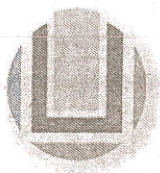


Carlos Dullius

**Sistemas Dinâmicos Bayesianos com INLA
Aplicados a Séries Temporais Incompletas de
Preço de Pescado**

Brasil

2017



Universidade Federal do Rio Grande – FURG

Instituto de Matemática, Estatística e Física

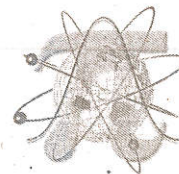
Curso de Bacharelado em Matemática Aplicada

Av. Itália km 8 Bairro Carreiros

Rio Grande-RS CEP: 96.201-900 Fone (53)3233.5411

e-mail: imef@furg.br

Sítio: www.imef.furg.br



Ata de Defesa de Monografia

Aos sete dias do mês de dezembro de 2017, às 17h30min, na sala 4110, no Campus Carreiros, foi realizada a apresentação pública da defesa do Trabalho de Conclusão de Curso pelo acadêmico **Carlos Edgar Prill Dullius**, sob a orientação do Prof. Dr. Paul Gerhard Kinas, deste instituto, e intitulada **Sistemas Dinâmicos Bayesianos com INLA, aplicados a séries temporais incompletas de preço de pescado**. Para participar da banca avaliadora, junto ao orientador, foram convidados a Prof.^a Dr.^a Raquel da Fontoura Nicolette, docente do IMEF/FURG, e o Prof. Dr. Juliano César Marangoni, docente do IMEF/FURG. Concluídos os trabalhos de apresentação e arguição, o candidato foi: (X) aprovado por unanimidade; () aprovado somente após satisfazer as exigências que constam na folha de modificações, no prazo fixado pela banca; () reprovado. Na forma regulamentar, foi lavrada a presente ata, que é abaixo assinada pelos membros da banca, na ordem acima relacionada.

Prof. Dr. Paul Gerhard Kinas
Orientador

Prof.ª Dr.ª Raquel da Fontoura Nicolette
Membro

Prof. Dr. Juliano César Marangoni
Membro

Agradecimentos

Agradeço a minha namorada Gabriela Jardim pelo amor e paciência; ao Paul Gerhard Kinas pelas correções; ao Juliano César Marangoni pelas horas de conversas e dicas; a Raquel da Fontoura Nicolette pelos conselhos; a todos meu amigos e colegas do **LEA** pela companhia e risadas. Muito obrigado.

Resumo

O objetivo do estudo é o ajuste de modelos dinâmicos estocásticos a séries temporais com dados ausentes que, por esta característica, não permitem o ajuste por modelos frequentistas convencionais. Foi usada uma análise *Bayesiana* em que as distribuições posteriores marginais das estimativas foram obtidas com o método de aproximação de Laplace integradas e aninhadas (Integrated Nested Laplace Approximation - **INLA**). Selecionou-se o melhor modelo a partir do Critério de Informação da Deviância (Deviance Information Criterium - **DIC**). Como estudo de caso foi utilizada uma série temporal da evolução mensal do preço de varejo do Bagre no município de Rio Grande, ao longo de 11 anos (2006 a 2016). Esta série, contendo 22% dos meses sem informação (dados ausentes), foi disponibilizada pela Unidade de Pesquisa em Economia Costeira e Marinha do Instituto de Ciências Econômicas Administrativas e Contábeis (**ICEAC**) da **FURG**. O melhor modelo selecionado via **DIC** foi o modelo de Markov Oculto (Hidden Markov Model - **HMM**) de ordem 3, **HMM(3)**. A série latente de preços foi ajustada a partir deste modelo, fornecendo estimativas pontuais de preço com os respectivos intervalos de credibilidade de 95% para todos os meses do período analisado.

Palavras-chaves: INLA, Modelo Dinâmico, séries de Tempo, Analise Bayesiana.

Sumário

1	INTRODUÇÃO	5
1.1	Processo estocástico e Séries de Tempo	5
1.2	Modelos Dinâmicos	5
1.3	Revisão dos Modelos	6
1.4	Inferência Bayesiana	7
1.5	Objetivos do Trabalho	8
1.5.1	Objetivos Específicos	8
2	MATERIAL E MÉTODOS	9
2.1	Banco de Dados	9
2.1.1	Procedimento da coleta	9
2.2	Modelos Dinâmicos e Modelo de Markov Oculo	10
2.3	Campos Aleatórios Gaussianos Markovianos	12
2.4	Aproximação de Laplace e Laplace-Simplificada	14
2.5	Integrated Nested Laplace Approximation	15
2.6	Seleção de Modelo	18
2.7	Implementação do RINLA	18
3	APLICAÇÃO E DISCUSSÃO	21
	Conclusão	25
	REFERÊNCIAS	26
	APÊNDICES	27
	APÊNDICE A – SCRIPT COMPLETO R	28

1 Introdução

1.1 Processo estocástico e Séries de Tempo

Um processo controlado por leis de probabilidades, o processo estocástico é uma família Y de funções onde cada par ordenado $(Y(t), \omega)$ segue uma função distribuição de probabilidade (**F.D.P.**). É definido formalmente em [Morettin e Toloi \(2006\)](#) como:

Definição 1. *Seja T um conjunto arbitrário, um processo estocástico é uma família $y = \{Y(t), t \in T\}$, tal que, para cada $t \in T$, $Y(t)$ é uma variável aleatória.*

Dado um certo $\omega \in \Omega$, onde Ω é o espaço amostral do ponto t , então, obtemos uma trajetória $Y^{(\omega)}(t)$, esta função é conhecida como *Trajetoária* ou *Série Temporal*, usaremos a notação \mathbf{y}_t em simplificação a $Y_t = \{y_1, y_2, \dots, y_n\}$, onde n é o número de observações, e y_t quando escalar.

Em séries temporais geralmente o foco principal é a previsão (forecast), onde queremos saber Y_{t+h} para $h, t \in \mathbb{Z}^+$. Como é intuitivo, supõe-se que quanto menor for o h mais precisa será a nossa previsão. Podemos também fazer uma previsão retrocedendo no tempo quando tomamos $\{h \in \mathbb{Z}^-; t \in \mathbb{Z}^+; |h| < t\}$ para preencher lacunas desconhecidas (*missing data*), nesse caso, chamamos o processo de *backcast*. Além de previsões, o processo da série em si é também um objeto de estudo. A decomposição de série em tendência, sazonalidade e ruído pode nos dar muitas informações sobre o seu comportamento, e o modelo ajustado pode nos permitir criar hipóteses sobre o processo.

1.2 Modelos Dinâmicos

Os Modelos Dinâmicos ou *Espaço estado*, são um conjunto de modelos onde o tempo é a variável explicativa dada a forma funcional. Permitindo descrever a evolução do fenômeno em relação aos parâmetros do modelo.

Usaremos um Modelo Dinâmico Linear Gaussiano (Gaussian Linear Dynamic Model - **GLDM**), definido em [West e Harrison \(1989\)](#) como:

$$\text{Equação de Observações} \quad y_t = F'_t x_t + v_t, \quad v_t \stackrel{iid}{\sim} N(0, V_t) \quad (1.1)$$

$$\text{Equação do Processo} \quad x_t = G_t x_{t-1} + w_t, \quad w_t \stackrel{iid}{\sim} N(0, W_t) \quad (1.2)$$

$$(1.3)$$

- i) Onde y_t é um vetor $n \times 1$ dos dados observados, é a variável observada ou estado, explicada ou pelo processo.

