

# Rumo a um novo filtro de anomalias baseadas em distribuições de probabilidade

Thiago Marques<sup>1</sup> Arthur Vaz<sup>2</sup>

<sup>1</sup>Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ

thiago.marques@aluno.cefet-rj.br

arthur.vaz@aluno.cefet-rj.br

**Resumo.** Há muitos métodos de detecção de anomalias ~~na literatura de detecção de anomalias~~ em séries temporais, ~~os mesmos~~ podem se subdividir em supervisionados e não supervisionados.

No que tange aos métodos não-supervisionados, existem diversos algoritmos disponíveis, alguns exemplos incluem o DBSCAN, K-Means e Autoencoder.

Esses algoritmos podem detectar anomalias, no entanto, há algumas limitações, uma delas é que só podem detectar a sequência anômala, não as observações individuais dentro da sequência que são anômalas.

O presente trabalho visa propor um novo filtro de anomalias baseado em distribuições de probabilidade, se utilizando de testes de aderência ~~qui-quadrado~~ para detecção da melhor distribuição que se adere aos dados das janelas deslizantes e posterior estimação das probabilidades para filtragem das observações anômalas.

O novo filtro **será** aplicado em series sintéticas que tem três comportamentos distintos, quais sejam: Estacionárias com sequências anômalas, estacionárias com observações anômalas e não-estacionárias com observações anômalas, comparando os resultados das sequências anômalas encontradas utilizando os métodos não supervisionados: Dbscan, K-means e autoencoder.

## 1. Introdução

Há muitos métodos de detecção de anomalias ~~na literatura de detecção de anomalias~~ em séries temporais, ~~os mesmos~~ podem se subdividir em supervisionados e não supervisionados, os primeiros quando existe um rótulo já definido previamente da variável alvo, já no segundo quando não há rótulos e a exploração se dá no contexto da distância entre variáveis, agrupamento das mesmas, ou até mesmo na redução de dimensionalidade.

Apesar dos métodos não supervisionados serem mais simples que os métodos supervisionados, podem ser especialmente úteis quando o contexto é detecção de sequências anômalas, independente da área que esta anomalia se apresenta.

há aplicações no contexto aeronáutico segundo [Corrado et al., 2021] que por meio da matriz de densidade do espaço aéreo para cada intervalo, aplicou o procedimento DBSCAN recursivo, classificando os estados operacionais como nominal, transitório ou anômalo.

Assim como no setor industrial destacado por [Princz et al., 2024] que trouxe um estudo em dados de sensores e desafios em ambientes de manufatura reais, trazendo sugestões práticas para implementação na indústria para melhorar a eficiência e confiabilidade do sistema.

Outra aplicação na indústria de gás segundo [Zhou et al., 2023] que propõe o D2AE, um autoencoder aprimorado por destilação de dados, para detectar consumo anômalo de gás em estabelecimentos comerciais (restaurantes, cantinas etc.). O D2AE considera diversas causas, como medidores defeituosos, vazamentos e roubos.

Dessa forma, este trabalho vai trazer um novo filtro de anomalias baseadas em distribuições de probabilidade, com o objetivo de potencializar ainda mais os métodos não supervisionados que não somente detectam sequências anômalas, em detrimento a observações anômalas, como é o caso do DBSCAN, K-means e autoencoder.

Será realizada uma fundamentação teórica do método na seção 2), falaremos sobre os trabalhos relacionados na seção 3), falaremos sobre o método desenvolvido na seção 4), Mostraremos os resultados da aplicação do mesmo nos algoritmos mencionados na seção 5) e traremos a conclusão dos experimentos, limitações do método na conclusão e possibilidades futuras na seção 6)

## 2. Fundamentação Teórica

Para entendermos como filtrar probabilidades em séries temporais, precisamos entender o que são séries temporais, como identificar distribuições de probabilidade por meio de testes de aderência e os métodos não supervisionados que foram escolhidos neste artigo para serem desenvolvidos.

### SÉRIES TEMPORAIS

Uma série temporal é qualquer conjunto de observações ordenadas no tempo.[Morettin and Toloi, 2022]

Há basicamente dois enfoques usados em análises de séries temporais. Em ambos o objetivo é construir modelos para as séries, com propósitos determinados. No primeiro enfoque, a análise é feita no domínio temporal e os modelos propostos são modelos paramétricos (Número finito de parâmetros). No segundo, a análise é conduzida no domínio de frequências e os modelos são não paramétricos.[Morettin and Toloi, 2022]

~~DBSCAN Segundo [Ester et al., 1996] é uma técnica baseada em densidade. Essa técnica identifica grupos (*clusters*) de dados procurando áreas separadas por regiões de alta e baixa densidade. três pontos de classificação são utilizados nessa abordagem, quais sejam: Pontos centrais Possuem pelo menos *m* outros dados dentro da vizinhança. Pontos de fronteira Estão dentro da vizinhança de um ponto central, mas não possuem *m* dados ao seu redor. Pontos *outliers* (isolados) Não estão dentro da vizinhança de nenhum *cluster*.~~

~~O algoritmo DBSCAN seleciona aleatoriamente um dado e o classifica como ponto central, de fronteira ou *outlier*. Este processo se repete para todos os dados até que todos estejam agrupados ou classificados como outliers.~~

~~Uma vantagem desse tipo de técnica é a capacidade de lidar com dados ruidosos, especialmente em regiões esparsas do espaço de dados.~~

~~K MEANS Segundo [Ferreira et al., 2022] Além de outras técnicas, métodos de agrupamento como DBSCAN e K-means são usados para segmentar e prever trajetórias~~

### ~~AUTOENCODER~~

~~É usado para reconstruir sinais de entrada com a ajuda de métodos do kernel. O processo de treinamento de um autoencoder não requer rótulos de dados, por isso é considerado um modelo e método de aprendizagem não supervisionado [Pei et al., 2018]~~

## TESTES DE ADERÊNCIA

São métodos matemático-estatísticos que avaliam se uma série teórica de dados (distribuição de probabilidade) descreve o comportamento de uma série real de dados. [Souza et al., 2013]

### ~~Teste do Qui-Quadrado~~

~~É um teste de aderência clássico, sendo, talvez, o mais utilizado na Estatística como testes de aderência.~~

~~Sua estatística de teste é dada por:~~

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

~~onde:~~

~~$\chi^2$  é a Estatística do teste~~

~~$k$  é o número de categorias ou células na tabela de contingência.~~

~~$O_i$  o valor da frequência observada na categoria  $i$ .~~

~~$E_i$  o valor da frequência esperada na categoria  $i$  sob a hipótese nula do teste.~~

### Teste de Kolmogorov Smirnov

A metodologia do teste baseia-se na diferença absoluta máxima (D) entre os valores obtidos de uma distribuição empírica de tamanho N e os valores obtidos de uma distribuição teórica específica, segundo [Souza et al., 2013].

