



Estatística Aplicada a Ciências Ambientais

Introdução

Daniel Detzel
detzel@ufpr.br

INTRODUÇÃO

apresentação da disciplina

Introdução | generalidades

Apresentações individuais

- nome

- área de concentração

- conhecimento de estatística

- conhecimento de programação (R, Python, Matlab, etc.)

- o que espera da disciplina

Por que estatística?

- sintetizar e descrever dados

- inferir atributos

- avaliar incertezas, riscos e probabilidades

- identificar padrões e relações



Por que é tão difícil para nós pensarmos estatisticamente?

pensamos associativamente com facilidade, pensamos metaforicamente, pensamos casualmente, mas estatísticas requerem que pensemos sobre muitas coisas de uma vez, coisa que o sistema intuitivo da mente humana **não está projetado** para fazer.

Daniel Kahneman

Rápido e Devagar: duas formas de pensar



A legislação ambiental/hídrica/energética faz uso de estatística

“(…) não deverá ser excedido um limite de 200 coliformes termotolerantes por 100 mililitros em 80% ou mais, de pelo menos 6 amostras, coletadas durante o período de um ano, com frequência bimestral.”
(CONAMA Nº357/2005)

“A vazão mínima a ser garantida para jusante da barragem deve ser aquela correspondente à 95% da curva de permanência.”
(Manual de Outorga IAT)

“(…) compatibilizar a energia assegurada do sistema (...) ao critério de garantia de suprimento de risco de déficit máximo de 5%.”
(NT MME/SPD/05)

Introdução | ciências ambientais e estatística

Estatística e a verdade

Tomadores de decisão querem **fatos**, mas...

incertezas nas medições produzem dados imperfeitos

os fatores que influenciam os fenômenos não são conhecidos em sua totalidade

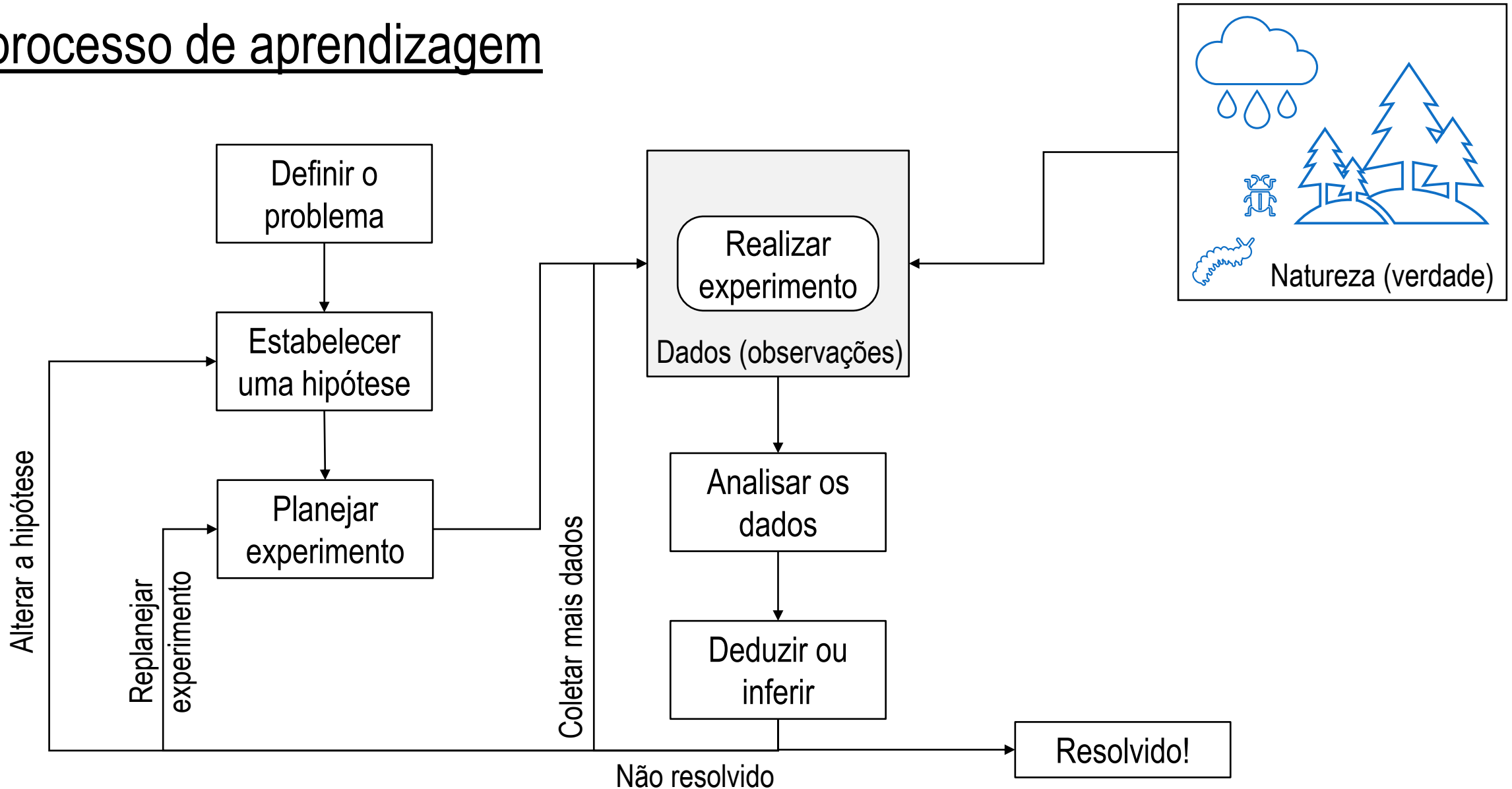
A noção de verdade (ou realidade) pode ser **aproximada** por meio da estatística



Recomendação: não
utilizem estes termos!

