Imagem: bit.ly/2XhcVWd

ERHA7016 – Hidrologia Estocástica

Introdução (pt. 1)

Daniel Detzel
detzel@ufpr.br

Agenda

Apresentação da disciplina

Conceitos iniciais

- definições
- natureza estocástica das variáveis hidrológicas

Revisão de estatística

- variáveis aleatórias
- distribuições de probabilidades
- testes de hipótese

APRESENTAÇÃO DA DISCIPLINA

Hidrologia Estocástica | conteúdos

Introdução: conceitos iniciais; natureza estocástica das variáveis hidrológicas; revisão de conceitos básicos de estatística; processos estocásticos e suas propriedades.

Análise de séries temporais: séries temporais; características das séries hidrológicas; persistência; tendências; transformações numéricas; decomposição de séries.

Geração de séries sintéticas (foco em vazões): modelos estocásticos lineares (Box & Jenkins); modelos sazonais; extensão para o caso multivariado; validação das séries sintéticas; incerteza.

Simulação hidrológica: método de Monte Carlo aplicado a problemas de hidrologia (dimensionamentos, curvas de regularização etc.).

Cadeias de Markov: Matrizes de transição de estados; teoria estocástica dos reservatórios; regularização interanual e intra-anual.

Regressão: modelo de regressão simples e múltipla; inferência e testes estatísticos; estimação de parâmetros; regionalização; aplicações em hidrologia.

Hidrologia Estocástica | avaliação

A avaliação é composta por:

1. Conjunto de atividades proposta nas aulas: peso 30%

    atividades semanais (ou quase!)

2. Trabalho: peso 70%

    entrega em formato de artigo/comunicação

Comunicação centrada no Teams da disciplina!

CONCEITOS INICIAIS

definições

Conceitos iniciais | definições

O que é um modelo?

Modelo: qualquer representação simplificada de um sistema

Modelo hidrológico: qualquer representação simplificada de um sistema hidrológico

- bacias hidrográficas
- rios
- reservatórios
- atmosfera, etc.

Conceitos iniciais | definições

Modelo de séries temporais: representação de observações de variáveis registradas em intervalos discretos de tempo

Um bom modelo deve *[Hipel & McLeod, 1994, p. 16]*:

- capturar matematicamente as principais características do sistema
- ser concebido de tal forma que possa ser compreendido, manipulado e interpretado

A construção de um modelo passa por duas etapas:

- análise exploratória de dados: estatísticas descritivas, gráficos, etc.
- análise confirmatória de dados: uso de modelos para melhor entender seu comportamento

Conceitos iniciais | definições

Modelos determinísticos e estocásticos

dois grandes ramos de modelos usados em problemas

considere o seguinte modelo:

$$y = f(x)$$

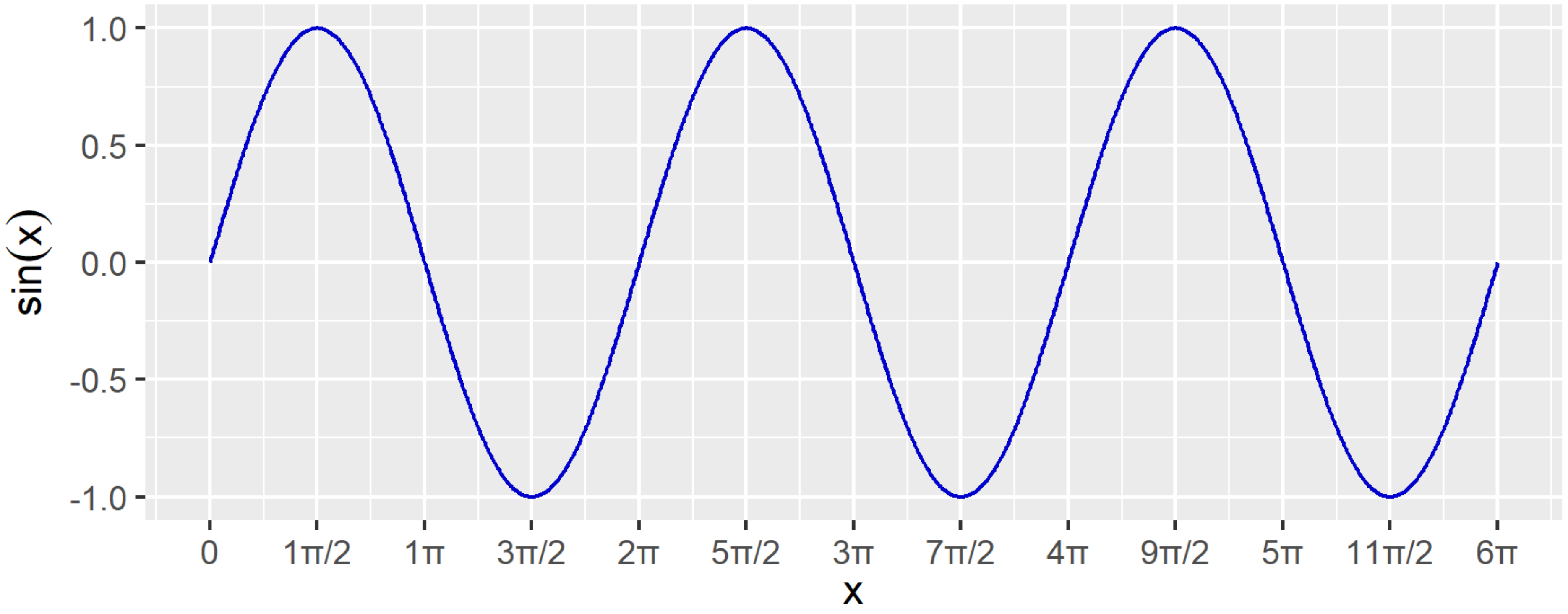
Se $f(\quad)$ consegue definir exatamente y → modelo determinístico

Se $f(\quad)$ for descrita com base em conceitos probabilísticos → modelo estocástico

nesse caso não se sabe precisamente o valor de y

Conceitos iniciais | definições

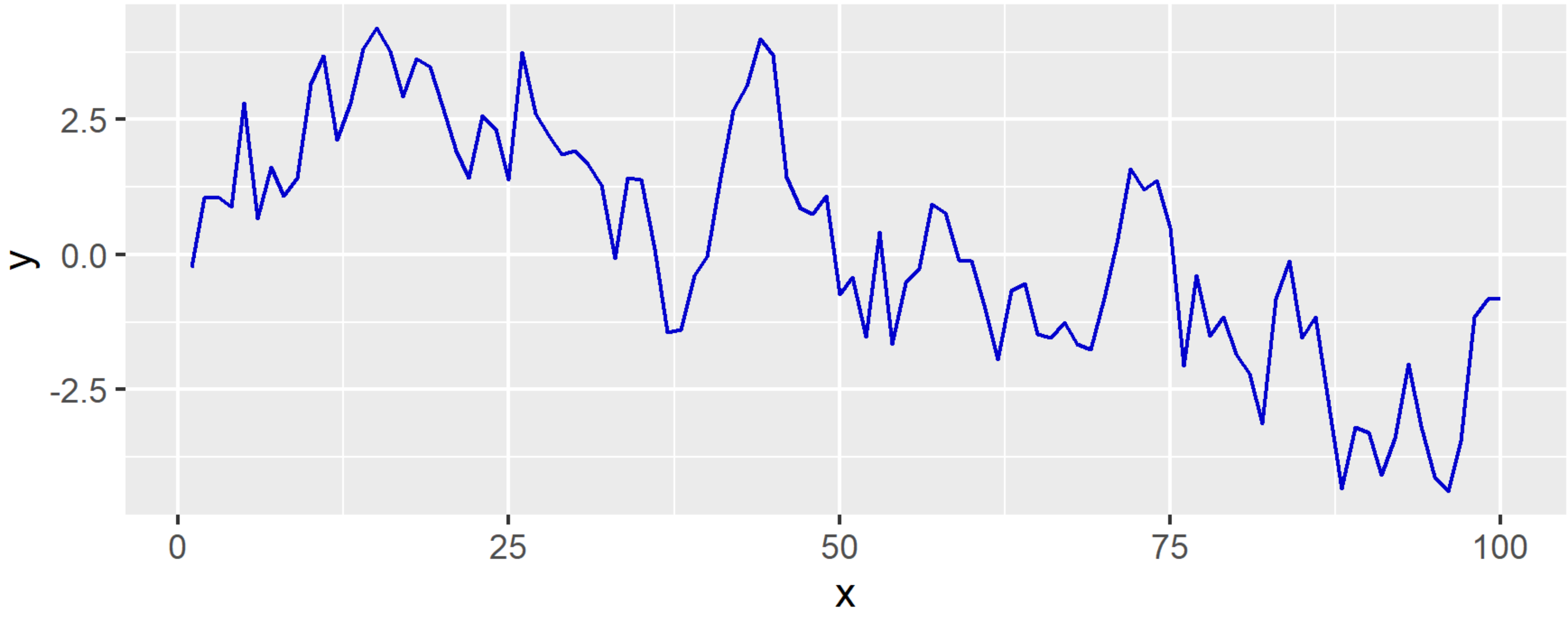
[Exemplo 1] O modelo $y = \sin(x)$ é determinístico:



A variável x consegue definir exatamente o valor de y.

Conceitos iniciais | definições

[Exemplo 2] O modelo $y = 0{,}8\,x + a$ é estocástico:



A variável x não consegue definir exatamente o valor de y devido à presença da variável aleatória a

Conceitos iniciais | definições

"A maioria dos fenômenos naturais (...) parecem se comportar aleatoriamente ou de acordo com leis probabilísticas"

[Hipel e McLeod, 1994, p. 23]

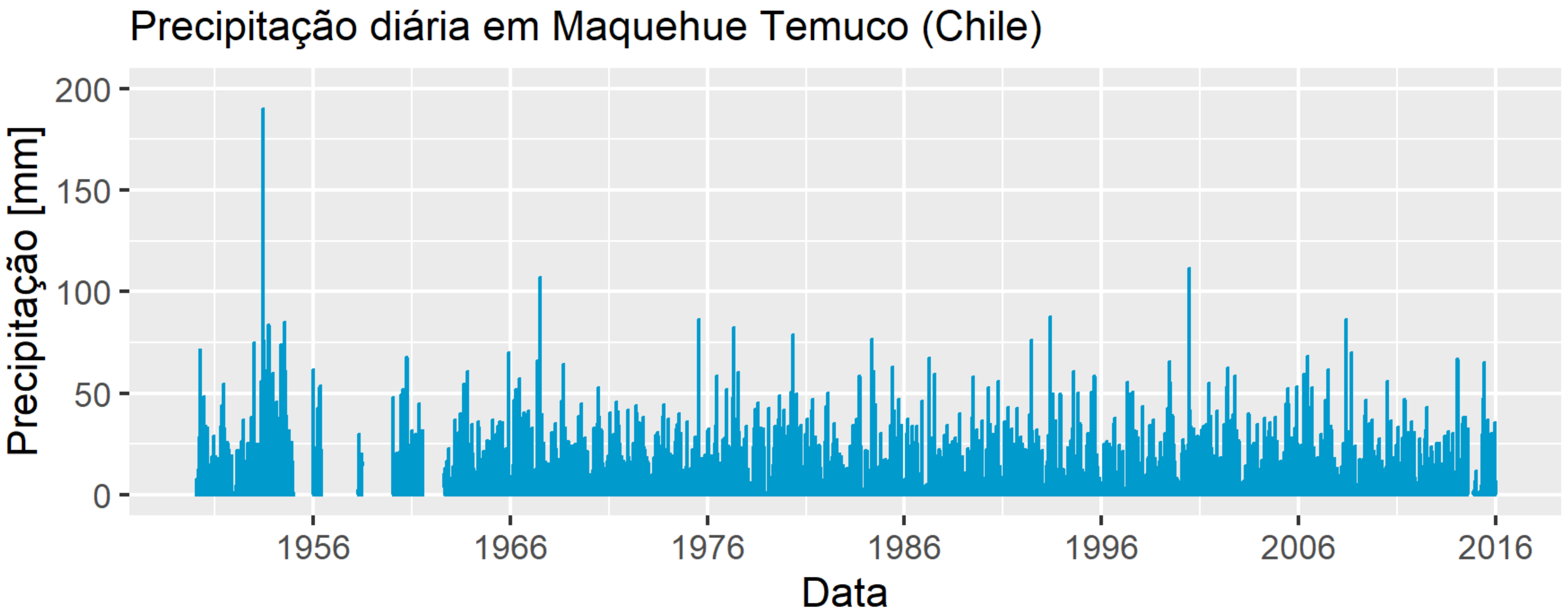
Não é possível prever exatamente a evolução dos fenômenos
É possível aproximá-los através de modelos estocásticos apropriados

Variáveis hidrológicas, enquanto manifestações de fenômenos naturais, seguem essas premissas

CONCEITOS INICIAIS

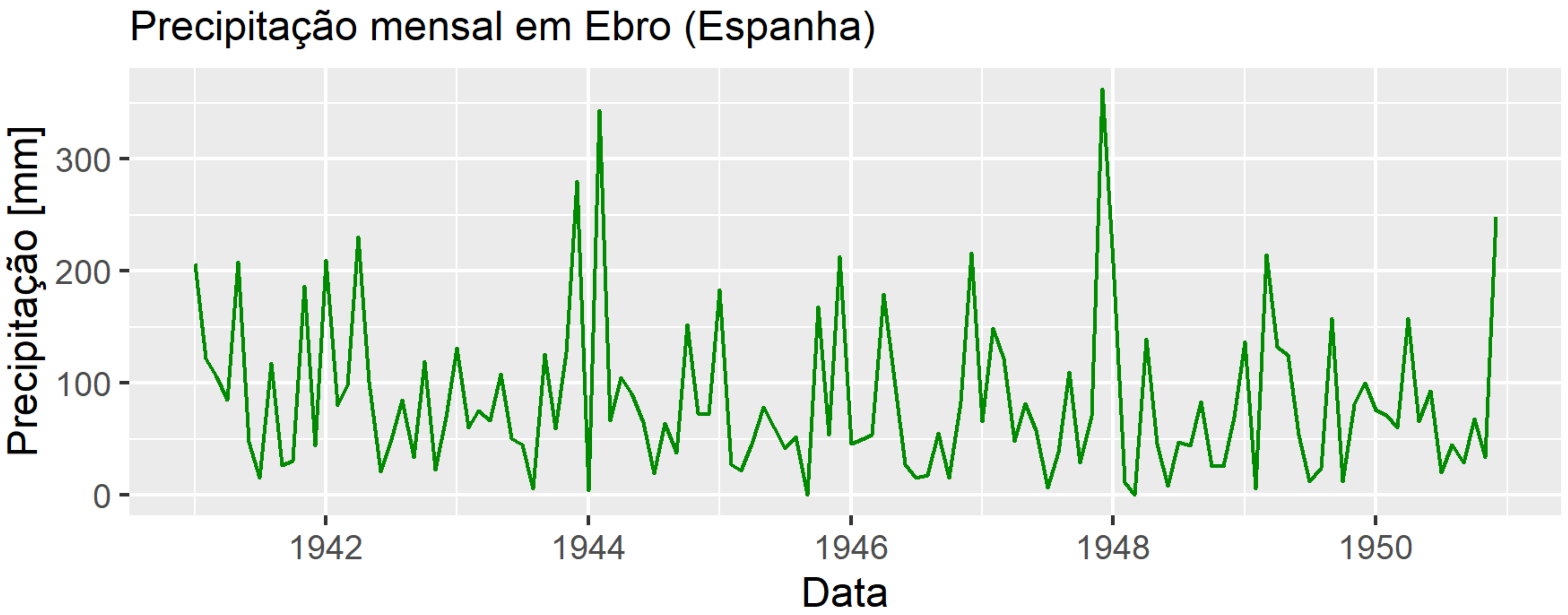
natureza estocástica das variáveis hidrológicas

[Exemplo 3] Observe as séries temporais dos próximos slides. Discuta as principais características que detectou.



