



Trabalho de Conclusão de Curso

**Aplicação de Aprendizado de Máquina na  
Identificação de Reações de Pragas a Compostos  
Semioquímicos**

Elias Nogueira da Silva  
ens@ic.ufal.br

**Orientadores:**

Prof. Dr. André Luiz Lins de Aquino  
Me. Danilo Fernandes Costa

Maceió, Dezembro de 2024

Elias Nogueira da Silva

**Aplicação de Aprendizado de Máquina na  
Identificação de Reações de Pragas a Compostos  
Semioquímicos**

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação do Instituto de Computação da Universidade Federal de Alagoas.

Orientadores:

Prof. Dr. André Luiz Lins de Aquino

Me. Danilo Fernandes Costa

Maceió, Dezembro de 2024

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**  
Bibliotecária: Girlaine da Silva Santos – CRB-4 – 1127

S586a Silva, Elias Nogueira da.  
Aplicação de aprendizado de máquina na identificação de reações de pragas a compostos semioquímicos / Elias Nogueira da Silva. – 2024.  
46 f.: il

Orientador: André Lins de Aquino.  
Coorientador: Danilo Fernandes Costa.  
Monografia (Trabalho de Conclusão de Curso em Ciência da Computação: Bacharelado) - Universidade Federal de Alagoas, Instituto de Computação, Maceió, 2024.

Bibliografia: f. 43- 46.

1. Aprendizagem de Máquina. 2. Séries temporais. 3. Cromatografia gasosa. 4. Eletroantenografia. I. Título.

CDU: 004.8:632

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação do Instituto de Computação da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina.

---

Prof. Dr. André Luiz Lins de Aquino - Orientador  
Universidade Federal de Alagoas

---

Me. Danilo Fernandes Costa - Coorientador  
Universidade Federal de Alagoas

---

Douglas Leite Leal Moura - Examinador  
Departamento de Ciência da Computação  
Universidade Federal de Minas Gerais

---

Eliane dos Santos - Examinador  
Instituto de Química e Biotecnologia  
Universidade Federal de Alagoas

---

Josias Jordão Andrade Alves - Examinador  
Instituto de Química e Biotecnologia  
Universidade Federal de Alagoas

# Agradecimentos

Meus agradecimentos vão, em primeiro lugar, à minha mãe, que sempre esteve ao meu lado, me apoiando em todas as etapas da vida e em cada decisão que tomei. Expresso também minha gratidão aos meus orientadores e amigos, André e Danilo, pelo suporte e orientação ao longo deste caminho.

Agradeço à Universidade Federal de Alagoas e ao Instituto de Computação, que se tornaram minha segunda casa durante minha formação, proporcionando um ambiente de aprendizado e crescimento.

Aos meus amigos do V.S., que estiveram comigo ao longo de toda a graduação e que levarei comigo para a vida, minha mais profunda gratidão.

E, é claro, um agradecimento especial aos colegas do laboratório ORION, que constantemente contribuem para minha formação acadêmica e pessoal, e com quem tenho aprendido tanto ao longo desses anos.

Elias Nogueira da Silva

# Resumo

Este trabalho investiga a aplicação de técnicas de aprendizado de máquina para identificar reações de pragas a compostos semioquímicos. As séries temporais analisadas foram obtidas a partir de dados de reações de insetos por eletroantenografia (EAG) e respostas de um sistema de cromatografia gasosa com detector de ionização de chama (GC-FID). O *pipeline* metodológico desenvolvido inclui etapas de pré-processamento, como a remoção de dados não utilizáveis, padronização e equalização de frequências entre as séries, utilizando duas abordagens distintas: remoção de dados excedentes e interpolação.

Para unificar os sinais de cada experimento, foi aplicado o valor mínimo absoluto entre as séries, técnica que preserva picos simultâneos — potencialmente correspondentes às reações dos insetos aos semioquímicos — enquanto reduz picos isolados considerados ruídos. As séries unificadas foram então analisadas por modelos de aprendizado de máquina para detecção de anomalias, incluindo K-Means e Isolation Forest. A análise comparativa das abordagens de equalização indicou que o impacto da interpolação foi baixo, com influência mínima sobre o desempenho do K-Means. Ambos os modelos apresentaram bons resultados na identificação das reações, destacando-se o K-Means com janelas deslizantes como o mais eficaz para a detecção das respostas.

**Palavras-chave:** Aprendizado de máquina, Séries temporais, Eletroantenografia, Cromatografia gasosa.

# Abstract

This study explores the application of machine learning techniques to identify pest reactions to semiochemical compounds. The analyzed time series consist of insect reaction data obtained through electroantennography (EAG) and responses from a gas chromatography system with flame ionization detector (GC-FID). The developed methodological pipeline includes preprocessing steps such as removal of unusable data, data standardization, and frequency equalization between the series using two distinct approaches: removal of excess data and interpolation.

To unify the signals in each experiment, the minimum absolute value between the series was applied, a technique that preserves simultaneous peaks—potentially corresponding to insect reactions to semiochemicals—while reducing isolated peaks considered noise. The unified series were then analyzed using machine learning models for anomaly detection, including K-Means and Isolation Forest. A comparative analysis of the equalization approaches showed that the impact of interpolation was low, with minimal influence on K-Means performance. Both models achieved good results in detecting reactions, with K-Means using sliding windows being the most effective for identifying the responses.

**Keywords:** Machine learning, Time series, Electroantennography, Gas chromatography.

# Lista de Figuras

|      |  |    |
|------|--|----|
| 2.1  | Fluxograma da metodologia utilizada. . . . .   | 4  |
| 2.2  | Esquema simplificado do funcionamento do GC-FID acoplado ao EAG. . . . .                     | 6  |
| 2.3  | Dados brutos utilizados. . . . .   | 7  |
| 2.4  | Dados brutos do experimento com a espécie <i>Rhynchophorus</i> . . . . .                     | 8  |
| 3.1  | Exemplo de remoção do início dos dados. . . . .  | 16 |
| 3.2  | Todos experimentos após remoção de dados do hexano, com marcação nas reações. . . . .        | 17 |
| 3.3  | Sinal padronizados - EAG. . . . .  | 18 |
| 3.4  | Sinal padronizados - FID. . . . .  | 18 |
| 3.5  | Exemplo de equalização por redução. . . . .  | 20 |
| 3.6  | Equalização por interpolação dos sinais de FID. . . . .                                      | 21 |
| 3.7  | Espectro de Frequência da série . . . . .  | 22 |
| 3.8  | Suavização da série equalizada . . . . .   | 22 |
| 3.9  | Ampliação da suavização da série equalizada . . . . .  | 23 |
| 3.10 | Unificação da série equalizada por redução. . . . .  | 24 |
| 3.11 | Unificação da série equalizada por interpolação. . . . .                                     | 24 |
| 3.12 | Aplicação Isolation Forest - <i>Rhynchophorus</i> . . . . .                                  | 26 |
| 3.13 | Aplicação do Isolation Forest - <i>Spodoptera</i> . . . . .                                  | 26 |
| 3.14 | Aplicação do Isolation Forest - <i>Spodoptera</i> amostra 2. . . . .                         | 27 |
| 3.15 | Distância do centroide <i>Rhynchophorus</i> experimento 1. . . . .                           | 28 |
| 3.16 | Aplicação K-means agrupamento direto - <i>Rhynchophorus</i> experimento 1. . . . .           | 28 |
| 3.17 | Distância do centroide <i>Rhynchophorus</i> experimento 1. . . . .                           | 29 |
| 3.18 | Aplicação do K-means agrupamento direto - <i>Rhynchophorus</i> experimento 2. . . . .        | 29 |
| 3.19 | Distância ao centroide - <i>Spodoptera</i> experimento 1. . . . .                            | 30 |
| 3.20 | Aplicação do K-means agrupamento direto - <i>Spodoptera</i> experimento 1. . . . .           | 30 |
| 3.21 | Distância ao centroide - <i>Spodoptera</i> experimento 2. . . . .                            | 31 |
| 3.22 | Aplicação do K-means agrupamento direto - <i>Spodoptera</i> experimento 2. . . . .           | 31 |
| 3.23 | Distância ao centroide - <i>Spodoptera</i> amostra 2 experimento 1. . . . .                  | 32 |
| 3.24 | Aplicação do K-means agrupamento direto - <i>Spodoptera</i> amostra 2 experimento 1. . . . . | 32 |
| 3.25 | Distância ao centroide - <i>Spodoptera</i> amostra 2 experimento 2. . . . .                  | 33 |
| 3.26 | Aplicação do K-means agrupamento direto - <i>Spodoptera</i> amostra 2 experimento 2. . . . . | 33 |
| 3.27 | Distância do centroide - <i>Rhynchophorus</i> experimento 1. . . . .                         | 35 |
| 3.28 | Aplicação K-means com janelas deslizantes - <i>Rhynchophorus</i> experimento 1. . . . .      | 35 |
| 3.29 | Distância ao centroide - <i>Rhynchophorus</i> experimento 2. . . . .                         | 36 |
| 3.30 | Aplicação do K-means com janelas deslizantes - <i>Rhynchophorus</i> experimento 2. . . . .   | 36 |
| 3.31 | Distância ao centroide - <i>Spodoptera</i> experimento 1. . . . .                            | 37 |
| 3.32 | Aplicação do K-means com janelas deslizantes - <i>Spodoptera</i> experimento 1. . . . .      | 37 |
| 3.33 | Distância ao centroide - <i>Spodoptera</i> experimento 2. . . . .                            | 38 |

---

|      |  |    |
|------|--|----|
| 3.34 | Aplicação do K-means com janelas deslizantes - <i>Spodoptera</i> experimento 2. . .                    | 38 |
| 3.35 | Distância ao centroide - <i>Spodoptera</i> amostra 2 experimento 1. . . . .                            | 39 |
| 3.36 | Aplicação do K-means com janelas deslizantes - <i>Spodoptera</i> amostra 2 experi-<br>mento 1. . . . . | 39 |
| 3.37 | Distância ao centroide - <i>Spodoptera</i> amostra 2 experimento 2. . . . .                            | 40 |
| 3.38 | Aplicação do K-means com janelas deslizantes - <i>Spodoptera</i> amostra 2 experi-<br>mento 2. . . . . | 40 |

# Lista de Algoritmos

|     |   |    |
|-----|---|----|
| 2.1 | Equalização de Frequências pela Remoção de Amostras . . . . . | 9  |
| 2.2 | Detecção de anomalias com Isolation Forest . . . . .          | 13 |
| 2.3 | K-Means . . . . .   | 14 |
| 2.4 | Detecção de anomalias com K-Means . . . . .                   | 15 |

# Conteúdo

|  |            |
|--|------------|
| <b>Lista de Figuras</b>  | <b>iii</b> |
| <b>1 Introdução</b>  | <b>1</b>   |
| 1.1 Motivação . . . . .  | 1          |
| 1.2 Trabalhos relacionados . . . . .   | 2          |
| <b>2 Metodologia</b>   | <b>4</b>   |
| 2.1 Pre processamento . . . . .  | 4          |
| 2.1.1 Dados . . . . .  | 5          |
| 2.1.1.1 Cromatografia Gasosa com Detector de Ionização de Chama (GC-FID) . . . . . | 5          |
| 2.1.1.2 Eletroantenografia (EAG) . . . . .   | 5          |
| 2.1.1.3 Integração GC-FID e EAG (GC-EAD) . . . . .                                 | 5          |
| 2.1.1.4 Conjunto de Dados . . . . .  | 6          |
| 2.1.2 Remoção do início dos dados . . . . .  | 7          |
| 2.1.3 Padronização . . . . .   | 8          |
| 2.1.4 Equalização da Frequência das Séries . . . . .                               | 9          |
| 2.1.4.1 Redução da Frequência da Série de Maior Amostragem . . . . .               | 9          |
| 2.1.4.2 Interpolação Linear . . . . .  | 10         |
| 2.1.5 Suavização . . . . .   | 10         |
| 2.1.5.1 Filtro passa-baixa . . . . .   | 11         |
| 2.2 Unificação dos Sinais . . . . .  | 11         |
| 2.3 Avaliação com Modelos de Machine Learning . . . . .                            | 12         |
| 2.3.1 Isolation Forest . . . . .   | 12         |
| 2.3.2 K-means . . . . .  | 14         |
| 2.3.2.1 Detecção de Anomalias com K-Means . . . . .                                | 14         |
| <b>3 Resultados e Discussões</b>   | <b>16</b>  |
| 3.1 Remoção de dados não utilizáveis . . . . .                                     | 16         |
| 3.2 Resultado da padronização dos dados . . . . .                                  | 17         |
| 3.3 Equalização de frequências . . . . .   | 19         |
| 3.3.1 Equalização por redução . . . . .  | 19         |
| 3.3.2 Equalização por interpolação . . . . .                                       | 20         |
| 3.4 Suavização com filtro passa-baixa . . . . .                                    | 21         |
| 3.5 Unificação de sinais . . . . .   | 24         |
| 3.6 Resultado dos modelos . . . . .  | 25         |
| 3.6.1 Isolation forest . . . . .   | 25         |
| 3.6.2 K-means utilizando valores individuais . . . . .                             | 27         |

---

|  |           |
|--|-----------|
| 3.6.3 K-means utilizando janela deslizante . . . . . | 34        |
| <b>4 Conclusão</b>                                   | <b>42</b> |
| <b>Referências bibliográficas</b>                    | <b>43</b> |

# 1

## Introdução

### 1.1 Motivação

A comunicação química é um dos pilares fundamentais da ecologia e do comportamento dos insetos. Essa interação é mediada por substâncias químicas conhecidas como semioquímicos, um grupo amplo de compostos que influenciam respostas comportamentais e fisiológicas entre organismos. Dentre os semioquímicos, destacam-se os feromônios, substâncias emitidas por um indivíduo e detectadas por outros da mesma espécie, modulando comportamentos como reprodução, forrageamento e defesa (Rya, 2002; Yew and Chung, 2015). Conforme destacado em Ezzat et al. (2020), os semioquímicos são utilizados pelos insetos para transmitir mensagens químicas específicas que influenciam sua fisiologia e comportamento, oferecendo oportunidades para a manipulação desses processos em aplicações práticas.

Na agricultura, o uso de semioquímicos é uma abordagem inovadora e sustentável para o manejo de pragas. Substituindo ou complementando os pesticidas convencionais, os semioquímicos permitem o controle de pragas com menor impacto ambiental e maior especificidade, como observado em Ezzat et al. (2020). Essa característica é particularmente importante em um cenário global que exige práticas agrícolas mais sustentáveis e eficientes.

Apesar do seu potencial, a identificação e aplicação de semioquímicos enfrentam desafios significativos. A análise de compostos orgânicos voláteis (VOCs), frequentemente complexos e presentes em baixas concentrações, requer técnicas avançadas e sensíveis. Além disso, a interação entre os compostos químicos e os receptores olfativos dos insetos ainda não é completamente compreendida, como apontado em Barbosa-Cornelio et al. (2019).

Nesse contexto, o uso de aprendizado de máquina apresenta-se como uma ferramenta poderosa e promissora para superar esses desafios. Técnicas de aprendizado de máquina podem ser aplicadas para analisar grandes volumes de dados químicos e biológicos, identificar padrões complexos e correlacionar respostas comportamentais de pragas a estímulos químicos. Essas

abordagens permitem uma análise mais robusta e eficiente, acelerando a descoberta de compostos relevantes e otimizando sua aplicação prática.

Com isso, o presente trabalho se propõe a explorar o uso de aprendizagem de máquina para identificar reações de pragas agrícolas a compostos semioquímicos. A aplicação de técnicas de aprendizado de máquina em biotecnologia, como a detecção de anomalias e diagnósticos baseados em dados históricos, tem mostrado um grande potencial. Um exemplo disso é o trabalho de [Zeberli et al. \(2021\)](#), que apresenta uma abordagem de dois passos para a detecção de anomalias e diagnóstico em processos de manufatura farmacêutica, usando dados históricos para identificar falhas e suas causas. A integração do entendimento do processo com a análise de dados ajudou a reduzir falhas operacionais e aumentar a eficiência da produção.

Além disso, a aplicação de aprendizado de máquina na engenharia química também está se expandindo. Como discutido em [Schweidtmann et al. \(2021\)](#), o aprendizado de máquina é usado para otimizar processos químicos, prever propriedades moleculares e melhorar a eficiência na produção de produtos químicos.

Com isso, o presente trabalho se propõe a explorar o uso de aprendizado de máquina para identificar reações de pragas a compostos semioquímicos.

## 1.2 Trabalhos relacionados

Pesquisas anteriores demonstrando o potencial de técnicas como GC-FID, GC-EAD e EAG no manejo sustentável de pragas agrícolas a partir da detecção da reação dos insetos. Alguns exemplos relevantes incluem:

O estudo de [Būda et al. \(2016\)](#) utilizou GC-FID, GC-EAD e GC-MS para identificar compostos voláteis orgânicos (VOCs) emitidos por grãos de trigo, tanto infectados quanto não infectados por fungos, que despertam respostas comportamentais e olfativas em fêmeas acasaladas de *Plodia interpunctella*. Entre os compostos identificados, 1-hexanol e nonanal, emitidos por grãos não infectados, mostraram ser atrativos em baixas concentrações. Já o composto 4-oxoisoforona apresentou propriedades repelentes, sugerindo que ele pode ser empregado como um desestimulador de oviposição.

[Fürstenau et al. \(2016\)](#) investigaram a interação entre o parasitoide *Holepyris sylvanidis* e voláteis associados ao habitat de larvas hospedeiras (*Tribolium confusum*). Utilizando bioensaios com olfatômetros e análise por GC-MS acoplada à eletroantenografia (EAG), foi demonstrado que compostos presentes nas fezes das larvas, como (E)-2-nonenal e 1-pentadeceno, induzem respostas comportamentais e eletrofisiológicas significativas no parasitoide.

Em [Hassemer et al. \(2015\)](#), foram identificados 23 compostos emitidos pelas glândulas abdominais de *Alphitobius diaperinus*, sendo seis deles quinonas, como 2-metil-1, 4-benzoquinona e 2-etil-1, 4-benzoquinona, que se destacaram como os principais constituintes. A análise por GC-EAD revelou respostas significativas das antenas tanto de machos quanto de

fêmeas a essas quinonas.

O estudo de [Miano et al. \(2022\)](#) investigou a interação entre a traça-do-tomateiro (*Tuta absoluta*) e compostos voláteis orgânicos (VOCs) emitidos por uma variedade selvagem de tomate, *Solanum lycopersicum var. cerasiforme*, em comparação com variedades cultivadas. Por meio de ferramentas analíticas multivariadas, como Floresta Aleatória, análise discriminante de mínimos quadrados parciais esparsos e escala multidimensional não métrica (NMDS), foram identificados 20 VOCs como os principais discriminantes entre os voláteis emitidos por plantas selvagens e cultivadas, tanto durante o dia quanto à noite.

Adicionalmente, a análise por cromatografia gasosa acoplada à eletroantenografia (GC-EAG) revelou que compostos como (Z)-3-hexenol, cânfora, citronelal e limoneno óxido são detectáveis pelas antenas das fêmeas de *T. absoluta*. Testes em túnel de vento com uma mistura desses compostos mostraram que eles têm propriedades repelentes dependentes da dose, reduzindo a infestação nas plantas selvagens.

# 2

## Metodologia

Esse capítulo descreve as tecnologias utilizadas na pesquisa, incluindo o pré-processamento dos dados, que abrange as etapas de padronização e equalização de frequências, unificação dos sinais por meio do valor mínimo absoluto até finalmente a aplicação de modelos de aprendizado de máquina para detecção das respostas dos insetos para com os compostos, como podemos visualizar na Figura 2.1.

