



CEFET/RJ

Detecção de Pontos de Mudança e Desvios de Conceito



Eduardo Ogasawara

eduardo.ogasawara@cefet-rj.br

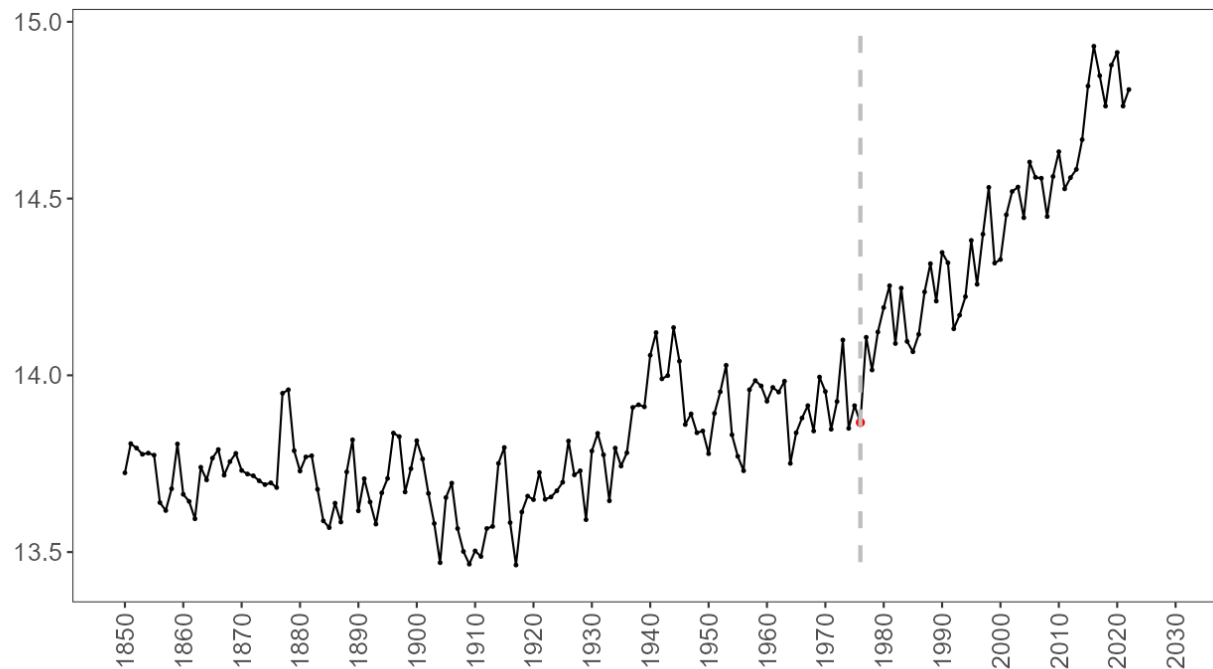
<https://eic.cefet-rj.br/~eogasawara>

Introdução e Aplicações

- A detecção de pontos de mudança em séries temporais permite identificar transições importantes no comportamento de sistemas dinâmicos
- Essas mudanças podem indicar:
 - Quebras em tendências econômicas ou financeiras
 - Falhas em máquinas e equipamentos
 - Mudanças no comportamento de usuários
 - Alterações em sinais biomédicos ou ambientais
 - Detectar essas transições é essencial para monitoramento, diagnóstico e tomada de decisão

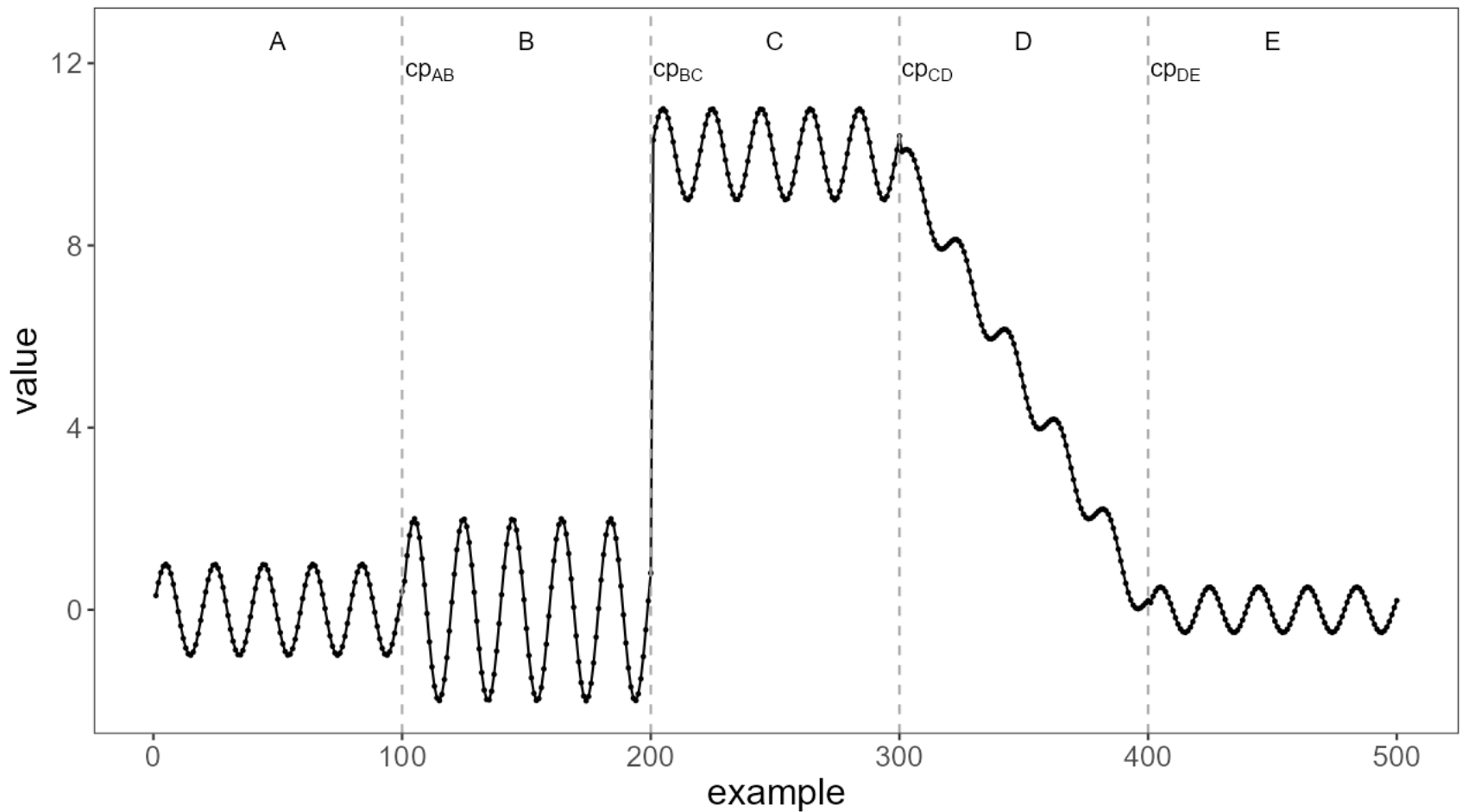
O que são pontos de mudança?

