturados e os efeitos temporalmente estruturados, fazendo com que haja dependência entre áreas vizinhas e entre períodos de tempo vizinhos, ao mesmo tempo.

Os modelos foram ajustados e as estimativas obtidas foram exibidas nos mapas. Também foram calculados os DICs dos modelos bayesianos. A partir das comparações

realizadas, observou-se que o modelo proposto por Martínez-Beneito et al. é o que parece se ajustar melhor aos dados de LV humana em Belo Horizonte. Portanto, decidimos adotar este modelo para explicar a distribuição espaço-temporal da LV humana em Belo Horizonte. Dessa forma, esse modelo poderá ser utilizado daqui em diante pelas Secretarias de Saúde como uma ferramenta para direcionar e priorizar estratégias de controle.

2.6. Referências Bibliográficas

[1] Assunção RM, Reis IA, Oliveira CL. Diffusion and prediction of leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. *Statistics in Medicine.* 2001; **20**:2391–2335.

[2] Besag J. Spatial interaction and the statistical analysis of lattice systems (with discussions). *Journal of the Royal Statistical Society, Series B.* 1974; **36**:192–236.

[3] Besag J, Kooperberg C. On conditional and intrinsic autoregressions. *Biometrika.* 1995; **82**:733–746.

[4] Bernardinelli L, Montomoli C. Empirical Bayes Versus Fully Bayesian Analisys of Geographical Variation in Disease Risk. *Statistics in Medicine.* 1992; **11**:983–1007.

[5] Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis (2nd ed.).* 2004; Chapman & Hall/CRC, Boca Raton.

[6] Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine.* 2000; **19**:2555–2567.

[7] Martinez-Beneito MA, Lopez-Quilez A, Botella-Rocamora P. An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine.* 2008; **27**:2874–2889.

[8] Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. Manual de Vigilância e Controle da Leishmaniose Visceral. *Série A. Normas e Manuais Técnicos.* 2003; Editora MS, Brasília - DF.

[9] Spiegelhalter D, Thomas A, Best N, Lunn D. *WinBUGS User Manual (Version 1.4)* 2003; http://www.mrc-bsu.cam.ac.uk/bugs

2.7. Apêndice

- Código implementado no WinBugs para estimar as taxas de incidência de LV em Belo Horizonte para os anos de 2000 a 2008 utilizando o método de taxas brutas e o proposto por Assunção et al.

```
model {

# Likelihood
for (i in 1 : N) {
for (t in 1 : T) {
O[i,t]  ~ dpois(mu[i,t])
log(mu[i,t]) <- log(pop[i,t]) + theta[i,t]
theta[i,t] <- alpha0 + alpha[i] + beta[i]*(t-1)
}
  taxa1[i] <- exp(theta[i,1])*10000;
      taxa2[i] <- exp(theta[i,2])*10000;
      taxa3[i] <- exp(theta[i,3])*10000;
      taxa4[i] <- exp(theta[i,4])*10000;
  taxa5[i] <- exp(theta[i,5])*10000;
  taxa6[i] <- exp(theta[i,6])*10000;
  taxa7[i] <- exp(theta[i,7])*10000;
  taxa8[i] <- exp(theta[i,8])*10000;
  taxa9[i] <- exp(theta[i,9])*10000;
  taxa10[i] <- exp(theta[i,10])*10000;

      bruta1[i]<-(O[i,1]/pop[i,1])*10000;
      bruta2[i]<-(O[i,2]/pop[i,2])*10000;
      bruta3[i]<-(O[i,3]/pop[i,3])*10000;
  bruta4[i]<-(O[i,4]/pop[i,4])*10000;
      bruta5[i]<-(O[i,5]/pop[i,5])*10000;
      bruta6[i]<-(O[i,6]/pop[i,6])*10000;
      bruta7[i]<-(O[i,7]/pop[i,7])*10000;
  bruta8[i]<-(O[i,8]/pop[i,8])*10000;
  bruta9[i]<-(O[i,9]/pop[i,9])*10000;
  bruta10[i]<-(O[i,10]/pop[i,10])*10000;
}

for(k in 1:sumNumNeigh) {
weights[k] <- 1
}

 alpha0  ~ dflat()

# CAR prior distribution for alpha:
alpha[1:N] ~ car.normal(adj[], weights[], num[], tau.alpha)
```

```
# CAR prior distribution for betha:
beta[1:N] ~ car.normal(adj[], weights[], num[], tau.beta)


# Other priors:
tau.alpha  ~ dgamma(0.5, 0.0005)  # prior on precision for alpha
sigma.alpha <- sqrt(1 / tau.alpha) # standard deviation for alpha
tau.beta  ~ dgamma(0.5, 0.0005)  # prior on precision for betha
sigma.beta <- sqrt(1 / tau.beta) # standard deviation for betha

}
```

