ERHA7016 – Hidrologia Estocástica

Geração de Séries Sintéticas

Análise de Séries Temporais (pt. 3)

Geração de Séries Sintéticas (pt. 1)

Daniel Detzel
detzel@ufpr.br

Imagem: bit.ly/2WhSPdF

Agenda

Análise de séries temporais

decomposição

Geração de séries sintéticas

modelos ARIMA

modelos ARIMA: identificação

Análise de séries temporais

decomposição

Análise de séries temporais | decomposição

Uma série temporal z_t pode ser compreendida como a composição de três termos:

$$z_t = S_t + T_t + a_t$$

onde

S_t	componente sazonal
T_t	componente de tendência/ciclo
a_t	componente residual

Análise de séries temporais | decomposição

O equacionamento mostrado assume uma forma aditiva de representação

há também a forma multiplicativa, mas é menos comum em séries hidrológicas

A forma do modelo pode variar de acordo com a escala considerada

ex.: para série anuais, a componente S_t é inexistente

A componente residual a_t representa tudo aquilo que as demais componentes não conseguem extrair da série

idealmente possui características aleatórias

depende dos modelos utilizados para representar S_t e T_t

Análise de séries temporais | decomposição

Portanto, a decomposição de uma série diz respeito à aplicação de técnicas estatísticas para isolar S_t e T_t do sinal original z_t

Aqui são consideradas três técnicas

- decomposição clássica
- decomposição STL
- decomposição CEEMDAN

As formulações detalhadas de cada método podem ser conferidas em:

- decomposições clássica e STL: https://otexts.com/fpp3/decomposition.html
- CEEMDAN: *[Zhang et al., 2022]*

Análise de séries temporais | decomposição

Decomposição clássica:

Procedimento simplificado que assume:

sazonalidade constante no decorrer dos anos

ciclo/tendência modelado a partir de médias móveis

1. Estimar a componente $\hat{T}_t$

utilizar média móvel com janela equivalente a um ciclo sazonal completo

ex.: para séries mensais, adota-se janela temporal de 12 meses

Análise de séries temporais | decomposição

2. Remover a componente de tendência do modelo

fazer: $z_t - \hat{T}_t$

a série resultante possui as componentes S_t e a_t

3. Estimar a componente $\hat{S}_t$

determinar a média de cada período sazonal da série $(z_t - \hat{T}_t)$

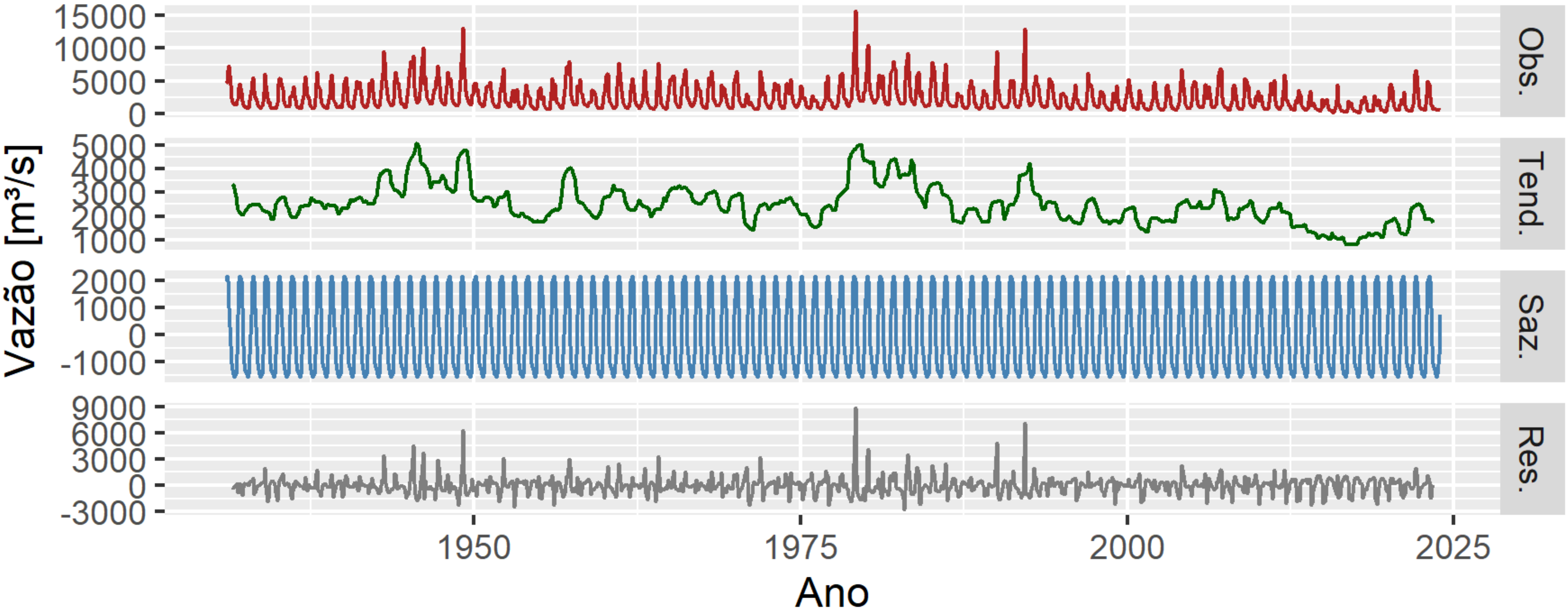
ex.: para séries mensais, calcular as médias de todos os janeiros, depois todos os fevereiros, etc.