Sua estatística de teste é dada por:

$$D = \max |F(x) - S(x)|$$

D é a estatística de teste. F(x) é a distribuição acumulada da distribuição empírica. S(x) é a distribuição acumulada da distribuição teórica sob a hipótese nula do teste.

## 3. Trabalhos relacionados

Para se entender mais a fundo sobre os métodos que serão utilizados e identificar lacunas na literatura para posterior contribuição, iremos realizar uma revisão de artigos relacionados a métodos não supervisionados de detecção de anomalias em séries temporais.

[Corrado et al., 2021] viram a necessidade de modernizar o tráfego aéreo aumentando a capacidade, eficiência e sem deixar de lado a segurança envolvida no processo. Afirmam que análises anteriores não consideravam estados operacionais de transição ou anomalias. No estudo os autores propõem um método que agrega dados de voos de todas as aeronaves em determinado terminal aéreo para gerar uma representação do estado operacional do espaço aéreo. Em seguida, foi utilizado o método DBSCAN recursivo para classificar o estado operacional como nominal, transitório ou anômalo, identificando também padrões nominais distintos.

[Princz et al., 2024] analisaram a detecção de anomalias em dados binários de série temporal na manufatura inteligente. A otimização da produção é crucial para garantir alta qualidade e produtividade. Afirmam que a detecção e prevenção precoces de falhas minimizam paradas, reduzem custos e garantem o desempenho ideal. avaliaram diferentes modelos de aprendizado de máquina para detectar anomalias nesses dados. O processo inclui pré-processamento, normalização e avaliação do desempenho dos modelos. Precisão, taxa de detecção e pontuação F1 são usadas como medidas. Também é analisado o tempo de execução de cada modelo. Os modelos Isolation Forest, Local Outlier Factor, DBSCAN e k-Means apresentaram alta precisão e detecção. Autoencoder e Variational Autoencoder tiveram alta precisão, mas menor detecção. O modelo Support Vector Machine alcançou desempenho equilibrado. O AutoML mostrou ótimo reconhecimento, mas não funciona em tempo real.

[Zhou et al., 2023] propõe um novo método para detectar consumo anormal de gás natural, chamado D2AE (Autoencoder Aprimorado por Destilação de Dados). O D2AE tem base no autoencoder e foi utilizado para evitar problemas com vazamentos e roubos, foi pensado para sanar a dificuldade detecção de anomalias com dados rotulados.

[Magdalene and Zoraida, 2022] analisaram a previsão de Consumo de Energia em Casas Inteligentes usando IWKMCA, um novo método proposto para prever o consumo de energia em casas inteligentes, considerando variações sazonais. O IWKMCA melhora o algoritmo K-means para criar grupos (clusters) mais precisos. Atribui pesos a dados dentro de cada grupo, priorizando aqueles mais próximos do centro do grupo (centróide). Utiliza o modelo ARIMA para prever o consumo com base nesses centróides. O IWKMCA apresentou maior precisão do que o ARIMA tradicional e o K-means Clustering ARIMA e obteve menores erros de previsão (RMSE, MAPE, AIC, AICC) e maior valor de log-verossimilhança, indicando melhor desempenho que o ARIMA padrão.

[Ferreira et al., 2022] afirma que o tráfego marítimo gera dados valiosos através do Sistema de Identificação Automática (AIS), que registra informações sobre a trajetória das embarcações. Analisar esses dados permite detectar comportamentos incomuns e padrões de tráfego suspeitos. Modelos como ARIMA (Média Móvel Integrada Autorregressiva) e Ornstein-Uhlenbeck são usados para prever o comportamento das embarcações. Em seguida, técnicas de agrupamento como DBSCAN e DBSCAN hierárquico identificam padrões de comportamento incomuns, garantindo a segurança da navegação. <sup>1</sup>. <sup>2</sup>. <sup>3</sup>. <sup>4</sup>. <sup>5</sup>.

#### 4. Método

Bases sintéticas Foram Desenvolvidas funções para gerar séries temporais sintéticas em formatos específicos, a fim de compreender a eficácia da detecção de janelas anômalas e, consequentemente, o ponto anômalo dentro dessas janelas. Em cima das bases produzidas, foram aplicadas em três modelos não supervisionados, quais sejam: K-means,

---

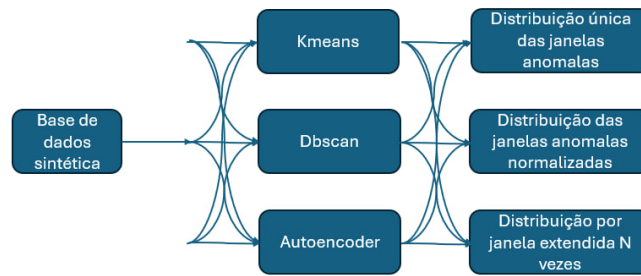
<sup>1</sup>Characterizing Terminal Airspace Operational States and Detecting Airspace-Level Anomalies

<sup>2</sup>Anomaly Detection in Binary Time Series Data: An unsupervised Machine Learning Approach for Condition Monitoring

<sup>3</sup>A data distillation enhanced autoencoder for detecting anomalous gas consumption

<sup>4</sup>Predicting the Usage of Energy in a Smart Home Using Improved Weighted K-Means Clustering ARIMA Model

<sup>5</sup>A novel machine learning approach to analyzing geospatial vessel patterns using AIS data



**Figura 1. Desenvolvimento do método**

DBSCAN e Autoencoder e a posteriori a aplicação dos métodos, as sequências apontadas pelos mesmos como anômalas serão objetos de filtros realizados em três abordagens propostas a seguir:

**Distribuição Única:** Todas as janelas anômalas são agrupadas em uma única distribuição, em seguida O tipo de distribuição é identificado (normal, uniforme, etc.) e ao definir um limiar, pontos acima do mesmo são considerados anômalos.

**Distribuição Única das janelas normalizadas:** Todas as janelas são normalizadas por observações na janela e as sequências anômalas são agrupadas em uma única distribuição, em seguida O tipo de distribuição é identificado (normal, uniforme, etc.) e ao definir um limiar, pontos acima do mesmo são considerados anômalos.

**Janelas Estendidas:** As janelas anômalas são aumentadas para um tamanho maior, Essa técnica visa melhorar a estimativa das distribuições de probabilidade e pontos anômalos são identificados com base na distribuição estimada da janela estendida.

## 5. Avaliação experimental

Nesta seção serão aplicados os métodos e filtros apresentados na seção de metodologia a três bases sintéticas que representam três diferentes contextos, quais sejam: Série estacionária com ponto anômalo, Série estacionária com sequências anômalas e Serie não estacionária.

### Série estacionária com ponto anômalo

Foram aplicados nessa série os modelos: K- means, autoencoder e dbscan, respectivamente.