- Código implementado no WinBugs para estimar as taxas de incidência de LV em Belo Horizonte para os anos de 2000 a 2008 utilizando o método de taxas brutas e o proposto por Assunção et al.

```
model{
  for(i in 1:N){
    for(j in 1:T){
      O[i,j]~dpois(mu[i,j])
#Modelling of the mean for every municipality and period
      log(mu[i,j])<-log(pop[i,j])+mediainter+inter[j]+theta.ST[i,j
#SMR for every municipality and period
      SMR[i,j]<-10000*exp(mediainter+inter[j]+theta.ST[i,j])
    }
  }
for(i in 1:N){
  SMR1[i]<-SMR[i,1];
  SMR2[i]<-SMR[i,2];
  SMR3[i]<-SMR[i,3];
  SMR4[i]<-SMR[i,4];
  SMR5[i]<-SMR[i,5];
  SMR6[i]<-SMR[i,6];
  SMR7[i]<-SMR[i,7];
  SMR8[i]<-SMR[i,8];
  SMR9[i]<-SMR[i,9];
  SMR10[i]<-SMR[i,10];
}

#Spatio-temporal effect for the first period
    theta.S[1,1:N]~car.normal(adj[],weights[],num[],prec.spat)
    for(i in 1:N){
      BYM[i,1]~dnorm(theta.S[1,i],prec.het)
    }
    for(i in 1:N){
theta.ST[i,1]<-pow(1-ro*ro,-0.5)*BYM[i,1]
```

```
}
#Spatio-temporal effect for the subsequent periods
    for(j in 2:T){
        for(i in 1:N){
            theta.ST[i,j]<-ro*theta.ST[i,j-1]+BYM[i,j]
            BYM[i,j]~dnorm(theta.S[j,i],prec.het)
        }
        theta.S[j,1:N]~car.normal(adj[],weights[],num[],prec.spat)
    }

#Prior distribution for the mean risk for every
#municipality and period
    mediainter~dnorm(0,0.01)
#Prior distribution for the global time trend
    inter[1:T]~car.normal(adjT[],weightsT[],numT[],prec.inter)
#Prior distribution for the precision parameters in the model
    prec.inter~dgamma(0.5,0.005)
    prec.het~dgamma(0.5,0.005)
    prec.spat~dgamma(0.5,0.005)
#Prior distribution for the temporal dependence parameter
    ro~dunif(-1,1)
}
```

Capítulo 3

Teste de Independência Entre Dois Padrões de Pontos Espaço-Temporais

3.1. Introdução

O estudo de métodos estatísticos para análise de dados espaciais tem importância cada vez maior em diversas áreas, como ecologia, epidemiologia, demografia, geografia, entre outros. Talvez a principal razão para essa crescente procura por métodos de estatística espacial seja o frequente interesse em responder "quanto está em que local", ao invés de responder somente "quanto". Os dados espaciais contêm, além dos valores do atributo de interesse, as localizações espaciais relativas às observações.

Podemos classificar os dados espaciais em três tipos [1]:

- Dados pontuais referenciados, sendo $Y(s)$ um vetor aleatório em uma localização $s \in \Re^r$, onde s varia continuamente em D, um subconjunto fixo de $\Re^r$ que contem um retângulo r-dimensional de volume positivo. Suponha, por exemplo, que sejam implantadas algumas estações de monitoramento para medir o nível de poluição do ar em um estado. Cada estação tem sua localização s e um valor $Y(s)$, que pode ser a média dos níveis medidos a cada mês durante o ano de 2008. Dessa forma, temos $Y(s)$ vetor aleatório que receberá um valor para cada ponto s.
- Dados de área, onde D é novamente um subconjunto fixo (de forma regular ou irregular) mas agora particionado em um número finito de unidades de área com fronteiras bem definidas. Suponha, por exemplo, que temos um estado e consideraremos sua divisão em cidades. Para cada cidade, temos uma medida da quantidade de furtos ocorridos no ano de 2007. Nesse caso, não temos a localização pontual de onde ocorreu cada furto, mas temos, para cada área (cada cidade), o número total de furtos ocorridos.
- Dados de padrões pontuais, onde D agora é aleatório. Seu conjunto de índices dá a localização dos eventos aleatórios que compõem o padrão de pontos espacial. $Y(s)$ pode ser simplesmente igual a 1 para todo $s \in D$ (indicando a ocorrência do evento), ou pode dar alguma informação adicional de uma covariável (produzindo o denominado processo pontual marcado).

Neste trabalho, estamos interessados no terceiro tipo, os padrões de pontos espaciais, que podem ser exemplificados pelas localizações de residências das pessoas que têm uma determinada doença em uma cidade ou de árvores de determinada espécie em uma floresta. Temos que a resposta Y frequentemente é fixa (e, no caso de nosso interesse, igual a 1) e apenas as localizações s são aleatórias.

Em estatística espacial é comum considerar simultaneamente dois ou mais padrões de pontos espaciais [8]. Duas aplicações usuais podem ser citadas:

- Considere um padrão composto pelas localizações de residência de casos de uma doença em uma região de um plano e outro com o conjunto de localizações de residências de indivíduos classificados como controles. Geralmente, o interesse está em comparar as distribuições marginais dos dois processos, decidindo se os casos têm algum grau de aglomeração espacial em relação ao padrão dos controles [4, 6]. Se os casos e os controles tiverem padrões espaciais semelhantes, é razoável que a hipótese nula de que eles são amostras aleatórias independentes da mesma população de risco não seja rejeitada. É usual executar o teste condicionado ao número observado de casos e controles.
- Suponha que estamos estudando duas espécies de árvores em uma mesma região e que, de alguma forma, sabemos que elas têm configurações espaciais diferentes. Agora temos que analisar a distribuição conjunta dos processos. O interesse é testar a independência de dois padrões pontuais ou, alternativamente, se existe interação entre os dois processos. Sob a hipótese de independência, o número esperado de indivíduos de uma espécie em um disco centrado em $\mathbf{x} = (x_1, x_2)$ é independente da presença de um indivíduo da outra espécie nesse disco. Um teste desse tipo é a chamada função K_{12} [7].

Nesse trabalho, estamos interessados no segundo caso, no qual queremos testar a independência entre dois padrões pontuais. Porém, gostaríamos de fazer isso não para dados espaciais, mas para dados espaço-temporais. Em diversos estudos epidemiológicos é frequente a observação dos eventos espaciais em diferentes períodos de tempo, de forma a obter dados denominados espaço-temporais. Podemos pensar nesse tipo de dado como definido em três dimensões: as coordenadas x e y no espaço e a coordenada t no tempo. Por exemplo, além das localizações de residência dos indivíduos com determinada doença, podemos considerar também a data dos primeiros sintomas.