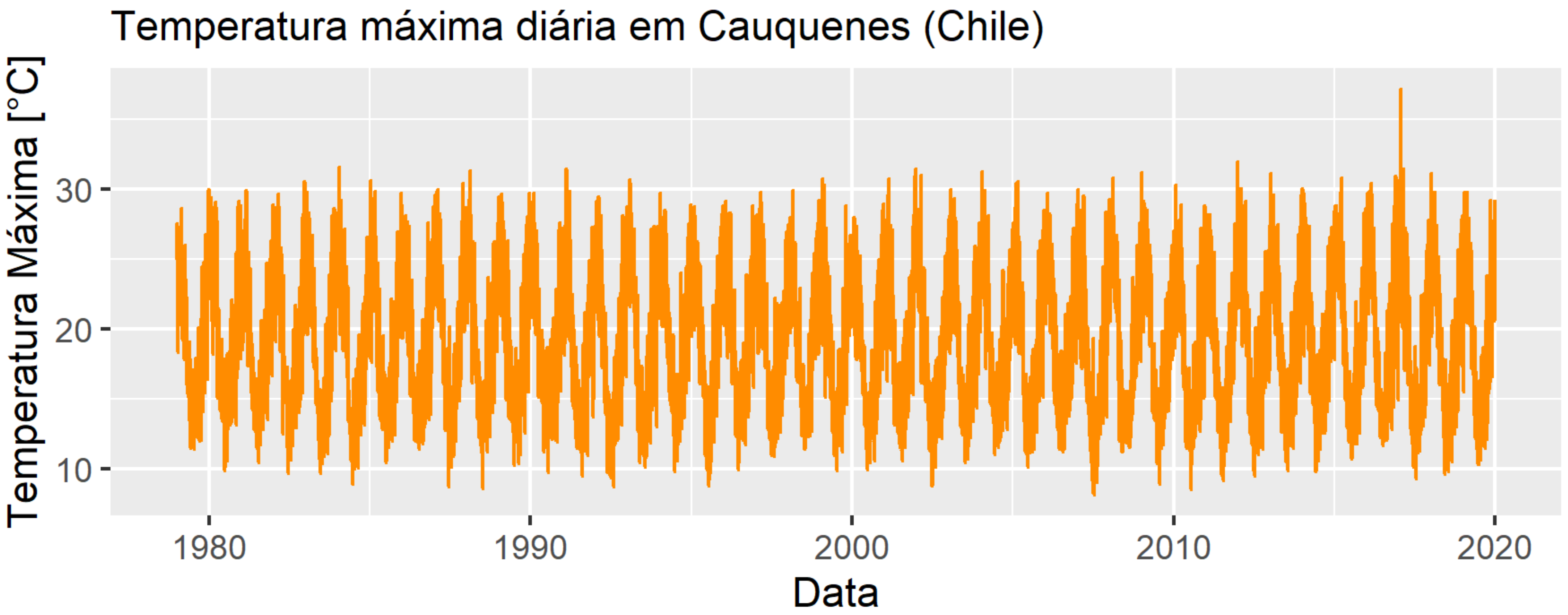
Conceitos iniciais | natureza estocástica das variáveis hidrológicas

[Exemplo 3] Observe as séries temporais dos próximos slides. Discuta as principais características que detectou.



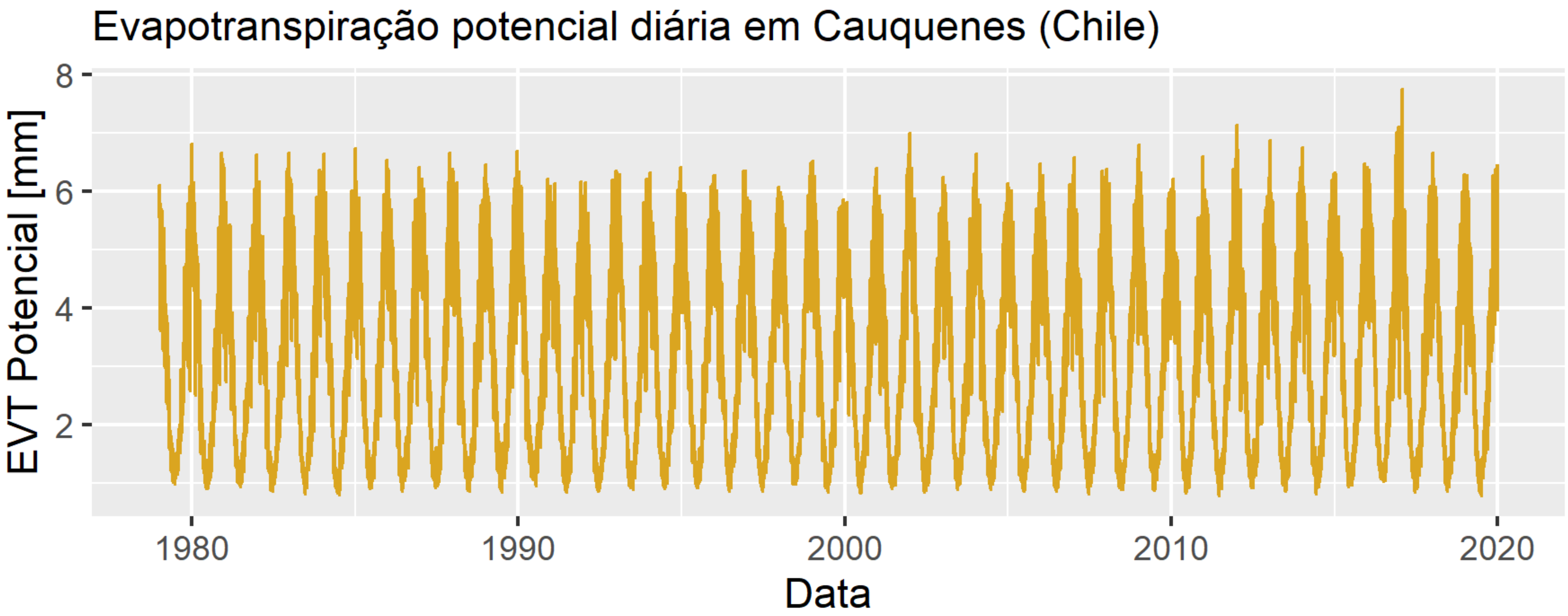
Conceitos iniciais | natureza estocástica das variáveis hidrológicas

[Exemplo 3] Observe as séries temporais dos próximos slides. Discuta as principais características que detectou.



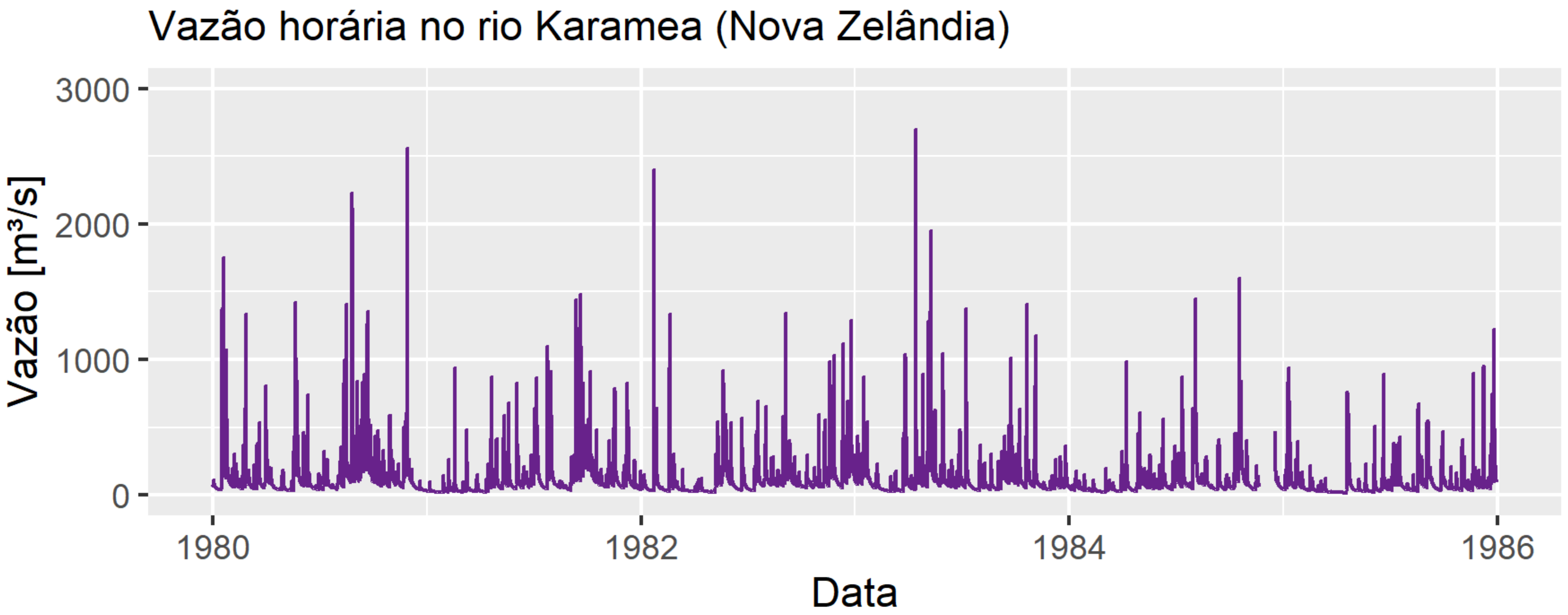
Conceitos iniciais | natureza estocástica das variáveis hidrológicas

[Exemplo 3] Observe as séries temporais dos próximos slides. Discuta as principais características que detectou.



Conceitos iniciais | natureza estocástica das variáveis hidrológicas

[Exemplo 3] Observe as séries temporais dos próximos slides. Discuta as principais características que detectou.



Conceitos iniciais | natureza estocástica das variáveis hidrológicas

Características das variáveis hidrológicas

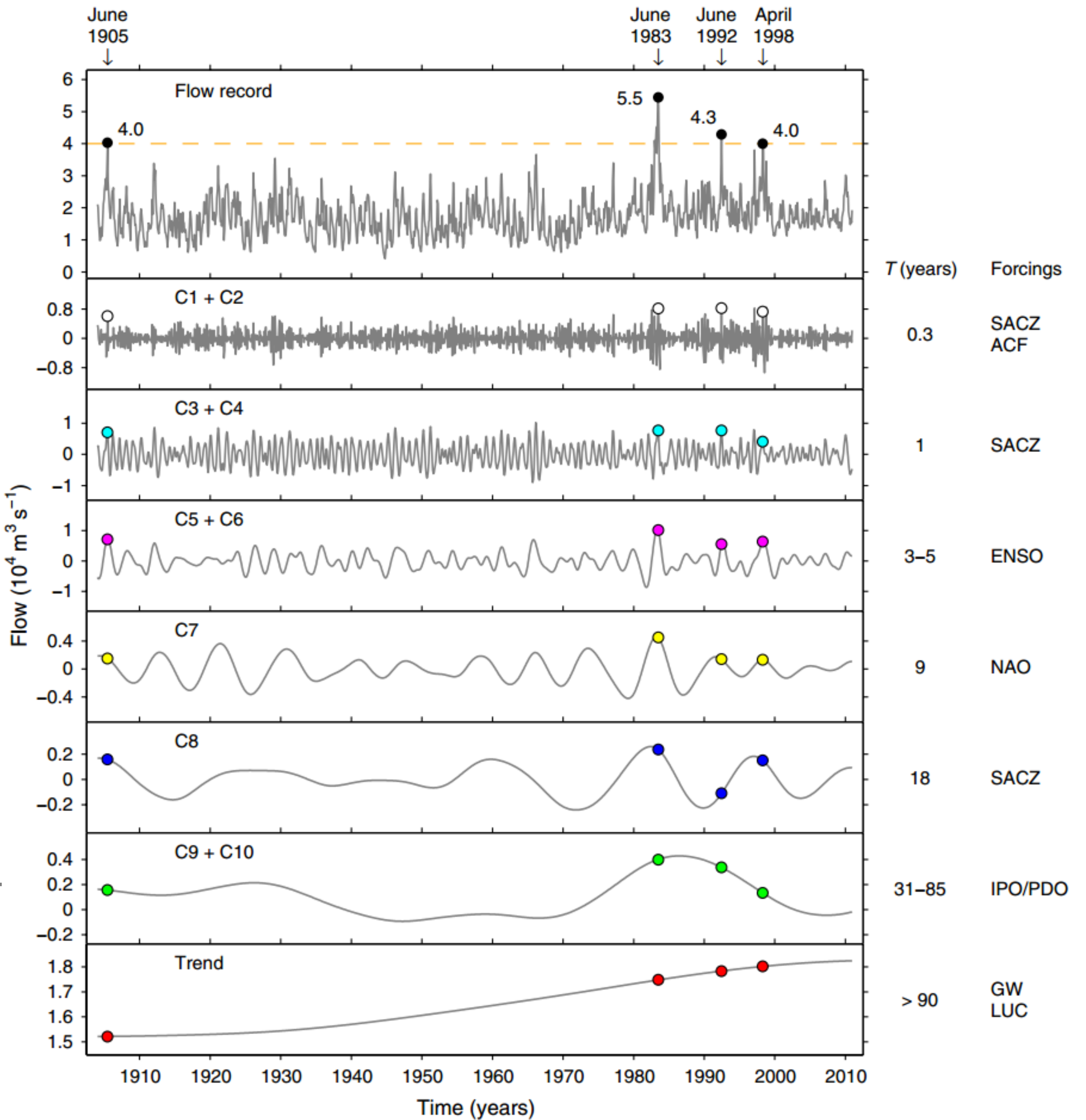
1. ciclos sazonais e outras variações quase periódicas
2. tendências e alterações nos regimes
3. outros efeitos (ex.: marés)
4. componentes aleatórios provocados por fenômenos atmosféricos:

   turbulência
   vorticidade de larga escala
   transferência de calor
   emissão/absorção de ondas de radiação, etc.