Introdução | ciências ambientais e estatística

O processo de aprendizagem



Introdução | ciências ambientais e estatística

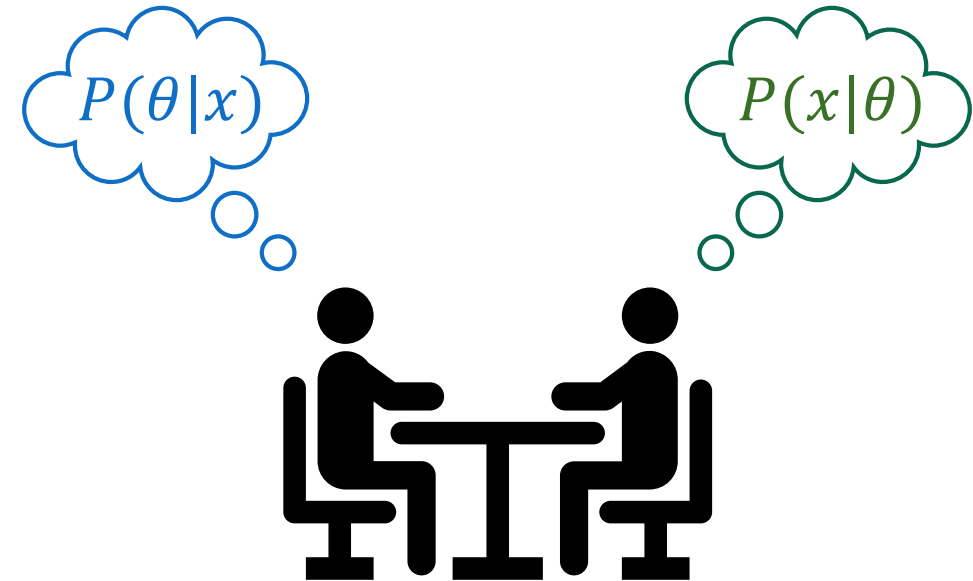
Abordagem frequentista

a incerteza está nos **dados** e não na hipótese
requer experimentação/coleta de dados

Abordagem bayesiana

a incerteza está na **hipótese** e não nos dados
requer conhecimento *a priori*
pode mudar/adaptar a hipótese

Neste curso: abordagem **frequentista**



Introdução

processos hidrológicos e ambientais; variáveis hidrológicas e de qualidade da água; séries temporais; população e amostra; introdução ao RStudio e à linguagem R

Análise preliminar de dados

características dos dados hidrológicos e de qualidade da água; representação gráfica de dados; medidas de tendência central, medidas de variabilidade, medidas de assimetria

Fundamentos da Teoria de Probabilidades

eventos aleatórios; medidas de probabilidade; probabilidade condicional e independência; variáveis aleatórias; medidas descritivas de variáveis aleatórias

Distribuições de Probabilidade

distribuições discretas; distribuições contínuas; estimação de parâmetros; incerteza

Testes de Hipótese

estrutura dos testes de hipótese; testes paramétricos para diferenças entre grupos independentes (médias e variâncias); testes não paramétricos para aleatoriedade, independência, homogeneidade e estacionariedade; testes para adequação de ajustes de distribuições de probabilidade

Correlação e Regressão

medidas de correlação (Pearson, Spearman e Kendall); regressão linear simples; construção do modelo de regressão; diagnóstico do modelo de regressão

Objetivo: curso aplicado!

(a teoria estatística não será aprofundada)

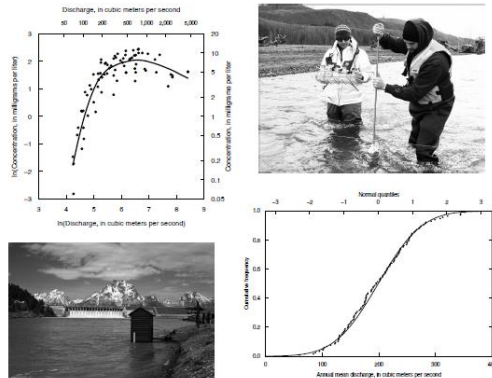


Introdução | bibliografia recomendada



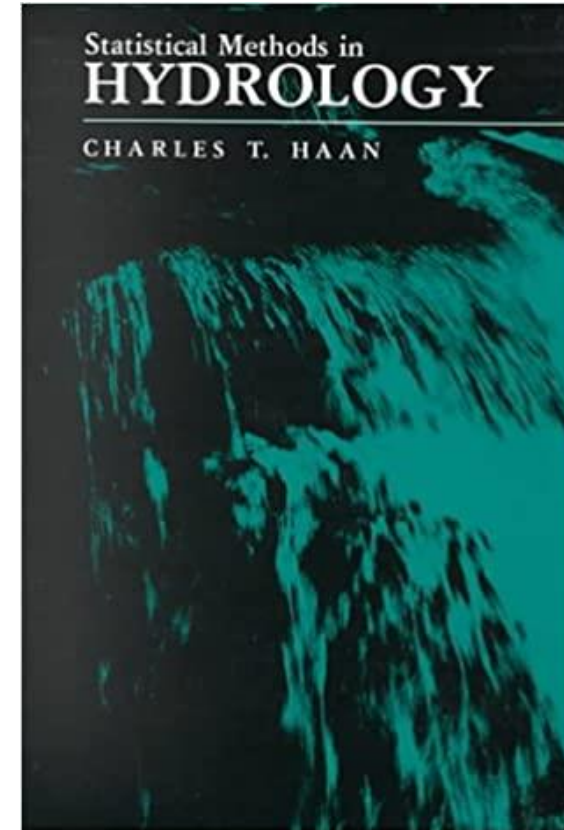
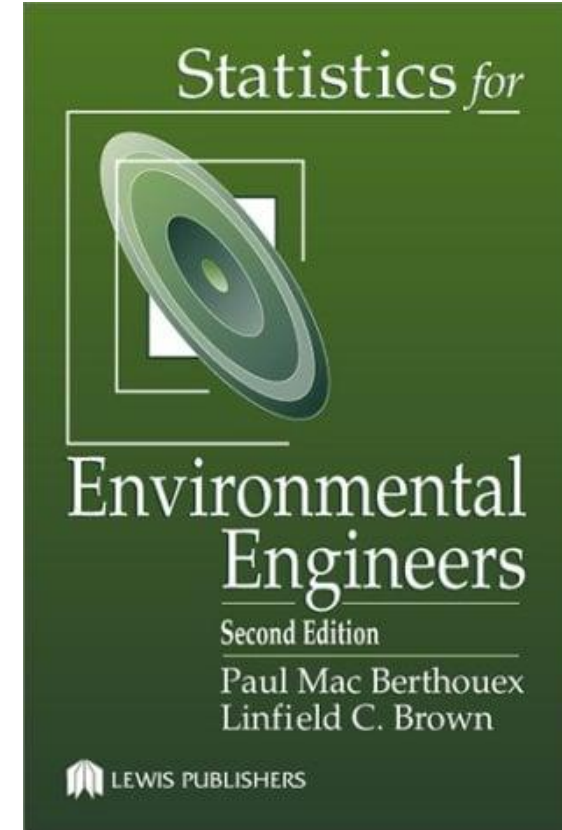
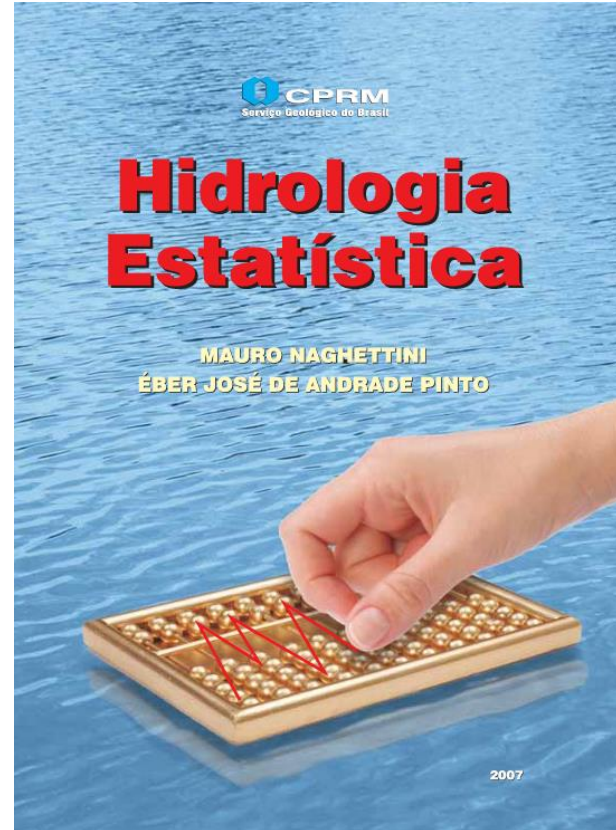
Statistical Methods in Water Resources

