

Filtering in Preprocessing for Anomaly Detection in Time Series

Ricardo Fernandes de Araujo · Lais Baroni · Eduardo Bezerra · Gustavo Guedes · Esther Pacitti · Larissa Pinheiro · Diego Salles · Fabio Porto · Eduardo Ogasawara

Motivação e Lacuna de Pesquisa

A detecção de anomalias é essencial em sistemas industriais, financeiros, energéticos, de saúde e de redes. A filtragem é frequentemente aplicada para reduzir ruído antes da modelagem — mas seus efeitos sobre os padrões de anomalia raramente são avaliados de forma sistemática.

Detecção via ARIMA

ARIMA é amplamente utilizado para detecção de anomalias baseada em resíduos. Outliers interferem diretamente no ajuste e na estabilidade do modelo.

Risco da Filtragem

A filtragem altera mais do que apenas o ruído. Padrões de anomalia podem ser atenuados ou distorcidos nas etapas seguintes do pipeline de análise.

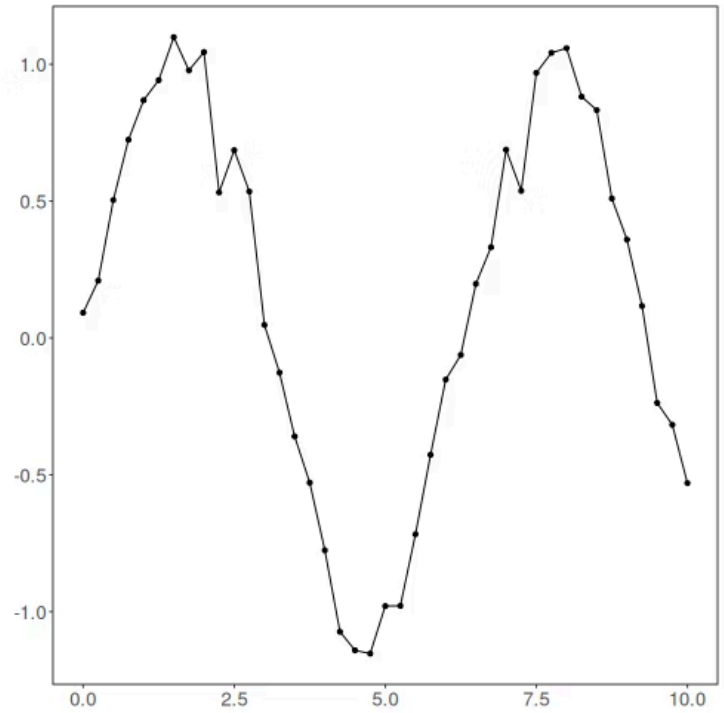
Lacuna Identificada

A literatura carece de evidências experimentais sistemáticas sobre quando a filtragem ajuda ou prejudica dentro de um pipeline ARIMA fixo com ground truth rotulado.

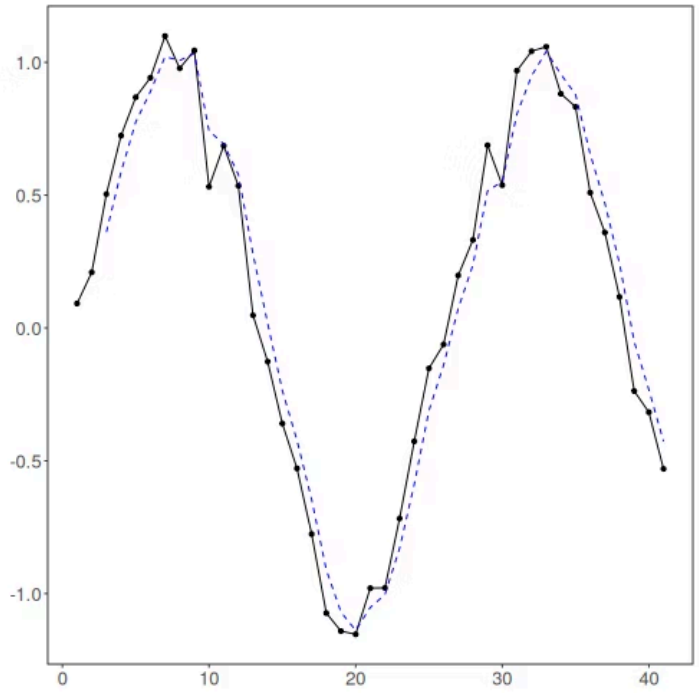
⚠ Este trabalho não propõe um novo detector — isola e avalia o pré-processamento como etapa de design de pipeline.