Conceitos iniciais | natureza estocástica das variáveis hidrológicas

Características das variáveis hidrológicas (cont.)

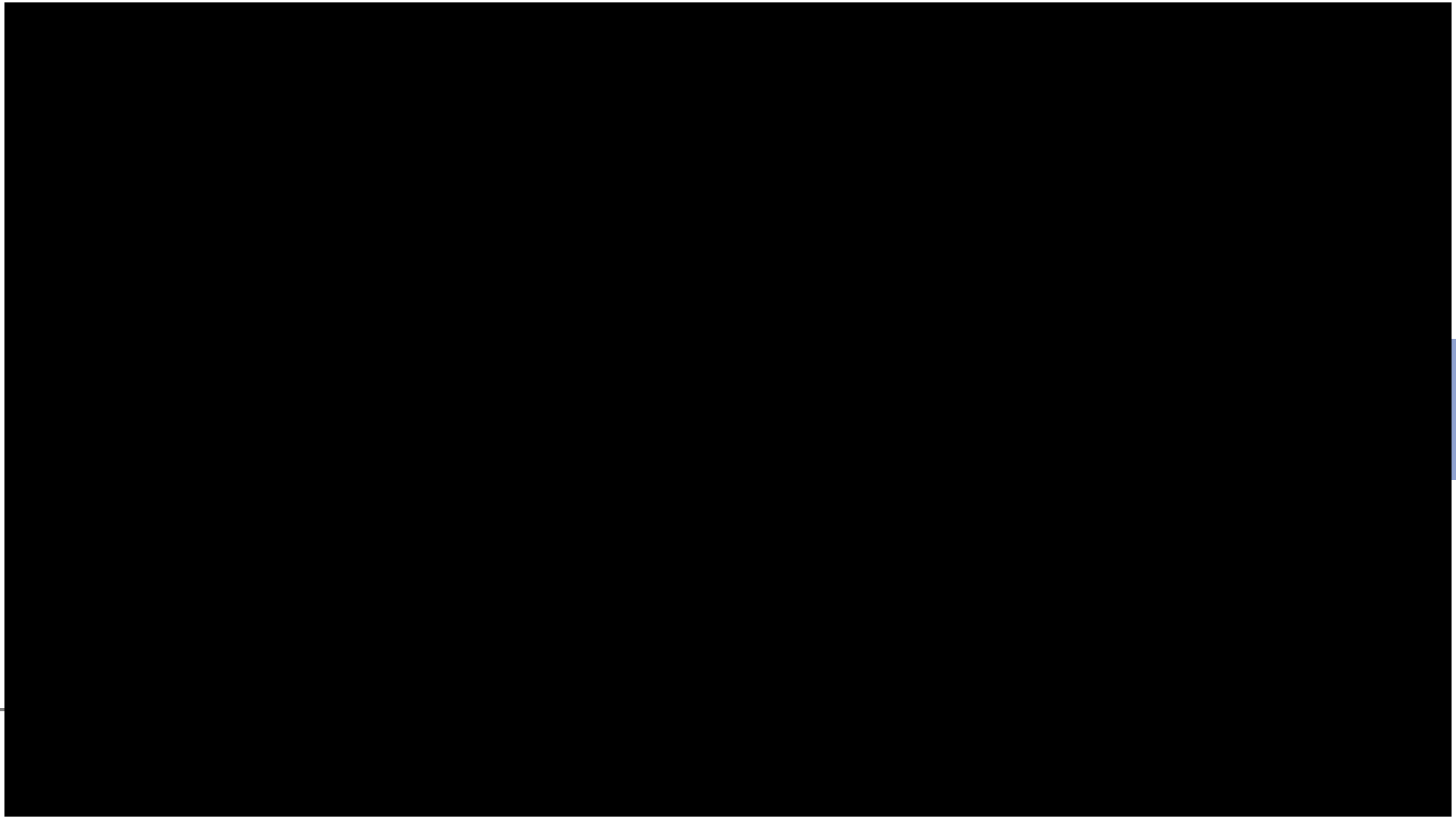
5. combinação de fenômenos em diversas escalas temporais



Antico, A., Torres, M.E. & Diaz, H.F. (2016) Contributions of different time scales to extreme Paraná floods. *Clim Dyn*, 46: 3785.

Características das variáveis hidrológicas (cont.)

6. persistência

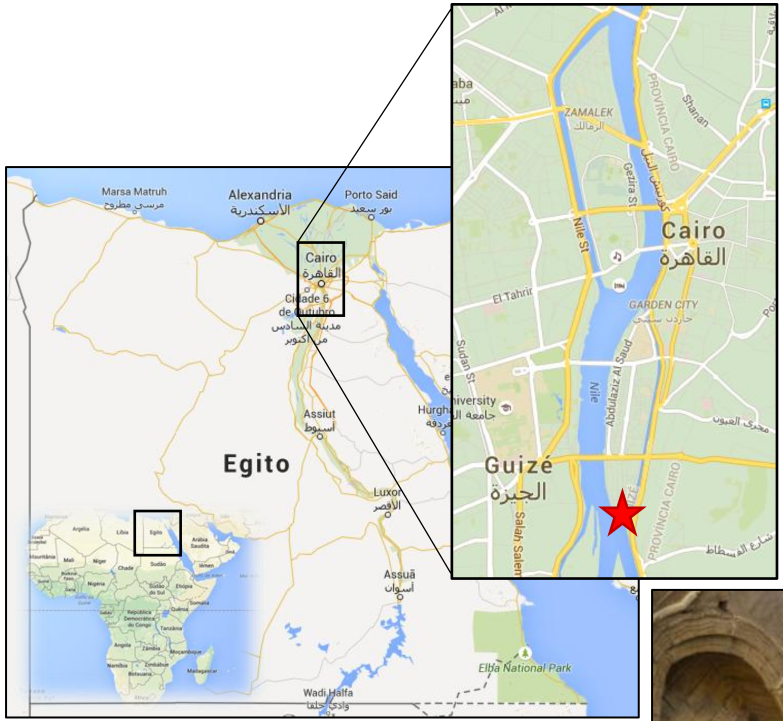


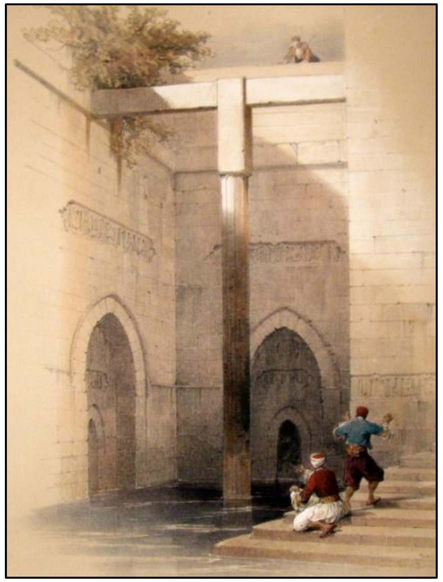
Rio Elba, Hamburgo, Alemanha

https://www.youtube.com/watch?v=tddi7sM6CIU

Características das variáveis hidrológicas (cont.)

6. persistência (cont.)





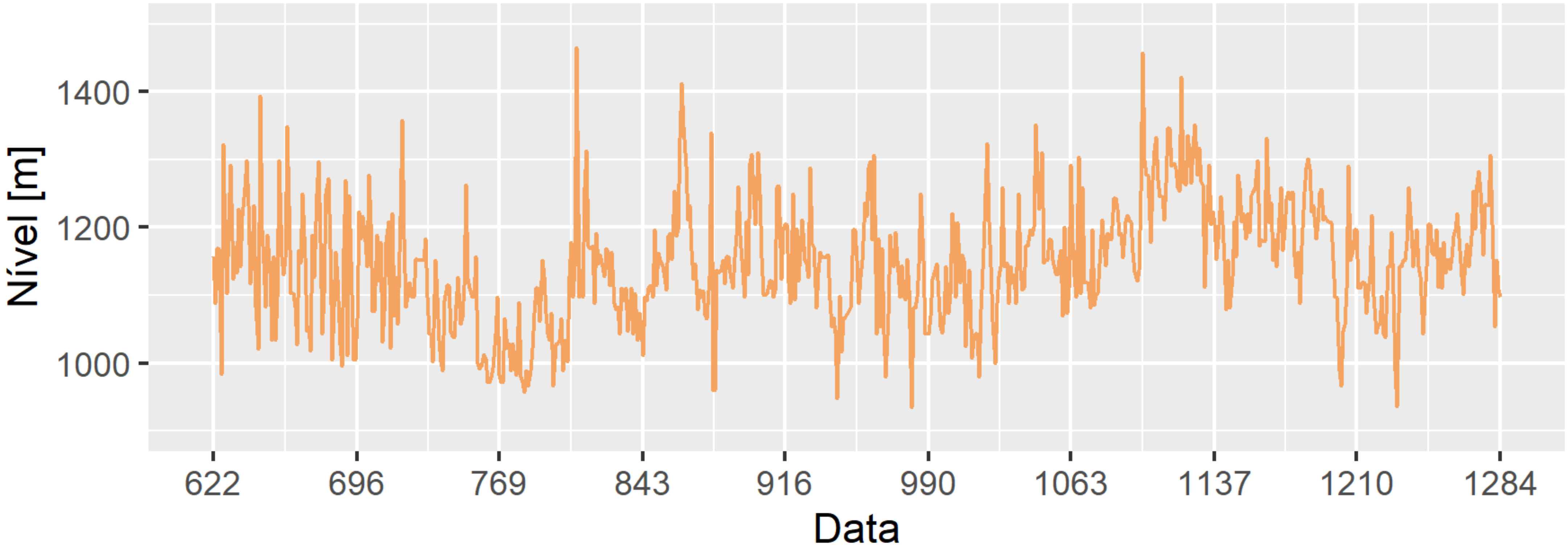
Nilômetro de Rhoda

622 a 1284

Características das variáveis hidrológicas (cont.)

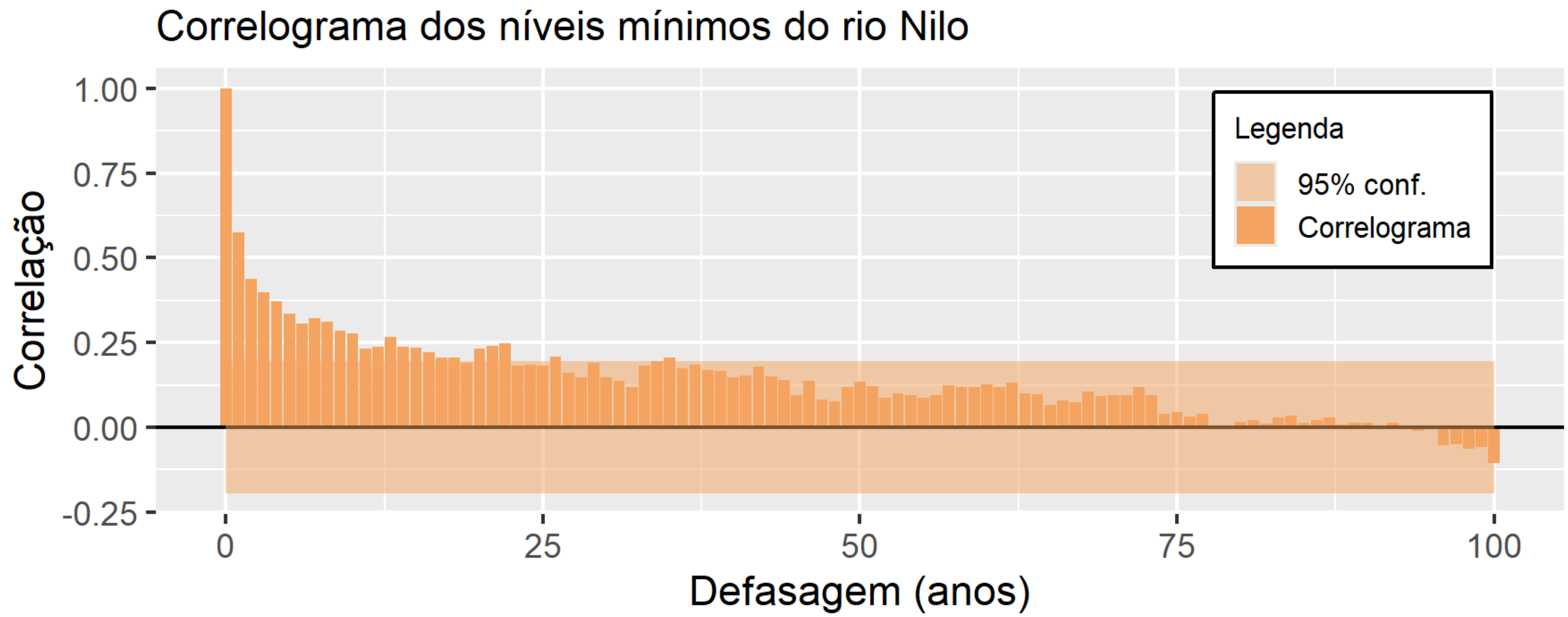
6. persistência (cont.)

Nível mínimo no rio Nilo (estação de Roda, Cairo)



Características das variáveis hidrológicas (cont.)

6. persistência (cont.)



Conceitos iniciais | natureza estocástica das variáveis hidrológicas

Processos hidrológicos considerados estocásticos

- Precipitação
- Evapotranspiração
- Escoamento superficial e sub-superficial
- Fluxos de sedimentos em suspensão
- Concentrações de oxigênio dissolvido
- Formas dos leitos fluviais
- Temperatura da água
- Capacidade de infiltração, etc.