Com a intenção de testar a independência entre dois padrões pontuais de dados espaço-temporais, propomos uma extensão da função K_{12}, chamada aqui de Função Kt_{12}. Essa função será descrita detalhadamente na Seção 3.3 deste capítulo. Na Seção 3.4, serão apresentados testes da Função Kt_{12} em três possíveis cenários, utilizando dados gerados computacionalmente. Na Seção 3.5, aplicaremos a função desenvolvida a dados reais referentes a casos de Leishmaniose Visceral na cidade de Belo Horizonte. Finalmente, na Seção 2.5, apresentaremos as conclusões do trabalho.

3.2. A Função K_{12}

A função K_{12} [3, 7] é um teste comumente utilizado para verificar se dois processos pontuais estacionários [12], observados em uma janela finita A com área $|A|$, são independentes. Suponha que esses dois processos sejam N_1 e N_2. A função $K_{12}(r)$ é definida como o número de pontos do padrão N_1 que se localizam a uma distância espacial menor que r de um ponto arbitrário do padrão N_2, dividido pela intensidade de pontos no padrão N_1. Essa intensidade de pontos costuma ser representada por λ e é o número de pontos esperado por unidade de área em um padrão de pontos. Sob a hipótese nula de independência entre os padrões N_1 e N_2, teremos $K_{12}(r) = \pi r^2$.

Quando A é um retângulo, a função K_{12} é baseada em testes condicionais de Monte Carlo [2, 10]. Caso os processos sejam observados em uma área cuja forma é diferente de um retângulo, devemos inscrevê-la dentro de um. Suponha que n_1 e n_2 representam o número de eventos em N_1 e N_2, respectivamente, observados na janela retangular A. Seja $I_r(u) = 1$ se $u \leq r$ e 0 caso contrário, e seja u_{ij} a distância entre o i-ésimo evento de N_1 e o j-ésimo evento de N_2. Temos então a função empírica utilizada para estimar $K_{12}(r)$, representada por $\widetilde{K}_{12}$ [5] e definida da seguinte forma:

$$\widetilde{K}_{12}(r) = \frac{|A|}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I_r(u_{ij})$$

Para não haver problemas nas bordas do retângulo A, podemos transformá-lo em um toro, como mostra a Figura 3.1. Dessa forma, não é necessário fazer nenhum tipo de correção ao calcular os valores de $\widetilde{K}_{12}$, pois não existirão mais bordas, ou seja, áreas fora de A não serão abrangidas pelo círculo de raio r.

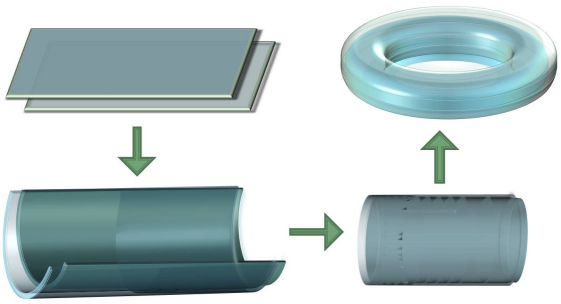


Figura 3.1. Tranformação do retângulo em toro

Definida a função empírica na qual está baseada a estatística de teste, podemos descrever o algoritmo que caracteriza o teste da função K_{12}.

- Estabeleça uma quantidade R de raios distintos r's.
- Para cada um dos valores de r, calcule o valor da função $\widetilde{K}_{12}(r)$.
- Faça um diagrama de dispersão com os valores de r ordenados contra os valores de $\widetilde{K}_{12}(r)$ correspondentes. Trace uma curva que passe por esses pontos.
- Repita s vezes o seguinte:
  - Deixe fixo o padrão de pontos N_1, por exemplo, e desloque o padrão N_2 aleatoriamente no toro.

  - Para cada um dos valores de r recalcule o valor da função $\widetilde{K}_{12}(r)$.
- Depois de uma quantidade s suficiente de deslocamentos aleatórios de um dos padrões em relação ao outro, teremos a distribuição empírica de $\widetilde{K}_{12}(r)$ sob a hipótese nula de que os processos pontuais estacionários N_1 e N_2 são independentes. Escolha o valor de confiança do teste (por exemplo, 95%).
- No gráfico onde foi traçada a curva da função com os padrões na posição original, trace o envelope com o grau de confiança escolhido, a partir dos valores encontrados para a distribuição empírica de $\widetilde{K}_{12}$.
- Caso a curva construída para os dados nas posições originais esteja dentro do envelope encontrado, o teste conclui que há evidências de que os processos pontuais N_1 e N_2 sejam espacialmente independentes. Caso contrário, conclui-se pela evidência de que deve haver dependência espacial entre os dois padrões pontuais. Neste caso, pode haver uma relação positiva (há tendência de ter mais pontos do tipo 1 próximos a um ponto de 2) ou negativa (há tendência de ter mais pontos do padrão 1 distantes de um ponto de 2). No caso da relação positiva, a curva construída pela função aparece acima da linha superior do envelope. No caso da negativa, ela aparece abaixo da linha inferior do envelope.

