

Engenharia Elétrica

PolySense Station

Sistema de Monitoramento Ambiental

Coleta de dados multi-sensor, análise temporal e aprendizado de máquina
aplicados a monitoramento climático

Autor: Jonas Souza

Formação: Engenheiro Elétrica

Data: 19 de dezembro de 2025

Sumário

1 Sumário Executivo	2
1.1 Principais Realizações	2
2 Introdução	2
2.1 Contexto	2
2.2 Objetivos	2
2.3 Localização e Período	3
3 Metodologia de Coleta de Dados	3
3.1 Arquitetura de Hardware	3
3.2 Suite de Sensores Implementada	6
3.3 Firmware de Aquisição	7
3.4 Validação Contra Estação INMET	8
4 Módulo 1: Análise Exploratória de Dados (Notebooks 01-05)	9
4.1 Notebook 01: Análise Exploratória Inicial	9
4.1.1 Distribuições de Temperatura	9
4.1.2 Distribuições de Umidade	11
4.1.3 Distribuições de Pressão Barométrica	13
4.2 Notebook 02: Análise de Correlações	15
4.2.1 Matriz de Correlação Global	15
4.2.2 Relações Bivariadas com KDE	16
4.3 Notebook 03: Análise de Dados Faltantes	19
4.4 Notebook 04: Validação de Sensores	22
4.4.1 Análise de Bland-Altman - Temperatura	22
4.4.2 Análise de Bland-Altman - Umidade e Pressão	24
4.5 Notebook 05: Análise Temporal	26
4.5.1 Séries Temporais Completas	26
4.6 Notebook 06: Decomposição de Séries Temporais	29
5 Módulo 2: Machine Learning (Notebooks 07-11)	32
5.1 Notebook 07: Detecção de Anomalias	32
5.1.1 Isolation Forest	32
5.1.2 Detecção de Valores Congelados	34
5.2 Notebook 08: Regressão com Árvore de Decisão	35
5.3 Notebook 09: Clustering com Gaussian Mixture Model (GMM)	37
5.4 Notebook 10: Clustering com KMeans	39
5.5 Notebook 11: Previsão com Redes Neurais LSTM	41
6 Módulo 3: Processamento Digital de Sinais (Notebooks 12-13)	43
6.1 Notebook 12: Filtros Digitais	43
6.1.1 Comparação de Filtros - Temperatura	43
6.1.2 Comparação de Filtros - Umidade	64
6.1.3 Comparação de Filtros - Pressão	70
6.2 Notebook 13: Análise de Frequência (FFT)	75
7 Dataset Público	76

8 Estrutura do Repositório	76
9 Conclusões	77
9.1 Principais Conquistas	77
9.2 Limitações do Estudo	77
9.3 Trabalhos Futuros	77
10 Referências	78

1 Sumário Executivo

Este relatório apresenta a análise completa do projeto **PolySense Station**, um sistema autônomo de aquisição e análise de dados ambientais desenvolvido do zero. O sistema combina hardware customizado com 7 sensores integrados, microcontrolador Raspberry Pi Pico e um pipeline completo de análise de dados com 13 notebooks especializados.

1.1 Principais Realizações

- **Coleta de dados real:** 82.430 medições ao longo de 30 dias (setembro de 2025)
- **Resolução temporal:** Intervalo de amostragem de 30 segundos
- **Redundância de sensores:** 7 sensores de temperatura, 2 de umidade, 2 de pressão
- **Validação contra estação INMET:** Instituto Nacional de Meteorologia
- **Acurácia do modelo LSTM:** MAE $\pm 1^{\circ}\text{C}$ para previsão de temperatura 1 hora à frente
- **Clustering climático:** 4 regimes distintos identificados via GMM
- **Dataset público:** Disponibilizado no Kaggle com 82.430 registros

2 Introdução

2.1 Contexto

O monitoramento ambiental é fundamental para compreender padrões climáticos locais, validar modelos preditivos e desenvolver sistemas de automação sensíveis ao clima. Este projeto implementa uma estação meteorológica completa com coleta automática, armazenamento em cartão SD e análise avançada de dados.