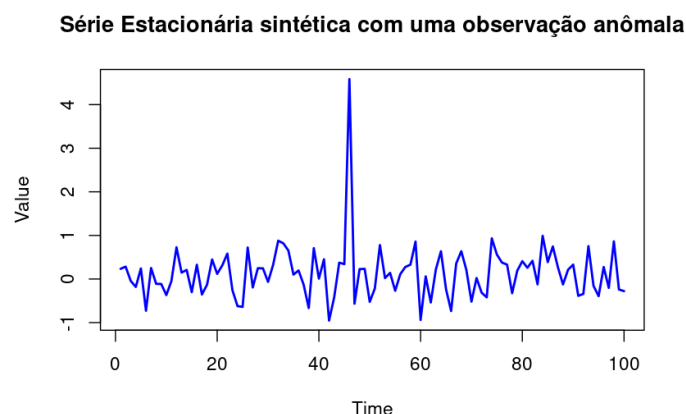
#### K-means

O método k-means foi aplicado fixando o tamanho da janela deslizante em 10 e o número de clusters igual a 1.

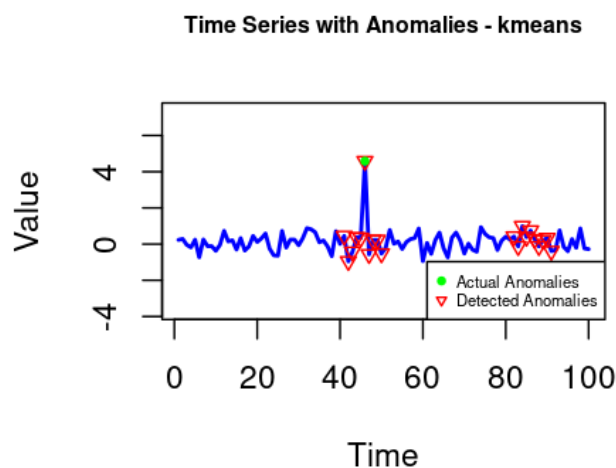
Segue o resultado encontrado:

Apesar de ter detectado a observação anômala, também tiveram muitos falso positivos, o que nos motiva a utilizar os filtros baseados em distribuição de probabilidade desenvolvidos para a diminuição dos falsos positivos.

Vamos aplicar o filtro nas três abordagens que foram propostas na metodologia, quais sejam: distribuição única, distribuição única das janelas normalizadas e janelas estendidas.



**Figura 2. Série estacionária com ponto anômalo**



**Figura 3. K-means - Série estacionária com ponto anômalo**

Todos os filtros serão aplicados considerando um limiar de probabilidade de 0.25, ou seja, só comporão a sequência àquelas observações que apresentarem probabilidade menor ou igual a 0.25 de pertencer a distribuição das sequências anômalas.

Em especial no filtro das janelas estendidas foi considerado a extensão da janela de tamanho 10 para mais ou para menos.

distribuição única

Podemos observar que o filtro de distribuição única atuou satisfatoriamente, pois não só detectou a observação anômala, como também reduziu significativamente os falsos positivos, só teve um falso positivo.

distribuição única das janelas normalizadas

Podemos observar que o filtro de distribuição única das janelas normalizadas atuou satisfatoriamente também, pois não só detectou a observação anômala, como também

Time Series with Anomalies - kmeans com filtro dist unica

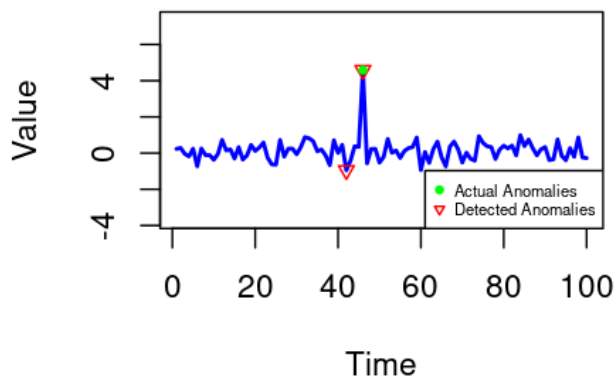


Figura 4. Filtro Distribuição única

Time Series with Anomalies - kmeans com filtro dist unica n

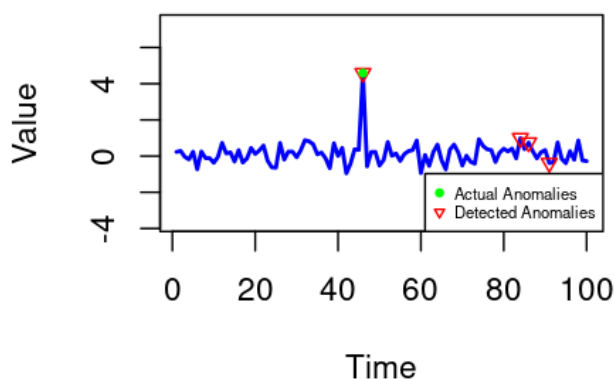


Figura 5. Filtro Distribuição única das janelas normalizadas

reduziu significativamente os falsos positivos, só tiveram três falso positivos, entretanto o primeiro filtro proporcionou uma maior redução dos falso positivos.

janelas estendidas

Podemos observar que o filtro de distribuição única das janelas estendidas atuou satisfatoriamente também, pois não só detectou a observação anômala, como também reduziu significativamente os falsos positivos, só tiveram três falso positivos, teve o mesmo desempenho do primeiro filtro.

DBSCAN

O método DBSCAN foi aplicado fixando o tamanho da janela deslizando em 10,

Time Series with Anomalies - kmeans com janela extendida di:

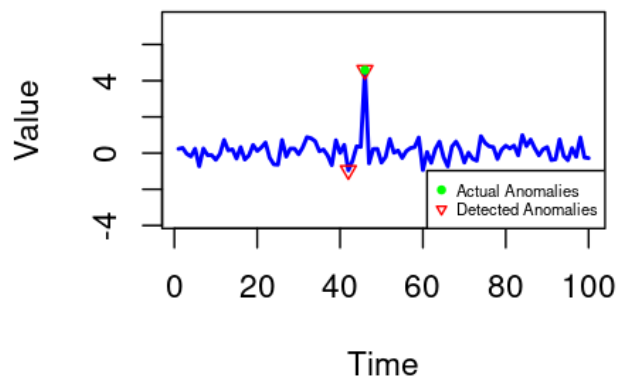


Figura 6. Filtro das janelas estendidas

minPts igual a 2 e o EPS é detectado automaticamente pelo algoritmo.

Segue o resultado encontrado:

Time Series with Anomalies - dbscan

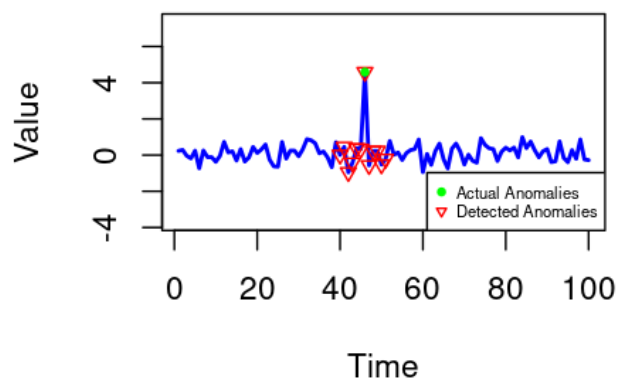


Figura 7. DBSCAN - Série estacionária com ponto anômalo

Apesar de ter detectado a observação anômala, também tiveram muitos falso positivos, o que nos motiva a utilizar os filtros baseados em distribuição de probabilidade desenvolvidos para a diminuição dos falsos positivos.

distribuição única

Podemos observar que o filtro de distribuição única atuou satisfatoriamente, pois não só detectou a observação anômala, como também reduziu significativamente os falsos positivos, só teve um falso positivo.