- Pontos de mudança indicam momentos em que a série temporal sofre uma transição significativa
- Essas transições podem afetar a média, variância ou distribuição dos dados
- Detectá-los é essencial para análise, previsão e tomada de decisão em séries temporais
- $cp(X, k, \sigma) = \{t, |tc(x_t) - ep(x_t, k)| > \sigma \vee |tc(x_t) - ef(x_t, k)| > \sigma \vee |ep(x_t, k) - ef(x_t, k)| > \sigma\}$



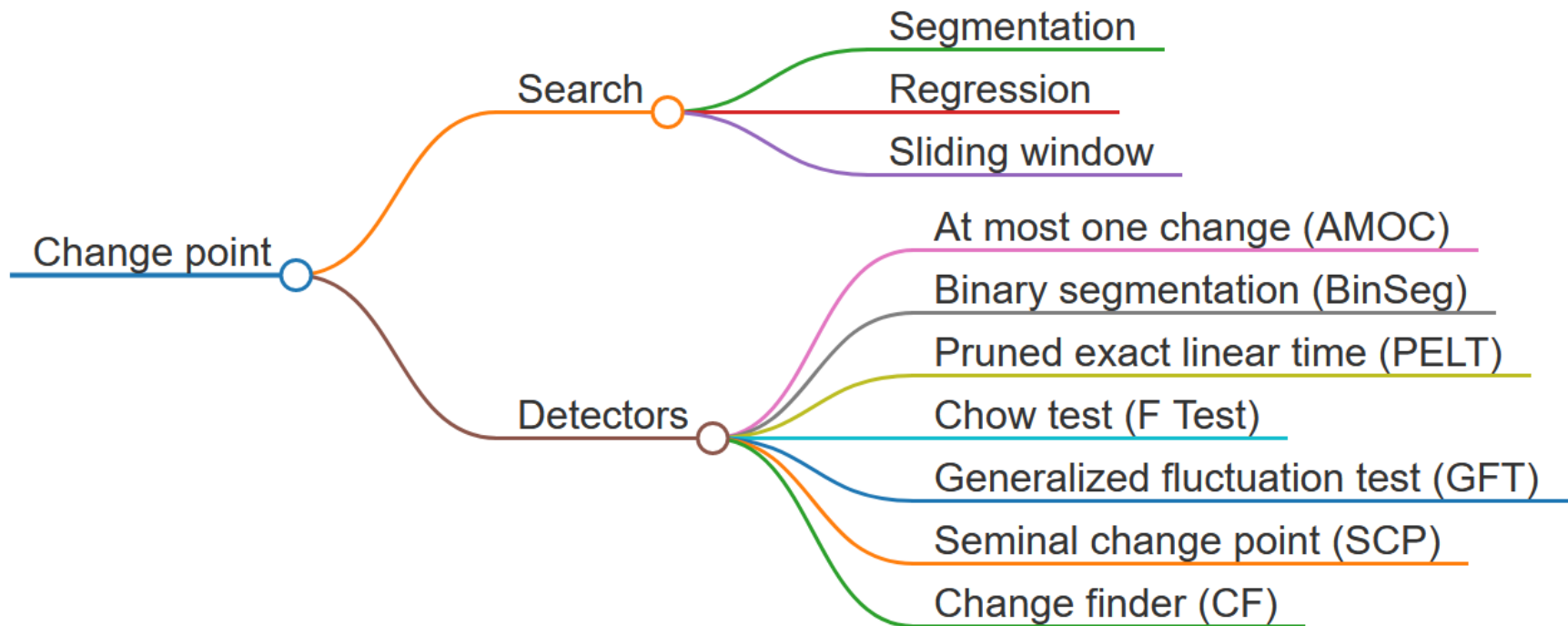
Exemplo de ponto de mudança

- Série sintética com quatro pontos de mudança



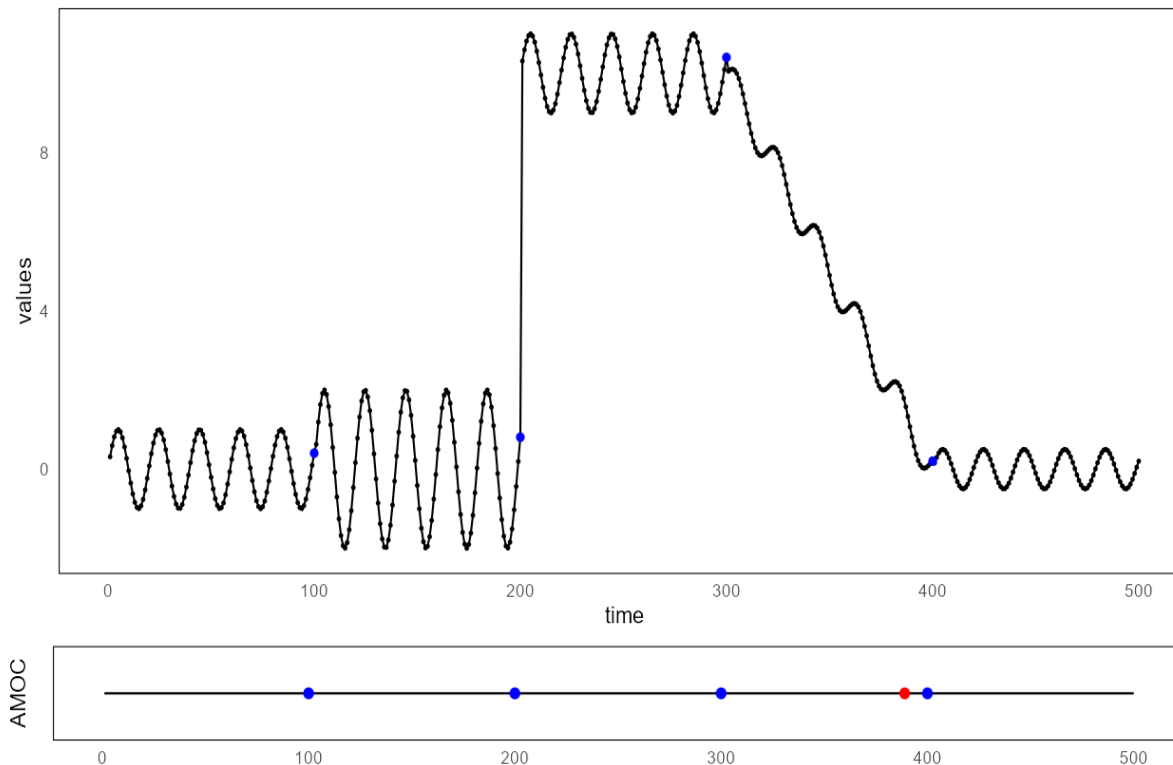
Taxonomia de detectores de mudança

- A detecção de pontos de mudança pode ser classificada de acordo com:
 - Número de mudanças: uma (AMOC) ou múltiplas (PELT, BinSeg)
 - Tipo de abordagem: segmentação, teste estatístico, modelagem preditiva
 - Eficiência: métodos exatos ou aproximado
- A escolha do detector depende da complexidade da série e do objetivo da análise



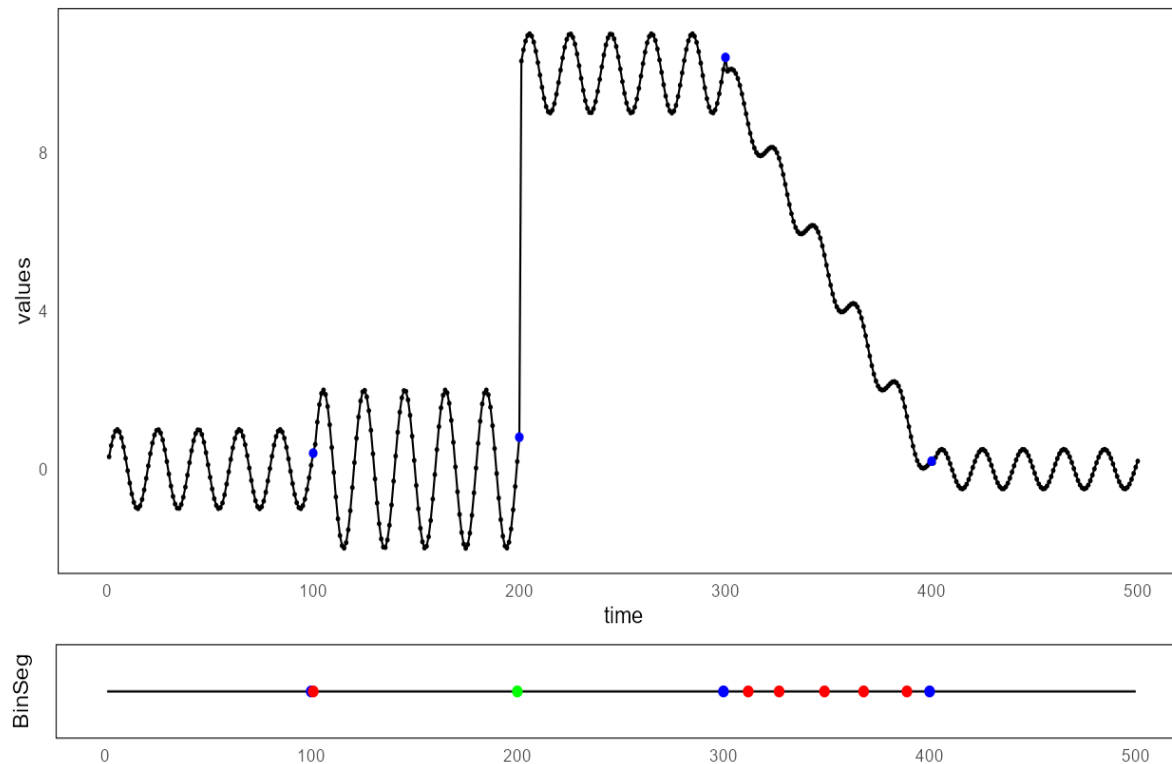
At most one change (AMOC)