Exemplo: Efeito do Filtro EMA sobre Anomalias

Série original (com ruído e spikes)



Após filtro EMA (janela = 3)



⊗ O filtro EMA atenua o ruído, mas também reduz a amplitude dos picos de anomalia — o que pode prejudicar a detecção baseada em resíduos do ARIMA.

Questões de Pesquisa

Melhora o desempenho?

A filtragem melhora a detecção de anomalias baseada em ARIMA?

Quais são confiáveis?

Quais filtros apresentam resultados estáveis entre diferentes séries?

Quais têm potencial?

Quais filtros oferecem ganho com maior risco associado?

Regime temporal?

Como o efeito varia conforme o regime temporal da série?

O artigo avalia o pré-processamento, não um novo detector.

Metodologia

Para responder às perguntas de pesquisa, adotamos um experimento controlado que isola o efeito da filtragem mantendo fixos o detector e os benchmarks.

14 Filtros

Implementados via TSPredIT, agrupados por comportamento (médias, adaptativos, suavização, decomposição).

2 Benchmarks

GECCO Industrial Challenge 2018 e Yahoo! Webscope S5 — dados reais e sintéticos com ground truth rotulado.

Detector fixo

Pipeline Harbinger com arima_filter: ARIMA ajustado via auto.arima, detecção por resíduos com critério L1 (k=3).

Métricas de comparação

F1-Score por execução, com medidas de melhoria, degradação e potencial relativo ao baseline não filtrado.

Filtros de Pré-processamento Avaliados

Médias (Average)

- MA (Moving Average)
- EMA (Exponential Moving Average)

Adaptativos

- Kalman
- RFilter

Robusto

- Winsor

Suavização (Smoothing)

- HP (Hodrick–Prescott)
- LOWESS
- Smooth
- Spline

Decomposição

- EMD (Empirical Mode Decomposition)
- FFT (Fast Fourier Transform)
- REMD
- Seasonal_Adj
- Wavelet

Todos os filtros utilizam parâmetros padrão, sem ajuste por série ou domínio. Agrupados por comportamento para análise de risco versus potencial.

Conjuntos de Dados

GECCO Industrial Challenge 2018

Dados de sensores de tratamento de água (planta industrial). Fatia de 24 horas com 9 variáveis avaliadas individualmente. Ground truth derivado diretamente do campo EVENT. Resulta em 135 execuções (126 excluindo baseline).

Yahoo! Webscope S5

363 séries univariadas válidas (4 removidas por inconsistências) distribuídas em 4 subconjuntos (A1–A4) com dados reais e sintéticos sob diferentes níveis de ruído e mudanças de regime. Resulta em 5.445 execuções (5.082 excluindo baseline).

Dataset	Séries	Execuções (sem baseline)
GECCO Industrial Challenge 2018	9 variáveis	126
Yahoo! Webscope S5	363 séries	5.082
Total	—	**5.208**

Protocolo Experimental

Estrutura de cada execução

- Em cada execução `e = (s, f)`, o filtro *f* é aplicado à série *s*
- Um modelo ARIMA é ajustado à série filtrada usando `auto.arima()`
- Os resíduos são calculados comparando a série **original** com os valores ajustados in-sample — preservando os sinais de anomalia que o filtro poderia atenuar
- A detecção é feita no espaço de resíduos com o critério `arima_filter` do Harbinger (distância L1, k=3)
- Todas as execuções são comparadas com a linha de base não filtrada da mesma série

Fórmulas-chave

Definição de execução:

$$e = (s, f)$$

Variação absoluta de F1:

$$\Delta F1(e) = F1_{\text{filter}}(e) - F1_{\text{baseline}}(s)$$

Variação percentual de F1 (quando $F1_{\text{baseline}} > 0$):

$$\Delta F1_{\%}(e) = 100 \cdot \frac{\Delta F1(e)}{F1_{\text{baseline}}(s)}$$

Os resíduos são calculados contra a série original para preservar os sinais de anomalia.

Configuração Experimental

01

Aplicar filtro

Parâmetros padrão

02

Selecionar ARIMA

auto.arima automático

03

Calcular resíduos

Original − ajustado

04

Detectar anomalias

L1 com k=3

Pipeline: Harbinger arima_filter + TSPredIT

Para cada execução e = (s, f):

1. s_f ← apply_filter(f, s) # filtro com parâmetros padrão
2. model ← auto.arima(s_f) # seleção automática do modelo
3. ŝ ← fitted_values(model) # valores ajustados in-sample
4. r ← s - ŝ # resíduos contra série ORIGINAL
5. anomalias ← L1_threshold(r, k=3) # critério Harbinger padrão
6. ΔF1(e) ← F1(anomalias) - F1(baseline_s)

Todos os filtros usam parâmetros padrão — sem ajuste por série ou domínio. O detector ARIMA permanece fixo em todas as execuções, isolando o efeito do filtro.