2.2 Objetivos

1. Desenvolver hardware de aquisição de dados com redundância de sensores
2. Coletar dados ambientais (30 segundos de intervalo)
3. Validar medições contra estação meteorológica oficial (INMET)
4. Realizar análise exploratória completa dos dados coletados
5. Aplicar técnicas de machine learning para identificação de padrões
6. Desenvolver modelos preditivos de temperatura com redes neurais LSTM
7. Aplicar processamento digital de sinais para redução de ruído

2.3 Localização e Período

- **Local:** Vitória da Conquista, Bahia, Brasil
- **Altitude:** 923 metros
- **Período de coleta:** 31 de agosto a 30 de setembro de 2025
- **Total de registros:** 82.430 medições
- **Tamanho do dataset:** 4.8 MB (CSV bruto)

3 Metodologia de Coleta de Dados

3.1 Arquitetura de Hardware

O sistema foi desenvolvido utilizando o microcontrolador **Raspberry Pi Pico** (RP2040) programado em MicroPython, com as seguintes interfaces:

- **Dual I2C:** Dois barramentos I2C para minimizar interferências elétricas
- **SPI:** Interface para gravação em cartão SD
- **OneWire:** Protocolo para sensor de temperatura DS18B20
- **Analógico:** Conversor ADC para termistor NTC
- **Display OLED:** Tela 128x64 para monitoramento em tempo real
- **RTC:** Relógio de tempo real para timestamps precisos

Design de Hardware: O sistema inclui uma PCB customizada para integração dos sensores e um esquemático detalhado para replicação:

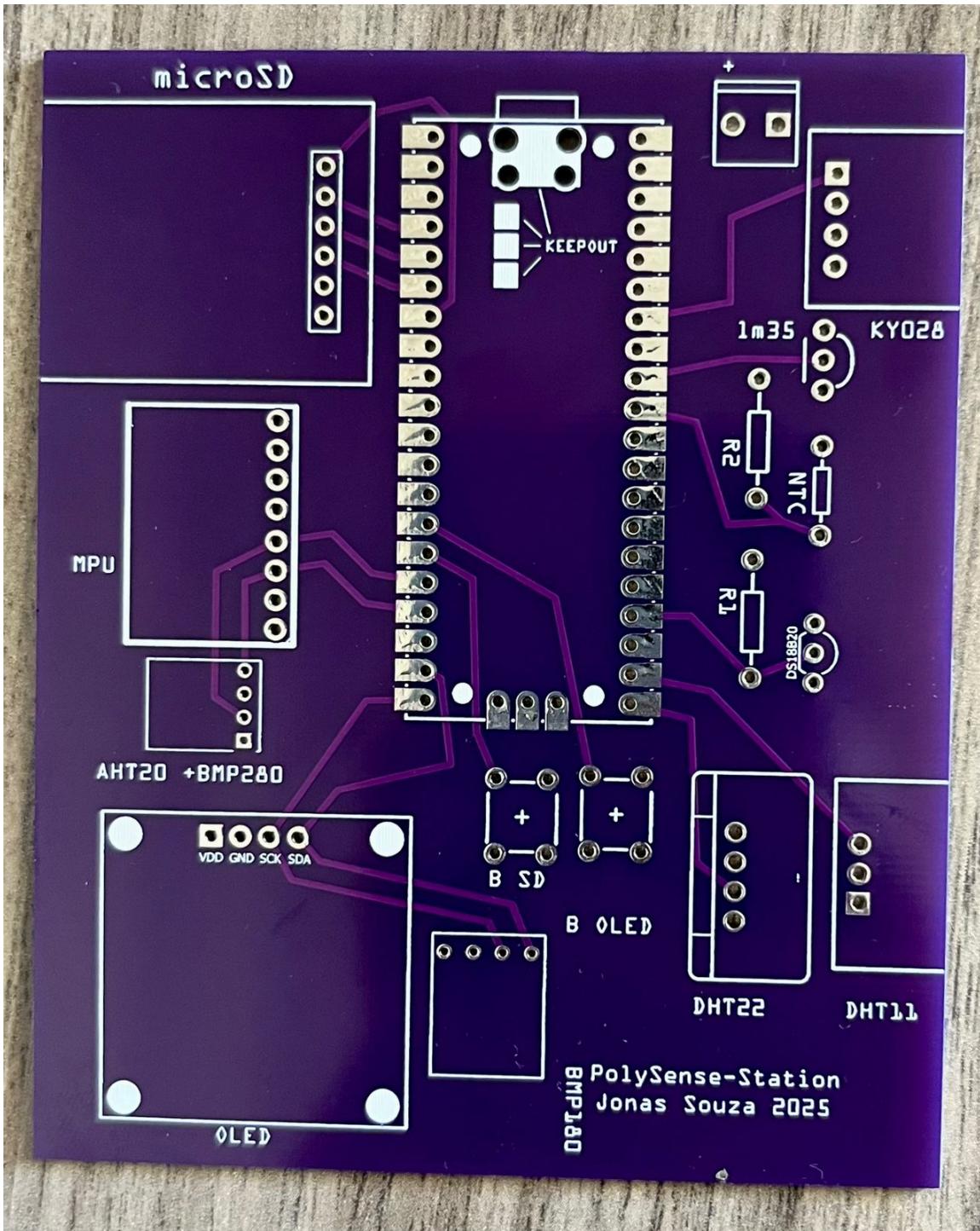


Figura 1: Fotografia da PCB front (Printed Circuit Board) do PolySense Station

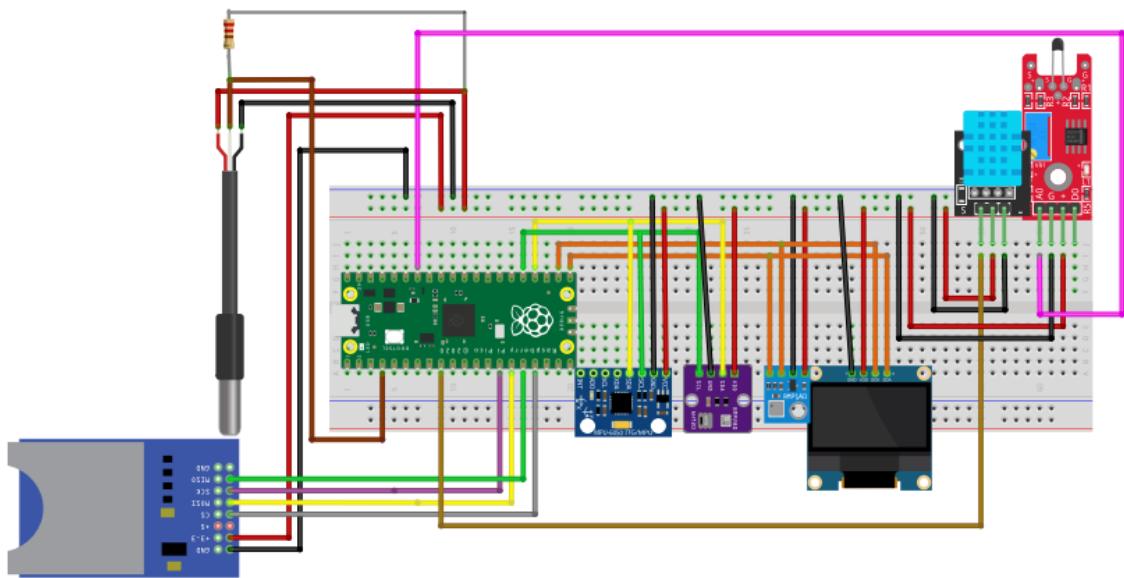


Figura 2: Esquemático do circuito protoboard do PolySense Station

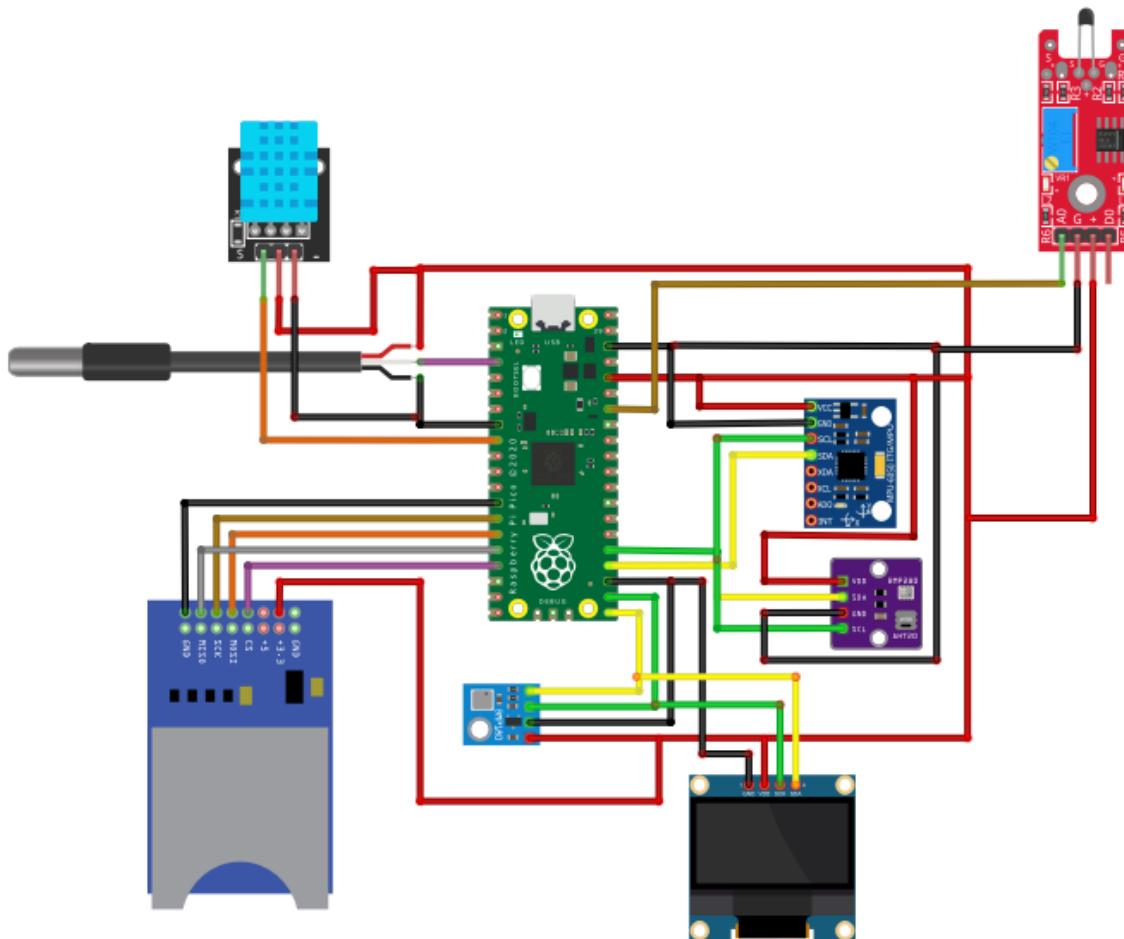


Figura 3: Esquemático detalhado dos sensores integrados

3.2 Suite de Sensores Implementada

Tabela 1: Sensores integrados no PolySense Station

Sensor	Tipo	Medições	Protocolo
MPU6050	Giroscópio/Acelerômetro	Temperatura	I2C
AHT20	Ambiental	Temperatura, Umidade	I2C
BMP280	Barométrico	Temperatura, Pressão	I2C
BMP180	Barométrico	Temperatura, Pressão	I2C
DS18B20	OneWire Digital	Temperatura	OneWire
NTC (KY-028)	Termistor Analógico	Temperatura	ADC
DHT11	Ambiental	Temperatura, Umidade	Digital

Justificativa da redundância:

- Validação cruzada entre múltiplos sensores
- Detecção de falhas instrumentais
- Referências (DS18B20, BMP280, AHT20)

3.3 Firmware de Aquisição

O firmware principal (`main.py`, 226 linhas) implementa:

- Leitura síncrona de todos os sensores a cada 30 segundos
- Gravação em cartão SD com timestamp RTC
- Display OLED rotativo com 3 telas de informação