- Modelo que assume no máximo uma mudança na série
- Detecta uma transição abrupta única, como uma mudança de nível
- Simples e eficiente para séries curtas com uma única ruptura



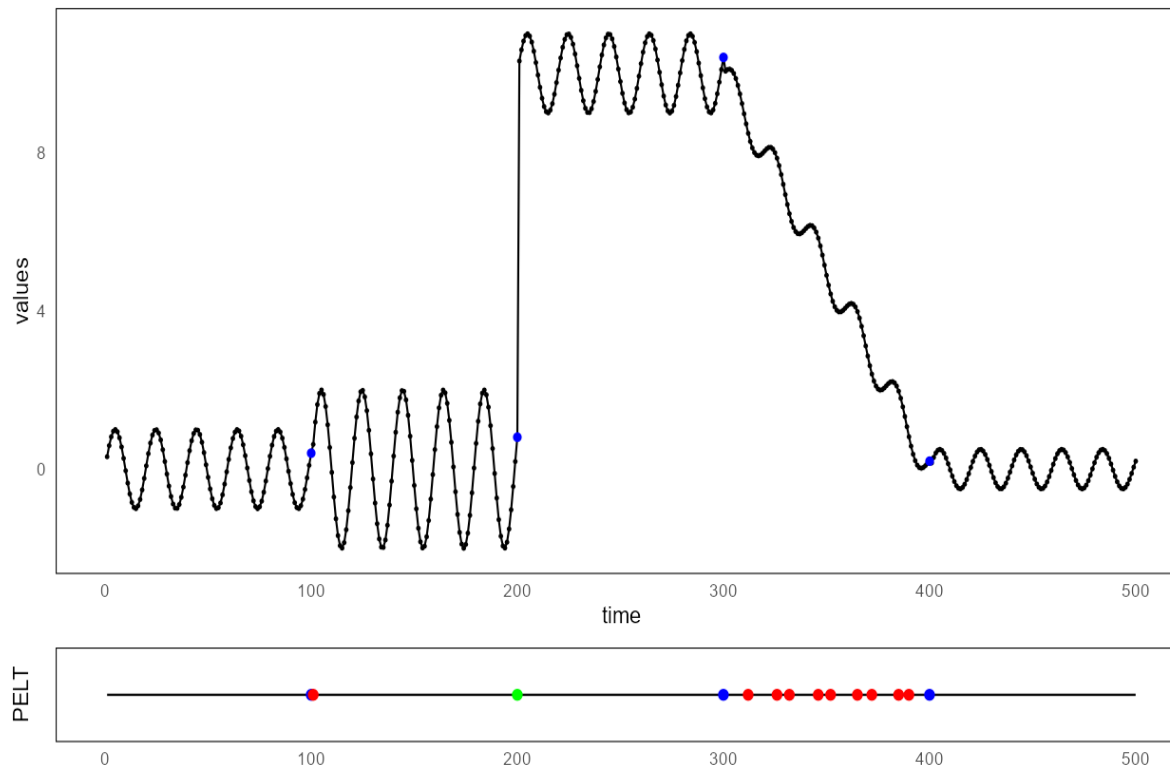
Binary Segmentation (BinSeg)

- Algoritmo recursivo para detectar múltiplos pontos de mudança
- Divide a série em segmentos, identificando o melhor ponto de ruptura em cada um
- Após cada divisão, aplica o mesmo processo nas subsequências
- Simples e eficiente, mas pode perder mudanças próximas



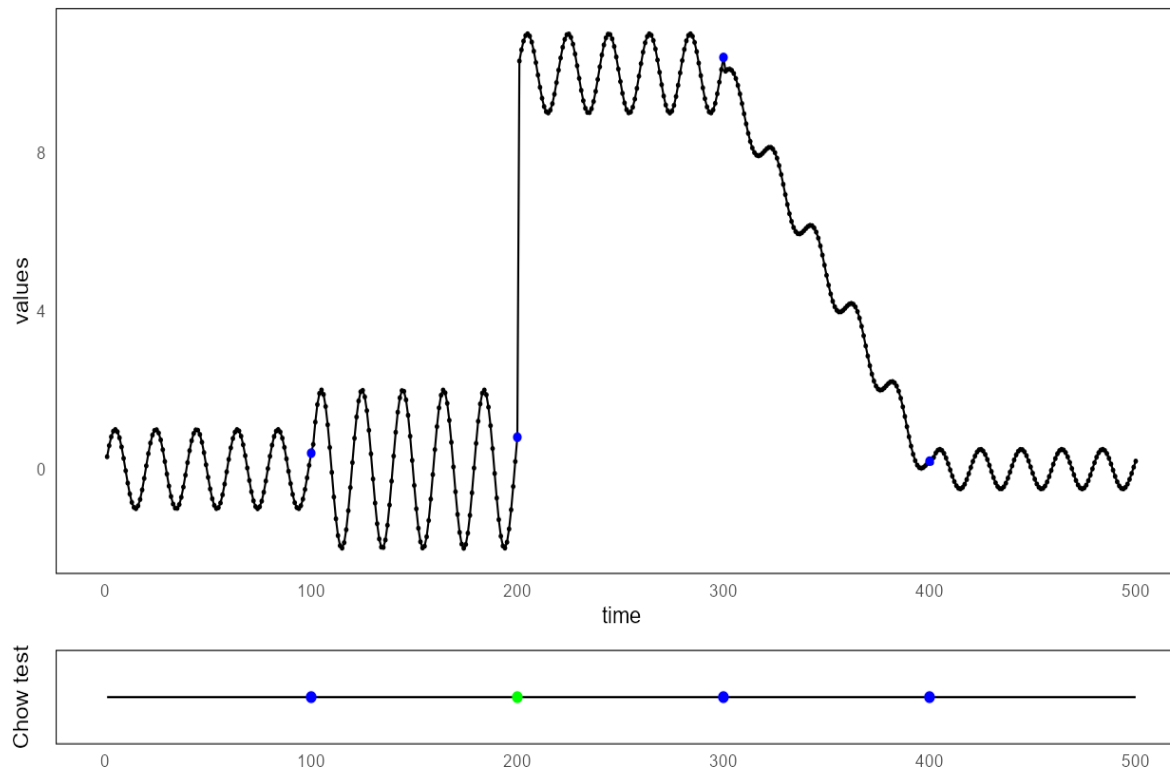
Pruned Exact Linear Time (PELT)

- Algoritmo eficiente para detectar múltiplos pontos de mudança.
- Usa poda para garantir complexidade linear no número de observações.
- Equilíbrio entre precisão e desempenho computacional.