Chapter 3 of
Section A, Statistical Analysis
Book 4, Hydrologic Analysis and Interpretation



Techniques and Methods 4-A3
Supersedes USGS Techniques of Water-Resources Investigations, book 4,
chapter A3

U.S. Department of the Interior
U.S. Geological Survey



(ou qualquer outro livro de estatística...)

Introdução | avaliação

Tarefas diversas propostas nas aulas
peso 30%

Trabalho 1: análise de dados, representação gráfica e estatísticas
descritivas
peso 40%

Trabalho 2: probabilidades, distribuições e testes de hipótese
peso 40%

Sugestão: tragam os seus dados!
séries temporais diversas serão disponibilizadas



INTRODUÇÃO

conceitos iniciais

Introdução | conceitos iniciais

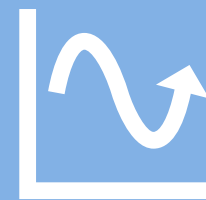
Processos
naturais



Variáveis
naturais



Séries
temporais



Introdução | processos naturais

Processos naturais

conjunto de mecanismos que definem um fenômeno natural

Exemplos:

Processos hidrológicos: mecanismos de armazenamento e transporte de água entre as fases do ciclo hidrológico

Eutrofização natural: transporte de material mineral e orgânico para uma massa d'água, aumentando a capacidade de produção biológica local

Águas residuárias: aporte de nutrientes provenientes de atividades humanas

Introdução | processos naturais

Intensidade de ocorrência de processos naturais são função do **tempo**, **espaço** e **escala geográfica**

