

Modelos Dinâmicos Hierárquicos Espaço-Temporais para Dados na Família Exponencial

Ana Carolina C. da Costa Marina S. Paez Flávia M. P. F. Landim

Universidade Federal do Rio de Janeiro
Instituto de Matemática
Departamento de Métodos Estatísticos

Resumo

Neste trabalho foi proposta uma classe de modelos dinâmicos hierárquicos para respostas pertencentes à família exponencial de distribuições, com o principal objetivo de obter um modelo capaz de capturar a estrutura espaço-temporal presente nos dados, permitindo fazer interpolação espacial e previsão temporal para processos de natureza não gaussiana. Os modelos foram estruturados de forma hierárquica com parâmetros variando no tempo e/ou no espaço. A variação temporal foi definida através da especificação de estruturas dinâmicas no tempo e a variação espacial através de funções de correlação paramétricas. Foram propostos quatro modelos para processos de contagem e contínuos, contemplando também a possibilidade de conjuntos de dados inflacionados de zeros. Em particular, foram considerados os casos em que a variável resposta tem distribuição Poisson e uma mistura de duas outras distribuições: Gama e Bernoulli. O procedimento de inferência foi feito sob o enfoque bayesiano e uma metodologia de estimação foi desenvolvida através dos métodos de Monte Carlo via cadeias de Markov (MCMC), utilizando para a validação do algoritmo de estimação, conjuntos de dados artificiais gerados a partir dos modelos propostos.

A aplicação dos modelos espaço-temporais e do procedimento de inferência foi realizada para dois conjuntos de dados reais. O primeiro referente às observações trimestrais da quantidade de chuva coletadas ao longo de 103 anos na região leste da Austrália e o segundo relativo ao número de tornados reportados anualmente em 32 estados norte americanos, ao longo de 60 anos.

Palavras-Chaves: modelos espaço-temporais, inferência bayesiana, modelos dinâmicos, família exponencial, modelos de mistura, Monte Carlo via cadeias de Markov.

1 Introdução

Ao longo da década de 90, houve um aumento substancial no número de pesquisas sobre modelagem espaço-temporal. Isto ocorreu devido à importância do entendimento de fenômenos espaço-temporais como soluções de questões científicas em diversas áreas, tais como ciências ambientais, economia, mapeamento de doenças, etc. O poder crescente dos computadores tem tornado possível a análise de grandes bancos de dados que variam no espaço e no tempo, através de modelos cada vez mais realistas.

Harrison e Stevens (1976) propõem o uso de uma ampla classe de modelos para tratar séries temporais, denominada modelos dinâmicos. Diversos modelos espaço-temporais baseados em modelos dinâmicos têm sido propostos na literatura para uma gama de aplicações como, por exemplo, em meteorologia (Ghil et al., 1981), análise de ozônio (Guttorp et al., 1994) e meio ambiente (Wikle et al. (1998) e Huerta et al. (2004)). A maior classe destes modelos é aquela que se baseia no uso da família exponencial de distribuições (West e Harrison (1997)) para a modelagem das observações.

Zellner et al. (1991) e Min e Zellner (1993) mostraram que a introdução de uma estrutura hierárquica na modelagem de séries temporais multivariadas com parâmetros que variam no tempo produz ganhos na previsão. Uma combinação entre os modelos hierárquicos de Lindley e Smith (1972) e os modelos

dinâmicos de [Harrison e Stevens \(1976\)](#) foi proposta por [Gamerman e Migon \(1993\)](#), introduzindo os modelos dinâmicos hierárquicos.

As estruturas dos modelos dinâmicos hierárquicos podem ser estendidas em várias direções. Por exemplo, quando se deseja modelar processos indexados ao longo do tempo, é natural pensar que observações em instantes de tempo subsequentes sejam altamente correlacionadas. Por outro lado, se as observações são espacialmente referenciadas, é natural esperar que para localizações próximas entre si, o processo se comporte de forma semelhante e, conforme a distância entre as localizações aumenta, as observações vão se tornando independentes. Especificamente, o interesse está em modelar processos que variam simultaneamente ao longo do tempo e espaço, e que, além disso, possam ser modelados por uma distribuição pertencente à família exponencial.

Na seção 2, uma formulação geral dos modelos dinâmicos hierárquicos com estrutura espaço-temporal para dados na família exponencial será apresentada. Posteriormente, serão discutidos os procedimentos de inferência e os esquemas de amostragem utilizados nestes modelos para os casos específicos em que a variável resposta tem distribuição Poisson, além de uma mistura de duas outras distribuições: Gama e Bernoulli. Na seção 3, os modelos propostos serão aplicados a dois conjuntos de dados reais. A primeira aplicação é referente à modelagem do número de tornados reportados nos Estados Unidos e a segunda se refere à modelagem da quantidade de chuva na região leste da Austrália. Por fim, na seção 4 serão apresentadas as conclusões e possíveis extensões deste trabalho.

2 Formulação Geral dos Modelos

Considere um conjunto de períodos de tempo discretos, $t = 1, 2, \dots, T$, em que para cada t , observa-se a variável aleatória $\mathbf{Y}_t(\mathbf{s})$ em um conjunto de N locais no espaço $\mathbf{s} = (s_1, s_2, \dots, s_N)$. Para simplificar a notação, $\mathbf{Y}_t(\mathbf{s})$ será denotado por \mathbf{Y}_t . Suponha que \mathbf{Y}_t tenha distribuição pertencente à família exponencial com média $\boldsymbol{\mu}_t$, ou seja,

$$\mathbf{Y}_t \sim \mathcal{FE}(\boldsymbol{\mu}_t). \quad (1)$$

Suponha também que uma função dessa média possa ser modelada através de uma equação de regressão, de forma que

$$g(\boldsymbol{\mu}_t) = \mathbf{F}_{1,t}\boldsymbol{\theta}_{1,t}, \quad (2)$$

em que $g(\cdot)$ é uma função de ligação apropriada que liga o preditor linear a uma função da média da distribuição, $\mathbf{F}_{1,t}$ é uma matriz de regressão com dimensão $N \times Np$ e $\boldsymbol{\theta}_{1,t}$ é um vetor Np -dimensional de estados no tempo t , com p representando o número de covariáveis acrescido de uma unidade. Adicionalmente, as equações estrutural e de sistema são dadas, respectivamente, por

$$\boldsymbol{\theta}_{1,t} = \mathbf{F}_{2,t}\boldsymbol{\theta}_{2,t} + \mathbf{v}_{2,t} \quad \text{com } \mathbf{v}_{2,t} \sim PG(\mathbf{0}, \mathbf{V}_{2,t}), \quad (3)$$

$$\boldsymbol{\theta}_{2,t} = \mathbf{G}_t\boldsymbol{\theta}_{2,t-1} + \boldsymbol{\omega}_t \quad \text{com } \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}_t), \quad (4)$$

em que $\mathbf{F}_{2,t}$ é uma matriz de regressão com dimensão dada por $Np \times p$ e $\boldsymbol{\theta}_{2,t}$ é um vetor p -dimensional que evolui no tempo de acordo com a matriz \mathbf{G}_t de dimensão $p \times p$. Além disso, $\mathbf{V}_{2,t}$ é uma matriz $N.p \times N.p$ de covariância dos $\mathbf{v}_{2,t}$'s que incorpora a dependência espacial no modelo e \mathbf{W}_t é uma matriz de covariância $p \times p$.

Outro modelo com formulação semelhante àquela definida acima também será considerado. Neste modelo, comparativamente à equação de observação definida em (2) e com o objetivo de capturar alguma variabilidade não explicada pelas covariáveis, um efeito aleatório é incluído na função que modela a média da distribuição. Desta forma, tem-se que

$$g(\boldsymbol{\mu}_t) = \mathbf{F}_{1,t}\boldsymbol{\theta}_{1,t} + \mathbf{v}_{1,t} \quad \text{com } \mathbf{v}_{1,t} \sim N(\mathbf{0}, \mathbf{V}_{1,t}), \quad (5)$$

em que $\mathbf{V}_{1,t}$ é a matriz de covariância $N \times N$ do vetor de erros $\mathbf{v}_{1,t}$. Em ambos os modelos, os vetores $\boldsymbol{\theta}_{1,t}$ e $\boldsymbol{\theta}_{2,t}$ são, por hipótese, independentes dos vetores de erros $\mathbf{v}_{1,t}$ e $\mathbf{v}_{2,t}$, respectivamente, assim como $\boldsymbol{\theta}_{2,t-1}$ e $\boldsymbol{\omega}_t$ são independentes. As matrizes $\mathbf{F}_{1,t}$, $\mathbf{F}_{2,t}$ e \mathbf{G}_t são conhecidas, com $\mathbf{F}_{1,t}$ e $\mathbf{F}_{2,t}$ podendo incorporar covariáveis.

Neste estudo será abordado um caso particular em que $\mathbf{W}_t = \mathbf{W}$ para todo t e $\mathbf{V}_{1,t} = \mathbf{V}_1 = \sigma^2\mathbf{I}_N$, com \mathbf{I}_N denotando a matriz identidade de ordem N . A estrutura de correlação espacial será definida

através da distribuição de $\mathbf{v}_{2,t}$. Assim, a correlação espacial é dada somente através dos parâmetros de regressão $\theta_{1,t}$. Trabalhando sob a hipótese de isotropia, a função de covariância depende somente de $d_{i,j}$, a distância euclidiana entre as localizações s_i e s_j . A matriz $\mathbf{V}_{2,t}$ define a dependência espacial no modelo através da função de correlação exponencial e é tal que $\mathbf{V}_{2,t} = \mathbf{V}_2 = (\tau^2 \exp\{-\phi \cdot d_{i,j}\}) \otimes \mathbf{I}_p$. Além da função de correlação exponencial, outras formas mais gerais também podem ser consideradas.

2.1 Distribuição Poisson

Modelos para dados de contagem são amplamente empregados na modelagem de fenômenos em diversas áreas de estudo. Nesta seção, serão apresentados modelos para variáveis que representam contagens de um determinado fenômeno distribuídas de forma discreta no espaço e no tempo. Particularmente, para a modelagem de dados com estrutura espaço-temporal, um dos modelos dinâmicos comumente utilizados é o modelo Poisson.

2.1.1 Modelo sem Efeito de Sobredispersão

Considere que a variável aleatória de interesse $Y_t(s_i)$, para $i = 1, 2, \dots, N$ e $t = 1, 2, \dots, T$, represente contagens, condicionalmente independentes, de um determinado fenômeno e possa ser assumido que tenha distribuição de Poisson com média $\mu_t(s_i)$ ou, especificamente, sua função de probabilidade seja da seguinte forma:

$$p(Y_t(s_i) | \mu_t(s_i)) = \frac{\mu_t(s_i)^{y_t(s_i)} \exp\{-\mu_t(s_i)\}}{y_t(s_i)!}.$$

Para a modelagem de dados de contagem com estrutura espaço-temporal, um dos modelos dinâmicos comumente utilizados é o modelo Poisson, que assume intrinsecamente que a média e a variância do processo sob observação são iguais e evoluem no espaço e no tempo através de uma estrutura dinâmica. Assim sendo, o primeiro modelo apresentado na seção 2 pode ser reescrito como

$$\mathbf{Y}_t | \boldsymbol{\mu}_t \sim Poi(\boldsymbol{\mu}_t), \quad t = 1, 2, \dots, T \quad (6a)$$

$$\log(\boldsymbol{\mu}_t) = \mathbf{F}_{1,t} \boldsymbol{\theta}_{1,t} \quad (6b)$$

$$\boldsymbol{\theta}_{1,t} = \mathbf{F}_{2,t} \boldsymbol{\theta}_{2,t} + \mathbf{v}_{2,t}, \quad \mathbf{v}_{2,t} \sim PG(\mathbf{0}, \tau^2 \exp\{-\phi \cdot d_{i,j}\} \otimes \mathbf{I}_p) \quad (6c)$$

$$\boldsymbol{\theta}_{2,t} = \mathbf{G}_t \boldsymbol{\theta}_{2,t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}). \quad (6d)$$

O intuito de usar modelos de variação espacialmente contínua para dados de contagem leva à idéia de dissolver os limites das áreas em superfícies contínuas.

Procedimento de Inferência

Assumindo que o processo \mathbf{Y}_t possa ser modelado como em (6), o objetivo então é estimar as quantidades desconhecidas envolvidas. Defina $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_T)'$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_T)'$, $\boldsymbol{\theta}_1 = (\boldsymbol{\theta}_{1,1}, \boldsymbol{\theta}_{1,2}, \dots, \boldsymbol{\theta}_{1,T})'$, $\boldsymbol{\theta}_2 = (\boldsymbol{\theta}_{2,1}, \boldsymbol{\theta}_{2,2}, \dots, \boldsymbol{\theta}_{2,T})'$ e $\boldsymbol{\Psi}$ como a coleção de todos os parâmetros. Sob o modelo (6) e considerando o caso em que $p > 1$, o vetor paramétrico a ser estimado é dado por

$$\boldsymbol{\Psi} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{2,0}, \phi, \tau^2, \mathbf{W}). \quad (7)$$

Como a inferência será feita sob o enfoque bayesiano precisa-se atribuir uma distribuição *a priori* para o vetor paramétrico $\boldsymbol{\Psi}$. É razoável assumir que os hiperparâmetros são independentes *a priori*, e portanto, a distribuição conjunta é dada pelo produto das distribuições marginais. As distribuições *a priori* para os vetores paramétricos $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ seguem diretamente da definição das estruturas dinâmicas em 6c e 6d, respectivamente. Para os demais elementos que compõem o vetor paramétrico, denominados hiperparâmetros, as distribuições *a priori* atribuídas estão especificadas a seguir.

A informação *a priori*, D_0 , a respeito do vetor $\boldsymbol{\theta}_{2,0}$, é quantificada no instante inicial $t = 0$, através do primeiro e segundo momentos da distribuição Normal, ou seja, $\boldsymbol{\theta}_{2,0} | D_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$. Para o parâmetro ϕ foi atribuída a *priori* Gama, ou seja, $\phi \sim Ga(\alpha_\phi, \beta_\phi)$, em que $Ga(\alpha_\phi, \beta_\phi)$ denota a distribuição Gama com média α_ϕ / β_ϕ e variância $\alpha_\phi / \beta_\phi^2$, para $\alpha_\phi, \beta_\phi > 0$. Para o parâmetro τ^2 , a variância do processo gaussiano, foi atribuída uma *priori* tal que $\tau^2 \sim GI(\alpha_\tau, \beta_\tau)$, em que $GI(\alpha_\tau, \beta_\tau)$ denota a distribuição Gama invertida com média $\beta_\tau / (\alpha_\tau - 1)$, para $\alpha_\tau > 1$ e variância $\beta_\tau^2 / (\alpha_\tau - 1)^2 (\alpha_\tau - 2)$, para $\alpha_\tau > 2$.

Para a matriz de covariância \mathbf{W} foi assumido *a priori* que $\mathbf{W} \sim WI(m, \mathbf{S})$, em que $WI(m, \mathbf{S})$ denota a distribuição Wishart invertida com m graus de liberdade e matriz de escala \mathbf{S} .

Sob o enfoque bayesiano, toda informação necessária sobre um parâmetro está descrita por sua distribuição *a posteriori*. Dada toda a informação $D_T = (D_0, y_1, y_2, \dots, y_T)$, a distribuição *a posteriori* do vetor paramétrico Ψ é dada por

$$\begin{aligned} p(\Psi | D_T) &\propto p(\theta_{2,0})p(\mathbf{W})p(\tau^2)p(\phi) \prod_{t=1}^T p(\theta_{2,t} | \theta_{2,t-1}, \mathbf{W}) \\ &\times \left[\prod_{t=1}^T \prod_{i=1}^N p(\theta_{1,t}(s_i) | \theta_{2,t}, \phi, \tau^2) p(y_t(s_i) | \theta_{1,t}(s_i)) \right]. \end{aligned}$$

Como a distribuição $p(\Psi | D_T)$ não tem solução analítica, uma forma de obter amostras desta distribuição é a utilização de métodos de simulação estocástica. Os métodos MCMC são uma alternativa em problemas complexos nos quais a distribuição *a posteriori* não pode ser explicitada. O conjunto de parâmetros Ψ é amostrado através de um algoritmo que utiliza o amostrador de Gibbs. A matriz \mathbf{W} é amostrada através de sua condicional completa *a posteriori*, cuja forma é conhecida. Os parâmetros $\theta_{2,0}$ e θ_2 são amostrados através do algoritmo FFBS (do inglês *Forward Filtering Backward Sampling*) (Frühwirth-Schnatter (1994) e Carter e Kohn (1994)). Por fim, θ_1 , τ^2 e ϕ são amostrados através de passos de Metropolis-Hastings, após a especificação de distribuições propostas adequadas. A distribuição proposta para $\theta_{1,t}$ é uma distribuição Normal centrada no valor corrente do parâmetro. Para a amostragem dos parâmetros τ^2 e ϕ foram propostas distribuições Lognormais centradas nos valores correntes dos respectivos parâmetros. As variâncias das distribuições propostas no algoritmo de Metropolis-Hastings são fixas para cada parâmetro e foram ajustadas de forma que as taxas de aceitação estivessem entre 20% e 50%.