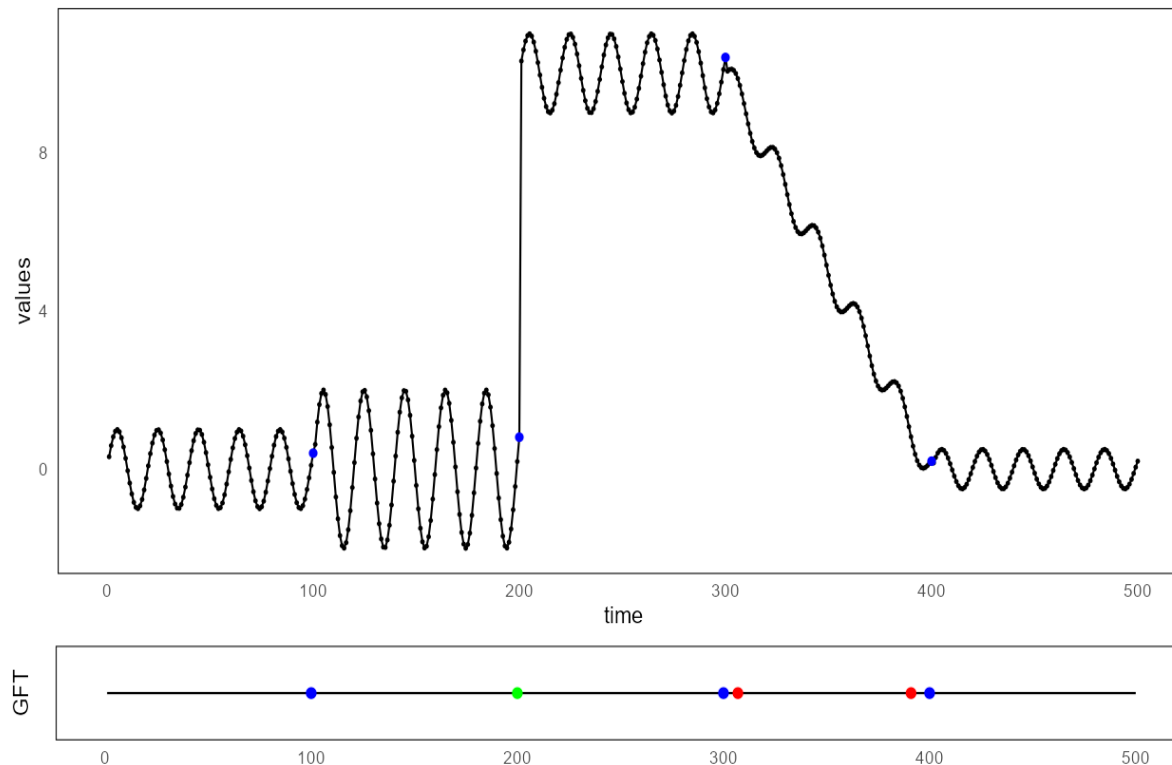
Chow test

- Teste estatístico para verificar se há quebra estrutural em modelos de regressão.
- Compara o ajuste antes e depois de um ponto candidato à mudança.
- Requer conhecimento prévio ou hipótese sobre o ponto de quebra.



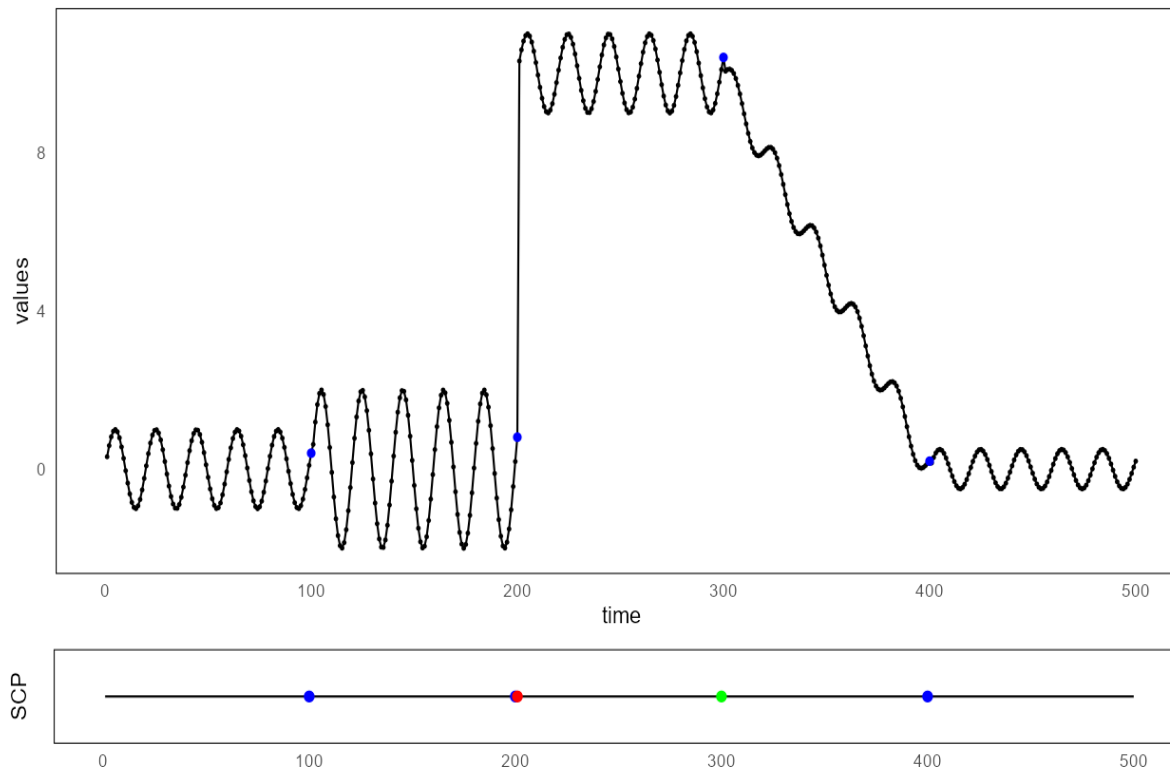
Generalized fluctuation test

- Detecta instabilidade nos parâmetros de modelos ao longo do tempo.
- Baseado em flutuações cumulativas dos resíduos ou estatísticas.
- Útil para análise visual de mudanças sutis em séries temporais.

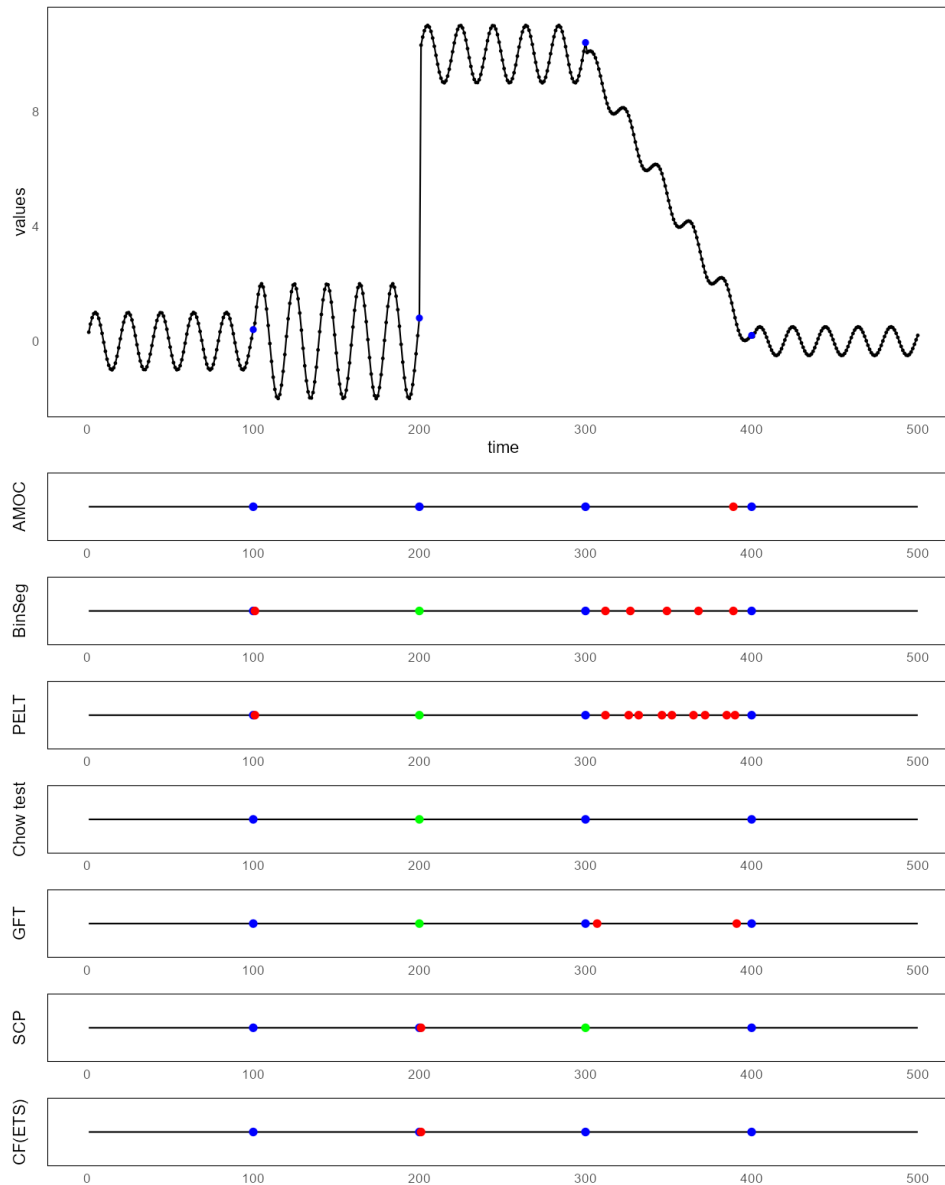


Seminal change point

- Refere-se a um ponto de ruptura fundamental na estrutura dos dados
- Muitas vezes marca o início de uma nova fase ou regime no processo
- Pode ser identificado como o ponto mais significativo entre vários candidatos