Time Series with Anomalies - dbscan com filtro dist unica

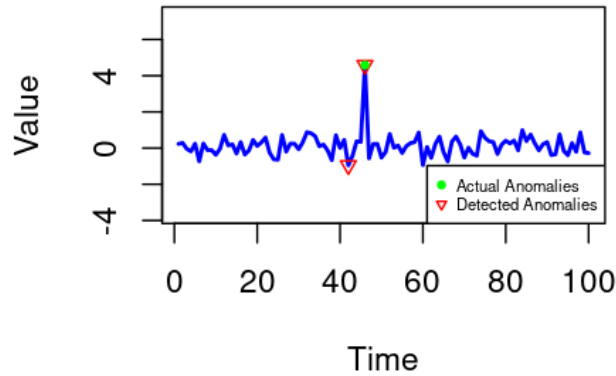


Figura 8. Filtro distribuição única

distribuição única das janelas normalizadas

Time Series with Anomalies - dbscan com filtro dist unica nc

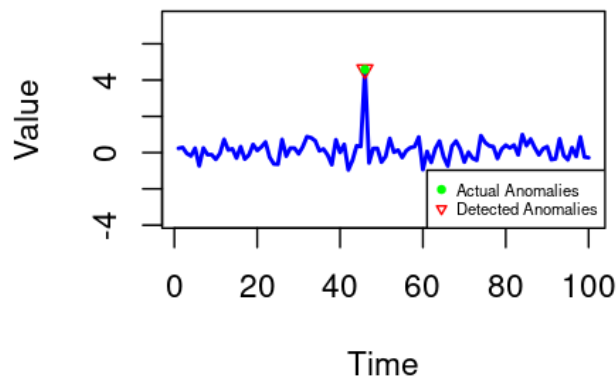


Figura 9. Filtro distribuição única das janelas normalizadas

Podemos observar que o filtro de distribuição única das janelas normalizadas atuou satisfatoriamente, pois não só detectou a observação anômala, como também zerou os falso positivos, mostrando melhor desempenho que o primeiro filtro aplicado.

janelas estendidas

Podemos observar que o filtro de distribuição única atuou satisfatoriamente, pois não só detectou a observação anômala, como também reduziu significativamente os falsos positivos, só teve um falso positivo, assim como o primeiro filtro aplicado, mas pior desempenho que o segundo filtro.

Time Series with Anomalies - dbscan com janela extendida di

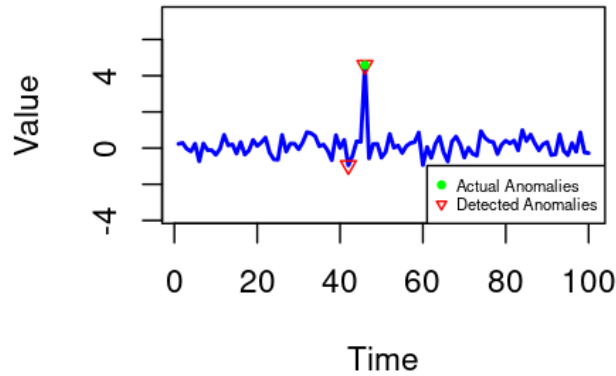


Figura 10. Filtro janelas estendidas

Autoencoder

O método Autoencoder foi aplicado fixando o tamanho da janela deslizante em 10 e encode igual a 1.

Segue o resultado encontrado:

Time Series with Anomalies - autoencoder

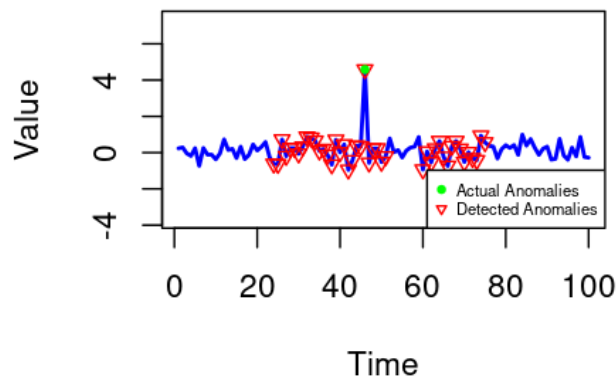


Figura 11. Autoencoder - Série estacionária com ponto anômalo

Apesar de ter detectado a observação anômala, também tiveram muitos falso positivos, o que nos motiva a utilizar os filtros baseados em distribuição de probabilidade desenvolvidos para a diminuição dos falsos positivos.

distribuição única

Podemos observar que o filtro de distribuição única atuou satisfatoriamente, pois

Time Series with Anomalies - autoencoder com filtro dist un

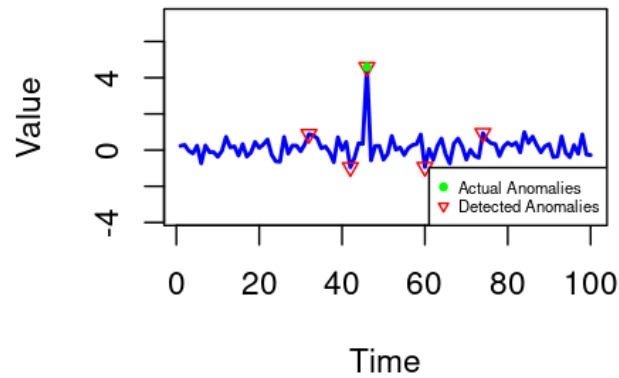


Figura 12. Filtro distribuição única

não só detectou a observação anômala, como também reduziu significativamente os falsos positivos, só tiveram quatro falso positivos.