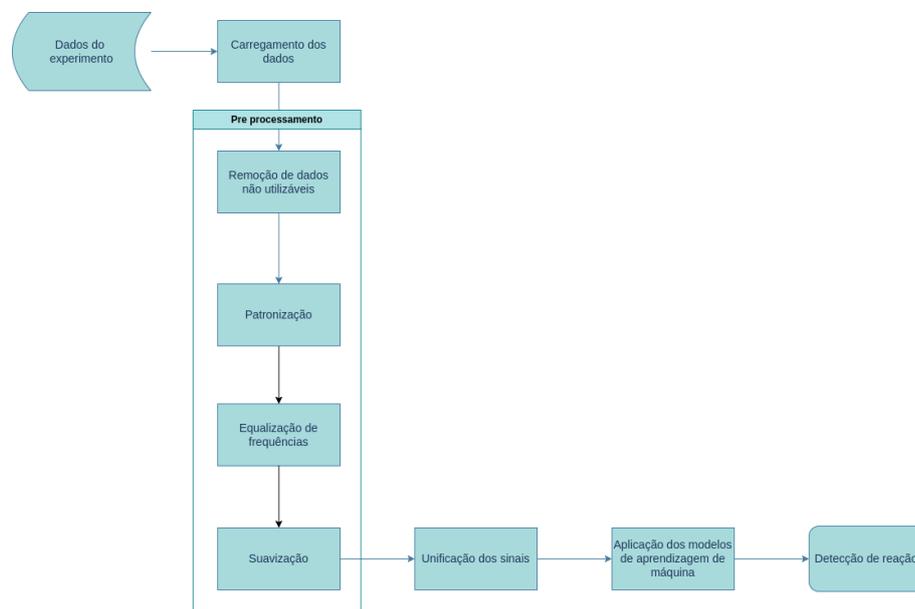


Figura 2.1: Fluxograma da metodologia utilizada.

### 2.1 Pre processamento

O pré-processamento de dados é uma etapa essencial no *pipeline* de aprendizado de máquina, sendo responsável por assegurar a qualidade, consistência e prontidão dos dados para análise

e modelagem. Conforme destacado na literatura, essa etapa engloba uma série de operações, como a identificação e tratamento de dados faltantes ou errôneos, normalização, escalonamento, seleção e transformação de características (Golazad et al., 2024). Essas práticas visam aprimorar a qualidade e a eficácia das análises subsequentes, garantindo que os modelos possam extrair padrões significativos de maneira confiável e eficiente.

Ao abordar questões como a presença de valores discrepantes, variações de escala entre atributos e incompatibilidades de frequência em séries temporais, o pré-processamento contribui diretamente para a confiabilidade dos modelos. Além disso, ele melhora a representação dos dados, reduzindo o impacto de ruídos ou inconsistências que poderiam comprometer os resultados. Como enfatizado por Golazad et al., 2024, o sucesso de métodos baseados em dados depende, em grande parte, de uma preparação adequada dos mesmos.

## 2.1.1 Dados

Os dados experimentais fornecidos para esta pesquisa foram coletados empregando técnicas amplamente reconhecidas e validadas na literatura para análise de compostos voláteis e suas interações com insetos:

### 2.1.1.1 Cromatografia Gasosa com Detector de Ionização de Chama (GC-FID)

A técnica de GC-FID é amplamente utilizada para identificar e quantificar compostos voláteis. O detector de ionização de chama (FID) é altamente sensível para compostos orgânicos, permitindo análise quantitativa. Trabalhos como Brezolin et al. (2018); Hassemer et al. (2015) demonstram a eficácia dessa abordagem na análise de semioquímicos e na identificação de VOCs associados ao comportamento dos insetos.

### 2.1.1.2 Eletroantenografia (EAG)

A eletroantenografia mede as respostas elétricas das antenas dos insetos a estímulos químicos, permitindo identificar compostos que afetam diretamente o comportamento dos mesmos (Olsson and Hansson, 2013). Estudos como o de Barbosa-Cornelio et al. (2019) e destacam a relevância dessa técnica em análises comportamentais, especialmente quando integrada ao GC (GC-EAG), que permite correlacionar respostas eletrofisiológicas com perfis químicos.

### 2.1.1.3 Integração GC-FID e EAG (GC-EAD)

A combinação GC-FID e EAG utiliza um conector em "Y" para dividir o fluxo da coluna cromatográfica, direcionando uma parte para o detector FID e outra para a antena do inseto. Essa abordagem integrada, ilustrada na Figura 2.2, é amplamente utilizada para identificar compostos bioativos relevantes no comportamento de insetos pragas (Myrick and Baker, 2012). Trabalhos como Shiota et al. (2018); Rya (2002) ressaltam que esse tipo de dados são fundamentais para

correlacionar estímulos químicos e comportamentos, especialmente em pragas de importância agrícola.

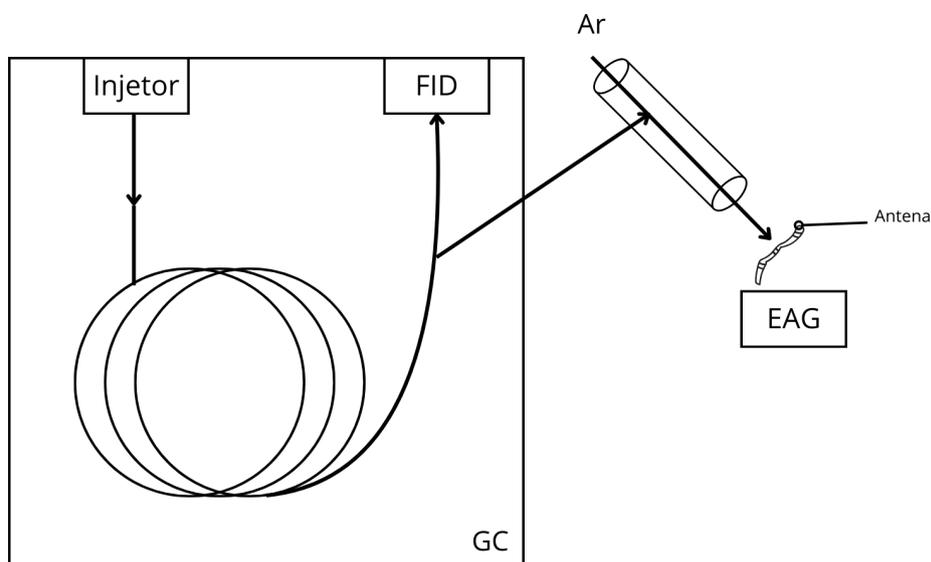


Figura 2.2: Esquema simplificado do funcionamento do GC-FID acoplado ao EAG.

#### 2.1.1.4 Conjunto de Dados

Para esse estudo tivemos acesso a seis experimentos, dos quais 2 envolveram a espécie *Rhynchophorus* e 4 a espécie *Spodoptera*. Cada experimento gerou duas séries temporais distintas: uma correspondente aos sinais do GC-FID e outra às respostas do EAG conforme ilustrado na Figura 2.3. Pode-se observar que os experimentos apresentam durações variadas e que, em todos os experimentos, os dados iniciais exibem valores muito altos provenientes do FID. Esses valores iniciais serão devidamente tratados em etapas posteriores da análise.

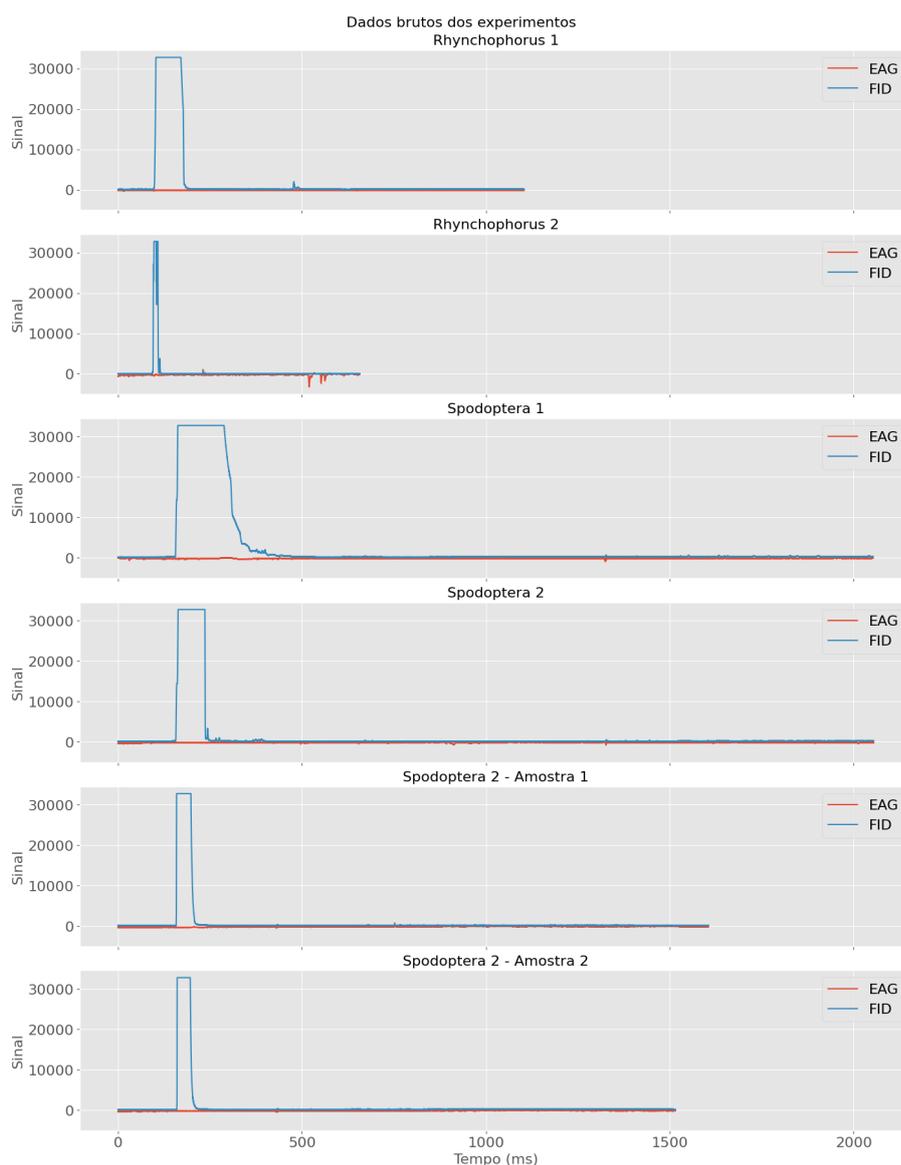


Figura 2.3: Dados brutos utilizados.

### 2.1.2 Remoção do início dos dados

O início dos dados capturados contém predominantemente informações relacionadas à presença de hexano (Figura 2.4), que é utilizado como solvente durante a extração de compostos voláteis semioquímicos, sendo empregado devido à sua baixa polaridade e alta eficiência na extração de compostos voláteis a partir de fontes biológicas, incluindo insetos e plantas, por métodos como extração assistida por solventes (SAE) (Barbosa-Cornelio et al., 2019).

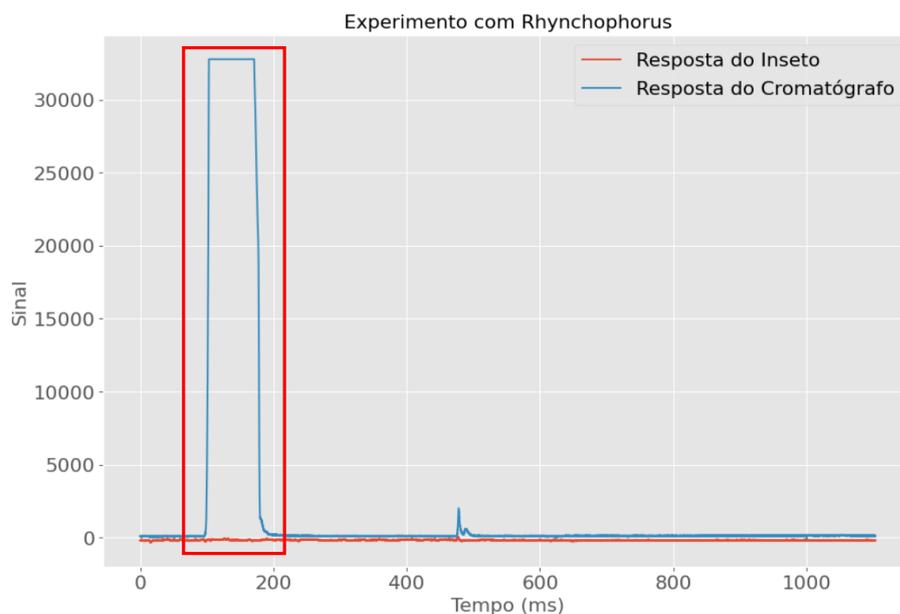


Figura 2.4: Dados brutos do experimento com a espécie *Rhynchophorus*

Analisando os dados podemos notar que nenhuma reação relevante ocorre durante esse período, portanto, os dados referentes a essa fase inicial não contribuem para a análise do comportamento do sistema ou dos resultados das reações. Em vez disso, eles podem introduzir ruídos ou influências desnecessárias nos modelos analíticos.

Portanto, foi optado pela exclusão dessa parte inicial dos dados, permitindo uma análise mais precisa e confiável.

### 2.1.3 Padronização

A padronização dos dados é uma etapa importante para garantir o desempenho adequado de algoritmos de aprendizado de máquina, especialmente em casos onde há grande variação entre os valores das características. Diferenças significativas, como valores que tenham grande variação, podem afetar negativamente o comportamento de modelos, comprometendo sua capacidade de identificar padrões relevantes e produzir resultados precisos (Alexandropoulos et al., 2019).

Este processo, visa transformar os valores das características para uma escala apropriada e comparável. Essa técnica é necessária para diversos algoritmos, incluindo redes neurais, máquinas de vetores de suporte (SVMs), algoritmos k-NN e classificadores fuzzy, que apresentam dificuldades em lidar com dados desbalanceados ou com magnitudes discrepantes (Alexandropoulos et al., 2019).

Para este trabalho, foi utilizada a padronização dos dados por meio do *z-score*. Este método ajusta os valores de forma que tenham uma média igual a 0 e desvio padrão igual a 1. O cálculo do *z-score*, conforme destacado na literatura, mede exatamente quantos desvios padrão um dado se encontra acima ou abaixo da média (Elen and Avuçlu, 2021). A fórmula para a padronização

é dada por:

$$z = \frac{x - \mu}{\sigma} \quad (2.1)$$

Onde  $x$  é valor original do dado  $\mu$  média da variável e  $\sigma$  desvio padrão da variável. A aplicação do  $z$ -score é especialmente importante neste estudo, já que promove uma escala uniforme entre os dados provenientes de diferentes séries temporais. Isso não apenas promove maior estabilidade nos cálculos internos dos modelos, mas também melhora a eficácia de algoritmos sensíveis à escala, como o K-Means.

## 2.1.4 Equalização da Frequência das Séries

Os dados utilizados para essa pesquisa apresentam frequências diferentes e como estamos lidando com a identificação de respostas simultâneas em duas series diferentes, é necessário que esses dados estejam na mesma frequência. Para alcançar essa sincronização, duas abordagens foram empregadas neste estudo: a redução da frequência da série de maior amostragem e a interpolação da série de menor amostragem.

### 2.1.4.1 Redução da Frequência da Série de Maior Amostragem

Esta abordagem visa igualar a frequência das séries temporais, reduzindo a amostragem da série de maior frequência ( $H$ ) ao mesmo número de amostras da série de menor frequência ( $L$ ). O método utilizado seleciona amostras de maneira uniforme ao longo da série, preservando as principais características e o comportamento geral dos dados originais.

O processo é detalhado no Algoritmo 2.1, onde a série  $H$  é ajustada calculando-se índices proporcionalmente espaçados, que correspondem ao número de amostras de  $L$ . Para determinar esses índices é gerada uma sequência de  $N_L$  valores igualmente espaçados no intervalo entre 0 e  $N_H - 1$ . Esses valores são então arredondados para os índices inteiros mais próximos, com isso, obtemos uma lista de  $N_L$  índices distribuídos uniformemente na série  $H$ . A partir desses índices, as amostras correspondentes são selecionadas, formando a série ajustada  $\mathcal{H}_{eq}$ .

---

#### Algoritmo 2.1 Equalização de Frequências pela Remoção de Amostras

---

**Função** *EqualizarFrequencias*( $\mathcal{H}, \mathcal{L}$ )

**Entrada:**  $\mathcal{H}$ : Série de alta frequência,  $\mathcal{L}$ : Série de baixa frequência

**Saída:**  $\mathcal{H}_{eq}$ : Versão de  $\mathcal{H}$  ajustada ao tamanho de  $\mathcal{L}$

$N_H \leftarrow |\mathcal{H}|$

$N_L \leftarrow |\mathcal{L}|$

$indices \leftarrow [\text{arredondar}(i) \mid i \in \text{sequencia}(0, N_H - 1, N_L)]$

$\mathcal{H}_{eq} \leftarrow \mathcal{H}[indices]$

---

Essa abordagem apresenta a vantagem de preservar exclusivamente os dados originais, como destacado por Golzad et al. (2024), que identificou em seu trabalho que a maioria dos estudos

envolvendo séries temporais utilizam dados originais, sem a introdução de dados sintéticos. Contudo, ela possui limitações, uma vez que a redução da frequência pode resultar na perda de informações importantes, o que pode comprometer a precisão de análises futuras.

#### 2.1.4.2 Interpolação Linear

A interpolação linear é uma técnica amplamente utilizada para preencher valores ausentes ou redistribuir amostras em séries temporais com base em uma aproximação linear entre dois pontos consecutivos conhecidos. Esse método é particularmente útil quando os dados variam de forma aproximadamente linear em pequenos intervalos.

O funcionamento da interpolação linear pode ser descrito da seguinte maneira: dado dois pontos consecutivos  $(x_1, y_1)$  e  $(x_2, y_2)$  em um conjunto de dados, a interpolação linear calcula o valor  $y$  para o novo ponto  $x$  no intervalo  $[x_1, x_2]$  usando a Equação 2.2:

$$y = y_1 + (x - x_1) \cdot \frac{y_2 - y_1}{x_2 - x_1} \quad (2.2)$$

Esse método gera novos pontos proporcionalmente à distância relativa entre  $x$  e os valores conhecidos  $x_1$  e  $x_2$ , promovendo uma transição suave entre os dados originais. No caso em estudo, onde a série de menor frequência possui aproximadamente metade do tamanho da série de maior frequência, a interpolação linear adicionará um ponto interpolado entre cada par de dados consecutivos da série menor, ajustando-a para uma frequência equivalente à série de maior amostragem.

Embora simples, a interpolação linear é robusta e eficiente para diversos cenários, sendo amplamente utilizada em aplicações práticas de séries temporais. Por exemplo, a interpolação tem sido empregada como uma solução eficaz para expandir pequenos conjuntos de dados, permitindo análises mais robustas com métodos baseados em aprendizado de máquina como mostrado por [Bu et al. \(2023\)](#). No entanto, é importante observar que ela pode introduzir imprecisões em casos onde os dados apresentam variações não lineares significativas.

#### 2.1.5 Suavização

A suavização de séries temporais é uma técnica essencial para reduzir a interferência de ruídos, permitindo uma melhor identificação de padrões e tendências subjacentes nos dados. Séries temporais frequentemente apresentam flutuações de alta frequência, que podem mascarar informações relevantes ou introduzir ruído em análises subsequentes. Métodos de suavização, como o uso de filtros passa-baixa, desempenham um papel crucial na remoção de componentes de alta frequência, preservando os elementos mais importantes da série.

### 2.1.5.1 Filtro passa-baixa

Os filtros passa baixa são projetados para atenuar frequências altas e permitir que apenas frequências abaixo de uma frequência de corte especificada sejam preservadas. Em um filtro ideal, todas as frequências na banda de passagem seriam transmitidas completamente, enquanto todas as frequências na banda de rejeição seriam completamente eliminadas, com uma transição abrupta entre essas bandas. Entretanto, como filtros ideais não existem, a aplicação prática de filtros passa-baixa requer um equilíbrio: minimizar o ruído sem perder informações significativas do sinal original (Elisa Coraggio, 2023).