Visual comparativo de pontos de mudança



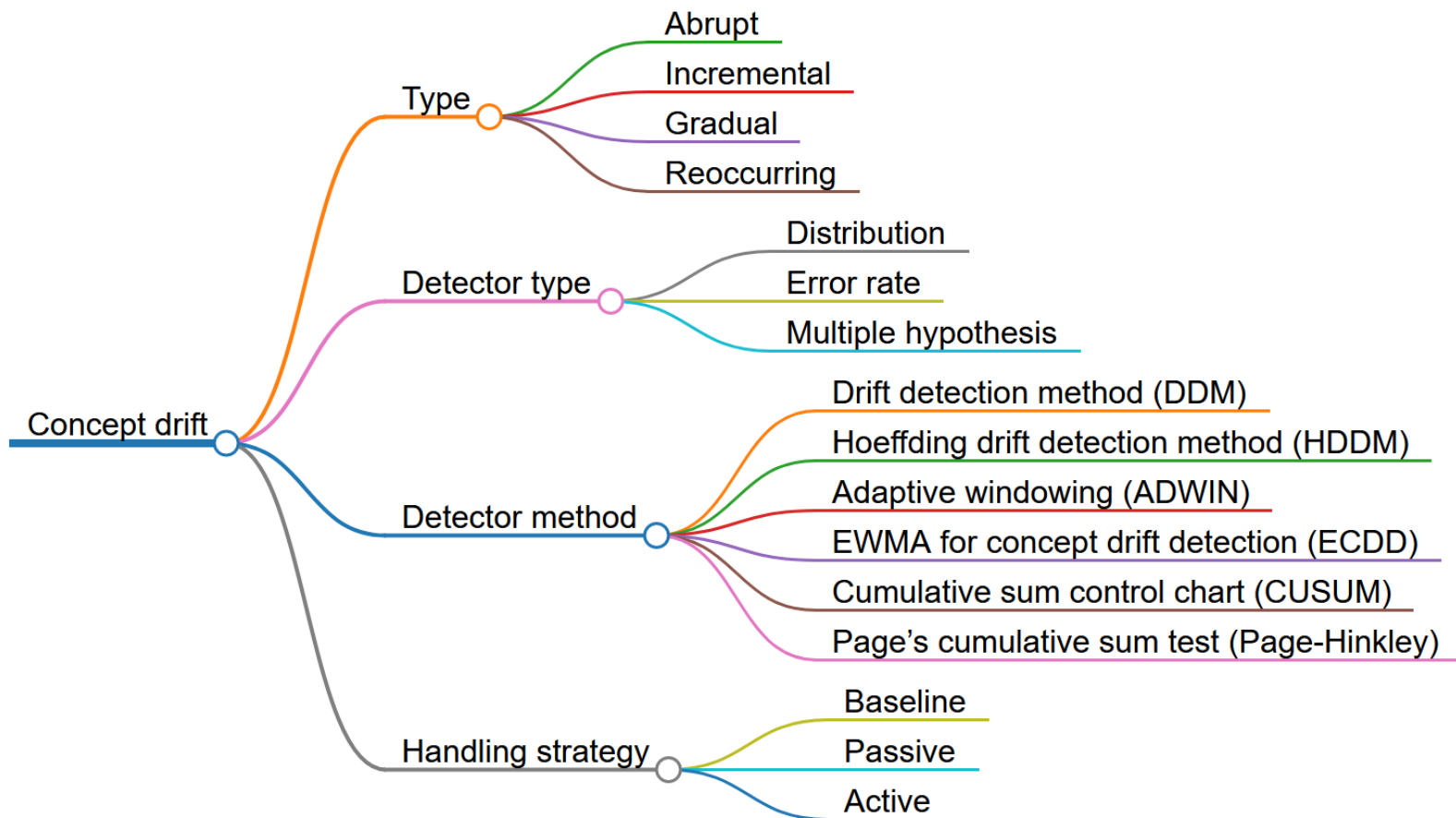
Acertos: verde
FP: vermelho
FN: azul

Desvio de conceito

- Ocorre quando a distribuição dos dados muda ao longo do tempo, invalidando modelos aprendidos previamente.
- Comum em fluxos contínuos de dados (streaming), especialmente em sistemas de decisão automática.
- Requer técnicas específicas para detecção e adaptação.
- Drift real (real drift): muda a relação entre atributos e rótulo, ou seja, afeta diretamente o modelo preditivo.
 - Exemplo: em um sistema de diagnóstico, o mesmo sintoma passa a indicar outra doença.
- Drift virtual (virtual drift): muda apenas a distribuição dos atributos, mas sem alterar a relação com o rótulo.
 - Exemplo: a distribuição de temperaturas muda, mas a classe “normal” ainda é a mesma para os mesmos valores.

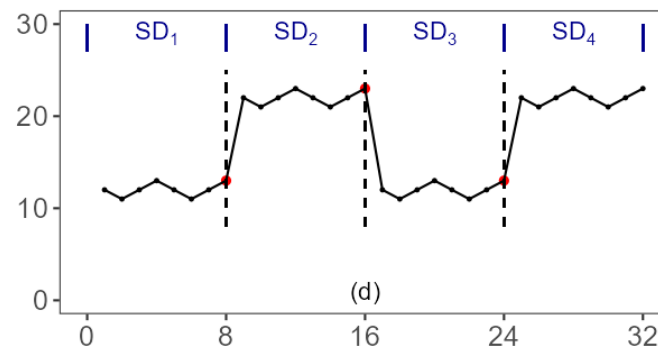
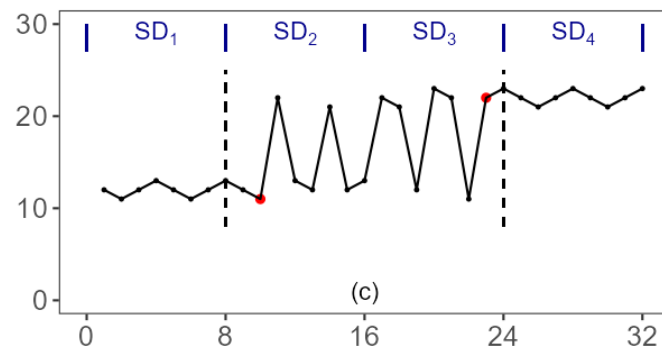
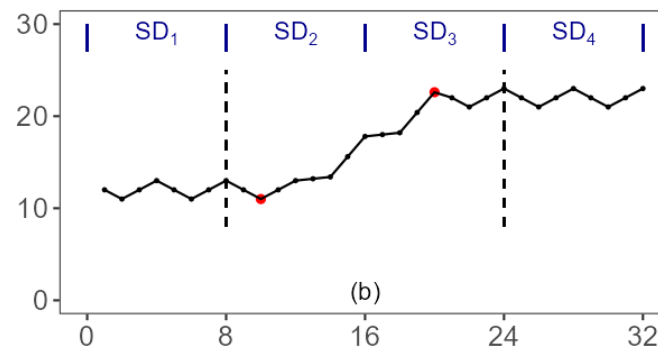
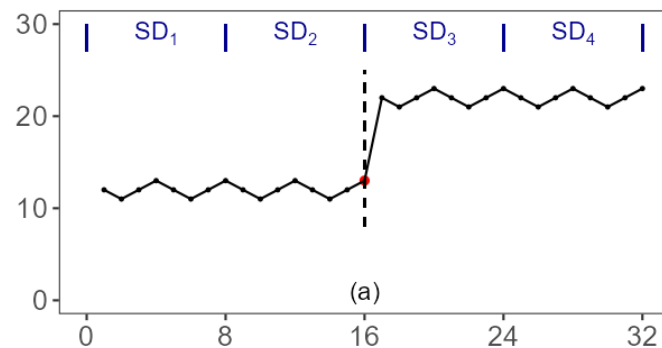
Taxonomia de desvio de conceito