Previsão Temporal

Suponha que se deseja trabalhar com um horizonte de previsão de h passos a frente, a partir de um instante T qualquer. Dessa forma, o interesse é obter amostras da distribuição $(\mathbf{Y}_{T+h} | \mathbf{Y})$. A distribuição conjunta *a posteriori* das observações passadas pode ser escrita como

$$\begin{aligned} p(\mathbf{Y}_{T+h}, \theta_{1,T+h}, \theta_{2,T+h}, \Psi | \mathbf{Y}) &= p(\mathbf{Y}_{T+h} | \theta_{1,T+h})p(\theta_{1,T+h} | \theta_{2,T+h}, \tau^2, \phi) \\ &\times p(\theta_{2,T+h} | \theta_{2,T}, \mathbf{W})p(\Psi | \mathbf{Y}). \end{aligned}$$

Da equação acima tem-se que $p(\Psi | \mathbf{Y})$ pode ser amostrada conforme descrito no último parágrafo da subseção 2.1.1. As distribuições $p(\mathbf{Y}_{T+h} | \theta_{1,T+h})$ e $p(\theta_{1,T+h} | \theta_{2,T+h}, \tau^2, \phi)$ são conhecidas, e $p(\theta_{2,T+h} | \theta_{2,T}, \mathbf{W})$ pode ser facilmente obtida através do uso repetido da equação de evolução de $\theta_{2,T}$, tal que

$$\theta_{2,T+h} = \mathbf{G}_{T+h}\theta_{2,T+h-1} + \omega_{T+h} = \mathbf{H}_{T+h}(h)\theta_{2,T} + \sum_{k=1}^h \mathbf{H}_{T+h}(h-k)\omega_{T+k},$$

em que $\mathbf{H}_{T+h}(k) = \mathbf{G}_{T+h}\mathbf{G}_{T+h-1} \dots \mathbf{G}_{T+h-k+1}$ para todo t e inteiro $k \leq h$, com $\mathbf{H}_{T+h}(0) = \mathbf{I}_p$. Assim, a partir das suposições de independência feitas na seção 2 e utilizando propriedades da distribuição Normal, segue que

$$\theta_{2,T+h} | \theta_{2,T}, \mathbf{W} \sim N \left(\mathbf{H}_{T+h}(h)\theta_{2,T}, \sum_{k=1}^h \mathbf{H}_{T+h}(h-k)\mathbf{W}\mathbf{H}'_{T+h}(h-k) \right). \quad (9)$$

Portanto, para gerar amostras da distribuição *a posteriori* de \mathbf{Y}_{T+h} dado $\theta_{1,T+h}$, além de amostrar os demais parâmetros do modelo como detalhado na subseção 2.1.1, é necessário adicionar um passo ao algoritmo MCMC e amostrar $\theta_{2,T+h}$ de $p(\theta_{2,T+h} | \theta_{2,T}, \mathbf{W})$, cuja distribuição está explicitada em (9). Em seguida, $\theta_{1,T+h}$ é amostrado da distribuição $p(\theta_{1,T+h} | \theta_{2,T+h}, \tau^2, \phi)$ e \mathbf{Y}_{T+h} de $p(\mathbf{Y}_{T+h} | \theta_{1,T+h})$.

Interpolação Espacial

Suponha que os dados são observados em um conjunto de locais $S^o = (s_1, s_2, \dots, s_N) \subset \mathbb{R}^2$ e o interesse seja interpolar a variável resposta para outras r localidades $S^n = (s_{N+1}, s_{N+2}, \dots, s_{N+r})$ em

um dado período de tempo t . Seja \mathbf{Y}_t^o o vetor N -dimensional de valores que foram observados no tempo t e $\boldsymbol{\mu}_t^o$ o vetor de mesma dimensão representando suas respectivas médias. Considere também \mathbf{Y}_t^n , o vetor r -dimensional de valores a serem interpolados no tempo t e $\boldsymbol{\mu}_t^n$ suas respectivas médias. Analogamente, defina $\mathbf{F}_{1,t}^o$ como uma matriz de dimensão $N \times N.p$ incorporando possíveis covariáveis relacionadas à variável de interesse \mathbf{Y}_t^o e $\boldsymbol{\theta}_{1,t}^o$ os coeficientes associados a estas covariáveis. Utilizando o mesmo raciocínio, $\mathbf{F}_{1,t}^n$ é uma matriz $r \times r.p$ também incorporando possíveis covariáveis relacionadas à variável de interesse em locais não observados \mathbf{Y}_t^n e $\boldsymbol{\theta}_{1,t}^n$ são seus coeficientes. Defina também $\mathbf{v}_{2,t}^o = (\mathbf{v}_{2,t}^o(s_1), \mathbf{v}_{2,t}^o(s_2), \dots, \mathbf{v}_{2,t}^o(s_N))'$ e $\mathbf{v}_{2,t}^n = (\mathbf{v}_{2,t}^n(s_{N+1}), \mathbf{v}_{2,t}^n(s_{N+2}), \dots, \mathbf{v}_{2,t}^n(s_{N+r}))'$.

Dessa forma, o modelo (6) pode ser escrito da seguinte maneira

$$\begin{aligned} \mathbf{Y}_t^o(s_i) &\sim Poi(\boldsymbol{\mu}_t^o(s_i)) \quad \text{para } i = 1, 2, \dots, N \quad \text{e } t = 1, 2, \dots, T, \\ \mathbf{Y}_t^n(s_{N+k}) &\sim Poi(\boldsymbol{\mu}_t^n(s_{N+k})) \quad \text{para } k = 1, 2, \dots, r. \end{aligned}$$

As demais componentes observadas e não observadas do modelo podem ser descritas em notação matricial, de forma que

$$\begin{aligned} \begin{pmatrix} \log(\boldsymbol{\mu}_t^o) \\ \log(\boldsymbol{\mu}_t^n) \end{pmatrix} &= \begin{pmatrix} \mathbf{F}_{1,t}^o & \mathbf{0}_{N \times r.p} \\ \mathbf{0}_{r \times N.p} & \mathbf{F}_{1,t}^n \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_{1,t}^o \\ \boldsymbol{\theta}_{1,t}^n \end{pmatrix} \\ \begin{pmatrix} \boldsymbol{\theta}_{1,t}^o \\ \boldsymbol{\theta}_{1,t}^n \end{pmatrix} &= \begin{pmatrix} \mathbf{F}_{2,t}^o \\ \mathbf{F}_{2,t}^n \end{pmatrix} \boldsymbol{\theta}_{2,t} + \begin{pmatrix} \mathbf{v}_{2,t}^o \\ \mathbf{v}_{2,t}^n \end{pmatrix} \quad \text{com } \begin{pmatrix} \mathbf{v}_{2,t}^o \\ \mathbf{v}_{2,t}^n \end{pmatrix} \sim PG \left[\mathbf{0}, \begin{pmatrix} \mathbf{V}_2^o & \mathbf{V}_2^{on} \\ \mathbf{V}_2^{no} & \mathbf{V}_2^n \end{pmatrix} \otimes \mathbf{I}_p \right], \\ \boldsymbol{\theta}_{2,t} &= \mathbf{G}_t \boldsymbol{\theta}_{2,t-1} + \boldsymbol{\omega}_t \quad \text{com } \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}), \end{aligned}$$

em que \mathbf{V}_2^o , \mathbf{V}_2^{on} , \mathbf{V}_2^{no} e \mathbf{V}_2^n são partições de \mathbf{V}_2 , obtidas a partir da função que define a estrutura espacial de $\boldsymbol{\theta}_{1,t}$.

Seja $\boldsymbol{\Psi}^*$ o vetor formado pelos hiperparâmetros do modelo, ou seja, $\boldsymbol{\Psi}^* = (\tau^2, \phi, \boldsymbol{\theta}_{2,0}, \mathbf{W})$. Sob a hipótese de $\mathbf{F}_{1,t}^o$ e $\mathbf{F}_{1,t}^n$ serem conhecidas, a distribuição conjunta *a posteriori* é dada por

$$p(\mathbf{Y}_t^n, \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* | \mathbf{Y}_t^o) = p(\mathbf{Y}_t^n | \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^*) p(\boldsymbol{\theta}_{1,t}^n | \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^*) p(\boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* | \mathbf{Y}_t^o).$$

Para amostrar desta distribuição precisa-se amostrar de cada uma das distribuições da equação acima separadamente. Inicialmente, uma amostra de $p(\boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* | \mathbf{Y}_t^o)$ é obtida através do procedimento detalhado na subseção 2.1.1. Em seguida, uma amostra da distribuição $p(\boldsymbol{\theta}_{1,t}^n | \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^*)$ é obtida utilizando propriedades da distribuição Normal, de tal forma que $\boldsymbol{\theta}_{1,t}^n | \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* \sim N(\mathbf{M}_1, \mathbf{H}_1)$, em que $\mathbf{M}_1 = \mathbf{F}_{2,t}^n \boldsymbol{\theta}_{2,t} + (\mathbf{V}_2^{no} \otimes \mathbf{I}_p)(\mathbf{V}_2^o \otimes \mathbf{I}_p)^{-1}(\boldsymbol{\theta}_{1,t}^o - \mathbf{F}_{2,t}^o \boldsymbol{\theta}_{2,t})$ e $\mathbf{H}_1 = [\mathbf{V}_2^n - \mathbf{V}_2^{no}(\mathbf{V}_2^o)^{-1}\mathbf{V}_2^{on}] \otimes \mathbf{I}_p$. Por fim, uma amostra de $p(\mathbf{Y}_t^n | \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^*)$ é obtida utilizando o fato de que $p(\mathbf{Y}_t^n | \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^*) = \prod_{k=1}^r p(Y_t^n(s_{N+k}) | \boldsymbol{\theta}_{1,t}^n(s_{N+k}), \boldsymbol{\Psi}^*)$ e $Y_t^n(s_{N+k}) | \boldsymbol{\theta}_{1,t}^n(s_{N+k}), \boldsymbol{\Psi}^* \sim Poi(\boldsymbol{\mu}_t^n(s_{N+k}))$.

A validação do algoritmo de estimação foi realizada através de um conjunto de dados artificiais gerado a partir do modelo proposto em (6). A geração dos dados artificiais foi feita com o auxílio do *software* R versão 2.13.0 (R Development Core Team, 2009) e as simulações estocásticas foram realizadas com o *software* Ox versão 5.1 (Doornik (2008)). Os detalhes da análise de dados artificiais não serão apresentados aqui.

2.1.2 Modelo com Efeito de Sobredispersão

A distribuição Poisson é frequentemente utilizada para a modelagem de dados de contagem e, com isso, a média e a variância do processo sob observação são assumidas iguais, o que raramente ocorre. Na prática, o que geralmente se observa é que a variância do processo é maior que a média, fenômeno este denominado de sobredispersão. De forma particular, a estrutura dinâmica imposta à média da distribuição de Poisson no modelo dinâmico é capaz de capturar algum efeito de sobredispersão, porém pode não ser suficiente para explicar toda a variação do processo de interesse. Assim, para capturar a variação extra pode-se considerar modelos de mistura, ou equivalentemente, incluir efeitos aleatórios na média ou em alguma função da média da distribuição. Esta última alternativa será usada aqui para contornar possíveis efeitos de sobredispersão, de forma que o segundo modelo apresentado na seção 2 será reescrito como

$$\mathbf{Y}_t | \boldsymbol{\mu}_t^* \sim Poi(\boldsymbol{\mu}_t^*), \quad t = 1, 2, \dots, T \quad (10a)$$

$$\log(\boldsymbol{\mu}_t^*) = \mathbf{F}_{1,t} \boldsymbol{\theta}_{1,t} + \mathbf{v}_{1,t}, \quad \mathbf{v}_{1,t} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N) \quad (10b)$$

$$\boldsymbol{\theta}_{1,t} = \mathbf{F}_{2,t} \boldsymbol{\theta}_{2,t} + \mathbf{v}_{2,t}, \quad \mathbf{v}_{2,t} \sim PG(\mathbf{0}, \tau^2 \exp\{-\phi \cdot \mathbf{d}_{i,j}\} \otimes \mathbf{I}_p) \quad (10c)$$

$$\boldsymbol{\theta}_{2,t} = \mathbf{G}_t \boldsymbol{\theta}_{2,t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}). \quad (10d)$$

A diferença na estrutura dinâmica de μ_t é a inclusão de um erro aleatório com distribuição Normal na equação em (6b). Da equação em (10b), tem-se que

$$\mu_t^* = \exp(\mathbf{F}_{1,t}\boldsymbol{\theta}_{1,t} + \mathbf{v}_{1,t}) \quad (11a)$$

$$= \exp(\mathbf{F}_{1,t}\boldsymbol{\theta}_{1,t}) \exp(\mathbf{v}_{1,t}) \quad (11b)$$

$$= \mu_t \delta_t, \quad (11c)$$

em que $\mu_t = \exp(\mathbf{F}_{1,t}\boldsymbol{\theta}_{1,t})$ segue a mesma estrutura dinâmica vista em (6) e $\delta_t = \exp(\mathbf{v}_{1,t})$, para $t = 1, 2, \dots, T$, segue uma distribuição Lognormal com vetor de médias igual a $\mathbf{0}$ e matriz de covariância $\mathbf{V}_1 = \sigma^2 \mathbf{I}_N$, isto é, $\delta_t | \mathbf{V}_1 \sim LN(\mathbf{0}, \mathbf{V}_1)$.

O vetor paramétrico δ_t , neste modelo, tem como finalidade capturar algum efeito de sobredispersão que possa existir nos dados e que não é explicado pela estrutura dinâmica em 6b, 6c e 6d. A distribuição marginal de \mathbf{Y}_t com respeito à δ_t , para $t = 1, 2, \dots, T$, não possui forma conhecida. No entanto, utilizando propriedades da esperança e da variância condicional, pode-se mostrar que a série temporal para cada localidade i é tal que $E(Y_t | \mu_t, V_1) = \mu_t \exp\left(\frac{V_1}{2}\right)$ e $Var(Y_t | \mu_t, V_1) = \mu_t \exp\left(\frac{V_1}{2}\right) + \mu_t^2 \exp(V_1)[\exp(V_1) - 1]$. Portanto, o modelo Poisson-lognormal é capaz de capturar a sobredispersão através de um efeito aditivo positivo na média do processo.

Procedimento de Inferência

Assumindo que o processo \mathbf{Y}_t possa ser modelado como em (10), o objetivo então é estimar as quantidades desconhecidas envolvidas. Sob este modelo e considerando o caso em que $p > 1$, o vetor paramétrico a ser estimado é dado por

$$\boldsymbol{\Psi} = (\boldsymbol{\mu}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{2,0}, \sigma^2, \phi, \tau^2, \mathbf{W}). \quad (12)$$

As distribuições *a priori* para os vetores paramétricos $\boldsymbol{\mu}$, $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ seguem diretamente da definição das estruturas dinâmicas em 10b, 10c e 10d, respectivamente. Para os hiperparâmetros, $\boldsymbol{\theta}_{2,0}$, ϕ , τ^2 e \mathbf{W} , as mesmas distribuições *a priori* atribuídas aos parâmetros do modelo da subseção 2.1.1 também foram atribuídas neste caso. Para σ^2 atribuiu-se uma distribuição *a priori* dada por $\sigma^2 \sim GI(\alpha_\sigma, \beta_\sigma)$.

Dada toda a informação D_T , a distribuição *a posteriori* do vetor paramétrico $\boldsymbol{\Psi}$ é dada por

$$\begin{aligned} p(\boldsymbol{\Psi} | D_T) &\propto p(\boldsymbol{\theta}_{2,0})p(\mathbf{W})p(\tau^2)p(\phi)p(\sigma^2) \prod_{t=1}^T [p(\boldsymbol{\theta}_{2,t} | \boldsymbol{\theta}_{2,t-1}, \mathbf{W})] \\ &\times \left[\prod_{t=1}^T \prod_{i=1}^N p(\boldsymbol{\theta}_{1,t}(s_i) | \boldsymbol{\theta}_{2,t}, \phi, \tau^2) p(\mu_t(s_i) | \boldsymbol{\theta}_{1,t}(s_i), \sigma^2) p(y_t(s_i) | \mu_t(s_i)) \right]. \end{aligned}$$

Novamente, a distribuição $p(\boldsymbol{\Psi} | D_T)$ não tem solução analítica. Os parâmetros $\boldsymbol{\theta}_1$, \mathbf{W} e σ^2 são amostrados através de suas condicionais completas *a posteriori*, cujas formas são conhecidas. Os parâmetros $\boldsymbol{\theta}_{2,0}$ e $\boldsymbol{\theta}_2$ são amostrados através do algoritmo FFBS. Por fim, $\boldsymbol{\mu}$, τ^2 e ϕ são amostrados através de passos de Metropolis-Hastings, após a especificação de distribuições propostas adequadas. As distribuições propostas para a amostragem dos parâmetros μ_t , τ^2 e ϕ são Lognormais centradas nos valores correntes dos respectivos parâmetros e as variâncias das distribuições propostas são fixas e foram ajustadas de forma que as taxas de aceitação estivessem entre 20% e 50%.

Previsão Temporal

No caso do modelo (10), a distribuição conjunta *a posteriori* dadas as observações passadas pode ser escrita como

$$\begin{aligned} p(\mathbf{Y}_{T+h}, \boldsymbol{\mu}_{T+h}, \boldsymbol{\theta}_{1,T+h}, \boldsymbol{\theta}_{2,T+h}, \boldsymbol{\Psi} | \mathbf{Y}) &= p(\mathbf{Y}_{T+h} | \boldsymbol{\mu}_{T+h}) p(\boldsymbol{\mu}_{T+h} | \boldsymbol{\theta}_{1,T+h}, \sigma^2) \\ &\times p(\boldsymbol{\theta}_{1,T+h} | \boldsymbol{\theta}_{2,T+h}, \tau^2, \phi) p(\boldsymbol{\theta}_{2,T+h} | \boldsymbol{\theta}_{2,T}, \mathbf{W}) p(\boldsymbol{\Psi} | \mathbf{Y}). \end{aligned}$$

Logo, tem-se que $p(\boldsymbol{\Psi} | \mathbf{Y})$ pode ser amostrada através de um algoritmo que utiliza o amostrador de Gibbs, como descrito na subseção 2.1.2. As distribuições $p(\mathbf{Y}_{T+h} | \boldsymbol{\mu}_{T+h})$, $p(\boldsymbol{\mu}_{T+h} | \boldsymbol{\theta}_{1,T+h}, \sigma^2)$ e $p(\boldsymbol{\theta}_{1,T+h} | \boldsymbol{\theta}_{2,T+h}, \tau^2, \phi)$ são conhecidas. Assim, para gerar amostras da distribuição *a posteriori* de \mathbf{Y}_{T+h} dado $\boldsymbol{\mu}_{T+h}$, além de amostrar os demais parâmetros do modelo como detalhado na subseção

2.1.2, é necessário adicionar um passo ao algoritmo MCMC e amostrar $\theta_{2,T+h}$ de $p(\theta_{2,T+h} | \theta_{2,T}, \mathbf{W})$, cuja distribuição é a mesma explicitada em (9). Em seguida, $\theta_{1,T+h}$ é amostrado da distribuição $p(\theta_{1,T+h} | \theta_{2,T+h}, \tau^2, \phi)$ e μ_{T+h} é amostrado de $p(\mu_{T+h} | \theta_{1,T+h}, \sigma^2)$. Por fim, \mathbf{Y}_{T+h} é gerado de $p(\mathbf{Y}_{T+h} | \mu_{T+h})$.