- ii) \mathbf{x}_t é um vetor $p \times 1$, uma simplificação para $X_t = \{x_1, x_2, \dots, x_n\}$, do processo de estado no tempo t , parâmetro estocástico, com p sendo o número de defasagens (lag's) do modelo.
- iii) v_t e w_t são os erros, considerados não correlacionados temporalmente nem entre si.
- iv) \mathbf{G}_t é chamado de matriz do sistema e \mathbf{F}'_t a matriz de observações. Onde \mathbf{G}_t é $p \times p$, e \mathbf{F}'_t é $n \times p$ onde $n \geq p$
- v) Sendo onde n numero de coletas por observação e os hiperparâmetro $\boldsymbol{\theta} = \{\mathbf{F}'_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$

1.3 Revisão dos Modelos

Existe uma ampla gama de modelos utilizados para a análise de séries temporais. Dos modelos dinâmicos univariados clássicos, propostos em [E.P.Box e M.Jenkins \(1976\)](#) temos:

- i) O Autorregressivo (*Autoregressive* - **AR**) de ordem h , **AR**(p):

$$\mathbf{AR}(p) \ y_t = \rho_0 + \rho_1 y_{t-1} + \dots + \rho_p y_{t-p} + w_t \quad w_t \stackrel{iid}{\sim} N(0, \sigma^2) \quad (1.4)$$

- ii) O Médias Móveis (*Moving Average* - **MA**) de ordem q , **MA**(q)

$$\mathbf{MA}(q) \ y_t = w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q} + w_t \quad w_t \stackrel{iid}{\sim} N(0, \sigma^2) \quad (1.5)$$

- iii) O Autorregressivo de Médias Móveis (*Autoregressive Moving Average* - **ARMA**) de ordem h e q , **ARMA**(h, q).

$$\mathbf{ARMA}(p, q) \ y_t = \rho_0 + \rho_1 y_{t-1} + \dots + \rho_p y_{t-p} + w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q} \quad w_t \stackrel{iid}{\sim} N(0, \sigma^2) \quad (1.6)$$

onde $\theta_1, \dots, \theta_q, \rho_1, \dots, \rho_p$ são constantes reais, h e q são números inteiros não negativos que indicam a ordem (número de defasagens) dos processos **AR** e **MA**, respectivamente. Esses modelos são estimados pelo método frequentista dos Mínimos Quadrados Ordinários (**MQO**), possuem muitas limitações, dentre elas:

- i) O problema da micro-numerosidade, quando se tem poucas observações e necessita-se estimar um número grande de parâmetros;
- ii) Séries com lacunas, acarretam a impossibilidade de ajustar o modelo inferencial;
- iii) As observações são produzidas pelo processo sem o erro do processo latente.

As limitações (i) e (ii), se devem ao método **MQO**, a limitação (iii) é um erro estrutural dos modelos, por não levar em consideração os erros observacionais v_t .

1.4 Inferência Bayesiana

Na abordagem *Bayesiana* de inferência estatística, a probabilidade é definida como métrica de incerteza. Modelando um **GLDM**, faremos inferência sobre os parâmetros calculando graus de incerteza sobre os mesmos, tanto observados como latentes, por meio de uma **F.D.P. Posteriori**. O teorema de *Bayes* é a peça central dessa abordagem, nos dando uma relação de proporcionalidade para a distribuição a *Posteriori* em função da: **i)** função de *verossimilhança* $\pi(y_t|\theta)$, que é o modelo de probabilidade que descreve os dados; **ii)** distribuição a *Priori* $\pi(\theta)$, o que já sabíamos sobre os parâmetros antes de vermos os dados; **iii)** E Probabilidade Marginal dos dados $\pi(y_t)$, que é a probabilidade de obtermos esse dado dentre todos os θ 's possíveis:

$$\pi(\theta|y_t) = \frac{\pi(y_t|\theta)\pi(\theta)}{\pi(y_t)} \quad (1.7)$$

Note que tanto os dados futuros quanto ausentes também são incertos e podem ser representados por distribuições *Posteriori*'s, estas que em alguns casos particulares podem ser escritas analiticamente como uma conjugação de distribuições já conhecidas, método conhecido como famílias conjugadas, definido em [Kinas e Andrade \(2010\)](#), como:

Definição 2. *Seja F uma família de distribuições para a verossimilhança $\pi(y | \theta)$ e P uma família de distribuições para a priori $\pi(\theta)$. Dizemos que \mathcal{F} e P são famílias conjugadas de distribuições se a distribuição posterior $\pi(\theta | y)$ também for membro de P .*

Em casos genéricos existem dois caminhos para se utilizar, um de simulação e o outro de aproximação. Dentro dessas sub-áreas existem alguns métodos como:

- i) Simulação por Cadeias de Markov de Monte Carlos (**MCMC**) (Markov chain Monte Carlo) para esse propósito temos por exemplo o *Amostrador de Gibbs* e *Metropolis-Hastings*. Uma introdução a esses métodos pode ser vista em [D. Cemgil A.T. \(2011\)](#);
- ii) A Aproximação de Laplace Aninhada Integrada (*Integrated Nested Laplace Approximation* - **INLA**) ([RUE; MARTINO, 2005](#)) e Aproximação Gaussiana ([RUE; HELD, 2005](#))

Ambos os caminhos tem seus respectivos pacotes implementados no *software R*. A principal diferença prática entre eles é o custo computacional. O **MCMC**, por exemplo, depende de critérios de convergência, embora sua robustez seja indiscutível. Para Modelos Dinâmicos, em particular os **GLDM**, onde o tempo computacional do **MCMC** deixa muito a desejar, os métodos de aproximação como o **INLA** vem ganhando mais adeptos nos últimos anos.

1.5 Objetivos do Trabalho

Este trabalho tem por objetivo o estudo das séries de tempo, com dados ausentes, usando uma abordagem de Modelos Dinâmicos com Inferência Bayesiana.

1.5.1 Objetivos Específicos

- i) Modelar as observações do preço do Bagre como função da variável latente com ruído em um **GLDM**;
- ii) Implementar o modelo no *software R* usando o pacote INLA partindo de uma abordagem *Bayesiana*.
- iii) Selecionar o melhor modelo ajustado dando um grau de incerteza para cada ponto, permitindo estudos futuros sobre a série.
- iv) Levantar hipóteses sobre o modelo ajustado que possam servir de motivação para trabalhos futuros.

2 Material e Métodos

2.1 Banco de Dados

Com o intuito de entender como se dá a variação do preço dos pescados, foram coletados os preços de diferentes espécies vendidas no mercado de Rio Grande sob orientação da Professora Dra. Patrícia Raggi Abdallah, do Instituto de Ciências Econômicas, Administrativas e Contábeis (**ICEAC**) da Universidade Federal do Rio Grande (**FURG**), constituindo o Índice de Preço Pago Pelo Consumidor de Pescado no Município de Rio Grande (**IPPPC-RG**).

O banco de dados utilizado foi coletado pela Unidade de Pesquisa Costeira e Marinha (**UPEC-MAR**), durante os períodos de Agosto de 2006 a Março de 2016, aproximadamente 11 anos, com 116 observações. O índice utilizou-se da metodologia do cálculo do Índice de Preços Pagos ao Consumidor Amplo do Instituto Brasileiro de Geografia e Estatística (**IPCA-IBGE**). Foram coletados preços de vários pescados vendidos na cidade de Rio Grande. Preço dado com base no preço pago por quilograma (**Kg**) de pescado pelo consumidor final, categorizado: **i**) pela espécie; **ii**) pelo processamento (Inteiro sujo, Inteiro Limpo, filé, posta...). Os preços estão disponíveis mensalmente no site **UPEC-MAR** <www.upec.furg.br>.

2.1.1 Procedimento da coleta

i) Determinação das Regiões coletadas

Os pontos selecionados para as coletas amostrais foram pontos comerciais de peixe em diferentes bairros no município incluindo Centro e Balneário Cassino.

ii) Preço Coletado

Os preços coletados são referentes ao preço praticado no mês anterior, coletados no primeiro dia de cada mês.

iii) Espécies Coletadas

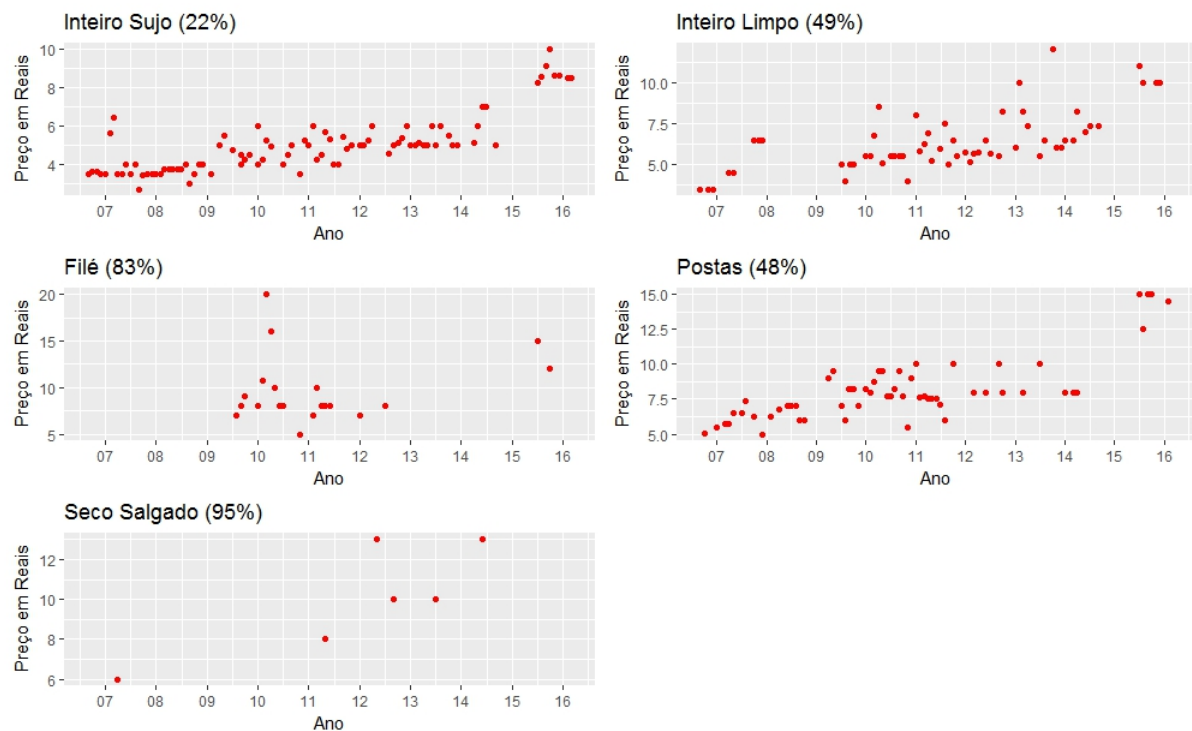
As espécies coletadas para a formação do Índice são as seguintes: abrótea, batata, cação, cação anjo, cação anequim, caçonete, castanha, cherne, congro rosa, enchova, olhete, garoupa, mangona, marisco, merluza, miragaia, pacu, namorado, pampo, papaterra, pargo, peixe-porco, peixe-sapo (tamoril), pescada olhuda, pescadiha, polvo, tilápia, viola, atum, espadarte-meca, lula, lula (anéis), lula (canudo), bagre, corvina, linguado (grande, médio e pequeno), ova de peixe, peixe rei, tainha, camarão-rosa/oce (VG), camarão-rosa/oce (GG), camarão-rosa/lag. (grande,