- Tipos: abrupto, gradual, incremental, recorrente
- Tipos de detectores: análise de erro, comparação de janelas, mudanças na média
- Métodos de detecção: DDM, HDDDM, ADWIN, etc.
- Estratégias de tratamento: baseline, passiva, ativa



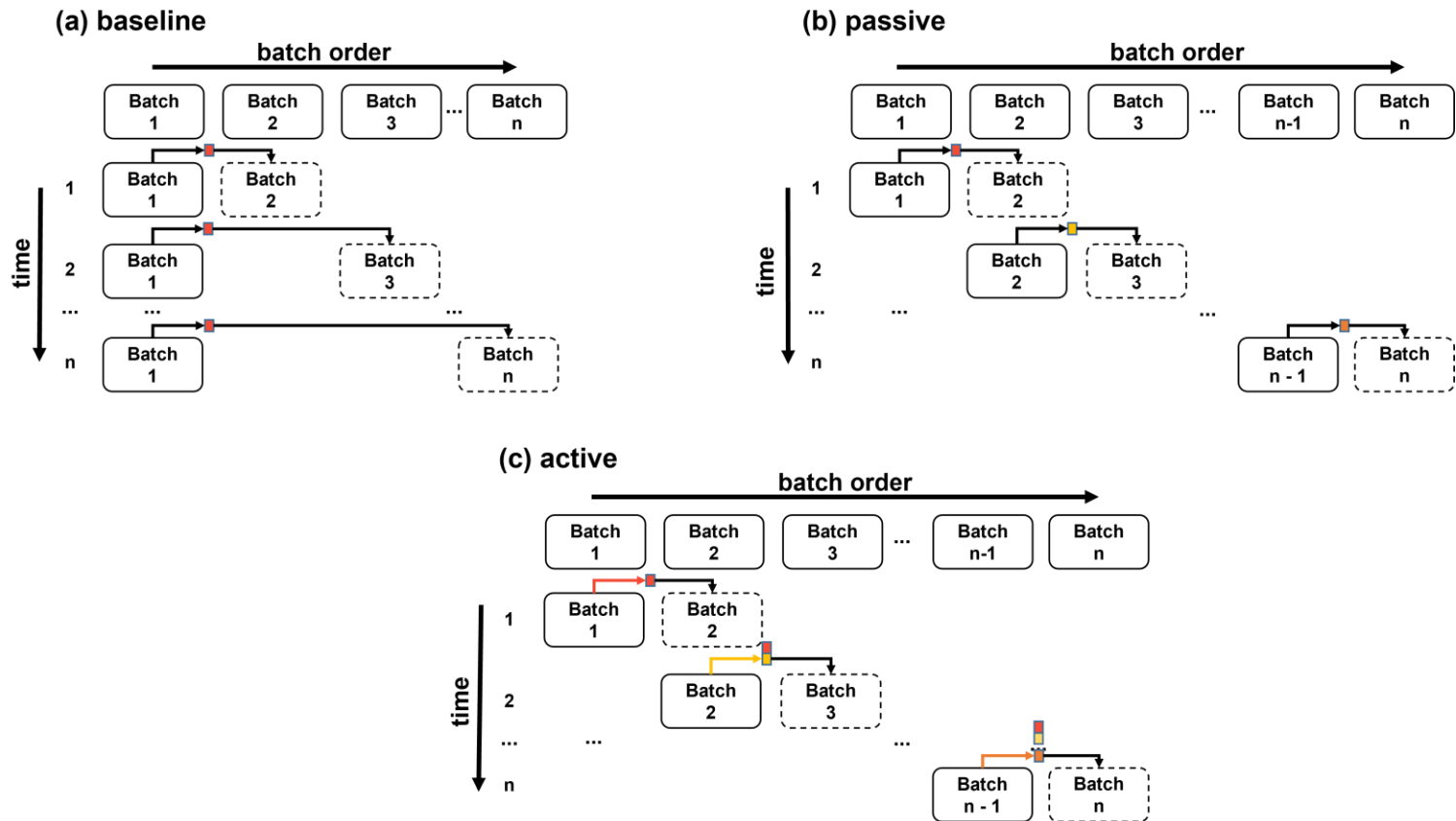
Tipos de desvio

- Abrupto: mudança repentina e clara entre conceitos
- Incremental: alteração lenta ao longo do tempo
- Gradual: coexistência temporária dos conceitos antigo e novo
- Recorrente: o conceito antigo reaparece após certo tempo



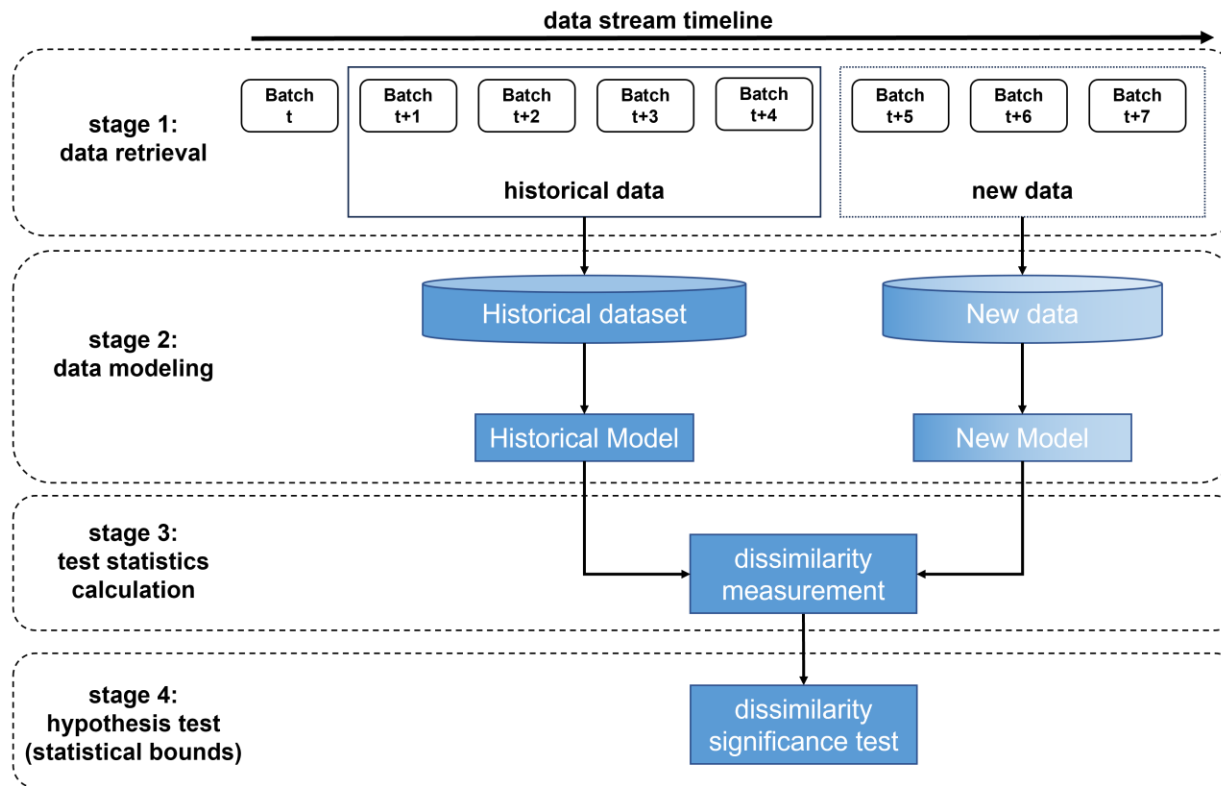
Estratégias para tratamento de desvio

- Baseline (sem ação):
 - O modelo continua operando normalmente, sem adaptação. Serve como referência para comparação, mas tende a perder desempenho com o tempo
- Resposta Passiva
 - O modelo se ajusta gradualmente ao novo conceito, usando estratégias como: aprendizado incremental, esquecimento de dados antigos, janelas deslizantes
- Resposta Ativa
 - A detecção de desvio aciona uma adaptação imediata, como: reset do modelo, retreinamento com novos dados e troca de parâmetros ou algoritmos



