

Modelos Espaço-Temporais para Dados de Área

Marco A. R. Ferreira (marco@im.ufrj.br)
Universidade Federal do Rio de Janeiro
<http://www.dme.ufrj.br/marco>

Colaboração com Juan C. Vivar-Rojas

Organização da Palestra

- O que são dados espaço-temporais?
- Alguns exemplos
- Uma nova classe de modelos espaço-temporais
- Aplicações
- Discussão

O que são dados espaço-temporais?

- São dados observados em alguma região de interesse ao longo do tempo
- Aqui vamos considerar um caso particular de dados espaciais: dados de área
- A região de interesse é particionada em sub-regiões, e a cada sub-região é relacionada uma medição a cada tempo

Exemplo: Vento no Pacífico

Componente leste-oeste do vetor de velocidade de vento em uma região do Oceano Pacífico nos trópicos no período de Novembro de 1992 a Fevereiro de 1993, medida em metros/segundo (m/s).

Esta porção do Oceano Pacífico faz parte da região conhecida como "warm pool", e é crítica para a formação e manutenção de vários fenômenos climáticos. Vamos considerar ventos a cada 24 horas em uma resolução espacial de 2 graus, ou seja, aproximadamente 210km por 210 km.

Exemplo: Violência no RJ

Número anual de homicídios por município do estado do Rio de Janeiro de 1979 a 1998

Dados fornecidos pelo Dr. Oswaldo Gonçalves Cruz
(FIOCRUZ)

Modelo Espaço-Temporal

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{F}'_t \mathbf{X}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t &\sim N(\mathbf{0}, \mathbf{V}_t), \\ \mathbf{X}_t &= \mathbf{G}_t \mathbf{X}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t &\sim CAMG(\mathbf{0}, \mathbf{Q}), \end{aligned}$$

- \mathbf{Y}_t -> campo observado no tempo t vetorizado;
- \mathbf{X}_t -> processo latente para o qual queremos fazer inferência;
- \mathbf{F}_t -> relaciona o campo observado ao processo latente;
- \mathbf{G}_t -> matriz de evolução;
- $CAMG$ -> Campo Aleatório Markoviano Gaussiano;
- \mathbf{Q} -> matriz de precisão do $CAMG$.

Modelando F_t

- Caso mais simples: F_t é a matriz identidade. Processo latente é interpretado como a média do processo observado
- Casos mais interessantes: Dimensão do vetor de estados X_t muito menor do que a dimensão do campo observado Y_t
 - Modelo fatorial: F_t seria a matriz de pesos. X_t seriam os fatores ao longo do tempo
 - Representação multi-escala para o processo latente; neste caso, F_t faria a conexão entre as diferentes escalas

Modelando G_t

- G_t pode incorporar conhecimento a respeito do processo físico na modelagem (e.g. Wikle e Cressie 1999)
- G_t pode fazer o modelo estacionário ou não-estacionário
- Modelo de crescimento
- Modelo de contaminação

Dependência Espacial

ω_t , o distúrbio da equação de sistema, é espacialmente correlacionado. Sua função de densidade de probabilidade é:

$$p(\omega_t) \propto \exp\left(-\frac{1}{2}\omega_t^T \mathbf{Q}\omega_t\right)$$

$$\mathbf{Q} = \tau(\mathbf{I}_n + \phi\mathbf{H})$$

$$\tau > 0 \text{ e } \phi > 0$$

Dependência Espacial

$$(H)_{kl} = \begin{cases} h_k, & k = l \\ -g_{kl}, & k \in N_l \\ 0, & \text{caso contrário,} \end{cases}$$

N_l é o conjunto de regiões vizinhas à região l
 $g_{kl} > 0$ mede a similaridade entre as regiões k e l

$$h_k = \sum_{l \in N_k} g_{kl}$$

A matriz H é simétrica e de posto $n - 1$
 H permite a incorporação de sistemas de vizinhança bastante flexíveis (Ferreira e De Oliveira 2004).