4. Estimar a componente $\hat{a}_t$

fazer: $\hat{a}_t = z_t - \hat{T}_t - \hat{S}_t$

Análise de séries temporais | decomposição

[Exemplo] Decomposição clássica no rio São Francisco (usina Sobradinho)



Decomposição STL:

STL: "*seasonal and trend decomposition using Loess*"

permite com que o comportamento sazonal varie com o passar dos anos
ciclo/tendência pode ser não linear
considera que a componentes S_t e T_t estão interrelacionadas

Loess: "*locally weighted scatterplot smoothing*"

é um método de regressão local
em uma série, o método aplica sucessivas regressões em janelas móveis
são ponderadas de modo que os pontos mais próximos do centro da janela
tenham maior representatividade

Análise de séries temporais | decomposição

A decomposição STL aplica um processo iterativo alternando ajustes Loess distintos para as componentes S_t e T_t

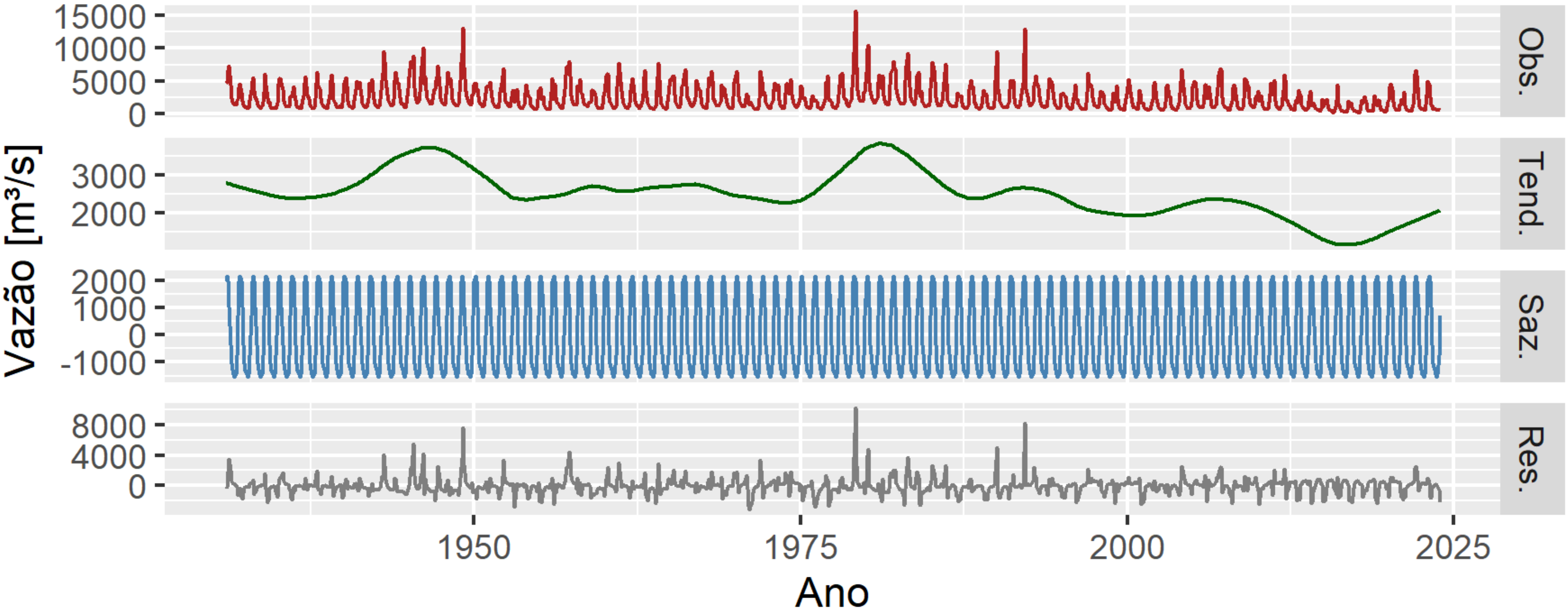
1ª iteração:
idem à decomposição clássica para obter $\hat{S}_t^1$ e $\hat{T}_t^1$

2ª iteração:
obtém $z_t - \hat{T}_t^1$ e aplica Loess para determinar $\hat{S}_t^2$;
obtém $z_t - \hat{S}_t^2$ e aplica Loess para determinar $\hat{T}_t^2$

3ª iteração:
obtém $z_t - \hat{T}_t^2$ e aplica Loess para determinar $\hat{S}_t^3$;
obtém $z_t - \hat{S}_t^3$ e aplica Loess para determinar $\hat{T}_t^3$

[repete até a convergência]

[Exemplo] Decomposição STL no rio São Francisco (usina Sobradinho)



Decomposição CEEMDAN:

CEEMDAN: "*complementary ensemble empirical mode decomposition with adaptive noise*"

variante aprimorada do método EMD (*empirical mode decomposition*)

Diferentemente das técnicas anteriores, o CEEMDAN assume que a série possui mais componentes do que S_t e T_t

as componentes adicionais são consideradas padrões oscilatórios denominados IMFs (*intrinsic mode functions*)

Análise de séries temporais | decomposição

Portanto, a decomposição pode ser representador por:

$$z_t = IMF_t^1 + IMF_t^2 + \cdots + IMF_t^k + a_t$$

O processo de determinação das IMFs continua até que o que resta da série possa ser considerado como resíduo