Interpolação Espacial

Suponha novamente que os dados são observados em um conjunto de locais $S^o = (s_1, \dots, s_N) \subset \mathbb{R}^2$ e o interesse seja interpolar a variável resposta para outras r localidades $S^n = (s_{N+1}, \dots, s_{N+r})$ em um dado período de tempo t . Considere também $\mathbf{v}_{1,t}^o = (\mathbf{v}_{1,t}^o(s_1), \dots, \mathbf{v}_{1,t}^o(s_N))'$ e $\mathbf{v}_{1,t}^n = (\mathbf{v}_{1,t}^n(s_{N+1}), \dots, \mathbf{v}_{1,t}^n(s_{N+r}))'$, vetores de erros aleatórios observados e não observados, respectivamente. Assim como no modelo (6), o modelo (10) pode ser escrito da seguinte maneira

$$\begin{aligned} Y_t^o(s_i) &\sim Poi(\mu_t^o(s_i)) \quad \text{para } i = 1, 2, \dots, N \quad \text{e } t = 1, 2, \dots, T, \\ Y_t^n(s_{N+k}) &\sim Poi(\mu_t^n(s_{N+k})) \quad \text{para } k = 1, 2, \dots, r. \end{aligned}$$

As demais componentes observadas e não observadas do modelo podem ser descritas compactamente em notação matricial, tal que

$$\begin{aligned} \begin{pmatrix} \log(\mu_t^o) \\ \log(\mu_t^n) \end{pmatrix} &= \begin{pmatrix} \mathbf{F}_{1,t}^o & \mathbf{0}_{N \times r \cdot p} \\ \mathbf{0}_{r \times N \cdot p} & \mathbf{F}_{1,t}^n \end{pmatrix} \begin{pmatrix} \theta_{1,t}^o \\ \theta_{1,t}^n \end{pmatrix} + \begin{pmatrix} \mathbf{v}_{1,t}^o \\ \mathbf{v}_{1,t}^n \end{pmatrix}, \quad \begin{pmatrix} \mathbf{v}_{1,t}^o \\ \mathbf{v}_{1,t}^n \end{pmatrix} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{N+r}), \\ \begin{pmatrix} \theta_{1,t}^o \\ \theta_{1,t}^n \end{pmatrix} &= \begin{pmatrix} \mathbf{F}_{2,t}^o \\ \mathbf{F}_{2,t}^n \end{pmatrix} \theta_{2,t} + \begin{pmatrix} \mathbf{v}_{2,t}^o \\ \mathbf{v}_{2,t}^n \end{pmatrix}, \quad \begin{pmatrix} \mathbf{v}_{2,t}^o \\ \mathbf{v}_{2,t}^n \end{pmatrix} \sim PG \left[\mathbf{0}, \begin{pmatrix} \mathbf{V}_2^o & \mathbf{V}_2^{on} \\ \mathbf{V}_2^{no} & \mathbf{V}_2^n \end{pmatrix} \otimes \mathbf{I}_p \right]. \\ \theta_{2,t} &= \mathbf{G}_t \theta_{2,t-1} + \omega_t, \quad \omega_t \sim N(\mathbf{0}, \mathbf{W}). \end{aligned}$$

Seja Ψ^* o vetor formado pelos hiperparâmetros do modelo, ou seja, $\Psi^* = (\sigma^2, \tau^2, \phi, \theta_{2,0}, \mathbf{W})$. Sob a hipótese de $\mathbf{F}_{1,t}^o$ e $\mathbf{F}_{1,t}^n$ serem conhecidas, a distribuição conjunta *a posteriori* é dada por

$$\begin{aligned} p(\mathbf{Y}_t^n, \mu_t^n, \mu_t^o, \theta_{1,t}^n, \theta_{1,t}^o, \theta_{2,t}, \Psi^* | \mathbf{Y}_t^o) &= p(\mathbf{Y}_t^n | \mu_t^n, \theta_{1,t}^n, \Psi^*) p(\log(\mu_t^n) | \log(\mu_t^o), \theta_{1,t}^n, \Psi^*) \\ &\times p(\theta_{1,t}^o | \theta_{1,t}^n, \theta_{2,t}, \Psi^*) p(\mu_t^o, \theta_{1,t}^o, \theta_{2,t}, \Psi^* | \mathbf{Y}_t^o). \end{aligned}$$

Para amostrar desta distribuição precisa-se amostrar de cada distribuição da equação acima separadamente. Inicialmente, uma amostra da distribuição $p(\mu_t^o, \theta_{1,t}^o, \theta_{2,t}, \Psi^* | \mathbf{Y}_t^o)$ é obtida através do procedimento detalhado na subseção 2.1.2. Em seguida, uma amostra de $p(\theta_{1,t}^n | \theta_{1,t}^o, \theta_{2,t}, \Psi^*)$ é obtida utilizando propriedades da distribuição Normal, de tal forma que $\theta_{1,t}^n | \theta_{1,t}^o, \theta_{2,t}, \Psi^* \sim N(\mathbf{M}_1, \mathbf{H}_1)$, em que \mathbf{M}_1 e \mathbf{H}_1 são como especificados na subseção 2.1.1. Analogamente, uma amostra de $p(\log(\mu_t^n) | \log(\mu_t^o), \theta_{1,t}^n, \Psi^*)$ pode ser obtida utilizando o seguinte resultado: $\log(\mu_t^n) | \log(\mu_t^o), \theta_{1,t}^n, \Psi^* \sim N(\mathbf{M}_2, \mathbf{H}_2)$, com $\mathbf{M}_2 = \mathbf{F}_{1,t}^n \theta_{1,t}^n$ e $\mathbf{H}_2 = \sigma^2 \mathbf{I}_r$. Por fim, uma amostra de $p(\mathbf{Y}_t^n | \mu_t^n, \theta_{1,t}^n, \Psi^*)$ é obtida utilizando o fato de que $p(\mathbf{Y}_t^n | \mu_t^n, \theta_{1,t}^n, \Psi^*) = \prod_{k=1}^r p(Y_t^n(s_{N+k}) | \mu_t^n(s_{N+k}), \theta_{1,t}^n(s_{N+k}), \Psi^*)$ e $Y_t^n(s_{N+k}) | \mu_t^n(s_{N+k}), \theta_{1,t}^n(s_{N+k}), \Psi^* \sim Poi(\mu_t^n(s_{N+k}))$.

Com o intuito de validar o algoritmo de estimação, um conjunto de dados artificiais foi gerado a partir do modelo proposto em (10). O procedimento de geração dos dados artificiais e as simulações estocásticas foram realizados da mesma forma descrita no modelo sem efeito de sobredispersão da subseção 2.1.1. De acordo com o modelo 10, deve-se especificar ainda uma distribuição *a priori* para o parâmetro σ^2 . Foi atribuída uma distribuição Gama invertida, *a priori*, com $\alpha_\sigma = 2,001$ e $\beta_\sigma = 1,001$.

2.2 Distribuição Gama Inflacionada de Zeros

Nesta seção, serão considerados modelos para variáveis não negativas distribuídas continuamente no tempo e espaço, contemplando também a ocorrência de valores nulos. Exemplos comuns incluem a modelagem de poluentes do ar em grandes centros urbanos, a velocidade do vento, etc. Dados desta natureza podem conter muitas observações assumindo o valor zero, fazendo com que haja uma grande sobredispersão. Assim, é necessária a utilização de modelos que considerem a inflação de zeros, os quais são conhecidos na literatura como modelos espaço-temporais inflacionados de zeros. A idéia é incluir uma massa de probabilidade no ponto zero, inflacionando suas possibilidades de existir no modelo. Vale ressaltar que esse tipo de conceituação pode ser estendida, em especial, para qualquer outro valor que se deseje modelar. Entretanto, de maneira geral, o interesse está no valor zero.

2.2.1 Modelo sem Efeito de Sobredispersão

Considere que a variável aleatória de interesse $Y_t(s_i)$, para $i = 1, 2, \dots, N$ e $t = 1, 2, \dots, T$, represente quantidades contínuas e não negativas de um determinado fenômeno, em que ocasionalmente podem ser medidos valores nulos. Defina agora uma variável aleatória latente $X_t(s_i)$, seguindo uma distribuição Bernoulli representando a “presença” ($X_t(s_i) = 1$) do processo que está sendo observado, o que ocorre com probabilidade ξ_j , ou “ausência” ($X_t(s_i) = 0$) deste processo, o que ocorre com probabilidade $(1 - \xi_j)$, para $j = 1, 2, \dots, J$, em que a dimensão do vetor $\xi = (\xi_1, \xi_2, \dots, \xi_J)'$ é definida a partir de características observadas do fenômeno de interesse em cada tempo t e localidade s_i . Assim,

$$X_t(s_i) | \xi_j \sim Ber(\xi_j). \quad (14)$$

Como exemplo de definição da dimensão do vetor ξ , considere o caso em que se deseja modelar a quantidade de chuva em localidades e períodos de tempo específicos. Suponha que a região de estudo possa ser caracterizada por dois períodos bem definidos, seco e chuvoso, identificando um fenômeno sazonal. Além disto, sabe-se que a quantidade de chuva é influenciada por diversos fatores, tais como temperatura, umidade e pressão. A influência destes fatores pode agir de forma diferenciada nas localidades de interesse, de forma que algumas localidades, em média, são mais secas do que outras independentemente do período de tempo considerado. Neste caso, tem-se $j = 4$ cenários distintos que podem ser representados a partir da combinação de períodos e localidades secos ou chuvosos. São eles:

Cenário 1: Estação seca e período seco.

Cenário 2: Estação seca e período chuvoso.

Cenário 3: Estação chuvosa e período seco.

Cenário 4: Estação chuvosa e período chuvoso.

Quando o processo está presente, será assumido que $Y_t(s_i)$ tem distribuição Gama com parâmetros α e $\beta_t(s_i)$ ou, especificamente, se sua função densidade de probabilidade é dada por

$$p(Y_t(s_i) | \alpha, \beta_t(s_i)) = \frac{\beta_t(s_i)^\alpha}{\Gamma(\alpha)} y_t(s_i)^{\alpha-1} \exp\{-\beta_t(s_i)y_t(s_i)\}, \quad \text{com} \quad (15)$$

$$E(Y_t(s_i)) = \mu_t(s_i) = \frac{\alpha}{\beta_t(s_i)} \quad \text{e} \quad \text{Var}(Y_t(s_i)) = \frac{\alpha}{\beta_t(s_i)^2}. \quad (16)$$

Seja ν o coeficiente de variação desta distribuição, que informa a dispersão dos dados em torno da média. Por definição,

$$\nu = \frac{\sqrt{\text{Var}(Y_t(s_i))}}{E(Y_t(s_i))} = \frac{1}{\sqrt{\alpha}}. \quad (17)$$

Portanto, a densidade em (15) poderá ser reescrita em termos de sua média $\mu_t(s_i)$ e do coeficiente de variação ν . Ainda, de acordo com (16) e (17), os novos parâmetros α^* e $\beta_t^*(s_i)$ da distribuição Gama serão dados por

$$\alpha^* = \frac{1}{\nu^2} \quad \text{e} \quad \beta_t^*(s_i) = \frac{1}{\nu^2 \mu_t(s_i)}.$$

A dependência espaço-temporal novamente aparece no modelo sob a forma de efeitos aleatórios espacial e temporalmente estruturados, de forma que o modelo mais simples definido na seção 2 será reescrito como

$$\mathbf{Y}_t | \boldsymbol{\mu}_t, \nu \sim Ga(\alpha^*, \boldsymbol{\beta}_t^*), \quad t = 1, 2, \dots, T \quad (18a)$$

$$\log(\boldsymbol{\mu}_t) = \mathbf{F}_{1,t} \boldsymbol{\theta}_{1,t} \quad (18b)$$

$$\boldsymbol{\theta}_{1,t} = \mathbf{F}_{2,t} \boldsymbol{\theta}_{2,t} + \mathbf{v}_{2,t}, \quad \mathbf{v}_{2,t} \sim PG(\mathbf{0}, \tau^2 \exp\{-\phi \cdot \mathbf{d}_{i,j}\} \otimes \mathbf{I}_p) \quad (18c)$$

$$\boldsymbol{\theta}_{2,t} = \mathbf{G}_t \boldsymbol{\theta}_{2,t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}) \quad (18d)$$

Por definição, tem-se com probabilidade ξ_j , um valor que virá da distribuição $p(Y_t(s_i) | X_t(s_i) = 1, \mu_t(s_i), \nu)$ que, neste caso, é a distribuição Gama. Os eventos $Y_t(s_i) = 0$ serão provenientes apenas

da distribuição Bernoulli e a variável latente $X_t(s_i)$ terá uma relação determinística com a variável resposta $Y_t(s_i)$, que pode ser verificada a partir das seguintes probabilidades condicionais:

$$\begin{aligned} P(X_t(s_i) = 1 | Y_t(s_i) = 0) &= 0 ; \\ P(X_t(s_i) = 0 | Y_t(s_i) = 0) &= 1 ; \\ P(Y_t(s_i) = 0 | X_t(s_i) = 1, \mu_t(s_i), \nu) &= 0 ; \\ P(Y_t(s_i) = 0 | X_t(s_i) = 0, \mu_t(s_i), \nu) &= 1 . \end{aligned}$$

A partir das probabilidades descritas acima e da expressão em (14), pode-se escrever a função de probabilidade conjunta de $X_t(s_i)$ e $Y_t(s_i)$, para $t = 1, 2, \dots, T$, $i = 1, 2, \dots, N$ e $j = 1, 2, \dots, J$, como

$$\begin{aligned} p(Y_t(s_i), X_t(s_i) | \xi_j, \mu_t(s_i), \nu) &= p(Y_t(s_i) | X_t(s_i), \xi_j, \mu_t(s_i), \nu) p(X_t(s_i) | \xi_j, \mu_t(s_i), \nu) \\ &= p(Y_t(s_i) | X_t(s_i), \mu_t(s_i), \nu) p(X_t(s_i) | \xi_j) \\ &= \begin{cases} p(Y_t(s_i) | \mu_t(s_i), \nu) \xi_j, & (X_t(s_i) = 1, Y_t(s_i) > 0) \\ 1 - \xi_j, & (X_t(s_i) = 0, Y_t(s_i) = 0) \\ 0, & \text{caso contrário} \end{cases} \end{aligned}$$

Equivalentemente, a função de probabilidade conjunta $p(Y_t(s_i), X_t(s_i) | \xi_j, \mu_t(s_i), \nu)$ é dada por

$$\{p(Y_t(s_i) | \mu_t(s_i), \nu) \xi_j\}^{X_t(s_i)} (1 - \xi_j)^{1 - X_t(s_i)}, \quad \begin{cases} (X_t(s_i) = 1, Y_t(s_i) > 0) \\ (X_t(s_i) = 0, Y_t(s_i) = 0) \end{cases} \quad (19)$$

e assume o valor zero, caso contrário. Adicionalmente, a função de probabilidade marginal de $Y_t(s_i)$ com relação a $X_t(s_i)$, para $t = 1, 2, \dots, T$, $i = 1, 2, \dots, N$ e $j = 1, 2, \dots, J$, é da forma

$$p(Y_t(s_i) | \xi_j, \mu_t(s_i), \nu) = p(Y_t(s_i) | \mu_t(s_i), \nu) \xi_j + (1 - \xi_j) I_{(Y_t(s_i)=0)}, \quad (20)$$

em que $I_{(Y_t(s_i)=0)}$ é uma função indicadora da ocorrência de um valor zero na observação $Y_t(s_i)$. Note que, desta forma, a probabilidade de obter um valor 0 (zero) está sendo inflacionada. Esta modelagem se comporta de maneira especial, pois todas as ocorrências do evento $Y_t(s_i) = 0$ estarão associadas à distribuição de Bernoulli.

O modelo de mistura para dados inflacionados de zeros não está restrito à distribuição Gama, isto é, outras distribuições contínuas ou discretas podem ser utilizadas em uma mistura com a distribuição de Bernoulli.

Procedimento de Inferência

Assumindo que o processo \mathbf{Y}_t possa ser modelado como uma mistura entre a distribuição Gama descrita em (18) e a distribuição de Bernoulli em (14), o objetivo agora é estimar as quantidades desconhecidas envolvidas.