médio e pequeno), peixe-galo, carne de siri (garra), carne de siri (branca), jundiá, carpa capim, carpa húngara, carpa prateada, salmão, dourado (oceano), sardinha, mexilhão, trilha, lagosta sapateira (grande e média)

Das espécies coletadas escolhemos o pescado Bagre, que no varejo de peixe é composto por três espécies: (i) *Genidens genidens*, (ii) *Genidens barbatus* e (iii) *Genidens planifrons*. As espécies (ii) e (iii) estão em foco pela recente discussão sobre a sustentabilidade de sua pesca. Elas foram categorizadas respectivamente como Espécies em Perigo (**EN**) e Espécies Criticamente em Perigo (**CR**), na portaria 445 de 2014 e MMA 2016 que tratam sobre o tema. (citar portarias bib)

Na coleta Figura 1, podemos visualizar a série do Bagre categorizada por processamento bem como a proporção de dados ausentes.

Figura 1 – Preço mensal do Bagre em Rio Grande agosto de 2006 a Março de 2016, parenteses indicam a proporção de dados ausentes.



Fonte: Elaborada pelo autor, dados **ICEAC - FURG**

2.2 Modelos Dinâmicos e Modelo de Markov Oculo

Dentro dos Modelos Dinâmicos optamos por utilizar um processo da família dos Autorregressivos como na Equação 1.4 no campo latente, também conhecido como modelo de de *Markov* Oculo (Hidden Markov Model - **HMM**). Um **AR(1)** adicionado à variável

latente seria um **HMM(1)**, considerando o erro observacional v_t . Ou seja, um **AR(p)** e um **HMM(p)** teriam a mesma forma dinâmica, considerando que o erro da observação é nulo $v_t = 0$. Reescrevendo **GLDM** da Equação 1.2 com $k=1$ observações por mês temos:

$$\text{Equação de Observações} \quad y_t = F'x_t + v_t, \quad v_t \stackrel{iid}{\sim} N(0, V) \quad (2.1)$$

$$\text{Equação do Processo} \quad x_t = G_t x_{t-1} + w_t, \quad w_t \stackrel{iid}{\sim} N(0, W) \quad (2.2)$$

$$(2.3)$$

$$x_t = \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{bmatrix}_{px1} \quad x_{t-1} = \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-h} \end{bmatrix}_{px1} \quad F = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{px1} \quad w_t = \begin{bmatrix} w_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{px1} \quad v_t = \begin{bmatrix} v_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{px1}$$

$$G = \begin{bmatrix} \rho_1 & \rho_2 & \dots & \rho_{h-1} & \rho_h \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}_{pxp}$$

Um modelo Dinâmico pode ser visto na forma de modelo Hierárquico Bayesiano. Nesses modelos existem pelo menos dois níveis de dependência entre as variáveis.

$$\text{Hierarquia} \quad (2.4)$$

$$\text{Nível 3} \quad \boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}) \quad (2.5)$$

$$\text{Nível 2} \quad x \sim \pi(x | \boldsymbol{\theta}) = N(x; 0, \mathbf{Q}^{-1}(\boldsymbol{\theta})) \quad (2.6)$$

$$\text{Nível 1} \quad \mathbf{y} | \mathbf{x} \sim \pi(y_i | x_i), \quad i = 1, \dots, n \quad (2.7)$$

Onde os *hiperparâmetros* do **HMM** da Equação 2.1 são $\theta = \{V, W, \rho_t\}$. Estes são estimados como base para encontrar o processo gerador para as variáveis latentes e observadas. A distribuição *Posteriori* da forma Hierárquica 2.4 é escrita como:

$$\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y}) \propto \pi(\boldsymbol{\theta}) \pi(\mathbf{x} | \boldsymbol{\theta}) \prod_{i=1}^n (y_i | x_i).$$

Essas redes de informação *hierárquica* são melhor representadas como Grafos Dirigidos Acíclicos (Directed Aciclic Graph -**DAG**). Formalmente definido em Rue e Held (2005) como:

Definição 3. Um grafo dirigido \mathbf{G} é o par $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, onde \mathbf{V} é o conjunto de nós e \mathbf{E} é o conjunto de arcos $\{i, j\}$, onde $i, j \in \mathbf{E}$, e existe uma ligação dirigida entre os nós i e j . Caso contrário não há ligação entre os elementos i e j . Um grafo é completamente conectado se $\{i, j\} \in E$ para todo $i, j \in \mathbf{V}$ com $i \neq j$.

A forma *hierárquica* 2.4 usando **DAG** é representada na Figura 2. Onde a direção dos arcos representa a dependência condicional, e os nós são variáveis com incerteza associada.

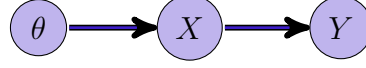
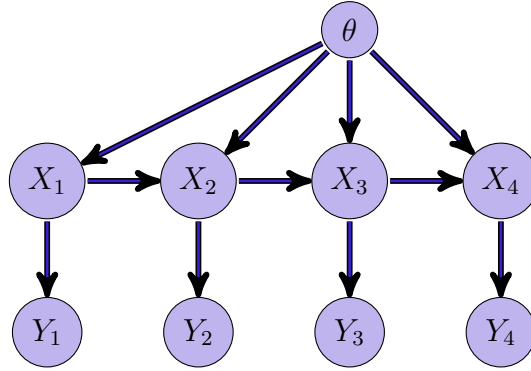


FIGURA. 2 – Grafo Dirigido Acíclico

Cada um dos níveis contém ainda outras dependências simplificadas nesse grafo. Em particular a do **GLDM** escolhido na Equação 2.1 considerando $h = 1$ em um **DAC** seria escrito como na Figura 3.

Figura 3 – Grafo Dirigido Acíclico de um **HMM(1)**



2.3 Campos Aleatórios Gaussianos Markovianos

Uma das bases do **INLA** são os Campos Aleatórios Gaussianos Markovianos (Gaussian Markov Random Field - **GMRF**), que provém dos *Processo de Markov* que é o processo de crescimento da variável latente do **HMM**, e possui uma propriedade muito interessante de independência condicional. Um **GMRF** é formalmente definido em [Rue e Held \(2005\)](#) como:

Definição 4. (*Campos Aleatórios Gaussianos Markovianos*) O vetor aleatório $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}$ é um **GMRF** com relação ao grafo $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ com média $\boldsymbol{\mu}$ e matriz de precisão $\mathbf{Q} = \boldsymbol{\Sigma}^{-1} > 0$, onde $\boldsymbol{\Sigma}$ é matriz de covariância, se e somente se sua densidade tem a forma

$$\pi(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} |\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q}(\mathbf{x} - \boldsymbol{\mu})\right)$$

$$\mathbf{Q}_{ij} \neq 0 \Leftrightarrow \{i, j\} \in \mathbf{E} \quad \forall i \neq j$$

Ou seja é um vetor normalmente distribuído com a propriedade de independência condicional associada. Uma forma de ver a independência condicional se dá pelo Teorema 1 abaixo:

Teorema 1. *Seja \mathbf{x} normalmente distribuído com média μ e precisão \mathbf{Q} . Então para $i \neq j$ temos:*

$$x_i \perp x_j \mid \mathbf{x}_{-ij} \iff \mathbf{Q}_{ij} = 0$$

Onde \perp significa independência, ou seja, podemos verificar em \mathbf{Q} quando x_i e x_j são condicionalmente independentes, observando apenas se a entrada \mathbf{Q}_{ij} é zero. \mathbf{Q} é uma matriz esparsa, ou seja com muitos zeros. A demonstração desse teorema pode ser vista Rue e Held (2005, p. -33) .