técnicas estatísticas são selecionadas de acordo com esses aspectos

Processos determinísticos

aplicações diretas de leis da Física, Química ou Biologia

Processos estocásticos

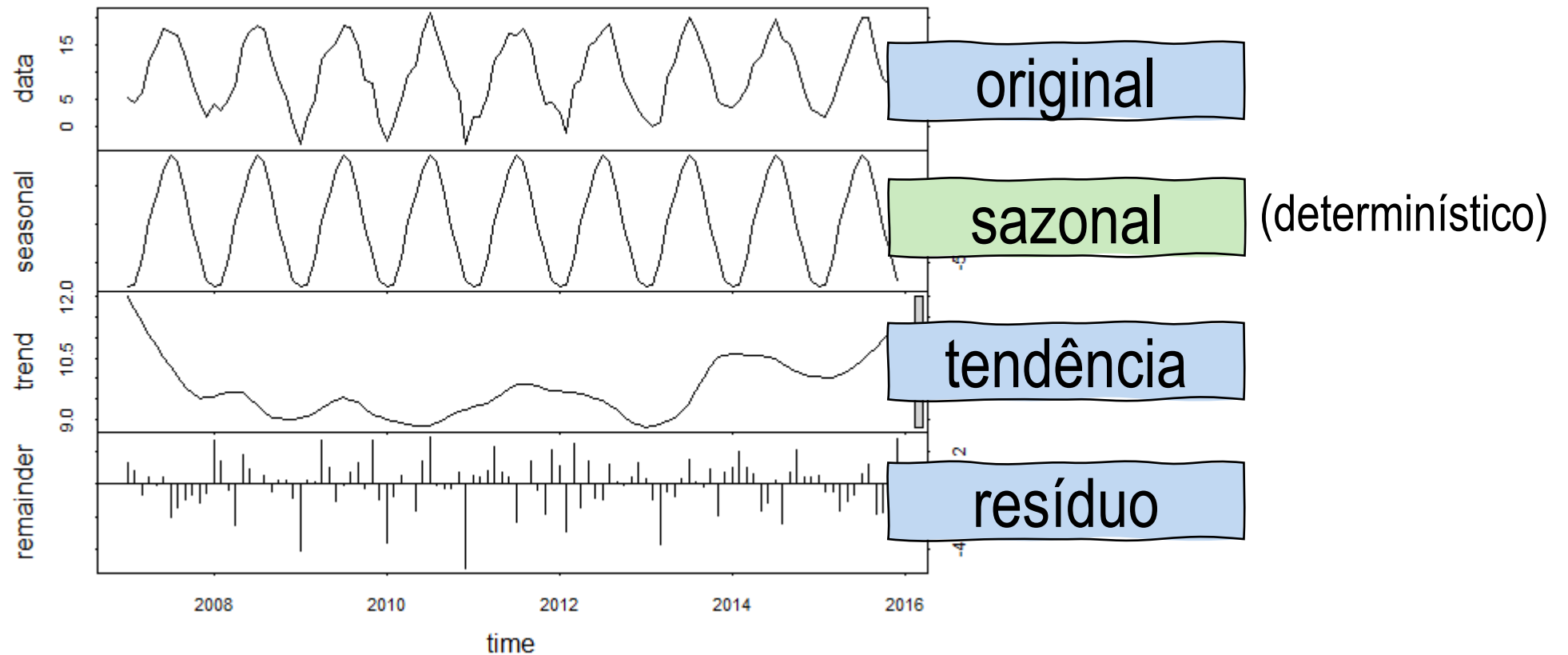
governados por leis probabilísticas

possuem componentes aleatórias

Processos naturais são formados por uma **combinação de ambos**

Introdução | processos naturais

Ex.: componentes da temperatura média do ar em Cölbe (Alemanha)



Introdução | variáveis naturais

Variáveis naturais

descrevem as variações temporais e espaciais dos processos naturais

Exemplos:

Variáveis hidrológicas: vazão média anual na seção de um rio; nº de dias chuvosos em um ano; evaporação líquida mensal em um reservatório; etc.

Variáveis de qualidade da água (em uma amostra): temperatura; pH; oxigênio dissolvido; turbidez; condutividade elétrica; conteúdo de matéria orgânica; etc.

Introdução | variáveis naturais

São quantificadas ou categorizadas por meio de observações, executadas de acordo com procedimentos padronizados

Exemplos:

Chuva média diária em um local: média das leituras feitas às 7:00 e 17:00 em um pluviômetro

Nível de oxigênio dissolvido em uma seção de rio: leitura proveniente de sonda em dia e horário definidos

Introdução | variáveis naturais

Variáveis qualitativas: expressas por **atributos**

nominais: não podem ser classificadas ou ordenadas

ex.: estados do tempo {'claro', 'chuvoso', 'nublado'}

ordinais: permitem classificação

ex.: níveis de reservatórios {'baixo', 'médio', 'alto'}

Variáveis quantitativas: expressas por **números**

discretas: números inteiros

ex.: n^o de semanas em que a turbidez da água permaneceu elevada

contínuas: números reais

ex.: altura diária máxima anual de precipitação

Introdução | variáveis naturais

Variáveis quantitativas: expressas por números (cont.)

limitadas: restritas a um intervalo, por razões físicas ou de medição

ex.: concentração de oxigênio dissolvido em um corpo hídrico

ilimitada: não possuem restrições inferiores e/ou superiores

ex.: vazão máxima em uma seção de um rio

univariadas: recebem um atributo de quantidade ou qualidade

ex.: radiação solar global em um ponto

multivariadas: recebem mais de um atributo de quantidade ou qualidade

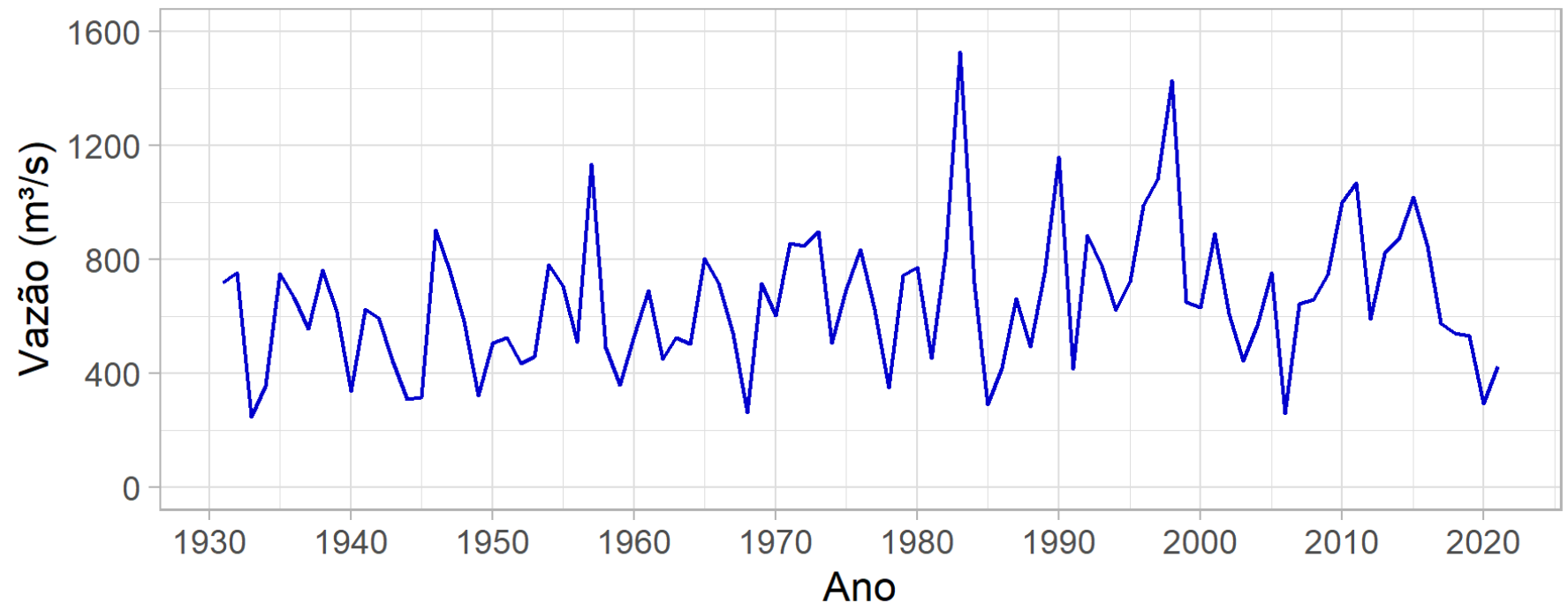
ex.: concentração de fósforo total em diferentes reservatórios de uma bacia

Introdução | séries temporais

Sequenciamento temporal da ocorrência de variáveis naturais

Podem ser organizadas em **intervalos fixos** de tempo comuns em variáveis hidrometeorológicas