Defina $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_T)'$, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T)'$, $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_J)'$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_T)'$, $\boldsymbol{\theta}_1 = (\boldsymbol{\theta}_{1,1}, \boldsymbol{\theta}_{1,2}, \dots, \boldsymbol{\theta}_{1,T})'$, $\boldsymbol{\theta}_2 = (\boldsymbol{\theta}_{2,1}, \boldsymbol{\theta}_{2,2}, \dots, \boldsymbol{\theta}_{2,T})'$ e Ψ como a coleção de todos os parâmetros. Note que, se $Y_t(s_i) > 0$, então $X_t(s_i) = 1$ e se $Y_t(s_i) = 0$, então $X_t(s_i) = 0$. Dessa forma, não é necessário estimar \mathbf{X} . Considerando o caso em que $p > 1$, o vetor paramétrico a ser estimado é dado por

$$\Psi = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{2,0}, \phi, \tau^2, \mathbf{W}, \nu, \boldsymbol{\xi}). \quad (21)$$

Como a inferência será feita sob o enfoque bayesiano precisa-se atribuir uma distribuição *a priori* para o vetor paramétrico Ψ . Será assumido que os hiperparâmetros são independentes *a priori*, e portanto, a distribuição conjunta será dada pelo produto das distribuições marginais. As distribuições *a priori* para os vetores paramétricos $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ seguem diretamente da definição das estruturas dinâmicas em 18c e 18d, respectivamente. A informação *a priori*, D_0 , a respeito do vetor $\boldsymbol{\theta}_{2,0}$ é quantificada no instante inicial $t = 0$, através do primeiro e segundo momentos da distribuição Normal, ou seja, $\boldsymbol{\theta}_{2,0} | D_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$. Para o parâmetro ϕ foi atribuída a seguinte distribuição *a priori*: $\phi \sim Ga(\alpha_\phi, \beta_\phi)$. Para o parâmetro τ^2 , da matriz de correlação espacial foi atribuída uma *a priori* tal que $\tau^2 \sim GI(\alpha_\tau, \beta_\tau)$. Para a matriz de covariância \mathbf{W} foi assumido *a priori* que $\mathbf{W} \sim WI(m, \mathbf{S})$. Para o coeficiente de variação ν , foi atribuída uma distribuição Gama, com parâmetros α_ν e β_ν , ou seja, $\nu \sim Ga(\alpha_\nu, \beta_\nu)$. Por fim, para o vetor de probabilidades $\boldsymbol{\xi}$, assumiu-se *a priori* que $\boldsymbol{\xi} \sim Dir(\boldsymbol{\alpha})$, em que *Dir* denota a distribuição Dirichlet com parâmetros $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_J)$.

Dada toda a informação $D_T = (D_0, y_1, y_2, \dots, y_T, x_1, x_2, \dots, x_T)$, a distribuição *a posteriori* do vetor paramétrico Ψ é dada por

$$\begin{aligned} p(\Psi | D_T) &\propto p(\theta_{2,0})p(\mathbf{W})p(\tau^2)p(\phi)p(\nu)p(\xi) \prod_{t=1}^T [p(\theta_{2,t} | \theta_{2,t-1}, \mathbf{W})] \\ &\times \prod_{t=1}^T \prod_{i=1}^N p(\theta_{1,t}(s_i) | \theta_{2,t}, \phi, \tau^2) \prod_{t=1}^T \prod_{i=1}^N \prod_{j=1}^J p(y_t(s_i), x_t(s_i) | \xi_j, \theta_{1,t}(s_i), \nu). \end{aligned}$$

Novamente, a distribuição $p(\Psi | D_T)$ não tem solução analítica e os métodos MCMC serão utilizados para obter amostras desta distribuição. A matriz \mathbf{W} e as componentes do vetor ξ são amostradas através de suas condicionais completas *a posteriori*, cujas formas são conhecidas. Os parâmetros $\theta_{2,0}$ e θ_2 são amostrados através do algoritmo FFBS. Por fim, θ_1 , τ^2 , ϕ e ν são amostrados através de passos de Metropolis-Hastings, após a especificação de distribuições propostas adequadas. As distribuições propostas para $\theta_{1,t}$, τ^2 e ϕ são as mesmas atribuídas no modelo sem efeito de sobredispersão da subseção 2.1.1. Para a amostragem do parâmetro ν foi proposta uma distribuição Lognormal centrada no valor corrente do parâmetro. As variâncias das distribuições propostas no algoritmo de Metropolis-Hastings foram estabelecidas segundo o critério especificado nas subseções 2.1.1 e 2.1.2.

Previsão Temporal

Suponha que se deseja trabalhar novamente com um horizonte de previsão h passos a frente, a partir de um instante T qualquer. Dessa forma, o objetivo é obter amostras da distribuição $(\mathbf{Y}_{T+h} | \mathbf{Y})$. A distribuição conjunta *a posteriori* dadas as observações passadas pode ser escrita como

$$\begin{aligned} p(\mathbf{Y}_{T+h}, \mathbf{X}_{T+h}, \theta_{1,T+h}, \theta_{2,T+h}, \Psi | \mathbf{Y}) &= p(\mathbf{Y}_{T+h} | \mathbf{X}_{T+h}, \theta_{1,T+h}, \nu) p(\mathbf{X}_{T+h} | \xi) \\ &\times p(\theta_{1,T+h} | \theta_{2,T+h}, \tau^2, \phi) p(\theta_{2,T+h} | \theta_{2,T}, \mathbf{W}) \\ &\times p(\Psi | \mathbf{Y}). \end{aligned}$$

Da equação acima tem-se que $p(\Psi | \mathbf{Y})$ pode ser amostrada como descrito na subseção 2.2.1. As distribuições $p(\mathbf{Y}_{T+h} | \mathbf{X}_{T+h}, \theta_{1,T+h}, \nu)$, $p(\mathbf{X}_{T+h} | \xi)$ e $p(\theta_{1,T+h} | \theta_{2,T+h}, \tau^2, \phi)$ são conhecidas, e $p(\theta_{2,T+h} | \theta_{2,T}, \mathbf{W})$ pode ser facilmente obtida assim como nos modelos anteriores, de tal forma que

$$\theta_{2,T+h} | \theta_{2,T}, \mathbf{W} \sim N \left(\mathbf{H}_{T+h}(h) \theta_{2,T}, \sum_{k=1}^h \mathbf{H}_{T+h}(h-k) \mathbf{W} \mathbf{H}_{T+h}(h-k)' \right). \quad (23)$$

Portanto, para gerar amostras da distribuição *a posteriori* de \mathbf{Y}_{T+h} dados \mathbf{X}_{T+h} , $\theta_{1,T+h}$ e ν , além de amostrar os demais parâmetros do modelo como detalhado na subseção 2.2.1, é necessário adicionar dois passos ao algoritmo MCMC amostrando \mathbf{X}_{T+h} de $p(\mathbf{X}_{T+h} | \xi)$ e, se $\mathbf{X}_{T+h}(s_i) = 1$, para $i = 1, 2, \dots, N$ e $t = 1, 2, \dots, T$, então amostra-se $\theta_{2,T+h}$ de $p(\theta_{2,T+h} | \theta_{2,T}, \mathbf{W})$, cuja distribuição é a mesma explicitada em (23). Em seguida, $\theta_{1,T+h}$ é amostrado da distribuição $p(\theta_{1,T+h} | \theta_{2,T+h}, \tau^2, \phi)$ e \mathbf{Y}_{T+h} de $p(\mathbf{Y}_{T+h} | \mathbf{X}_{T+h}, \theta_{1,T+h}, \nu)$.

Interpolação Espacial

Suponha agora que os dados são observados em um conjunto de locais $S^o = (s_1, s_2, \dots, s_N) \subset \mathbb{R}^2$ e o interesse é interpolar a variável resposta para outras r localidades $S^n = (s_{N+1}, s_{N+2}, \dots, s_{N+r})$ em um dado período de tempo t .

Seja \mathbf{Y}_t^o o vetor N -dimensional de valores observados no tempo t e $\boldsymbol{\mu}_t^o$ o vetor de mesma dimensão representando suas respectivas médias. Considere também \mathbf{Y}_t^n , o vetor r -dimensional de valores a serem interpolados no tempo t e $\boldsymbol{\mu}_t^n$ suas respectivas médias. Analogamente, defina $\mathbf{F}_{1,t}^o$ como uma matriz de dimensão $N \times N.p$ incorporando possíveis covariáveis relacionadas à variável de interesse \mathbf{Y}_t^o e $\boldsymbol{\theta}_{1,t}^o$ os coeficientes associados a estas covariáveis. Utilizando o mesmo raciocínio, $\mathbf{F}_{1,t}^n$ é uma matriz $r \times r.p$ também incorporando possíveis covariáveis relacionadas à variável de interesse em locais não medidos \mathbf{Y}_t^n e $\boldsymbol{\theta}_{1,t}^n$ são seus coeficientes. Defina também $\mathbf{v}_{2,t}^o = (\mathbf{v}_{2,t}^o(s_1), \mathbf{v}_{2,t}^o(s_2), \dots, \mathbf{v}_{2,t}^o(s_N))'$ e $\mathbf{v}_{2,t}^n = (\mathbf{v}_{2,t}^n(s_{N+1}), \mathbf{v}_{2,t}^n(s_{N+2}), \dots, \mathbf{v}_{2,t}^n(s_{N+r}))'$.

Quando o processo está presente, o modelo (18) pode ser escrito da seguinte maneira

$$Y_t^o(s_i) \sim Ga\left(\frac{1}{\nu^2}, \frac{1}{\nu^2 \mu_t^o(s_i)}\right) \quad \text{para } i = 1, 2, \dots, N \quad \text{e } t = 1, 2, \dots, T,$$

$$Y_t^n(s_{N+k}) \sim Ga\left(\frac{1}{\nu^2}, \frac{1}{\nu^2 \mu_t^n(s_{N+k})}\right) \quad \text{para } k = 1, 2, \dots, r.$$

As demais componentes observadas e não observadas do modelo podem ser descritas em notação matricial, de tal forma que

$$\begin{pmatrix} \log(\boldsymbol{\mu}_t^o) \\ \log(\boldsymbol{\mu}_t^n) \end{pmatrix} = \begin{pmatrix} \mathbf{F}_{1,t}^o & \mathbf{0}_{N \times r \cdot p} \\ \mathbf{0}_{r \times N \cdot p} & \mathbf{F}_{1,t}^n \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_{1,t}^o \\ \boldsymbol{\theta}_{1,t}^n \end{pmatrix}$$

$$\begin{pmatrix} \boldsymbol{\theta}_{1,t}^o \\ \boldsymbol{\theta}_{1,t}^n \end{pmatrix} = \begin{pmatrix} \mathbf{F}_{2,t}^o \\ \mathbf{F}_{2,t}^n \end{pmatrix} \boldsymbol{\theta}_{2,t} + \begin{pmatrix} \mathbf{v}_{2,t}^o \\ \mathbf{v}_{2,t}^n \end{pmatrix} \quad \text{com } \begin{pmatrix} \mathbf{v}_{2,t}^o \\ \mathbf{v}_{2,t}^n \end{pmatrix} \sim PG\left[\mathbf{0}, \begin{pmatrix} \mathbf{V}_2^o & \mathbf{V}_2^{on} \\ \mathbf{V}_2^{no} & \mathbf{V}_2^n \end{pmatrix} \otimes \mathbf{I}_p\right],$$

$$\boldsymbol{\theta}_{2,t} = \mathbf{G}_t \boldsymbol{\theta}_{2,t-1} + \boldsymbol{\omega}_t \quad \text{com } \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}),$$

em que \mathbf{V}_2^o , \mathbf{V}_2^{on} , \mathbf{V}_2^{no} e \mathbf{V}_2^n são partições de \mathbf{V}_2 , obtidas a partir da função que define a estrutura espacial de $\boldsymbol{\theta}_{1,t}$.

Seja $\boldsymbol{\Psi}^*$ o vetor formado pelos hiperparâmetros do modelo, ou seja, $\boldsymbol{\Psi}^* = (\xi, \nu, \tau^2, \phi, \boldsymbol{\theta}_{2,0}, \mathbf{W})$. Sob a hipótese de $\mathbf{F}_{1,t}^o$ e $\mathbf{F}_{1,t}^n$ serem conhecidas, a distribuição conjunta *a posteriori* é dada por

$$\begin{aligned} p(\mathbf{Y}_t^n, \mathbf{X}_t^n, \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* | \mathbf{Y}_t^o) &= p(\mathbf{Y}_t^n | \mathbf{X}_t^n, \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^*) p(\mathbf{X}_t^n | \boldsymbol{\Psi}^*) \\ &\times p(\boldsymbol{\theta}_{1,t}^n | \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^*) p(\boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* | \mathbf{Y}_t^o). \end{aligned}$$

Para amostrar desta distribuição precisa-se amostrar de cada uma das distribuições da equação acima separadamente. Inicialmente, são obtidas amostras de $p(\mathbf{X}_t^n | \boldsymbol{\Psi}^*)$ e $p(\boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* | \mathbf{Y}_t^o)$, através do procedimento detalhado na subseção 2.2.1. Em seguida, uma amostra da distribuição $p(\boldsymbol{\theta}_{1,t}^n | \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^*)$ é obtida utilizando propriedades da distribuição Normal, de tal forma que

$$\boldsymbol{\theta}_{1,t}^n | \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* \sim N(\mathbf{M}_1, \mathbf{H}_1),$$

em que \mathbf{M}_1 e \mathbf{H}_1 são como especificados no modelo sem efeito de sobredispersão da subseção 2.1.1. Por fim, uma amostra de $p(\mathbf{Y}_t^n | \mathbf{X}_t^n, \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^*)$ é obtida utilizando o fato de que $p(\mathbf{Y}_t^n | \mathbf{X}_t^n, \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^*) = \prod_{k=1}^r p(Y_t^n(s_{N+k}) | X_t^n(s_{N+k}), \boldsymbol{\theta}_{1,t}^n(s_{N+k}), \boldsymbol{\Psi}^*)$ e $Y_t^n(s_{N+k}) | X_t^n(s_{N+k}), \boldsymbol{\theta}_{1,t}^n(s_{N+k}), \boldsymbol{\Psi}^* \sim Ga\left(\frac{1}{\nu^2}, \frac{1}{\nu^2 \mu_t^n(s_{N+k})}\right)$.

A validação do algoritmo de estimação foi realizada através de um conjunto de dados artificiais gerado a partir do modelo proposto em (18). A geração dos dados artificiais foi feita com o auxílio do *software* R versão 2.13.0 (R Development Core Team, 2009) e as simulações estocásticas foram realizadas com o *software* Ox versão 5.1 (Doornik (2008)). Os detalhes dessa análise de dados artificiais também não serão apresentados aqui.

2.2.2 Modelo com Efeito de Sobredispersão

Assim como em 2.2.1, considere novamente que a variável aleatória de interesse $Y_t(s_i)$, para $i = 1, 2, \dots, N$ e $t = 1, 2, \dots, T$, represente quantidades contínuas e não negativas de um determinado fenômeno, em que ocasionalmente são medidos valores nulos. Neste modelo, a função que modela a média da distribuição também pertence à família exponencial e contempla, além de covariáveis, um efeito aleatório que traz mais incerteza ao modelo.

Quando o processo está presente, o segundo modelo da seção 2 será reescrito como

$$\mathbf{Y}_t | \boldsymbol{\mu}_t, \nu \sim Ga(\boldsymbol{\alpha}^*, \boldsymbol{\beta}_t^*), \quad t = 1, 2, \dots, T \quad (24a)$$

$$\log(\boldsymbol{\mu}_t) = \mathbf{F}_{1,t} \boldsymbol{\theta}_{1,t} + \mathbf{v}_{1,t}, \quad \mathbf{v}_{1,t} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N) \quad (24b)$$

$$\boldsymbol{\theta}_{1,t} = \mathbf{F}_{2,t} \boldsymbol{\theta}_{2,t} + \mathbf{v}_{2,t}, \quad \mathbf{v}_{2,t} \sim PG(\mathbf{0}, \tau^2 \exp\{-\phi \cdot \mathbf{d}_{i,j}\} \otimes \mathbf{I}_p) \quad (24c)$$

$$\boldsymbol{\theta}_{2,t} = \mathbf{G}_t \boldsymbol{\theta}_{2,t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}). \quad (24d)$$

Procedimento de Inferência

Assumindo que o processo \mathbf{Y}_t possa ser modelado como uma mistura entre a distribuição Gama descrita em (24) e a distribuição de Bernoulli em (14), o próximo passo é estimar as quantidades

desconhecidas envolvidas. Sob o modelo (24) e considerando o caso em que $p > 1$, o vetor paramétrico a ser estimado é dado por

$$\Psi = (\boldsymbol{\mu}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{2,0}, \sigma^2, \phi, \tau^2, \mathbf{W}, \nu, \boldsymbol{\xi}). \quad (25)$$

As distribuições *a priori* para os vetores paramétricos $\boldsymbol{\mu}$, $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ seguem diretamente da definição das estruturas dinâmicas em 24b, 24c e 24d, respectivamente. Para os hiperparâmetros, $\boldsymbol{\theta}_{2,0}$, ϕ , τ^2 , \mathbf{W} , ν e $\boldsymbol{\xi}$, as mesmas distribuições *a priori* atribuídas aos parâmetros do modelo da subseção 2.2.1 também foram atribuídas neste caso. Adicionalmente, para σ^2 atribuiu-se uma distribuição *a priori* tal que $\sigma^2 \sim GI(\alpha_\sigma, \beta_\sigma)$.