Modelo polinomial de primeira ordem

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{X}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \\ \mathbf{X}_t &= \rho \mathbf{X}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t &\sim \text{CAMG}(\mathbf{0}, \mathbf{Q}) \end{aligned}$$

Modelo polinomial de segunda ordem

$$\mathbf{Y}_t = \mathbf{X}_{1t} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{V}_t)$$

$$\mathbf{X}_{1t} = \rho_1(\mathbf{X}_{1,t-1} + \mathbf{X}_{2,t-1}) + \boldsymbol{\omega}_{1t}, \quad \boldsymbol{\omega}_{1t} \sim CAMG(\mathbf{0}, \mathbf{Q}_1)$$

$$\mathbf{X}_{2t} = \rho_2 \mathbf{X}_{2,t-1} + \boldsymbol{\omega}_{2t}, \quad \boldsymbol{\omega}_{2t} \sim CAMG(\mathbf{0}, \mathbf{Q}_2)$$

Modelo com Sazonalidade

$$\mathbf{Y}_t = \mathbf{X}_t + \mathbf{S}_t + \boldsymbol{\epsilon}_t,$$

$$\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{V}_t)$$

$$\mathbf{X}_t = \mathbf{X}_{t-1} + \boldsymbol{\omega}_t,$$

$$\boldsymbol{\omega}_t \sim CAMG(\mathbf{0}, \mathbf{Q}_1)$$

$$\mathbf{S}_t = -(\mathbf{S}_{t-1} + \dots + \mathbf{S}_{t-p+1}) + \boldsymbol{\nu}_t,$$

$$\boldsymbol{\nu}_t \sim CAMG(\mathbf{0}, \mathbf{Q}_2)$$

Modelo de Contaminação

$$\mathbf{F}_t = \mathbf{I}_n$$

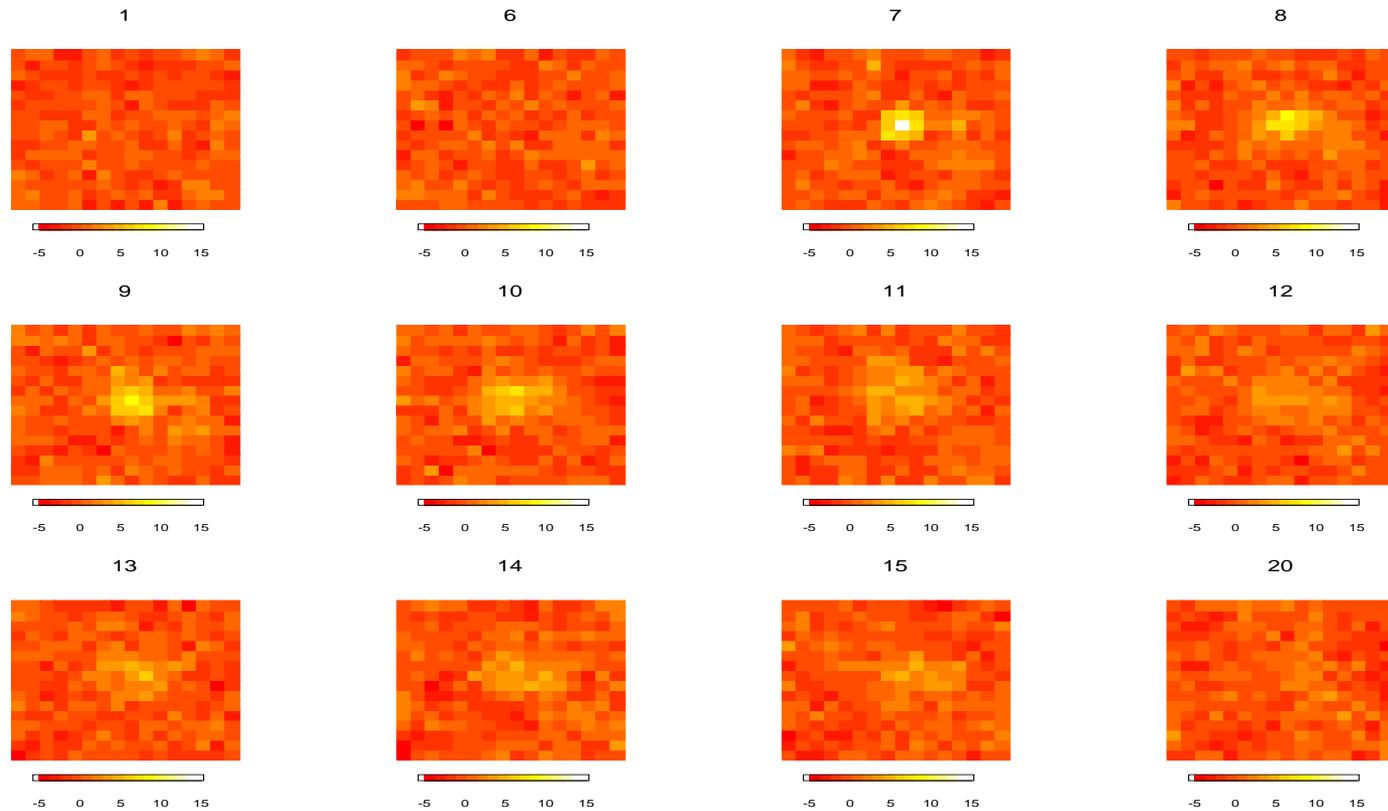
$$(\mathbf{G}_t)_{kl} = \rho \begin{cases} 1, & k = l, \\ \beta, & k \in C_l, \\ 0, & \text{caso contrário,} \end{cases}$$