Variáveis hidrológicas → variáveis aleatórias

REVISÃO DE ESTATÍSTICA

VARIÁVEIS ALEATÓRIAS

Variáveis aleatórias (VA):

Função que associa um valor numérico a cada resultado de um experimento aleatório proveniente de um espaço amostral

VA discretas *vs.* contínuas

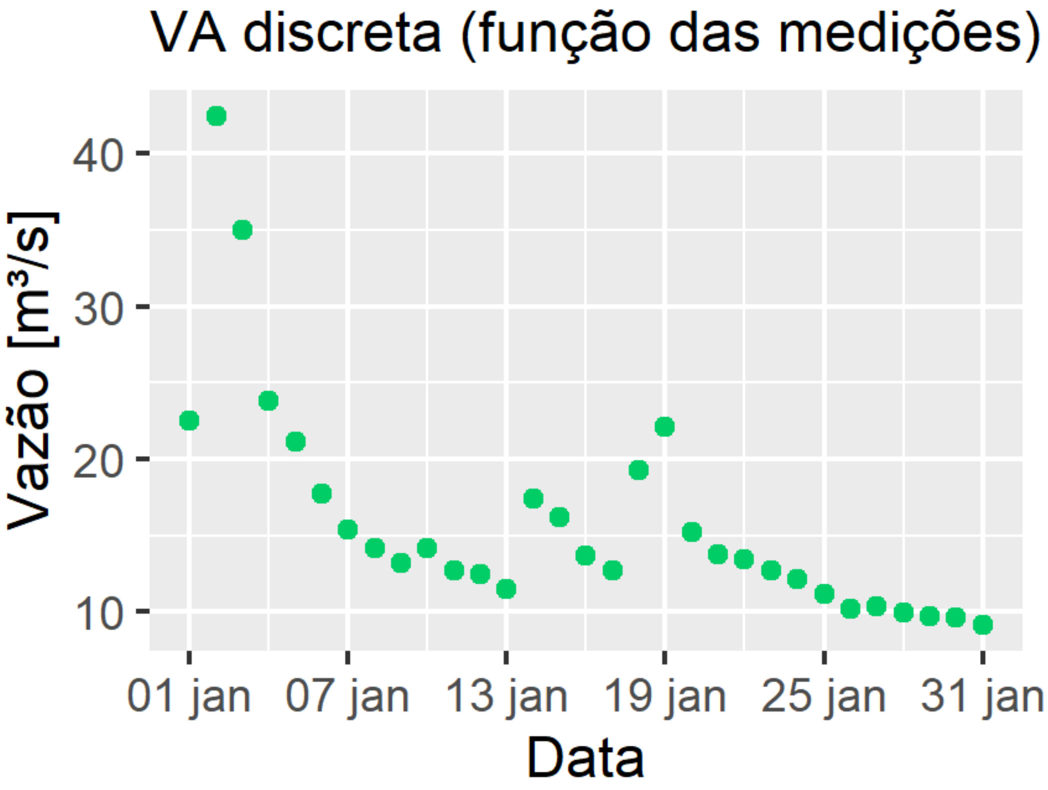
- VA discreta: assume somente valores numéricos inteiros associados a um espaço amostral finito e numerável
  - Ex.: nº de dias chuvosos em um ano; nº de vezes que a vazão de um rio superou um limite
- VA contínua: assume valores possíveis em um intervalo, ou uma coleção de intervalos (intervalo de valores reais)
  - Ex.: vazão em um rio; volumes de chuva; nível de água subterrânea; temperatura da água; conteúdo de oxigênio na água; etc.

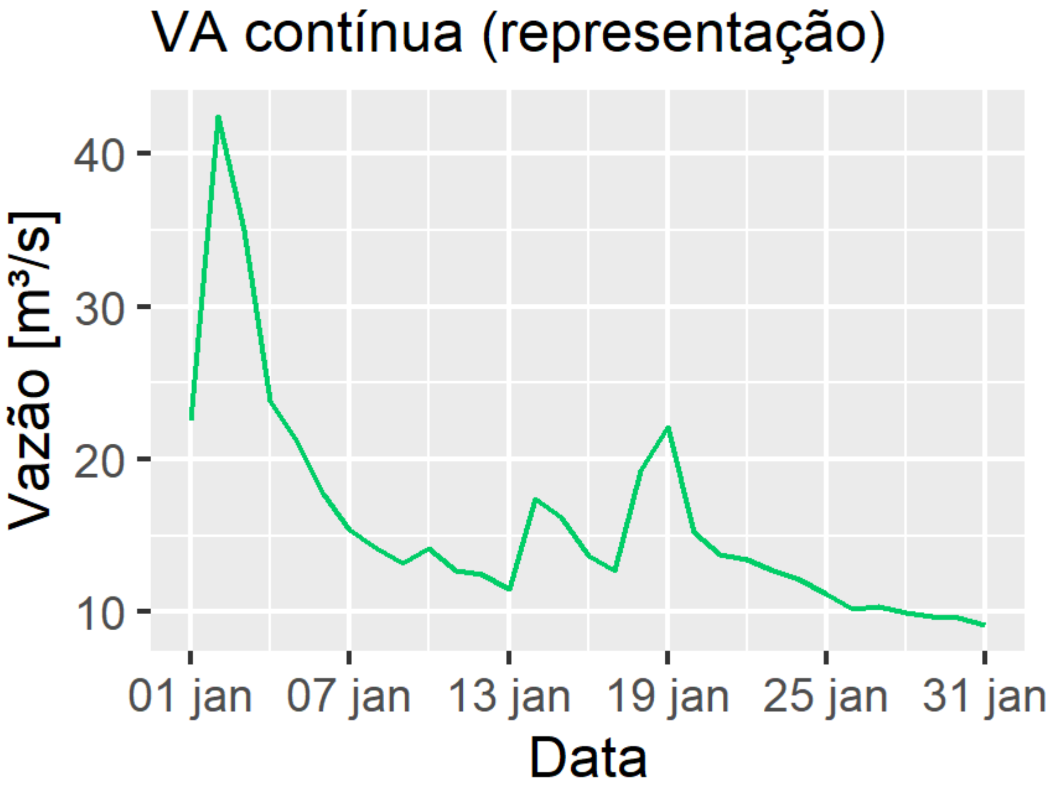
Revisão de estatística | variáveis aleatórias

Representação de VA contínuas por meio de VA discretas

Ex.: registro diário das vazões em um rio

considerada uma simplificação





Revisão de estatística | variáveis aleatórias

VA qualitativas *vs.* quantitativas

VA qualitativa: resultados expressos por atributo ou qualidade

Ex.: Estado do tempo {‘bom’, ’chuvoso’, ’nublado’} (nominal);
Nível de reservatório {‘muito alto’, ‘alto’, ‘médio’, ‘baixo’, ‘muito baixo’} (ordinal)

VA quantitativa: resultados expressos por números

Ex.: Nº de dias chuvosos em um ano (discreta);
Altura máxima de chuva anual (contínua)

Revisão de estatística | variáveis aleatórias

VA univariadas *vs.* multivariadas

VA univariadas: associada a um único atributo de qualidade ou quantidade

Ex.: vazão observada no rio Iguaçu em União da Vitória

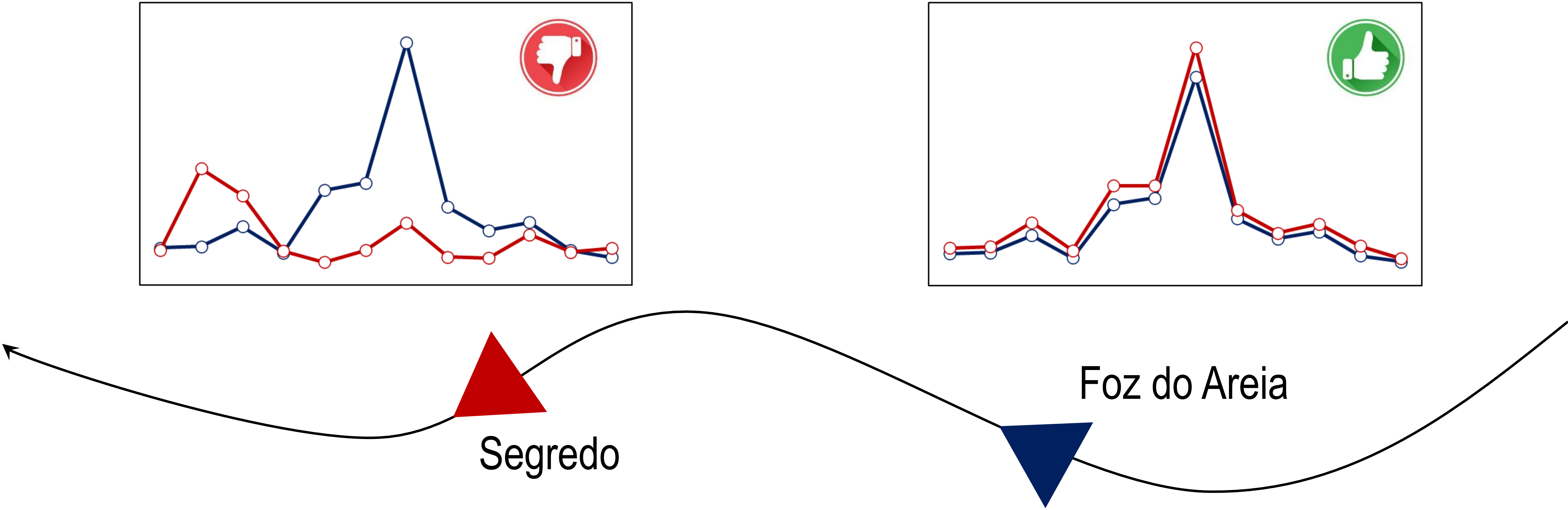
VA multivariadas: associada a mais de um atributo de qualidade ou quantidade

Ex.: vazões observadas no rio Iguaçu em União da Vitória e em Porto Amazonas (simultaneamente)

Revisão de estatística | variáveis aleatórias

Importância de considerar VA multivariadas

Ex: vazões do rio Iguaçu em Foz do Areia e em Segredo

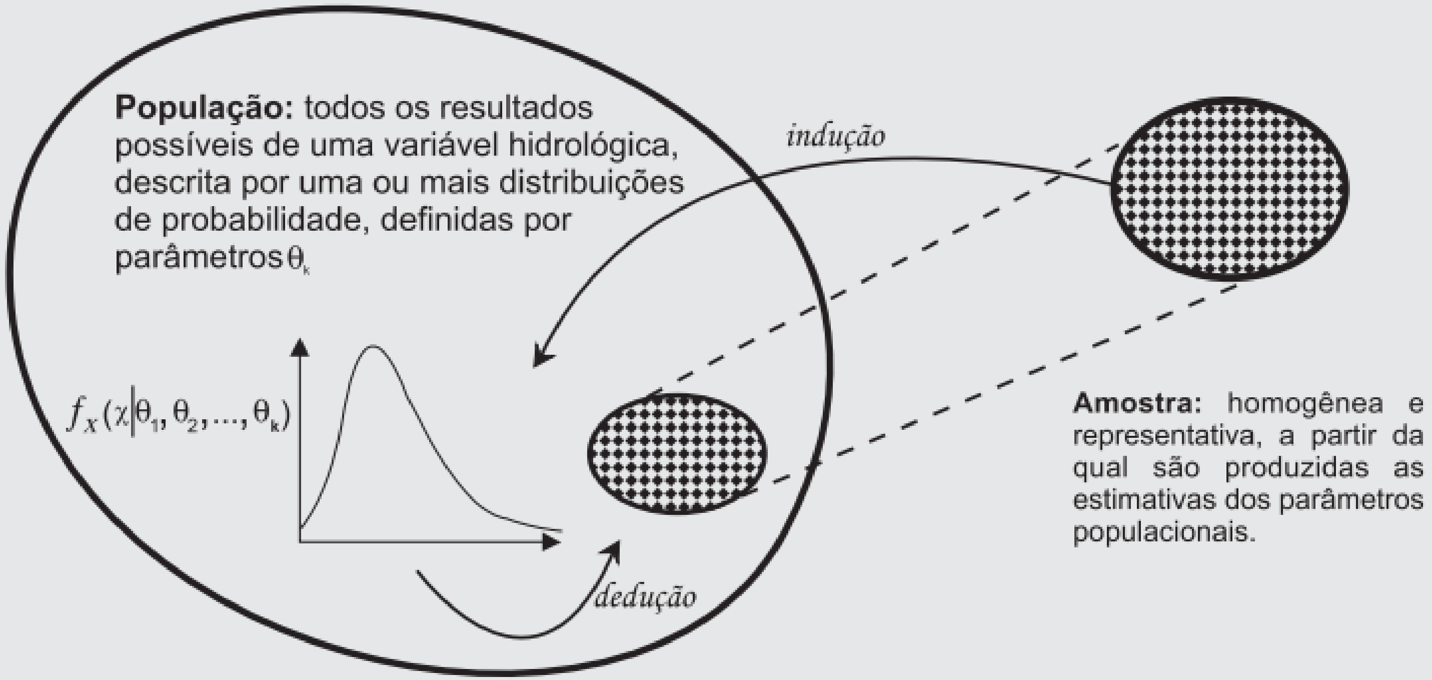


Revisão de estatística | variáveis aleatórias

<u>População e amostra</u>

População: conjunto de todas as realizações possíveis de uma variável hidrológica

Amostra: subconjunto extraído da população

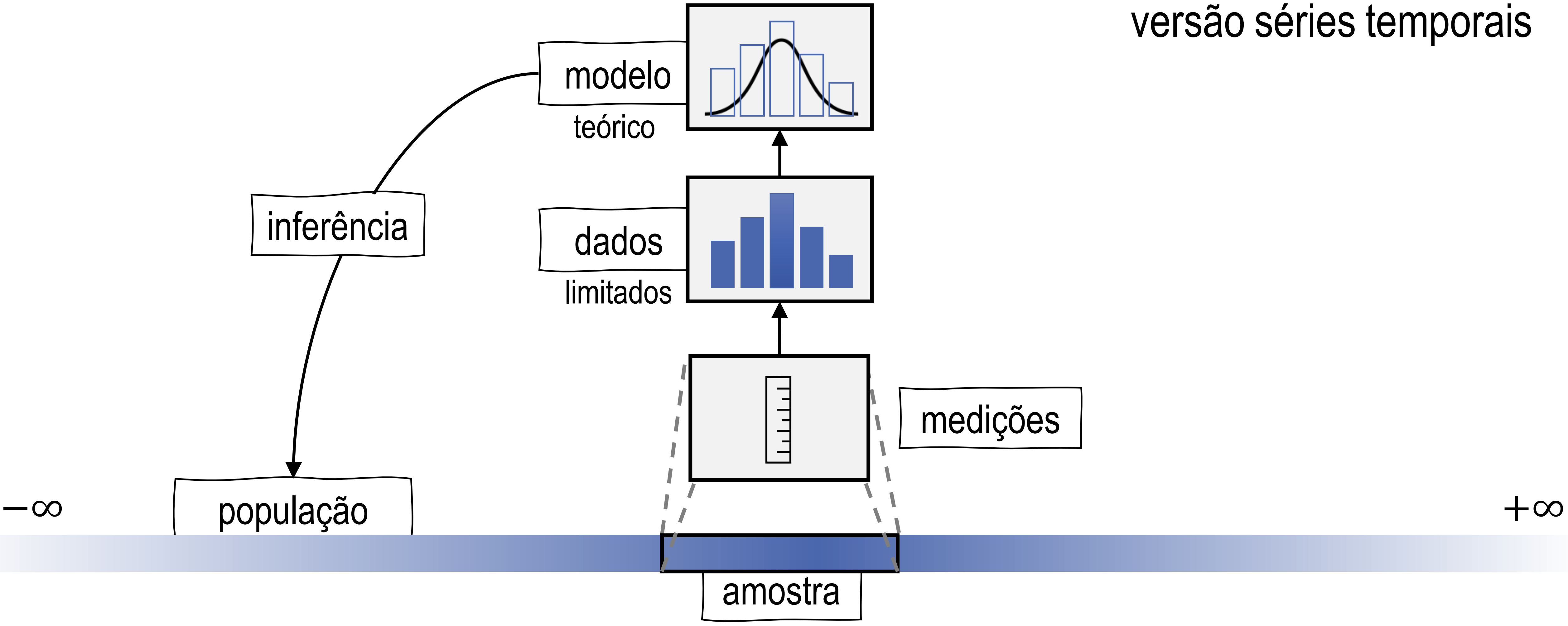


Adaptado de Naghettini e Pinto (2007, p. 12)