Dada toda a informação $D_T = (D_0, y_1, y_2, \dots, y_T, x_1, x_2, \dots, x_T)$, a distribuição *a posteriori* do vetor paramétrico Ψ é dada por

$$\begin{aligned} p(\Psi | D_T) &\propto p(\boldsymbol{\theta}_{2,0})p(\mathbf{W})p(\sigma^2)p(\tau^2)p(\phi)p(\nu)p(\boldsymbol{\xi}) \prod_{t=1}^T [p(\boldsymbol{\theta}_{2,t} | \boldsymbol{\theta}_{2,t-1}, \mathbf{W})] \\ &\times \left[\prod_{t=1}^T \prod_{i=1}^N p(\boldsymbol{\theta}_{1,t}(s_i) | \boldsymbol{\theta}_{2,t}, \phi, \tau^2) p(\boldsymbol{\mu}_t(s_i) | \boldsymbol{\theta}_{1,t}(s_i), \sigma^2) \right] \\ &\times \prod_{t=1}^T \prod_{i=1}^N \prod_{j=1}^J p(y_t(s_i), x_t(s_i) | \boldsymbol{\xi}_j, \boldsymbol{\mu}_t(s_i), \nu). \end{aligned}$$

As componentes dos vetores $\boldsymbol{\theta}_{1,t}$ e $\boldsymbol{\xi}$, a matriz \mathbf{W} e o parâmetro de variância σ^2 são amostrados através de suas condicionais completas *a posteriori*, cujas formas são conhecidas. Os parâmetros $\boldsymbol{\theta}_{2,0}$ e $\boldsymbol{\theta}_2$ são amostrados através do algoritmo FFBS. Finalmente, $\boldsymbol{\mu}$, τ^2 , ϕ e ν são amostrados através de passos de Metropolis-Hastings, após a especificação de distribuições propostas adequadas. As distribuições propostas para $\boldsymbol{\mu}_t$, τ^2 , ϕ e ν são distribuições Lognormais centradas nos valores correntes destes parâmetros. As variâncias das distribuições propostas no algoritmo de Metropolis-Hastings foram novamente estabelecidas segundo o critério especificado nas subseções 2.1.1 e 2.1.2.

Previsão Temporal

Suponha que se deseja trabalhar novamente com um horizonte de previsão h passos a frente, a partir de um instante T qualquer. Dessa forma, o objetivo é obter amostras da distribuição $(\mathbf{Y}_{T+h} | \mathbf{Y})$. A distribuição conjunta *a posteriori* dadas as observações passadas pode ser escrita como

$$\begin{aligned} p(\mathbf{Y}_{T+h}, \mathbf{X}_{T+h}, \boldsymbol{\mu}_{T+h}, \boldsymbol{\theta}_{1,T+h}, \boldsymbol{\theta}_{2,T+h}, \Psi | \mathbf{Y}) &= p(\mathbf{Y}_{T+h} | \mathbf{X}_{T+h}, \boldsymbol{\mu}_{T+h}, \nu) p(\mathbf{X}_{T+h} | \boldsymbol{\xi}) \\ &\times p(\boldsymbol{\mu}_{T+h} | \boldsymbol{\theta}_{1,T+h}, \sigma^2) p(\boldsymbol{\theta}_{1,T+h} | \boldsymbol{\theta}_{2,T+h}, \tau^2, \phi) \\ &\times p(\boldsymbol{\theta}_{2,T+h} | \boldsymbol{\theta}_{2,T}, \mathbf{W}) p(\Psi | \mathbf{Y}). \end{aligned}$$

Logo, tem-se que $p(\Psi | \mathbf{Y})$ pode ser amostrada através de um algoritmo que utiliza o amostrador de Gibbs, como descrito na subseção 2.2.2. As distribuições $p(\mathbf{Y}_{T+h} | \mathbf{X}_{T+h}, \boldsymbol{\mu}_{T+h}, \nu)$, $p(\mathbf{X}_{T+h} | \boldsymbol{\xi})$, $p(\boldsymbol{\mu}_{T+h} | \boldsymbol{\theta}_{1,T+h}, \sigma^2)$, $p(\boldsymbol{\theta}_{1,T+h} | \boldsymbol{\theta}_{2,T+h}, \tau^2, \phi)$ e $p(\boldsymbol{\theta}_{2,T+h} | \boldsymbol{\theta}_{2,T}, \mathbf{W})$ são conhecidas. Assim, para gerar amostras da distribuição *a posteriori* de \mathbf{Y}_{T+h} dados \mathbf{X}_{T+h} , $\boldsymbol{\mu}_{T+h}$ e ν , além de amostrar os demais parâmetros do modelo como detalhado na subseção 2.2.2, é necessário adicionar dois passos ao algoritmo MCMC e amostrar \mathbf{X}_{T+h} de $p(\mathbf{X}_{T+h} | \boldsymbol{\xi})$. Se $\mathbf{X}_{T+h}(s_i) \neq 0$, para $i = 1, 2, \dots, N$ e $t = 1, 2, \dots, T$, então amostra-se $\boldsymbol{\theta}_{2,T+h}$ de $p(\boldsymbol{\theta}_{2,T+h} | \boldsymbol{\theta}_{2,T}, \mathbf{W})$, cuja distribuição é a mesma explicitada em (23). Em seguida, $\boldsymbol{\theta}_{1,T+h}$ é amostrado da distribuição $p(\boldsymbol{\theta}_{1,T+h} | \boldsymbol{\theta}_{2,T+h}, \tau^2, \phi)$ e $\boldsymbol{\mu}_{T+h}$ é amostrado de $p(\boldsymbol{\mu}_{T+h} | \boldsymbol{\theta}_{1,T+h}, \sigma^2)$. Por fim, \mathbf{Y}_{T+h} é gerado de $p(\mathbf{Y}_{T+h} | \mathbf{X}_{T+h}, \boldsymbol{\mu}_{T+h}, \nu)$.

Interpolação Espacial

Suponha agora que os dados são observados em um conjunto de locais $S^o = (s_1, s_2, \dots, s_N) \subset \mathbb{R}^2$ e deseja-se interpolar a variável resposta para outras r localidades $S^n = (s_{N+1}, s_{N+2}, \dots, s_{N+r})$ em um dado período de tempo t . Considere também $\mathbf{v}_{1,t}^o = (\mathbf{v}_{1,t}^o(s_1), \dots, \mathbf{v}_{1,t}^o(s_N))'$ e $\mathbf{v}_{1,t}^n = (\mathbf{v}_{1,t}^n(s_{N+1}), \dots, \mathbf{v}_{1,t}^n(s_{N+r}))'$, vetores de erros aleatórios observados e não observados, respectivamente.

Quando o processo está presente, o modelo (24) pode ser escrito da seguinte maneira

$$Y_t^o(s_i) \sim Ga\left(\frac{1}{\nu^2}, \frac{1}{\nu^2 \mu_t^o(s_i)}\right) \quad \text{para } i = 1, 2, \dots, N \quad \text{e } t = 1, 2, \dots, T,$$

$$Y_t^n(s_{N+k}) \sim Ga\left(\frac{1}{\nu^2}, \frac{1}{\nu^2 \mu_t^n(s_{N+k})}\right) \quad \text{para } k = 1, 2, \dots, r.$$

As demais componentes observadas e não observadas do modelo podem ser descritas em notação matricial, de tal forma que

$$\begin{pmatrix} \log(\boldsymbol{\mu}_t^o) \\ \log(\boldsymbol{\mu}_t^n) \end{pmatrix} = \begin{pmatrix} \mathbf{F}_{1,t}^o & \mathbf{0}_{N \times r.p} \\ \mathbf{0}_{r \times N.p} & \mathbf{F}_{1,t}^n \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_{1,t}^o \\ \boldsymbol{\theta}_{1,t}^n \end{pmatrix} + \begin{pmatrix} \mathbf{v}_{1,t}^o \\ \mathbf{v}_{1,t}^n \end{pmatrix}, \quad \begin{pmatrix} \mathbf{v}_{1,t}^o \\ \mathbf{v}_{1,t}^n \end{pmatrix} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{N+r}),$$

$$\begin{pmatrix} \boldsymbol{\theta}_{1,t}^o \\ \boldsymbol{\theta}_{1,t}^n \end{pmatrix} = \begin{pmatrix} \mathbf{F}_{2,t}^o \\ \mathbf{F}_{2,t}^n \end{pmatrix} \boldsymbol{\theta}_{2,t} + \begin{pmatrix} \mathbf{v}_{2,t}^o \\ \mathbf{v}_{2,t}^n \end{pmatrix}, \quad \begin{pmatrix} \mathbf{v}_{2,t}^o \\ \mathbf{v}_{2,t}^n \end{pmatrix} \sim PG\left[\mathbf{0}, \begin{pmatrix} \mathbf{V}_2^o & \mathbf{V}_2^{on} \\ \mathbf{V}_2^{no} & \mathbf{V}_2^n \end{pmatrix} \otimes \mathbf{I}_p\right],$$

$$\boldsymbol{\theta}_{2,t} = \mathbf{G}_t \boldsymbol{\theta}_{2,t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}).$$

Seja $\boldsymbol{\Psi}^*$ o vetor formado pelos hiperparâmetros do modelo, ou seja, $\boldsymbol{\Psi}^* = (\boldsymbol{\xi}, \nu, \sigma^2, \tau^2, \phi, \boldsymbol{\theta}_{2,0}, \mathbf{W})$. Sob a hipótese de $\mathbf{F}_{1,t}^o$ e $\mathbf{F}_{1,t}^n$ serem conhecidas, a distribuição conjunta *a posteriori* é dada por

$$\begin{aligned} p(\mathbf{Y}_t^n, \mathbf{X}_t^n, \boldsymbol{\mu}_t^n, \boldsymbol{\mu}_t^o, \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* | \mathbf{Y}_t^o) &= p(\mathbf{Y}_t^n | \mathbf{X}_t^n, \boldsymbol{\mu}_t^n, \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^*) p(\mathbf{X}_t^n | \boldsymbol{\Psi}^*) \\ &\times p(\log(\boldsymbol{\mu}_t^n) | \log(\boldsymbol{\mu}_t^o), \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^*) p(\boldsymbol{\theta}_{1,t}^n | \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^*) \\ &\times p(\boldsymbol{\mu}_t^o, \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* | \mathbf{Y}_t^o). \end{aligned}$$

Inicialmente, são obtidas amostras de $p(\mathbf{X}_t^n | \boldsymbol{\Psi}^*)$ e $p(\boldsymbol{\mu}_t^o, \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* | \mathbf{Y}_t^o)$, através do procedimento detalhado na subseção 2.2.2. Em seguida, uma amostra de $p(\boldsymbol{\theta}_{1,t}^n | \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^*)$ é obtida utilizando propriedades da distribuição Normal, de tal forma que

$$\boldsymbol{\theta}_{1,t}^n | \boldsymbol{\theta}_{1,t}^o, \boldsymbol{\theta}_{2,t}, \boldsymbol{\Psi}^* \sim N(\mathbf{M}_1, \mathbf{H}_1),$$

em que \mathbf{M}_1 e \mathbf{H}_1 são como especificados na subseção 2.1.1. Analogamente, uma amostra de $p(\log(\boldsymbol{\mu}_t^n) | \log(\boldsymbol{\mu}_t^o), \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^*)$ pode ser obtida utilizando o seguinte resultado:

$$\log(\boldsymbol{\mu}_t^n) | \log(\boldsymbol{\mu}_t^o), \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^* \sim N(\mathbf{M}_2, \mathbf{H}_2),$$

em que \mathbf{M}_2 e \mathbf{H}_2 são como especificados na subseção 2.1.2. Por fim, uma amostra de $p(\mathbf{Y}_t^n | \mathbf{X}_t^n, \boldsymbol{\mu}_t^n, \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^*)$ é obtida utilizando o fato de que $p(\mathbf{Y}_t^n | \mathbf{X}_t^n, \boldsymbol{\mu}_t^n, \boldsymbol{\theta}_{1,t}^n, \boldsymbol{\Psi}^*) = \prod_{k=1}^r p(Y_t^n(s_{N+k}) | X_t^n(s_{N+k}), \boldsymbol{\mu}_t^n(s_{N+k}), \boldsymbol{\theta}_{1,t}^n(s_{N+k}), \boldsymbol{\Psi}^*)$ e $Y_t^n(s_{N+k}) | X_t^n(s_{N+k}), \boldsymbol{\mu}_t^n(s_{N+k}), \boldsymbol{\theta}_{1,t}^n(s_{N+k}), \boldsymbol{\Psi}^* \sim Ga\left(\frac{1}{\nu^2}, \frac{1}{\nu^2 \mu_t^n(s_{N+k})}\right)$.

3 Aplicações

Nesta seção, os modelos propostos serão aplicados a dois conjuntos de dados reais: o primeiro referente às observações trimestrais da quantidade de chuva coletadas ao longo de 103 anos na região leste da Austrália e o segundo relativo ao número de tornados reportados anualmente em 32 estados norte americanos, ao longo de 60 anos.

3.1 Modelando o Número de Tornados nos Estados Unidos

Do ponto de vista estatístico, a natureza não-gaussiana dos dados de contagem, os eventos raros e a presença de correlação espacial e temporal são alguns dos fatores que podem dificultar o processo de estimação. Nos modelos propostos, essas características são explicitamente acomodadas na estrutura do modelo. Em particular, a estrutura hierárquica bayesiana será considerada para avaliar a variabilidade espaço-temporal de um processo de contagem de tornados subjacente (não observável) dadas as observações com ruído (número de tornados reportado). Além disso, pode-se considerar os efeitos de processos climáticos exógenos sobre este processo de contagem. Na literatura, índices de atividade El Niño são frequentemente associados ao número de tornados reportado nos Estados Unidos e há uma variabilidade regional substancial nesta relação. Tendo em vista essa associação, será considerada a relação desse processo de contagem com o índice do El Niño/Oscilação Sul (ENSO), dado por Niño 3.4. Os valores deste índice utilizados no estudo foram obtidos como médias anuais, de modo a corresponder às contagens também anuais do número de tornados. Os dados referentes ao índice Niño 3.4, foram obtidos em dezembro de 2010 a partir do endereço eletrônico <http://www.cgd.ucar.edu/cas/catalog/climind> e as contagens de tornado a partir de <http://www.spc.noaa.gov/wcm>.

Neste estudo, considerou-se a distribuição espaço-temporal do número de tornados reportados anualmente em trinta e dois dos cinquenta estados norte americanos mais afetados por este fenômeno no período de 1950 a 2009, tendo como base a sua posição inicial de formação. A representação espacial é feita em termos das coordenadas geográficas latitude e longitude dos centróides dos estados e a dependência espacial é quantificada através da distância euclidiana entre os centróides. Para fazer a interpolação espacial, foi utilizada uma grade regular 20×20 (400 pontos), de forma a sobrepor a região de estudo. A Tabela 5 no apêndice A apresenta o nome dos estados considerados, bem como suas referências espaciais (latitude e longitude).

A estrutura temporal foi representada através da matriz bloco diagonal $\mathbf{F}_{1,t}$ que incorpora o logaritmo do índice ENSO denotado por $\log(\mathbf{Z}_t)$, além do intercepto, ou seja, $\mathbf{F}_{1,t} = \text{diag}(1, \log(\mathbf{Z}_t))$, para $t = 1, 2, \dots, T = 60$. A matriz $\mathbf{F}_{2,t}$ é especificada de forma que $\mathbf{F}_{2,t} = \mathbf{1}_N \otimes \mathbf{I}_2$, com $\mathbf{1}_N$ denotando o vetor coluna de tamanho $N = 32$ com todos os elementos iguais a 1 e \mathbf{I}_2 denota a matriz identidade de ordem 2. Por fim, a matriz de evolução \mathbf{G}_t é igual a \mathbf{I}_2 .

A Tabela 1 apresenta as principais estatísticas descritivas para os estados de Montana e Texas. Nota-se que o valor da variância amostral, em ambos os casos, é substancialmente maior que o valor da média amostral fornecendo indícios de sobredispersão nos dados.

Tabela 1: Estatísticas descritivas para as séries temporais referentes ao número de tornados reportado nos estados de Montana e Texas.

Estado	Mínimo	Máximo	Mediana	Média	Variância
Montana	0	30	4,50	6,13	33,85
Texas	13	235	132,00	130,63	2520,88

O interesse na climatologia dos tornados está diretamente relacionado ao seu potencial destrutivo. Claramente, a perda de vidas causadas por tornados é, por si só, um motivo suficiente para a obtenção de um entendimento maior acerca deste fenômeno. A Figura 1 mostra a evolução temporal do número de tornados ao longo de 60 anos, independente do estado considerado. Observa-se uma ligeira tendência de crescimento e um aumento na variabilidade associada a esse número, principalmente a partir de meados da década de 70.

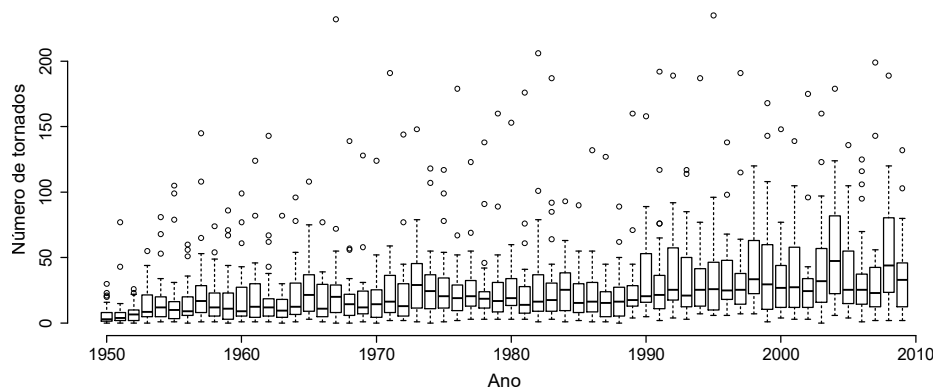


Figura 1: *Boxplots* do número de tornados por ano, considerando trinta e dois estados norte americanos.

Outra questão de extrema relevância é a influência da localização dos estados no número de tornados reportado. A Figura 2 ilustra o mapa dos Estados Unidos com os centróides de cada estado discriminados. O diâmetro dos círculos e as cores identificam os quintis da distribuição empírica do número total de tornados reportado por estado ao longo do período de estudo. Quanto maior o diâmetro dos círculos, maiores serão os totais de tornados. As cores auxiliam na diferenciação dos quintis. Pode-se observar que os estados na parte central do país apresentam maior incidência de tornados.