Quando o **GMRF** é escrito na forma dinâmica a propriedade de independência condicional não fica tão clara. Por exemplo considere um **AR(1)** na forma dinâmica teríamos:

$$\begin{array}{lll} \text{Equação de Observações} & y_t = 1x_t + v_t, & v_t \stackrel{iid}{\sim} N(0, V) \\ \text{Equação do Processo} & x_t = \rho_1 x_{t-1} + w_t, & w_t \stackrel{iid}{\sim} N(0, 1) \end{array}$$

Simplificando:

$$x_t = \rho_1 x_{t-1} + w_t, \quad w_t \stackrel{iid}{\sim} N(0, 1), \quad \text{com } |\rho_1| < 1$$

Escrevendo na forma condicional:

$$x_t \mid x_1, \dots, x_{t-1} \sim N(\rho_1 x_{t-1}, 1) \quad t = 2, \dots, n$$

Como o processo é estacionário sua variância pode ser facilmente obtida. Usando $\mathbb{V}(\cdot)$ como símbolo para função Variância e $\mathbb{E}(\cdot)$ como símbolo para função Esperança.

$$\begin{aligned} \mathbb{V}(x_t) &= \mathbb{V}(\rho_1 x_{t-1}) + \mathbb{V}(w_t) \Rightarrow \mathbb{V}(x_t) - \rho_1^2 \mathbb{V}(x_{t-1}) = 1 \Rightarrow \mathbb{V}(x_t) = \frac{1}{1 - \rho_1^2} \\ \mathbb{E}(x_t) &= \mathbb{E}(\rho_1 x_{t-1}) + \mathbb{E}(w_t) \Rightarrow \mathbb{E}(x_t) = 0 \end{aligned}$$

Sua densidade conjunta é da forma:

$$\begin{aligned} \pi(\mathbf{x}) &= \pi(x_1) \prod_{t=2}^n \pi(x_t \mid x_{1:t-1}) = \pi(x_1) \pi(x_2 \mid x_1) \dots \pi(x_n \mid x_{n-1}) = \\ &= \frac{1}{(2\pi)^{n/2}} \mid \mathbf{Q}(\boldsymbol{\theta}) \mid^{1/2} \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x}\right) \end{aligned}$$

Um **GMRF** como pode ser visto na Definição 3.

Na Matriz de precisão Q para um $n=7$. Como no Teorema 1 ela tem forma esparsa, o que indica as independências condicionais. As entradas em branco são zeros.

$$Q = \Sigma^{-1} = \begin{bmatrix} 1 & -\rho & & & & & \\ -\rho & 1 + \rho^2 & -\rho & & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -\rho & 1 + \rho^2 & -\rho & \\ & & & & -\rho & 1 & \end{bmatrix} \quad (2.8)$$

Já a matriz de covariância para é da forma:

$$\Sigma = \frac{1}{1 - \rho_1^2} \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \quad (2.9)$$

O que não nos fornece muitas informações sobre o processo analisado.

2.4 Aproximação de Laplace e Laplace-Simplificada

A aproximação de Laplace-Simplificada é uma aproximação de uma curva qualquer $g(x)$ por uma família de funções exponenciais. Para isso é usada uma expansão em séries de Taylor até a segunda ordem centrada na moda x^* , pois a moda é um ponto de máximo local (global se unimodal), portanto sua derivada naquele ponto é zero.

$$\begin{aligned} \log g(x) &\approx \log g(x^*) + \frac{\partial \log g(x^*)}{\partial x}(x - x^*) + \frac{1}{2} \frac{\partial^2 \log g(x^*)}{\partial x^2}(x - x^*)^2 \\ &= \log g(x^*) + \frac{1}{2} \frac{\partial^2 \log g(x^*)}{\partial x^2}(x - x^*)^2 \\ \text{Tomando} \quad \hat{\sigma}^2 &= -\frac{1}{\frac{\partial^2 \log g(x^*)}{\partial x^2}} \end{aligned}$$

Podemos escrever

$$\int e^{\log g(x)} dx \approx g(x^*) \int \exp \left[-\frac{(x - x^*)^2}{2\hat{\sigma}^2} \right] dx = \text{Normal}(x^*, \hat{\sigma}^2)$$

A aproximação de Laplace para com \mathbf{x} multivariado pode ser vista em [Tierney e Kadane \(1986\)](#) como:

$$\lim_{n \rightarrow \infty} \int e^{ng(x)} dx = \lim_{n \rightarrow \infty} \sqrt{\frac{(2\pi)^n}{n |-\mathbf{H}|}} \quad (2.10)$$

$$\text{onde } \mathbf{H} \text{ é a matriz Hessiana na moda } x^* \quad (2.11)$$

$$H_{ij} = \frac{\partial^2}{\partial x_i \partial x_j} \Big|_{x=x^*} \quad (2.12)$$

As propriedades do **GMRF** são preservadas em ambos os casos.

2.5 Integrated Nested Laplace Aproximation

O principal objetivo do **INLA** é a aproximação das marginais as *Posteriori*'s tanto das variáveis latentes (isto é, não observadas) quanto dos *Hiperparâmetros* do **GMRF** os dados observados (coletados), permitindo assim realizar inferência *Bayesiana* sobre as variáveis. Ele produz uma aproximação numérica para as nossas *Posteriori*'s de Interesse usando a Aproximação de Laplace ([TIERNEY; KADANE, 1986](#)). As Marginais são escritas como:

$$\pi(x_i | \mathbf{y}) = \int \pi(x_i | \boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad (2.13)$$

$$\pi(\theta_j | \mathbf{y}) = \int \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-j} \quad (2.14)$$

Seja $\tilde{\pi}(\cdot | \cdot)$ a densidade condicional aproximada de Laplace dos argumentos, o INLA tenta construir a aproximação aninhada (*Nested*) das marginais como:

$$\tilde{\pi}(x_i | \mathbf{y}) = \int \tilde{\pi}(x_i | \boldsymbol{\theta}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad (2.15)$$

$$\tilde{\pi}(\theta_j | \mathbf{y}) = \int \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-j} \quad (2.16)$$

Já que não temos uma forma analítica para $\pi(\boldsymbol{\theta} | \mathbf{y})$ e para $\pi(x_i | \boldsymbol{\theta}, \mathbf{y})$ para o cálculo de [2.16](#) e [2.15](#) o **INLA** faz aproximações analíticas e integrações numéricas em duas etapas:

- i) Etapa - Se calcula $\pi(\boldsymbol{\theta} | \mathbf{y})$ e todas as outras marginais $\pi(\theta_j | \mathbf{y})$ que queremos também podem ser calculadas.
- ii) Etapa - Calculamos $\pi(x_i | \boldsymbol{\theta}, \mathbf{y})$ que é necessária para o cálculo das marginal *Posteriori*'s dos parâmetros.

Primeiramente é feita uma aproximação da *Posteriori* $\pi(\boldsymbol{\theta} | \mathbf{y})$ dos Hiperparâmetros $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_m\} \in \mathbb{R}^m$. Usando a definição de probabilidade conjunta temos:

$$\pi(\boldsymbol{\theta} | \mathbf{y}) = \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\pi(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})} \propto \frac{\pi(\boldsymbol{\theta}) \pi(\mathbf{x} | \boldsymbol{\theta}) \pi(\mathbf{y} | \mathbf{x})}{\pi(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})} \simeq \frac{\pi(\boldsymbol{\theta}) \pi(\mathbf{x} | \boldsymbol{\theta}) \pi(\mathbf{y} | \mathbf{x})}{\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})} \Big|_{x=x^*(\boldsymbol{\theta})} \quad (2.17)$$

Onde $\tilde{\pi}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})$ é a aproximação gaussiana para a condicional completa $\pi(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})$, também conhecida como aproximação por **GMRF**, com $\mathbf{x} = \mathbf{x}^*(\boldsymbol{\theta})$ sua moda ([RUE; MARTINO; CHOPIN, 2009](#)), ([RUE; HELD, 2005](#)).

$$\begin{aligned}\tilde{\pi}(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y}) &\propto \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x} \Sigma_i \log \pi(y_i \mid x_i)\right) \approx \\ &\approx \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{Q}(\boldsymbol{\theta}) + \text{diag}(C_i))(\mathbf{x} - \boldsymbol{\mu})\right) = \tilde{\pi}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})\end{aligned}$$

Onde C_i é o coeficiente da expansão de $\log \pi(y_i \mid x_i)$ em Taylor até a segunda ordem, com centro na moda, x_i^* , como visto em [Rue e Held \(2005, p.-211\)](#). O que é equivalente a aproximação de Laplace para $\pi(\boldsymbol{\theta} \mid \mathbf{y})$, mostrado em [Tierney e Kadane \(1986\)](#).

