

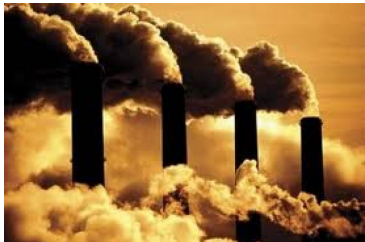
# Modelos de espaço de estados para processos não gaussianos

**Mariane Branco Alves**

IM/UFRJ - mariane@im.ufrj.br

Rio, Outubro/2014

# Motivação: poluição atmosférica...



## ... e seus efeitos sobre a saúde



# Objetivo

- ✓ Do ponto de vista da aplicação em tela → construção de modelos estatísticos que permitam avaliar a associação entre níveis de poluição atmosférica e desfechos epidemiológicos.
- ✓ Do ponto de vista da modelagem estatística → classe de modelos que:
  - contemple respostas não-gaussianas
  - trate formalmente autocorrelação temporal
  - admita variação temporal nos efeitos de componentes estruturais
  - permita tratar efeito cumulativo de uma ou mais regressoras.

# Elementos de inferência bayesiana

## Teorema de Bayes

Observações  $y$ : descritas por uma densidade  $f(y|\theta)$

Verossimilhança:  $l(\theta) = f(y|\theta)$

$\theta$ : índice de  $f$  (parâmetro)

Situação Canônica: amostra aleatória

$y = (y_1, \dots, y_n)$  obtida de  $f(y|\theta)$ .

### Exemplo (Alexandra M. Schmidt)

Medidas de uma quantidade física  $\theta$  com erros  $e_i \sim N(0, \sigma^2)$ ,  $\sigma^2$  conhecido.

$$y_i = \theta + e_i, \quad i = 1, \dots, n \text{ e } f(y|\theta) =$$

$$\prod_{i=1}^n f(y_i; \theta, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \frac{(y_i - \theta)^2}{\sigma^2} \right\}$$

- $\theta$  é mais do que um índice
- A situação se repete em casos mais gerais
- Muito provável que o pesquisador tenha informação *a priori* sobre  $\theta$
- Esta informação pode ser modelada através da densidade  $p(\theta)$

Processo de inferência baseado na distribuição de  $\theta$  após observar  $y \rightarrow$  **distribuição a posteriori** (em **contraste priori**).

Posteriori obtida através do teorema de Bayes como

$$p(\theta|y) = \frac{g(\theta, y)}{f(y)} = \frac{p(\theta)f(y|\theta)}{f(y)} \quad \text{ou}$$

$$\pi(\theta) \propto p(\theta) l(\theta)$$

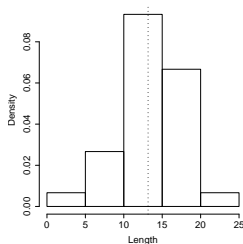
dist. a posteriori  $\propto$  dist. a priori  $\times$  verossimilhança

e

$$f(y) = \int_{\theta} f(y|\theta)p(\theta)d\theta$$

# Um exemplo hipotético

O comprimento de  $n = 30$  peixes observados no Lago St. Pierre no último verão. Observou-se  $\bar{y} = 13.28$ . Assuma  $Y_i | \theta \sim N(\theta, \sigma^2)$  com  $\sigma^2 = 4$  e  $\theta \sim N(\mu, \tau^2)$

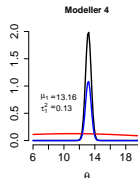
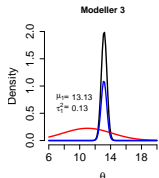
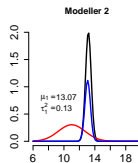
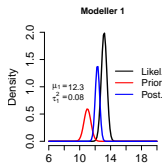


Pode-se mostrar que  $\theta | \mathbf{y} \sim N(\mu_1, \tau_1^2)$  onde  $\mu_1 = \frac{n\sigma^{-2}\bar{y} + \tau^{-2}\mu}{n\sigma^{-2} + \tau^{-2}}$  e  $\tau_1^{-2} = n\sigma^{-2} + \tau^{-2}$ .



# Uma análise bayesiana...

Todos os pesquisadores acreditam que  $\mu = 11$ , mas suas **incertezas** sobre a média variam...



# Previsões

Uma vez obtida a distribuição a posteriori de  $\theta$ , pode-se obter a distribuição preditiva de futuras mensurações:

$$\begin{aligned}f(y^f | \mathbf{y}) &= \int_{\theta} g(y^f, \theta | \mathbf{y}) d\theta \\ &= \int_{\theta} f(y^f | \theta, \mathbf{y}) p(\theta | \mathbf{y}) d\theta \\ &= \int_{\theta} f(y^f | \theta) p(\theta | \mathbf{y}) d\theta\end{aligned}$$

## Contexto temporal: notação

$D_0$ : conjunto contendo a informação (subjéitiva ou não) de que dispomos antes de observar a série  $y_t$  a partir de  $t=1$ .

$D_t$ : informação disponível no tempo  $t$ .

Modelo observacional:  $f(y_t|\theta_t)$ .

A cada instante  $t$ , interesse em:

- $p(\theta_t|D_{t-1})$  (priori)
- $p(\theta_t|D_t)$  (posteriori)
- $f(y_{t+h}|D_t)$ ,  $h > 0$  (preditiva  $h$  passos à frente).