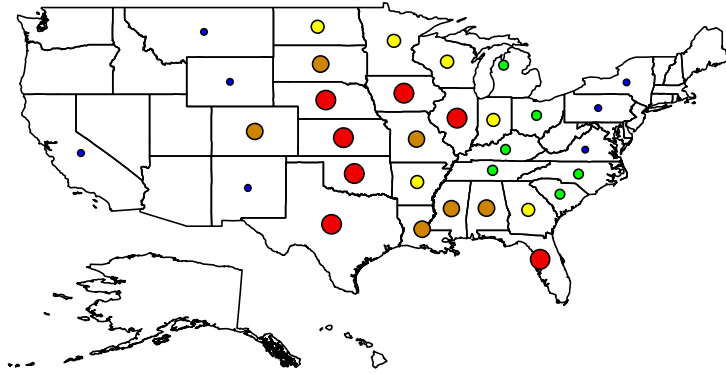


Figura 2: Mapa dos Estados Unidos com a localização dos centróides de trinta e dois estados. O diâmetro dos círculos e as cores identificam os quintis da distribuição empírica do número total de tornados reportado por estado ao longo do período de estudo.

Assim, fica evidente a dependência temporal e a variação espacialmente estruturada no número de tornados reportados nos Estados Unidos. Uma vez verificada esta dependência, pode-se aplicar os modelos descritos nas subseções 2.1.1 e 2.1.2.

3.1.1 Resultados

Os modelos com distribuição Poisson serão ajustados para o conjunto de dados em questão. Entretanto, como o modelo mais simples (sem efeito de sobredispersão) apresentou uma melhor performance, somente os resultados deste modelo serão apresentados a seguir.

A Figura 3 mostra os histogramas das amostras da distribuição *a posteriori* da variância τ^2 e do alcance espacial do processo, ϕ . O parâmetro ϕ representa a taxa de decaimento da correlação com a distância. A linha vermelha representa a média *a posteriori* de cada um desses parâmetros.

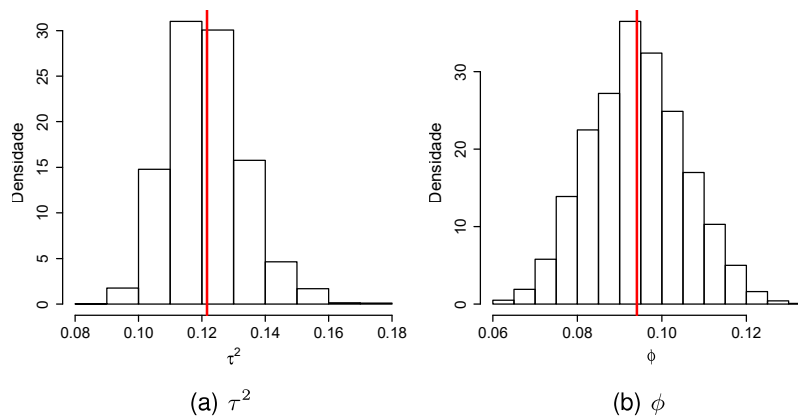


Figura 3: Histograma das amostras da distribuição *a posteriori* dos parâmetros da matriz de correlação espacial.

A Figura 4 ilustra o gráfico da função de correlação espacial estimada obtida a partir da média *a posteriori* de ϕ , com os seus respectivos intervalos de 95% de credibilidade. Pode-se verificar que a correlação decai para zero conforme a distância aumenta, o que se espera de um fenômeno com dependência espacial. A função de correlação estimada decai lentamente para zero, indicando uma alta dependência espacial do processo. Vale ressaltar que a distância máxima entre os centróides dos trinta e dois estados considerados é aproximadamente igual a 44,46. A Figura 5 apresenta os histogramas das amostras da distribuição *a posteriori* das componentes de \mathbf{W} e a Tabela 2 mostra um sumário da distribuição *a posteriori* e o valor da redução potencial de escala estimada (Gelman e Rubin, 1992) para cada um dos parâmetros fixos do modelo (6). De acordo com este critério, todas as cadeias atingiram a convergência.

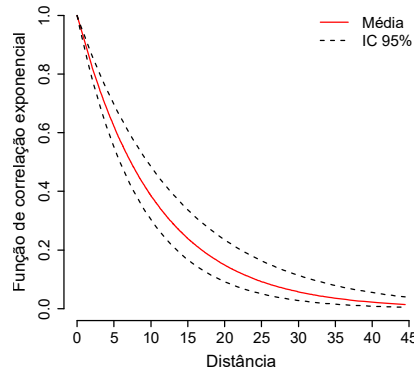


Figura 4: Função de correlação espacial estimada com seu respectivo intervalo de 95 % de credibilidade.

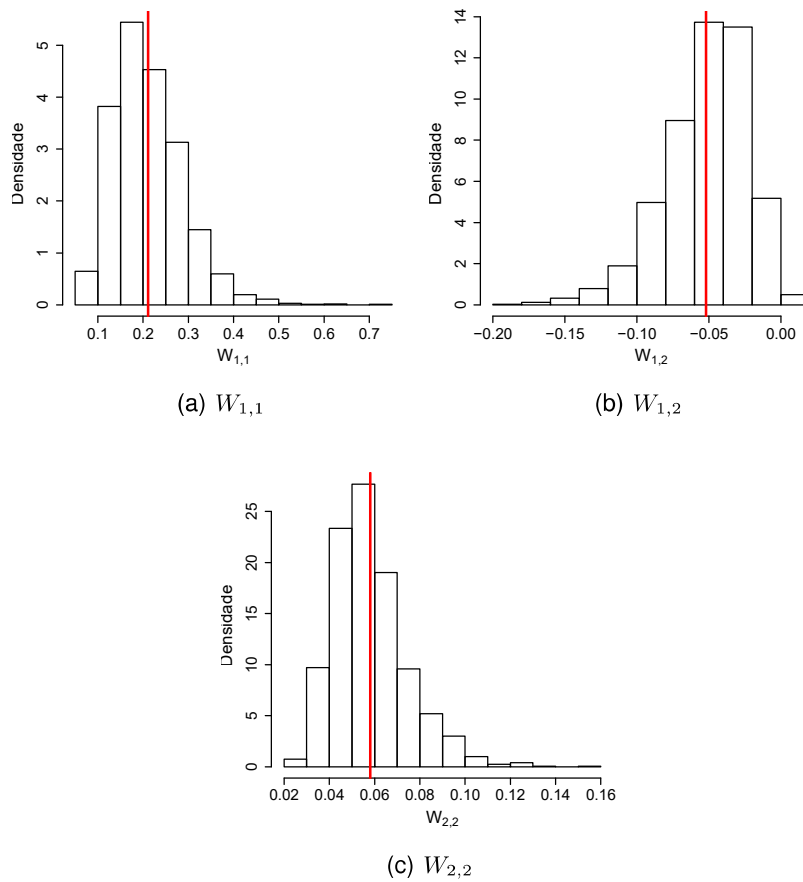


Figura 5: Histograma da distribuição *a posteriori* dos elementos da matriz de covariância W .

Tabela 2: Sumário da distribuição *a posteriori* para os parâmetros fixos do modelo sem efeito de sobre-dispersão.

Parâmetro	Média	Mediana	Quantil 2,5%	Quantil 97,5%	$\sqrt{\hat{R}}$
τ^2	0,1216	0,1206	0,1009	0,1482	1,002
ϕ	0,0941	0,0940	0,0726	0,1168	1,002
$W_{1,1}$	0,2113	0,2012	0,0961	0,3845	1,040
$W_{1,2}$	-0,0521	-0,0484	-0,1201	-0,0063	1,020
$W_{2,2}$	0,0581	0,0562	0,0331	0,0962	1,005

A Figura 6 mostra a evolução temporal das componentes do vetor de estados $\theta_{2,t}$. Os efeitos das componentes $\theta_{2.1,t}$ e $\theta_{2.2,t}$ são, em geral, significativos, ou seja, o intervalo de credibilidade *a posteriori* de 95% não compreende o valor zero indicando que as componentes variam de forma significativa ao longo da maioria dos instantes de tempo. Em particular, a série temporal de $\theta_{2.2,t}$ apresentou um efeito crescente, sugerindo que o efeito médio do índice ENSO no número de tornados reportado nos Estados Unidos vem aumentando ao longo do tempo.

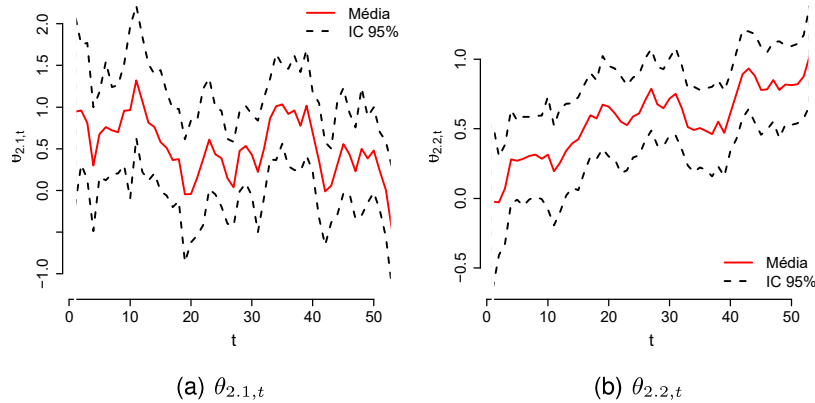


Figura 6: Evolução temporal das componentes do vetor de estados $\theta_{2,t}$.

Na Figura 7 está representada a evolução temporal das componentes do vetor de estados $\theta_{1,t}$ para o estado do Texas. A primeira componente, $\theta_{1.1,t}(s_{26})$, representa o intercepto e a segunda, $\theta_{1.2,t}(s_{26})$, o efeito do índice ENSO no logaritmo da média do número de tornados reportados para a localidade 26 (Texas). Nota-se que as séries temporais se mostraram significativas para a maioria dos instantes de tempo, sendo que a segunda componente apresentou uma tendência de crescimento ao longo dos anos, indicando uma influência direta do índice ENSO no número de tornados reportados. Para os demais estados estudados, essas componentes se comportaram de maneira similar.

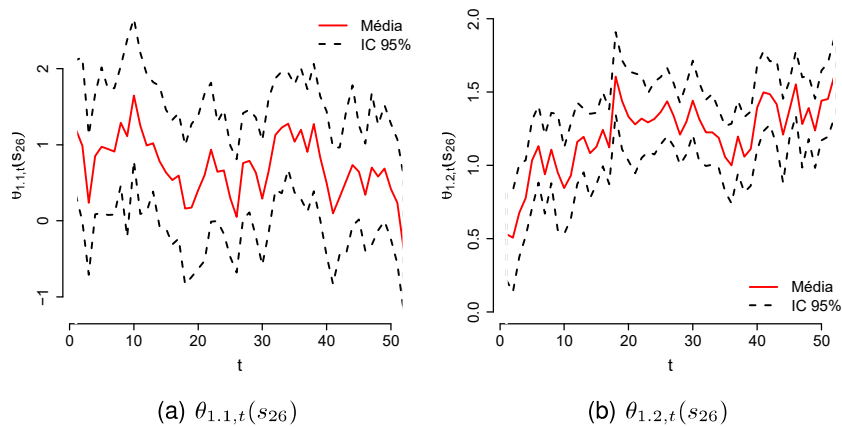


Figura 7: Evolução temporal das componentes do vetor de estados $\theta_{1,t}$ para o estado do Texas.

As superfícies interpoladas padronizadas de $\theta_{1.1}$ e $\theta_{1.2}$ para o tempo $t = 53$, correspondente à previsão um passo a frente são mostradas na Figura 8. A magnitude dos efeitos de $\theta_{1.1,53}$ e $\theta_{1.2,53}$ foi estimada a partir da média *a posteriori*. A Figura 9 ilustra o comportamento dos valores replicados e previsões obtidas para o estado do Texas, a partir da média *a posteriori*, com seus respectivos intervalos de 95% de credibilidade para um horizonte de previsão de 8 passos a frente. Pode-se notar que, para os dados replicados, os valores verdadeiros das observações estão todos dentro do intervalo de 95% de credibilidade e a média *a posteriori* captura bem a estrutura dos dados. Por outro lado, a capacidade preditiva do modelo é razoável, uma vez que ele consegue capturar o padrão dos valores previstos porém, mesmo para horizontes de previsão poucos passos a frente, as estimativas se afastam dos valores reais observados.

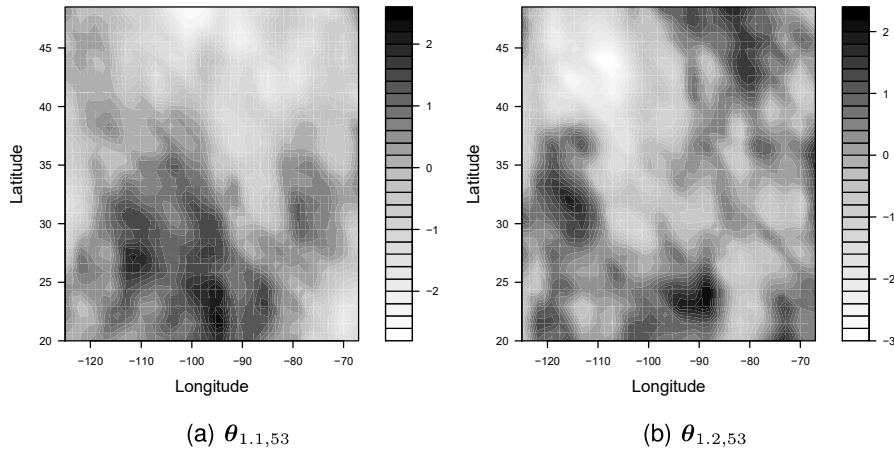


Figura 8: Superfícies interpoladas padronizadas de $\theta_{1.1}$ e $\theta_{1.2}$ para o instante de tempo $t = 53$, correspondente à previsão um passo a frente.

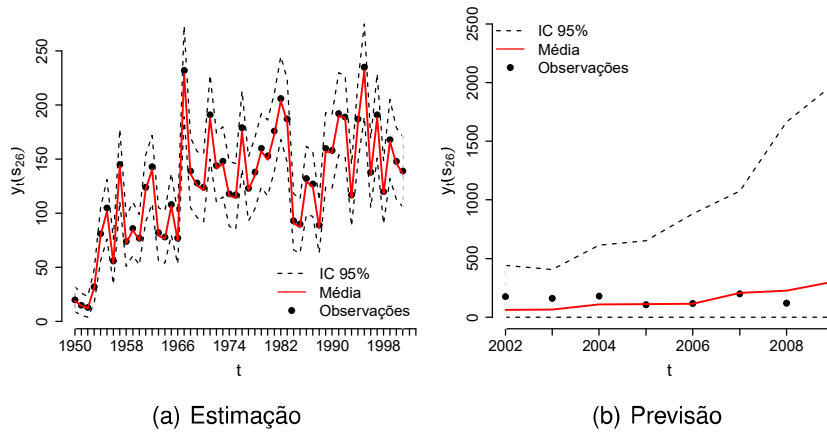


Figura 9: Valores replicados e previsões obtidas a partir da média *a posteriori*, com seus respectivos intervalos de 95% de credibilidade para um horizonte de previsão de 8 passos a frente no estado do Texas.

A Figura 10 ilustra a superfície interpolada padronizada no instante de tempo correspondente à previsão um passo a frente. Nota-se que o modelo conseguiu captar a chamada “Alameda dos Tornados”(do inglês *Tornado Alley*), um termo comum para designar a região central dos Estados Unidos onde os tornados ocorrem frequentemente.

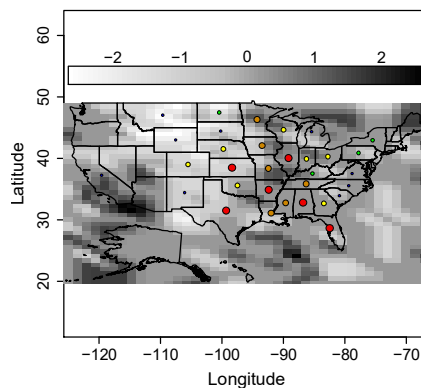


Figura 10: Mapa da superfície interpolada padronizada no instante de tempo $t = 53$, correspondente à previsão um passo a frente.

3.2 Modelando a Chuva na Região Leste da Austrália

Como dados ambientais, em geral, variam de forma suave no espaço e no tempo, é natural pensar que sejam correlacionados em ambas as dimensões. Em particular, a precipitação é um fenômeno caracterizado por alta variabilidade espacial e temporal, traduzindo a vulnerabilidade dos ecossistemas da região. Esse fenômeno, apesar do seu caráter regional, apresenta variações espaciais devido aos efeitos orográficos difíceis de serem conhecidos detalhadamente. Tendo em vista que a precipitação influencia, por exemplo, a fauna, a flora e a agricultura, deseja-se modelar a quantidade de chuva na região sob estudo. Uma vez que informações hidrológicas e meteorológicas precisas são obtidas, um efetivo planejamento de recursos hídricos, incluindo a adoção de medidas de prevenção a enchentes ou avisos de alerta para períodos de seca, pode ser adotado.

Os dados utilizados nesta aplicação foram gentilmente cedidos pela Dr^a Aline Maia da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e são referidos em [Maia et al. \(2007\)](#). O conjunto de dados é relativo à quantidade trimestral de chuva em 15 estações pluviométricas na Austrália de 1900 a 2002. A representação espacial é feita em termos das coordenadas geográficas latitude e longitude das estações pluviométricas e a dependência espacial é quantificada através da distância euclidiana entre as estações. A Tabela 6 no apêndice B apresenta o nome das estações pluviométricas consideradas, bem como suas referências espaciais (latitude e longitude). Para fazer a interpolação espacial, foi utilizada uma grade regular 10×10 (200 pontos), de forma a sobrepor a região de estudo.