Agora precisamos calcular $\pi(x_i \mid \boldsymbol{\theta}, \mathbf{y})$. Uma maneira direta é para $\tilde{\pi}(x_i \mid \boldsymbol{\theta}, \mathbf{y})$ usando uma distribuição Normal, onde o cálculo da matriz de precisão é baseada na decomposição de Cholesky ([RUE; HELD, 2005](#)). A aproximação foi testada em [Rue e Martino \(2005\)](#) mas apresenta erros de assimetria e de alocação. A solução encontrada por eles foi usar a aproximação de Laplace novamente. Primeiramente decompomos $\mathbf{x} = (x_i, \mathbf{x}_{-i})$. Temos então:

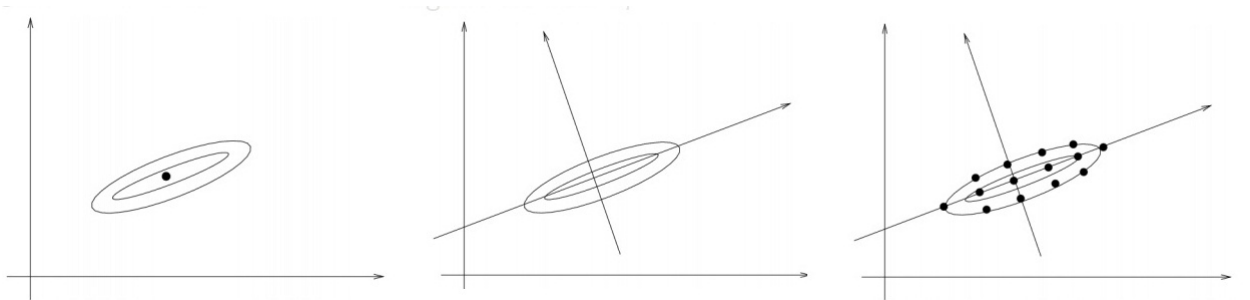
$$\pi(x_i \mid \boldsymbol{\theta}, \mathbf{y}) = \frac{\pi((x_i, \mathbf{x}_{-i}) \mid \boldsymbol{\theta}, \mathbf{y})}{\pi(\mathbf{x}_{-i} \mid x_i, \boldsymbol{\theta}, \mathbf{y})} \simeq \frac{\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y})}{\tilde{\pi}_G(\mathbf{x}_{-i} \mid x_i, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}_{-i} = \mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})} \quad (2.18)$$

Onde $\tilde{\pi}_G(\mathbf{x}_{-i} \mid x_i, \boldsymbol{\theta}, \mathbf{y})$ é a aproximação gaussiana para a $\pi(\mathbf{x}_{-i} \mid x_i, \boldsymbol{\theta}, \mathbf{y})$ e $\mathbf{x}_{-i} = \mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})$ sua moda.

Como a aproximação de Laplace é computacionalmente muito "cara", o **INLA** opta pela aproximação simplificada de Laplace por *default*, para $\pi(x_i \mid \boldsymbol{\theta}, \mathbf{y})$.

Na prática o **INLA** primeiro procura na marginal *Posteriori* conjunta do hiperparâmetro pela moda de $\pi(\boldsymbol{\theta} \mid \mathbf{y})$. Disto é montada uma malha de pontos, deixando a nova base com origem na moda, e se escolhe um conjunto **G** dos pontos relevantes $\{\boldsymbol{\theta}^*\}$ juntamente com um conjunto de pesos para os pontos mais relevantes $\{W_{\boldsymbol{\theta}^*}\}$, de modo a gerar uma distribuição aproximada como na Figura 4.

Figura 4 – Moda Encontrada; Mudança de Base; Pontos Relevantes

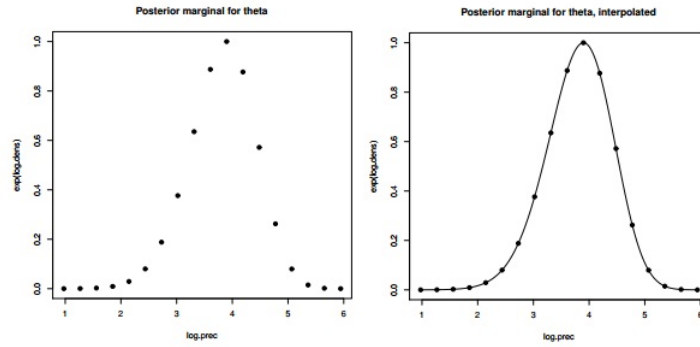


Fonte: Minicurso Håvard Rue

Cada Marginal *Posteriori* $\tilde{\pi}(\boldsymbol{\theta} \mid \mathbf{y})$ pode ser obtida interpolando os pontos calculados usando uma taxa de correção de assimetria. Como mostra na Figura 5.

Para cada θ^* , a *Posteriori* condicional $\tilde{\pi}(x_i \mid \theta^*, \mathbf{y})$ é então verificada na malha dos va-

Figura 5 – Pontos Encontrados; Interpolação da Marginal *Posteriori*.

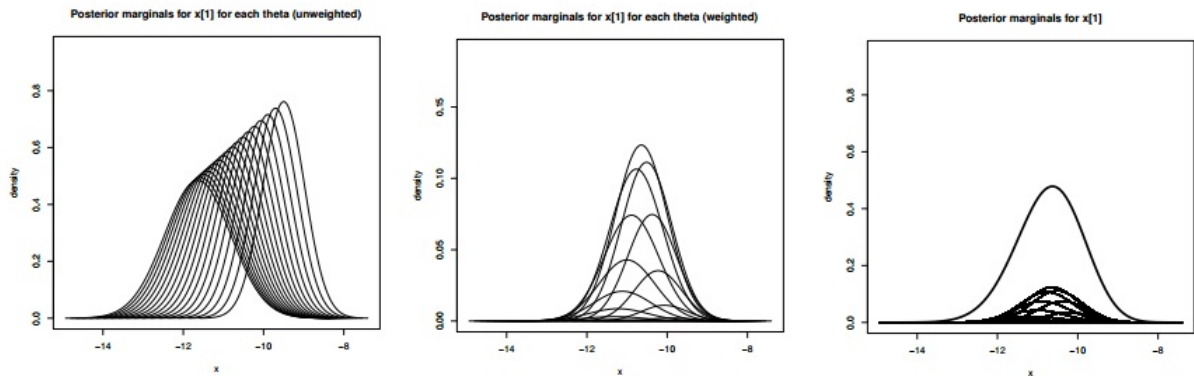


Fonte: Minicurso Håvard Rue

lores selecionados x_i e a marginal *Posteriori* $\tilde{\pi}(x_i \mid \mathbf{y})$ é obtida por integração numérica, como na Equação 2.19, considerando o conjunto de pesos W_{θ^*} , como na Figura 6 .

$$\tilde{\pi}(x_i \mid \mathbf{y}) = \sum_{\theta^* \in G} \tilde{\pi}(x_i \mid \theta^*, \mathbf{y}) \tilde{\pi}(\theta^* \mid \mathbf{y}) W_{\theta^*} \quad (2.19)$$

Figura 6 – Conjuntos de Marginais *Posteriori*; Ajuste Com Pesos; Integração Numérica.



Fonte: Minicurso Håvard Rue

Comparações com o **MCMC**, comparação entre as aproximações de Laplace e Gaussiana para as *Marinais Posterior's* podem ser encontradas em: [Smedt et al. \(2015\)](#), [Rue e Martino \(2005\)](#), [Rue, Martino e Chopin \(2009\)](#), [Rue e Held \(2005\)](#).

2.6 Seleção de Modelo

O grau de ajuste aos dados é muito importante, em analogia com o R^2 em uma regressão linear, queremos comparar diferentes modelos em uma mesma escala. Com esse propósito foi criado o Critério de Informação da Deviância (Deviance information criterion **DIC**), calculado como:

$$D(\mathbf{x}; \boldsymbol{\theta}) = 2\sum_i \log(y_i | x_i, \boldsymbol{\theta})$$

$$DIC = 2\mathbb{E}(D(\mathbf{x}; \boldsymbol{\theta})) - D(\mathbb{E}(\mathbf{x}); \boldsymbol{\theta}^*)$$

Ou seja **DIC**= "Qualidade do ajuste"+ "penalidade pela complexidade".