Revisão de estatística | variáveis aleatórias

População e amostra

versão séries temporais



Revisão de estatística

análise exploratória de dados

Revisão de estatística | análise exploratória de dados

Análise exploratória de dados (EDA): coleção de técnicas estatísticas utilizadas para a caracterizar a população com base na amostra em mãos

A escolha das técnicas depende dos dados em análise

conhecimento prévio do comportamento físico da variável

importante para evitar o uso de métodos não adequados, a partir dos quais as interpretações seriam incorretas ou inconclusivas

(ex.: aplicar técnicas que requerem normalidade a uma série com distribuição assimétrica)

Revisão de estatística | análise exploratória de dados

Caracterização geral da série

intervalo de observação (início e fim da série)
número de observações
escala temporal
resolução espacial
presença de falhas
dados suspeitos (outliers)

Medidas de tendência central

média
mediana*
*técnica robusta (imune a outliers)

Revisão de estatística | análise exploratória de dados

Medidas de dispersão

- variância
- desvio padrão
- amplitude interquartil*
- desvio absoluto da mediana*
- coeficiente de variação

Assimetria

- coeficiente de assimetria
- assimetria interquartílica*

*técnicas robustas (imune a outliers)

Revisão de estatística | análise exploratória de dados

Determinação de percentis

tipicamente 5%, 25%, 50%, 75%, 95%

Representações gráficas

série temporal

histograma

box-plot, etc.

Outras análises aplicáveis

correlações

tendências

ajuste a distribuições

Revisão de estatística | análise exploratória de dados

Em resumo, a EDA reúne um grupo de análises realizadas previamente ao estudo ou aplicação desejados

etapa fundamental para a escolha adequada dos procedimentos metodológicos subsequentes

[Exemplo]
EDA em artigo

RBRH

Revista Brasileira de Recursos Hídricos
Brazilian Journal of Water Resources
Versão On-line ISSN 2318-0331
RBRH, Porto Alegre, v. 24, e14, 2019
Scientific/Technical Article

https://doi.org/10.1590/2318-0331.241920180031

Data assimilation using the ensemble Kalman filter in a distributed hydrological model on the Tocantins River, Brasil

Assimilação de dados por filtro de Kalman por conjunto em um modelo hidrológico distribuído na bacia do rio Tocantins, Brasil

Karena Quiroz Jiménez[1], Walter Collischonn[1] and Rodrigo Cauduro Dias de Paiva[1]

https://www.scielo.br/j/rbrh/a/hxZ6BCX5zLxqjGbLkcW3HJr/?lang=en

Revisão de estatística | análise exploratória de dados

https://www.scielo.br/j/rbrh/a/hxZ6BCX5zLxqjGbLkcW3HJr/?lang=en

Study area description

The Tocantins River basin is an area of study located in the central region of Brazil, with a drainage area of 310.000 km² up to the confluence with the Araguaia River (see Figure 1). The monthly mean temperature of the study area varies from 20 °C to 25 °C, approximately. The monthly mean maximums occur in the months of August and September while the monthly mean minimum occurs in July and August. The mean rainfall is 1480 $mm.year^{-1}$ and streamflow is 3300 $m^3.s^{-1}$ at the Estreito station according to the estimated calculation for the 2008-2014. The basin topography elevations range from 83 to 1640 meters.

The Tocantins River basin was selected as area of study because it has a very important hydropower system formed by the Serra da Mesa, Cana Brava, São Salvador, Peixe Angical and Estreito hydroelectric plants. The Cana Brava and Estreito have an installed capacity for electricity generation of 1275 MW and 1087 MW, respectively. It was also chosen because it is a region that periodically suffers extreme events, typical situations in other regions of Brazil. Currently, the National Center of Monitoring and Alerts of Natural Disasters (*CEMADEN Centro Nacional de Monitoramento e Alertas de Desastres Naturais*) monitors the municipalities of Goiatins and Porto Nacional, located within the Tocantins River basin, classified as being vulnerable to hydrological risks. Likewise, the town of Imperatriz do Maranhão, downstream from Estreito suffers constant floods, with an impact on the populations living in the riparian areas.

The basin was discretized into 410 cells, 45 sub-basins and the integration of the use and soil type maps generated 6 different types of hydrological response units which are: forest in medium

Available data

The data on streamflow at hourly time intervals were extracted from 16 stations, 10 (ten) of them from hydroelectric power companies, as well as from the National Water Agency (ANA - *Agência Nacional de Águas*) and the other 6 (six) stations are from National Operator of the Electric System (ONS - *Operador Nacional de Sistema Elétrico*). The naturalized streamflow data of ONS is available at daily time intervals and for this work they were interpolated linearly to obtain hourly data. The drainage areas for the sites with a hydroelectric plant located along the mainstream of the river are more than 50.000 km² and less than 289.000 km². For the stream stations located in the Southest and Northeast region of the basin the drainage areas range from 3.000 – 44.000 km².

The data on mean air temperature, relative humidity, wind velocity, atmospheric pressure and insolation were obtained from 15 climate gauging stations located around the basin, supplied by ANA and interpolated for hourly data. The precipitation was obtained from 50 rain-gauging station supplied by hydroelectric power companies and by the National Institute of Meteorology (*INMET Instituto Nacional de Meteorologia*). Considering the low density of the rain-gauging stations (1 station at every 6,200 km²) it was chosen to combine with TRMM satellite precipitation product. This option was performed to attempt to improve the response in the simulated streamflows at several basin stations using a methodology based on the work by Rozante et al. (2010). Quiroz (2017) used that methodology to determine temporal series of rainfall at hourly time intervals in the Tocantins River basin for the period of 1998-2014 called MergeHQ.

Revisão de estatística | análise exploratória de dados

Referências recomendadas:

Helsel et al. (2020): Statistical Methods in Water Resources
capítulos 1, 2 e 16

Naghettini & Pinto (2007): Hidrologia Estatística
capítulo 2

Detzel (2025): Estatística Aplicada a Ciências Ambientais (notas de aula)
aulas 2 e 3
disponível em: https://detzeldhm.github.io/estatistica_ciencias_ambientais/

REVISÃO DE ESTATÍSTICA

distribuições de probabilidades

Distribuição de probabilidades:

Modelos teóricos para inferência de informações populacionais a partir da amostra

Definidos por funções densidade de probabilidade (FDP) e por funções densidade de probabilidade acumulada (FDA) (caso de VA contínuas)

A FDP de uma variável aleatória X é expressa por $p_X(x)$

A FDA de uma variável aleatória X é expressa por $P_X(x)$

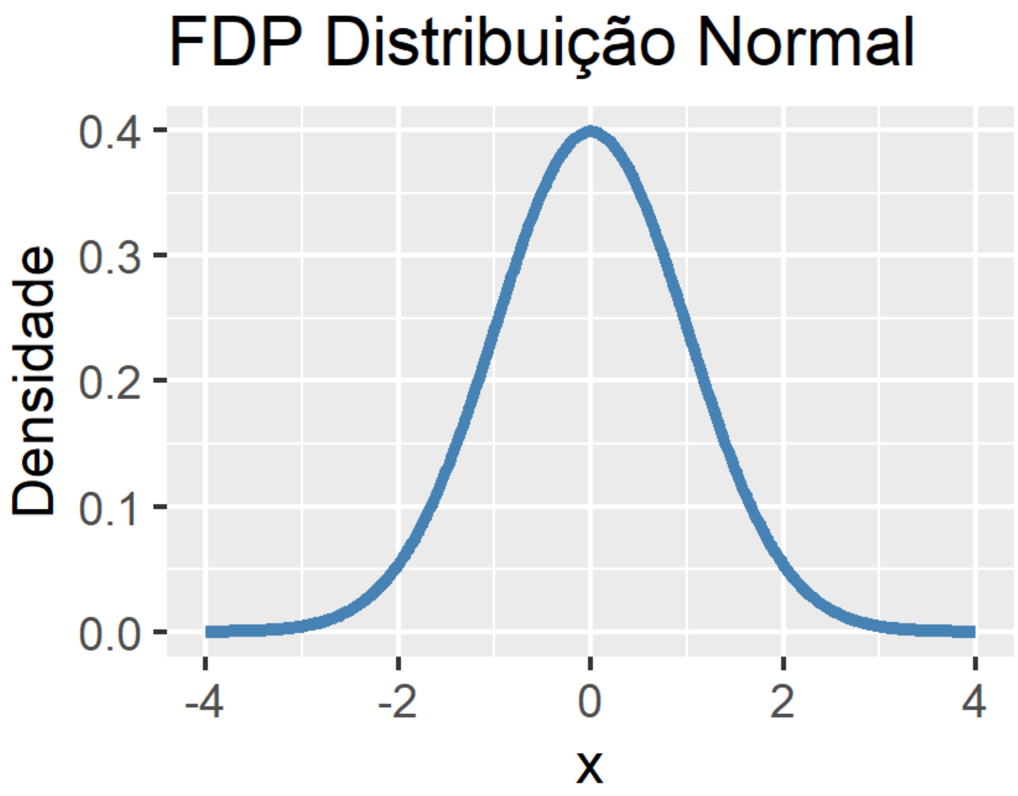
$p_X(x)$ ou $P_X(x)$

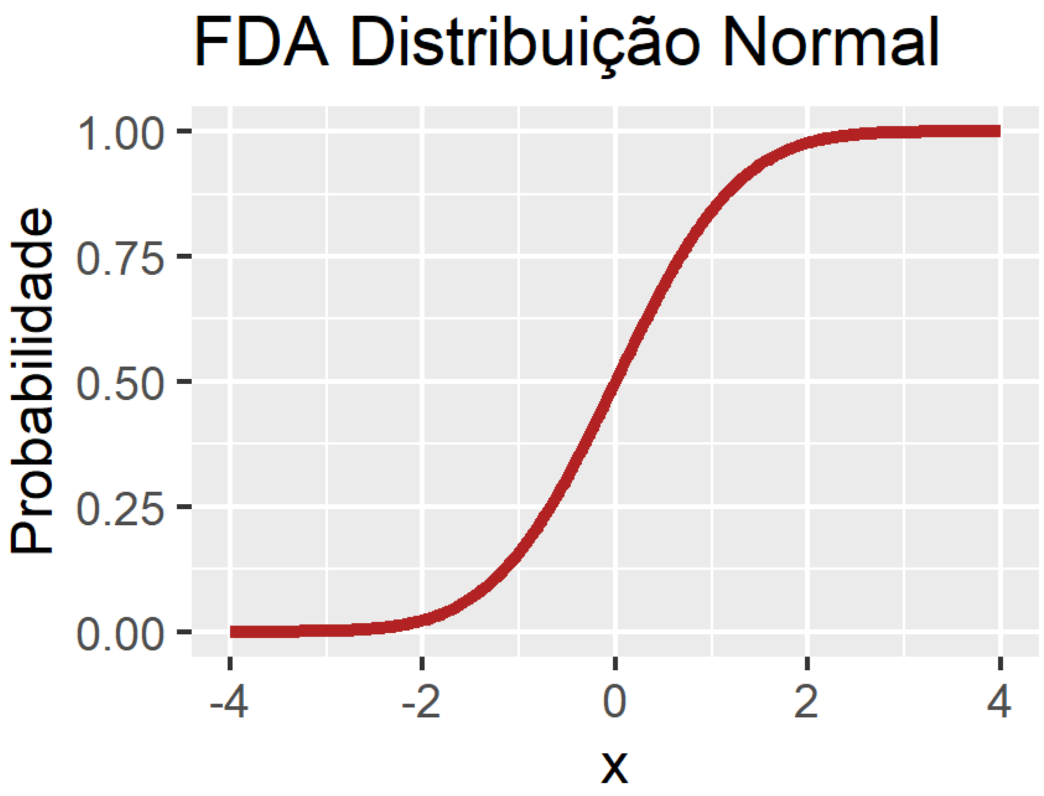
Variável aleatória

Valor assumido pela VA

Revisão de estatística | distribuições de probabilidades

A FDA e a FDP são relacionadas por $P_X(x) = \int_{-\infty}^{x} p_X(x)\, dx$





Revisão de estatística | distribuições de probabilidades

Propriedades básicas da FDP

$$\begin{cases} p_X(x) \geq 0; \text{ para todos os valores de } x \\ \displaystyle\int_{-\infty}^{+\infty} p_X(x)\,dx = 1 \end{cases}$$

A FDA representa a probabilidade de X ser menor ou igual a x:

$$P_X(x) = \mathrm{prob}(X \leq x)$$

Revisão de estatística | distribuições de probabilidades

Distribuições marginais para vazões médias mensais e anuais

Raciocínio 1: o mecanismo gerador de vazões médias é aditivo
(composto pela soma de diversas causas físicas representadas por VA)

Pelo Teorema do Limite Central*, assume-se uma distribuição Normal (N)
(*a média da soma de VA independentes e igualmente distribuídas tende a uma distribuição Normal conforme o tamanho da amostra aumenta)