A transformação do retângulo em um toro, como apresentado, soluciona muito bem o problema que poderia ocorrer nas bordas do retângulo. Porém, para o desenvolvimento dos procedimentos da Seção 3.3, precisamos contornar esse problema de uma forma diferente. Para isso, tomamos um dos padrões pontuais, N_1 por exemplo e o copiamos de forma a tê-lo repetido quatro vezes. Posicionamos essas quatro réplicas da forma vista na Figura 3.2, que mostra o plano N_1 replicado quatro vezes e o plano N_2 sendo deslocado "sobre"ele. Dessa forma, é possível realizar o deslocamento do padrão N_2, em todas as direções, sem que haja problema nas bordas. Utilizaremos essa abordagem na Seção 3.3, como base para a elaboração da Função Kt_{12}.

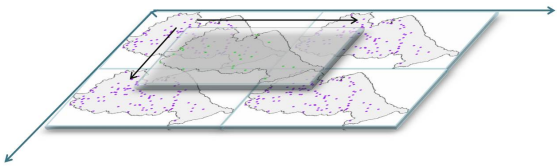


Figura 3.2. Um dos padrões de pontos replicado e o outro sendo deslocado

3.3. A Função Kt_{12}

Quando os dados a serem analisados variam também no tempo, não podemos utilizar a função K_{12} original descrita na seção anterior, a não ser que a intenção seja compará-los em um período de tempo fixo. Suponha, por exemplo, que queremos comparar padrões pontuais que representem os endereços dos hospedeiros e dos humanos infectados por um parasita causador de determinada doença. Suponha ainda que essa doença tem um período médio de incubação que varia de 1 a T dias. Então o humano infectado hoje pode vir a manisfestar e ser detectado com a doença dentro de um intervalo de tempo T a partir de hoje. O teste feito baseado na função K_{12} não é adequado nesse caso, já que temos que comparar padrões de pontos variando no tempo.

Para resolver esse problema, desenvolvemos uma função baseada na K_{12} para lidar com dados espaço-temporais. Chamaremos essa função de Kt_{12} de agora em diante. Considere dois processos pontuais estacionários N_1 e N_2 nos quais os eventos são espaço-temporais, ou seja, cada evento é caracterizado por sua localização em três dimensões (x, y e t). Fixe um intervalo de tempo T de forma que um evento aleatório de um dos padrões seja considerado próximo no tempo de um evento do outro padrão sempre que a distância temporal entre eles for menor que T. $Kt_{12}(r)$ é definida como o número de pontos do padrão N_1 que se localizam a uma distância espacial menor que r e a uma distância temporal menos que T de um ponto arbitrário do padrão N_2, dividido pela intensidade de pontos no padrão N_1. Nesse caso de três dimensões, a intensidade λ é o número de pontos esperado por unidade de volume em um padrão de pontos. Sob a hipótese nula de independência entre os padrões N_1 e N_2, teremos $Kt_{12}(r) = \pi r^2 T$.

Suponha que n_1 e n_2 representam o número de eventos em N_1 e N_2, respectivamente, observados na janela retangular A (lembrando que agora A tem três dimensões). Ao invés de um retângulo, teremos agora um cubo, como mostra a Figura 3.3, e $|A|$ corresponde ao volume desse cubo.

Figura 3.3. Representação de um cubo englobando todos os eventos espaço-temporais envolvidos no problema

Seja $I_r(u,t) = 1$ se $u \leq d$ e $t \leq T$ e 0, caso contrário, e sejam u_{ij} e t_{ij} as distâncias espacial e temporal, respectivamente, entre o i-ésimo evento de N_1 e o j-ésimo evento de N_2. Observe que no caso apenas espacial da função K_{12} eram contados os eventos de um dos padrões que estavam dentro de um círculo de raio r centrado em um evento aleatório do outro padrão. Agora, no caso espaço-temporal, esse círculo dá lugar a um cilindro de raio r e altura T. Temos então a função empírica $\widetilde{Kt}_{12}$ para estimar Kt_{12}, definida da seguinte forma:

$$\widetilde{Kt}_{12} = \frac{|A|}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I_r(u_{ij}, t_{ij})$$

.

Para não haver problemas nas bordas do cubo A, utilizaremos um procedimento semelhante ao que foi visto na Seção 3.2, na qual copiamos um dos padrões e deslocamos o outro. Faremos o mesmo com os cubos, como podemos observar na Figura 3.4, lembrando que agora os deslocamentos aleatórios serão realizados nas três direções (x, y e t). Dessa forma, não é necessário fazer nenhum tipo de correção ao calcular os valores de $\widetilde{Kt}_{12}$.

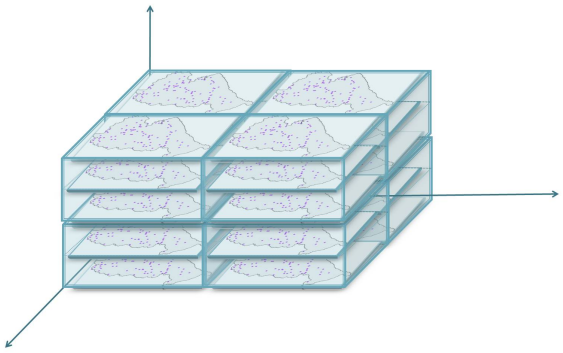


Figura 3.4. Solução segundo a abordagem de replicação de um dos cubos

Definida a função empírica na qual está baseada a estatística de teste, descreveremos o algoritmo que caracteriza o teste da função Kt_{12}.