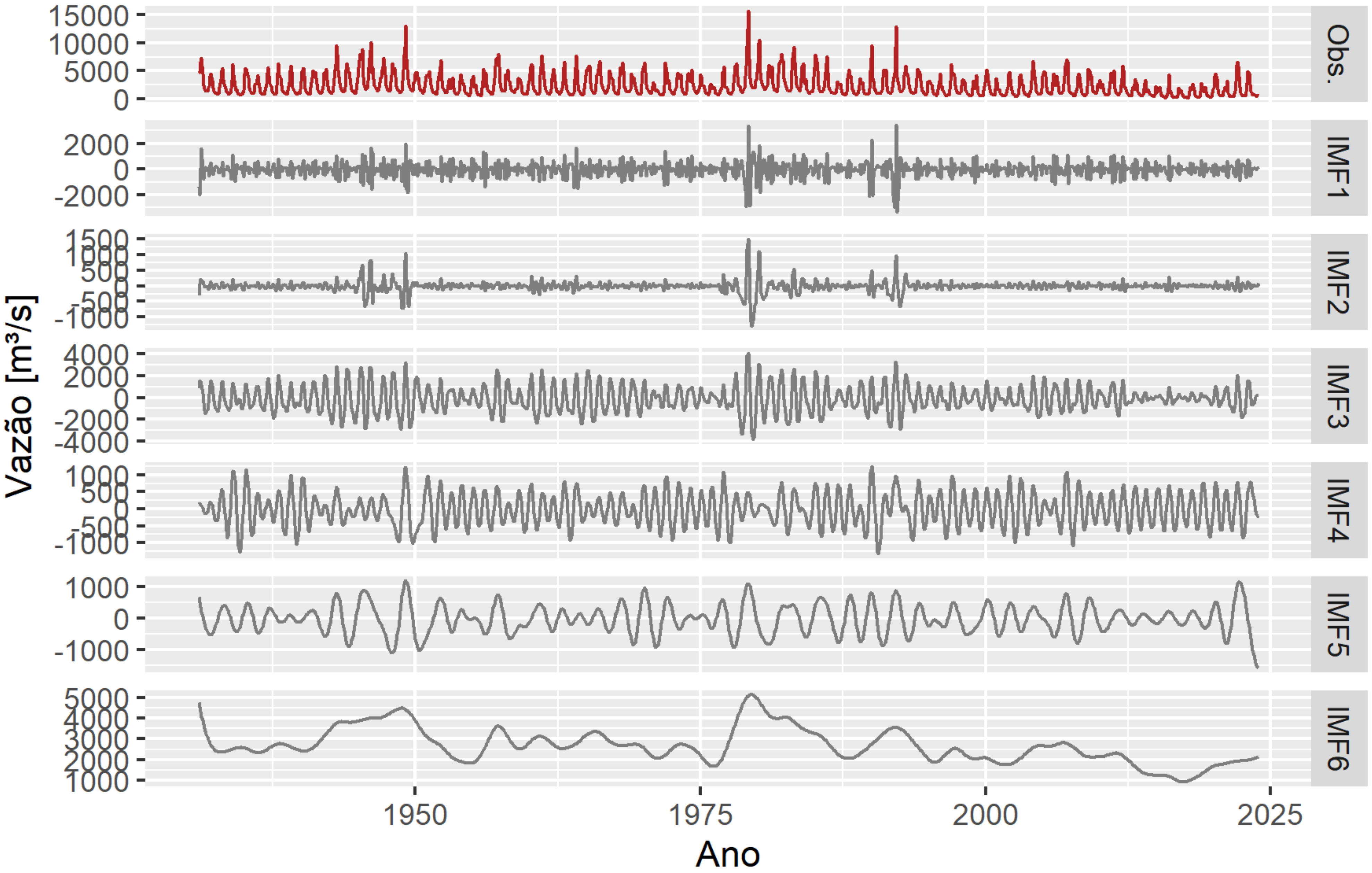
o próprio algoritmo faz isso automaticamente

A partir de técnicas numéricas adicionais, é possível converter cada IMF em um frequência no domínio do tempo

ex.: oscilação com período de X anos

ver *[Antico et al., 2016]*

Análise de séries temporais | decomposição



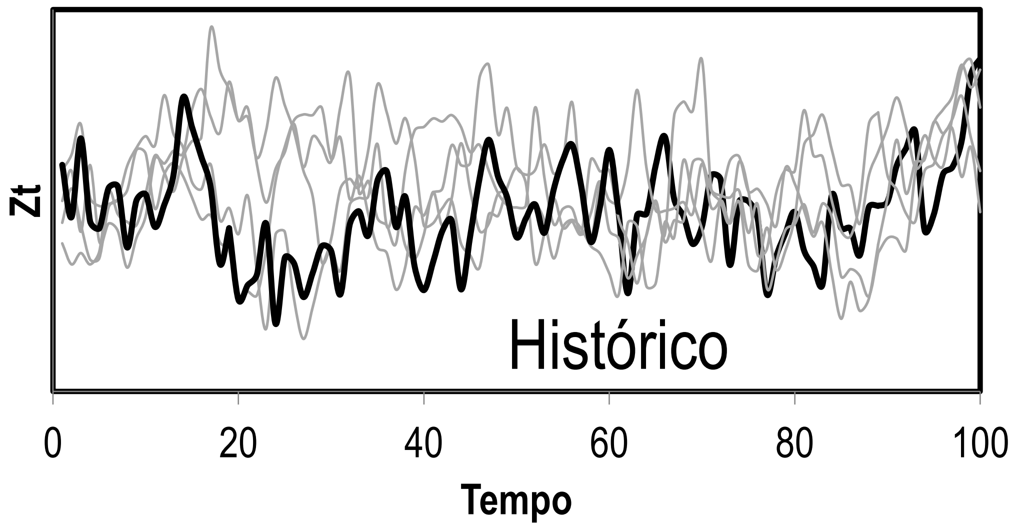
[Exemplo]
Decomposição CEEMDAN no rio São Francisco
(usina Sobradinho)

Geração de Séries Sintéticas

modelos ARIMA

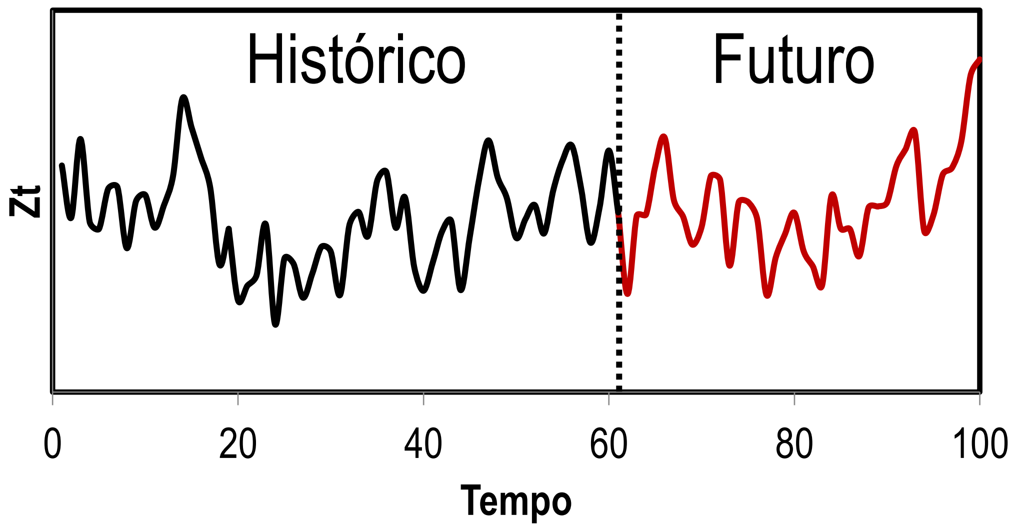
Geração de séries sintéticas | modelos ARIMA