# Modelo dinâmico linear normal

$$\begin{aligned}
 Y_t &= \mathbf{F}'_t \theta_t + \nu_t, & \nu_t &\sim N[0, V] \\
 \theta_t &= \mathbf{G}_t \theta_{t-1} + \omega_t, & \omega_t &\sim N[0, V\mathbf{W}_t^*] \\
 (\theta_0 \mid D_0, V) &\sim N[\mathbf{m}_0, V\mathbf{C}_0^*] \\
 (V \mid D_0) &\sim Gl[n_0/2, d_0/2].
 \end{aligned} \tag{1}$$

- $y_t$  variável resposta;
- $\theta_t$  vetor de parâmetros  $n \times 1$ ;
- $\mathbf{F}_t$  vetor conhecido  $n \times 1$ ;
- $\mathbf{G}_t$  matriz conhecida  $n \times n$ ;
- $V$  variância observacional;
- $\mathbf{W}_t = V\mathbf{W}_t^*$  matriz de covariâncias  $n \times n$ ;

# Modelo dinâmico linear normal

Modelo definido por  $\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t$ . Pode ser escrito como:

$$(\mathbf{y}_t | \theta_t) \sim N[\mu_t, \mathbf{V}_t], \mu = \mathbf{F}'_t \theta_t$$

$$(\theta_t | \theta_{t-1}) \sim N[\mathbf{G}_t \theta_{t-1}, \mathbf{W}_t].$$

Casos particulares:

(i) Modelo de 1a. Ordem :

$$F_t = 1, G_t = 1, \theta_t = \mu_t$$

(ii) Modelo de regressão dinâmica através da origem com  $n$  regressores :

$\mathbf{F}'_t = (x_{1,t}, \dots, x_{n,t})$ ,  $\mathbf{G} = \mathbf{I}_n$ ,  $\theta'_t = (\beta_{t,1} \dots, \beta_{t,n})$  onde  $\mathbf{I}_n$  é matriz identidade de ordem  $n$ .

# Ciclo inferencial sequencial (Filtro de Kalman)

- Para a precisão  $\tau = V^{-1}$  :

$$(\tau \mid D_{t-1}) \sim G(n_{t-1}/2, d_{t-1}/2),$$

$$(\tau \mid D_t) \sim G(n_t/2, d_t/2),$$

$$\text{onde : } n_t = n_{t-1} + 1 \quad \text{e} \quad d_t = d_{t-1} + e_t^2 Q_t^{*-1}$$

## Ciclo inferencial sequencial (Filtro de Kalman)

- para os estados e observações futuras :

$$(\theta_{t-1} | D_{t-1}) \sim t_{n_{t-1}}[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}],$$

$$(\theta_t | D_{t-1}) \sim t_{n_{t-1}}[\mathbf{a}_t, \mathbf{R}_t],$$

$$(Y_t | D_{t-1}) \sim t_{n_{t-1}}[f_t, Q_t],$$

$$(\theta_t | D_t) \sim t_{n_t}[\mathbf{m}_t, \mathbf{C}_t],$$

onde :

$$\mathbf{C}_{t-1} = S_{t-1} \mathbf{C}_{t-1}^* \quad , \quad \mathbf{R}_t = S_{t-1} \mathbf{R}_t^* \quad , \quad Q_t = S_{t-1} Q_t^*$$

$$\mathbf{C}_t = S_t \mathbf{C}_t^* \quad , \quad S_{t-1} = d_{t-1}/n_{t-1} \quad e \quad S_t = d_t/n_t$$

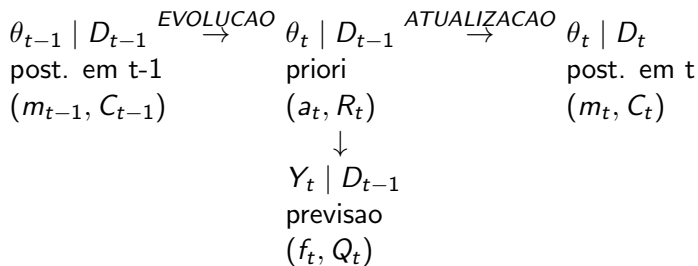
$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}, \quad \mathbf{R}_t^* = \mathbf{G}_t \mathbf{C}_{t-1}^* \mathbf{G}_t' + \mathbf{W}_t^*$$

$$f_t = \mathbf{F}_t' \mathbf{a}_t, \quad Q_t^* = 1 + \mathbf{F}_t' \mathbf{R}_t^* \mathbf{F}_t$$

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t, \quad \mathbf{C}_t^* = \mathbf{R}_t^* - \mathbf{A}_t \mathbf{A}_t' Q_t^*$$

$$\mathbf{A}_t = \mathbf{R}_t^* \mathbf{F}_t Q_t^{*-1}, \quad e_t = Y_t - f_t$$

# Esquema sequencial de inferência





## Porém...

Deseja-se, por exemplo, modelar contagens epidemiológicas ou níveis de certo poluente, com distribuição assimétrica.

Em particular: óbitos diários de crianças menores de 5 anos, por doenças respiratórias em SP e níveis de material particulado no RJ.

Modelo normal é inadequado.

# Modelo Dinâmico Generalizado

Definido (West, Harrison, Migon, 1985) por:

$$\begin{aligned}
 Y_t &\sim \mathcal{F}(\chi_t) \\
 g(\mu_t) &= \eta_t = \mathbf{F}'_t \boldsymbol{\theta}_t \\
 \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N[0, \mathbf{W}_t]
 \end{aligned}$$

em que :

- $\mathcal{F}$  é uma distribuição pertencente à família exponencial e parâmetro natural  $\chi_t$ , isto é:  $p(y_t|\chi_t) \propto \exp\left\{\frac{y_t \chi_t - b(\chi_t)}{\gamma_t}\right\}$ ,  $t = 1, \dots, T$ ,
- $g$  é uma função de ligação monótona e diferenciável;
- $\gamma_t$  é um parâmetro de escala;
- $\mu_t$  é a resposta média, dada por  $\mu_t = E[Y_t|\chi_t] = \dot{b}(\chi_t) = \frac{db(\chi_t)}{d\chi_t}$ .

## Tratamento do efeito cumulativo de uma regressora

Assuma que o efeito de certa regressora sobre a resposta média não seja apenas imediato, mas se propague no tempo.

Admita que um dos componentes de  $\theta_t$  seja  $E_t$ , que denota o efeito cumulativo de uma regressora  $X$  sobre a resposta esperada. Forma mais simples possível para  $E_t$  são defasagens distribuídas:

$$E_t = \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_s X_{t-s}.$$

com  $(\beta_0, \dots, \beta_s)$  geralmente seguindo alguma restrição, por exemplo, polinomial (Almon, 1965):

$$\beta_j = \sum_{k=0}^d \zeta_k j^k, \quad j=0, \dots, s, \quad d \ll s.$$

$$\Rightarrow E_t = \zeta_0 S_{t_0} + \zeta_1 S_{t_1} + \zeta_2 S_{t_2} + \dots + \zeta_d S_{t_d}.$$

# Aproximação polinomial para os efeitos defasados $\beta_j$

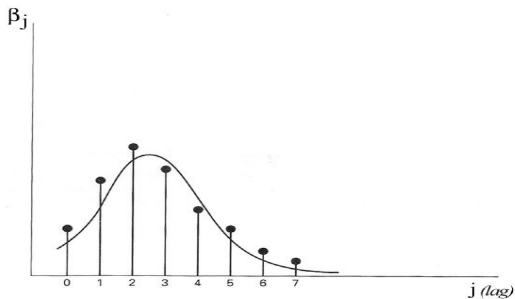


Figure: Aproximação para a trajetória de  $\beta_j, j = 1, 2, \dots, s$  por um polinômio nas defasagens  $1, 2, \dots, s$ .