distribuição única das janelas normalizadas

Time Series with Anomalies - autoencoder com filtro dist unica

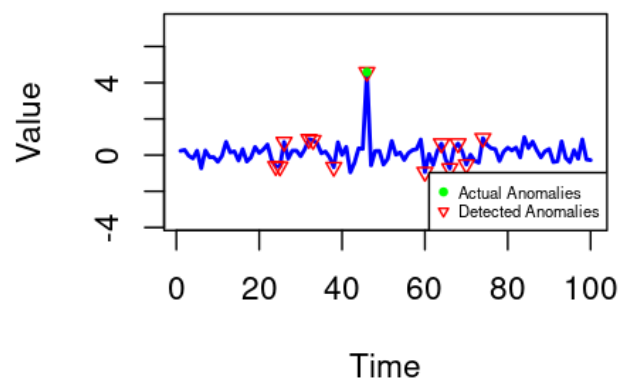


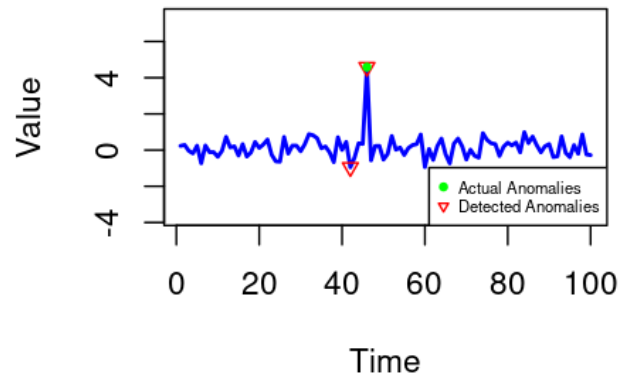
Figura 13. Filtro distribuição única das janelas normalizadas

Podemos observar que o filtro de distribuição única detectou a observação anômala, mas apesar de ter reduzido os falso positivos, ainda assim restaram 10 falso positivos.

janelas estendidas

Podemos observar que o filtro das janelas estendidas atuou satisfatoriamente, pois não só detectou a observação anômala, como também reduziu significativamente os falsos

**Time Series with Anomalies - autoencoder com janela estendida**

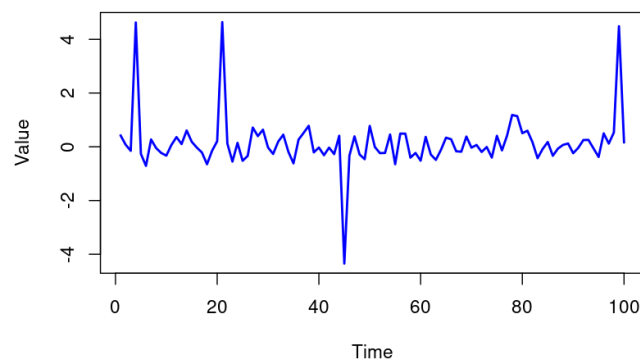


**Figura 14. Filtro janelas estendidas**

positivos, só teve um falso positivo e teve a melhor performance dos outros dois filtros aplicados.

Série estacionária com sequência anômala

**Série Estacionária com 4 observações anômalas**



**Figura 15. Série estacionária com sequência anômala**

Foram aplicados nessa série os modelos: K- means, autoencoder e dbscan, respectivamente.

K-means

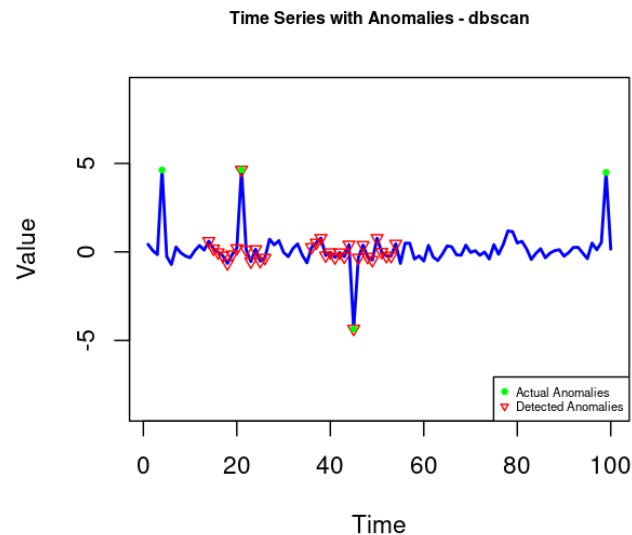
O método k-means foi aplicado fixando o tamanho da janela deslizante em 10 e o número de clusters igual a 1.

Não foi encontrada nenhuma observação anômala no k-means, logo não há como filtrar.

DBSCAN

O método DBSCAN foi aplicado fixando o tamanho da janela deslizante em 10, minPts igual a 2 e o EPS é detectado automaticamente pelo algoritmo.

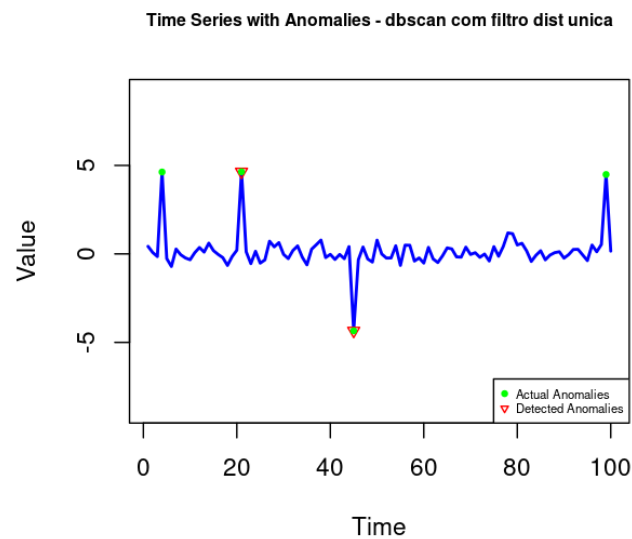
Segue o resultado:



**Figura 16. DBSCAN - Série estacionária sequência anômala**

Apesar de ter detectado duas das quatro observações anômalas, também tiveram muitos falso positivos, o que nos motiva a utilizar os filtros baseados em distribuição de probabilidade desenvolvidos para a diminuição dos falsos positivos.

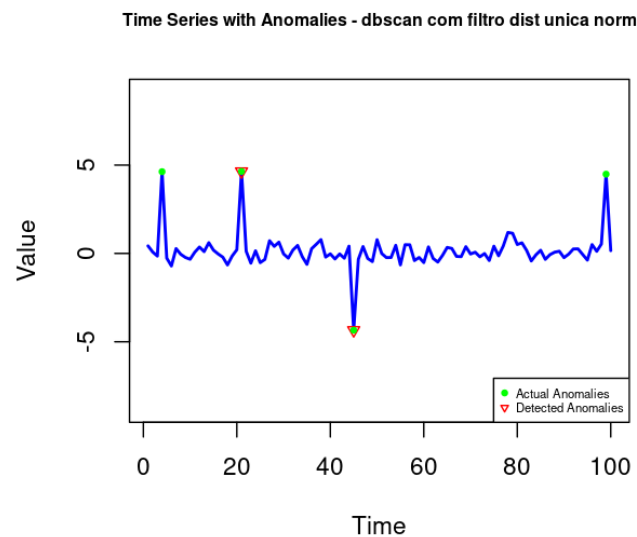
distribuição única



**Figura 17. Filtro distribuição única**

Podemos observar que o filtro de distribuição única atuou satisfatoriamente, pois não só detectou as duas observações anômalas que poderia detectar, como também zerou os falsos positivos.

distribuição única das janelas normalizadas



**Figura 18. Filtro distribuição única das janelas normalizadas**

Podemos observar que o filtro de distribuição única das janelas normalizadas atuou satisfatoriamente, pois não só detectou as duas observações anômalas que poderia detectar, como também zerou os falsos positivos.

janelas estendidas

Podemos observar que o filtro das janelas estendidas atuou satisfatoriamente, pois não só detectou as duas observações anômalas que poderia detectar, porém houve dois falso positivos, performando pior que o primeiro e segundo filtros.