Geração de séries sintéticas ≠ Previsão de séries



Geração de séries sintéticas

Previsão de uma série



Geração de séries sintéticas | modelos ARIMA

Principais modelos usados para geração de séries sintéticas

- modelos estocásticos lineares (ARIMA) *[Box et al., 2008]*
- *fractional gaussian noise* (FGN) *[Mandelbrot & Wallis, 1969]*
- reamostragem via bootstrap *[Srivastav & Simonovic, 2014]* ou *k*-vizinhos mais próximos *[Salas & Lee, 2010]*
- baseados em conceitos de entropia *[Hao & Singh, 2011]*
- redes neurais artificiais *[Can et al., 2012]*
- inferência *fuzzy* *[Luna et al., 2011]*
- *wavelets* *[Niu et al., 2013]*
- modelo de múltiplas escalas temporais (SSS) *[Koutsoyiannis, 2002]*

Geração de séries sintéticas | modelos ARIMA

Modelo de interesse: ARIMA

- proposto por George Box e Gwilym Jenkins em 1970, com grande difusão a partir da 2ª edição de 1976*
- o mais popular para a geração de séries sintéticas de vazão
- exploram a estrutura de persistência da série

Sigla para modelo AutoRregressivo Integrado de Médias Móveis, baseia-se na combinação linear entre três elementos

- componente autorregressivo: AR
- fator de integração: I
- componente de médias móveis: MA

---

*Box, G. E. P.; Jenkins, G. M. (1976). **Time Series Analysis: Forecasting and Control** (Revised ed.). San Francisco, CA: Holden-Day

Geração de séries sintéticas | modelos ARIMA

Para cada uso, o modelo deve ser construído por meio dos seguintes passos:

1. Identificação do modelo

Determinar quais componentes farão parte da formulação

2. Estimação dos parâmetros

Obter os valores dos parâmetros envolvidos

3. Validação do modelo

Avaliar, teoricamente, se o ajuste é satisfatório

Geração de séries sintéticas | modelos ARIMA

Originalmente, o processo era iterativo

caso o modelo não fosse válido, retornava-se ao passo 1 para determinar um novo modelo

Atualmente é possível aplicar diferentes métodos construir um modelo ARIMA sem iterações

técnicas calcadas no princípio da parcimônia

"A técnica de Box e Jenkins requer experiência e algum conhecimento além do uso automático de um pacote de computador"

*[Morettin & Toloi, 2019]**

*Morettin, P.; Toloi, C. (2019). **Análise de series temporais**. 3ª Ed. São Paulo: Blücher.

GERAÇÃO DE SÉRIES SINTÉTICAS

modelos ARIMA: identificação

Geração de séries sintéticas | modelos ARIMA: identificação

Objetivo: determinar os componentes da formulação e as suas ordens

A representação geral do modelo é: ARIMA (p, d, q)

- o componente autorregressivo AR possui ordem p
- o fator de integração I possui ordem d
- o componente de médias móveis MA possui ordem q

Uma série pode ser modelada pelos três elementos em conjunto, ou apenas parte deles

Geração de séries sintéticas | modelos ARIMA: identificação

Equação geral do modelo

$$\underbrace{\left(1-\varphi_1 B-\varphi_2 B^2-\cdots-\varphi_p B^p\right)}_{\text{AR}}\underbrace{(1-B)^d}_{\text{I}} z_t=\underbrace{\left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\right)}_{\text{MA}} a_t$$

onde

B operador defasagem tal que $z_{t-k}=B^k z_t$

z_t variável aleatória no tempo t

a_t resíduo no tempo t

Geração de séries sintéticas | modelos ARIMA: identificação

Modelos de quaisquer ordens são obtidos a partir da equação geral

Exemplos:

AR(1): $z_t - \varphi_1 z_{t-1} = a_t$

ARMA(1,1): $z_t - \varphi_1 z_{t-1} = a_t - \theta_1 a_{t-1}$

ARIMA(1,1,1): $(z_t - z_{t-1}) - \varphi_1 (z_{t-1} - z_{t-2}) = a_t - \theta_1 a_{t-1}$

Essencialmente, a etapa de identificação busca:

avaliar a pertinência de incluir as porções AR, I e/ou MA

definir as ordens dos parâmetros p, d e q

Geração de séries sintéticas | modelos ARIMA: identificação

Método convencional: análises das funções de autocorrelação (FAC) e autocorrelação parcial (FACP)

FAC (correlograma) foi definida na aula 2

FACP (correlograma parcial) é definida por:

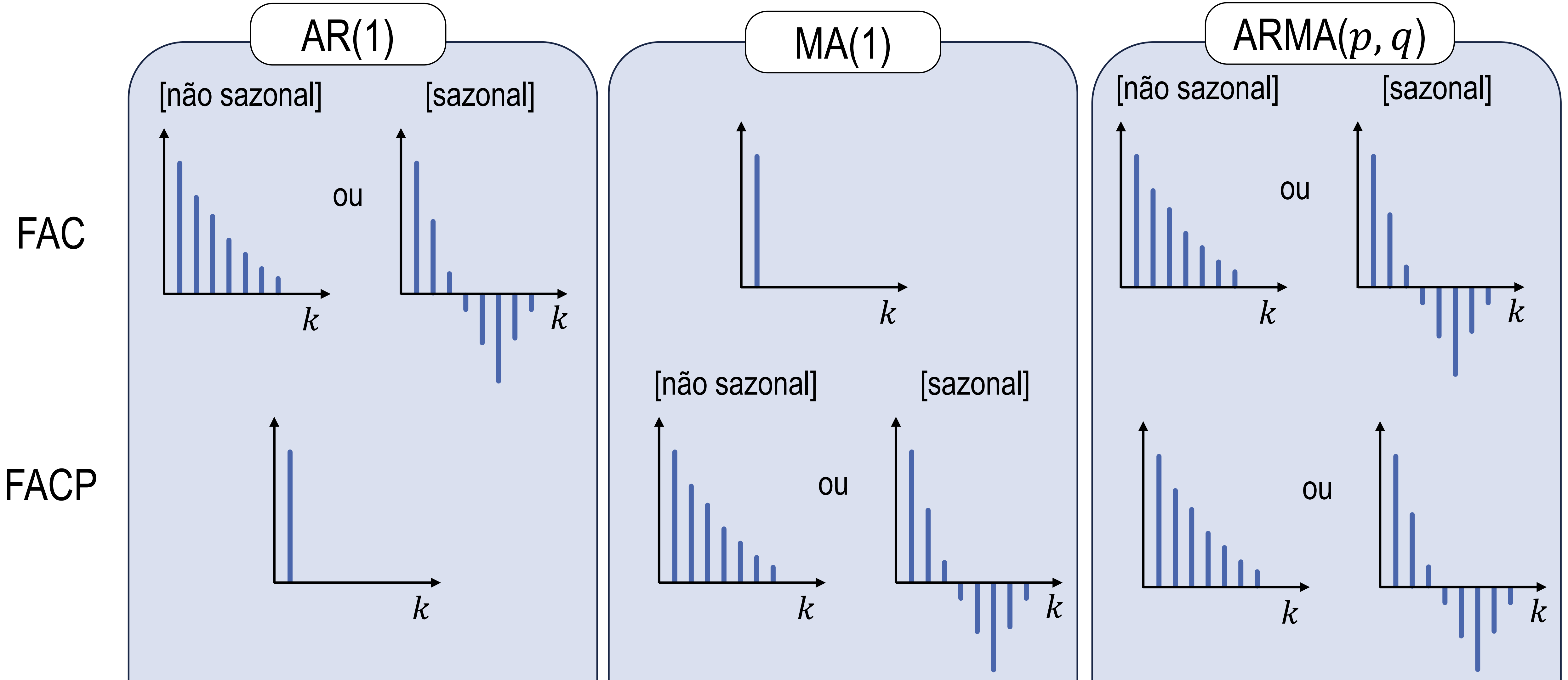
$$\rho_j = \varphi_{k1} \cdot \rho_{j-1} + \varphi_{k2} \cdot \rho_{j-2} + \cdots + \varphi_{k(k-1)} \cdot \rho_{j-k+1} + \varphi_{kk} \cdot \rho_{j-k}$$

Equivale ao ajuste de sucessivos modelos autorregressivos de ordens $p = 1,2, \ldots, k$ às séries

a FACP equivale aos últimos parâmetros φ_{kk} obtidos em cada ajuste

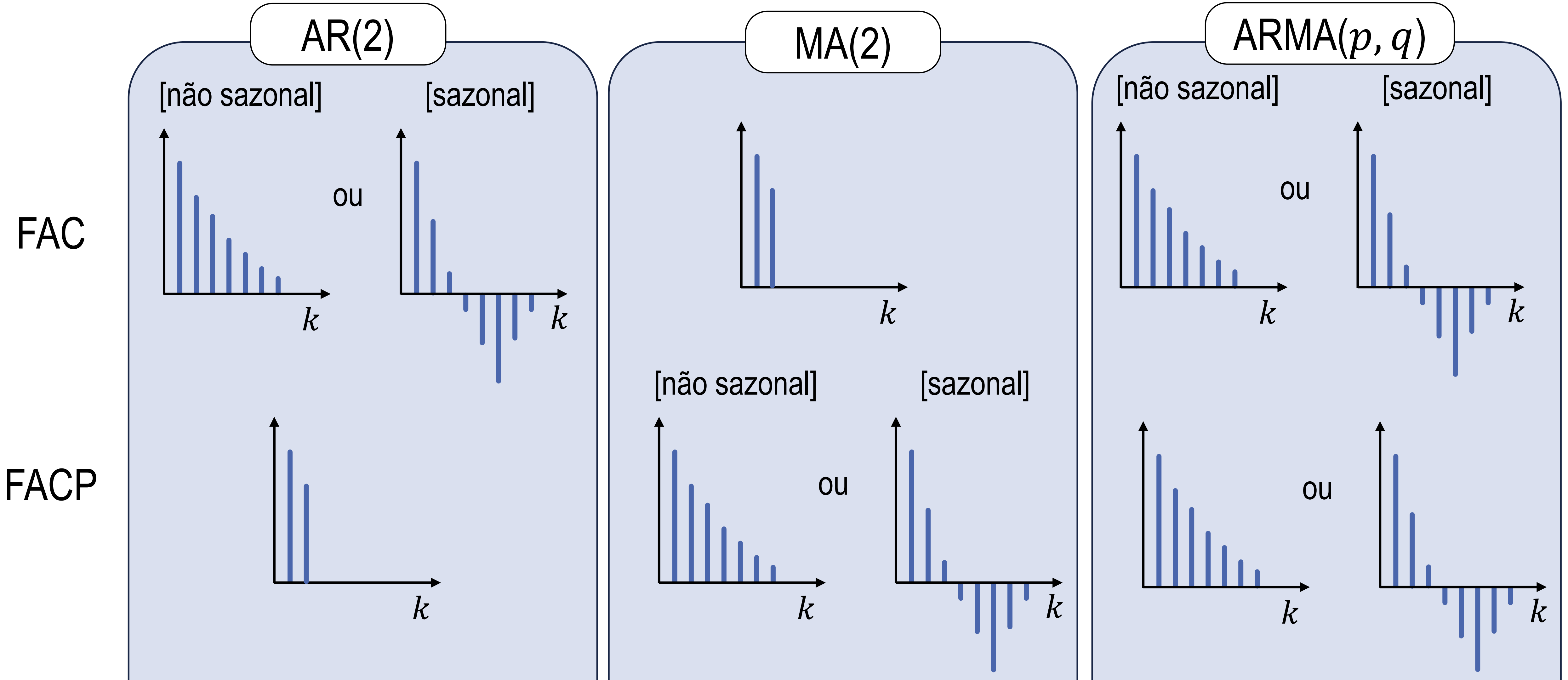
Geração de séries sintéticas | modelos ARIMA: identificação