Outras métricas também são utilizados com o mesmo intuito, por exemplo: **i)** Akaike information criterion **AIC**, Bayesian information criterion **BIC**, usados para análise frequentista; **ii)** **DIC** e mais recentemente Watanabe–Akaike information criterion **WAIC** para modelos *Bayesianos*.

2.7 Implementação do RINLA

O pacote com o método já implementado pode ser instalado facilmente usando o comando `install.packages(INLA)` no *prompt* do *R* ou *Rstudio*. Pode-se atualizar usando o comando `inla.upgrade()`, ou caso convenha a versão de teste `inla.upgrade(testing = TRUE)`. O **INLA** está disponível para Linux, Windows e Mac. Depois de instalado, para carregar utiliza-se o comando `library(INLA)`.

O modo de estruturar o código é bem parecido com o da função *lm* e o *GLM*

```
formula <- dados ~ x1 + x2 + f(i, model=...) +1
```

Onde $(x1 + \dots)$ são as co-variáveis que possuem relações lineares; +1 ou -1 indicam a existência ou ausência de efeitos fixos (constantes); e $f(\cdot)$ é o onde se adiciona os efeitos estruturados já implementados. Todos os modelos podem ser vistos executando o comando:

```
names(inla.models())$latent)$
```

No Site **INLA Project** ([RUE; FERKINGSTAD, 2017](#)) existe um documento técnico com uma pequena descrição de cada modelo implementado na pasta *models*.

Para carregar o modelo, deve-se utilizar a função `inla(.)`:

```
saida <- inla(formula, data = data.frame(x1,x2,i,dados),
family = "gaussian", control.compute = list(waic=T, dic=T ...))
```

Em `data.frame(.)` devem ser adicionadas todas as observações e índices criados. No caso de séries temporais uma das observações deve ser o índice para tempo. Por exemplo em 50 observações mensais teríamos $i = 1 : 50$.

Para um modelo onde os dados são uma função aleatória do tempo, teríamos:

```
i=1:length(dados)
formula <- dados ~f(i, model="rw1",constr=F) -1
saida <- inla(formula, data = data.frame(i,dados),
family = "gaussian")
```

Os modelos autoregressivos já estão implementados. Outros modelos dinâmicos podem ser ajustados aplicando o método proposto em (CARDENAS; KRAINSKI; RUE, 2010) e (CORTES, 2014).

Alguns critérios de seleção podem ser solicitados para serem computados como: **DIC**, **AIC**, **WAIC** entre outros:

```
inla(...,control.compute = list(waic=T, dic=T,...),...)
```

O **DIC** calculado pode ser obtido em:

```
saida$dic$dic
```

As distribuições marginais ficam armazenadas em:

```
saida$marginals.hyperpar$...
```

As distribuições estimadas podem ser encontradas em:

```
saida$marginals.fitted.values[[i]]
```

Os quantis calculados para o intervalo de credibilidade, por padrão de 95%, ficam armazenados em:

```
saida$summary.random$i
```

Usando o pacote *ggplot2*, para a análise gráfica do ajuste, produzimos um gráfico dos quantis com os dados observados:

```
banco <- subset(banco, select = c("dados","data"))
estimado<-as.data.frame(saida$summary.random$i)
banco <- cbind(banco, estimado) #salvar como data frame
ggplot(banco, aes(x=data )) +
```

```

geom_line(aes(y='0.5quant', colour="Quantil_0.5"))+
geom_line(aes(y='0.025quant', colour="Quantil_0.025"),
linetype="dashed", size=1) +
geom_line(aes(y='0.975quant', colour="Quantil_0.975"),
linetype="dashed", size=1)+
geom_point(aes(y=dados, colour="Observados"), size=1) +
scale_colour_manual("",breaks =c("Quantil_0.5",
"Quantil_0.025","Quantil_0.975","Observados"),
values = c("Quantil_0.5"="black","Quantil_0.025"="blue"
"Quantil_0.975"="blue","Observados"="red")) +
xlab("") + ylab("") + labs(fill = "") + ggtitle("") +
scale_x_date(breaks = date_breaks("1_year"),labels =
date_format("%y"))

```

As *Posteriori*'s estimadas para cada ponto, e o valor observado podem ser observados conjuntamente:

```

ggplot(data.frame(saida$marginals.fitted.values[[i]]),aes(x,y))+
geom_line(size=1 , aes(colour="Estimado"))+
geom_vline(aes(xintercept = dados1[i],colour="Observado")) +
scale_x_continuous(limits = c(0,100))+
scale_colour_manual("Legenda",breaks = c("Estimado","Observado"),
values = c("Estimado"="blue", "Observado"="red")) +
labs("")+ggtitle("")

```

O mesmo *script* pode ser usado para os hiperparâmetros marginais. Para mais informações e veja (RUE; FERKINGSTAD, 2017).

3 Aplicação e Discussão

Primeiramente consideraremos um Passeio Aleatório de ordem 1 (Random Walk - **RW**(1)) ajustado no processo latente. O que no **HMM**(p) da Equação 2.1 seria facilmente ajustado fazendo $p=1$ e $\rho = 1$. Em outras palavras, um **RW**(p) é um **AR**(p) com as variáveis ρ_t que multiplicam as defasagens constantes e iguais a 1 $\rho_t = 1$ (RUE; HELD, 2005). Ele é ajustado como:

```
formula <-dados ~f(i, model="rw1", constr=F, prior = "normal",
                    param = c(mean=0,prec=1e-5)) -1
```

Para um **RW**(2), basta modificar "rw1" por "rw2".

Escolhemos *Priori* não informativa deixando a precisão 1×10^{-5} (precisão é o inverso da variância) logo, escolhemos uma variância alta para não enviesar a estimativa. Podemos ver na saída do ajuste com o comando *summary(r)*:

Time used:

Pre-processing	Running inla	Post-processing	Total
0.1670	0.8640	0.0890	1.1201

Usando como *Default* do programa a aproximação de Laplace simplificada, podemos ver que o tempo total de processamento levou aproximadamente 1 segundo, usando um processador Intel core I5 com 4 Gb de memória Ram.

Com base nas *Posteriors* o **INLA** calcula um intervalo de credibilidade por padrão de 95% dando quantis para cada ponto amostral (dados ausentes ou não), como podemos visualizar na Figura 7. Onde os pontos vermelhos são as observações.

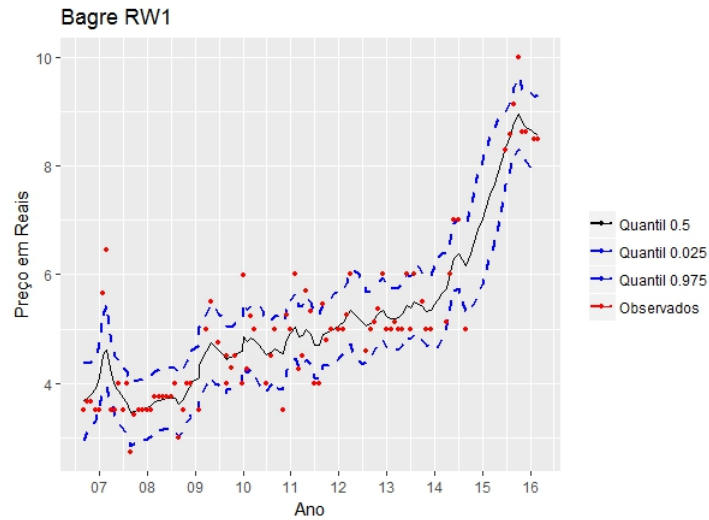
Para seleção de modelos calculamos o **DIC**. Ajustamos para comparação um **HMM**(1), Equação 2.1, com $h = 1$ e ρ um *hiperparâmetro* a ser estimado, no código:

```
formula <-dados ~f(i, model="ar", order=p, constr=F, prior =
"normal",param = c(mean=0,prec=1e-5)) -1
```

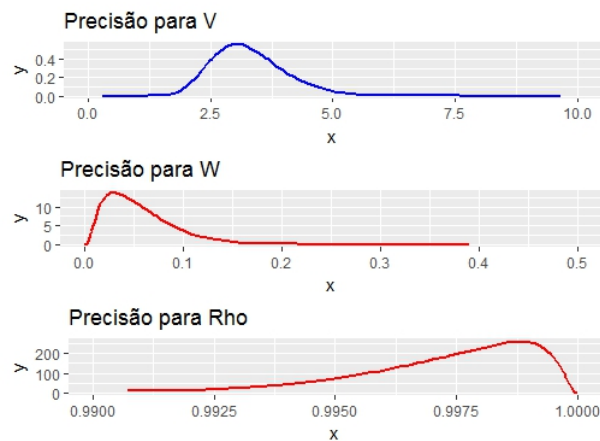
Onde p é a ordem que queremos do auto-regressivo.