# Algumas questões sobre modelos de defasagens distribuídas

- O horizonte de influência ( $s$ ) da regressora sobre a resposta deve ser especificado (assume-se efeito nulo além desse horizonte);
- a escolha do grau  $d$  do polinômio suavizador impõe a forma de propagação.

## Funções de transferência

\*Trabalho conjunto com Gamerman e Ferreira (Alves, Gamerman, Ferreira, 2010).

Assumimos que o efeito cumulativo de  $X$  sobre a resposta média possa ser expresso por:

$$\begin{aligned}
 E_t &= \frac{(\omega_0 B^b - \dots - \omega_s B^{b+s})}{(1 - \phi_1 B - \dots - \phi_r B^r)} X_t = \frac{\beta(B)}{\phi(B)} X_t \\
 &= \nu_0 X_t + \nu_1 X_{t-1} + \nu_2 X_{t-2} + \dots \\
 &= \nu(B) X_t,
 \end{aligned} \tag{2}$$

$B$  denota o operador defasagem:  $BX_t = X_{t-1}$ . Os pesos  $\nu_0, \nu_1, \nu_2, \dots$  são denominados **função de resposta ao impulso** e representam o efeito corrente of  $X$  e cada um de seus efeitos passados sobre a resposta atual. A soma desses pesos,  $\nu(B) = \nu_0 + \nu_1 B + \nu_2 B^2 + \dots$ , é denominada **função de transferência** e representa o impacto cumulativo de  $X$  sobre a resposta.

$$\text{TF}(1,0,0) \rightarrow E_t = \phi E_{t-1} + \beta X_t$$

Na forma (2), TF(1,0,0) se resume a:

$$E_t = \beta X_t + \phi \beta X_{t-1} + \phi^2 \beta X_{t-2} + \dots$$

Assuma  $\beta > 0$ . Então a função de resposta ao impulso segue os seguintes padrões, dependendo de  $\phi$ :

- $0 < \phi < 1$ : decaimento geométrico
- $-1 < \phi < 0$ : decaimento geométrico com alternância de sinal
- $\phi > 1$ : crescimento geométrico
- $\phi < -1$ : crescimento geométrico com alternância de sinal

# TF(1,0,0): Função de resposta ao impulso e de transferência

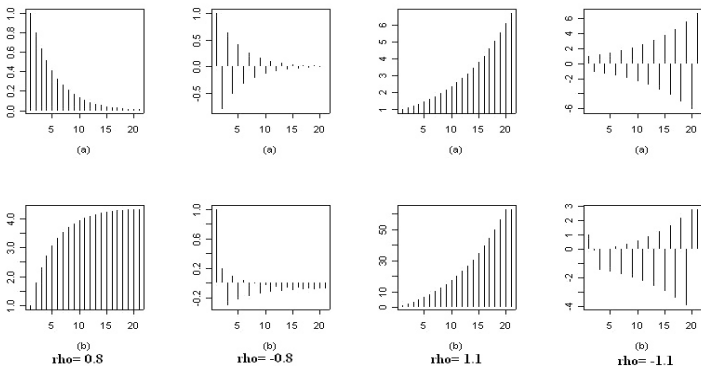


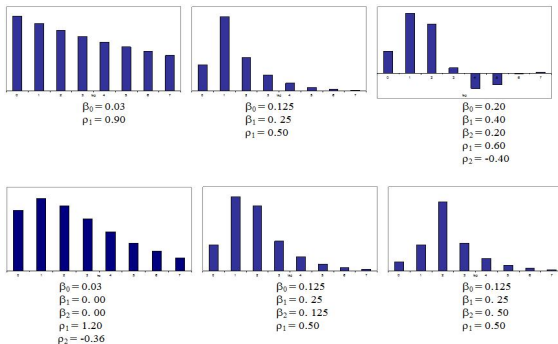
Figure: Impulse response (a) and transfer functions (b) associated to the first order model for different values of  $\phi$ .



$$TF(2, 0, 2): E_t = \phi_1 E_{t-1} + \phi_2 E_{t-2} + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2}$$

$$\leftrightarrow (1 - \gamma_1 B)(1 - \gamma_2 B)E_t = \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2}$$

Condição de estacionariedade:  $|\gamma_j| < 1, j = 1, \dots, r$ .



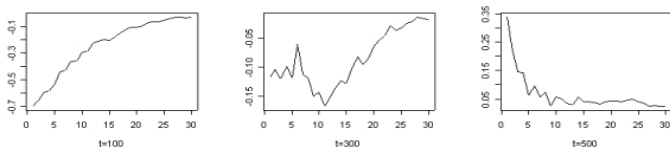
**Figure:** Examples of impulse response functions associated to the model  $TF(r,b,s)$ ,  $1 \leq r \leq 2, b = 0, 0 \leq s \leq 2$ .

# TF(1,0,0) com fatores de ganho dinâmicos

- TF (1,0,0) com fatores de ganho dinâmicos:

$$E_t = \phi E_{t-1} + \psi_t X_t$$

$$\psi_t = \psi_{t-1} + v_t, \quad v_t \sim N(0, Q). \quad (3)$$



**Figure:** Função de resposta ao impulso com fatores de ganho seguindo passeio aleatório com  $\phi = 0.9$  e variância evolutiva  $Q = 0.005$ , nos instantes 100, 300 and 500.