A sazonalidade foi considerada no modelo de forma trimestral através da matriz bloco diagonal $F_{1,t}$ que incorpora como covariáveis, ondas de seno e cosseno com periodicidade de um ano, ou seja, $F_{1,t} = \text{diag}(1, \text{sen}(2\pi t/4), \text{cos}(2\pi t/4))$, para $t = 1, 2, \dots, T = 412$. A matriz $F_{2,t}$ é especificada de forma que $F_{2,t} = \mathbf{1}_N \otimes \mathbf{I}_3$, com $\mathbf{1}_N$ denotando o vetor coluna de tamanho $N = 15$ com todos os elementos iguais a 1 e \mathbf{I}_3 denota a matriz identidade de ordem 3. Por fim, a matriz de evolução G_t é igual a \mathbf{I}_3 . Vale ressaltar ainda que o índice j foi fixado em quatro, de acordo com as características do fenômeno explicitado no exemplo da subseção 2.2.1.

A Tabela 3 apresenta as principais estatísticas descritivas para as séries temporais observadas em Marree e Te Kowai Exp Stn. Pode-se notar que a variância amostral de ambas as estações é relativamente grande. Além disso, a quantidade média de chuva por trimestre varia bastante de uma estação para outra.

Tabela 3: Estatísticas descritivas para as séries temporais referentes às quantidades de chuva nas estações pluviométricas Marree e Te Kowai Exp Stn.

Estação	Mín.	Máx.	Mediana	Média	Variância	% de zeros
Marree	0,00	293,10	29,60	41,30	1587,05	0,97
Te Kowai Exp Stn	1,40	2469,00	259,80	413,90	184496,10	0,00

É sabido que existe uma alternância entre a estação chuvosa e a estação seca. A Figura 11 ilustra melhor o padrão sazonal do fenômeno pluviométrico. Ela foi obtida a partir das médias trimestrais da quantidade de chuva, independente do ano e considerando todas as estações pluviométricas. Nota-se que, o trimestre de maior precipitação é o primeiro, referente aos meses de janeiro, fevereiro e março. O terceiro trimestre, referente aos meses de julho, agosto e setembro é o mais seco. Além disso, observa-se que o primeiro e o último trimestres apresentam os maiores volumes médios de chuva. Outra questão importante é a influência da localização das estações pluviométricas na quantidade de chuva. A Figura 12 ilustra o gráfico das médias de chuva, por estação. Observa-se que, em geral, estações próximas ao litoral apresentam medidas de precipitação maiores do que aquelas situadas no interior.

Uma vez verificada a dependência espacial e temporal presente nos dados e baseado nas análises preliminares descritas anteriormente, o próximo passo é aplicar os modelos propostos que levam em consideração essas características.

3.2.1 Resultados

Para o conjunto de dados em questão, serão ajustados os modelos com distribuição Gama inflacionada de zeros descritos nas subseções 2.2.1 e 2.2.2, respectivamente.

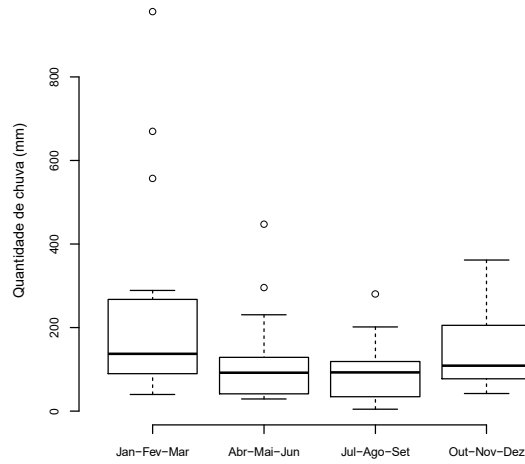


Figura 11: Boxplots das médias trimestrais da quantidade de chuva, independente do ano e considerando todas as estações pluviométricas.

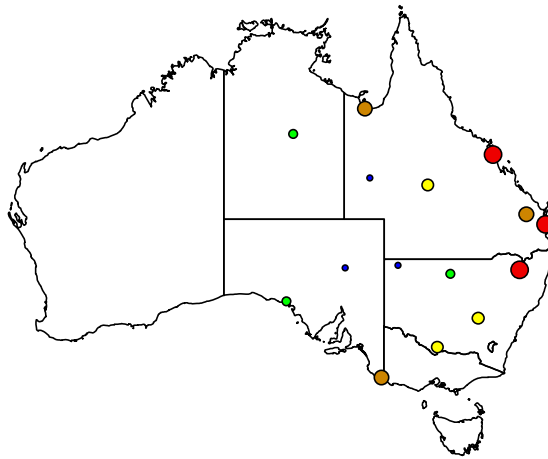


Figura 12: Mapa da Austrália com a localização das 15 estações pluviométricas. O diâmetro dos círculos e as cores identificam os quintis da distribuição empírica do volume médio de chuva por estação ao longo do período de estudo.

A Figura 13 apresenta o histograma das amostras da distribuição *a posteriori* das probabilidades de chuva (ξ_j) em quatro cenários distintos. O índice j foi fixado em quatro, de acordo com as características observadas do fenômeno pluviométrico na análise exploratória. Vale ressaltar que ξ_1 representa a probabilidade de chuva em um cenário com estação e período de tempo secos e ξ_2 é a probabilidade de chover dado que a estação é seca e o período chuvoso. Pode-se observar que os valores estimados de ξ_1 e ξ_2 estão em torno de 94% e 99,75%, respectivamente. Os parâmetros ξ_3 e ξ_4 representam, respectivamente, as probabilidades de chuva em uma estação chuvosa durante um período seco e em uma estação chuvosa ao longo de um período chuvoso. Observa-se que os valores estimados desses parâmetros estão em torno de 96% e 99,95%. A linha vermelha representa a média *a posteriori* para cada um desses parâmetros. Na Figura 14 está representado o histograma da distribuição *a posteriori* do coeficiente de variação. A estimativa pontual de ν , obtida através da média *a posteriori*, fornece evidências da heterogeneidade dos dados. A Figura 15 ilustra os histogramas das amostras da distribuição *a posteriori* da variância τ^2 e alcance ϕ do processo espacial que compõem a matriz de correlação espacial.

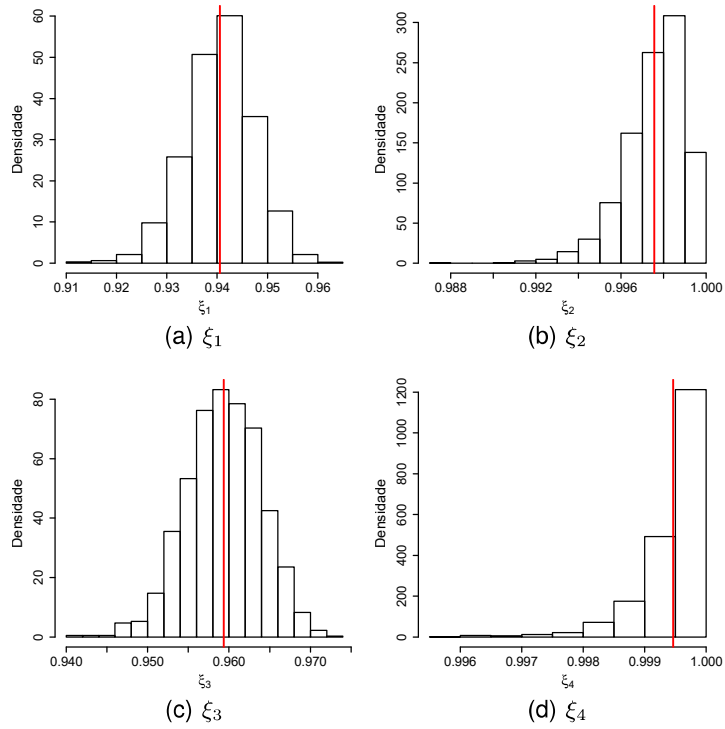


Figura 13: Histograma das amostras da distribuição *a posteriori* das probabilidades de chuva em quatro cenários distintos.

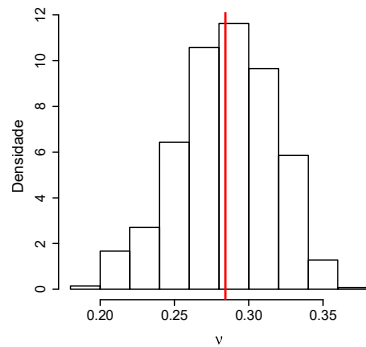


Figura 14: Histograma das amostras da distribuição *a posteriori* do coeficiente de variação.

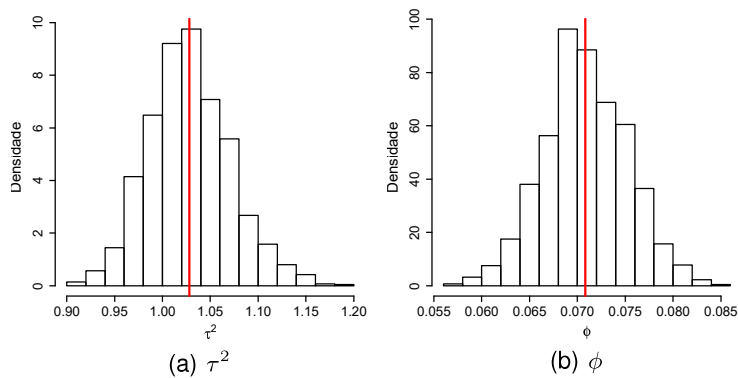


Figura 15: Histograma da distribuição *a posteriori* da variância e alcance do processo espacial.

A Figura 16 ilustra os gráficos da função de correlação espacial estimada obtida a partir da média *a posteriori* de ϕ para os modelos 3 e 4, juntamente com seus respectivos intervalos de 95% de credibilidade. Observa-se que as funções de correlação estimadas são muito semelhantes e decaem lentamente para zero, refletindo a alta dependência espacial e a igualdade aproximada das médias *a posteriori* de ϕ nos dois modelos. Vale ressaltar que a distância máxima observada entre as estações pluviométricas consideradas é aproximadamente igual a 20, 20.

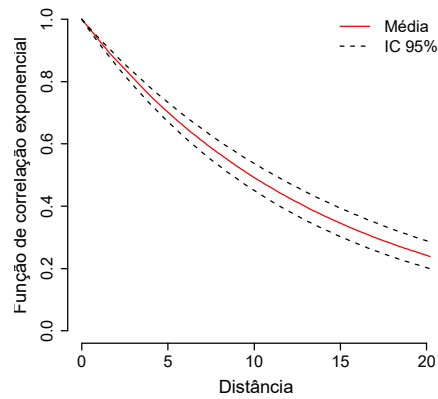
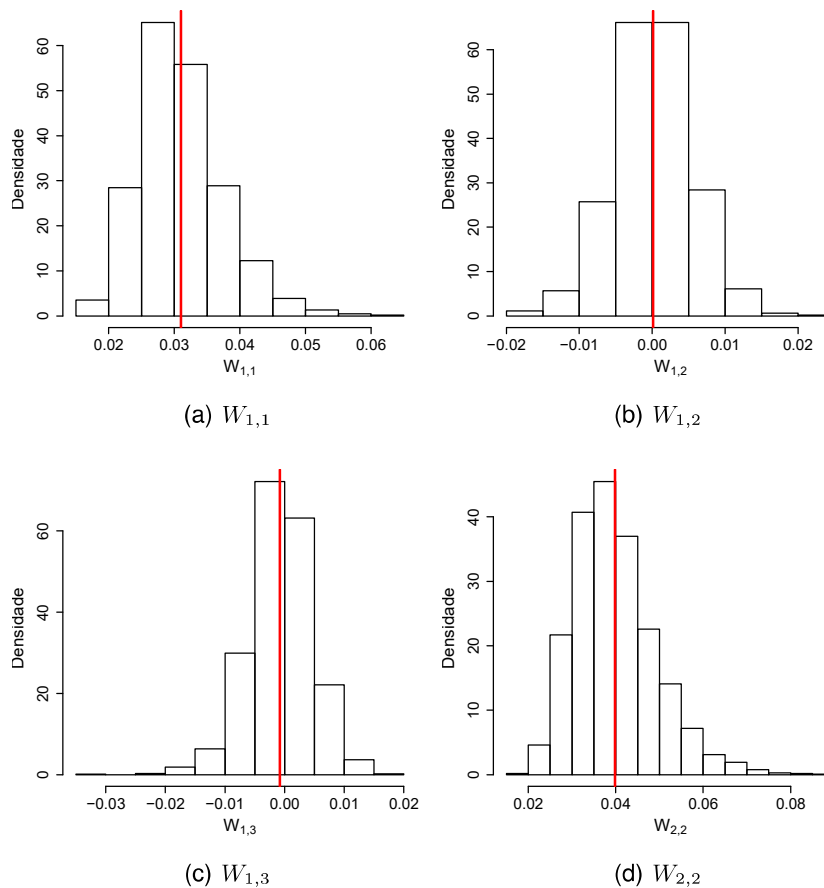


Figura 16: Gráfico da função de correlação espacial estimada com seu respectivo intervalo de 95 % de credibilidade.

Na Figura 3.2.1 estão os histogramas das amostras da distribuição *a posteriori* das componentes da matriz de covariâncias \mathbf{W} .



Histograma das amostras da distribuição *a posteriori* dos parâmetros da matriz de covariâncias \mathbf{W} .

A Tabela 4 mostra um sumário da distribuição *a posteriori* para os parâmetros fixos e o valor da redução potencial de escala estimada para cada um dos parâmetros fixos do modelo (18). De acordo

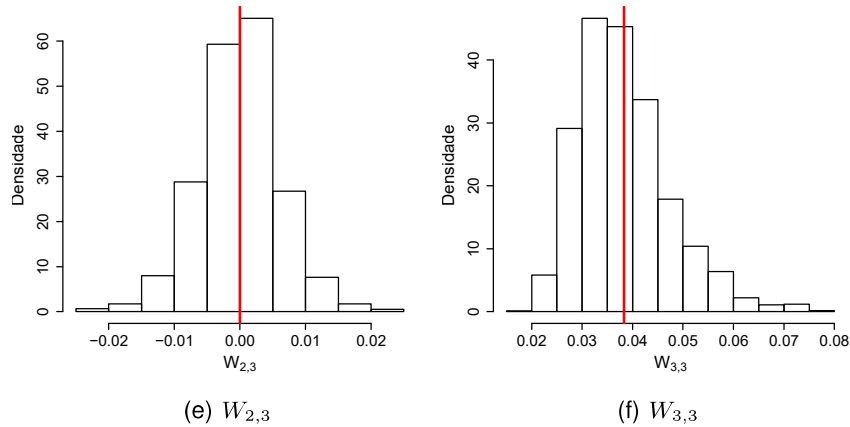


Figura 17: *

Histograma das amostras da distribuição *a posteriori* dos parâmetros da matriz de covariâncias W (continuação).

com este critério, todas as cadeias atingiram a convergência. Considerando as características pluviométricas no espaço e no tempo, pode-se dizer que a probabilidade de chover em cada um dos quatro cenários ajustados é bem alta. Assim, pode-se dizer que um percentual muito grande das observações é bem ajustado pelos modelos assumidos para descrever o processo sob estudo quando ele está presente. As maiores probabilidades de chuva foram estimadas para períodos de tempo considerados chuvosos, ou seja, o primeiro e o último trimestres do ano. A alta heterogeneidade dos dados refletida na estimativa do coeficiente de variação é possivelmente devida ao fato de que, no modelo sem o efeito de sobredispersão, o coeficiente de variação foi responsável por capturar toda a variabilidade remanescente das covariáveis, enquanto que no modelo com o efeito, o parâmetro σ^2 absorve parte desta variabilidade.

Tabela 4: Sumário da distribuição *a posteriori* para os parâmetros fixos do modelo sem efeito de sobredispersão.

Parâmetro	Média	Mediana	Quantil 2,5%	Quantil 97,5%	$\sqrt{\hat{R}}$
ξ_1	0,9406	0,9409	0,9267	0,9530	1,000
ξ_2	0,9976	0,9978	0,9941	0,9995	1,001
ξ_3	0,9594	0,9594	0,9502	0,9678	1,000
ξ_4	0,9995	0,9996	0,9980	1,0000	1,000
ν	0,2844	0,2852	0,2139	0,3412	1,010
τ^2	1,0280	1,0260	0,9502	1,1211	1,000
ϕ	0,0708	0,0707	0,0621	0,0797	1,003
$W_{1,1}$	0,0310	0,0303	0,0207	0,0455	1,000
$W_{1,2}$	0,0001	0,0001	-0,0106	0,0107	1,000
$W_{1,3}$	-0,0008	-0,0007	-0,0120	0,0093	1,002
$W_{2,2}$	0,0398	0,0387	0,0251	0,0615	1,014
$W_{2,3}$	0,0000	0,0000	-0,0123	0,0119	1,004
$W_{3,3}$	0,0383	0,0371	0,0249	0,0597	1,000

Na Figura 18, pode-se observar a evolução temporal das componentes de $\theta_{2,t}$ nos modelos 3 e 4. A componente $\theta_{2,2,t}$ apresentou, *a posteriori*, estimativas negativas ao longo de todos os períodos de tempo. Por outro lado, a componente $\theta_{2,3,t}$ apresentou somente estimativas positivas. Assim, tem-se indícios de que a quantidade de chuva tem uma correlação negativa com a componente seno e positiva com o cosseno. Entretanto, com exceção de $\theta_{2,1,t}$, as demais componentes de $\theta_{2,t}$ citadas anteriormente não se mostraram significativas em nenhum dos instantes de tempo.