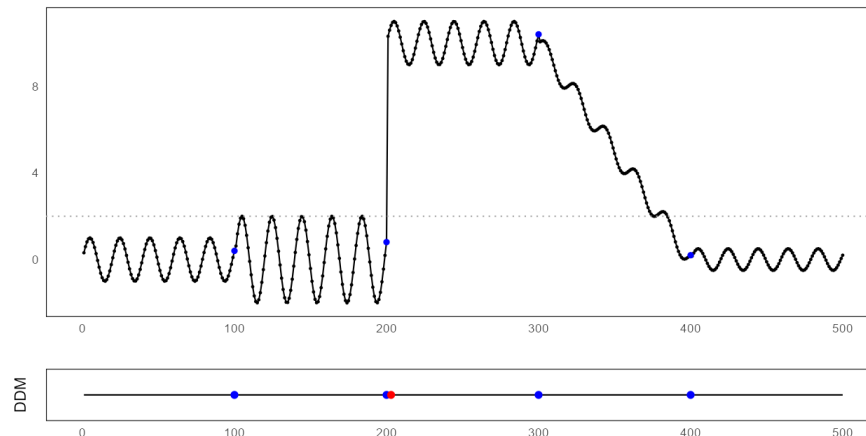
Detecção de desvio

- A detecção de desvio envolve:
 - Modelar o comportamento esperado da série
 - Comparar segmentos (janelas) ao longo do tempo
 - Aplicar testes estatísticos para detectar mudanças significativas
- Pode ser usada tanto para drift real (mudança no modelo) quanto virtual (mudança nos dados)



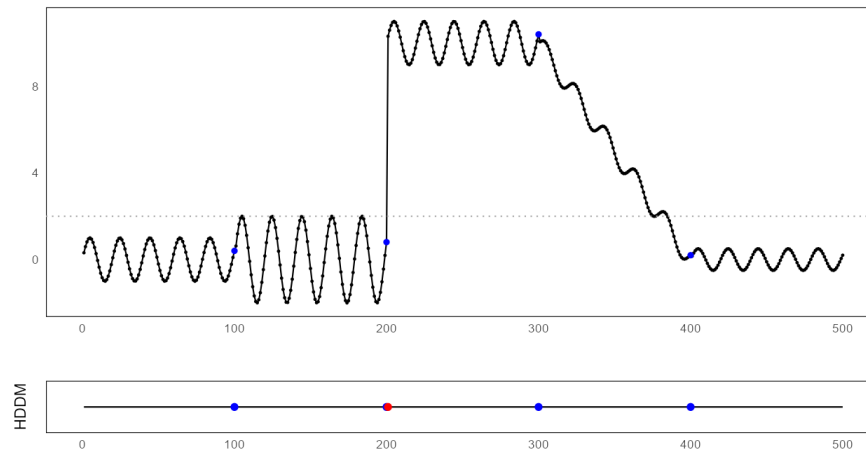
Drift Detection Method (DDM)

- Monitora a taxa de erro ao longo do tempo.
- Se a taxa excede limites estatísticos, um desvio é detectado.
- Simples e eficiente para detecção em fluxos supervisionados..
- Foco em drift real



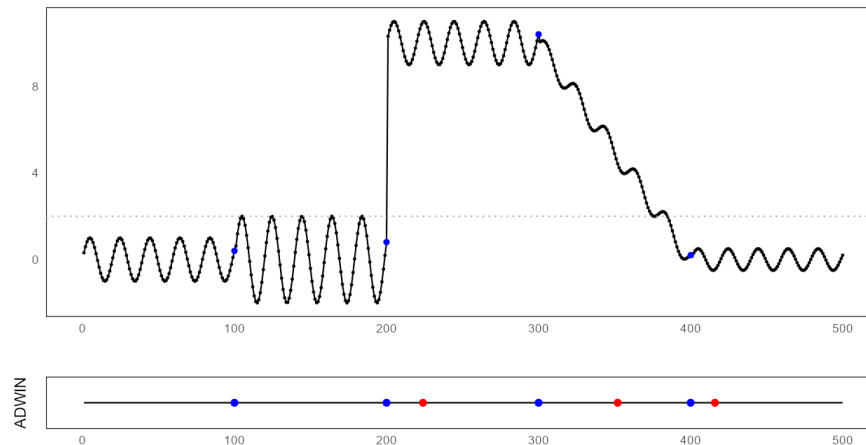
Hoeffding drift detection method (HDDM)

- Baseado no teorema de Hoeffding para detectar mudanças significativas.
- Sensível a desvios mesmo com pequenas janelas de observação.
- Adequado para aplicações em tempo real.
- Foco em drift real



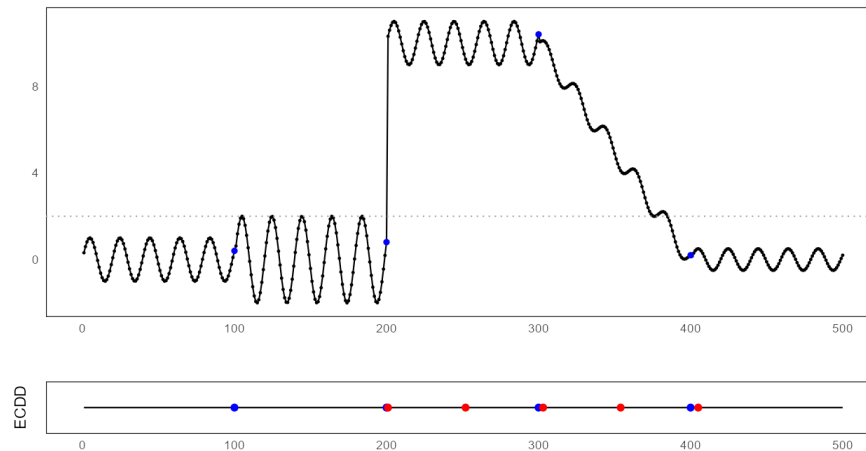
Adaptive windowing (ADWIN)

- Mantém uma janela adaptativa de dados.
- Detecta mudanças ao comparar estatísticas de diferentes partes da janela.
- Remove automaticamente dados antigos quando detecta mudança.
- Pode ser usado para drift real e virtual



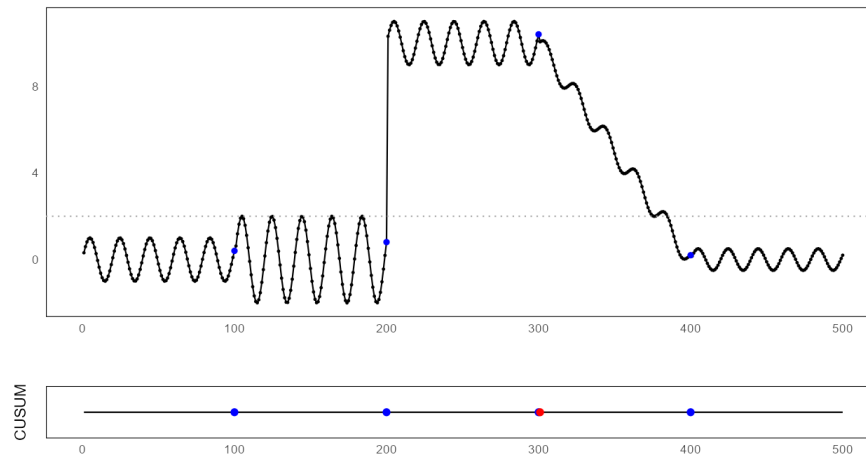
EWMA for concept drift detection (ECDD)

- Usa médias móveis exponencialmente ponderadas para rastrear desvios.
- Reage suavemente a pequenas mudanças e rapidamente a grandes rupturas.
- Bom equilíbrio entre sensibilidade e robustez.
- Foco em drift real