Autoencoder

O método Autoencoder foi aplicado fixando o tamanho da janela deslizante em 10 e encode igual a 1.

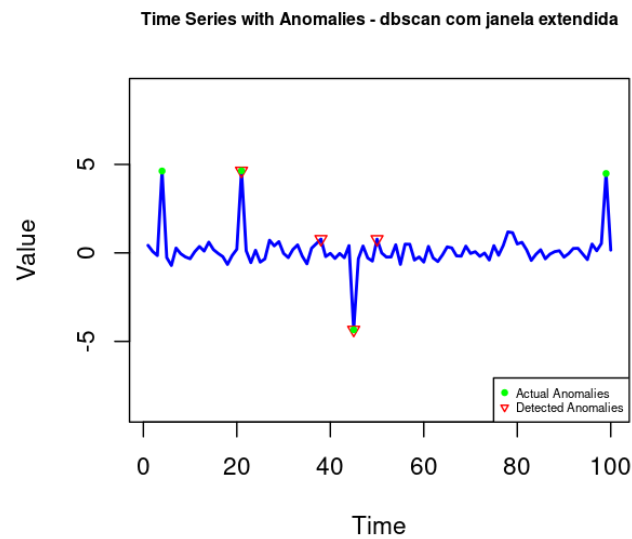
segue o resultado:

Apesar de ter detectado todas as quatro observações anômalas, também tiveram muitos falso positivos, o que nos motiva a utilizar os filtros baseados em distribuição de probabilidade desenvolvidos para a diminuição dos falsos positivos.

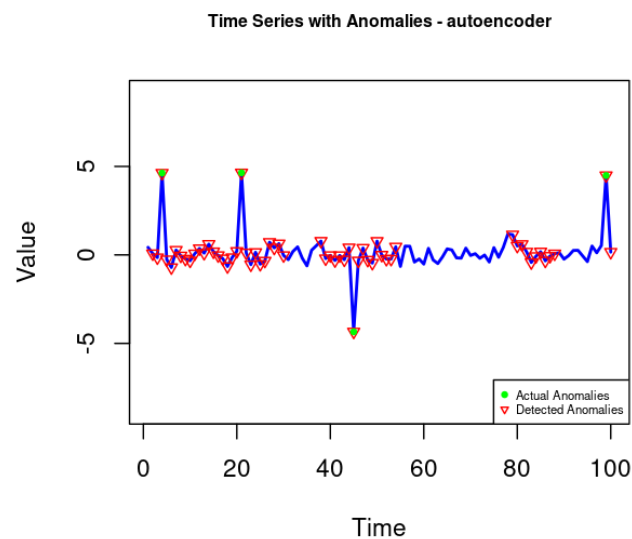
distribuição única

Podemos observar que o filtro de distribuição única atuou satisfatoriamente, pois detectou todas as quatro observações anômalas que poderia detectar, entretanto produziu um falso positivo.

distribuição única das janelas normalizadas



**Figura 19. Filtro das janelas estendidas**



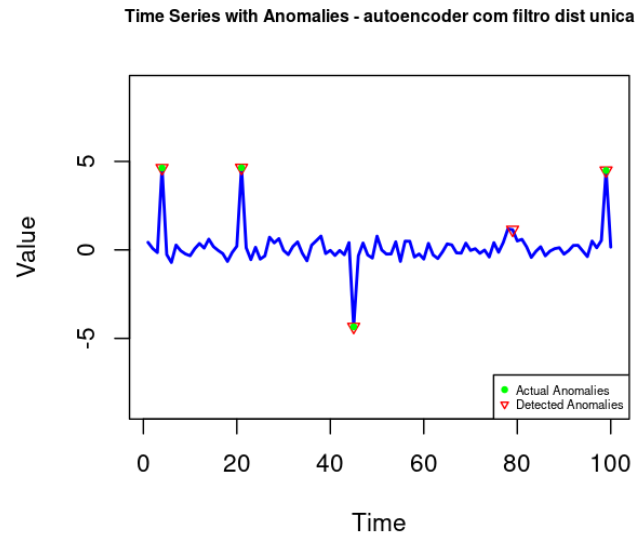
**Figura 20. Autoencoder - Série estacionária sequência anômala**

Podemos observar que o filtro de distribuição única das janelas normalizadas atuou satisfatoriamente, mas apesar de detectar três das quatro observações anômalas, produziu muitos falsos positivos.

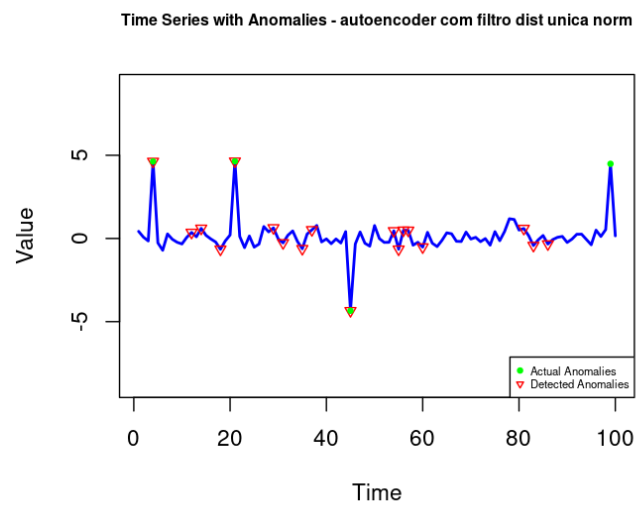
janelas estendidas

Podemos observar que o filtro das janelas estendidas atuou satisfatoriamente, pois detectou todas as quatro observações anômalas que poderia detectar, entretanto produziu um falso positivo, assim como no primeiro filtro.

Série não-estacionária com observação anômala



**Figura 21. Filtro distribuição única**



**Figura 22. Filtro distribuição única das janelas normalizadas**

Foram aplicados nessa série os modelos: K- means, autoencoder e dbscan, respectivamente.