Entre os diversos tipos de filtros passa-baixa, o filtro Butterworth é amplamente utilizado devido à sua resposta de frequência na banda de passagem ser maximamente plana, sem oscilações. Este filtro foi introduzido por Butterworth (1930) é projetado para ter uma resposta de magnitude que seja completamente plana na banda de passagem e diminua gradualmente na banda de rejeição, sem oscilações. A função de transferência é dada pela Equação 2.3 onde  $w$  é a frequência angular do sinal e  $w_e$  é a frequência de corte,  $n$  é a ordem do filtro (quanto maior, mais acentuada é a transição entre bandas).

$$|H(w)| = \frac{1}{\sqrt{1 + \left(\frac{w}{w_e}\right)^{2n}}} \quad (2.3)$$

Butterworth apresenta uma transição mais suave ao redor da frequência de corte, com um menor declínio na transmissão. Essa característica o torna uma escolha ideal em aplicações que exigem preservação máxima do sinal útil e redução de ruídos, como processamento de áudio, sistemas de radar e redução de *aliasing* (Elisa Coraggio, 2023; Wu et al., 2010).

De maneira prática, o filtro Butterworth pode ser ajustado em termos de ordem, onde ordens mais altas proporcionam uma transição mais acentuada entre as bandas de passagem e rejeição, aproximando-se do filtro ideal. Ainda assim, seu objetivo principal permanece em oferecer uma solução eficaz para reduzir o ruído com o mínimo impacto possível na integridade do sinal original.

## 2.2 Unificação dos Sinais

A redução de dimensionalidade permite que algoritmos de aprendizado de máquina operem de forma mais precisa e eficiente, com menor tempo de execução e menor complexidade computacional. No entanto, é fundamental que a redução de dimensionalidade preserve as informações significativas dos dados originais, garantindo que as características importantes do conjunto de dados sejam mantidas (Ashraf et al., 2023).

O método escolhido para unificar os sinais analisados foi o mínimo absoluto. Este método compara os valores absolutos de dois sinais, preservando o menor valor em cada ponto da série temporal. Ele é especialmente útil em contextos onde há sobreposição de respostas entre os

sinais. Quando os dois sinais apresentam picos simultâneos, o método mantém esses picos, garantindo que as respostas mais significativas sejam preservadas. Por outro lado, em situações onde os picos não são síncronos, o método reduz os valores, minimizando o impacto de discrepâncias entre os sinais.

A unificação por mínimo absoluto pode ser descrita matematicamente da seguinte forma: dados dois sinais  $f(t)$  e  $g(t)$  o sinal unificado  $u(t)$  é calculado como da seguinte forma:

$$u(t) = \begin{cases} f(t), & \text{se } |f(t)| < |g(t)| \\ g(t), & \text{caso contrário.} \end{cases} \quad (2.4)$$

## 2.3 Avaliação com Modelos de Machine Learning

A detecção de anomalias é uma tarefa fundamental amplamente aplicada em diferentes áreas, como processos industriais, análise de séries temporais, sistemas de saúde e segurança. Essa abordagem permite identificar pontos de dados que diferem substancialmente dos padrões normais, indicando possíveis erros, falhas ou eventos excepcionais (Marcelli et al., 2024).

No contexto desta pesquisa, técnicas de detecção de anomalias baseadas em aprendizado de máquina foram utilizadas para identificar as reações das pragas aos compostos, dado que essa reação se manifesta na forma de picos nos dados, auxiliando na análise e identificação das reações. Foram utilizados dois algoritmos, Isolation Forest e K-means, escolhidos por sua simplicidade.

### 2.3.1 Isolation Forest

O algoritmo Isolation Forest (IF), é uma técnica não supervisionada para detecção de anomalias. Ele se baseia na premissa de que as anomalias são mais fáceis de isolar em comparação com os pontos normais devido à sua baixa frequência e distribuição esparsa nos dados (Liu et al., 2008, 2012). Esse método é amplamente utilizado por sua eficiência computacional e capacidade de lidar com grandes volumes de dados.

Inicialmente, ele recebe como entrada um conjunto de dados  $X$ , contendo  $n$  amostras, além de parâmetros como o número de árvores  $T$  a serem construídas, adicionalmente, é fornecida uma estimativa da proporção de anomalias no conjunto de dados, o parâmetro conhecido como contaminação, que ajuda na definição do limiar para a classificação final.

Na etapa de inicialização, o modelo cria uma floresta de isolamento composta por  $T$  árvores independentes. Cada árvore é construída de forma aleatória a partir de uma amostragem do conjunto de dados. Durante a construção das árvores, cada nó realiza uma divisão aleatória do espaço dos atributos. Isso envolve a seleção de um atributo específico e de um valor de corte no intervalo de valores desse atributo. O objetivo é particionar recursivamente o espaço dos dados em subconjuntos menores, até que a profundidade máxima seja alcançada ou que reste apenas

uma amostra no nó terminal.

O isolamento de uma amostra é representado pelo número de divisões necessárias para isolá-la completamente, conhecido como o comprimento do caminho  $h(x)$ . Anomalias, por definição, tendem a estar em regiões esparsas dos dados e, portanto, são isoladas mais rapidamente, resultando em comprimentos de caminho menores. Em contrapartida, pontos normais requerem mais divisões devido à sua localização em áreas mais densas do espaço de atributos.

Para cada ponto no conjunto de dados, o algoritmo calcula o comprimento médio do caminho percorrido ao longo de todas as árvores na floresta. Esse valor médio, denotado por  $\mathbb{E}[h(x)]$ , é utilizado para calcular uma pontuação de anomalia  $s(x)$ .

A pontuação é obtida pela fórmula 2.5 onde  $c(n)$  é o valor esperado do comprimento do caminho para uma amostra de tamanho  $n$ . Essa normalização transforma o comprimento do caminho em uma pontuação que varia entre 0 e 1. Pontuações próximas a 1 indicam alta probabilidade de a amostra ser uma anomalia, enquanto pontuações próximas a 0 indicam que a amostra é normal.

$$s(x) = 2 - \frac{\mathbb{E}[h(x)]}{c(n)} \quad (2.5)$$

Por fim, com base no parâmetro contaminação, o algoritmo define um limiar para classificar as amostras. Pontos com pontuações  $s(x)$  superiores ao limiar são rotulados como anômalos, enquanto os demais são considerados normais. Podemos ver esse processo de forma simplificada no Algoritmo 2.3.1.

Com sua simplicidade, destaca-se por sua eficiência, tendo o poder de lidar com alta dimensionalidade e grandes volumes de dados. Sua robustez e desempenho o tornam uma ferramenta amplamente utilizada em aplicações de detecção de anomalias em diversos domínios, sendo os principais diagnóstico de falhas, detecção de fraudes e detecção de intrusos (Al Farizi et al., 2021).

---

**Algoritmo 2.2** Detecção de anomalias com Isolation Forest

---

**Entrada:**  $X$ : Conjunto de dados,  $T$ : Número de árvores, Contaminação: Taxa de anomalias

**Saída:** Pontuação de anomalias  $s(x)$

Inicialize a floresta com  $T$  árvores

**para**  $i \leftarrow 1$  **em**  $T$  **faça**

    Construa a árvore  $t_i$   
    Compute  $h_{t_i}(x)$  para todos os pontos  $x \in X$

Compute  $\mathbb{E}[h(x)]$  e  $s(x)$  para cada  $x \in X$

Classifique os pontos com  $s(x)$  acima de um limiar como anomalias.

---

### 2.3.2 K-means

O K-Means é um algoritmo de agrupamento que particiona os dados em  $k$  grupos, ou *clusters*, com base na proximidade dos pontos em um espaço vetorial. O objetivo principal é minimizar a soma das distâncias  $d(x, c_i)$  de cada ponto  $x$  ao centroide  $c_i$  do grupo ao qual pertence (Hartigan and Wong, 1979). Essa distância é geralmente medida pela distância euclidiana (Equação 2.6).

$$d(x, c_i) = \sqrt{\sum_{j=1}^n (x_j - c_{ij})^2} \quad (2.6)$$

Os centroides representam o ponto médio de cada grupo. A distância entre um ponto  $x$  e um centroide  $c_i$  é tipicamente medida pela distância euclidiana, dada pela fórmula 2.6 onde  $x = (x_1, x_2, \dots, x_n)$  é um ponto nos dados com  $n$  dimensões,  $c_i = (c_{i1}, c_{i2}, \dots, c_{in})$  é o centroide do *cluster*  $i$  nas mesmas  $n$  dimensões.

---

#### Algoritmo 2.3 K-Means

**Entrada:**  $X$ : Conjunto de dados,  $k$ : Número de clusters,  $max_{iter}$ : Número máximo de iterações

**Saída:**  $C$ : Centros dos clusters,  $L$ : Rótulos dos clusters

Inicialize  $C = \{c_1, c_2, c_3, \dots, c_k\}$  com  $k$  centróides

**para**  $i \leftarrow 1$  **até**  $max_{iter}$  **faça**

Inicialize rótulo  $L = []$

**para cada**  $x \in X$  **faça**

Compute  $d(x, c_i)$  para todos os  $c_i \in C$

Atribua  $x$  ao cluster  $c_i$  com menor  $d(x, c_i)$

Adicione o rótulo correspondente a  $L$

Atualize cada  $c_i \in C$  como a média dos pontos atribuídos a  $c_i$

**se os centróides**  $C$  **não mudarem significativamente** **então**

└ **break**

**retorna**  $C, L$ ;

---

#### 2.3.2.1 Detecção de Anomalias com K-Means

O algoritmo K-Means pode ser aplicado no contexto da detecção de anomalias, conforme demonstrado em estudos como Kumari et al. (2016) e Princz et al. (2024). Nesta pesquisa, utilizamos o K-Means para identificar padrões "normais" nos dados, classificando como possíveis anomalias os pontos que se encontram distantes dos centroides. A identificação é realizada com base em um limiar definido pela Equação 2.7, onde  $\mu$  representa a média do sinal,  $\sigma$  o desvio padrão dos dados, e  $t$  é um multiplicador ajustável.

O cálculo do limiar fundamenta-se em técnicas estatísticas clássicas, amplamente utilizadas em controle de qualidade de processos e análise de dados (Chandola et al., 2009). Foi testado também a definição do limiar usando  $1.5 \cdot IQR$  sendo  $IQR$  o intervalo interquartil, porém, essa abordagem se mostrou muito sensível identificando boa parte dos dados como acima do limiar.

Neste estudo, adotaremos duas abordagens para a aplicação do modelo: o agrupamento direto, considerando o conjunto completo de dados, e o uso de janelas deslizantes, que permite uma análise mais local e sensível a mudanças dinâmicas nos padrões dos dados (Algoritmo 2.3.2.1).

$$\tau = u + t \cdot \text{std}(D) \quad (2.7)$$

---

**Algoritmo 2.4** Detecção de anomalias com K-Means

---

**Entrada:**  $X$ : Dados,  $k$ : Número de grupos,  $t$ : Multiplicador do limiar para anomalias

**Saída:** Índices dos pontos anômalos  $A$

Treine o K-Means com  $k$  grupos usando  $X$ ;

Compute as distâncias  $D$  de cada ponto em  $X$  ao centroide mais próximo

Defina o limiar de anomalia  $\tau = u + t \cdot \text{std}(D)$

Identifique os índices anômalos  $A = \{i \in [1, |X|] : D[i] > \tau\}$

**retorna**  $A$ ;

---

A abordagem com janelas deslizantes busca identificar padrões anômalos em subseções da série temporal. A série é dividida em segmentos (janelas) sendo deslizada sobre os dados. As subseções cuja distância ao centroide mais próximo excede um limiar definido são classificadas como anômalas. Todos os pontos contidos nessas janelas são considerados parte da anomalia.

A segunda abordagem aplica o algoritmo diretamente sobre os valores individuais da série temporal. Cada ponto é tratado como uma observação, e a distância de cada ponto ao centroide mais próximo é calculada. Pontos cuja distância excede o limiar são marcados como anômalos.

# 3

## Resultados e Discussões

Neste capítulo iremos mostrar os resultados obtidos aplicando o *pipeline* demonstrado em 2.1 e detalhado no capítulo anterior.

### 3.1 Remoção de dados não utilizáveis

A remoção dos dados iniciais relacionados ao solvente foi realizada para eliminar informações irrelevantes como podemos notar na Figura 3.1. Apesar das diferenças de duração entre os experimentos, a uniformização promovida por essa remoção resultou em maior consistência e relevância nos sinais analisados.

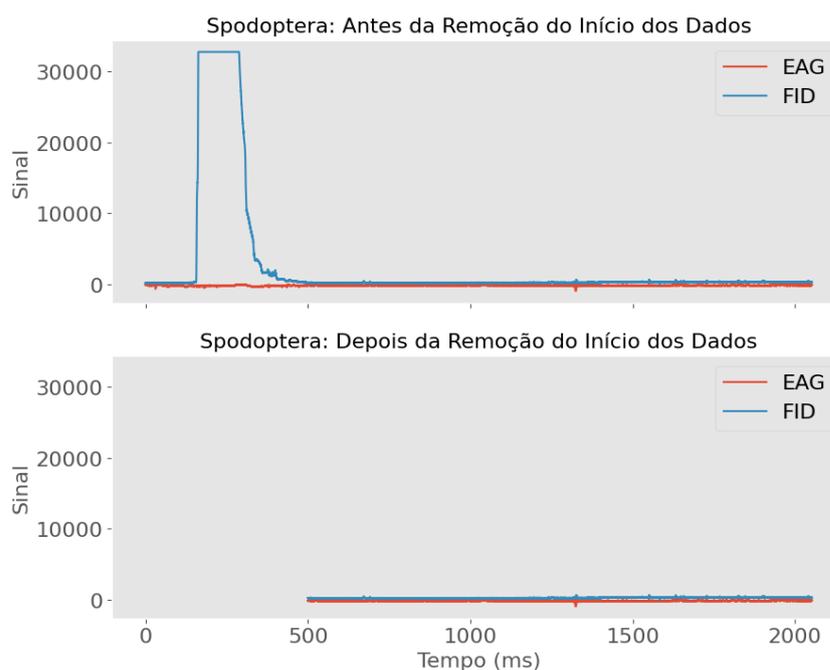


Figura 3.1: Exemplo de remoção do início dos dados.

Na Figura 3.2, são apresentados os dados após a remoção dos trechos iniciais. Para facilitar a interpretação, o eixo y não foi fixado, permitindo uma visualização mais clara tanto dos picos relacionados às reações dos insetos quanto dos picos associados à resposta do GC-FID na identificação de substâncias presentes nos compostos.

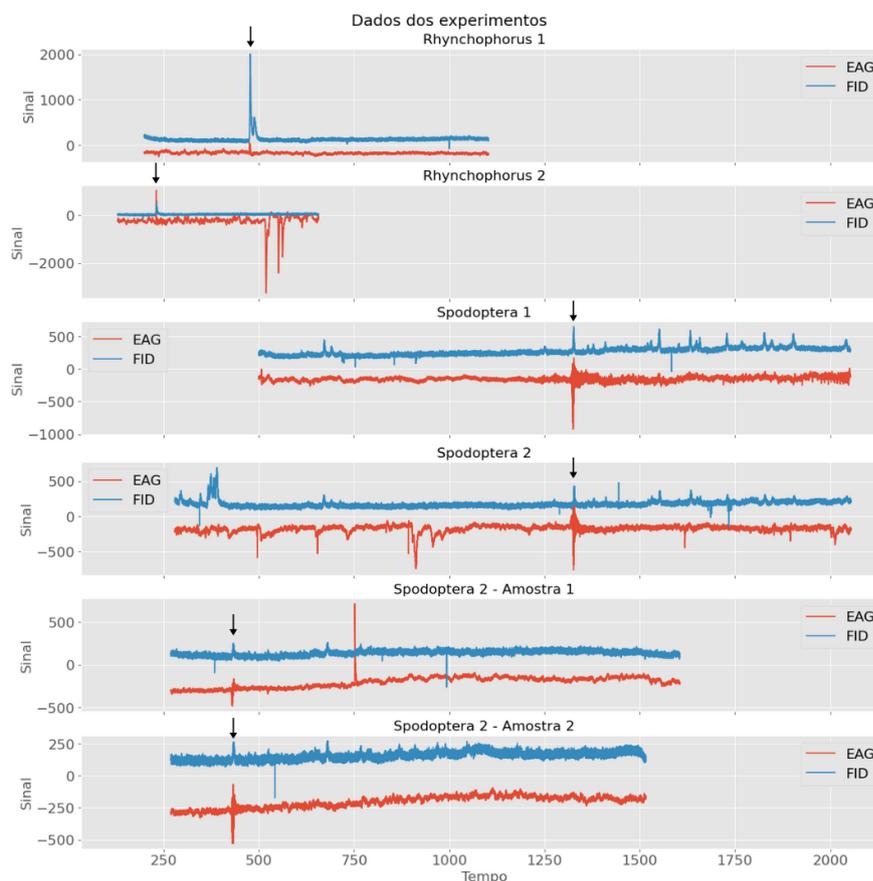


Figura 3.2: Todos experimentos após remoção de dados do hexano, com marcação nas reações.

## 3.2 Resultado da padronização dos dados

A padronização dos dados foi realizada para normalizar as séries temporais, ajustando-as a uma escala comum com média zero e desvio padrão unitário. Essa etapa é essencial para eliminar diferenças de magnitude entre as variáveis, melhorando a performance de algoritmos sensíveis à escala, como o K-Means.

Nas Figuras 3.3 e 3.4, observa-se a aplicação da padronização em duas séries da segunda amostra da espécie *Rhynchophorus*. Essa transformação reduz a escala dos sinais originais, preservando as características essenciais, como os picos associados às respostas dos insetos, enquanto atenua o impacto de flutuações extremas e facilita a comparação entre séries distintas.

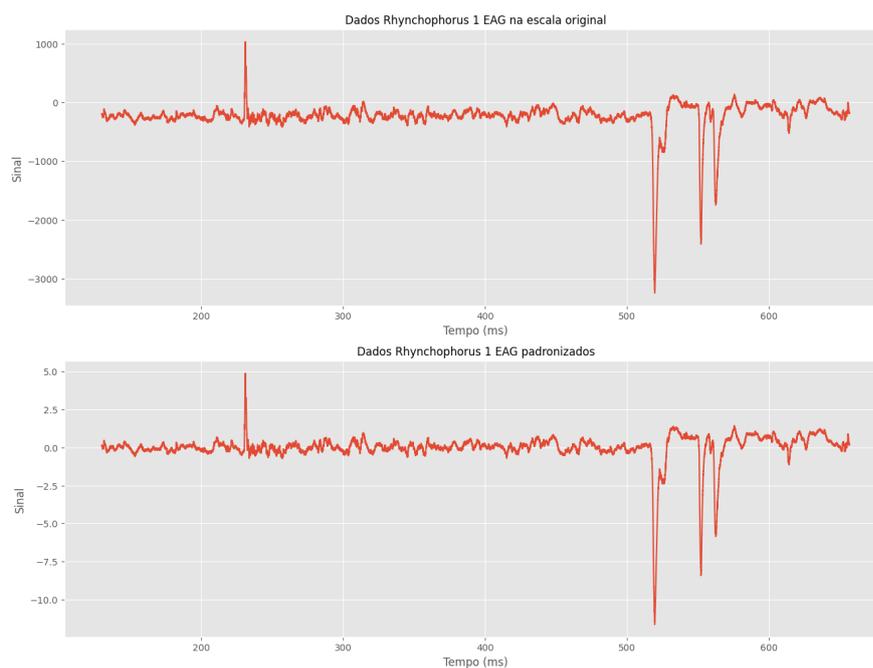


Figura 3.3: Sinal padronizados - EAG.