- C_l -> conjunto de vizinhos da região l através do tempo
- $\beta > 0$ -> coeficiente de contaminação
- $\rho > 0$ mede persistência temporal

Seja $h = \max_l \sum_{k=1, k \neq l}^n \mathbf{1}(k \in C_l)$

Se $\rho(1 + \beta h) \in (0, 1)$ então o processo é estacionário

Modelo de Contaminação: Exemplo



Distribuições a priori

Para o modelo polinomial de primeira ordem:

$$p(\sigma^2) = IG(n_0/2, n_0 s_0^2/2) \quad p(\rho) = U(-1, 1)$$

$$p(\tau, \phi) \propto \frac{\sqrt{\sum_{k=1}^{n-1} \lambda_k^2 (\phi \lambda_k + 1)^{-2} - \frac{1}{n-1} \left[\sum_{k=1}^{n-1} \lambda_k (\phi \lambda_k + 1)^{-1} \right]^2}}{\tau}$$

onde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{n-1} \geq \lambda_n = 0$ são os autovalores de \mathbf{H} .

Com estacionaridade: $p(\mathbf{X}_0 | \rho, \tau, \phi) = CAMG(\mathbf{0}, (1 - \rho^2)\mathbf{Q})$

Inferência

- Paradigma Bayesiano
- Métodos de Monte Carlo via cadeias de Markov
- Filtro de Informação Progressivo, Amostrador Regressivo

Filtro de Informação Progressivo

Filtro de informação progressivo no tempo (Anderson e Moore, 1979) para $t = 1, \dots, T$:

- Priori em t : $\mathbf{X}_t | \theta, \mathbf{D}_{t-1} \sim N(\mathbf{a}_t, \mathbf{R}_t)$,
- Posteriori em t : $\mathbf{X}_t | \theta, \mathbf{D}_t \sim N(\mathbf{m}_t, \mathbf{C}_t)$,

onde

$$\begin{aligned}\mathbf{a}_t &= \mathbf{G}_t \mathbf{m}_{t-1} \\ \mathbf{R}_t^{-1} &= \mathbf{W}_t^{-1} - \mathbf{W}_t^{-1} \mathbf{G}_t \\ &\quad \times \left(\mathbf{G}_t \mathbf{W}_t^{-1} \mathbf{G}_t' + \mathbf{C}_{t-1}^{-1} \right)^{-1} \mathbf{G}_t' \mathbf{W}_t^{-1} \\ \mathbf{C}_t^{-1} &= \mathbf{R}_t^{-1} + \mathbf{F}_t \mathbf{V}_t^{-1} \mathbf{F}_t' \\ \mathbf{C}_t^{-1} \mathbf{m}_t &= \mathbf{R}_t^{-1} \mathbf{a}_t + \mathbf{F}_t \mathbf{V}_t^{-1} \mathbf{Y}_t\end{aligned}$$

Amostrador Regressivo

Simular $\mathbf{X}_T^* | \theta, D_T \sim N(\mathbf{m}_T, \mathbf{C}_T)$.

Simular regressivamente para $t = T - 1, \dots, 0$
 $\mathbf{X}_t^* | \mathbf{X}_{t+1}^*, \theta, D_T \sim N(\mathbf{b}_t, \mathbf{B}_t)$

onde

$$\begin{aligned}\mathbf{B}_t^{-1} &= \mathbf{C}_t^{-1} + \mathbf{G}'_{t+1} \mathbf{W}_{t+1}^{-1} \mathbf{G}_{t+1} \\ \mathbf{B}_t^{-1} \mathbf{b}_t &= \mathbf{C}_t^{-1} \mathbf{m}_t + \mathbf{G}'_{t+1} \mathbf{W}_{t+1}^{-1} \mathbf{X}_{t+1}^*.\end{aligned}$$

Análise dos dados de vento no Pacífico

- Ruído foi adicionado aos campos de vento
- Modelo polinomial de primeira ordem foi utilizado para tentar recuperar os campos de vento originais
- Estrutura de primeira ordem para os erros de evolução

Análise da violência no RJ

- Transformação raiz quadrada para estabilizar a variância
- Modelo polinomial de segunda ordem
- Estrutura de primeira ordem para os erros de evolução

Discussão

- Novos modelos espaço-temporais
- Herdam a flexibilidade dos modelos de espaço de estados
- Aplicações a dados de vento e de violência