Note também, que na precisão para ρ , o valor se mantém abaixo de 1 diferentemente de um **RW**(1).

Testamos outros **HMM**(p), com diferentes defasagens $p = \{2, 3, 4\}$, mudando apenas no código $p = 2, 3, 4$. Os resultados dos critério de informação podem ser vistos na Tabela 1 abaixo:

Figura 7 – Intervalos de Credibilidade estimados para Bagre-Sujo com o modelo **RW(1)**.

Fonte: Elaborada pelo autor

Figura 8 – Marginais Posteriores Hiperparâmetros do modelo **HMM(1)**

Fonte: Elaborada pelo autor

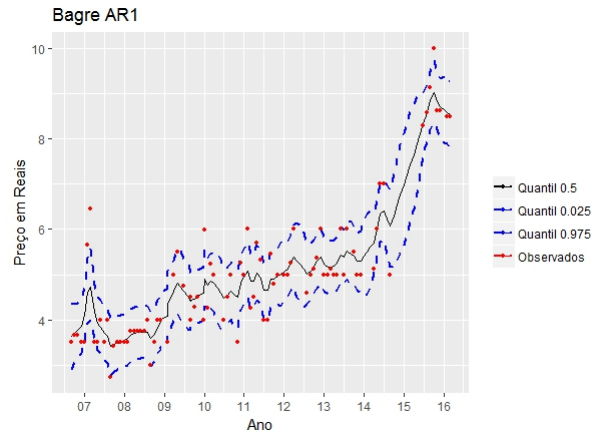
Tabela 1 – Critério de Informação da Deviância dos Modelos Testados

INTEIRO SUJO					
HMM(1)	HMM(2)	HMM(3)	HMM(4)	RW(1)	RW(2)
182.61	181.86	161.87	-515.60	182.08	195.77

O modelo escolhido foi o de menor **DIC** em valor absoluto, **(3)** com **DIC**=161.87, onde a qualidade do ajuste (*Deviance Mean*) foi de 116.34 e o número de parâmetros efetivos (complexidade) foi de 45.53. Note que um **HMM(4)** apresenta um **DIC** negativo, o que indica um sobreajuste (*Overfitting*) do modelo.

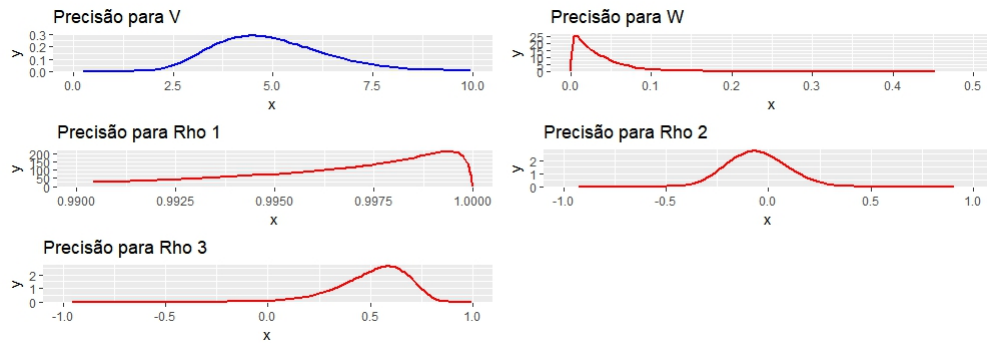
As posteriores podem ser vistas na Figura 10 e os intervalos de credibilidade de 95% ajustados podem ser visualizados na Figura 11.

Figura 9 – Intervalos de Credibilidade estimados para Bagre-Sujo com o modelo **HMM(1)**.



Fonte: Elaborada pelo autor

Figura 10 – Marginais *Posterioris* dos Hiperparâmetros do modelo **HMM(3)**



Fonte: Elaborada pelo autor

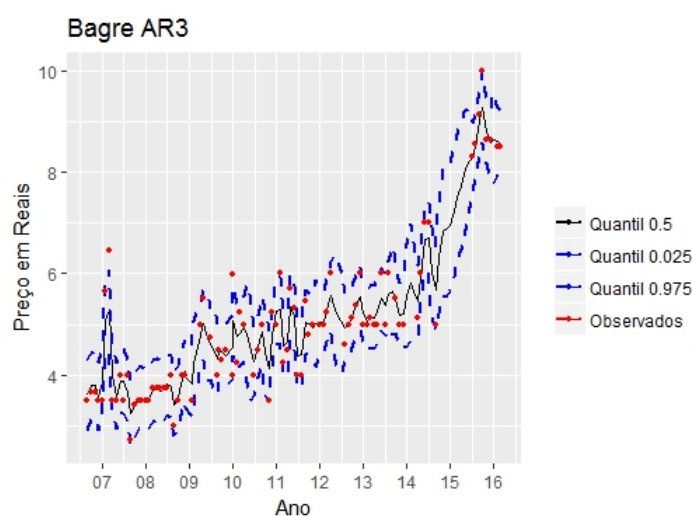
Com menor **DIC**, as diferenças de ajuste entre o **HMM(1)** Figura 9 e **HMM(3)** Figura 11 se tornam visíveis. O **HMM(3)** tem um refinamento maior, capturando além da tendência, as oscilações referentes ao preço.

Como o peixe é um bem inelástico, ou seja pequenas oscilações no preço não alteram a demanda, é provável que essas variações se devam então ao volume de oferta do Bagre ou às oscilações de espécies substitutas. Tal hipótese poderia ser testada adicionando co-variáveis de volume e preço ao modelo, e verificando através da melhora efetiva no **DIC** por exemplo.

Poderíamos procurar padrões, decompondo as séries de quantis em: tendência, sazonalidade e ruído.

Com as posteriores para cada ponto possuímos muito mais informações que originalmente; e mais hipóteses que podem ser levantadas e testadas do que com uma simples série completa. Algumas delas que antes seriam impensáveis agora poder ser propostas. Por exemplo poderíamos comparar as posteriores de dois meses como na Figura 12 abaixo. Visualmente

Figura 11 – Intervalos de Credibilidade estimados para Bagre-Sujo com o modelo **HMM(3)**.



Fonte: Elaborada pelo autor

Figura 12 – *Posteriori* para as Observações



Fonte: Elaborada pelo autor

temos o efeito de translação da **F.D.P.**, entre os períodos de Agosto de 2007 e agosto de 2008. Efeito que pode ser explicado parcialmente pela inflação. Removendo a inflação, o preço poderia ser explicado por uma possível maior escassez do pescado. Poderíamos também analisar o grau de aumento dessas incertezas, entre períodos, calculando a razão entre as *Posteriori's*. Essas e outras hipóteses podem ser levantadas no modelo com as *Posteriori's* estimadas e as observações. Outros estudos podem ser iniciados a partir do modelo ajustado.

O mesmo método pode ser ajustado a outras espécies do banco de dados.

Conclusão

O **INLA** tem um potencial muito promissor, principalmente em *Data Mining* e Inteligência Artificial, áreas em que se exige uma velocidade maior de processamento, constantemente se estimam *Posterioris* e as mesmas são ajustadas como *Prioris*.

O ajuste dele está muito entrelaçado com os **CAGM**, mesmo sendo uma ampla gama de modelos preditivos que se enquadram nessa categoria os **DLM's** que se enquadram são somente os Gaussianos. **HMM** é um deles, mas como o ajuste desses modelos por **MCMC** é muito proibitivo, o **INLA** é a saída mais recomendável.

Os materiais que explicam como o **INLA** funciona são difíceis de encontrar e muito pouco didáticos, o que torna o uso dele mais restrito.

Para a Inferência *Bayesiana* o **INLA** uma ferramenta que se tornará um de seus pilares, juntamente com o **MCMC**. Depois que as primeiras barreiras são ultrapassadas, a facilidade para o uso começa a aparecer.

Na área econômica, a análise *Bayesiana* ainda é pouco explorada. O **INLA** seria um bom começo, pela simplicidade de implementação do **RINLA**. Podem ser descobertas, a partir das *Posterioris* dos modelos ajustados, novas interpretações econômicas.