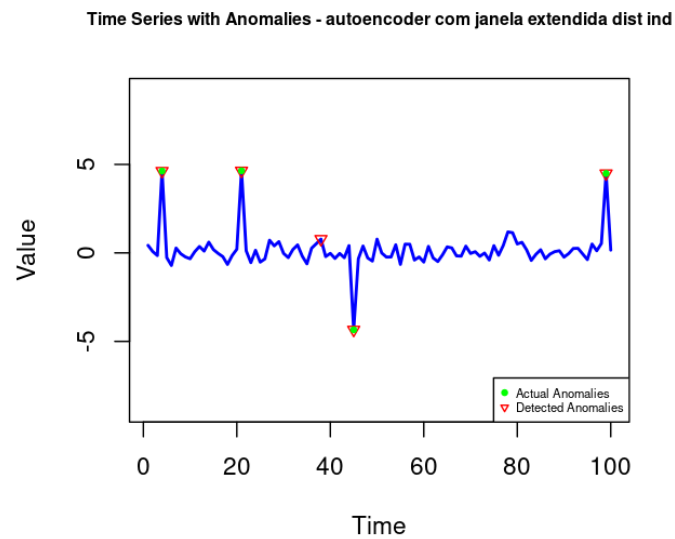
K-means

O método k-means foi aplicado fixando o tamanho da janela deslizante em 10 e o número de clusters igual a 1.

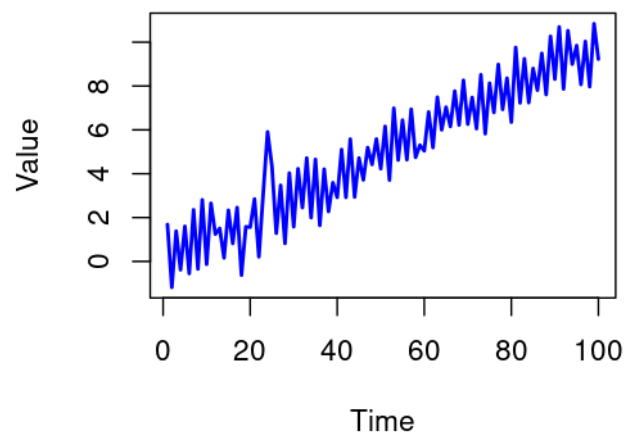
Não foi encontrada nenhuma observação anômala no k-means, logo não há como filtrar.

DBSCAN





**Figura 23. Filtro das janelas estendidas**



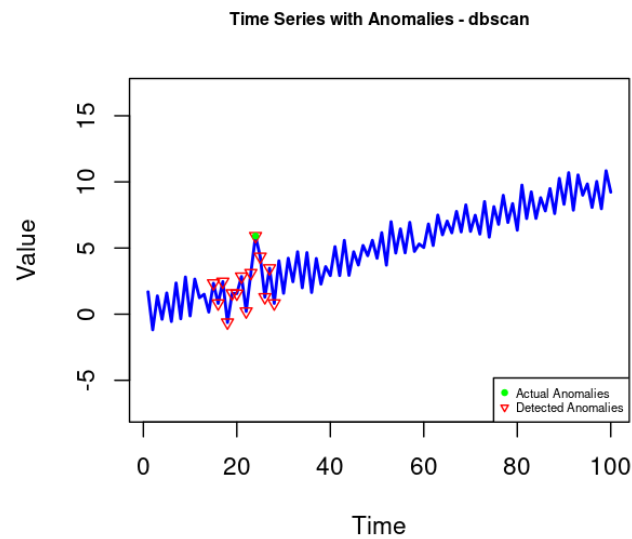
**Figura 24. Série não-estacionária com observação anômala**

O método DBSCAN foi aplicado fixando o tamanho da janela deslizante em 10, minPts igual a 2 e o EPS é detectado automaticamente pelo algoritmo.

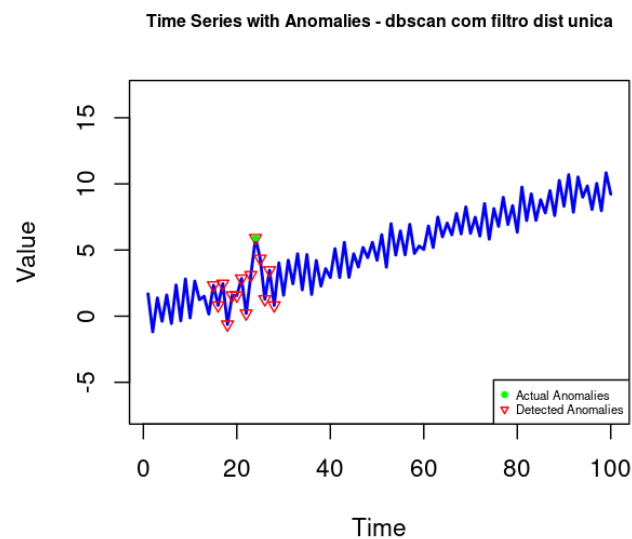
Segue o resultado:

Apesar de ter detectado a observação anômala, também tiveram muitos falso positivos, o que nos motiva a utilizar os filtros baseados em distribuição de probabilidade desenvolvidos para a diminuição dos falsos positivos.

distribuição única



**Figura 25. DBSCAN - Série não-estacionária observação anômala**



**Figura 26. Filtro distribuição única**

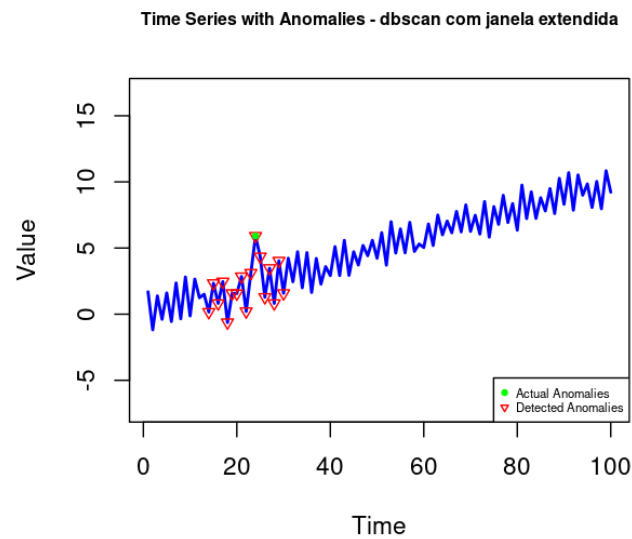
Podemos observar que o filtro de distribuição única não atuou satisfatoriamente, pois apesar de detectar observação anômala, continuou tendo muitos falso positivos

distribuição única com janelas normalizadas

Esse filtro não conseguiu filtrar observações anômalas

janelas estendidas

Podemos observar que o filtro das janelas estendidas não atuou satisfatoriamente, pois apesar de detectar observação anômala, continuou tendo muitos falso positivos.