- Estabeleça uma quantidade R de raios r's e um intervalo de tempo T.
- Para cada um dos valores de r, calcule o valor da função $\widetilde{Kt}_{12}$.
- Faça um diagrama de dispersão com os valores de r contra os valores de $\widetilde{Kt}_{12}$ correspondentes. Trace uma curva que passe por esses pontos.
- Repita s vezes o seguinte:
  - Deixe fixo o padrão de pontos N_1, por exemplo, e desloque o padrão N_2 aleatoriamente.
  - Para cada um dos valores de r, recalcule o valor da função $\widetilde{Kt}_{12}$.
- Depois de uma quantidade s suficiente de deslocamentos aleatórios de um dos padrões em relação ao outro, teremos a distribuição empírica de $\widetilde{Kt}_{12}$ sob a hipótese nula de que os processos pontuais N_1 e N_2 são independentes. Escolha o valor de confiança do teste (por exemplo, 95%).
- No gráfico onde foi traçada a curva da função com os padrões na posição original, trace o envelope com o grau de confiança escolhido, a partir dos valores encontrados para a distribuição empírica de $\widetilde{Kt}_{12}$.
- Caso a curva construída para os dados nas posições originais esteja dentro do envelope encontrado, o teste conclui que há evidências de que os processos pontuais N_1 e N_2 sejam espaço-temporalmente independentes. Caso contrário, conclui-se pela evidência de que há dependência espaço-temporal entre os dois padrões pontuais. Nesse caso, pode haver uma relação positiva (há tendência de ter mais

pontos do tipo 1 próximos a um ponto de 2) ou negativa (há tendência de ter mais pontos do padrão 1 distantes de um ponto de 2). No caso da relação positiva, a curva construída pela função aparece acima da linha superior do envelope. No caso da negativa, ela aparece abaixo da linha inferior do envelope.

3.4. Testes da Função Kt_{12}

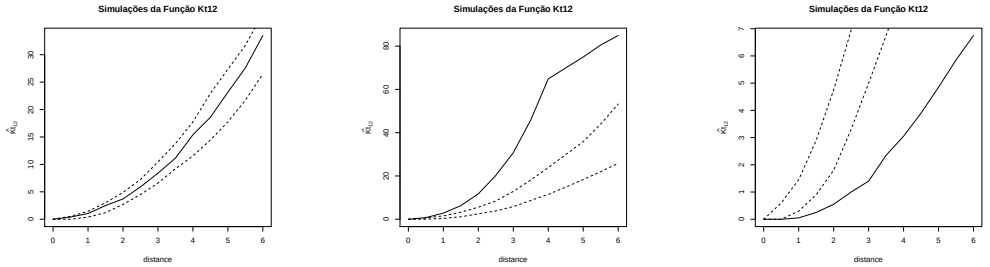
Com o objetivo de testar a função desenvolvida, foi realizada sua implementação utilizando o software R-2.9.1 [11]. Foram gerados os três cenários possíveis entre dois padrões distintos de pontos:

- No caso 1, supomos que não há dependência entre os padrões de pontos N_1 e N_2. Dessa forma, geramos aleatoriamente 20 pontos do tipo 1 e 10000 pontos do tipo 2 dentro de um cubo fictício de largura e comprimento igual a 100 e altura igual a 36.
- No caso 2, supomos que há dependência entre os padrões de pontos N_1 e N_2, sendo positiva a relação entre eles. O fato de um ponto do padrão N_1 estar localizado em uma determinada posição no espaço-tempo implica em uma maior chance de observar pontos do tipo 2 perto dessa posição. Pode-se dizer que os pontos do tipo 1 atraem os pontos do tipo 2.
  Para esse cenário, geramos aleatoriamente 20 pontos do tipo 1 e 9000 pontos do tipo 2 dentro de um cubo fictício de largura e comprimento igual a 100 e altura igual a 36. Posteriormente, para cada ponto do tipo 1, geramos mais 50 pontos do tipo 2, cujas distâncias entre cada um deles e seu ponto correspondente do tipo 1 fosse menor que 4 no espaço e menor que 6 no tempo. Dessa forma, geramos uma concentração maior de pontos do tipo 2 ao redor dos pontos do tipo 1.
- No caso 3, supomos que há dependência entre os padrões de pontos 1 e 2, sendo negativa a relação entre eles. O fato de um ponto do padrão A estar localizado em uma determinada posição no espaço-tempo implica uma menor chance de observar pontos do tipo 2 perto desta posição. Pode-se dizer que os pontos do tipo 1 repelem os pontos do tipo 2.
  Para esse cenário, geramos aleatoriamente 20 pontos do tipo 1 e 10000 pontos do tipo 2 dentro de um cubo fictício de largura e comprimento igual a 100 e altura igual a 36. Para cada ponto do tipo 2 cuja distância de algum ponto do tipo 1 seja menor que 4 no espaço e menor que 6 no tempo, fazemos sua exclusão do cenário com chance de 80%. Dessa forma, forçamos uma menor concentração de pontos do tipo 2 ao redor dos pontos do tipo 1.

Para cada um dos três cenários gerados, aplicamos a função Kt_{12}. O número de simulações realizadas para a construção do envelope foi de 200. Os gráficos resultantes podem ser observados na Figura 3.5.

Podemos observar que os resultados foram exatamente os esperados. No Caso 1, onde temos os pontos distribuídos aleatoriamente sem haver nenhuma relação espaço-temporal entre as distribuições dos padrões N_1 e N_2, a curva resultante da função Kt_{12} fica dentro do envelope obtido, indicando a independência entre os processos N_1 e N_2.

No Caso 2, onde há relação espaço-temporal entre as distribuições dos padrões N_1 e N_2 na forma de atração, a curva resultante da função Kt_{12} fica acima do limite



(a) Caso 1 (b) Caso 2 (c) Caso 3

Figura 3.5. Resultados dos testes da Função Kt_{12} nos possíveis cenários gerados

superior do envelope, indicando que há mais pontos de N_2 próximos aos de N_1 do que normalmente haveria caso houvesse independência entre espaço-temporal entre os dois processos. Portanto, a função indica que há evidências de relação espaço-temporal positiva entre os processos N_1 e N_2.