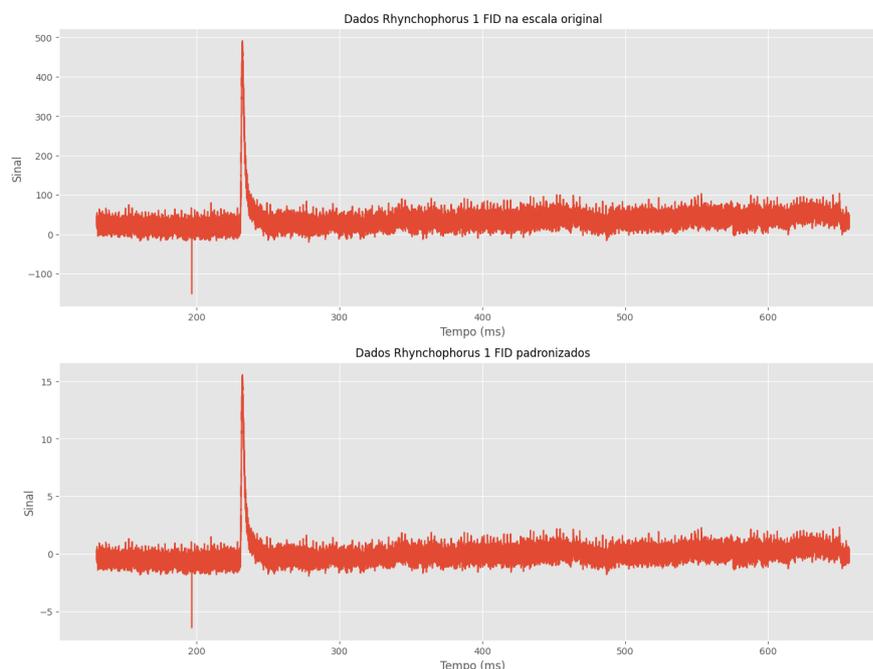


Figura 3.4: Sinal padronizados - FID.

Nas séries de dados da espécie *Rhynchophorus* apresentadas, é possível observar como os sinais originais, com amplitudes variáveis, foram transformados em uma escala uniforme. A padronização não altera a forma geral dos sinais, mantendo as características essenciais, como os picos relacionados às respostas dos insetos, enquanto reduz o impacto de flutuações extremas e facilita a análise comparativa entre diferentes séries. Esse comportamento se reflete em todos

os demais dados dos experimentos.

### **3.3 Equalização de frequências**

Os dados provenientes do FID têm uma frequência de 201Hz enquanto EAG possui uma frequência de 96Hz, logo, se faz necessário equalizar as frequências para manter a consistência dos dados, para tal usamos duas abordagens, descritas na sessão [2.1.4.1](#) e [2.1.4.1](#).

#### **3.3.1 Equalização por redução**

Podemos observar na Figura [3.5](#) que a seleção dos dados foi feita de forma uniforme, garantindo a preservação da estrutura temporal do sinal e evitando alterações significativas no padrão ou na dinâmica dos dados. O eixo  $x$  representa a quantidade de leituras realizadas durante o experimento, e o eixo  $y$  exibe os valores registrados em cada leitura, refletindo a resposta do sistema ao estímulo. Esse processo de seleção uniforme é crucial, pois mantém a representatividade dos dados ao longo do tempo, sem introduzir distorções que poderiam comprometer a análise posterior.

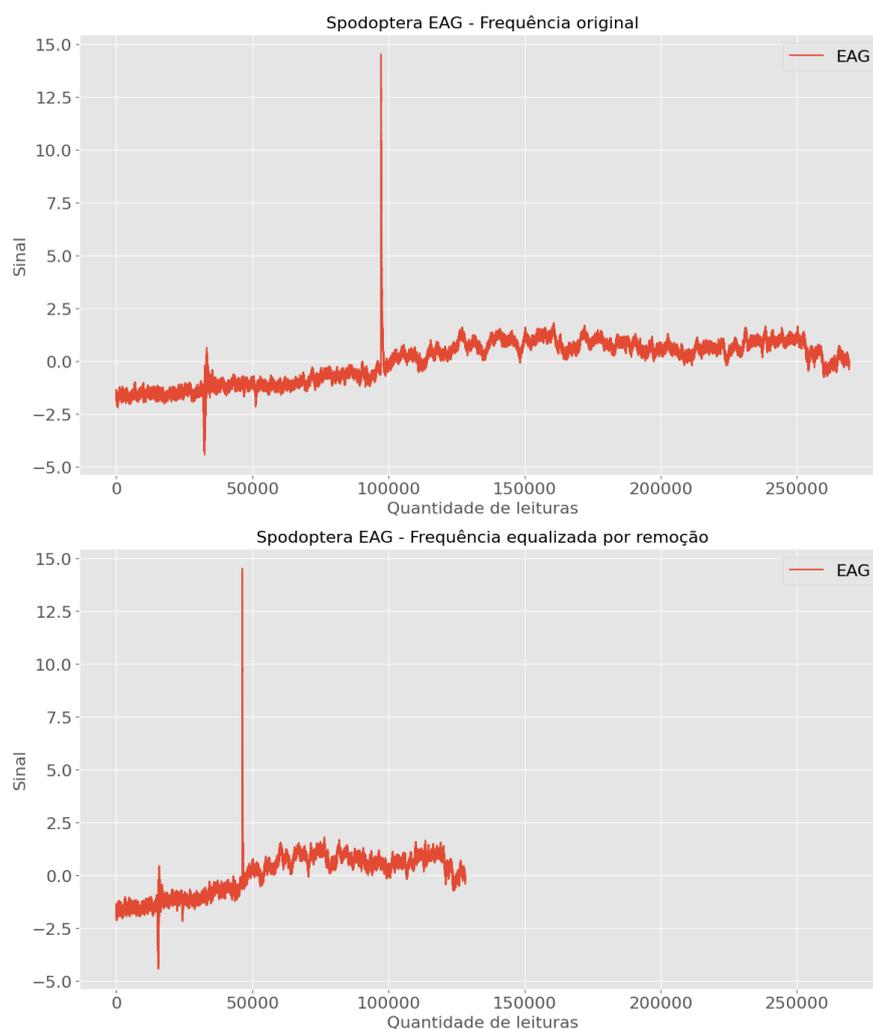


Figura 3.5: Exemplo de equalização por redução.

No gráfico superior (frequência original), o sinal exibe a densidade de amostras original. Já no gráfico inferior (após a equalização), a densidade de amostras foi reduzida para igualar a frequência de amostragem a 96 Hz. Esse processo foi realizado sem a inserção de dados adicionais, assegurando que ambos os sinais sejam compostos exclusivamente por dados reais. Essa consistência também aconteceu nos demais experimentos.

### 3.3.2 Equalização por interpolação

A figura 3.6 apresentada ilustra o processo de equalização por interpolação aplicado aos sinais de FID da espécie *Spodoptera*, comparando a frequência original com a frequência aumentada.

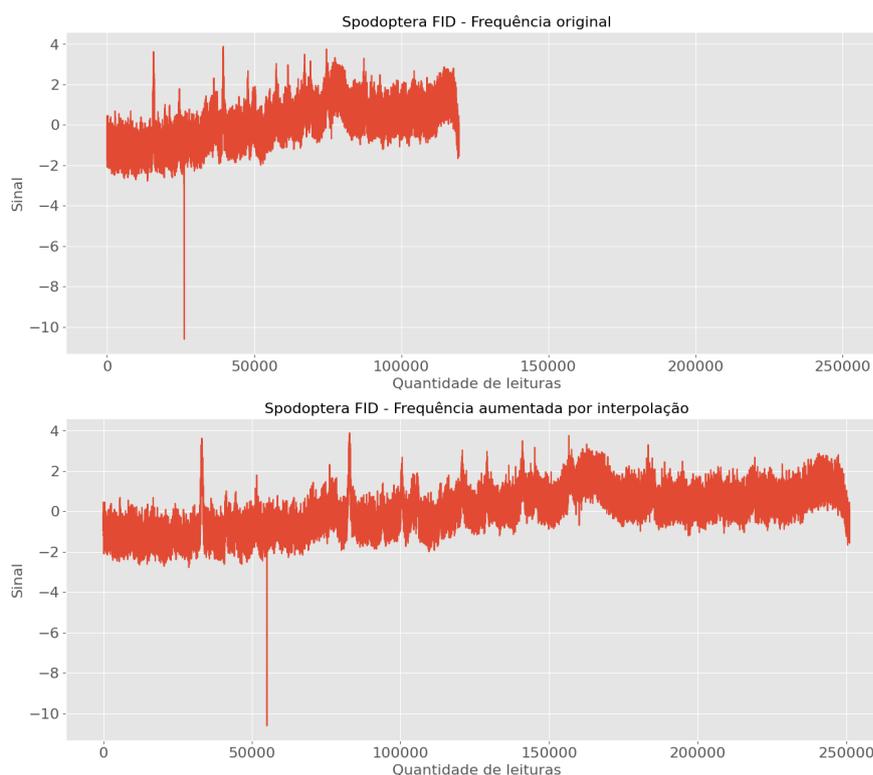
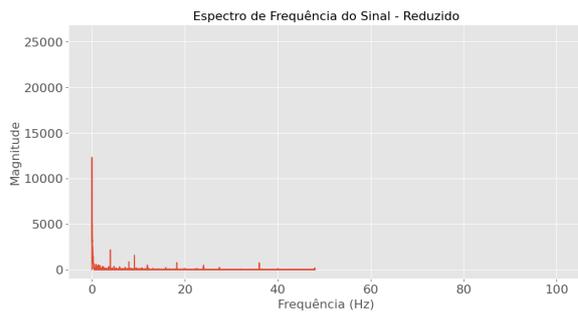


Figura 3.6: Equalização por interpolação dos sinais de FID.

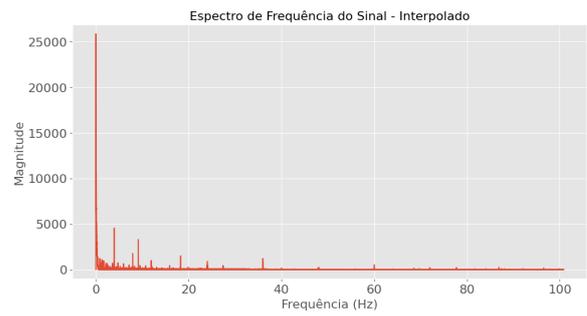
A interpolação foi utilizada para aumentar a densidade de amostragem da série original, permitindo compatibilidade com séries de maior frequência. Esse ajuste é visível pela maior uniformidade na distribuição dos dados na série interpolada, preservando os picos característicos relacionados às reações semioquímicas. A técnica foi essencial para alinhar as séries temporais de diferentes frequências, garantindo consistência analítica e favorecendo o desempenho dos modelos de aprendizado de máquina aplicados nos estágios subsequentes.

### 3.4 Suavização com filtro passa-baixa

As figuras 3.7, 3.8 e 3.9 apresentam a aplicação do filtro passa-baixa de Butterworth visa eliminar ruídos de alta frequência presentes nos sinais, destacando os padrões relevantes relacionados às respostas biológicas. Essa etapa foi aplicada aos sinais reduzidos e interpolados para garantir maior consistência e clareza nos dados.

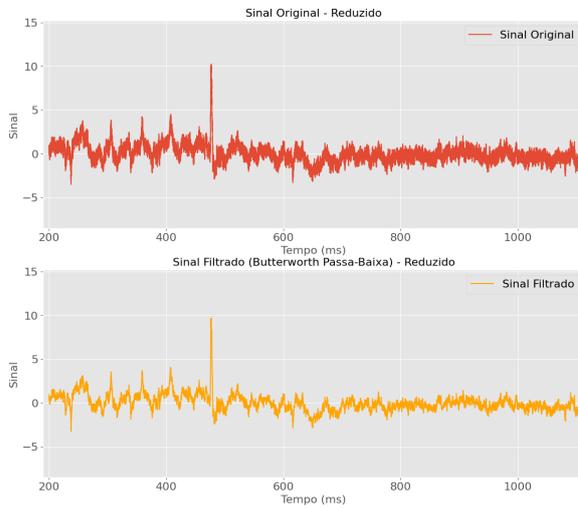


(a) Suavização da série equalizada por redução.

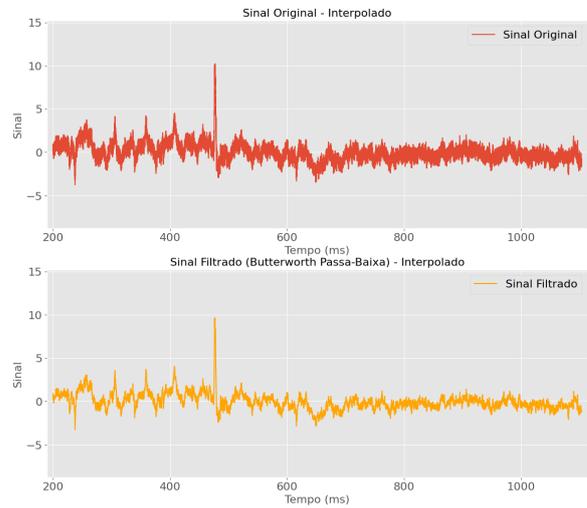


(b) Espectro de Frequência da série equalizada por interpolação.

Figura 3.7: Espectro de Frequência da série

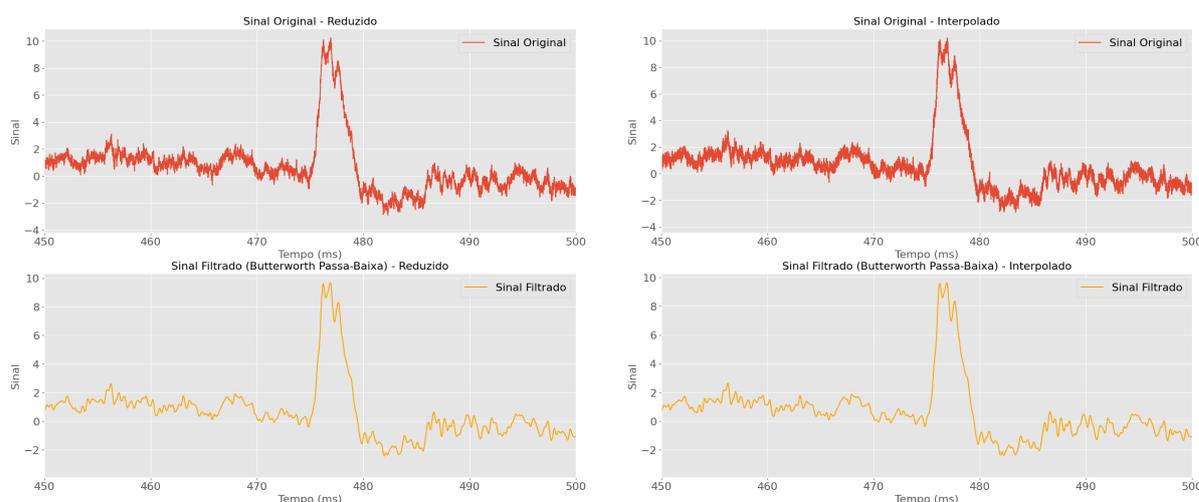


(a) Suavização da série equalizada por redução.



(b) Suavização da série equalizada por interpolação.

Figura 3.8: Suavização da série equalizada



(a) Ampliação da suavização da série equalizada por redução. (b) Ampliação suavização da série equalizada por interpolação.

Figura 3.9: Ampliação da suavização da série equalizada

O espectro de frequência dos sinais apresentados na Figura 3.7 para as séries reduzidas e interpoladas mostram padrões bastante semelhantes. A maior parte da energia dos sinais está concentrada em baixas frequências, enquanto as componentes de alta frequência, associadas a ruídos indesejáveis, são muito mais discretas. Com base nesse padrão, foi definido um valor de frequência de corte de 3 Hz, que foi utilizado no filtro passa-baixa de Butterworth para eliminar essas flutuações de alta frequência.

Tanto na redução quanto na interpolação, a aplicação do filtro passa-baixa de Butterworth demonstrou eficácia em suavizar os sinais, eliminando flutuações indesejadas e preservando os picos principais, essenciais para a identificação das respostas aos compostos semioquímicos.

No domínio do tempo, observa-se que o filtro atua de forma consistente em ambas as abordagens, reduzindo o impacto do ruído sem comprometer as características fundamentais do sinal como vemos na Figura 3.9.

Dessa forma, ambas as abordagens prepararam o sinal de forma adequada para as etapas de aprendizado de máquina, garantindo maior robustez na detecção de padrões relevantes. A escolha entre redução e interpolação pode ser guiada pela necessidade específica de compatibilidade entre séries de diferentes frequências ou pela simplicidade computacional desejada.

### 3.5 Unificação de sinais

As figuras 3.10 e 3.11 ilustram o processo de unificação dos sinais de EAG e FID para os experimentos realizados, apresentando os sinais originais e o sinal resultante da unificação.

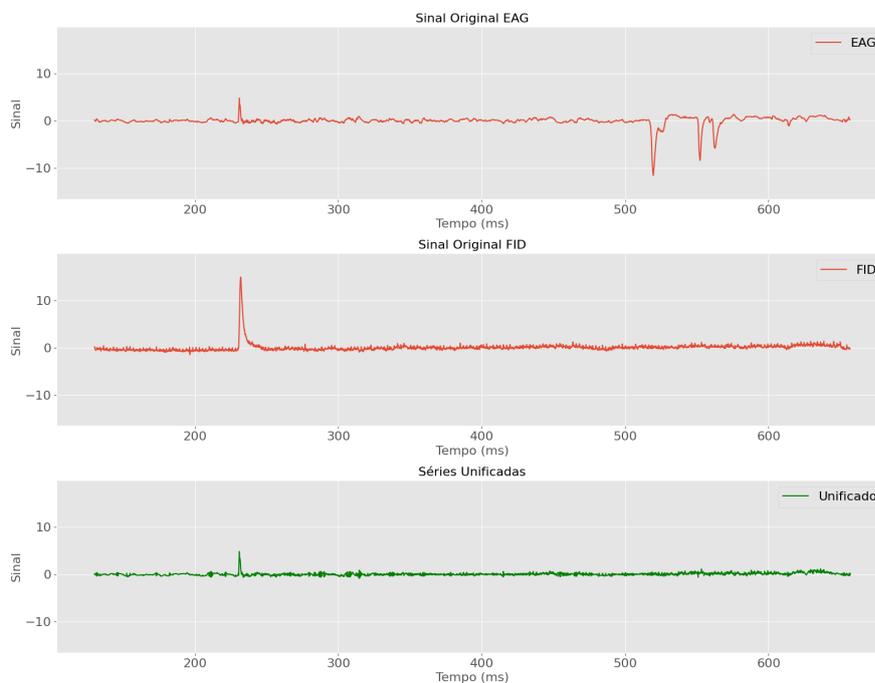


Figura 3.10: Unificação da série equalizada por redução.

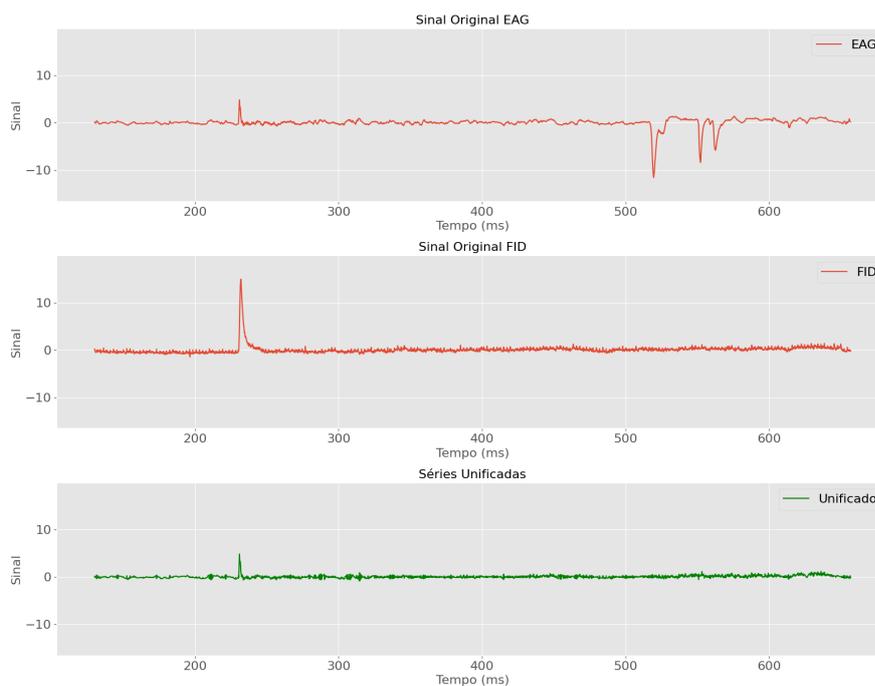


Figura 3.11: Unificação da série equalizada por interpolação.