Métricas de Avaliação

Métricas principais

Precisão, Revocação e **F1-Score** são calculados para cada execução. O F1 é a métrica principal de comparação entre filtro e linha de base.

Quando **F1_baseline = 0**, a variação percentual não é considerada — a execução é tratada como "sem variação" para evitar interpretações enganosas.

Frequência e Magnitude

- **Improvement:** proporção de execuções com $\Delta F1 > 0$
- **Degradation:** proporção de execuções com $\Delta F1 < 0$
- **Potential:** ganho percentual médio condicional às execuções que melhoraram (E^+)

Fórmulas (Eqs. 3–5)

$$\text{Improvement}(E) = \frac{1}{|E|} \sum_{e \in E} \mathbf{1}[\Delta F1(e) > 0]$$

$$\text{Degradation}(E) = \frac{1}{|E|} \sum_{e \in E} \mathbf{1}[\Delta F1(e) < 0]$$

$$\text{Potential}(E) = \frac{1}{|E^+|} \sum_{e \in E^+} \Delta F1_{\%}(e)$$

onde $E^+ = \{e \in E : \Delta F1(e) > 0\}$ — execuções com melhoria positiva.

Resultados: Visão Geral

A análise de **74,64%** das execuções revela que a aplicação do filtro não produz melhoria mensurável no F1-Score em relação à linha de base — resultado que contraria a hipótese de que a filtragem é uma etapa neutra ou universalmente benéfica.

25,36%	59,16%	15,48%
Execuções com melhoria de F1	Execuções com degradação de desempenho	Execuções sem variação mensurável

⊗ O resultado não sustenta o uso indiscriminado de filtragem. Filtrar não é uma etapa neutra.

Resultados: Impacto no F1-Score

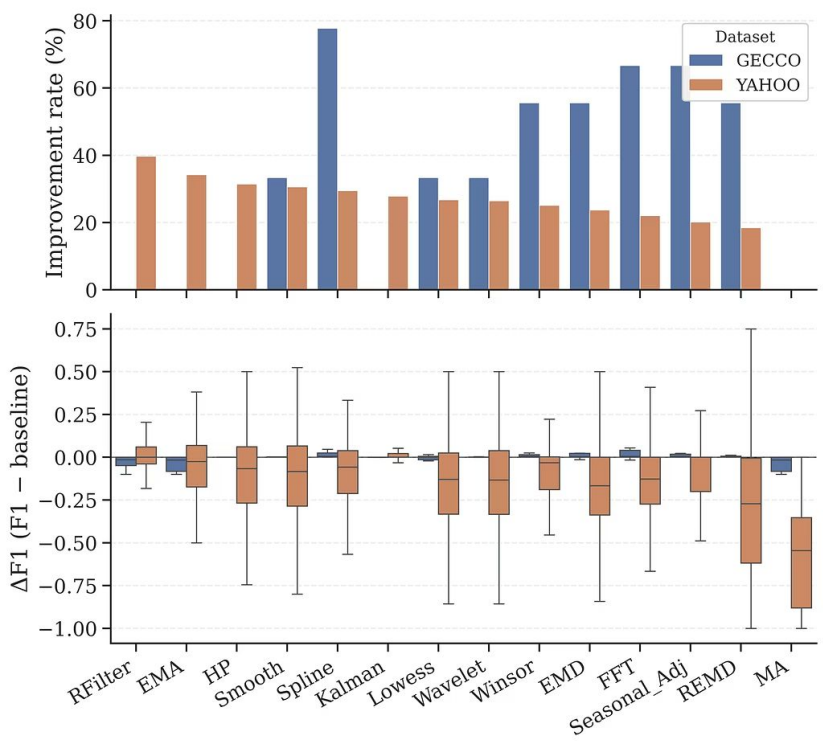
Tabela 1 — Impacto global da filtragem no F1-Score

Dataset	Execuções	▲ Melhoria	▼ Degradação	■ Sem variação
GECCO	126	26,98%	26,19%	46,83%
Yahoo!	5.082	25,32%	59,98%	14,70%
Total	**5.208**	**25,36% (1.321)**	**59,16% (3.081)**	**15,48% (806)**