Revisão de estatística | distribuições de probabilidades

Distribuição **Normal** $(-\infty \leq x \leq +\infty)$

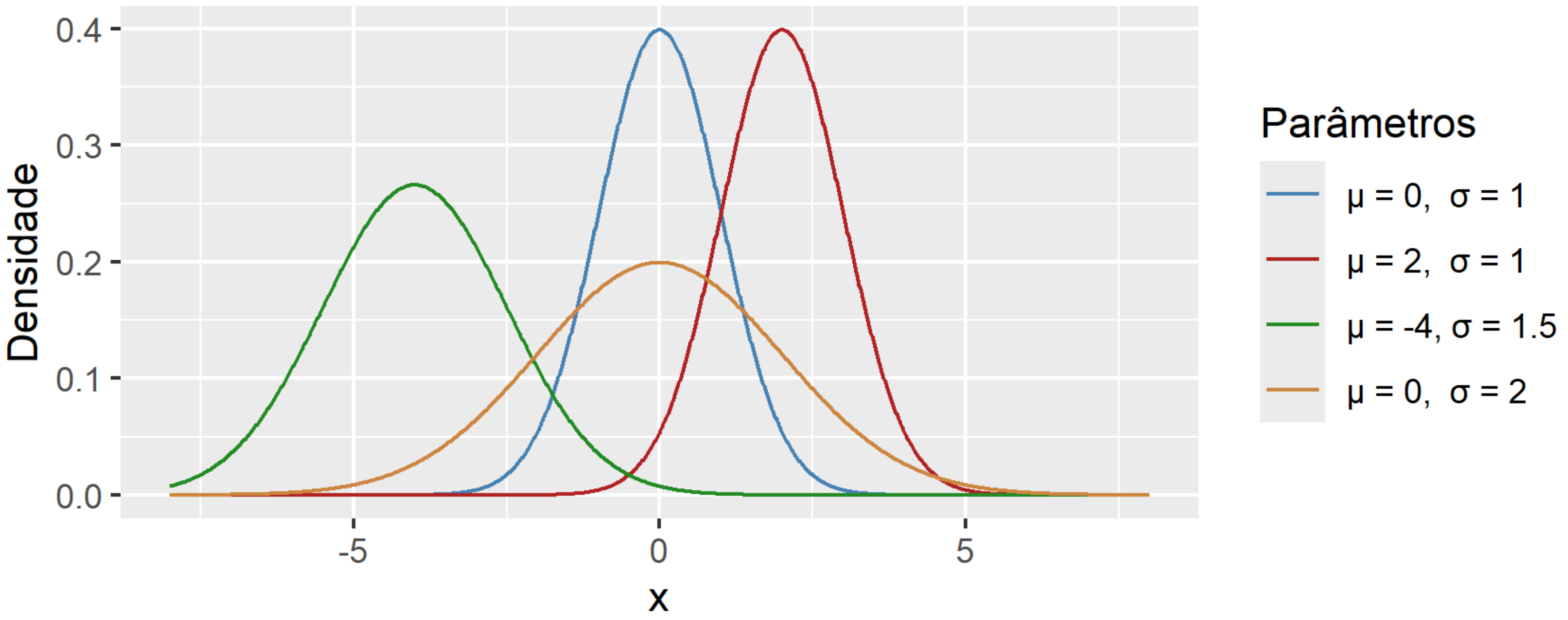
$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Parâmetros

$\mu = \hat{\mu}_x$

$\sigma = \hat{\sigma}_x$

$\xi = 0$



Revisão de estatística | distribuições de probabilidades

Contudo, a distribuição Normal aplicada a vazões aceita a existência de valores negativos

(mesmo que com baixa probabilidade)

Raciocínio 2: o mecanismo gerador de vazões médias é multiplicativo

(composto pelo produto de diversas causas físicas representadas por VA)

Assume-se uma distribuição Log-Normal (LN)

com 2 (LN2) ou 3 (LN3) parâmetros

Definida somente para valores positivos

Revisão de estatística | distribuições de probabilidades

Distribuição **Log-normal a 2 parâmetros** $(x > 0)$

$$p_X(x) = \frac{1}{x\sqrt{2\pi}\sigma_{ln}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \mu_{ln}}{\sigma_{ln}}\right)^2\right]$$

Parâmetros

$\mu = \exp[(\sigma_{ln}^2)/2 + \mu_{ln}]$

$\sigma^2 = \exp[2(\sigma_{ln}^2 + \mu_{ln})] - \exp(\sigma_{ln}^2 + 2\mu_{ln})$

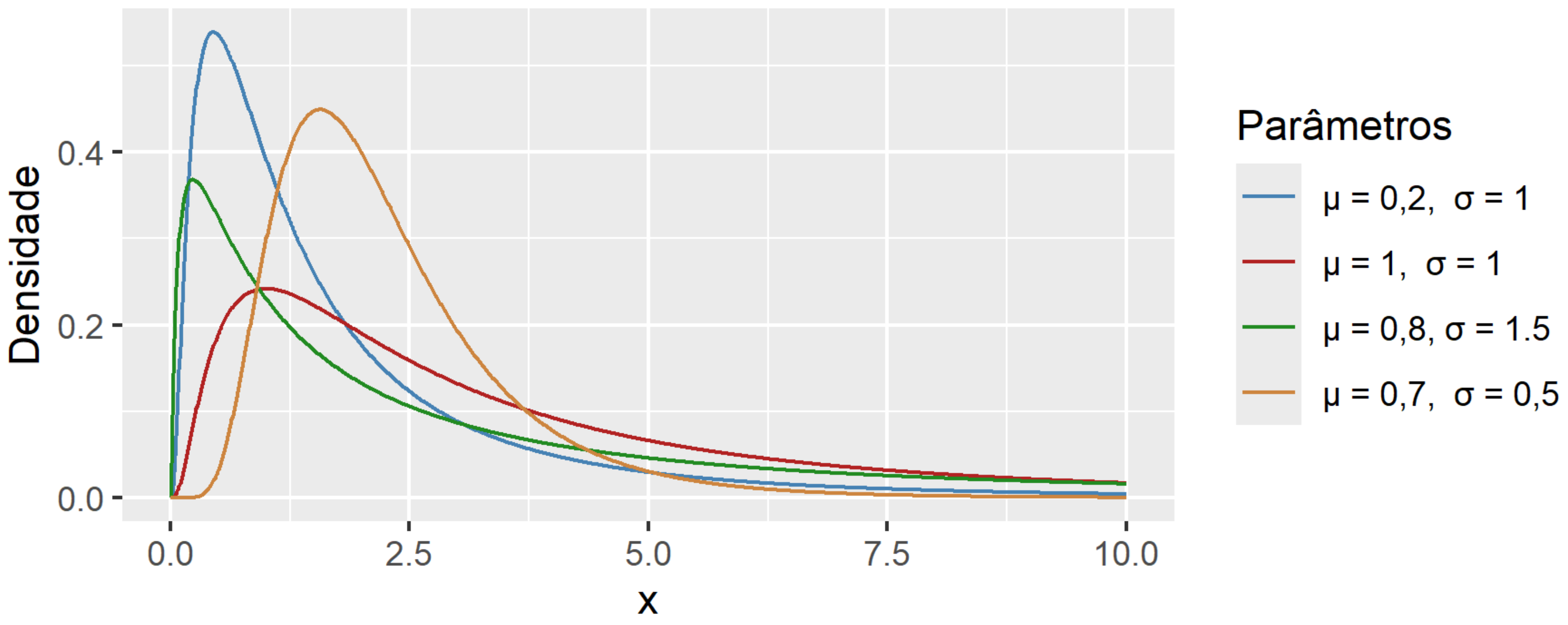
$\xi = (\hat{\sigma}_x/n)^3 + 3(\hat{\sigma}_x/n)$

$\psi = 1 + (\hat{\sigma}_x/\hat{\mu}_x)^2$

$\mu_{ln} = 0{,}5 \ln[\hat{\sigma}_x^2/(\psi^2 - \psi)]$

$\sigma_{ln}^2 = \ln \psi$

Distribuição **Log-normal a 2 parâmetros** $(x > 0)$



Revisão de estatística | distribuições de probabilidades

Distribuição **Log-normal a 3 parâmetros** $(x > \Delta)$

$$p_X(x) = \frac{1}{(x-\Delta)\sqrt{2\pi}\sigma_{ln}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x-\Delta)-\mu_{ln}}{\sigma_{ln}}\right)^2\right]$$

Parâmetros

$$\mu = \Delta + \exp[(\sigma_{ln}^2)/2 + \mu_{ln}]$$

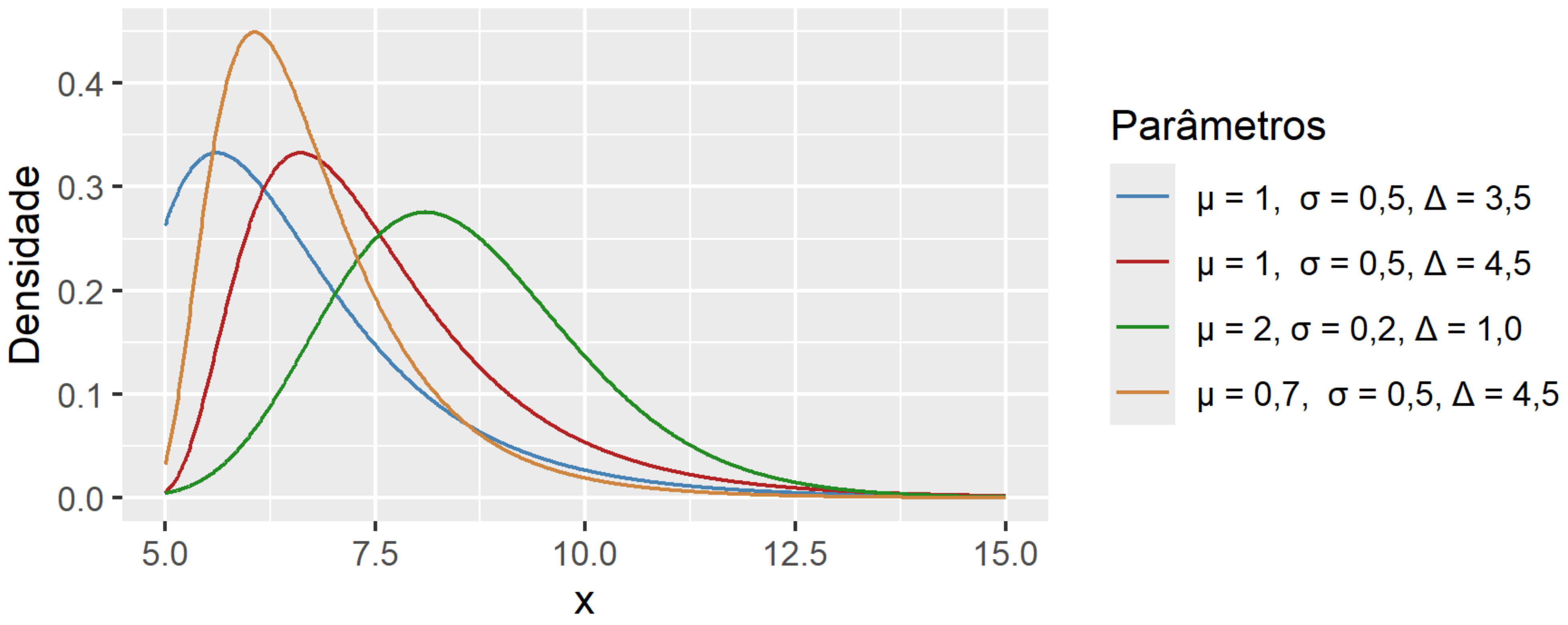
$$\sigma^2 = \exp[2(\sigma_{ln}^2 + \mu_{ln})] - \exp(\sigma_{ln}^2 + 2\mu_{ln})$$

$$\xi = \phi^3 + 3\phi$$

$$\phi = [\exp(\sigma_{ln}^2) - 1]^{0,5}$$

$$\hat{\Delta} = \frac{x_{mín}x_{máx} - \tilde{x}^2}{x_{mín} + x_{máx} - 2\tilde{x}}$$