A etapa de unificação combina as informações dos dois sinais em uma única série temporal, utilizando o valor mínimo absoluto entre as séries para preservar os picos simultâneos e eliminar discrepâncias. Ao aplicar essa abordagem podemos notar que reações correlacionadas, representadas por picos em ambas as séries, são destacadas, enquanto picos isolados, considerados ruídos, sejam suavizados ou removidos.

Comparando os resultados das séries equalizadas por redução (Figura 3.10) e por interpolação (Figura 3.11), percebemos que os sinais unificados apresentam alta similaridade, independentemente da técnica de equalização utilizada. Isso demonstra que ambas as abordagens são eficazes para preservar os padrões relevantes dos sinais originais durante o processo de unificação.

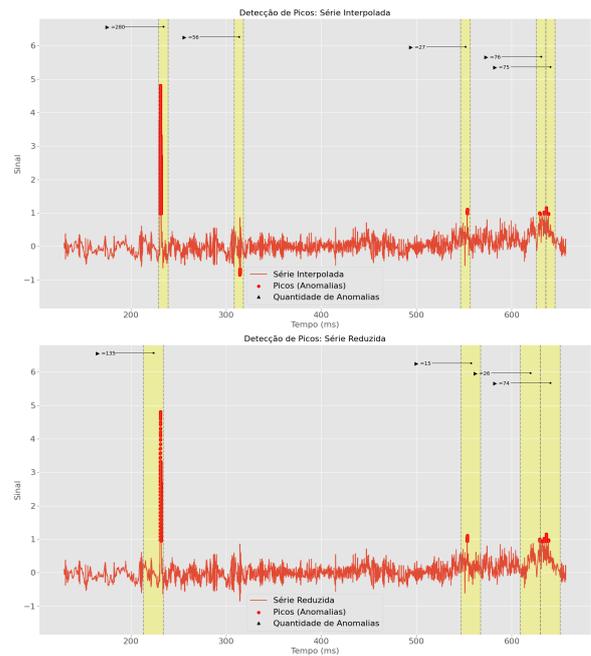
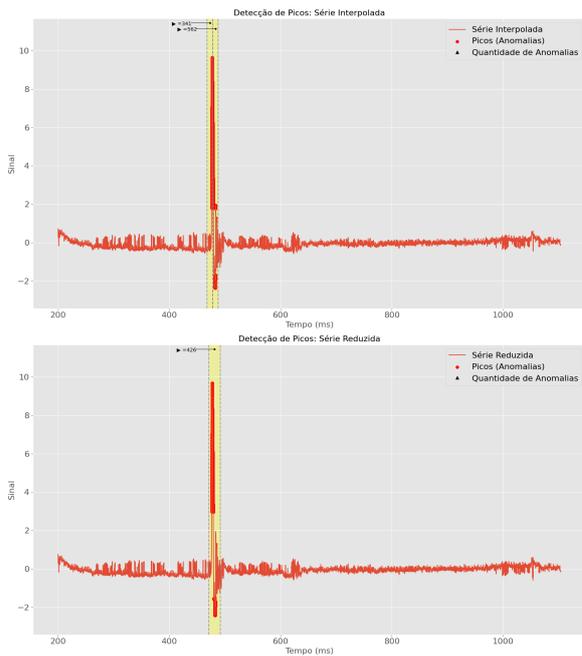
Nos demais experimentos, resultados consistentes foram observados. Em ambas as abordagens, a unificação destacou os picos simultâneos, assegurando que as informações mais relevantes de cada série fossem integradas sem perda significativa de dados importantes.

## 3.6 Resultado dos modelos

Os modelos de aprendizado de máquina foram aplicados para detecção de picos nas séries temporais, correspondendo a reações relevantes dos insetos aos compostos semioquímicos.

### 3.6.1 Isolation forest

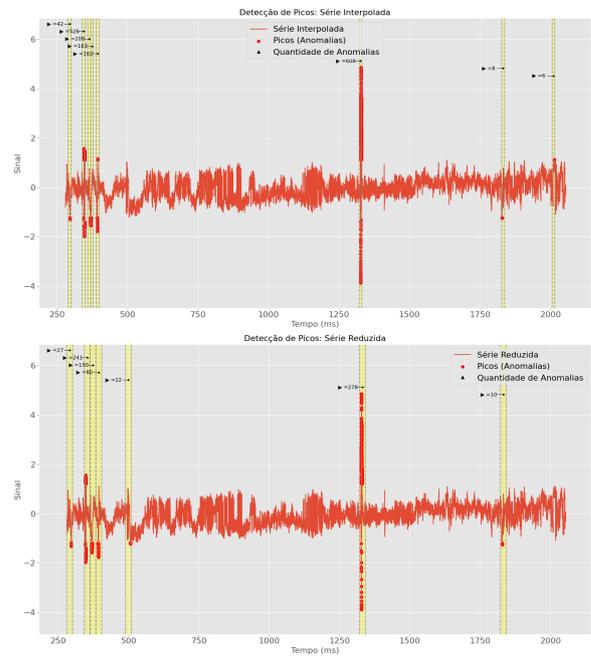
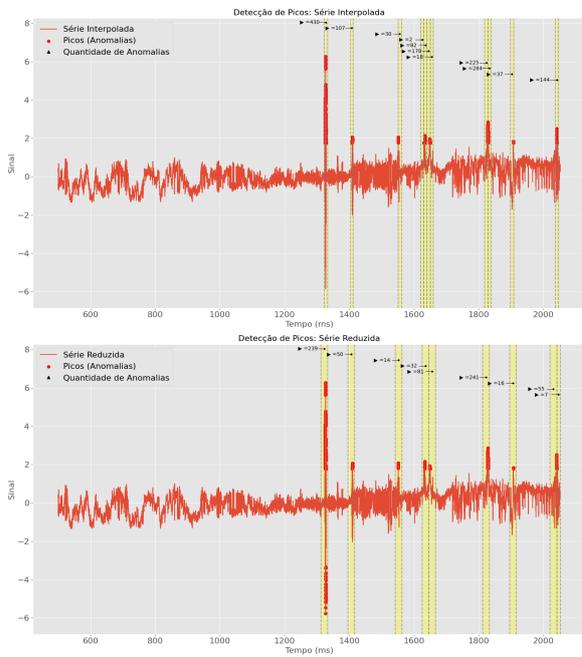
Os resultados da aplicação do Isolation Forest apresentados nas Figuras 3.12, 3.13 e 3.14, destacaram por sua eficiência na identificação de anomalias, permitindo isolar padrões significativos em meio ao ruído. O parâmetro de contaminação foi configurado como 0.01, uma vez que este define a porcentagem esperada de anomalias nos dados. Como as reações biológicas representam uma parcela muito pequena das séries temporais analisadas, um valor baixo foi necessário para garantir que o modelo focasse nos eventos raros e potencialmente significativos, minimizando a inclusão de ruídos ou eventos irrelevantes. O modelo foi configurado com 100 árvores. Este número foi escolhido para equilibrar o custo computacional e a robustez do modelo.



(a) Aplicação do Isolation Forest - Rhynchophorus experimento 1.

(b) Aplicação do Isolation Forest - Rhynchophorus experimento 2.

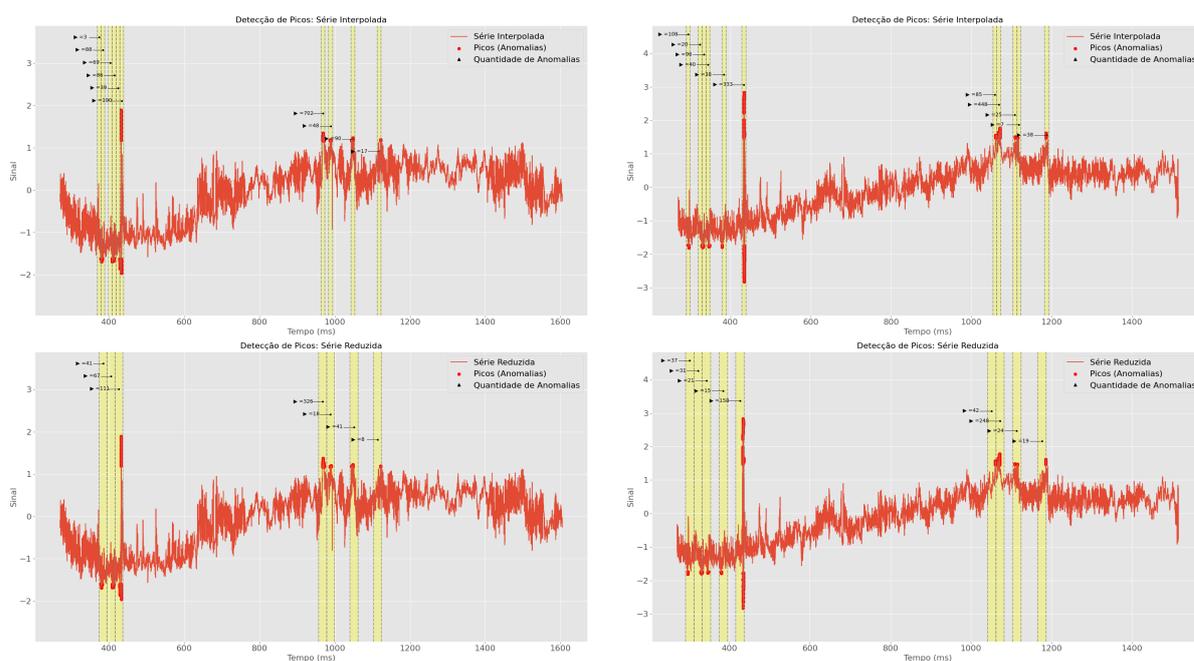
Figura 3.12: Aplicação Isolation Forest - Rhynchophorus



(a) Aplicação do Isolation Forest - Spodoptera experimento 1.

(b) Aplicação do Isolation Forest - Spodoptera experimento 2.

Figura 3.13: Aplicação do Isolation Forest - Spodoptera.



(a) Aplicação do Isolation Forest - *Spodoptera* amostra 1. (b) Aplicação do Isolation Forest - *Spodoptera* amostra 2.

Figura 3.14: Aplicação do Isolation Forest - *Spodoptera* amostra 2.

As séries foram divididas em janelas com 2000 dados que foi definida analisando os dados e observando que essa quantidade de dados consegue abranger todos os picos de resposta nos diferentes experimentos. Foi calculado a quantidade de pontos dentro dessas janelas que foram considerados anomalias pelo modelo.

Os resultados obtidos demonstram que o modelo foi capaz de identificar regiões de interesse com uma quantidade maior de anomalias, especialmente em áreas de maior variação do sinal, como nos picos característicos gerados por estímulos químicos específicos contendo a maior quantidade de pontos classificados como anômalos.

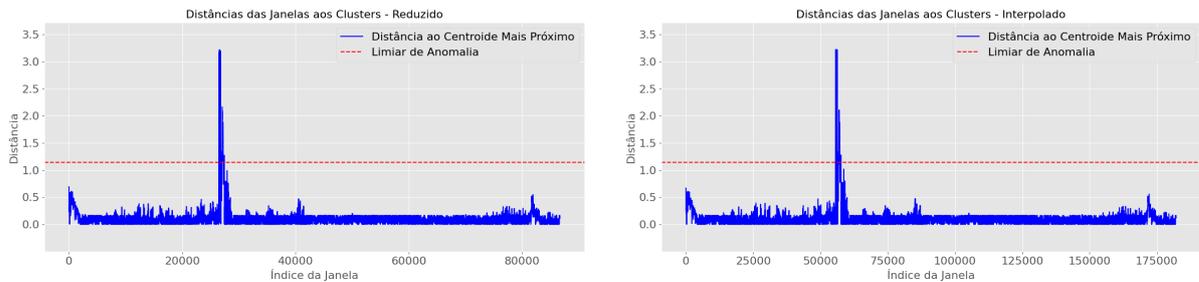
Essa identificação facilita a correlação entre as respostas fisiológicas dos insetos e os compostos analisados, facilitando a identificação do momento em que uma possível reação aconteceu, contribuindo para a interpretação biológica e química dos experimentos.

### 3.6.2 K-means utilizando valores individuais

O algoritmo K-means foi aplicado diretamente aos dados, sem o uso de janelas deslizantes, para identificar anomalias nas séries de *Rhynchophorus* e *Spodoptera*. O parâmetro  $k = 3$ , foi selecionado com base em testes realizados com diferentes valores de  $k$ , sendo este o que apresentou os melhores resultados. Esta escolha também reflete a hipótese de que os dados podem ser divididos em três grupos: ruído, picos menores e intermediários, enquanto os grandes picos, que são nosso foco, seriam identificados como anomalias.

Para detecção de anomalias, foi utilizado o limiar definido na Equação 2.7 usando  $t = 6$ , esse valor foi definido após experimentação de vários valores.

Podemos visualizar esse limiar nas Figuras 3.15, 3.17, 3.19, 3.21, 3.23, e 3.25 onde são exibidas as distâncias de cada ponto ao centroide do grupo mais próximo, considerando como anômalos os pontos que excedem esse limiar.



(a) Distância do centroide dados equalizados por redução - *Rhynchophorus* experimento 1.

(b) Distância do centroide dados equalizados por interpolação - *Rhynchophorus* experimento 1.

Figura 3.15: Distância do centroide *Rhynchophorus* experimento 1.

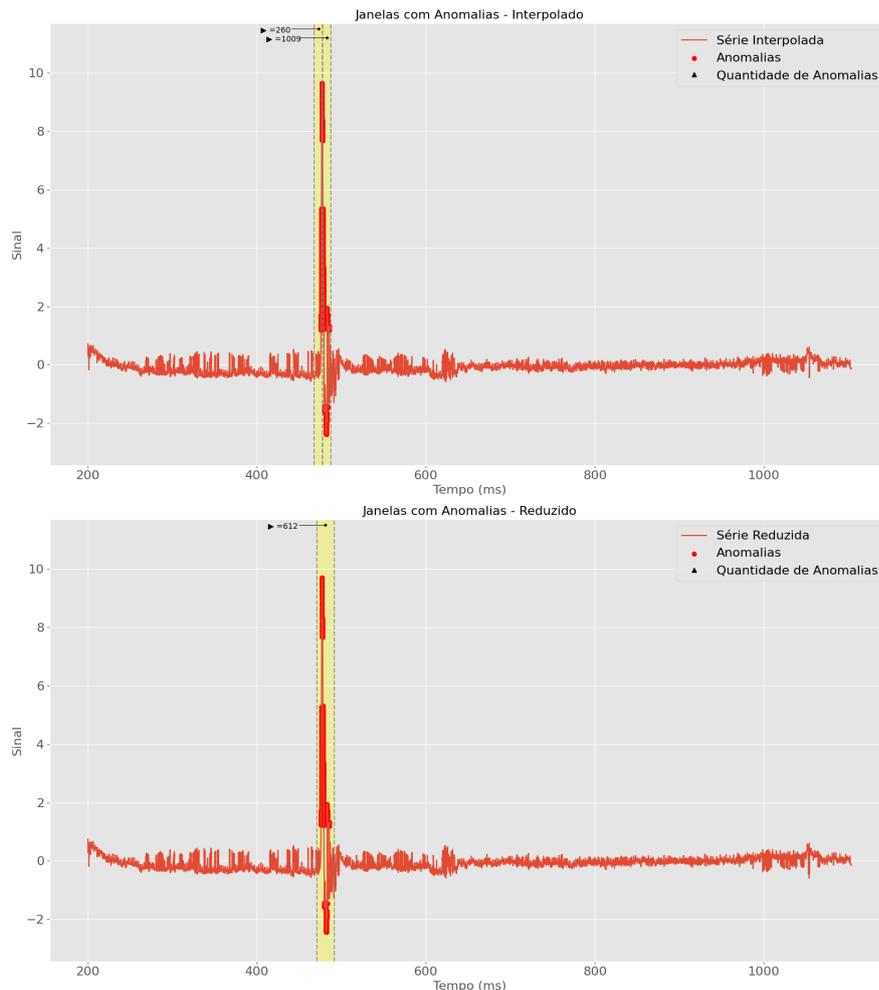
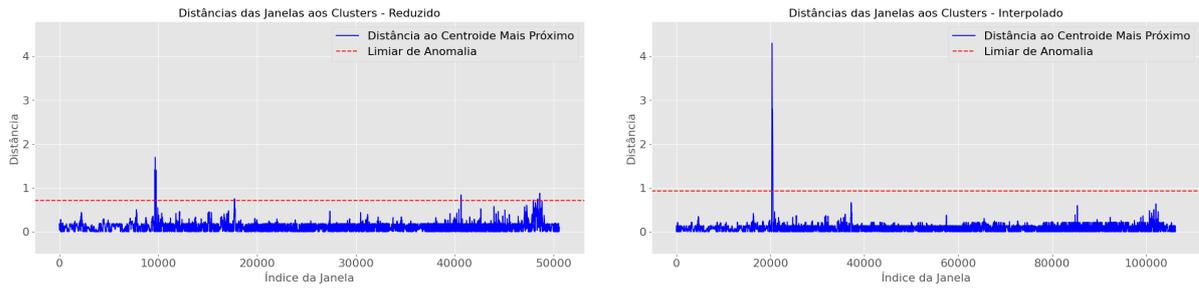


Figura 3.16: Aplicação K-means agrupamento direto - *Rhynchophorus* experimento 1.



(a) Distância do centroide dados equalizados por redução - *Rhynchophorus* experimento 1.

(b) Distância do centroide dados equalizados por interpolação - *Rhynchophorus* experimento 1.

Figura 3.17: Distância do centroide *Rhynchophorus* experimento 1.