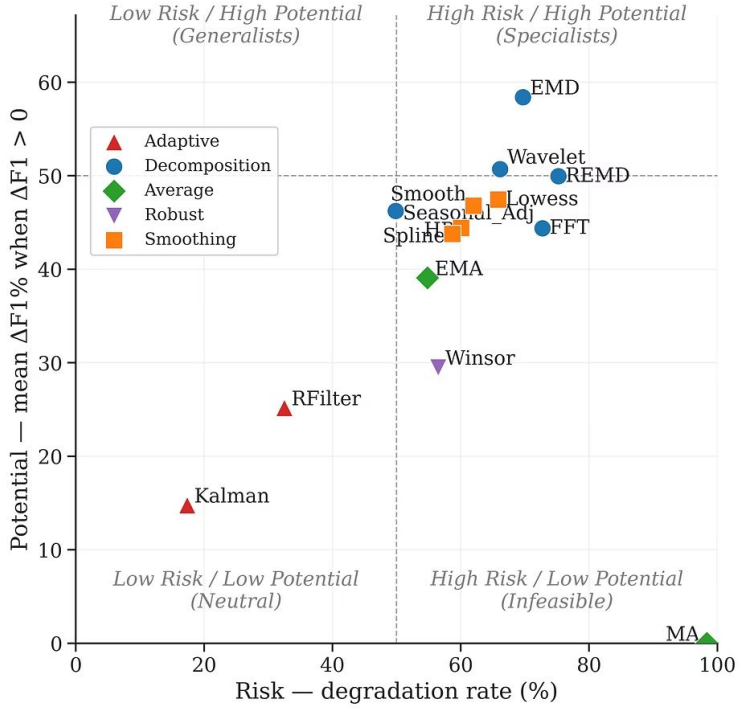
Em ambos os benchmarks, apenas 25,36% das execuções resultaram em melhoria do F1-Score.

⊗ Em ambos os benchmarks, a proporção de melhorias é similar (~25%), apesar dos perfis muito diferentes. O principal risco prático é a degradação, não a ausência de ganho.

Resultados: Confiabilidade e Potencial dos Filtros



(a) Improvement rate and ΔF1 distribution



(b) Risk vs Potential Matrix

O gráfico apresenta o posicionamento de todos os filtros avaliados em um quadrante Risco vs. Potencial, revelando o trade-off entre melhoria e degradação do F1-Score.

⊗ Nenhum filtro ocupa o quadrante ideal (Alto Potencial / Baixo Risco). O trade-off é universal no protocolo avaliado.

Resultados: Interpretação por Quadrante

Filtros Adaptativos (Generalistas)

Permanecem próximos a zero de variação. Kalman: degradação de apenas 17,36% no Yahoo — o mais confiável. Quadrante: Baixo Risco / Baixo Potencial (Neutro).

RFilter

Melhor equilíbrio entre os adaptativos: 39,67% de melhoria e 32,51% de degradação no Yahoo. Quadrante: Baixo Risco / Alto Potencial (Generalista).

Filtros de Decomposição (Especialistas)

EMD, Wavelet e REMD: potencial elevado, mas degradação acima de 65% no Yahoo. Quadrante: Alto Risco / Alto Potencial (Especialista).

Moving Average (MA)

Risco proibitivo: 98,35% de degradação no Yahoo. Quadrante: Alto Risco / Baixo Potencial (Inviável).

Resultados: Filtros de Alto Risco

EMD e REMD

Podem produzir grandes ganhos, mas degradam com frequência quando as premissas não correspondem à série. Superam **65% de degradação** no Yahoo.

Wavelet

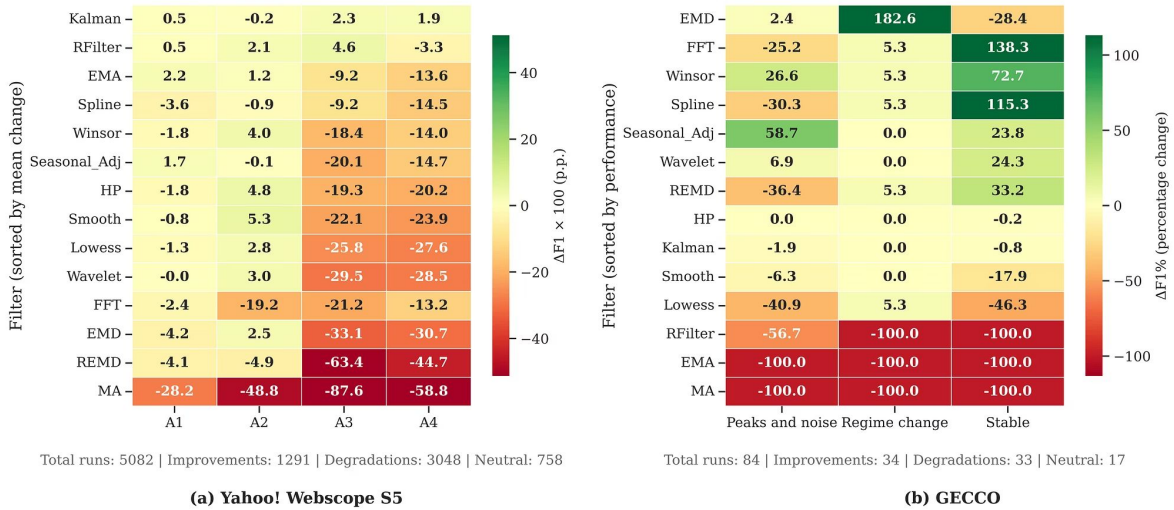
Filtro de decomposição especializado com potencial elevado, mas igualmente sujeito a alta degradação em séries incompatíveis com suas premissas.