MDG com função de transferência  $TF(p,b,0)$ 

$$p(y_t | \eta_t) \sim \mathcal{F}(\chi_t)$$

$$\eta_t = g(\mu_t) = \tilde{\mathbf{F}}_t' \tilde{\boldsymbol{\theta}}_t$$

$$\tilde{\boldsymbol{\theta}}_t = \tilde{\mathbf{G}}(\phi) \tilde{\boldsymbol{\theta}}_{t-1} + \boldsymbol{\psi}_{t-1} X_{t-b} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}_t)$$

$$\boldsymbol{\psi}_t = \boldsymbol{\psi}_{t-1} + \boldsymbol{\nu}_t \quad \boldsymbol{\nu}_t \sim N(\mathbf{0}, \mathbf{W}_{\nu t}),$$

com  $\boldsymbol{\psi}'_t = (\psi_t, 0, \dots, 0)$ ,  $\tilde{\boldsymbol{\theta}}'_t = (E_t^{(0)}, E_t^{(1)}, \dots, E_t^{(p-1)})$ ,  $E_t^{(0)} = E_t$ ,  
 $E_t^{(i)} = E_{t-1}^{(i-1)}$ ,  $i = 1, \dots, p-1$ .

$$\tilde{\mathbf{G}}_t(\phi) = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

## Modelo na forma de espaço de estados

Para obter o modelo na forma padrão de espaço de estados, basta tomar:

$$\theta'_{*t} = (\tilde{\theta}'_t, \psi'_t)$$

$$\mathbf{F}'_{*t} = (\tilde{\mathbf{F}}'_t, 0, \dots, 0)$$

$$\mathbf{G}_t = \begin{bmatrix} \tilde{\mathbf{G}}_t & X_t \mathbf{I}_p \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix}.$$

## Dificuldade para tratamento analítico

Pode-se adotar priori conjugada para  $\mu_t$ ,  $p(\mu_t|D_{t-1})$ , obtendo-se forma analítica fechada para:

- distribuição a posteriori de  $\mu_t$ :  $p(\mu_t|D_t)$ ;
- preditiva  $p(y_{t+h}|D_t)$ ,  $h > 0$ .

Entretanto, não há forma analítica fechada para as distribuições a posteriori:

- dos estados  $\theta_t$ ,  $p(\theta_t|D_t)$ ;
- das variâncias de evolução  $\mathbf{W}_t$ ,  $p(\mathbf{W}_t|D_t)$ ;
- dos parâmetros autorregressivos  $\phi_1, \dots, \phi_r$  e demais parâmetros envolvidos nas funções de transferência.

Alternativa: Aproximação da posteriori.

# Método de Monte Carlo via Cadeias de Markov - MCMC

Ideia é construir uma cadeia de Markov ergódica tendo como espaço de estados o espaço paramétrico e que tenha a distribuição a posteriori como distribuição estacionária.

Algoritmos:

**Amostrador de Gibbs:** apresentado por Geman e Geman (1984) - processamento de imagens e interesse na determinação da constante de normalização da distribuição de Gibbs. Gelfand e Smith (1990): potencial do amostrador de Gibbs para determinação de constantes de normalização em outras distribuições e difusão do algoritmo na comunidade estatística.

**Metropolis-Hastings:** deve-se aos trabalhos de Metropolis et al. (1953) - no contexto de definição de propriedades químicas de substâncias - e Hastings (1970), que generalizou o algoritmo de Metropolis para utilização no contexto estatístico.

Seja uma C.M. irredutível com função de transição  $q$ .

Considere a distribuição a posteriori  $\pi^*(\theta) = \pi(\theta | y)$  com núcleo  $n^*(\theta)$  e a C.M. com função de transição  $p$  dada por:

$$p(\theta_1, \theta_2) = q(\theta_1, \theta_2)\alpha(\theta_1, \theta_2),$$

$$\alpha(\theta_1, \theta_2) = \min \left\{ 1, \frac{\pi^*(\theta_2)q(\theta_2, \theta_1)}{\pi^*(\theta_1)q(\theta_1, \theta_2)} \right\} = \min \left\{ 1, \frac{n^*(\theta_2)q(\theta_2, \theta_1)}{n^*(\theta_1)q(\theta_1, \theta_2)} \right\}.$$

A função  $\alpha$  garante que  $p$  e  $\pi^*$  satisfaçam a condição de reversibilidade

$$\pi^*(\theta_1)p(\theta_1, \theta_2) = \pi^*(\theta_2)p(\theta_2, \theta_1), \quad \forall \theta_1, \theta_2 \in \Theta,$$

donde  $\pi^*$  é a distribuição invariante da cadeia com transição  $p$ .

# Simulação da cadeia

- 1 Escolhe-se um valor inicial  $\theta^{(0)} \in \Theta$  ( $i=0$ );
- 2 Dado  $\theta^{(i)}$ , amostra-se um valor  $\theta^*$  de  $q(\theta^{(i)}, \bullet)$ ;
- 3 Com probabilidade  $\alpha(\theta^{(i)}, \theta^*)$ , fazemos  $\theta^{(i+1)} = \theta^*$  e com probabilidade  $1 - \alpha(\theta^{(i)}, \theta^*)$  fazemos  $\theta^{(i+1)} = \theta^{(i)}$ .



# Amostrador de Gibbs

Sejam  $\theta = (\theta_1, \dots, \theta_n)$  e as  $n$  condicionais completas  $(\theta_1 | \theta_2, \dots, \theta_n), \dots, (\theta_n | \theta_1, \dots, \theta_{n-1})$ .

Algoritmo:

- 1 Escolha  $n$  valores iniciais  $\theta_1^{(0)}, \dots, \theta_n^{(0)}$  ( $i=0$ )
- 2 Sorteia-se um valor  $\theta_1^{(i+1)}$  de  $\theta_1 | \theta_2^{(i)}, \dots, \theta_n^{(i)}$ ;  
 Sorteia-se um valor  $\theta_2^{(i+1)}$  de  $\theta_2 | \theta_1^{(i+1)}, \theta_3^{(i)}, \dots, \theta_n^{(i)}$ ;  
 $\vdots$   
 Sorteia-se um valor  $\theta_n^{(i+1)}$  de  $\theta_n | \theta_1^{(i+1)}, \dots, \theta_{n-1}^{(i+1)}$ ;
- 3 Repete-se a etapa anterior para as iterações seguintes ( $i=1,2,3,\dots$ )

Obs.: Na etapa 2, aplica-se Metropolis-Hastings para cada condicional completa cuja forma analítica seja desconhecida.