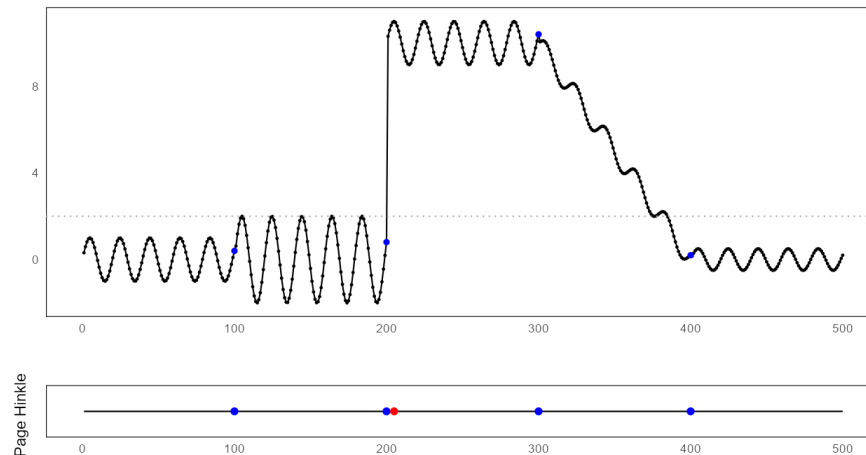
Cumulative sum control chart (CUSUM)

- Técnica de controle estatístico baseada na soma cumulativa dos desvios.
- Detecta mudanças sutis na média da série.
- Muito usada em controle de qualidade e monitoramento industrial.
- Foco em drift virtual, mas pode ser usado em real

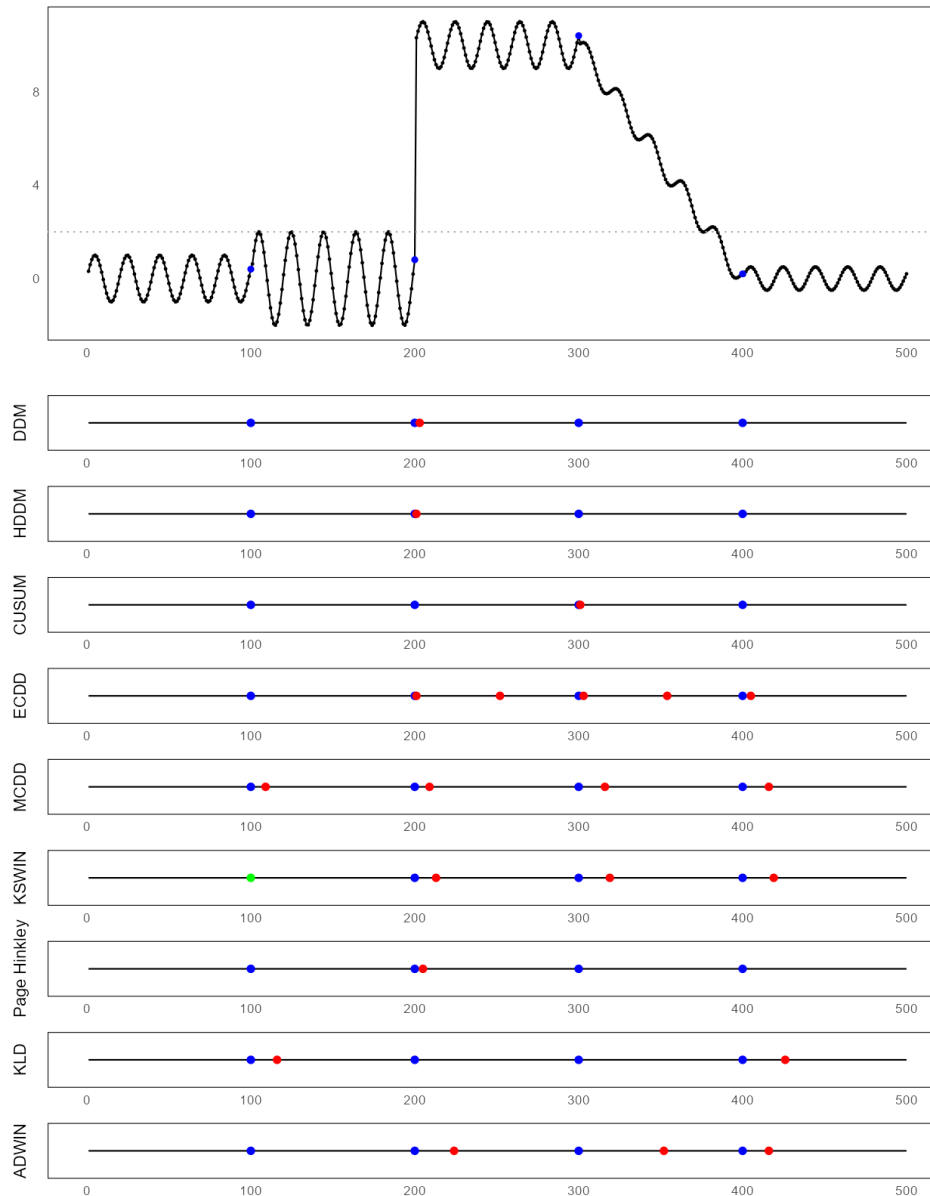


Page-Hinkley

- Algoritmo baseado em monitoramento de desvios acumulados da média esperada.
- Detecta mudanças sustentadas no comportamento da série.
- Eficiente para detecção online com baixo custo computacional.
- Foco em drift virtual, mas pode ser usado em real



Visual comparativo para desvios de conceito



Acertos: verde
FP: vermelho
FN: azul

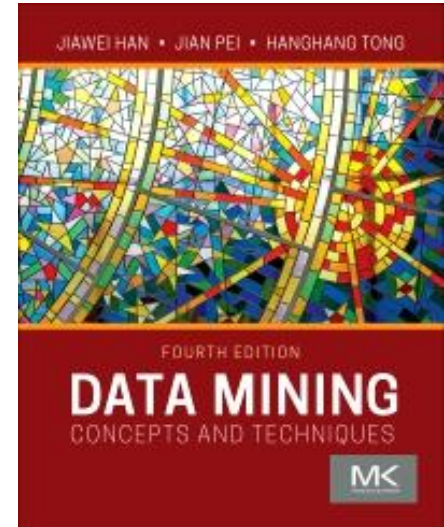
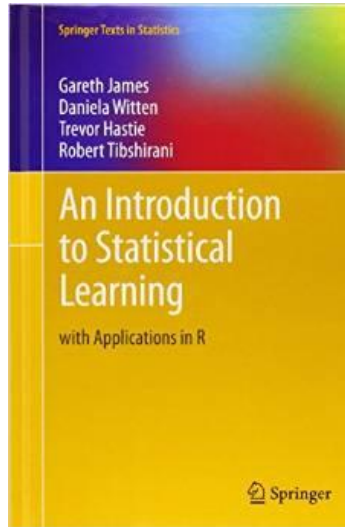
Resumo do Capítulo

- Pontos de mudança indicam transições importantes na série temporal.
- • Podem ocorrer na média, variância ou distribuição dos dados.
- • Detectores como AMOC, BinSeg e PELT são utilizados.
- • Visualizações ajudam a evidenciar essas mudanças na prática.

Exercício Harbinger

- Explore a página do Harbinger
 - <https://cefet-rj-dal.github.io/harbinger/>
- O Harbinger oferece suporte direto para experimentação e avaliação de métodos de detecção de ponto de mudança:
 - Implementa detectores clássicos como AMOC, BinSeg e PELT
 - Permite comparar os resultados com a série original, destacando:
 - Acertos (verde)
 - Falsos positivos (vermelho)
 - Falsos negativos (azul)
 - Facilita a validação visual e quantitativa com métricas automáticas
 - Suporta séries reais ou simuladas e diferentes funções de custo
 - Uma ferramenta prática para entender o comportamento dos detectores e avaliar sua sensibilidade
- Execute os detectores em séries sintéticas ou reais e observe como cada método reage a diferentes tipos de mudança

Referências



- [1] Ogasawara, E.; Salles, R.; Porto, F.; Pacitti, E. Event Detection in Time Series. 1. ed. Cham: Springer Nature Switzerland, 2025.
- [2] Cryer, J. D.; Chan, K.-S. Time Series Analysis: With Applications in R. Springer Science & Business Media, 2008.
- [3] Han, J.; Pei, J.; Tong, H. Data Mining: Concepts and Techniques. 4th edition ed. Cambridge, MA: Morgan Kaufmann, 2022.
- [4] James, G. M.; Witten, D.; Hastie, T.; Tibshirani, R. An Introduction to Statistical Learning: With Applications in R. [s.l.] Springer Nature, 2021.