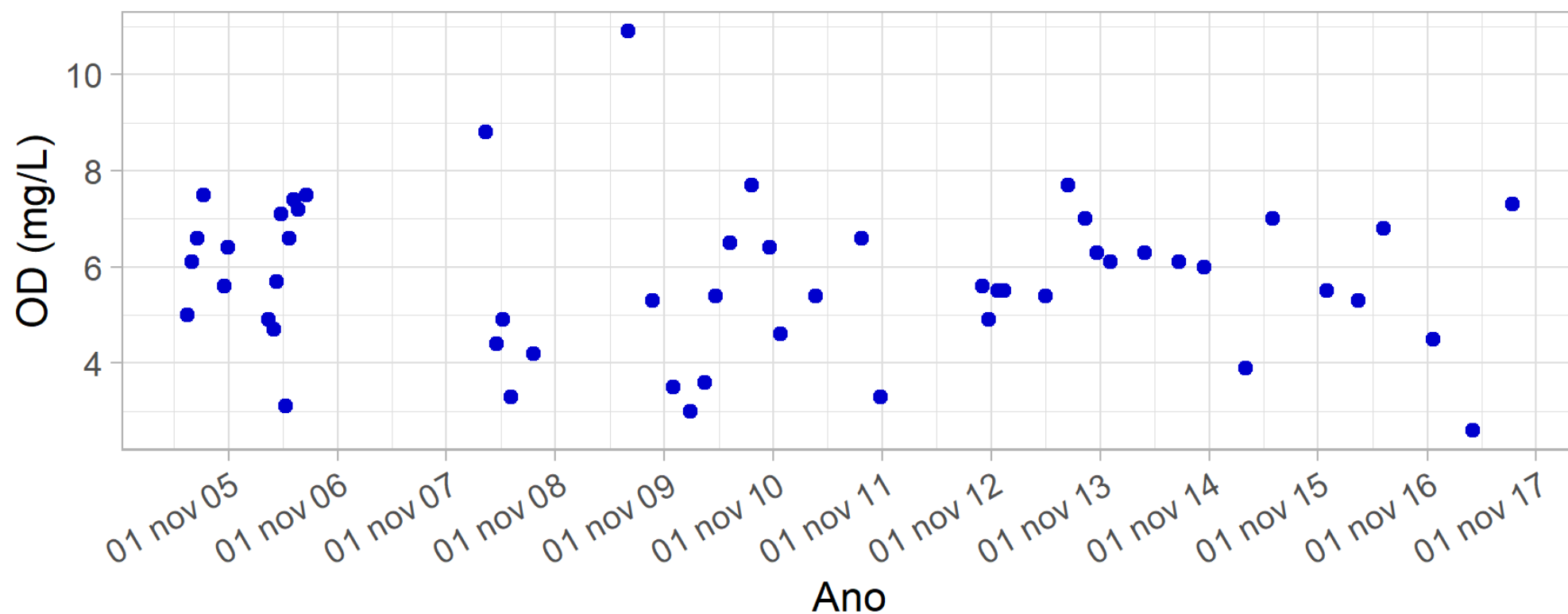
Ex.: vazões médias anuais em Foz do Areia (rio Iguaçu)



Introdução | séries temporais

Porém, dependendo do **método de medição** ou **estratégia de amostragem**, podem ser compostas por intervalos não equidistantes comuns em parâmetros de qualidade da água

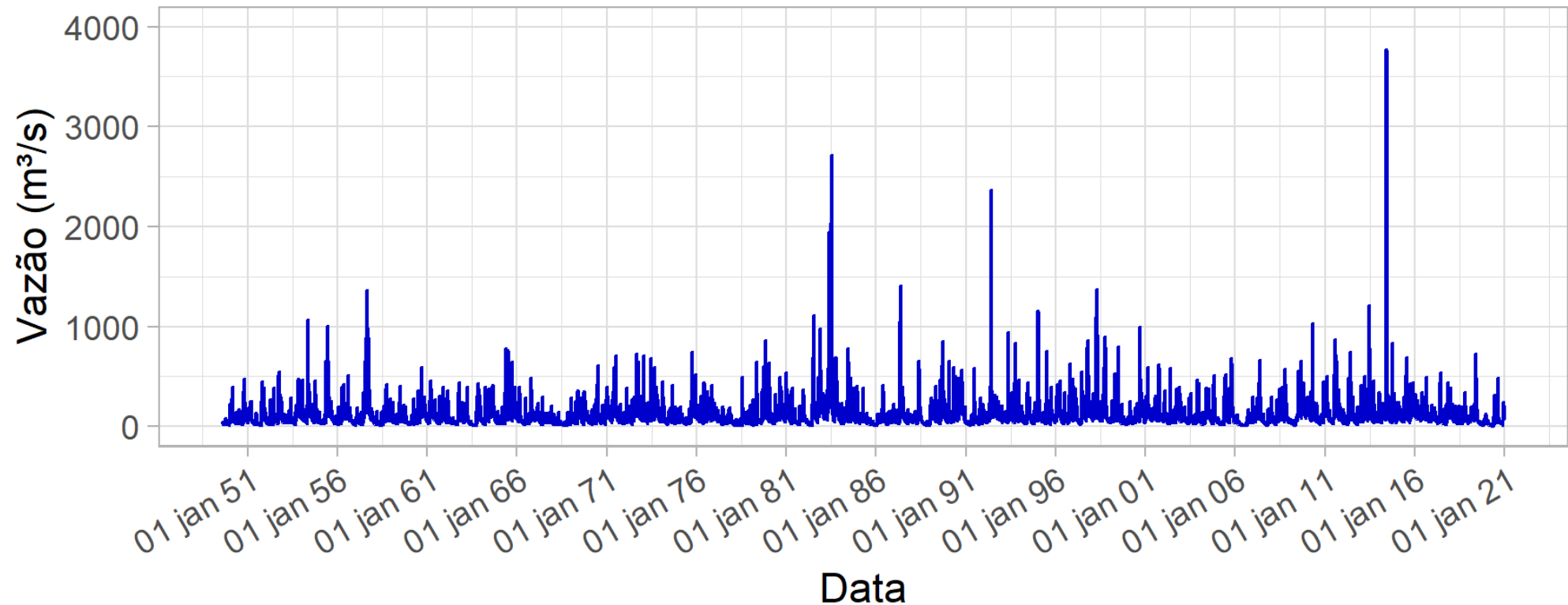
Ex.: oxigênio dissolvido no rio Iguaçu, posto IG1 (Alto Iguaçu)