# MCMC para modelos de espaço de estados

Em modelos com coeficientes dinâmicos, convergência das cadeias de Markov para a distribuição estacionária é dificultada por:

- dimensão do vetor paramétrico;
- correlação entre componentes do vetor paramétrico.

Diversas propostas buscando eficiência nesse contexto:

Frühwirth-Schnatter (1994);

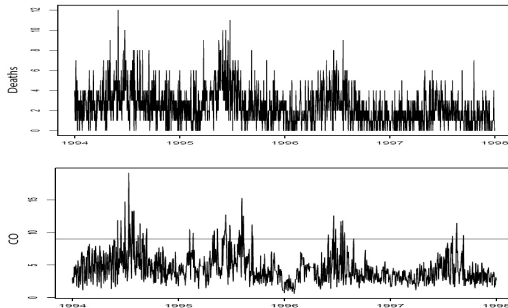
Carter & Kohn (1994);

Gamerman(1998);

Migon, Schmidt, Ravines, Pereira (2013).

Ainda assim, tempo computacional pode ser elevado.

# Application 1: Effect of Carbon Monoxide over under 5 aged childrens' counts of deaths in São Paulo, Brazil



**Figure:** Daily records of children deaths due to respiratory diseases (top) and levels of Carbon Monoxide in São Paulo, Brazil - Jan-1994 to Dec-1997

# Application 1: Meteorological variables

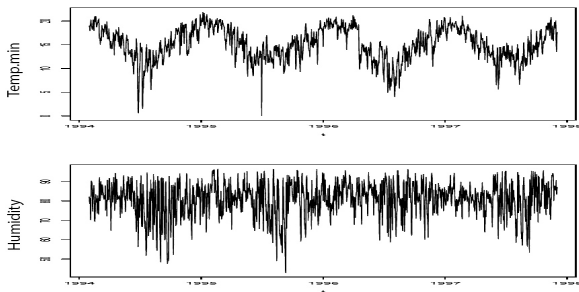


Figure: Daily records of minimum temperatures (top) and mean humidity levels

# Application 1: Chosen model

$$\text{Deaths}_t \sim \text{Poisson}(\lambda_t)$$

$$\begin{aligned} \log(\lambda_t) = \eta_t = & \alpha + E_t + \delta_1 \text{Temp}_t + \delta_2 \text{Humid}_t \\ & + \delta_3 \text{Cos} \left( \frac{2\pi t}{365} \right) + \delta_4 \text{Sen} \left( \frac{2\pi t}{365} \right) \\ & + \delta_5 \text{Cos} \left( \frac{4\pi t}{365} \right) + \delta_6 \text{Sen} \left( \frac{4\pi t}{365} \right) \end{aligned}$$

$$E_t = \phi E_{t-1} + \beta_t \text{CO}_t$$

$$\beta_t = \beta_{t-1} + v_t \quad v_t \sim N(0, Q)$$

## Application 1 - Chosen model: Model 11 - Posterior summary

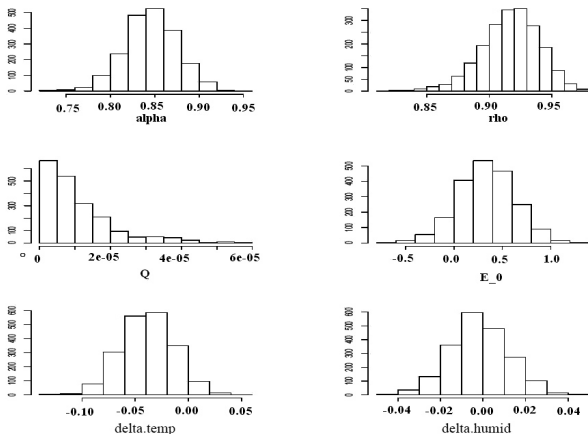


Figure: Histograms of the posterior distribution samples in model 11.



## Application 1 - Chosen model: Model 11 - Posterior summary (cont.)

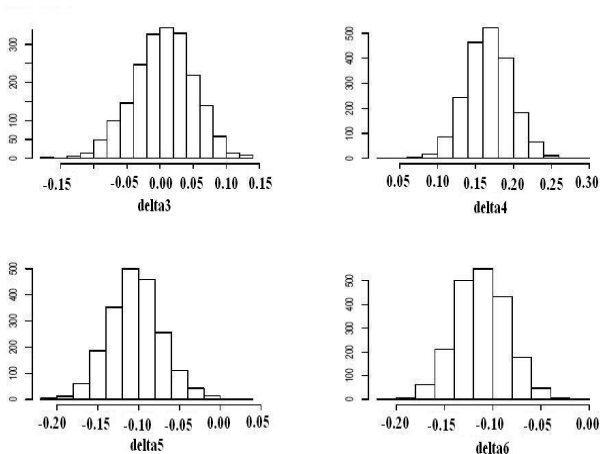
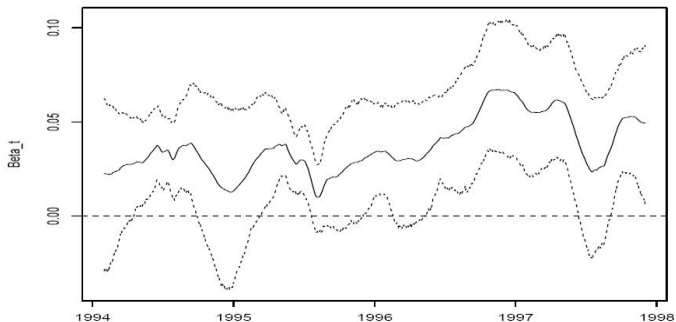


Figure: Histograms of the posterior distribution samples in model 11.