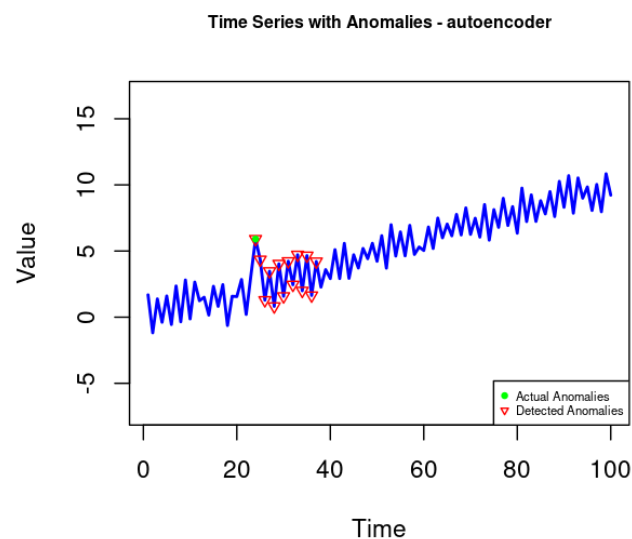


**Figura 27. Filtro das janelas estendidas**

#### Autoencoder

O método Autoencoder foi aplicado fixando o tamanho da janela deslizando em 10 e encode igual a 1.

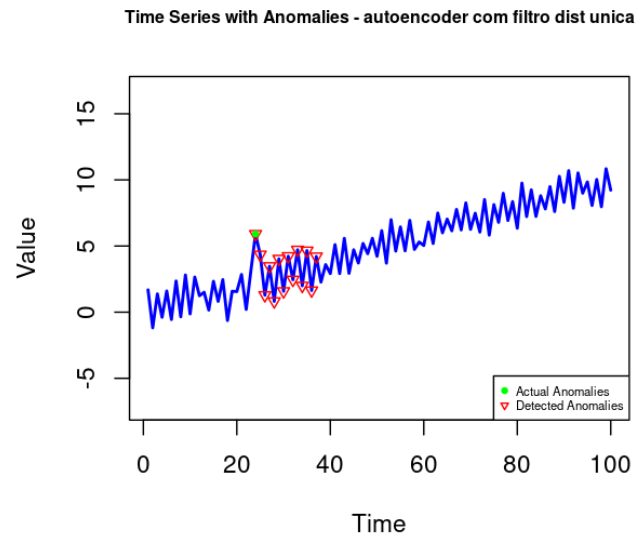
segue o resultado:



**Figura 28. Autoencoder - Série não-estacionária observação anômala**

Apesar de ter detectado a observação anômala, também tiveram muitos falso positivos, o que nos motiva a utilizar os filtros baseados em distribuição de probabilidade desenvolvidos para a diminuição dos falsos positivos.

distribuição única



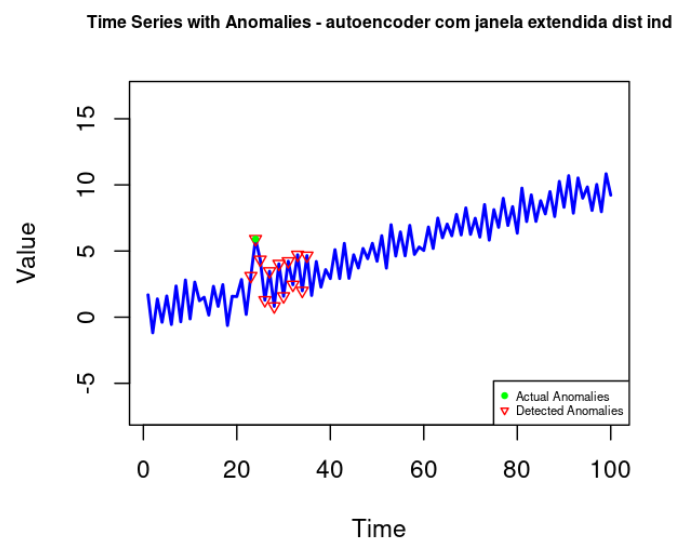
**Figura 29. Filtro distribuição única**

Podemos observar que o filtro de distribuição única não atuou satisfatoriamente, pois apesar de detectar observação anômala, continuou tendo muitos falso positivos.

distribuição única com janelas normalizadas

Esse filtro não conseguiu filtrar observações anômalas

janelas estendidas



**Figura 30. Filtro das janelas estendidas**

Podemos observar que o filtro das janelas estendidas não atuou satisfatoriamente,

pois apesar de detectar observação anômala, continuou tendo muitos falso positivos.

## 6. Conclusão

Esse trabalho teve como objetivo propor um método novo para potencializar a detecção de anomalias por meio da combinação da utilização de métodos não supervisionados com uma abordagem probabilística para filtrar não somente sequências anômalas, como também observações anômalas.

Esses métodos foram testados em bases sintéticas produzidas em cenários de estacionariedade e não estacionariedade, possuindo observações ou sequências anômalas, ambos os três algoritmos quando aplicados em séries com observações ou sequências anômalas se mostraram eficientes, minimizando ou até mesmo zerando os falso positivos e filtrando as observações que de fato eram anômalas, entretanto em cenários com bases sintéticas não estacionárias o filtro acaba não sendo tão eficiente na filtragem, uma vez que as distribuições acabam mudando com a tendência evidenciada na série testada.

Como trabalhos futuros, fica a sugestão de realizar simulações de monte-carlos produzindo um grande número de séries sintéticas com diferentes padrões para avaliar os pontos positivos e negativos do filtro probabilístico desenvolvido.

## Agradecimentos

Agradecimento as agências de fomento<sup>6</sup>. <sup>7</sup>.

## Referências

- Corrado, S. J., Puranik, T. G., and Mavris, D. N. (2021). Characterizing terminal airspace operational states and detecting airspace-level anomalies. *Engineering Proceedings*, 13(1):9.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231.
- Ferreira, M. D., Campbell, J. N., and Matwin, S. (2022). A novel machine learning approach to analyzing geospatial vessel patterns using ais data. *GIScience & Remote Sensing*, 59(1):1473–1490.
- Magdalene, J. J. C. and Zoraida, B. (2022). Predicting the usage of energy in a smart home using improved weighted k-means clustering arima model. *Journal of Algebraic Statistics*, 13(2):1770–1777.
- Morettin, P. A. and Toloi, C. M. d. C. (2022). Análise de séries temporais.
- Pei, Y. et al. (2018). A study on feature extraction of handwriting data using kernel method-based autoencoder. In *2018 9th International Conference on Awareness Science and Technology (iCAST)*, pages 1–6. IEEE.
- Princz, G., Shaloo, M., and Erol, S. (2024). Anomaly detection in binary time series data: An unsupervised machine learning approach for condition monitoring. *Procedia Computer Science*, 232:1065–1078.

---

<sup>6</sup>Agradecemos a CAPES pela disponibilização de acesso ao Portal de Periódicos para realizar este trabalho

<sup>7</sup>Agradecemos ao professor Eduardo Ogasawara do cefet do Rio de janeiro por todo apoio ao desenvolver este artigo

- Souza, D. N. d., Studart, T. M. d. C., and Lima Neto, I. E. (2013). Erro tipo ie “poder” de testes de aderência para postos pluviométricos no ceará.
- Zhou, Y., Jiang, J., Yang, S.-H., He, L., Ding, Y., Liu, K., Zhu, G., and Qing, Y. (2023). A data distillation enhanced autoencoder for detecting anomalous gas consumption. *IEEE Internet of Things Journal*.