Comportamentos teóricos conhecidos para FAC e FACP *[1ª ordem]*:



Geração de séries sintéticas | modelos ARIMA: identificação

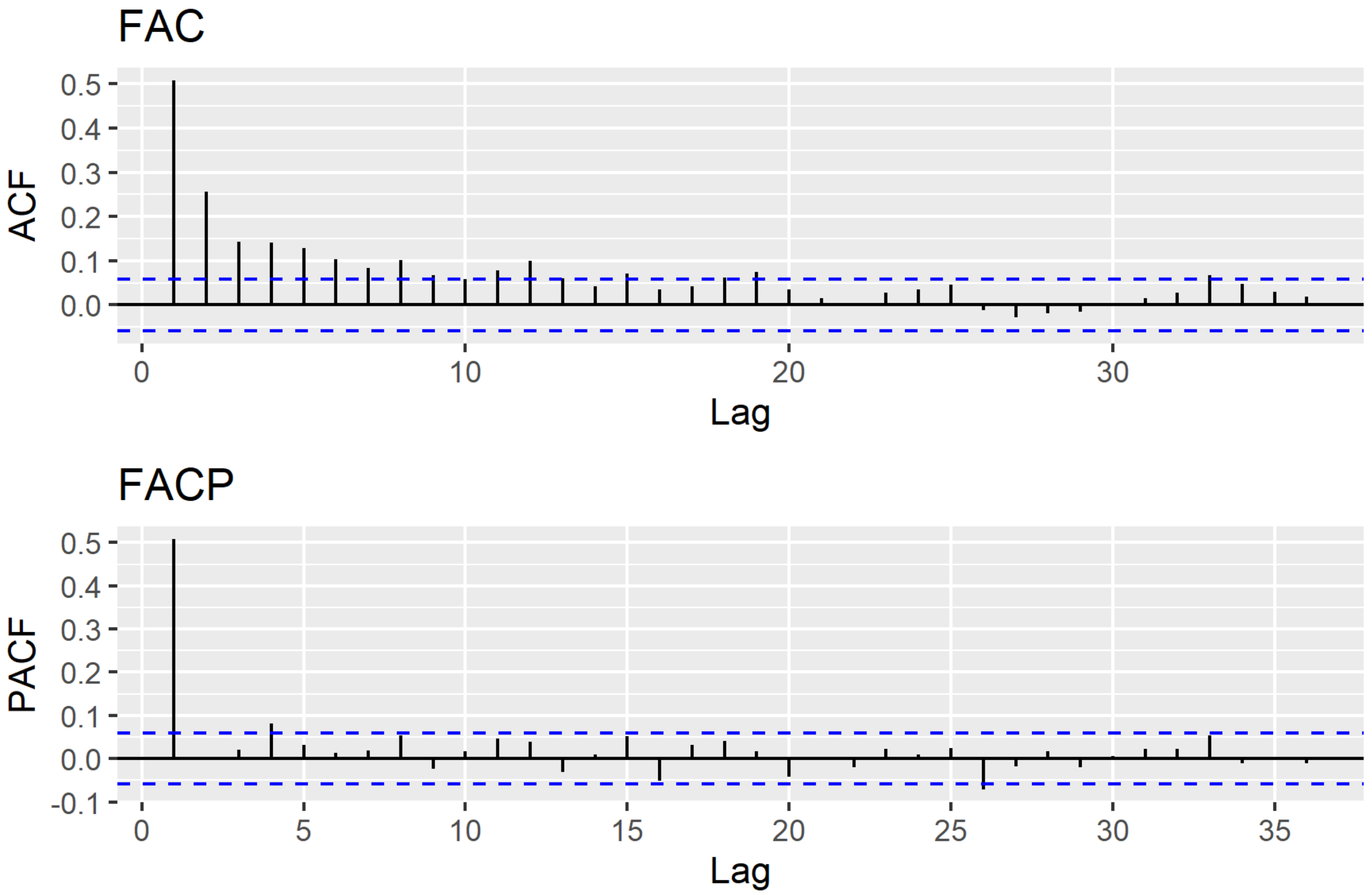
Comportamentos teóricos conhecidos para FAC e FACP *[2ª ordem]*:



Geração de séries sintéticas | modelos ARIMA: identificação

Por ser um critério visual, as análises via FAC e FACP podem ser subjetivas

Algumas séries são claras...
Ex.: Salto Caxias (rio Iguaçu)