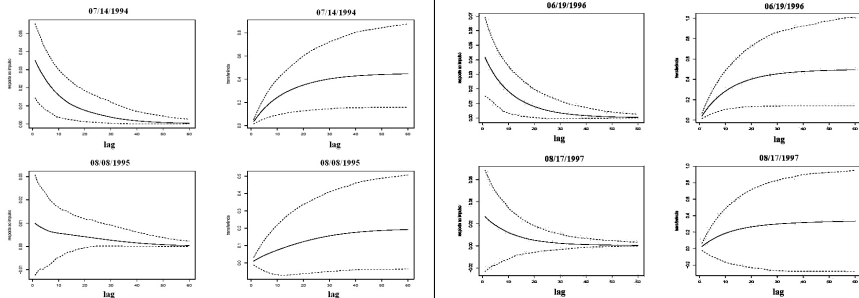


Application 1 - Model 11 - TF(1,0,0) - Estimation of the dynamic  $\beta_t$ 

**Figure:** Model 11 - Estimated trajectory of the dynamic gain factor  $\beta_t$  (the dynamic immediate impact of Carbon Monoxide over daily mean counts of deaths).

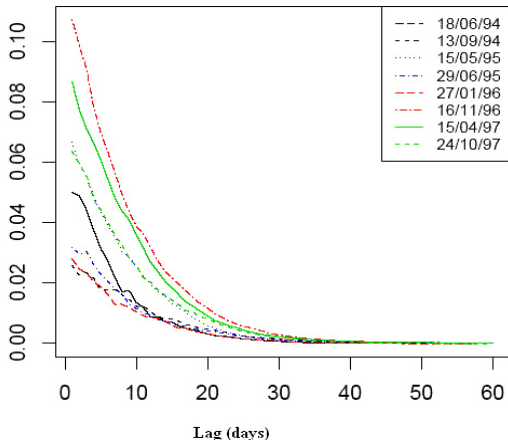


## Model 11 - Impulse and transfer functions associated to 1 s. d. raise on CO

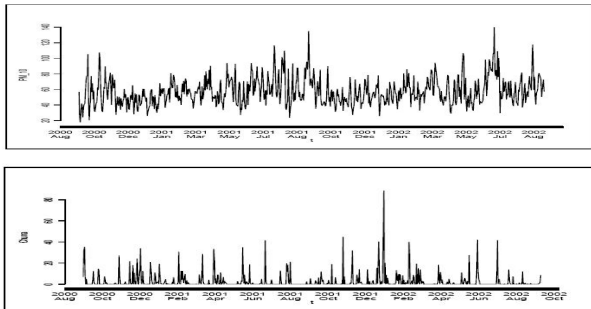


**Figure:** *Estimated impulse response and transfer functions associated to 1 standard error elevation on the pollutant levels at 4 different moments.*

Model 11 - Relative risks associated to a 12 ppm level of CO (maximum level at the period 1996-1997) as compared to the mean observed level(4 ppm)



# Application 2: Effect of rain volumes over $PM_{10}$ levels in Rio de Janeiro, Brazil



**Figure:** Daily records of particulate matter (top) and rain volumes in Rio de Janeiro, Brazil - Sep-2000 to Aug-2002

## Application 2: Meteorological variables

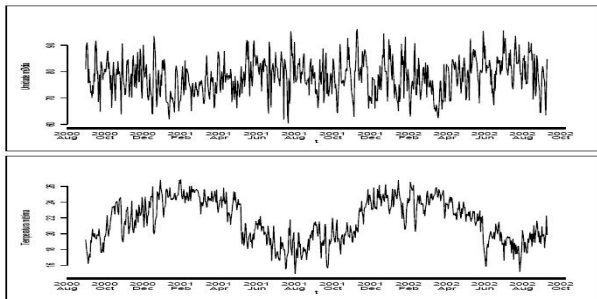


Figure: Daily records of mean humidity (top) and minimum temperature levels

## Application 2: General Structure of the proposed models

$$PM_t \sim \text{Gamma}(\varphi, \lambda_t)$$

$$\log\left(\frac{\varphi}{\lambda_t}\right) = \eta_t = \alpha + E_t + \delta'_C \mathbf{C}_t + \delta'_D \mathbf{D}_t + \delta'_S \mathbf{S}_t,$$

- $E_t$  is a structural block denoting the cumulative effect of rain volumes at time  $t$ , expressed through different transfer function proposals;
- $\mathbf{C}'_t = (\text{Humidity}_t, \text{Temperature}_t)$ ;
- $\mathbf{D}'_t = (\text{Mon}_t, \text{Tue}_t, \text{Wed}_t, \text{Thu}_t, \text{Fri}_t, \text{Sat}_t, \text{Hol}_t)$  is a vector of dummy calendar variables;
- $\mathbf{S}'_t = (\text{Cos}(\frac{2\pi t}{365}), \text{Sin}(\frac{2\pi t}{365}))$ .

# Effect of rain volumes over PM<sub>10</sub>: Comparison of models

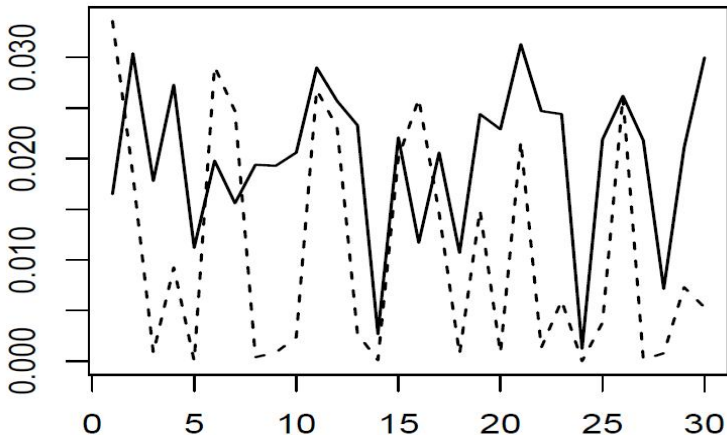
Table 1: Comparison of models fitted to daily PM<sub>10</sub>, considering cumulative impact of rain volumes.