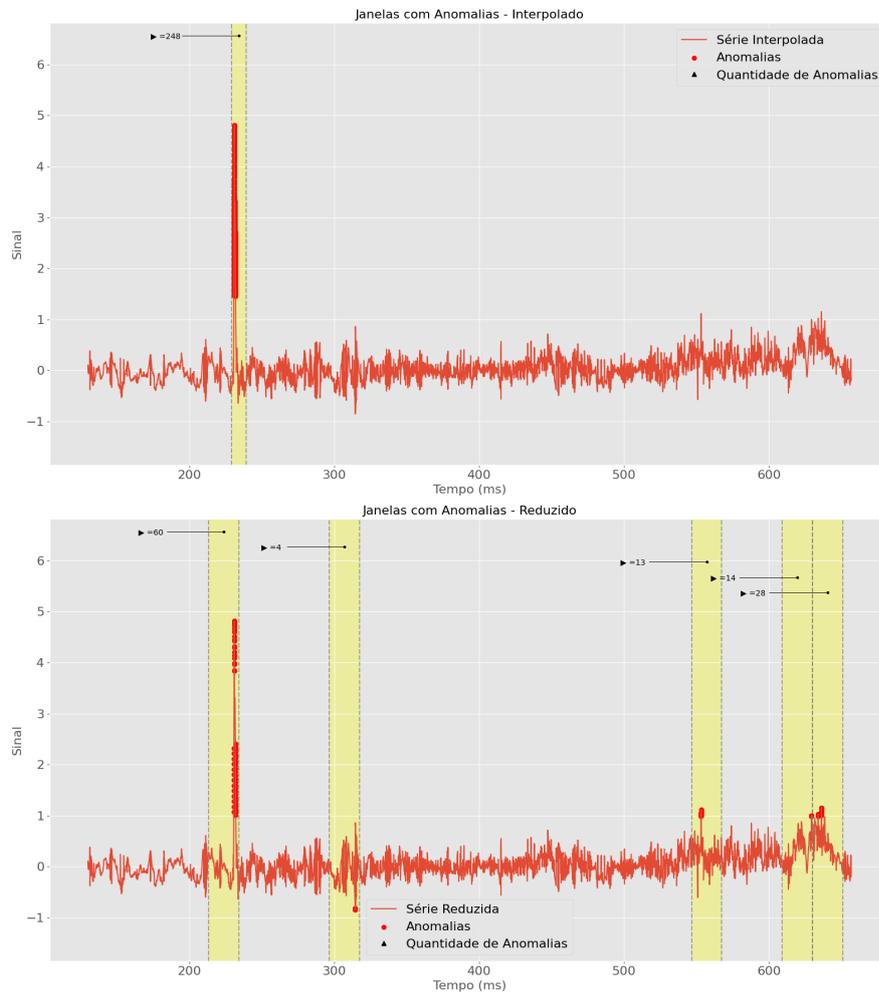
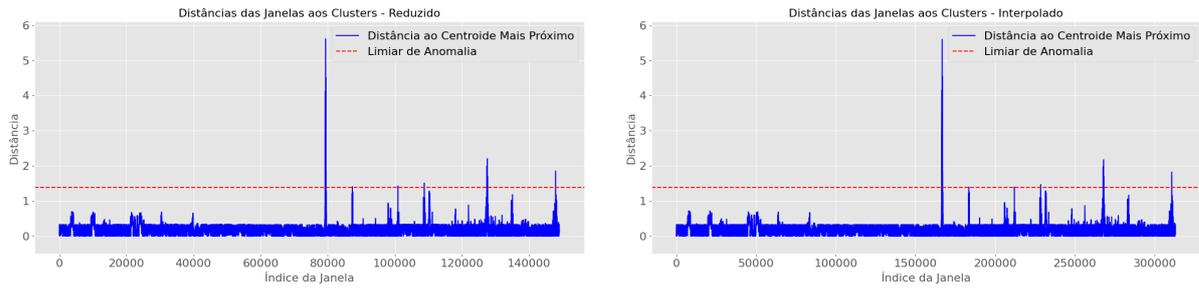


Figura 3.18: Aplicação do K-means agrupamento direto - *Rhynchophorus* experimento 2.



(a) Distância ao centroide - dados equalizados por redução - *Spodoptera* experimento 1.

(b) Distância ao centroide - dados equalizados por interpolação - *Spodoptera* experimento 1.

Figura 3.19: Distância ao centroide - *Spodoptera* experimento 1.

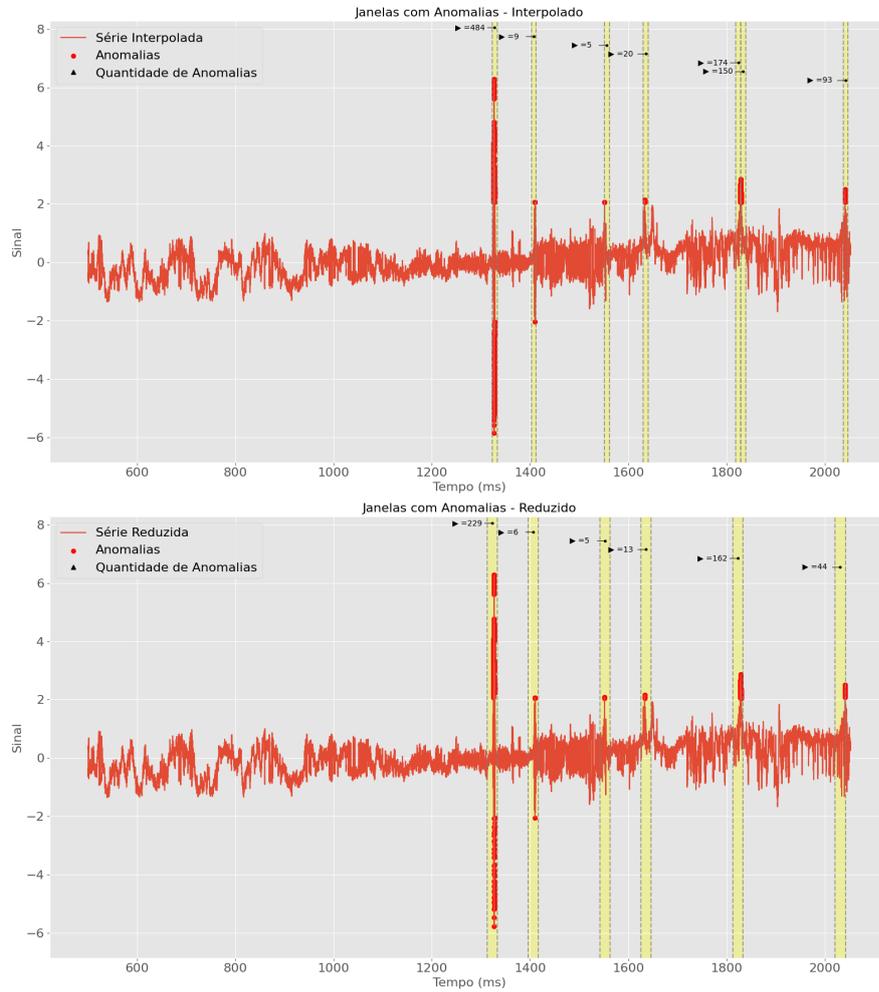
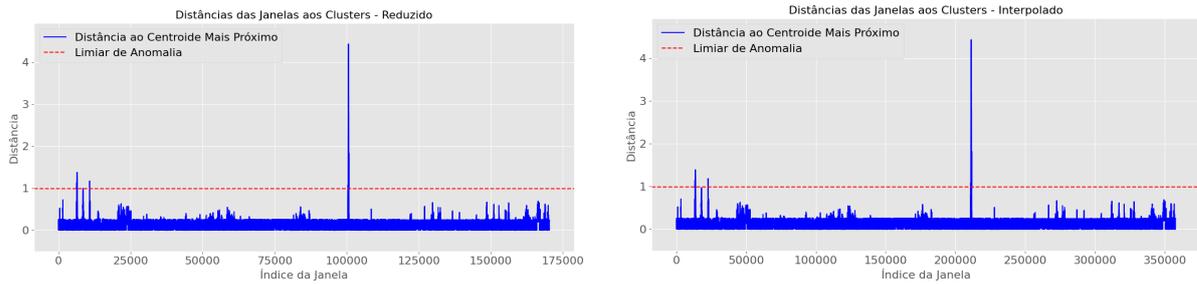


Figura 3.20: Aplicação do K-means agrupamento direto - *Spodoptera* experimento 1.



(a) Distância ao centroide - dados equalizados por redução - *Spodoptera* experimento 2. (b) Distância ao centroide - dados equalizados por interpolação - *Spodoptera* experimento 2.

Figura 3.21: Distância ao centroide - *Spodoptera* experimento 2.

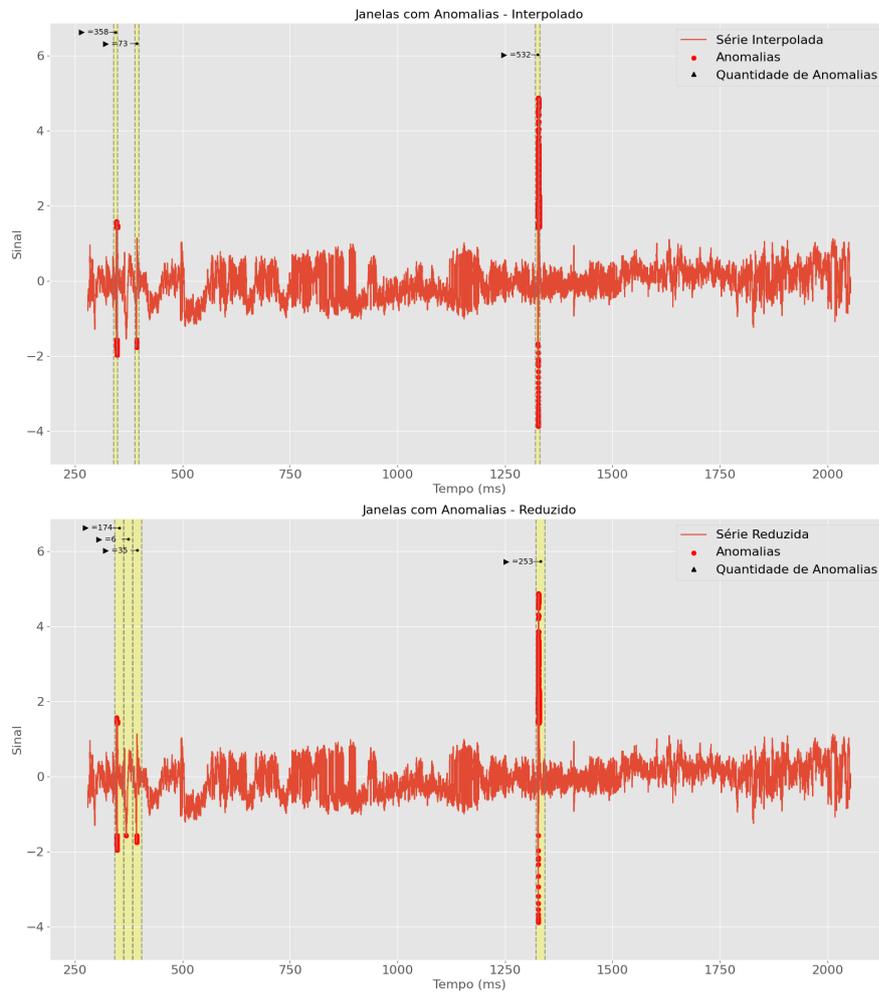
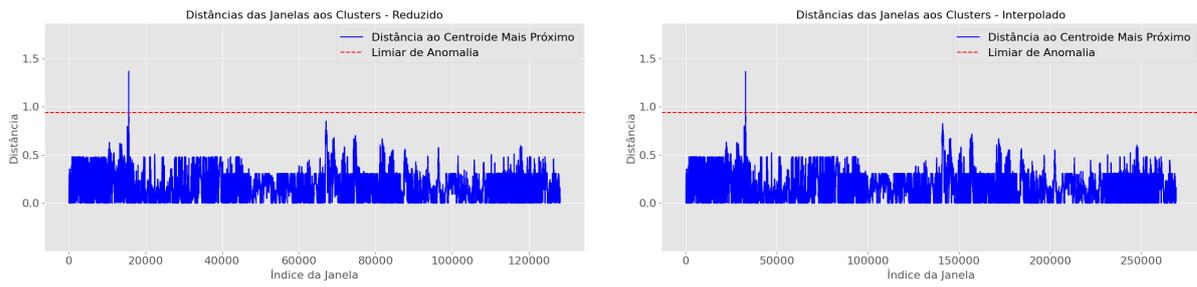


Figura 3.22: Aplicação do K-means agrupamento direto - *Spodoptera* experimento 2.



(a) Distância ao centroide - dados equalizados por redução - *Spodoptera* amostra 2 experimento 1.

(b) Distância ao centroide - dados equalizados por interpolação - *Spodoptera* amostra 2 experimento 1.

Figura 3.23: Distância ao centroide - *Spodoptera* amostra 2 experimento 1.

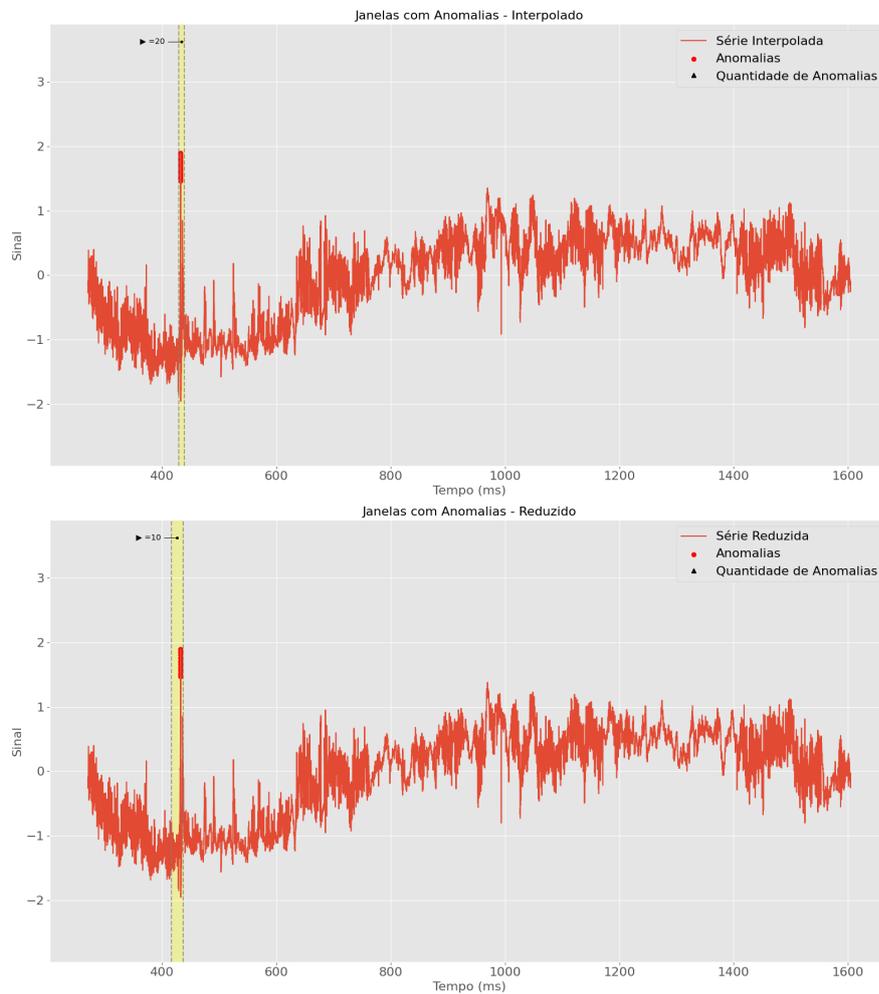
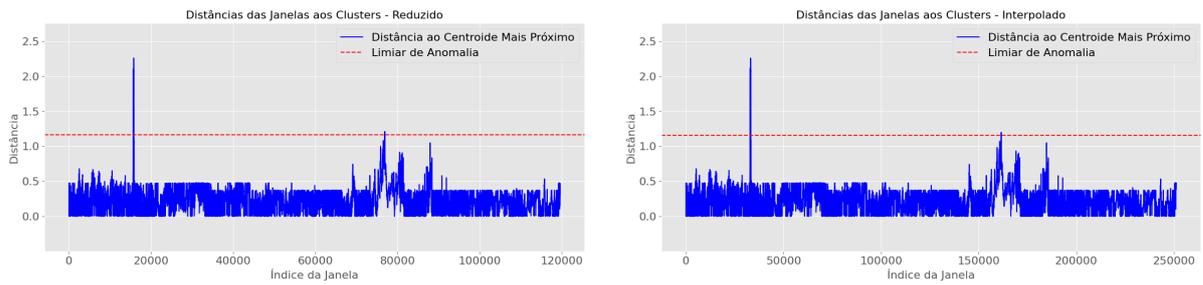


Figura 3.24: Aplicação do K-means agrupamento direto - *Spodoptera* amostra 2 experimento 1.



(a) Distância ao centroide - dados equalizados por redução - *Spodoptera* amostra 2 experimento 2.

(b) Distância ao centroide - dados equalizados por interpolação - *Spodoptera* amostra 2 experimento 2.

Figura 3.25: Distância ao centroide - *Spodoptera* amostra 2 experimento 2.

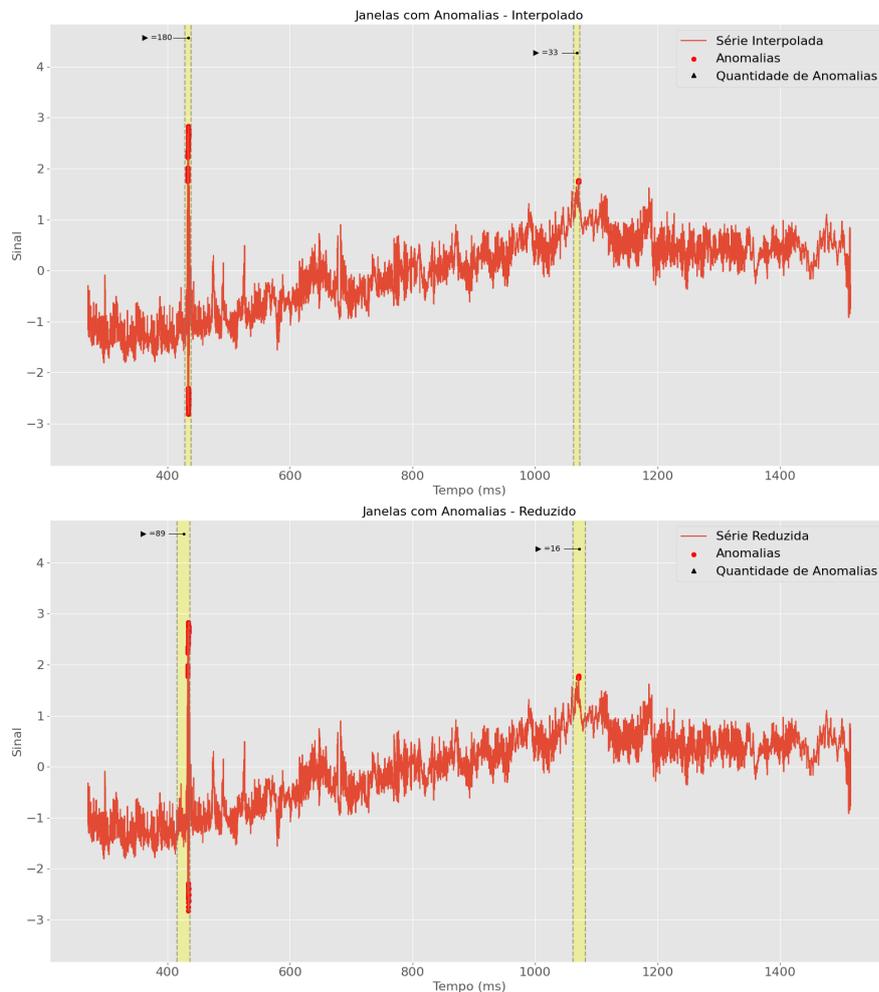


Figura 3.26: Aplicação do K-means agrupamento direto - *Spodoptera* amostra 2 experimento 2.

De forma análoga à análise realizada com o Isolation Forest, as séries foram divididas em janelas de 2000 amostras. Em cada janela, foi contabilizada a quantidade de pontos classificados como anômalos pelo modelo. Os resultados apresentados nas Figuras 3.16, 3.18, 3.20, 3.22,

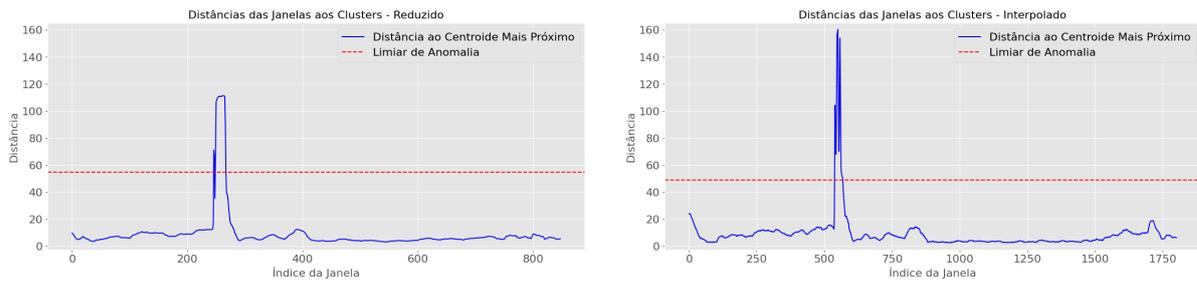
3.24, e 3.26 mostram que o K-means conseguiu identificar regiões de interesse com maior concentração de anomalias, obtendo resultados semelhantes ao Isolation Forest. Observa-se que ambas as técnicas de equalização produziram resultados consistentes.