Distribuição **Log-normal a 3 parâmetros** $(x > \Delta)$

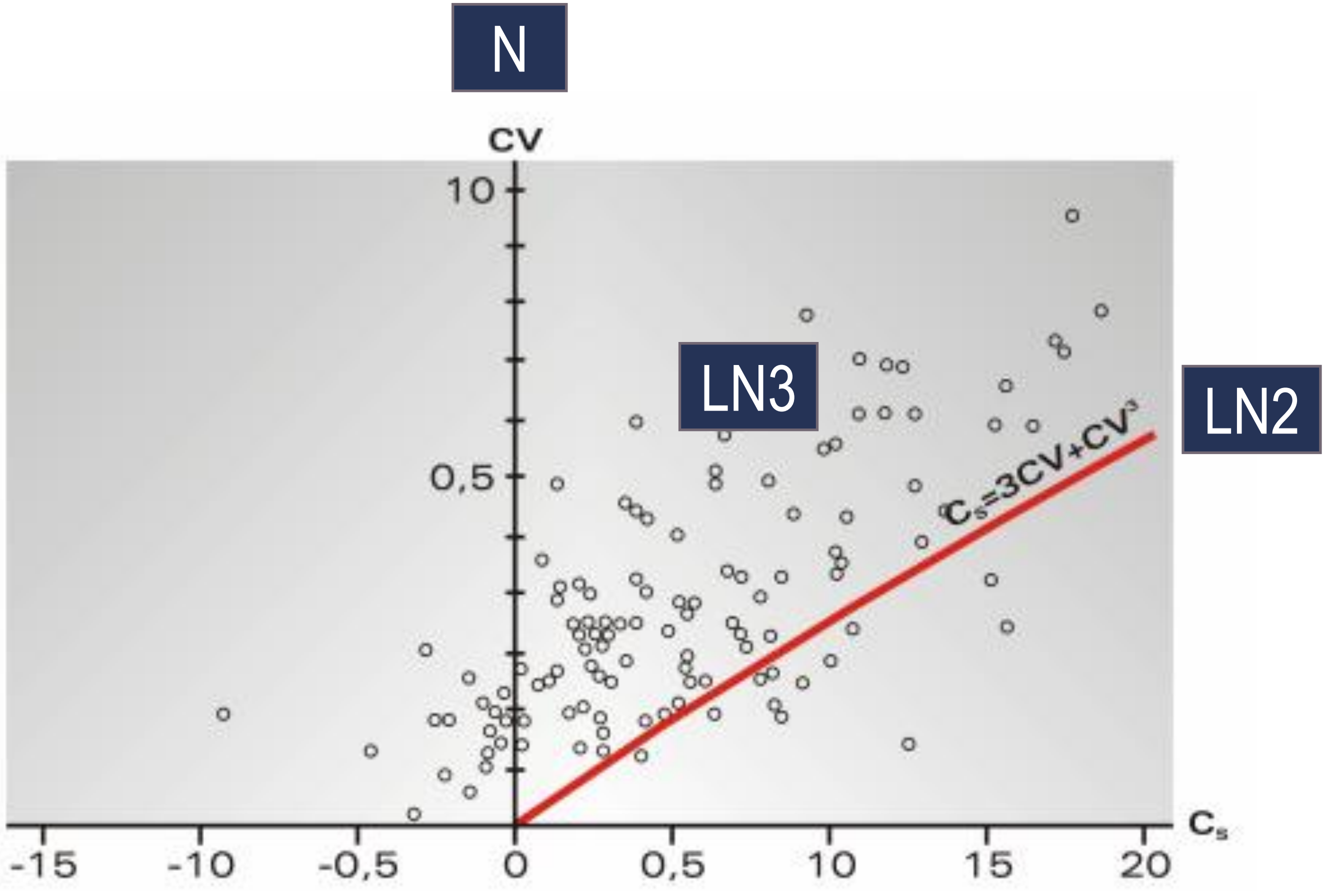


Revisão de estatística | distribuições de probabilidades

Klemes (1978) propôs o *Product Moment Diagram*

Gráfico coef. de variação vs. coef. de assimetria

Aplicado a cerca de 140 séries de vazão média anual distribuídas pelo mundo

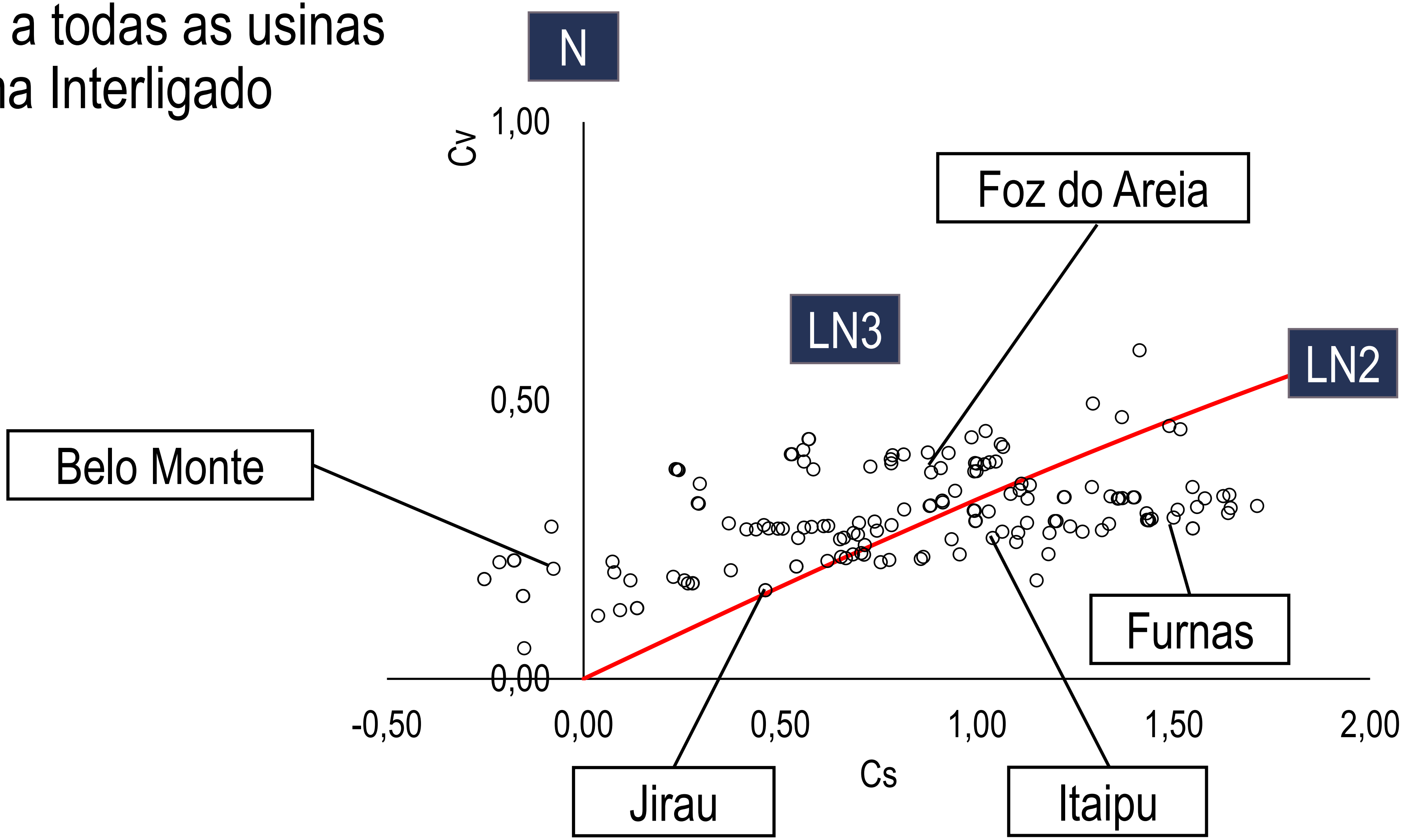


Fonte original: Klemes, V. Physically based stochastic hydrologic analysis. **Advances in Hydroscience**, n. 11, Academic Press, 1978.
Reproduzida em: Neira (2005)

Revisão de estatística | distribuições de probabilidades

Ochoa e Segundo (2016)* adaptaram para o caso brasileiro

vazões médias anuais a todas as usinas hidrelétricas do Sistema Interligado Nacional



*Estudo não publicado

Revisão de estatística | distribuições de probabilidades

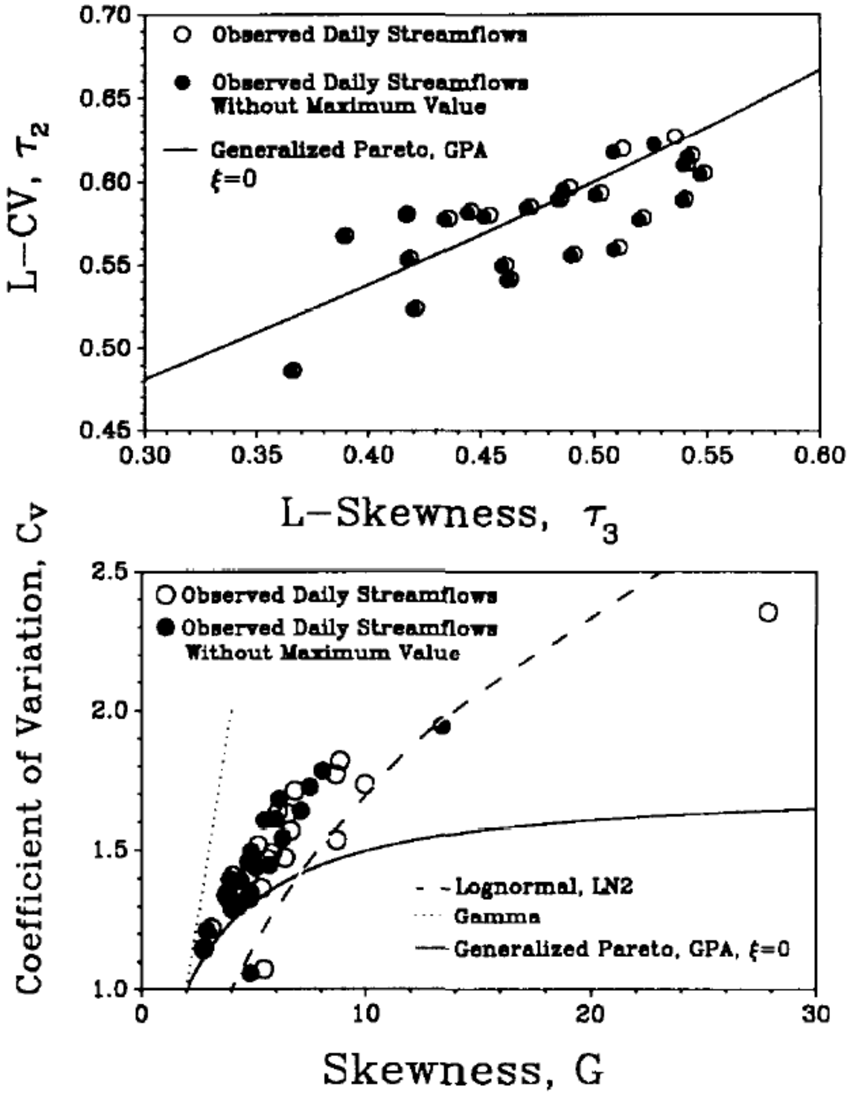
WATER RESOURCES RESEARCH, VOL. 29, NO. 6, PAGES 1745–1752, JUNE 1993

L Moment Diagrams Should Replace Product Moment Diagrams

RICHARD M. VOGEL AND NEIL M. FENNESSEY

Department of Civil and Environmental Engineering, Tufts University, Medford, Massachusetts

It is well known that product moment ratio estimators of the coefficient of variation C_v, skewness γ, and kurtosis κ exhibit substantial bias and variance for the small ($n \leq 100$) samples normally encountered in hydrologic applications. Consequently, L moment ratio estimators, termed L coefficient of variation τ_2, L skewness τ_3, and L kurtosis τ_4 are now advocated because they are nearly unbiased for all underlying distributions. The advantages of L moment ratio estimators over product moment ratio estimators are not limited to small samples. Monte Carlo experiments reveal that product moment estimators of C_v and γ are also remarkably biased for extremely large samples ($n \geq 1000$) from highly skewed distributions. A case study using large samples ($n \geq 5000$) of average daily streamflow in Massachusetts reveals that conventional moment diagrams based on estimates of product moments C_v, γ, and κ reveal almost no information about the distributional properties of daily streamflow, whereas L moment diagrams based on estimators of τ_2, τ_3, and τ_4 enabled us to discriminate among alternate distributional hypotheses.



Revisão de estatística | distribuições de probabilidades

Hydrol. Earth Syst. Sci., 21, 3093–3103, 2017
https://doi.org/10.5194/hess-21-3093-2017

Hydrology and Earth System Sciences
Open Access

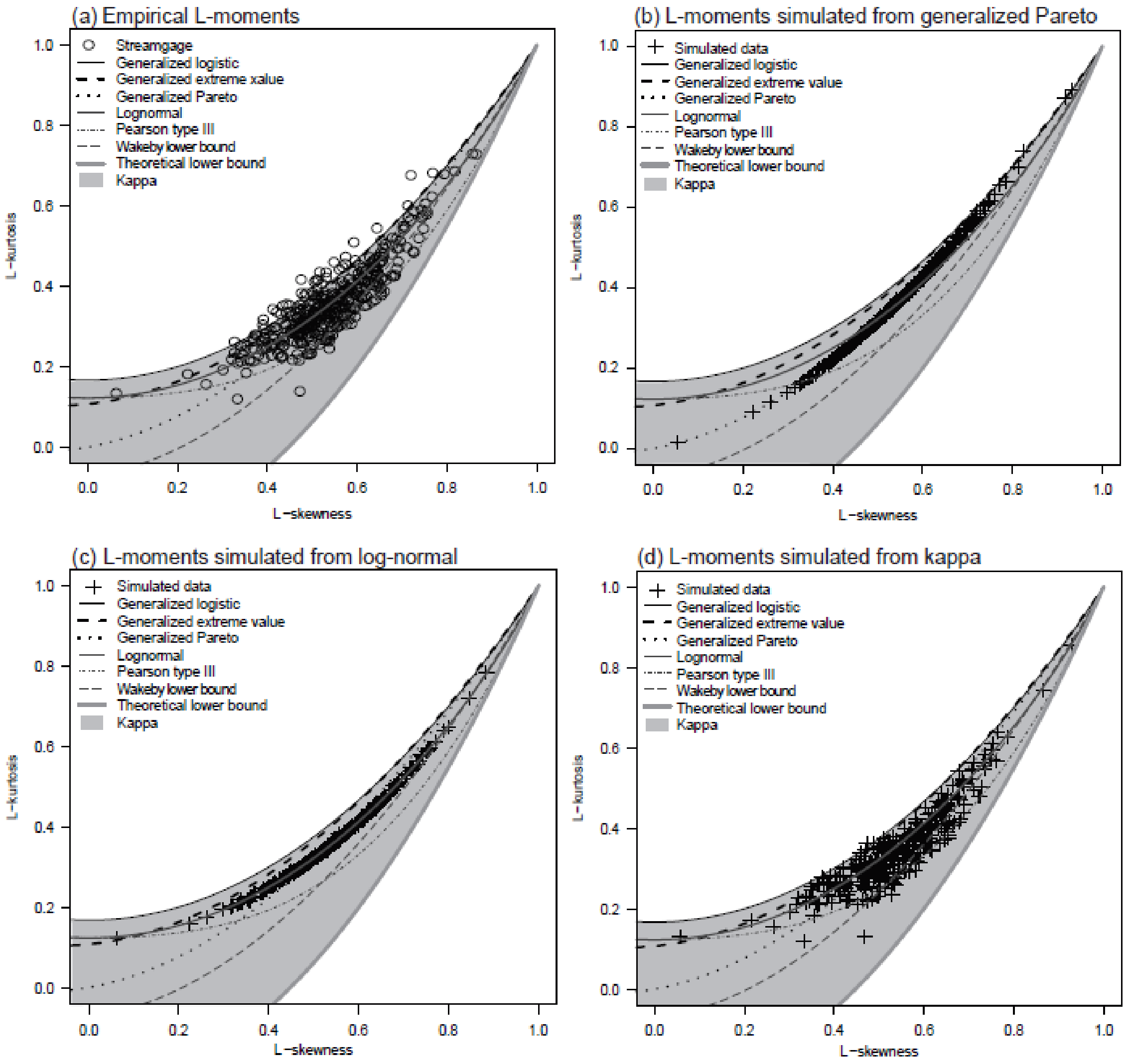
On the probability distribution of daily streamflow in the United States

Annalise G. Blum[1,2], Stacey A. Archfield[2], and Richard M. Vogel[1]

[1]Civil and Environmental Engineering, Tufts University, Medford, MA 02155, USA
[2]U.S. Geological Survey, Reston, VA 20192, USA

Correspondence to: Stacey A. Archfield (sarch@usgs.gov)

Received: 2 September 2016 – Discussion started: 22 September 2016
Revised: 24 March 2017 – Accepted: 19 May 2017 – Published: 28 June 2017



Revisão de estatística | distribuições de probabilidades

Como escolher a distribuição a ser adotada?