Predictor specification	Transfer Function	Predictive Likelihood
1: $(r = 1, s = 0), \delta_S = 0$	$E_t = \rho E_{t-1} + \beta \text{Rain}_t$	$2.24 \times 10^{-54}$
2: $(r = 1, s = 1), \delta_S = 0$	$E_t = \rho E_{t-1} + \beta_0 \text{Rain}_t + \beta_1 \text{Rain}_{t-1}$	$1.35 \times 10^{-54}$
3: $(r = 1, s = 0), \delta_S \neq 0$	$E_t = \rho E_{t-1} + \beta \text{Rain}_t$	$1.05 \times 10^{-53}$
4: $(r = 1, s = 1), \delta_S \neq 0$	$E_t = \rho E_{t-1} + \beta_0 \text{Rain}_t + \beta_1 \text{Rain}_{t-1}$	$8.65 \times 10^{-54}$
5: $(r = 0, s = 30, d = 2), \delta_S = 0$	$E_t = \sum_{j=0}^{30} \beta_j \text{Rain}_{t-j}$ $\beta_j = \sum_{k=0}^2 \zeta_k j^k$	$8.84 \times 10^{-55}$
6: $(r = 0, s = 30, d = 3), \delta_S = 0$	$E_t = \sum_{j=0}^{30} \beta_j \text{Rain}_{t-j}$ $\beta_j = \sum_{k=0}^3 \zeta_k j^k$	$1.54 \times 10^{-54}$
7: $(r = 1, s = 0), \delta_S = 0$ iid errors	$E_t = \rho E_{t-1} + \beta \text{Rain}_t + \varepsilon_t$ $\varepsilon_t \sim N(0, Q_\varepsilon)$	$2.89 \times 10^{-54}$

$r$ =autoregression order,  $s$ =number of lags in rain volume,  $d$ =degree of polynomial approximation to  $\beta_j$ ,  $j = 1, \dots, s$ . The reported predictive likelihoods are based on a horizon of 30 days.

Distributed lag model (large  $s$ ):  $\text{PredLik} = 6.73 \times 10^{-64}$

## Application 2 - Chosen model: Model 3 - Predictive comparison



## Application 2 - Chosen model: Model 3 - Posterior summary

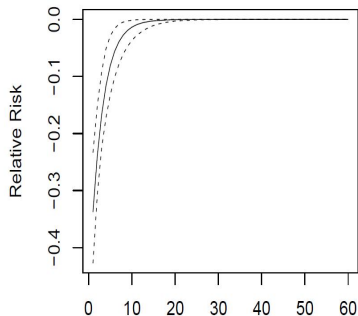
	Min	Q1	Median	Mean	Q3	Max.	MC error
$\varphi$	15.89	19.11	19.88	19.87	20.62	22.82	0.0147
$\alpha$	3.12	3.55	3.70	3.69	3.84	4.26	0.0685
$E_0$	0.04	0.65	0.82	0.85	1.03	2.34	0.0188
$\rho$	0.411	0.629	0.672	0.667	0.712	0.835	0.0052
$\beta$	-0.054	-0.038	-0.034	-0.034	-0.030	-0.015	0.0005
$\delta.Humid$	-0.093	-0.070	-0.062	-0.063	-0.057	-0.029	0.0016
$\delta.Temp$	0.074	0.104	0.121	0.121	0.138	0.170	0.0064
$\delta.cos$	-0.001	0.027	0.041	0.041	0.053	0.108	0.0052
$\delta.sin$	-0.317	-0.259	-0.240	-0.238	-0.211	-0.175	0.0086
$\delta.Mon$	0.010	0.079	0.102	0.103	0.127	0.194	0.0008
$\delta.Tue$	0.060	0.131	0.154	0.152	0.174	0.244	0.0008
$\delta.Wed$	0.077	0.160	0.182	0.180	0.201	0.295	0.0008
$\delta.Thu$	0.089	0.175	0.198	0.198	0.218	0.308	0.0008
$\delta.Fri$	0.120	0.196	0.217	0.217	0.238	0.314	0.0008
$\delta.Sat$	0.008	0.081	0.103	0.104	0.125	0.211	0.0008
$\delta.Hol$	-0.221	-0.108	-0.077	-0.077	-0.047	0.056	0.0011
$y_{478}$	21.24	39.39	45.28	46.00	51.94	87.61	0.2423
$y_{479}$	22.86	45.75	52.76	53.05	59.33	97.03	0.2828



## Application 2 - Chosen model: Model 3 - Posterior summary (cont.)

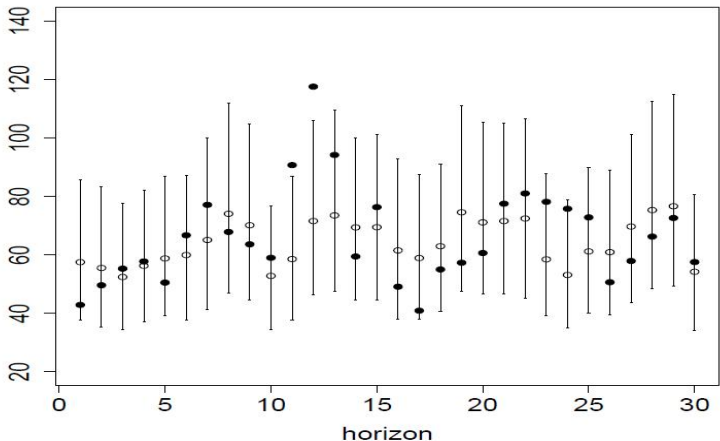
- We notice a growth trend of the pollutant levels towards Fridays, maybe due to the great volume of traffic usual in Rio de Janeiro on Fridays. Such levels decay on Saturdays.
- The pollutant levels are also significantly reduced on holidays.
- Temperature presents a significant, positive impact on  $PM_{10}$  levels according to this model and humidity presents negative effect. We are currently working on models with transfer function structure assigned to the climatic variables.
- The negative estimate of  $\beta$  indicates lower levels of  $PM_{10}$  on rainy days.
- The uncertainty due to 2 missing observations has been formally treated.

## Model 3 - Relative risks associated to the maximum observed rain volume as compared to non-rainy days



**Figure:** Relative risk (percentual decay on the expected level of  $PM_{10}$ ) associated to the maximum rain volume observed during the analysis period.

## Model 3 - Predictive distributions up to 30 days ahead



# MCMC: prós e contras

## Prós:

- inferência bayesiana completa para qualquer função  $g$  dos parâmetros do modelo
- Teoricamente aplicável a qualquer modelo. Não necessitaríamos nos restringir a equações observacionais na família exponencial

## Contras:

- custo computacional pode ser elevado em modelos que tenham parâmetros com dinâmica;
- perda do aspecto sequencial na atualização de informação sobre os estados.