### 3.6.3 K-means utilizando janela deslizante

Para aprofundar a análise, o algoritmo K-means foi aplicado utilizando janelas deslizantes com tamanho de 2000 amostras e deslocamento de 100 amostras entre as janelas. Essa abordagem permite considerar segmentos específicos das séries de *Rhynchophorus* e *Spodoptera*, destacando as janelas potencialmente anômalas.

O parâmetro  $k = 3$  foi mantido, alinhado à hipótese de separação dos dados em três grupos principais: ruído, picos menores e picos intermediários, com grandes picos sendo classificados como anomalias. Para a detecção, foi utilizado o limiar baseado na Equação 2.7, com  $t = 3$ , definido experimentalmente para otimizar a sensibilidade à identificação de anomalias.

Para cada janela, foram calculadas as distâncias dos pontos ao centroide mais próximo, permitindo a identificação de anomalias que excedem o limiar estabelecido. As Figuras 3.27, 3.29, 3.31, e 3.33 ilustram as distâncias calculadas para os dados equalizados por redução e interpolação. Podemos notar que para os dados interpolados os picos de distâncias estão mais destacados, possivelmente pela presença de mais dados os centroides podem estar mais distantes de janelas com valores maiores.



(a) Distância do centroide - dados equalizados por redução - *Rhynchophorus* experimento 1.

(b) Distância do centroide - dados equalizados por interpolação - *Rhynchophorus* experimento 1.

Figura 3.27: Distância do centroide - *Rhynchophorus* experimento 1.

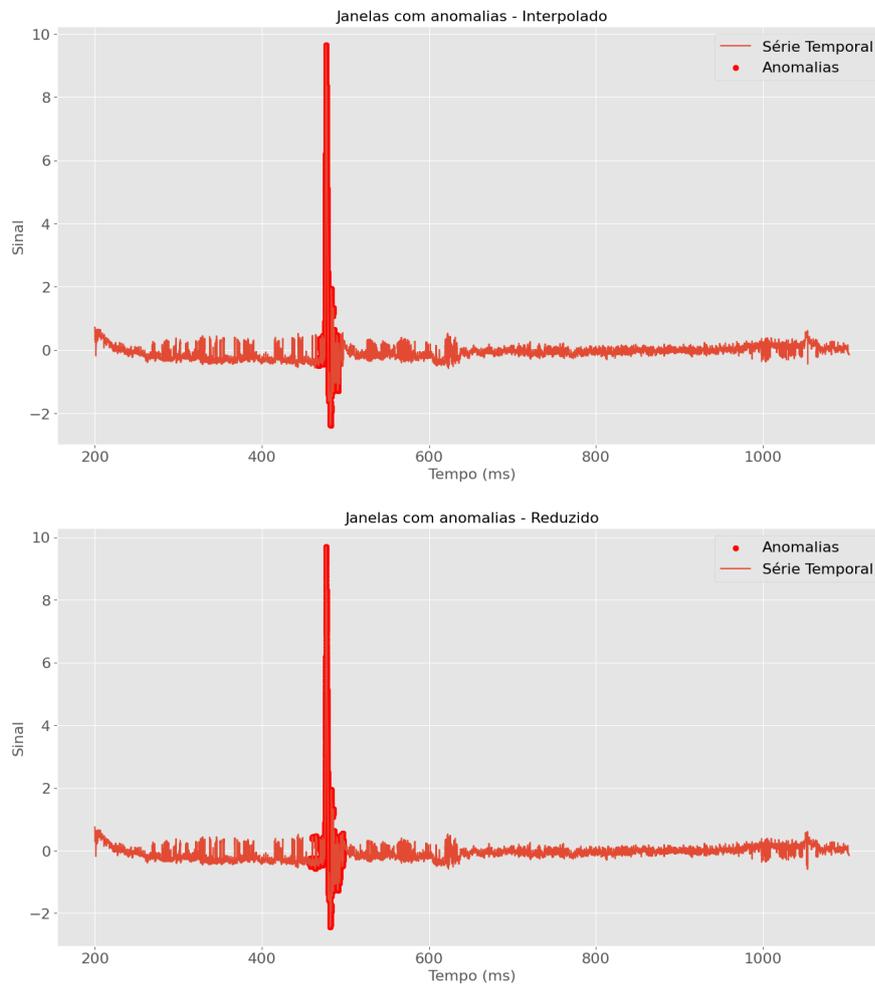
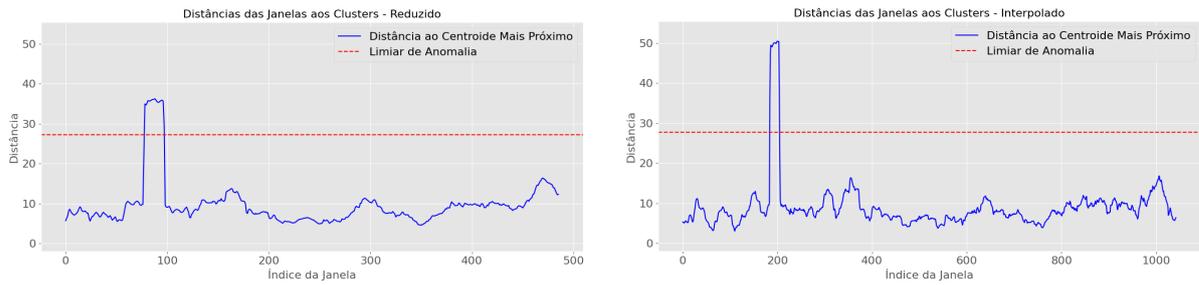


Figura 3.28: Aplicação K-means com janelas deslizantes - *Rhynchophorus* experimento 1.



(a) Distância ao centroide - dados equalizados por redução - *Rhynchophorus* experimento 2.

(b) Distância ao centroide - dados equalizados por interpolação - *Rhynchophorus* experimento 2.

Figura 3.29: Distância ao centroide - *Rhynchophorus* experimento 2.

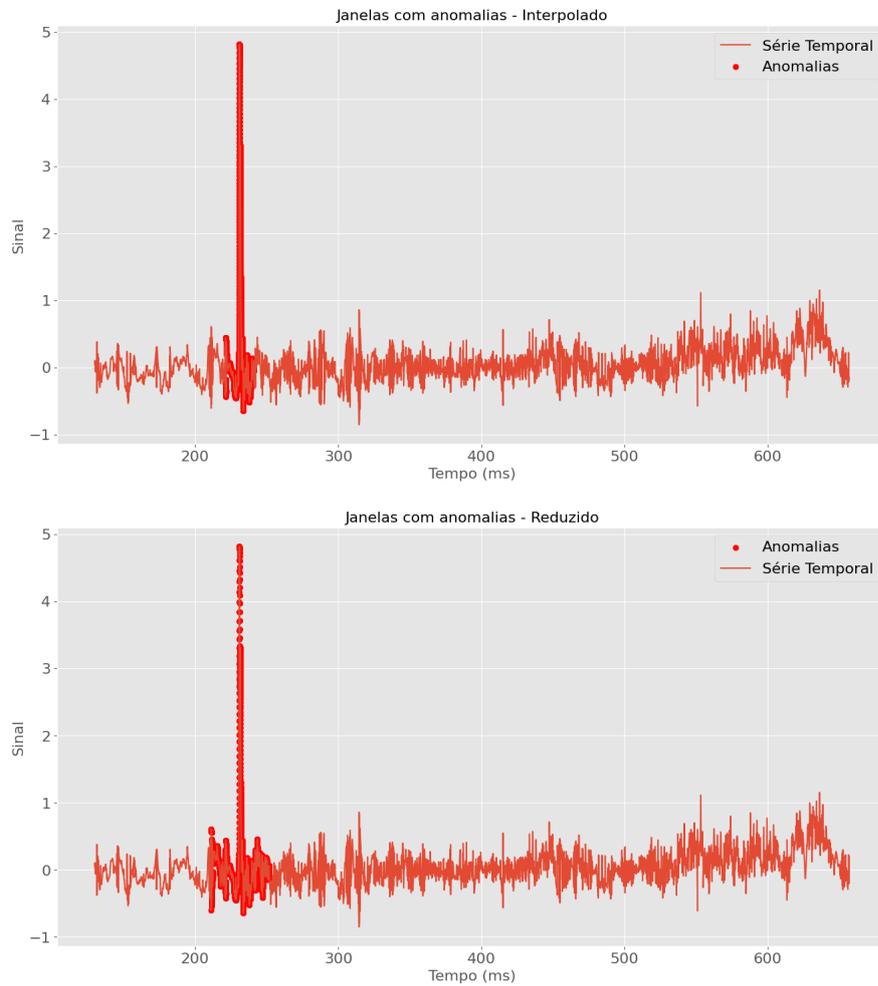
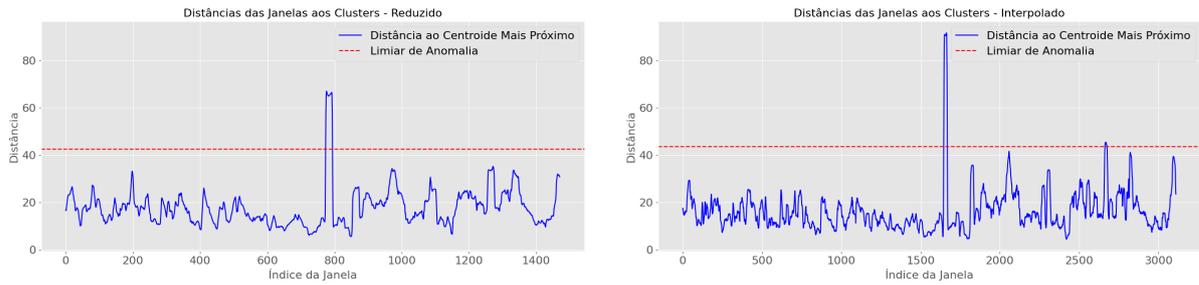


Figura 3.30: Aplicação do K-means com janelas deslizantes - *Rhynchophorus* experimento 2.



(a) Distância ao centroide - dados equalizados por redução - *Spodoptera* experimento 1.

(b) Distância ao centroide - dados equalizados por interpolação - *Spodoptera* experimento 1.

Figura 3.31: Distância ao centroide - *Spodoptera* experimento 1.

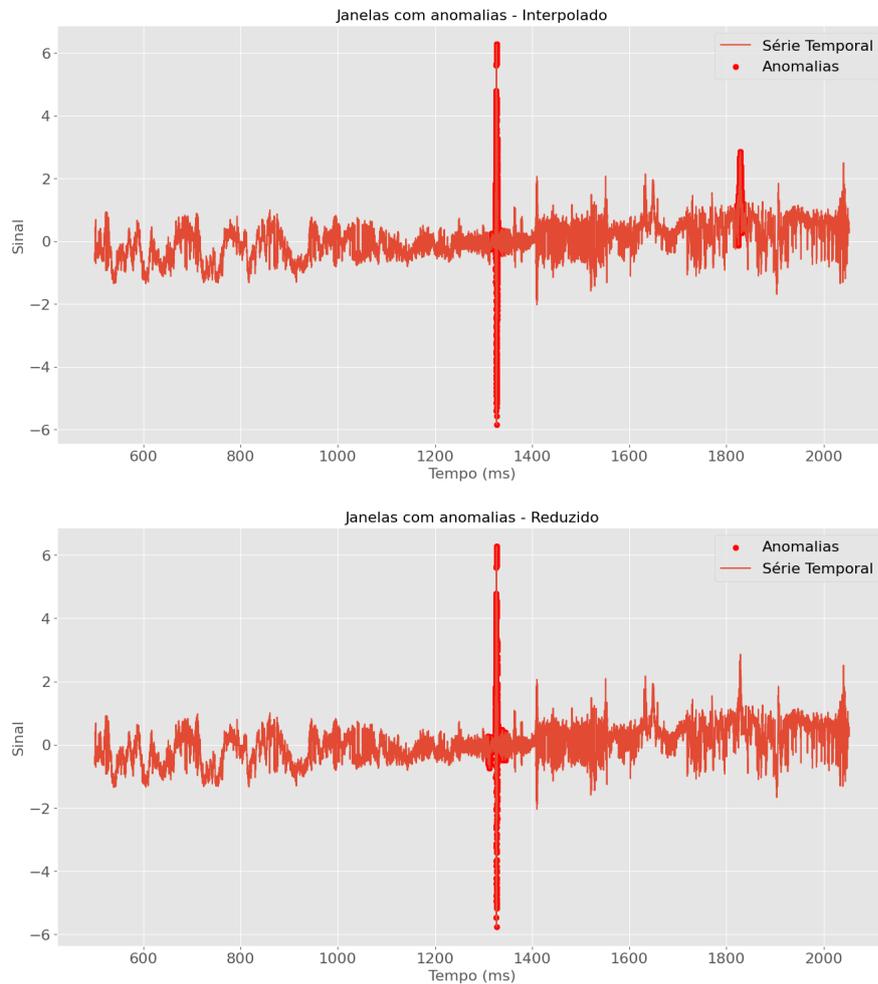
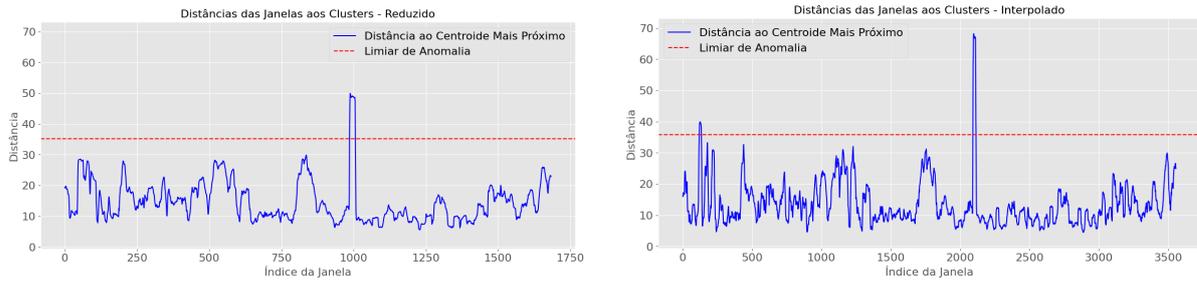


Figura 3.32: Aplicação do K-means com janelas deslizantes - *Spodoptera* experimento 1.



(a) Distância ao centroide - dados equalizados por redução - *Spodoptera* experimento 2.

(b) Distância ao centroide - dados equalizados por interpolação - *Spodoptera* experimento 2.

Figura 3.33: Distância ao centroide - *Spodoptera* experimento 2.

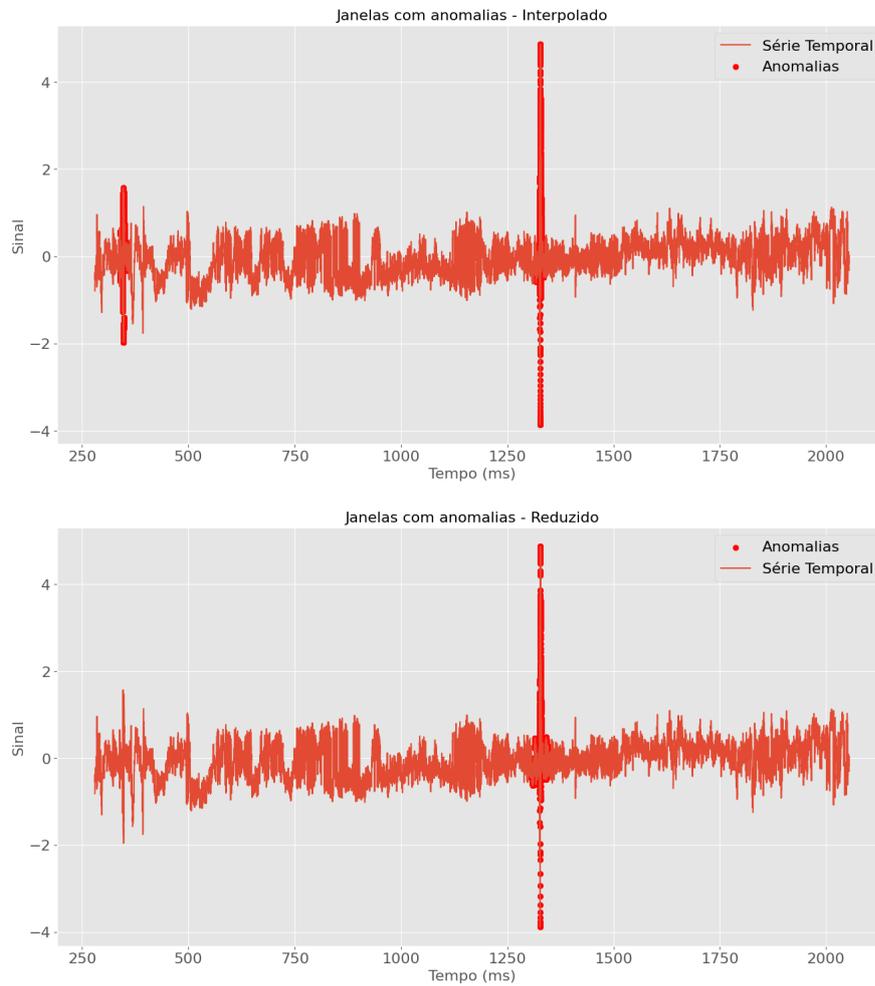
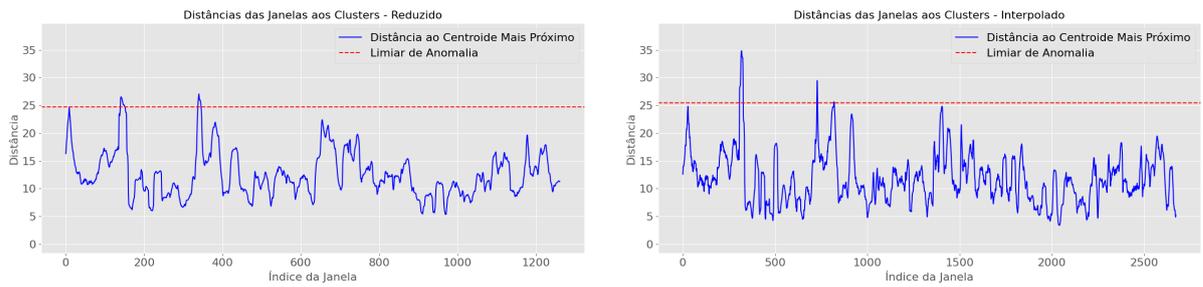


Figura 3.34: Aplicação do K-means com janelas deslizantes - *Spodoptera* experimento 2.



(a) Distância ao centroide - dados equalizados por redução - *Spodoptera* amostra 2 experimento 1.

(b) Distância ao centroide - dados equalizados por interpolação - *Spodoptera* amostra 2 experimento 1.

Figura 3.35: Distância ao centroide - *Spodoptera* amostra 2 experimento 1.