Introdução | séries temporais

Séries completas: formadas por todos os registros medidos

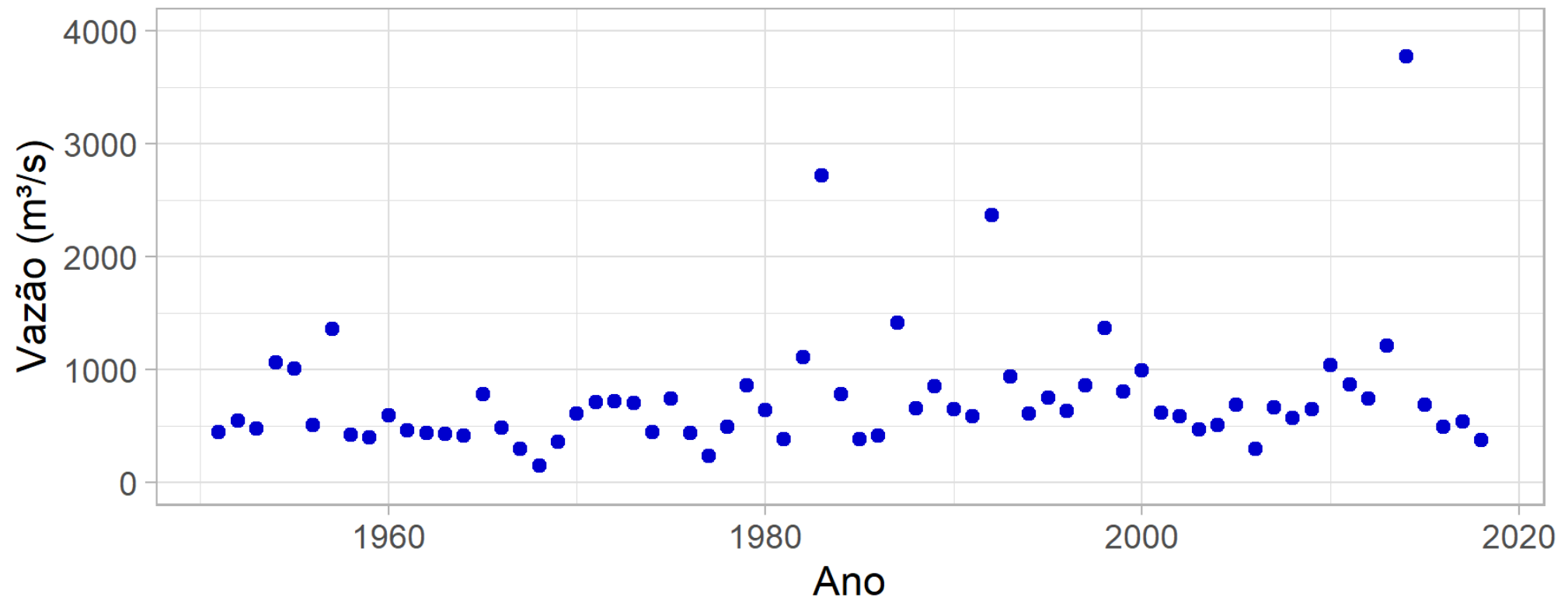
Ex.: vazões médias diárias em Santa Clara (rio Jordão, PR)



Introdução | séries temporais

Séries reduzidas: formadas parte dos registros, ou um resumo destes

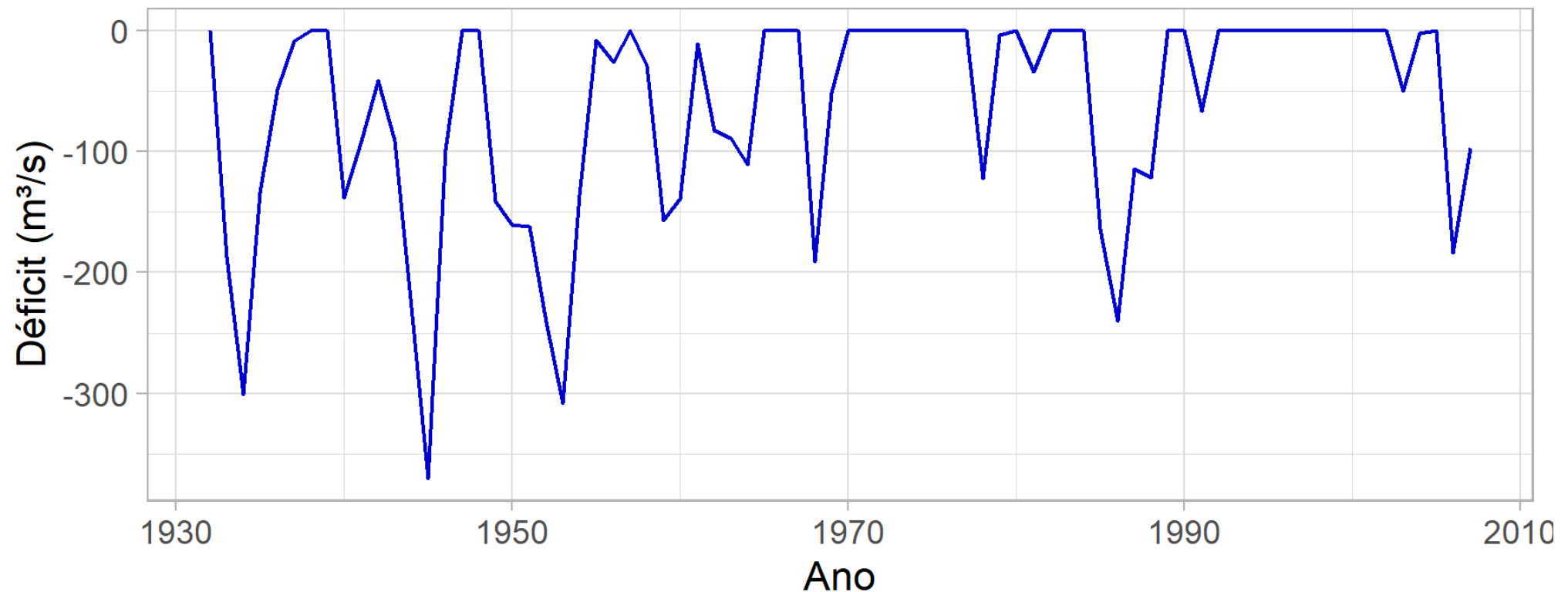
Ex.: vazões máximas diárias anuais em Santa Clara (rio Jordão, PR)