Referências

- CARDENAS, R. R.; KRAINSKI, E. T.; RUE, H. Fitting dynamic models using integrated nested laplace approximations – inla. 2010. Citado na página 19.
- CORTES, R. X. *Dissetação: Estimando modelos dinâmicos utilizando o INLA para Campos Aleatórios Markovianos Não Gaussianos*. Dissertação (Mestrado), 2014. Citado na página 19.
- D. CEMGIL A.T., C. S. e. B. *Bayesian Time Series Models*. [S.l.]: Cambridge University Press, 2011. ISBN 978-0-521-19676-5. Citado na página 7.
- E.P.BOX, G.; M.JENKINS, G. *Time Series Analysis: Forecasting and Control (Revised Edition)*. Revised. [S.l.]: Holden-Day, 1976. ISBN 0816211043,9780816211043. Citado na página 6.
- KINAS, P. G.; ANDRADE, H. A. *Introdução à Análise Bayesiana (Com R)*. 1º edição. ed. [S.l.]: MaisQnada Editora, 2010. ISBN 9788561797102. Citado na página 7.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. [S.l.]: Edgard Blucher, 2006. ISBN 8521203896,9788521203896. Citado na página 5.
- RUE, H.; FERKINGSTAD, E. *INLA Project*. [S.l.], 2017. [Http://www.r-inla.org/models/latent-models](http://www.r-inla.org/models/latent-models). Disponível em: www.r-inla.org. Citado 2 vezes nas páginas 18 e 20.
- RUE, H.; HELD, L. *Gaussian Markov Random Fields: Theory and Applications*. [S.l.: s.n.], 2005. ISBN 1584884320. Citado 7 vezes nas páginas 7, 11, 12, 13, 16, 18 e 21.
- RUE, H.; MARTINO, S. Approximate bayesian inference for hierarchical gaussian markov random fields models. 2005. Citado 3 vezes nas páginas 7, 16 e 18.
- RUE, H.; MARTINO, S.; CHOPIN, N. Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations. 2009. Citado 2 vezes nas páginas 16 e 18.
- SMEDT, T. D. et al. Comparing mcmc and inla for disease mapping with bayesian hierarchical models. 2015. Citado na página 18.
- TIERNEY, L.; KADANE, J. Accurate approximations for posterior moments and marginal densities. 1986. Disponível em: <https://www.cs.cmu.edu/~tom/10-702/laplace.pdf>. Citado 2 vezes nas páginas 15 e 16.
- WEST, M.; HARRISON, J. *Bayesian Forecasting and Dynamic Models*. [S.l.]: Springer Verlag, 1989. ISBN 0-387-97025-8. Citado na página 5.

Apêndices

APÊNDICE A – Script completo R

```
##-----
#Setando rea de Trabalho, Carregando Pacotes
library(INLA)
library(ggplot2)
require("scales")
library(gridExtra)

setwd("C:/Users/imef/Google_Drive/wdr/TCC")#Setar Diretório
rm(list=ls()) #Limpar Tudo
pescado<- read.csv2("Bagre.csv", stringsAsFactors = F) #Carregar Arquivos
pescado$DATA<-as.Date(pescado$DATA, format = "%d/%m/%y" ) #Ajustar Formato

##----- DADOS AUSENTES POR COLETA -----
nas<-data.frame(matrix(0,1,5));
colnames(nas)<-c("Inteiro_Sujo","Inteiro_Limpo","Fil ",
"Postas","Seco_Salgado");
row.names(nas)<- "Ausentes_"
for(i in 1:5){
  nas[1,i]=sum(is.na(pescado[,i])*1)/length(pescado[,i])
}
nas
#----- SERIE ORIGINAL -----
names(pescado)
p1<-ggplot(pescado, aes(x=DATA))+
  geom_point(aes( y=INTEIRO.SUJO),
  colour="red")+ ggtitle( "Inteiro_Sujo" ) +
  xlab("Ano") + ylab("Pre o_em_Reais") +
  scale_x_date(breaks = date_breaks("1_years"),
  labels = date_format("%y"));

p2<-ggplot(pescado, aes(x=DATA))+
  geom_point(aes( y=INTEIRO.LIMPO),
  colour="red")+ ggtitle( "Inteiro_Limpo" ) +
  xlab("Ano") + ylab("Pre o_em_Reais") +
  scale_x_date(breaks = date_breaks("1_years"),
  labels = date_format("%y"));
```

```

p3<-ggplot(pescado, aes(x=DATA))+
  geom_point(aes( y=FIL ), colour="red")+
  ggtitle( "Fil " ) +
  xlab("Ano") + ylab("Pre o em Reais") +
  scale_x_date(breaks = date_breaks("1 years"),
  labels = date_format("%y"));

p4<-ggplot(pescado, aes(x=DATA))+
  geom_point(aes( y=POSTAS), colour="red")+
  ggtitle( "Postas" ) +
  xlab("Ano") + ylab("Pre o em Reais") +
  scale_x_date(breaks = date_breaks("1 years"),
  labels = date_format("%y"));

p5<-ggplot(pescado, aes(x=DATA))+
  geom_point(aes( y=SECO.SALG),
  colour="red")+ ggtitle( "Seco Salgado" ) +
  xlab("Ano") + ylab("Pre o em Reais") +
  scale_x_date(breaks = date_breaks("1 years"),
  labels = date_format("%y"));

grid.arrange(p1,p2,p3,p4,p5)

##----- INLA -----
n<-length(pescado$DATA);t <- 1:n # indices for x_t
formula <-INTEIRO.SUJO ~f(t, model="ar", order=3 ,constr=F,
prior = "normal", param = c(mean=0,prec=1e-5)) -1
saida <- inla(formula, data = data.frame(t,pescado),
family = "gaussian", control.compute = list(dic=T), verbose=F)
summary(saida)
saida$dic$dic

##----- PLOT COM I.CRED -----
banco <- subset(banco, select = c("dados","data"))
estimado<-as.data.frame(saida$summary.random$i)
banco <- cbind(banco, estimado) #salvar como data frame
p<-ggplot(banco, aes(x=data )) +
  geom_line(aes(y='0.5quant', colour="Quantil_0.5"))+
  geom_line(aes(y='0.025quant', colour="Quantil_0.025"),
  linetype="dashed", size=1) +
  geom_line(aes(y='0.975quant', colour="Quantil_0.975"),
  linetype="dashed", size=1 )+

```

```

    geom_point(aes(y=dados, colour="Observados"), size=1) +
    scale_colour_manual("", breaks =
    c("Quantil_0.5", "Quantil_0.025", "Quantil_0.975", "Observados"),
values = c("Quantil_0.5"="black", "Quantil_0.025"="blue",
"Quantil_0.975"="blue", "Observados"="red")) +
    xlab("Ano") + ylab("Pre o em Reais") + labs(fill = "S RIE") +
    ggtitle("Bagre AR3") +
    scale_x_date(breaks = date_breaks("1_years"),
    labels = date_format("%y")); p2

#----- COMPLETANDO COM O PONTO M DIO -----
completo<-dados
rm<-matrix(NA,length(dados))
for(i in 1:length(dados)){if(is.na(dados))
    completo[i]<-r$summary.random$i$'0.5quant'[i] }
for(i in 1:length(dados)){if(is.na(dados))
    rm[i]<-r$summary.random$i$'0.5quant'[i] }
pescado$completo<-completo;pescado$rm<-rm
#----- POSTERIORIS -----
#Distribui o dos Pontos..
pos.dens<-data.frame(saida$marginals.fitted.values[13],
saida$marginals.fitted.values[25])
ggplot(pos.dens)+
geom_line(size=1 ,
aes(fitted.predictor.013.x,fitted.predictor.013.y,colour="31/08/07"))+
geom_line(size=1 ,
aes(fitted.predictor.025.x,fitted.predictor.025.y,colour="30/08/08"))+
scale_x_continuous(limits = c(0,7.5))+
scale_colour_manual("Meses",breaks = c("31/08/07","30/08/08"),
values = c("31/08/07"="blue", "30/08/08"="red")) +
labs(x="", y="")+ggtitle("Posterioris para as observa es")

#Poterioris hyperpar metros
p1<-ggplot(data.frame(
saida$marginals.hyperpar$'Precision for the Gaussian observations'),
aes(x,y))+
geom_line(size=1 , colour="blue", show.legend = NA)+
    scale_x_continuous(limits = c(0,10))+
    ggtitle("Precis o para V")
p2<-ggplot(
data.frame(saida$marginals.hyperpar$'Precision for i'),aes(x,y))+

```

```
  geom_line(size=1 , colour="red", show.legend = NA)+
  ggtitle("Precis o para W")+
  scale_x_continuous(limits = c(0,0.5))
p3<-ggplot(data.frame(
saida$marginals.hyperpar$'Rho for i'),aes(x,y))+
  geom_line(size=1 , colour="red", show.legend = NA)+
  ggtitle("Precis o para Rho")+
  scale_x_continuous(limits = c(0.99,1))
grid.arrange(p1,p2,p3)
```