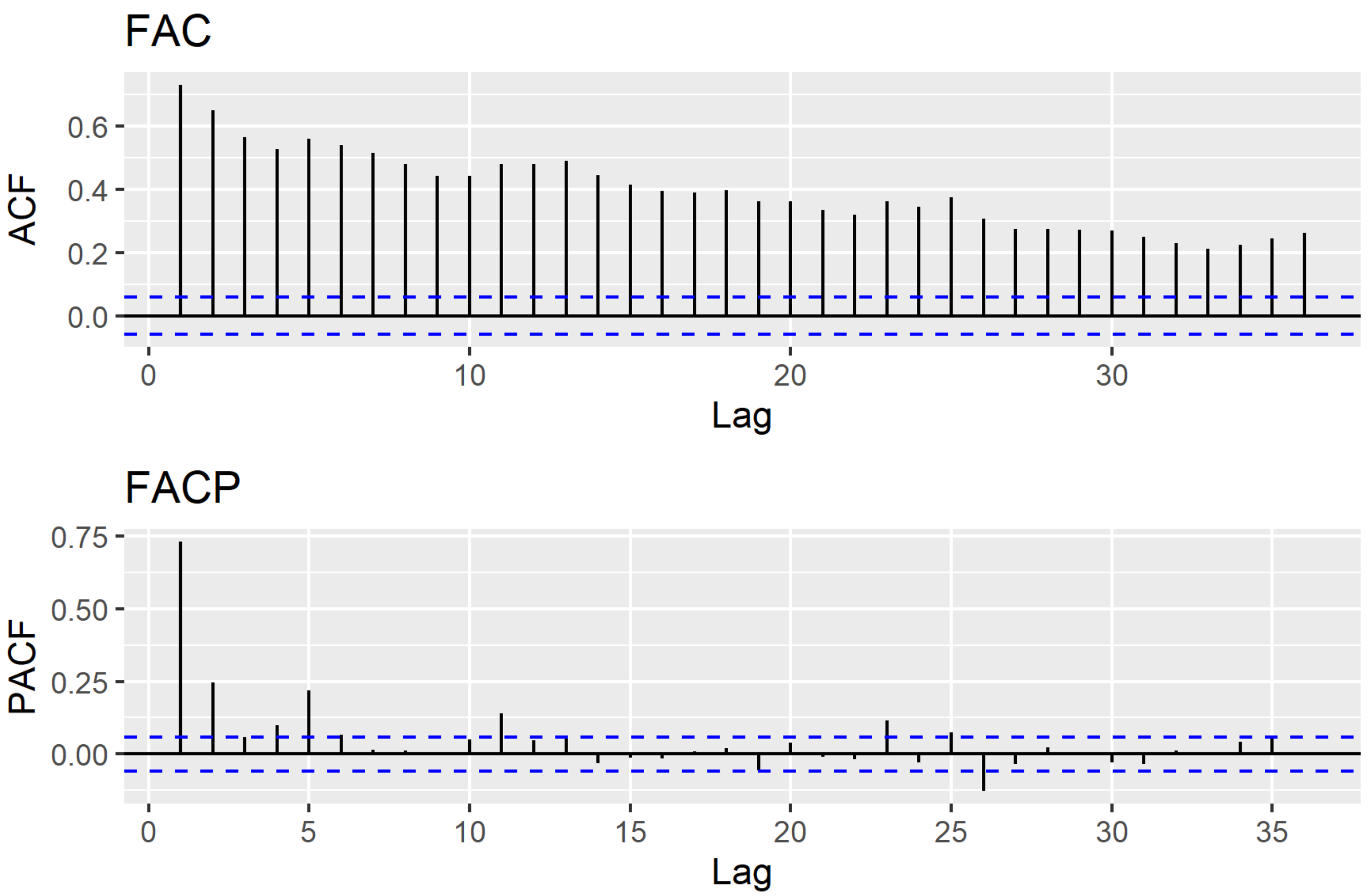


Geração de séries sintéticas | modelos ARIMA: identificação

Por ser um critério visual, as análises via FAC e FACP podem ser subjetivas

Outras nem tanto...

Ex.: Itiquira II (rio Verde – MT)



Geração de séries sintéticas | modelos ARIMA: identificação

Uma alternativa é calcular os chamados Critérios de Informação

procedimentos matemáticos baseados em funções penalizadoras dependentes do número de parâmetros do modelo

Critérios de Akaike (AIC) e Akaike corrigido (AICc)

objetivo: encontrar o modelo que mais bem representa a série

Critério Bayesiano (BIC)

objetivo: encontrar o modelo mais parcimonioso

São calculados para todos os modelos candidatos. O modelo selecionado é o que resulta no mínimo AIC/AICc/BIC

Geração de séries sintéticas | modelos ARIMA: identificação

Seja k o número total de parâmetros do modelo $(p + q + 1)$

o parâmetro adicional se refere à variância dos resíduos do ajuste $(\hat{\sigma}_a^2)$

Os resíduos podem ser entendidos como erros do ajuste

considerando estimadores de máxima verossimilhança, tem-se:

$$\hat{\sigma}_a^2 = \frac{1}{n}\sum_{t=1}^{n} a_t^2$$

onde

a_t diferença entre os valores observados z_t e estimados $\hat{z}_t$ $(a_t = z_t - \hat{z}_t)$

Geração de séries sintéticas | modelos ARIMA: identificação

Critério de Akaike (AIC) *[Akaike, 1973]**

$$AIC(k) = n \ln \hat{\sigma}_a^2 + 2k$$

Critério de Akaike corrigido (AICc) *[Hurvich & Tsai, 1989]***

$$AICc(k) = AIC(k) + \frac{2(k)(k+1)}{n-k-1}$$

para $n/k < 40$

---

*Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), Second International Symposium on Information Theory (pp. 267–281). Budapest: Akademiai Kiado.

** Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

Geração de séries sintéticas | modelos ARIMA: identificação

Critério Bayesiano (BIC) *[Schwartz, 1978]**

$$BIC(k) = n \ln \hat{\sigma}_a^2 + k \ln n$$

Quanto às penalidades:

AIC/AICc: fixa (baseada somente em k)

BIC: variável conforme o tamanho da amostra

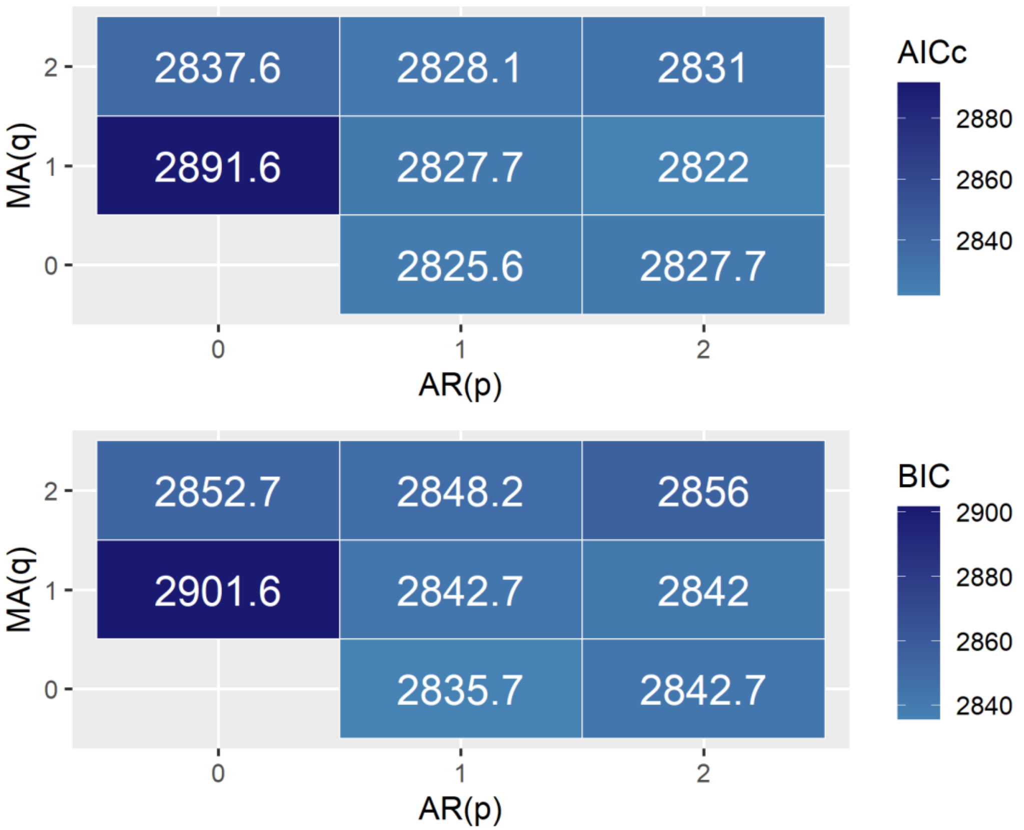
A tendência é o BIC priorizar simplicidade sobre qualidade de ajuste e vice-versa

*Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6(2), 461–464. https://doi.org/10.1214/aos/1176344136

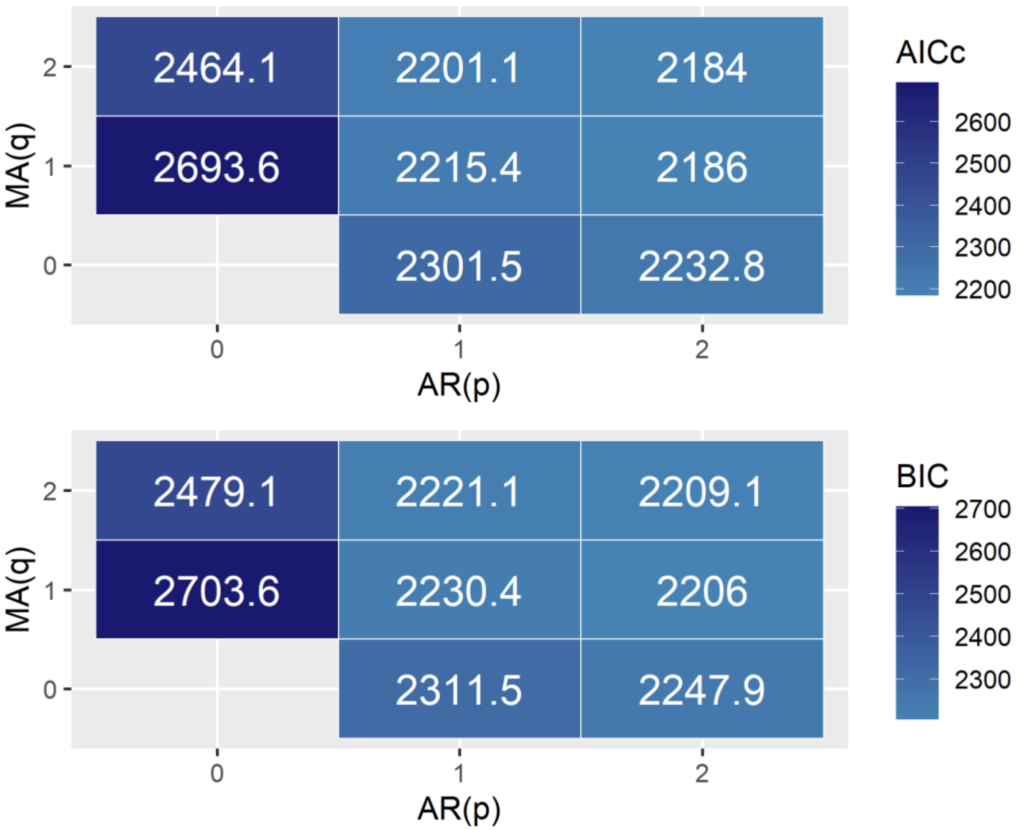
Geração de séries sintéticas | modelos ARIMA: identificação

[Exemplo] Determinação dos critérios de informação para Salto Caxias e Itiquira II.

Salto Caxias



Itiquira II



Resumo

Técnicas de decomposição de séries podem ajudar no entendimento aprofundado dos seus componentes

Os modelos de interesse para geração de séries sintéticas são da classe ARIMA (modelos estocásticos lineares de Box & Jenkins)

- representam muito bem a persistência das séries
- a construção do modelo passa pelas etapas: identificação, estimação, validação

Identificação: encontrar quais as componentes estarão presentes e as suas ordens

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

PPGERHA

ERHA7016 – Hidrologia Estocástica

Daniel Detzel
detzel@ufpr.br