Introdução | séries temporais

Séries censuradas: limitadas por valores inferiores e/ou superiores

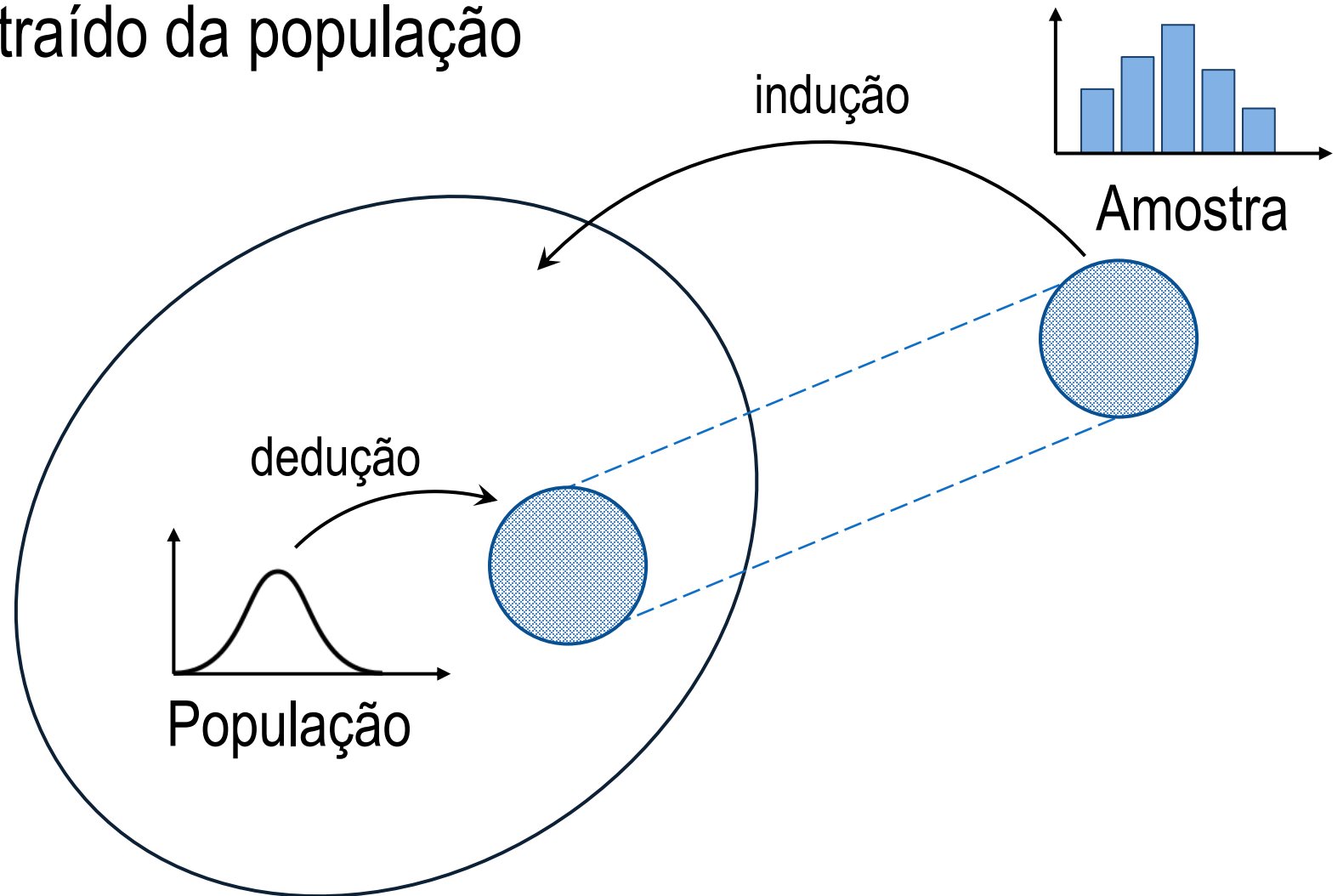
Ex.: déficits hídricos em União da Vitória (rio Iguaçu, PR)



Introdução | população e amostra

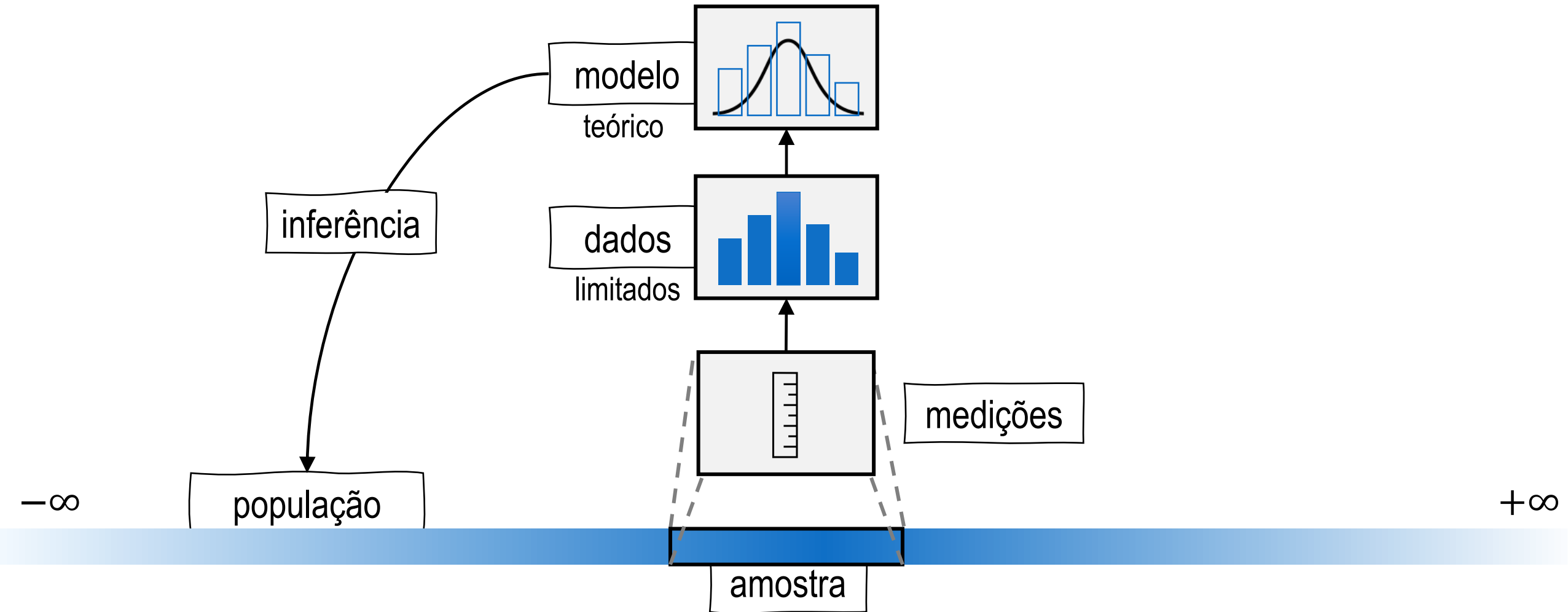
População: **todos** os resultados possíveis de uma variável