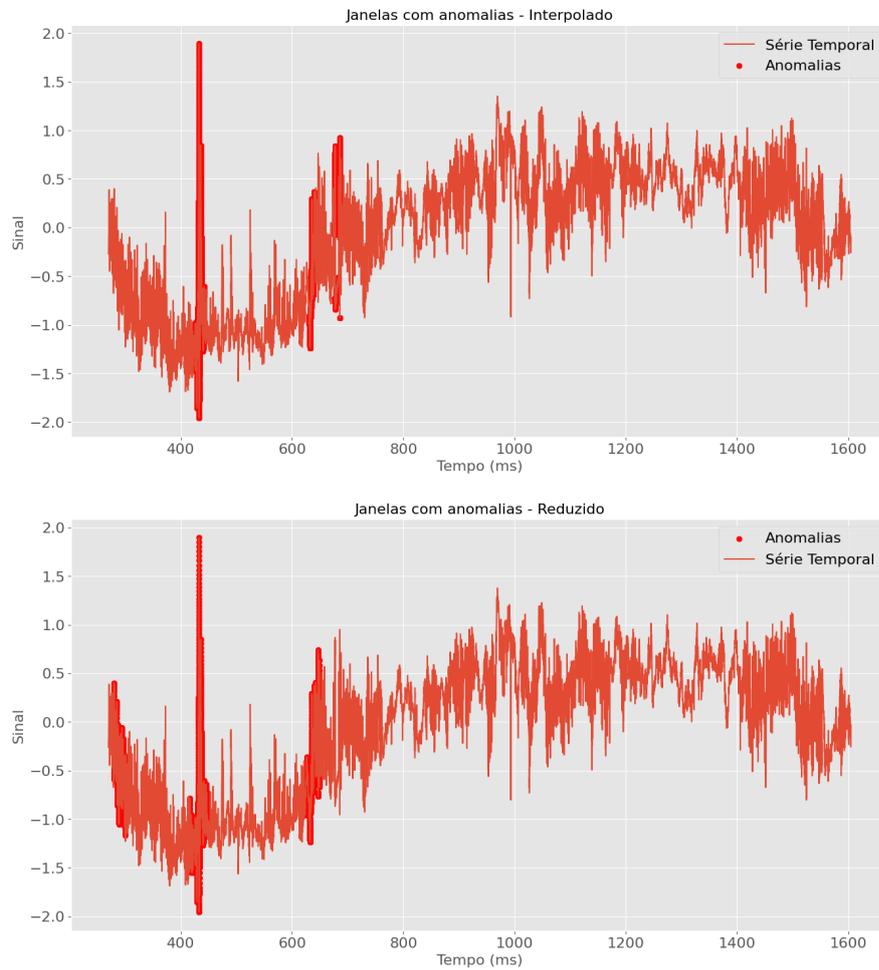
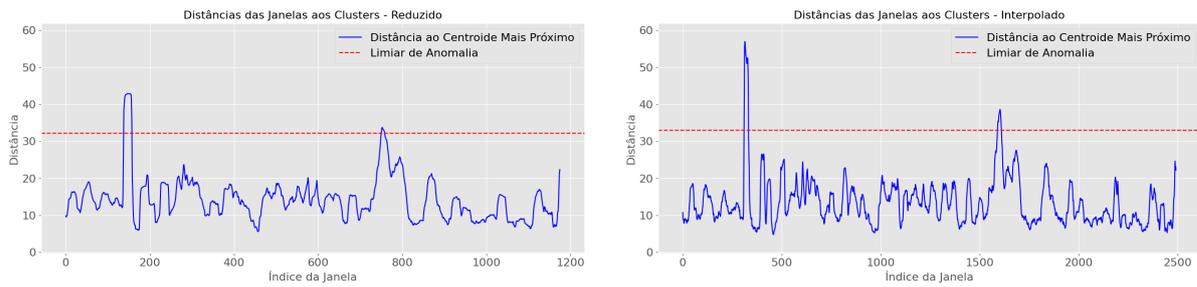


Figura 3.36: Aplicação do K-means com janelas deslizantes - *Spodoptera* amostra 2 experimento 1.



(a) Distância ao centroide - dados equalizados por redução - *Spodoptera* amostra 2 experimento 2.

(b) Distância ao centroide - dados equalizados por interpolação - *Spodoptera* amostra 2 experimento 2.

Figura 3.37: Distância ao centroide - *Spodoptera* amostra 2 experimento 2.

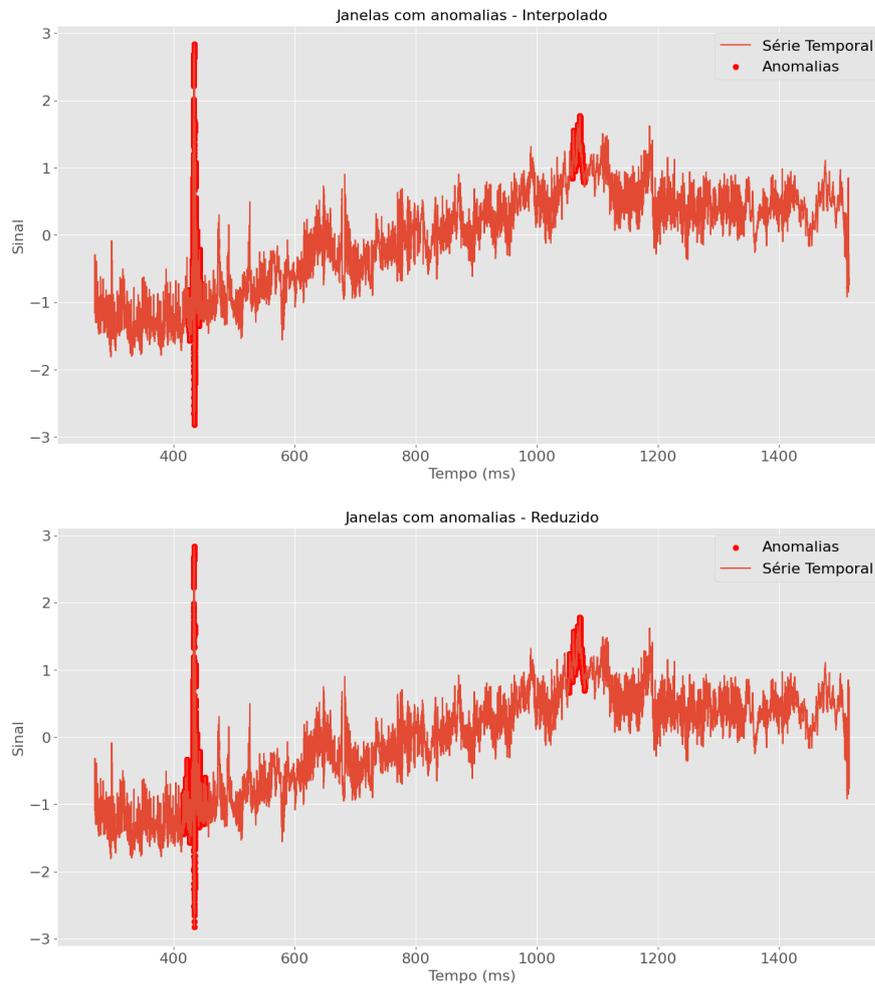


Figura 3.38: Aplicação do K-means com janelas deslizantes - *Spodoptera* amostra 2 experimento 2.

Os resultados evidenciam que a aplicação do K-means com janelas deslizantes conseguiu reduzir significativamente o número de janelas classificadas como anômalas, eliminando parte do ruído observado em abordagens anteriores. Além disso, essa abordagem identificou pelo menos em um dos tipos de equalização (Figuras 3.28, 3.30, 3.32, e 3.34) somente a janela

correspondente a reação das pragas analisadas que pode ser visualizada na Figura 3.2. Nos demais casos (Figura 3.36, e 3.38), a janela das reações também é identificada, porém, com mais algumas janelas.

Ao destacar com maior precisão as janelas relevantes, o método proporciona uma base mais confiável para identificar as reações das pragas, eliminando ruídos e otimizando a detecção de eventos críticos.



## Conclusão

O *pipeline* desenvolvido demonstrou que cada etapa contribuiu significativamente para aprimorar a qualidade dos dados e facilitar a detecção das reações de pragas a compostos semioquímicos. As etapas de pré-processamento, como a suavização para redução de ruídos e a unificação dos sinais por mínimo absoluto, mostraram-se eficazes na diminuição de picos irrelevantes e no destaque de picos simultâneos relevantes. Além disso, o uso de dados sintéticos gerados por interpolação linear apresentou um impacto mínimo no comportamento geral dos dados.

A aplicação de técnicas de aprendizagem de máquina revelou-se promissora na identificação dessas reações, destacando eventos de interesse e eliminando ruídos indesejados. Esse avanço promove uma melhoria significativa no processo de identificação de novos semioquímicos para o manejo sustentável de pragas.

Em trabalhos futuros, planejamos utilizar um maior número de experimentos, explorando ampliando a diversidade dos dados coletados. Um dos focos será a utilização dos dados provenientes do detector de ionização de chama (FID) sem a necessidade de estar acoplado ao eletroantenograma (EAG). Essa abordagem permitirá uma análise mais direta e precisa das leituras dos compostos, eliminando as perdas ocasionadas pela separação dos sinais entre os dois sistemas.

## Referências bibliográficas

*Pheromones*, pages 73–112. Springer Netherlands, Dordrecht, 2002. ISBN 978-0-306-47581-8. DOI [10.1007/0-306-47581-2<sub>3</sub>](https://doi.org/10.1007/0-306-47581-2_3). URL [https://doi.org/10.1007/0-306-47581-2\\_3](https://doi.org/10.1007/0-306-47581-2_3).

Wahid Salman Al Farizi, Indriana Hidayah, and Muhammad Nur Rizal. Isolation forest based anomaly detection: A systematic literature review. In *2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, pages 118–122, 2021. DOI [10.1109/ICITACEE53184.2021.9617498](https://doi.org/10.1109/ICITACEE53184.2021.9617498).

Stamatios-Aggelos N. Alexandropoulos, Sotiris B. Kotsiantis, and Michael N. Vrahatis. Data preprocessing in predictive data mining. *The Knowledge Engineering Review*, 34:e1, 2019. DOI [10.1017/S026988891800036X](https://doi.org/10.1017/S026988891800036X).

Mohsena Ashraf, Farzana Anowar, Jahangir H. Setu, Atiqul I. Chowdhury, Eshtiak Ahmed, Ashraful Islam, and Abdullah Al-Mamun. A survey on dimensionality reduction techniques for time-series data. *IEEE Access*, 11:42909–42923, 2023. DOI [10.1109/ACCESS.2023.3269693](https://doi.org/10.1109/ACCESS.2023.3269693).

Ricardo Barbosa-Cornelio, Fernando Cantor, Ericsson Coy-Barrera, and Daniel Rodríguez. Tools in the investigation of volatile semiochemicals on insects: From sampling to statistical analysis. *Insects*, 10(8), 2019. ISSN 2075-4450. DOI [10.3390/insects10080241](https://doi.org/10.3390/insects10080241). URL <https://www.mdpi.com/2075-4450/10/8/241>.

Alexandra Nava Brezolin, Janine Martinazzo, Daniela Kunkel Muenchen, Alana Marie de Cezaro, Aline Andressa Rigo, Clarice Steffens, Juliana Steffens, Maria Carolina Blassioli-Moraes, and Miguel Borges. Tools for detecting insect semiochemicals: a review. In *Anal Bioanal Chem*, volume 410, page 4091–4108, 2018. DOI <https://doi.org/10.1007/s00216-018-1118-3>.

Yinan Bu, Liangliang Shi, and Bin Ma. Method for establishing predictive models for total organic halogen based on piecewise interpolation and machine learning. *Journal of Environmental Chemical Engineering*, 11(3):109928, 2023. ISSN 2213-3437.

- DOI** <https://doi.org/10.1016/j.jece.2023.109928>. URL <https://www.sciencedirect.com/science/article/pii/S221334372300667X>.
- S. Butterworth. On the Theory of Filter Amplifiers. *Experimental Wireless & the Wireless Engineer*, 7:536–541, 1930.
- Vincas Būda, Violeta Apšegaitė, Laima Blažytė-Čereškienė, Rita Butkienė, Irena Nedveckytė, and Dalė Pečiulytė. Response of moth plodia interpunctella to volatiles of fungus-infected and uninfected wheat grain. *Journal of Stored Products Research*, 69:152–158, 2016. ISSN 0022-474X. **DOI** <https://doi.org/10.1016/j.jspr.2016.08.001>. URL <https://www.sciencedirect.com/science/article/pii/S0022474X16300522>.
- Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3), July 2009. ISSN 0360-0300. **DOI** [10.1145/1541880.1541882](https://doi.org/10.1145/1541880.1541882). URL <https://doi.org/10.1145/1541880.1541882>.
- Abdullah Elen and Emre Avuçlu. Standardized variable distances: A distance-based machine learning method. *Applied Soft Computing*, 98:106855, 2021. ISSN 1568-4946. **DOI** <https://doi.org/10.1016/j.asoc.2020.106855>. URL <https://www.sciencedirect.com/science/article/pii/S1568494620307936>.
- Theo Tryfonas et al. Elisa Coraggio, Claire Gronow. Low-pass filter application for noise removal in water quality data with high temporal resolution. *Research Square*, 2023. **DOI** <https://doi.org/10.21203/rs.3.rs-3284948/v1>.
- Shahira M. Ezzat, Jaison Jeevanandam, Chukwuebuka Egbuna, Rana M. Merghany, Muhammad Akram, Muhammad Daniyal, Jaweria Nisar, and Aamir Sharif. Chapter 7 - semiochemicals: A green approach to pest and disease control. In Chukwuebuka Egbuna and Barbara Sawicka, editors, *Natural Remedies for Pest, Disease and Weed Control*, pages 81–89. Academic Press, 2020. ISBN 978-0-12-819304-4. **DOI** <https://doi.org/10.1016/B978-0-12-819304-4.00007-5>. URL <https://www.sciencedirect.com/science/article/pii/B9780128193044000075>.
- Benjamin Fürstenau, Cornel Adler, Hartwig Schulz, and Monika Hilker. Host habitat volatiles enhance the olfactory response of the larval parasitoid holepyris sylvanidis to specifically host-associated cues. *Chem Senses*, pages 611–621, 2016. **DOI** [10.1093/chemse/bjw065](https://doi.org/10.1093/chemse/bjw065).
- SeyedeZahra Golazad, Abbas Mohammadi, Abbas Rashidi, and Mohammad Ilbeigi. From raw to refined: Data preprocessing for construction machine learning (ml), deep learning (dl), and reinforcement learning (rl) models. *Automation in Construction*, 168:105844, 2024. ISSN 0926-5805. **DOI** <https://doi.org/10.1016/j.autcon.2024.105844>. URL <https://www.sciencedirect.com/science/article/pii/S0926580524005806>.

- J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979. ISSN 00359254, 14679876. URL <http://www.jstor.org/stable/2346830>.
- M. J. Hassemer, J. Sant’Ana, M. W. M. de Oliveira, M. Borges, R. A. Laumann, M. Caumo, and M. C. Blassioli-Moraes. Chemical composition of alphetobius diaperinus (coleoptera: Tenebrionidae) abdominal glands and the influence of 1,4-benzoquinones on its behavior. *Journal of Economic Entomology*, 108(4):2107–2116, 06 2015. ISSN 0022-0493. DOI [10.1093/jee/tov147](https://doi.org/10.1093/jee/tov147). URL <https://doi.org/10.1093/jee/tov147>.
- R. Kumari, Sheetanshu, M. K. Singh, R. Jha, and N.K. Singh. Anomaly detection in network traffic using k-mean clustering. In *2016 3rd International Conference on Recent Advances in Information Technology (RAIT)*, pages 387–393, 2016. DOI [10.1109/RAIT.2016.7507933](https://doi.org/10.1109/RAIT.2016.7507933).
- Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422, 2008. DOI [10.1109/ICDM.2008.17](https://doi.org/10.1109/ICDM.2008.17).
- Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation-based anomaly detection. 6(1), March 2012. ISSN 1556-4681. DOI [10.1145/2133360.2133363](https://doi.org/10.1145/2133360.2133363). URL <https://doi.org/10.1145/2133360.2133363>.
- Elisa Marcelli, Tommaso Barbariol, Davide Sartor, and Gian Antonio Susto. Active learning-based isolation forest (alif): Enhancing anomaly detection with expert feedback. *Information Sciences*, 678:121012, 2024. ISSN 0020-0255. DOI <https://doi.org/10.1016/j.ins.2024.121012>. URL <https://www.sciencedirect.com/science/article/pii/S0020025524009265>.
- Raphael Njurai Miano, Pascal Mahukpe Ayelo, Richard Musau, Ahmed Hassanali, and Samira A. Mohamed. Electroantennogram and machine learning reveal a volatile blend mediating avoidance behavior by tuta absoluta females to a wild tomato plant. *Sci Rep*, 2022. DOI <https://doi.org/10.1038/s41598-022-13125-0>.
- Andrew J. Myrick and Thomas C. Baker. Chopper-stabilized gas chromatography-electroantennography: Part i. background, signal processing and example. *Biosensors and Bioelectronics*, 31(1):197–204, 2012. ISSN 0956-5663. DOI <https://doi.org/10.1016/j.bios.2011.10.017>. URL <https://www.sciencedirect.com/science/article/pii/S0956566311006853>.
- Shannon B. Olsson and Bill S. Hansson. *Electroantennogram and Single Sensillum Recording in Insect Antennae*, pages 157–177. Humana Press, Totowa, NJ, 2013. ISBN 978-1-62703-619-1. DOI [10.1007/978-1-62703-619-1\\_11](https://doi.org/10.1007/978-1-62703-619-1_11). URL [https://doi.org/10.1007/978-1-62703-619-1\\_11](https://doi.org/10.1007/978-1-62703-619-1_11).

- Gábor Princz, Masoud Shaloo, and Selim Erol. Anomaly detection in binary time series data: An unsupervised machine learning approach for condition monitoring. *Procedia Computer Science*, 232:1065–1078, 2024. ISSN 1877-0509.  
**DOI** <https://doi.org/10.1016/j.procs.2024.01.105>. URL <https://www.sciencedirect.com/science/article/pii/S1877050924001054>. 5th International Conference on Industry 4.0 and Smart Manufacturing (ISM 2023).
- Artur M. Schweidtmann, Erik Esche, Asja Fischer, Marius Kloft, Jens-Uwe Repke, Sebastian Sager, and Alexander Mitsos. Machine learning in chemical engineering: A perspective. *Chemie Ingenieur Technik*, 93(12):2029–2039, 2021.  
**DOI** <https://doi.org/10.1002/cite.202100083>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/cite.202100083>.
- Yusuke Shiota, Takeshi Sakurai, Takaaki Daimon, Hidefumi Mitsuno, Takeshi Fujii, Shigeru Matsuyama, Hideki Sezutsu, Yukio Ishikawa, and Ryohei Kanzaki. In vivo functional characterisation of pheromone binding protein-1 in the silkworm, *bombyx mori*. *Sci. Rep.*, 8(1):13529, September 2018.
- Xiao-chao Wu, Guo-liang Wang, Liao-liao Yan, and Li-xin Fang. Application of Butterworth Filter in Display of Radar Target Track . In *2014 Sixth International Conference on Measuring Technology and Mechatronics Automation*, volume 1, pages 605–607, Los Alamitos, CA, USA, March 2010. IEEE Computer Society.  
**DOI** [10.1109/ICMTMA.2010.297](https://doi.org/10.1109/ICMTMA.2010.297). URL <https://doi.ieeecomputersociety.org/10.1109/ICMTMA.2010.297>.
- Joanne Y. Yew and Henry Chung. Insect pheromones: An overview of function, form, and discovery. *Progress in Lipid Research*, 59:88–105, 2015. ISSN 0163-7827.  
**DOI** <https://doi.org/10.1016/j.plipres.2015.06.001>. URL <https://www.sciencedirect.com/science/article/pii/S0163782715000272>.
- Anicia Zeberli, Sara Badr, Christian Siegmund, Markus Mattern, and Hirokazu Sugiyama. Data-driven anomaly detection and diagnostics for changeover processes in biopharmaceutical drug product manufacturing. *Chemical Engineering Research and Design*, 167:53–62, 2021. ISSN 0263-8762.  
**DOI** <https://doi.org/10.1016/j.cherd.2020.12.018>. URL <https://www.sciencedirect.com/science/article/pii/S0263876220306146>.