# Método sequencial com expansão do vetor de estados

\*com Carlos Tadeu Pagani Zanini, Hélio Migon

Considere novamente o modelo

$$p(y_t | \eta_t) \sim \mathcal{F}(\chi_t)$$

$$\eta_t = g(\mu_t) = \mathbf{F}'_{*t} \boldsymbol{\theta}_{*t}$$

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

com

$$\boldsymbol{\theta}'_{*t} = (\tilde{\boldsymbol{\theta}}'_t, \boldsymbol{\psi}'_t),$$

$$\mathbf{F}'_{*t} = (\tilde{\mathbf{F}}'_t, 0, \dots, 0),$$

$$\mathbf{G}_t = \begin{bmatrix} \tilde{\mathbf{G}}_t & \chi_t \mathbf{I}_p \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix}.$$

## Expansão do vetor de estados

Incorporaremos a estimação dos parâmetros autoregressivos  $\phi_1, \dots, \phi_p$  ao vetor de estados, expandindo-o. Defina agora:

$$\theta'_t = (\tilde{\theta}'_t, \psi_t, \phi_1, \dots, \phi_p), \text{ com}$$

$$\tilde{\theta}'_t = (E_t^{(0)}, E_t^{(1)}, \dots, E_t^{(p-1)}), \quad E_t^{(0)} = E_t, \quad E_t^{(i)} = E_{t-1}^{(i-1)}, \quad i = 1, \dots, p-1;$$

$$F_t = (1, 0, \dots, 0)$$

## Linear Bayes

Se a distribuição de  $\theta_t$  é parcialmente especificada, em termos de primeiro e segundo momentos, para qualquer função de perda, não é possível obter estimador de Bayes para  $\theta_t$ .

Porém, se assumirmos perda quadrática e estimadores para  $\theta_t$  que sejam restritos à classe das funções lineares de  $Y$ , então o estimador de Bayes é a média de  $\theta_t$  e o risco de Bayes é o traço da matriz de covariâncias de  $\theta_t$ .

Mas  $\theta_t = f(\theta_{t-1}) + \varepsilon_t$ ,  $\varepsilon_t \sim [0, \mathbf{W}_t]$ ,  $f$  uma função não linear e  $\mathbf{W}_t = \text{diag}(0, \dots, 0, W_t, 0, \dots, 0)$

# Linearização da equação de evolução

Expansão de Taylor de ordem 1 para  $f$  centrada em  $m_{t-1} = E(\theta_{t-1} | D_{t-1})$ :

$$\begin{aligned}\theta_t &\approx f(m_{t-1}) + f'(m_{t-1})(\theta_{t-1} - m_{t-1}) + \varepsilon_t \\ &= h_t + H_t \theta_{t-1} + \varepsilon_t,\end{aligned}\tag{4}$$

onde  $h_t = f(m_{t-1}) - f'(m_{t-1})m_{t-1}$  e  $H_t = f'(m_{t-1})$ .

► MLDG



## Momentos dos estados, após linearização

$$\begin{aligned}a_t &= E[\theta_t | D_{t-1}] = E[h_t + H_t \theta_{t-1} + \omega_t | D_{t-1}] \\&= h_t + H_t m_{t-1} \\&= f(m_{t-1}) - f'(m_{t-1}) m_{t-1} + f'(m_{t-1}) m_{t-1} \\&= f(m_{t-1}) \\R_t &= V[\theta_t | D_{t-1}] = V[h_t + H_t \theta_{t-1} + \omega_t | D_{t-1}] \\&= H_t C_{t-1} H_t' + \mathbf{W}_t\end{aligned}$$

$\mathbf{W}_t$  tem uma componente não nula, especificada por meio de fator de desconto.

# Procedimento inferencial: Linear Bayes (simplificado)

Ligação a priori: estados, preditor e média

$$\theta_t \mid D_{t-1} \sim [a_t, R_t] \rightarrow \eta_t \mid D_{t-1} \sim [f_t, q_t]$$

$$\mu_t \mid D_{t-1} \sim \text{Priori Conj.}(r_t, s_t) \xrightarrow{g(\mu_t)=\eta_t} \eta_t \mid D_{t-1} \sim [h_1(r_t, s_t), h_2(r_t, s_t)]$$

Compatibilização dos momentos

$$\begin{cases} f_t = h_1(r_t, s_t), \\ q_t = h_2(r_t, s_t), \end{cases}$$

Conjugação

$$\mu_t \mid D_{t-1} \sim \text{Priori Conj.}(r_t, s_t) \Rightarrow \begin{cases} \mu_t \mid D_t \sim \text{Posteriori Conjugada}(r_t^*, s_t^*), \\ y_t \mid D_{t-1} \text{ forma analítica conhecida.} \end{cases}$$

# Procedimento inferencial: Linear Bayes (cont.)

Preditor a posteriori:

$$\eta_t = g(\mu_t) \mid D_t \sim [f_t^*, q_t^*] \quad \begin{cases} f_t^* = h_1(r_t^*, s_t^*), \\ q_t^* = h_2(r_t^*, s_t^*), \end{cases}$$

Posteriori dos estados (tempo  $t$ ):

$$\theta_t \mid D_t \sim [m_t, C_t] \quad \begin{cases} m_t = a_t + R_t F_t (f_t^* - f_t) / q_t, \\ C_t = R_t - R_t F_t F_t' (1 - q^* / q_t) / q_t. \end{cases}$$

## Método baseado em LB

- Alguma restrição sobre a forma do modelo observacional (membro da família exponencial)
- estados e parâmetros envolvidos na função de transferência especificados apenas em termos de primeiro e segundo momentos.

Em contrapartida:

- Recuperação do aspecto sequencial do processo de aprendizagem sobre os estados
- tempo computacional desprezível: 0.2 segundos, com processador I5, RAM 4GB.

## Comparação aos resultados MCMC - efeitos estáticos

Para a aplicação de efeito de poluentes sobre órbitos, obtivemos as seguintes estimativas pontuais:

	LB	MCMC
nivel	0.885	0.844
phi	0.855	0.916
temp	-0.034	-0.040

# Comparação aos resultados MCMC - efeitos dinâmicos