Amostra: **subconjunto** extraído da população



Introdução | população e amostra

Para séries temporais de variáveis naturais



Introdução | população e amostra

Ex.: considere a série de medições de oxigênio dissolvido no posto IG1:

Período de medições: 2005 a 2017

Tamanho da **amostra**: 59 elementos

Valor mínimo medido: 2,6 mg/L no dia 03 mar 17

Valor máximo medido: 10,9 mg/L no dia 01 jul 2009

Qual é a probabilidade de que os valores de OD sejam inferiores a 2,6 mg/L ou superiores a 10,9 mg/L no posto IG1?

R.: Nula!

(mas isso não quer dizer que tais valores nunca sejam observados na **população**!)

Introdução | população e amostra

A amostra precisa ser representativa da população

Propriedades a serem preservadas

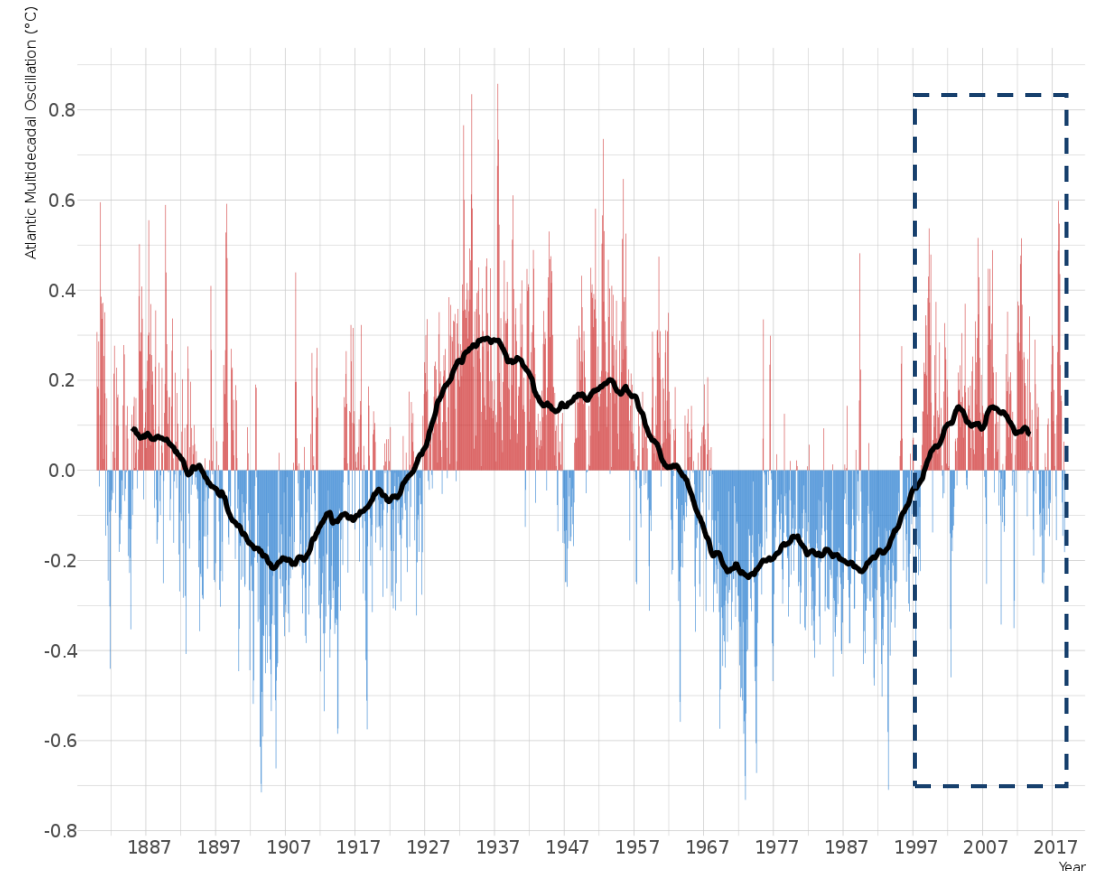
estacionariedade

homogeneidade

equiprobabilidade

independência entre elementos

Ex.: Índice Multidecadal do Atlântico
(anomalias da temperatura da superfície do mar,
medidas no Atlântico Norte)



Revisão

Processos naturais \neq variáveis naturais

Variáveis são categorizadas

- qualitativas (nominais ou ordinais)

- quantitativas (discretas, contínuas, ilimitadas, limitadas, uni-, multivariadas)

Séries temporais: organização cronológica das variáveis medidas

podem ser caracterizadas de diferentes maneiras

- intervalos de tempo equidistantes, ou não

- completas

- reduzidas

- censuradas

População \neq amostra

Para a próxima aula

RStudio instalado!
(ou software similar)

Selecionar pelo menos 3 séries temporais de livre escolha
algumas opções são fornecidas na pasta compartilhada da disciplina
idealmente séries com **pelo menos 30 elementos**
procurar séries **diferentes entre si**

Caracterizar de acordo com os critérios expostos em aula:
processo natural, variável natural, série temporal

Variáveis quantitativas, qualitativas, discretas, contínuas, limitada, ilimitada, reduzida, censurada, etc.



Estatística Aplicada a Ciências Ambientais

Daniel Detzel
detzel@ufpr.br