No Caso 3, onde há relação espaço-temporal entre as distribuições dos padrões N_1 e N_2 na forma de repulsão, a curva resultante da função Kt_{12} fica abaixo do limite inferior do envelope, indicando que há menos pontos de N_2 próximos aos de N_1 do que normalmente haveria caso houvesse independência espaço-temporal entre os dois processos. Portanto, a função indica que há evidências de relação espaço-temporal negativa entre os processos N_1 e N_2.

3.5. Aplicação

A Leishmaniose Visceral [9] (chamada de LV de agora em diante) é uma doença infecciosa, caracterizada por febre, aumento do fígado e baço, anemia, entre outros sintomas, levando o doente à morte se não for devidamente tratada. A transmissão se dá, principalmente, através da picada de flebotomíneo contaminado pelo parasita causador da doença. Atualmente, sabe-se que o cão doméstico tem importante papel na manutenção da doença, pois age como principal hospedeiro do parasita no meio urbano.

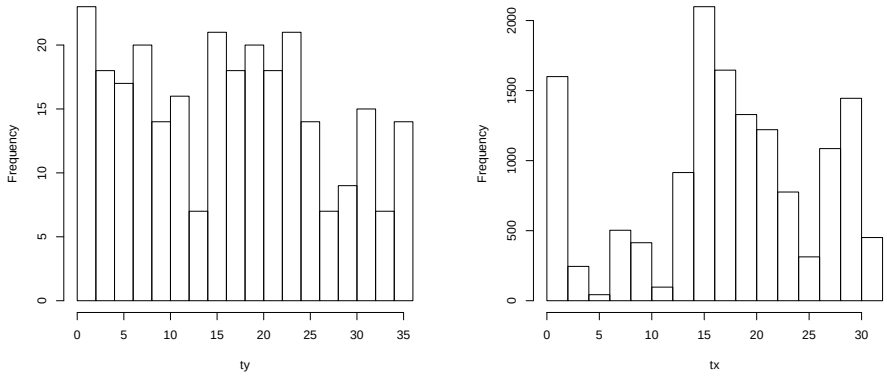
O período de incubação no homem (período entre a picada do mosquito transmissor e a manifestação dos sintomas) varia de 10 dias a 24 meses, com média de 2 a 6 meses.

Os dados disponíveis para análise são compostos pelas localizações pontuais e pela data dos casos de LV na cidade de Belo Horizonte (MG), tanto para os humanos quanto para os caninos, dos anos de 2006, 2007 e 2008. Os sistemas de coleta e armazenamento de dados das Secretarias de Saúde estão em fase de melhoria, de forma que os dados estão mais completos a cada ano. Porém, para o período estudado, há uma quantidade de dados faltantes no banco de dados de casos caninos, tanto devido a regiões onde não foi feito o inquérito censitário quanto devido a exames cujos resultados não foram armazenados no banco de dados. Mesmo cientes desses problemas, gostaríamos de verificar se há evidências de dependência espaço-temporal entre eles. Utilizaremos a Função Kt_{12} para fazer o teste, considerando $T = 6$ meses. O valor de T foi escolhido baseado nas informações obtidas sobre a LV, considerando o período de incubação médio

e dando uma margem de tempo para que um determinado cão seja constatado positivo para a doença após transmití-la a um humano.

O banco de dados utilizado consta de 2902, 7986 e 3295 casos caninos e 122, 109 e 154 casos humanos registrados nos anos de 2006, 2007 e 2008, respectivamente.

Os histogramas da Figura 3.6 mostram a quantidade de casos humanos e caninos, respectivamente, nos 36 meses analisados. Pode-se observar algumas semelhanças nas distribuições através do tempo. Em alguns momentos em que a quantidade de casos caninos diminui podemos observar uma diminuição dos casos humanos, assim como em alguns momentos em que a quantidade de casos caninos aumenta podemos observar um aumento na quantidade de casos humanos.



(a) Casos Humanos (b) Casos Caninos

Figura 3.6. Histogramas do número de casos humanos e de casos caninos de LV em Belo Horizonte durante os meses de Janeiro de 2006 a Dezembro de 2008

O mapa da Figura 3.7 mostra a distribuição dos casos caninos na forma de kernel e dos casos humanos na forma de pontos na cidade de Belo Horizonte durante os três anos citados. Podemos observar que as distribuições espaciais são semelhantes. As áreas onde há maior concentração de casos caninos também concentra um maior número de casos humanos, assim como as regiões onde há pouca concentração de casos caninos também parece ter poucos casos humanos.

Observando as Figuras 3.6 e 3.7, parece haver relação de dependência espacial e temporal entre os casos humanos e os casos caninos de LV em Belo Horizonte. Para verificar esses indícios, realizamos 100 simulações de Monte Carlo para obter a distribuição empírica $\widetilde{Kt}_{12}$. Escolhemos um intervalo de confiança de 95%. O resultado obtido na forma de gráfico é mostrado na Figura 3.8.