Moving Average

Apresenta risco proibitivo neste protocolo — não recomendado para uso geral em pipelines de detecção baseados em ARIMA.

Filtros de decomposição não são ruins — são especializados. Exigem correspondência com o regime da série.

Resultados: Influência do Regime Temporal



(a) Yahoo! Webscope S5 **(b) GECCO**

Yahoo A1 (Dados reais, picos e sazonalidade variada)

Mais casos em que a filtragem ajuda. Vários filtros produzem resultados positivos neste subconjunto.

Yahoo A2 (Sintético, baixo ruído, picos sazonais)

Também favorável à filtragem. Junto com A1, concentra a maioria das melhorias observadas.

Yahoo A3 (Sintético, alto ruído, picos sazonais)

Dominado por ruído — a maioria dos filtros tem desempenho ruim. Filtros adaptativos são os poucos que ainda se comportam razoavelmente.

Yahoo A4 (Mudanças de regime)

Kalman se destaca como uma das opções mais consistentes neste subgrupo de mudanças de regime.

GECCO (Industrial, 9 variáveis)

Resultados menos conclusivos pelo menor número de séries. EMD aparece com maiores ganhos em variáveis de mudança de regime. RFilter tem desempenho ruim em contextos ruidosos e estáveis.

O efeito da filtragem depende fortemente do regime temporal — não há padrão uniforme entre os casos.

Discussão: Implicações para o Design de Pipelines

Filtragem é decisão de design

Não é uma etapa neutra de pré-processamento — integra o pipeline com risco mensurável.

Prefira filtros adaptativos

Kalman e RFilter são opções mais estáveis para uso geral em produção.

Filtros especializados exigem cautela

Decomposição oferece ganhos maiores, mas requer tolerância ao risco de degradação.

Valide com rótulos antes do deploy

A validação contra rótulos reais é essencial antes de qualquer implantação em produção.

Guia Prático: Quando (e Como) Filtrar?

Ruído leve (baixa magnitude)

Filtros suaves com janela pequena (EMA, MA) são relativamente seguros. Mesmo assim, valide o impacto no F1 antes de adotar.

Ruído intenso ou estruturado

Risco de distorção das anomalias aumenta com qualquer filtro. Filtros de Kalman são os mais conservadores — degradação mínima observada (17,36% no Yahoo).

Série com padrões complexos (não-estacionária)

Filtros de decomposição (EMD, Wavelet) podem trazer ganhos expressivos, mas exigem correspondência com o regime da série. Use apenas com validação rigorosa.

Regra geral

Nenhum filtro é neutro. Sempre avalie o filtro dentro do pipeline completo, com ground truth rotulado, antes de qualquer decisão de deploy.

⊗ Os resultados deste trabalho sugerem que a omissão do filtro (baseline) é frequentemente a escolha mais segura na ausência de validação específica.

Conclusão

Filtragem raramente melhora o F1

Em ambos os benchmarks, a maioria das aplicações de filtros não melhora o F1.

Nenhum filtro universal identificado

Nenhum filtro se mostrou o melhor em todos os cenários avaliados.

Filtros adaptativos apresentam maior estabilidade

Kalman e RFilter oferecem maior estabilidade com menor risco de degradação.

Decomposição: maior potencial, maior risco

Filtros de decomposição têm maior potencial, mas também maior frequência de degradação.

As conclusões são restritas ao protocolo baseado em ARIMA avaliado — evidência controlada, não prescrição universal.



CEFET/RJ

Agradecimentos

FAPERJ

CAPES

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

Ricardo Fernandes de Araujo · Lais Baroni · Eduardo Bezerra · Gustavo Guedes · Esther Pacitti · Larissa Pinheiro · Diego Salles · Fabio Porto · Eduardo Ogasawara

https://eic.cefet-rj.br/dal