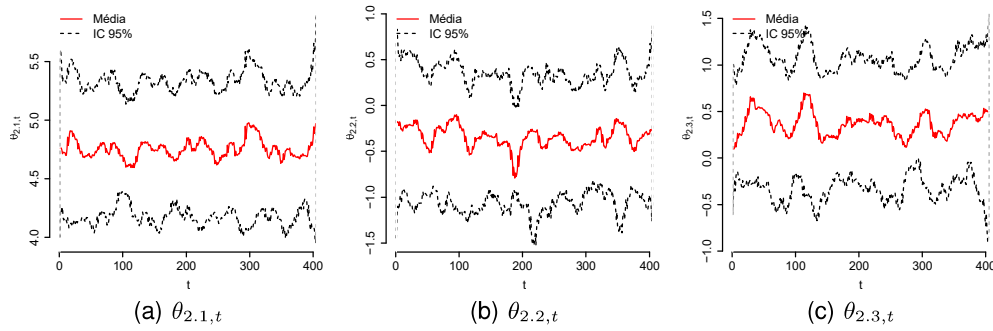


Figura 18: Evolução temporal das componentes de $\theta_{2,t}$.

Na Figura 19 está representada a evolução temporal das componentes do vetor $\theta_{1,t}$ para a estação pluviométrica Boulia Airport. A componente $\theta_{1,1,t}(s_4)$ representa o intercepto, $\theta_{1,2,t}(s_4)$ representa o coeficiente ou efeito das ondas de seno e $\theta_{1,3,t}(s_4)$, o efeito das ondas de cosseno no logaritmo da média da quantidade de chuva para a estação 4 (Boulia Airport). Note que, e com exceção da componente $\theta_{1,1,t}(s_4)$, as séries temporais não se mostraram significativas para a maioria dos instantes de tempo. Para as demais estações, não houve variação substancial no comportamento dessas componentes.

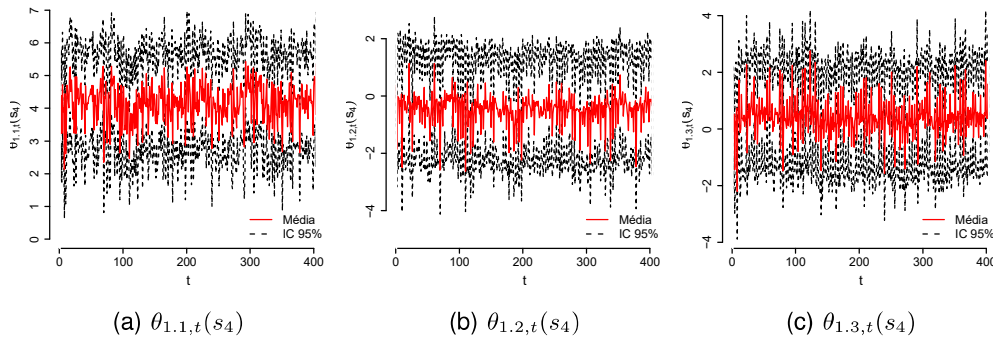


Figura 19: Evolução temporal das componentes do vetor $\theta_{1,t}$ para a estação Boulia Airport.

As superfícies interpoladas padronizadas de $\theta_{1,1}$, $\theta_{1,2}$ e $\theta_{1,3}$ para o instante de tempo $t = 405$, correspondente à previsão um passo a frente são mostradas na Figura 20. Pode-se verificar que as superfícies interpoladas para cada uma das componentes diferem substancialmente umas das outras, de acordo com o modelo considerado. Entretanto, com exceção de $\theta_{1,1,405}$, os efeitos das demais componentes não se mostraram significativos.

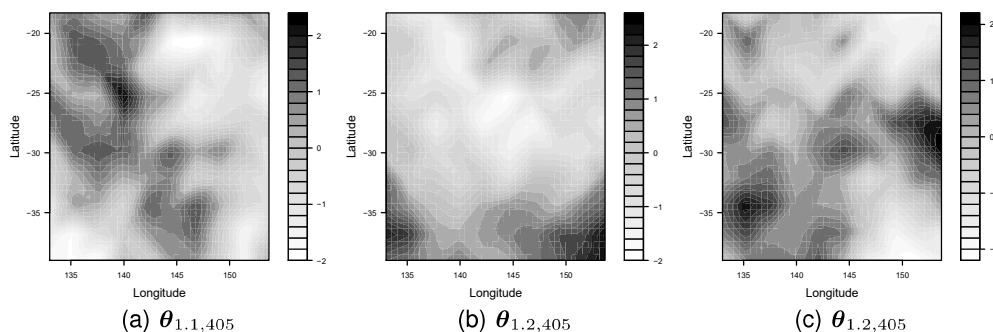


Figura 20: Superfícies interpoladas padronizadas de $\theta_{1,1}$, $\theta_{1,2}$ e $\theta_{1,3}$ para o instante de tempo $t = 405$, correspondente à previsão um passo a frente.

A Figura 21 ilustra o comportamento dos valores replicados da série temporal sob estudo para a estação Boulia Airport e as previsões obtidas para a mesma estação, a partir da mediana *a posteriori*, com seus respectivos intervalos de 95% de credibilidade para um horizonte de previsão de 8 passos a frente. Pode-se notar que, para os dados replicados, os valores verdadeiros das observações estão

todos dentro do intervalo de 95% de credibilidade *a posteriori* e a mediana *a posteriori* captura bem a estrutura dos dados. Observa-se também que a capacidade preditiva do modelo é razoável, uma vez que a mediana *a posteriori* captura bem o padrão da série, mas tende a superestimar os verdadeiros valores observados. Em geral, utilizando tanto a mediana como a média *a posteriori* para obter estimativas para as previsões, ocorre uma superestimação dos valores reais observados.

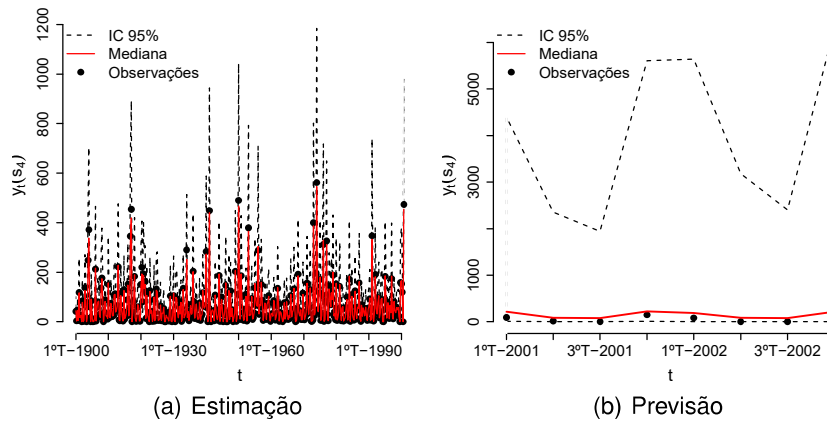


Figura 21: Valores replicados da série temporal para a estação Boulia Airport obtidos a partir da mediana *a posteriori*, com seus respectivos intervalos de 95% de credibilidade, nos modelos 3 e 4.

Na Figura 22 encontra-se o mapa da superfície interpolada padronizada no instante de tempo $t = 405$, correspondente à previsão um passo a frente. Como pode ser observado, o modelo conseguiu captar as áreas litorâneas como aquelas de maior concentração de chuva, estando de acordo com o que se sabe da literatura do fenômeno em questão.

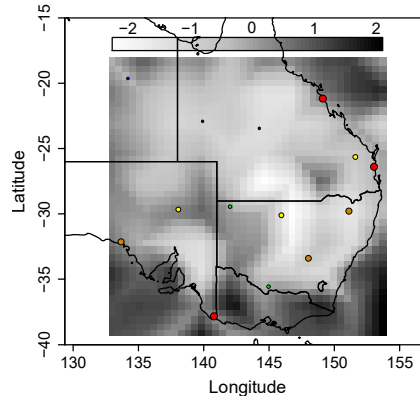


Figura 22: Mapas das superfícies interpoladas padronizadas no instante de tempo $t = 405$, correspondente à previsão um passo a frente.

4 Conclusões

Neste trabalho foi proposta uma classe de modelos dinâmicos hierárquicos para respostas pertencentes à família exponencial de distribuições, com o principal objetivo de obter um modelo capaz de verificar a estrutura espaço-temporal presente nos dados, permitindo fazer interpolação espacial e previsão temporal para processos de natureza não gaussiana. Os modelos propostos foram estruturados de forma hierárquica com parâmetros variando no tempo e/ou no espaço. A variação temporal foi definida através da especificação de estruturas dinâmicas no tempo, enquanto a variação espacial foi definida através de uma função de correlação paramétrica sob a suposição de que um dos termos de erro aleatório segue um processo gaussiano. Esses modelos têm a vantagem de permitir a interpolação

da variável resposta e dos coeficientes de regressão para qualquer conjunto de pontos não observados, levando em consideração a dependência espacial.

Com relação aos resultados encontrados na primeira aplicação, pode-se dizer que o modelo sem efeito de sobredispersão gerou estimativas mais estáveis dos parâmetros e obteve um desempenho muito próximo ao modelo com efeito de sobredispersão com respeito à estimação. Além disso, com relação à capacidade preditiva e à interpolação, o modelo sem efeito de sobredispersão também apresentou melhores resultados, apesar do menor número de parâmetros e da presença de sobredispersão nos dados. Isto sugere que a estrutura dinâmica imposta à média da distribuição Poisson foi capaz de capturar grande parte desse efeito. Os resultados da aplicação desse modelo indicaram a presença de uma alta dependência espacial, refletida principalmente através do índice ENSO. Este índice apresentou um efeito gradativamente crescente e significativo ao longo da grande maioria dos instantes de tempo, sugerindo uma influência direta no número de tornados reportados em quase todos os estados americanos, principalmente ao longo dos últimos 45 anos.

Na segunda aplicação foram considerados modelos de mistura para contemplar um cenário mais realista, no qual valores nulos para a quantidade de chuva podem ser observados. Os resultados da análise espaço-temporal indicaram uma probabilidade alta de chover em cada um dos cenários considerados, de forma que um percentual muito grande das observações é bem ajustado pelos modelos assumidos para descrever o processo sob estudo quando ele está presente. Assim, a distribuição Gama se mostrou uma boa alternativa para a modelagem do fenômeno em questão. Foi visto também que as maiores probabilidades de chuva foram estimadas para períodos de tempo chuvosos, que correspondem ao primeiro e ao último trimestres de cada ano. Os resultados indicaram uma alta dependência espacial do fenômeno pluviométrico, o que já era de se esperar. Quando a ocorrência de chuva é verificada, os Modelos 3 e 4 apresentaram desempenhos muito semelhantes. Com relação aos coeficientes da regressão, foi verificado para alguns instantes de tempo, que os efeitos sazonais incluídos na modelagem através das ondas de seno e cosseno foram significativos. Assim, a inclusão de parâmetros de estado específicos para capturar a sazonalidade dos dados não deve ser totalmente descartada. A estrutura espacial nos dados foi capturada, tendo o modelo se mostrado capaz de indicar áreas de menor e maior concentração de chuva em concordância com o que se conhece sobre o fenômeno pluviométrico na Austrália, de forma que as maiores concentrações de chuva ocorrem em áreas próximas ao litoral, onde também vive a maior parte da população. Por outro lado, o interior do país possui condições climáticas mais severas, compreendendo em sua maioria, áreas secas e esparsamente habitadas.

De um modo geral, a estimação paramétrica nos modelos propostos se mostrou bastante útil para explicar as variações espaço-temporais. Os modelos mais simples, ou seja, aqueles sem o termo de erro aleatório na função de ligação obtiveram melhores desempenhos com respeito à predição e à interpolação. Entretanto, previsões feitas a longo prazo tenderam a ser ineficientes, mostrando a necessidade de atualizar a informação proveniente dos dados para prever observações futuras. Isto pode ser devido à suposição de que a estrutura de correlação espacial seja a mesma em todos os parâmetros de regressão. Uma possível solução para esse problema é trabalhar com extensões desses modelos, considerando diferentes estruturas de correlação espacial. Uma limitação de algumas dessas extensões é o fato de que, em geral, torna-se difícil estimar os parâmetros de uma matriz de correlações espacialmente estruturada quando se tem pouca quantidade de informação sobre a variação espacial. Do ponto de vista computacional, um número muito grande de localidades pode ser um problema, principalmente se os modelos com o termo de erro aleatório na função de ligação forem considerados.

Uma série de hipóteses simplificadoras foram consideradas na especificação das estruturas de correlação espacial. Dentre elas estão: a hipótese de homogeneidade do processo espacial ($Var \{v_{2,t}(s)\} = \tau^2$ para todo $s \in \mathbb{R}^2$), estacionariedade e isotropia. Muitas outras possibilidades podem ser exploradas para abordar a variação espacial do processo. Outras famílias de funções de correlação podem ser utilizadas, tais como a família esférica e a família Matérn. A metodologia descrita aqui mostrou ser eficiente para a estimação paramétrica nos modelos propostos, os quais se mostraram flexíveis o suficiente para fornecer uma descrição razoável de processos que variam no espaço e no tempo.

A Aplicação 1

A Tabela 5 apresenta os códigos de identificação, o nome e as coordenadas geográficas longitude e latitude dos estados norte-americanos utilizados neste estudo.

Tabela 5: Localização Geográfica para trinta e dois estados norte americanos.

Código	Estado	Longitude	Latitude
1	Montana	-109,65180	47,03229
2	Wyoming	-107,55250	42,99831
3	Nova Iorque	-75,51261	42,94659
4	Pensilvânia	-77,80314	40,87481
5	Califórnia	-119,60020	37,24179
6	Ohio	-82,78750	40,28747
7	Kentucky	-85,29040	37,52644
8	Virginia	-78,85184	37,51851
9	Tenessi	-86,33410	35,84663
10	Novo México	-106,10180	34,42335
11	Carolina do Sul	-79,40156	35,54712
12	Michigan	-85,43860	44,33655
13	Dacota do Norte	-100,47030	47,44608
14	Dacota do Sul	-100,23490	44,43333
15	Wisconsin	-90,00605	44,63314
16	Minnesota	-94,30568	46,31569
17	Iowa	-93,50087	42,07600
18	Nebraska	-99,80740	41,52316
19	Indiana	-86,27616	39,91269
20	Illinóis	-89,20369	40,06397
21	Colorado	-105,54730	38,99727
22	Cansas	-98,38290	38,48264
23	Missouri	-92,48002	38,36395
24	Oklahoma	-97,50915	35,58391
25	Carolina do Norte	-79,40156	35,54712
26	Texas	-99,35174	31,49098
27	Alabama	-86,82636	32,79469
28	Mississippi	-89,66394	32,75903
29	Geórgia	-83,45433	32,65494
30	Arcansas	-92,43365	34,89786
31	Luisiana	-92,02902	31,09555
32	Flórida	-82,49574	28,65839

B Aplicação 2

A Tabela 6 apresenta o nome e as coordenadas geográficas das estações monitoradoras de chuva na Austrália, utilizados na Aplicação 2.

Tabela 6: Localização Geográfica para quinze estações monitoradoras de chuva na Austrália.

Código	Nome da Estação	Longitude	Latitude
1	Tennant Creek Post Office	134.190	-19.6300
2	Marree	138.060	-29.6500
3	Ceduna Post Office	133.670	-32.1300
4	Boulia Airport	139.910	-22.9100
5	Tibooburra Post Office	142.010	-29.4400
6	Bourke Post Office	145.940	-30.0900
7	Mount Gambier Post Office	140.780	-37.8300
8	Burketown Post Office	139.550	-17.7400
9	Te Kowai Exp Stn	149.120	-21.1600
10	Longreach Post Office	144.250	-23.4500
11	Gayndah Post Office	151.610	-25.6300
12	Tewantin Post Office	153.040	-26.3900
13	Inverell Comparison	151.110	-29.7800
14	Forbes (Camp Street)	148.010	-33.3900
15	Deniliquin (Wilkinson St)	144.950	-35.5600

Referências

- Carter, C. K. e Kohn, R. (1994) On gibbs sampling for state space models. *Biometrika*, **81**, 541–553.
- Doornik, A. (2008) *Object-Oriented Matrix Programming Using Ox*. Press and Oxford, London, England. URL <http://www.doornik.com>.
- Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**, 183–202.
- Gamerman, D. e Migon, H. (1993) Dynamic hierarchical models. *Journal of the Royal Statistical Society*, **55(3)**, 629–642.
- Gelman, A. e Rubin, D. R. (1992) A single series from the gibbs sampler provides a false sense of security. *Relatório técnico*, Oxford University Press.
- Ghil, M., Cohn, S., Tavantzis, J., Bube, K. e Isaacson, E. (1981) Applications of estimation theory to numerical weather prediction. Em *Dynamic Meteorology: Data Assimilation Methods* (eds. L. Bengtsson, M. Ghil e E. Källén), 139–224. New York: Springer-Verlag.
- Guttorp, P., Meiring, W. e Sampson, P. D. (1994) A space-time analysis of ground-level ozone data. *Environmetrics*, **5**, 241–254.
- Harrison, P. J. e Stevens, C. F. (1976) Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society*, **38**, 205–247.
- Huerta, G., Sansó, B. e Stroud, J. R. (2004) A spatio-temporal model for mexico city ozone levels. *Journal of the Royal Statistical Society*, **53(2)**, 231–248.
- Lindley, D. V. e Smith, A. F. M. (1972) Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, 1–41.
- Maia, A. H. N., Meinke, H. B., Lennox, S. e Stone, R. C. (2007) Inferential, nonparametric statistics to assess the quality of probabilistic forecast systems. *Monthly Weather Review*, **135**, 351–362.

- Min, C. e Zellner, A. (1993) Bayesian and non-bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics*, **56**, 89–118.
- R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- West, M. e Harrison, P. (1997) *Bayesian Forecasting and Dynamic Models*. Springer Verlag, New York.
- Wikle, C. K., Berliner, L. M. e Cressie, N. A. C. (1998) Hierarchical bayesian space-time models. *Environmental and Ecological Statistics*, **5(2)**, 117–154.
- Zellner, A., Hong, C. e Min, C. (1991) Forecasting turning points in international output growth rates using bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques. *Journal of Econometrics*, **49**, 275–304.