A curva obtida ficou acima do limite superior do envelope. Dessa forma, observamos evidência de haver dependência entre os dois padrões pontuais com 95% de confiança. Concluimos que há evidências de haver dependência espaço-temporal positiva entre os casos caninos e humanos de leishmaniose, ou seja, a proximidade espacial e temporal de cães positivos parece aumenta o risco de encontrarmos um caso humano positivo, e vice-versa. Esse resultado faz sentido se pensarmos com cuidado na forma de

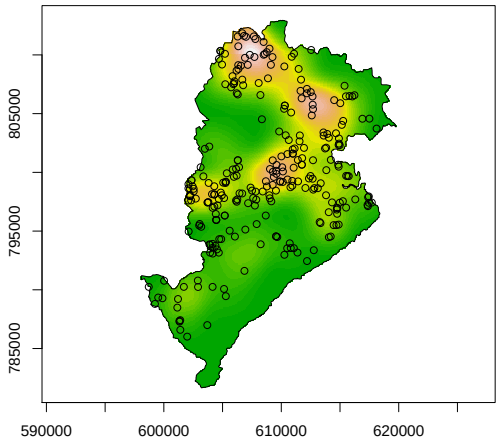


Figura 3.7. Mapa de Belo Horizonte com as localizações dos casos caninos de LV representados na forma de mapa de kernel e dos casos humanos representados na forma de pontos.

transmissão da LV. O mosquito pica um cão infectado e, ao picar um humano, transmite a doença a ele. Dessa forma, é esperado que humanos que vivem em regiões próximas espacialmente de locais onde há maior concentração de cães infectados tenham maior chance de contrair a doença. O mesmo ocorre em relação ao tempo, podendo haver uma certa variação devido ao tempo médio de incubação da doença nos humanos ser de 2 a 6 meses e devido ao tempo que os cães podem permanecer infectados sem serem detectados positivos.

3.6. Conclusões

Nesse capítulo, apresentamos o conceito de padrões de pontos espaciais e de dados espaço-temporais. Revisamos a função K_{12}, comumente utilizada para testar a independência espacial entre dois padrões pontuais. Mostramos que, em certos casos, quando a intenção não é verificar apenas a independência espacial, mas também a temporal entre os padrões, a função K_{12} não pode ser utilizada. Dessa forma, nos baseamos nela e propusemos uma nova função, chamada de Função Kt_{12}, com o objetivo de testar a independência espaço-temporal entre dois processos pontuais. Descrevemos a definição matemática e o algoritmo desenvolvidos.

Implementamos a função desenvolvida utilizando o software R. Geramos três possíveis cenários para testá-la, simulando claramente as três possíveis situações (ausência de correlação, presença de correlação positiva, presença de correlação negativa). Obtivemos os resultados esperados, indicando que a função atinge corretamente o objetivo proposto.

Na busca por dados reais para aplicar a função Kt_{12} desenvolvida, utilizamos os dados de casos humanos (padrão N_1) e casos caninos (padrão N_2) de Leishmaniose Visceral na cidade de Belo Horizonte, MG, ocorridos durante os anos de 2006, 2007 e

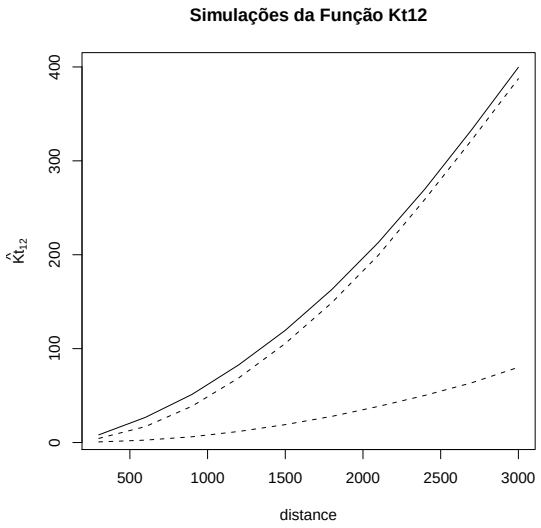


Figura 3.8. Resultado da aplicação da Função Kt_{12} aos dados de Leishmaniose Visceral em Belo Horizonte

2008. Consideramos próximos no tempo casos caninos distantes em até 6 meses dos casos humanos, levando em consideração o periodo de incubação da doença nos humanos e também a precariedade do sistema de detecção e eliminação de cães contaminados. O resultado obtido confirmou as suspeitas adquiridas na análise preliminar dos dados de que há evidências de que os dois padrões de pontos não são independentes, ou seja, onde há mais casos caninos há mais chance de ter um caso humano e vice-versa.

3.7. Referências Bibliográficas

[1] Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data.* 2004; Chapman & Hall/CRC Press, Boca Raton.

[2] Besag J, Diggle PJ. Simple Monte Carlo tests for spatial patterns. *Applied Statistics.* 1977; **26**:327–333.

[3] Diggle PJ. *Statistical Analysis of Spatial Point Patterns (2nd ed).* 2003; Oxford University Press Inc, New York.

[4] Diggle PJ. *Point process modelling in environmental epidemiology.* In: Barnett V, Turkman K (eds) Statistics for the Envirolment. 1993; pp 89–110.

[5] Hanisch KH, Stoyan D. Formulas for second-order analysis of marked point processes. *Mathematische Operationsforschung und Statistik, Series Statistics* 1982; **10**:555–560.

[6] Kelsall J, Diggle PJ. Kernel estimation of relative risk. *Bernoulli.* 1995; **1**:3–16.

[7] Lotwick HW, Silverman BW. Methods for analyzing spatial processes of several types of points. *Journal of the Royal Statistical Society B.* 1982; **44**:406–413.

[8] Mateu J. Parametric Procedures in the analysis of replicated pairwise interaction point patterns. *Biometrical Journal.* 2001; **43**:375–394.

[9] Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. Manual de Vigilância e Controle da Leishmaniose Visceral. *Série A. Normas e Manuais Técnicos.* 2003; Editora MS, Brasília - DF.

[10] Ripley BD. Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society B.* 1977; **39**:172–212.