1. Avaliação por métodos gráficos
2. Aplicação de testes de hipótese para adequação de ajustes
   (assunto do próximo tópico)

1. Métodos Gráficos

1.1. Histograma/densidades

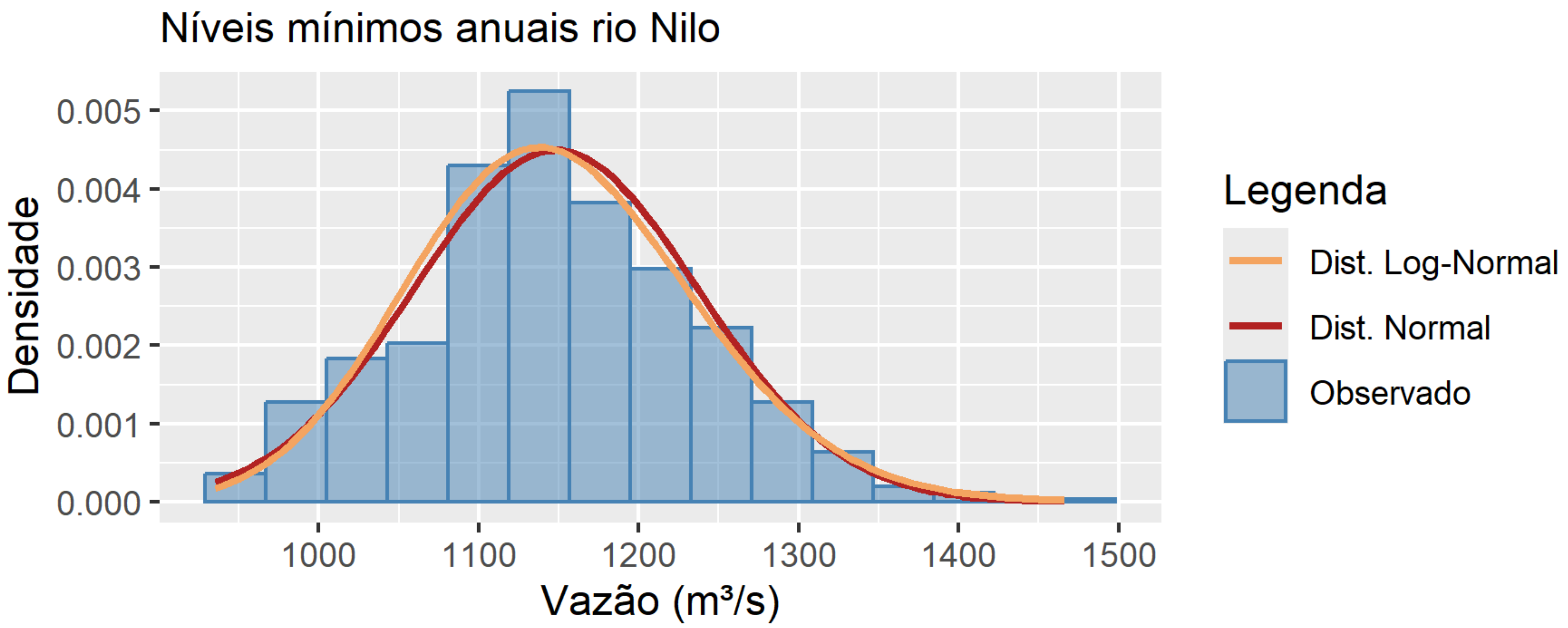
1.2. Frequências acumuladas

1.3. Gráficos quantil-quantil (*Q-Q plots*)

Revisão de estatística | distribuições de probabilidades

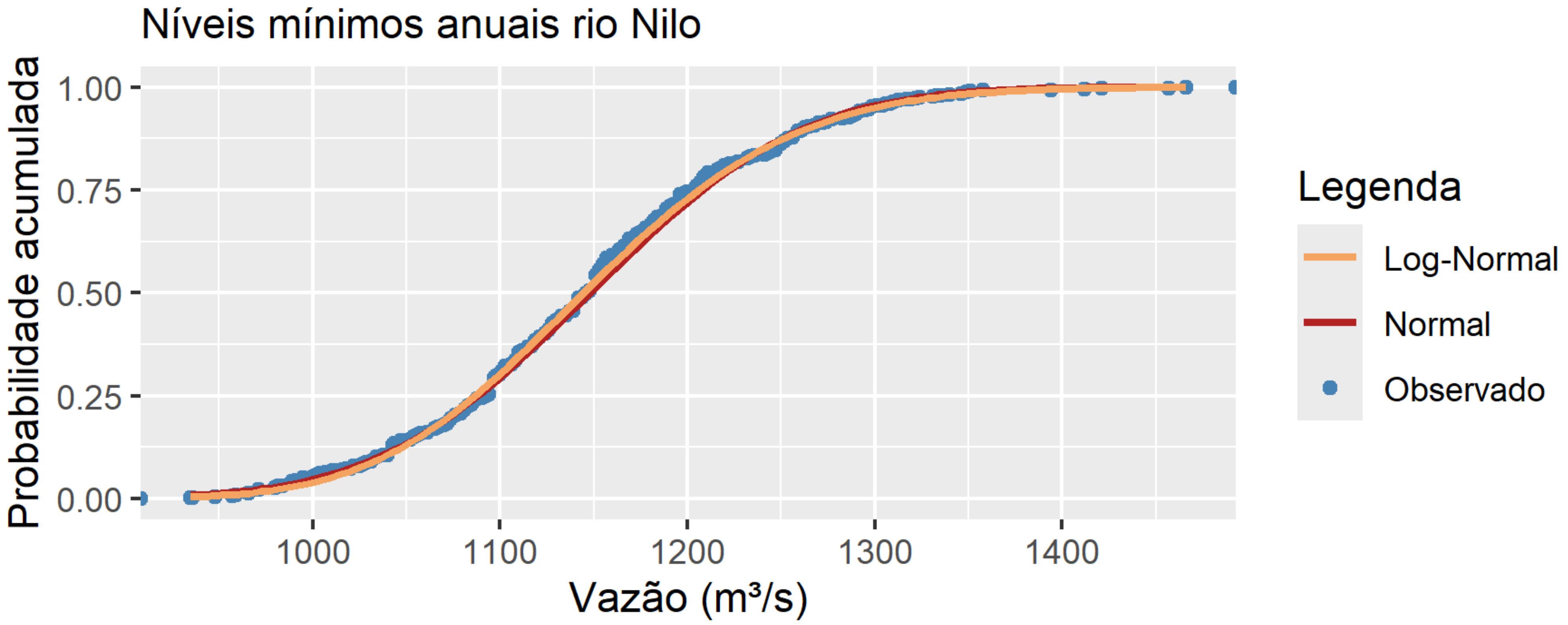
1.1. Histograma/densidades: comparação entre o histograma (empírico) e a função densidade de probabilidades (FDP)

o que se busca: adequação da densidade ao histograma



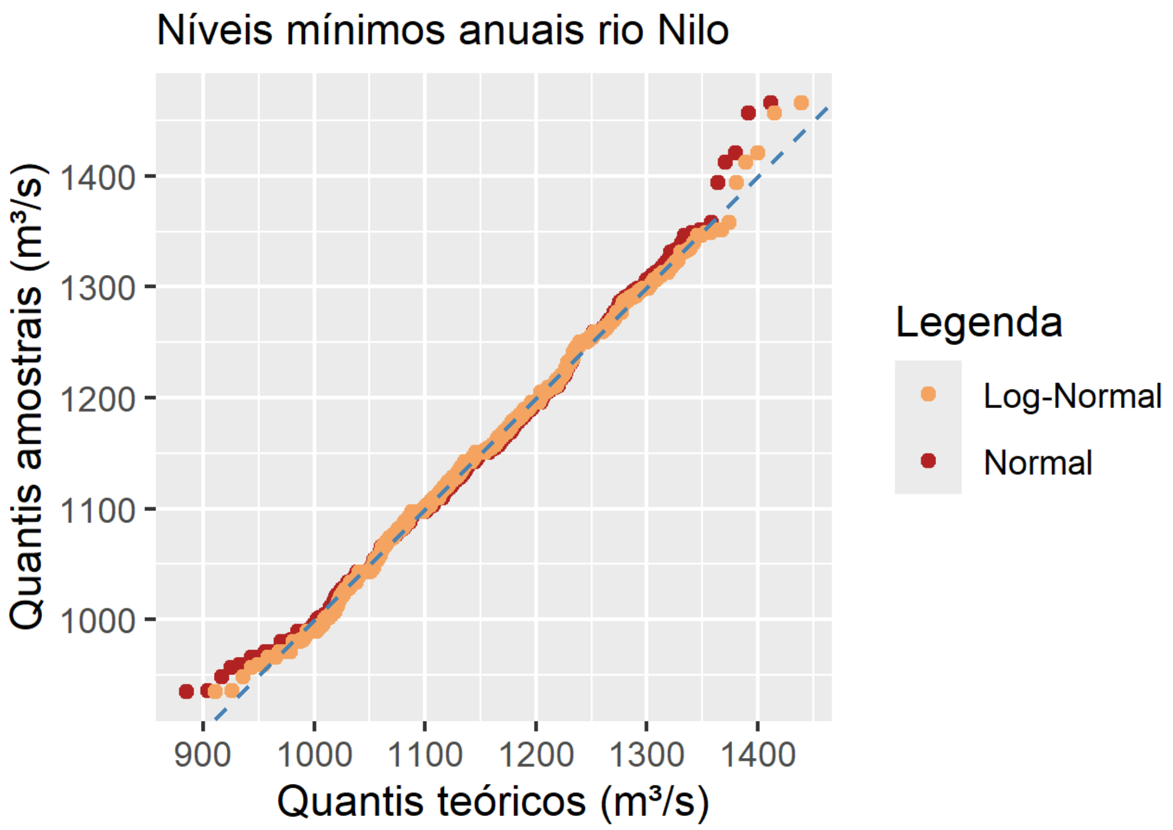
1.2. Frequências acumuladas: comparação entre as distribuições acumuladas empírica e teórica (FDA)

o que se busca: adequação da FDA aos dados observados



1.3. Gráficos quantil-quantil (*Q-Q plots*): comparação entre quantis empíricos e teóricos

o que se busca: dados observados sobre a linha reta (tipicamente de 45°)



REVISÃO DE ESTATÍSTICA

testes de hipótese

Testes de hipótese:

Regra de decisão para rejeitar, ou não, uma hipótese com base na amostra

Hipótese: suposição que algo pode ser verdade, ou não. Deve ser passível de testes

Hipóteses estatísticas:

H_0 – hipótese nula: o que se assume ser verdade

H_1 – hipótese alternativa: o que se assume ser verdade quando da rejeição de H_0

ex. de H_0: a distribuição dos dados é Normal; as amostras não têm a mesma média; a série não tem tendência; etc.

Revisão de estatística | testes de hipótese

Cada teste possui um procedimento matemático específico

Em comum, todos calculam a probabilidade (p-valor) de os dados terem sido observados assumindo que H_0 é verdadeira

$$P(\text{Dados} \mid H_0 \text{ verdadeira}) = \text{p-valor}$$

Por sua vez, o analista assume um risco de se tomar a decisão errada

- em particular, rejeitar H_0 quando ela é verdadeira
- o risco é quantificado pelo nível de significância α
- valores típicos de α: 1%, 5% e 10%

Revisão de estatística | testes de hipótese

Em suma:

p-valor $\neq \alpha$

Calculado usando a amostra

Atribuído pelo analista

Se o p-valor:

Baixo (p-valor $\leq \alpha$) → rejeita-se a hipótese nula

Alto (p-valor $> \alpha$) → falha-se em rejeitar a hipótese nula

Revisão de estatística | testes de hipótese

Testes de hipótese são construídos com base em premissas

- os dados são independentes e identicamente distribuídos
- as amostras possuem a mesma variância (homocedásticas)

Adicionalmente, testes do tipo paramétrico assumem que a distribuição de probabilidade dos dados é conhecida

- o analista precisa adotar uma distribuição para os dados
- perdem poder se a distribuição adotada não for adequada

Testes não paramétricos, não têm essa premissa

- perdem poder se a distribuição de probabilidades dos dados é conhecida

Revisão de estatística | distribuições de probabilidades

Como escolher a distribuição a ser adotada?

1. Avaliação por métodos gráficos
2. Aplicação de testes de hipótese para adequação de ajustes

2. Testes de adequação de ajustes

2.1. Shapiro-Wilk (para normalidade)

2.2. Anderson-Darling

2.3. Filliben/PPCC (*probability plot correlation coeficient*)

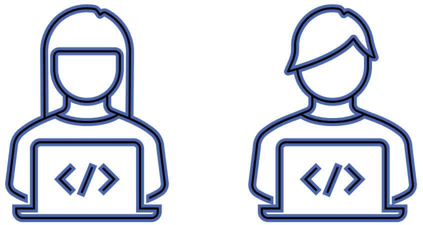
Revisão de estatística | distribuições de probabilidades

2.1. Shapiro-Wilk (para normalidade)

H_0: a amostra vem de uma distribuição normal

H_1: a amostra não vem de uma distribuição normal

```
Shapiro-Wilk normality test
W = 0.99329, p-value = 0.004614
```



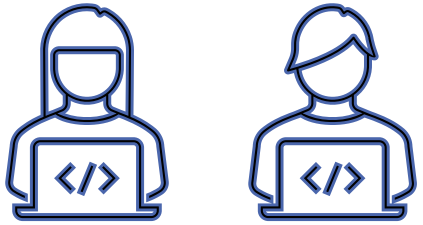
Veredicto: rejeita-se H_0 (a distribuição não é normal)

2.2. Anderson-Darling

H_0: a amostra vem de uma distribuição especificada
H_1: a amostra não vem de uma distribuição especificada

```
Anderson-Darling test of goodness-of-fit
Null hypothesis: Normal distribution
Anmax = 1.5552, p-value = 0.9905

Null hypothesis: log-normal distribution
Anmax = 3.3899, p-value = 0.3714
```



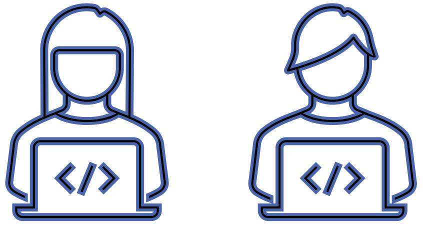
Veredicto: falha-se em rejeitar H_0 (a distribuição é normal)

2.3. PPCC

H_0: a amostra vem de uma distribuição especificada
H_1: a amostra não vem de uma distribuição especificada

```
Probability Plot Correlation Coefficient Test
Null hypothesis: Normal distribution
ppcc = 0.99677, n = 663, p-value = 0.0073

Null hypothesis: log-normal distribution
ppcc = 0.82211, n = 663, p-value = 0.0039
```



Veredicto: rejeita-se H_0 (a distribuição não é normal)

Resumo

A hidrologia estocástica é empregada como ferramenta para modelar fenômenos hidrológicos que parecem se comportar aleatoriamente

- a evolução temporal ocorre de acordo com leis probabilísticas

O foco está em variáveis aleatórias contínuas que evoluem em função do tempo

Base estatística

- análise exploratória de dados
- distribuição de probabilidades
- testes de hipótese

ERHA7016 – Hidrologia Estocástica

Daniel Detzel
detzel@ufpr.br