[11] R Development Core Team. *R: A Language and Environment for Statistical Computing.* 2009. R Foundation for Statistical Computing. **URL**:http://cran.r-project.org/doc/manuals

[12] Shiryaev AN. *Probability. (2nd ed.)* 1996. Springer. pp 405–409

3.8. Apêndices

- Código da Função Kt_{12} implementada no software R.

```
k12hatleish <- function (ptsdog, ptshuman, tdog, thuman,
                         polybh, r, tmax, quiet=FALSE)
{
    ptsdogx <- ptsdog[, 1]
    ptsdogy <- ptsdog[, 2]
    nptsdog <- npts(ptsdog)
    ptshumanx <- ptshuman[, 1]
    ptshumany <- ptshuman[, 2]
    nptshuman <- npts(ptshuman)
    nr <- length(r)
    r <- sort(r)
    np <- length(polybh[, 1])
    polyx <- c(polybh[, 1], polybh[1, 1])
    polyy <- c(polybh[, 2], polybh[1, 2])
    h12 <- rep(0, times = nr)
    dmax <- r[nr]

    for(i in 1:nptshuman){
xi <- rep(ptshumanx[i], times=nptsdog)
yi <- rep(ptshumany[i], times=nptsdog)
ti <- rep(thuman[i], times=nptsdog)
xi <- xi-ptsdogx
yi <- yi-ptsdogy
ti <- ti-tdog
d <- xi*xi + yi*yi
d<-sqrt(d)
dt <- sqrt(ti*ti)

for(j in 1:nptsdog){
if((d[j] < as.double(dmax)) && (dt[j]<tmax)){
k <- 0
while((d[j]<as.double(r[nr-k])) && (k<nr)){
h12[nr-k] <- h12[nr-k]+1
k <- k+1
}
}
}
   }
   h12 <- h12/nptshuman
   h12
}
```

- Código do cenário gerado para representar ausência de relação entre as distribuições espaço-temporal dos padrões A e B (Caso 1)

```
require(splancs)
quadrado <- as.points(c(0,100,100,0),c(0,0,100,100))

nA <- 20
pontosA <- as.points(100*runif(nA),100*runif(nA))
temposA <- (36*runif(nA))

nB <- 10000
pontosB <- as.points(100*runif(nB),100*runif(nB))
temposB <- (36*runif(nB))
```

- Código do cenário gerado para representar presença de relação positiva entre as distribuições espaço-temporal dos padrões A e B (Caso 2)

```
require(splancs)
quadrado <- as.points(c(0,100,100,0),c(0,0,100,100))

nA <- 20
pontosA <- as.points(100*runif(nA),100*runif(nA))
temposA <- (36*runif(nA))

nB <- 10000
pontosB <- as.points(100*runif(nB),100*runif(nB))
temposB <- (36*runif(nB))

num <- as.integer(1000/nA)
for(i in 1:nA){
x <- rep(pontosA[i,1], times=num)
y <- rep(pontosA[i,2], times=num)
t <- rep(temposA[i], times=num)
posneg <- runif(num)
posnegt <- runif(num)
sort <- 4*runif(num)
distaux <- sqrt(16-sort*sort)
sort2 <- distaux*runif(num)
sortt <- 6*runif(num)
ptsx <- rep(0, times=num)
ptsy <- rep(0, times=num)
tps <- rep(0, times=num)
for(j in 1:num){
if(posneg[j]<0.25){
ptsx[j] <- (x[j]+sort[j])
ptsy[j] <- (y[j]+sort2[j])
}
else if(posneg[j]<0.5){
ptsx[j] <- (x[j]-sort[j])
```

```
ptsy[j] <- (y[j]+sort2[j])
}
else if(posneg[j]<0.75){
ptsx[j] <- (x[j]+sort[j])
ptsy[j] <- (y[j]-sort2[j])
}
else{
ptsx[j] <- (x[j]-sort[j])
ptsy[j] <- (y[j]-sort2[j])
}
if(posnegt[j]<0.5){
tps[j] <- (t[i]+sortt[j])
}
else{
tps[j] <- (t[i]-sortt[j])
}
}
for(k in 1:num){
auxpos <- (nB-1000)+num*(i-1)+k
pontosB[auxpos,1] <- ptsx[k]
pontosB[auxpos,2] <- ptsy[k]
temposB[auxpos] <- tps[k]
}
}
```

- Código do cenário gerado para representar presença de relação negativa entre as distribuições espaço-temporal dos padrões A e B (Caso 3)

```
require(splancs)
quadrado <- as.points(c(0,100,100,0),c(0,0,100,100))

nA <- 20
pontosA <- as.points(100*runif(nA),100*runif(nA))
temposA <- (36*runif(nA))

nB <- 10000
pontosB <- as.points(100*runif(nB),100*runif(nB))
temposB <- (36*runif(nB))

decisao <- rep(0,nB)

for(i in 1:nA) {

var1 <- rep(pontosA[i,1],nB)
aux1 <- (var1-pontosB[,1])^2

var2 <- rep(pontosA[i,2],nB)
aux2 <- (var2-pontosB[,2])^2
```

```
var3 <- rep(temposA[i],nB)
aux3 <- (var3-temposB)^2

dist <- (aux1+aux2)

perto <- ((dist < 16) && (aux3<36))
sort <- runif(nB,0,1)
aux <- sort*perto
decisao <- decisao + (aux > 0.2)
}
decisao <- (decisao>0)
ptsBi <- pontosB[decisao==0,1]
ptsBj <- pontosB[decisao==0,2]
ptsB <- as.points(ptsBi,ptsBj)
tpsB <- temposB[decisao==0]
```