- LED de feedback (pisca a cada gravação bem-sucedida)
- Sistema de ejeção segura via botão
- Tratamento de erros e status de gravação

Formato de armazenamento CSV:

```
Timestamp, Temp_MP6050_C, Temp_AHT20_C, Umid_AHT20_pct,  
Temp_BMP280_C, Press_BMP280_hPa, Temp_BMP180_C,  
Press_BMP180_hPa, Temp_DS18B20_C, Temp_NTC_C,  
Temp_DHT11_C, Umid_DHT11_pct
```

```
-----  
MPU6050 : 26.38 °C  
BMP280 : 27.33 °C | 912.57 hPa  
AHT20 : 25.90 °C | 54.94 %  
DHT22 : 26.70 °C | 58.80 %  
DHT11 : 25.08 °C | 53.04 %  
-----
```

Figura 4: Saída serial do firmware durante operação do PolySense Station

3.4 Validação Contra Estação INMET

Os dados foram validados contra a estação meteorológica oficial do INMET:

- **Dataset de validação:** 719 registros da estação INMET
- **Variáveis INMET:** Temperatura, umidade, pressão, vento, radiação, precipitação
- **Dataset combinado:** 331 registros horários agregados

4 Módulo 1: Análise Exploratória de Dados (Notebooks 01-05)

4.1 Notebook 01: Análise Exploratória Inicial

4.1.1 Distribuições de Temperatura

Análise das distribuições de temperatura revelou que todos os 7 sensores apresentam padrões consistentes:

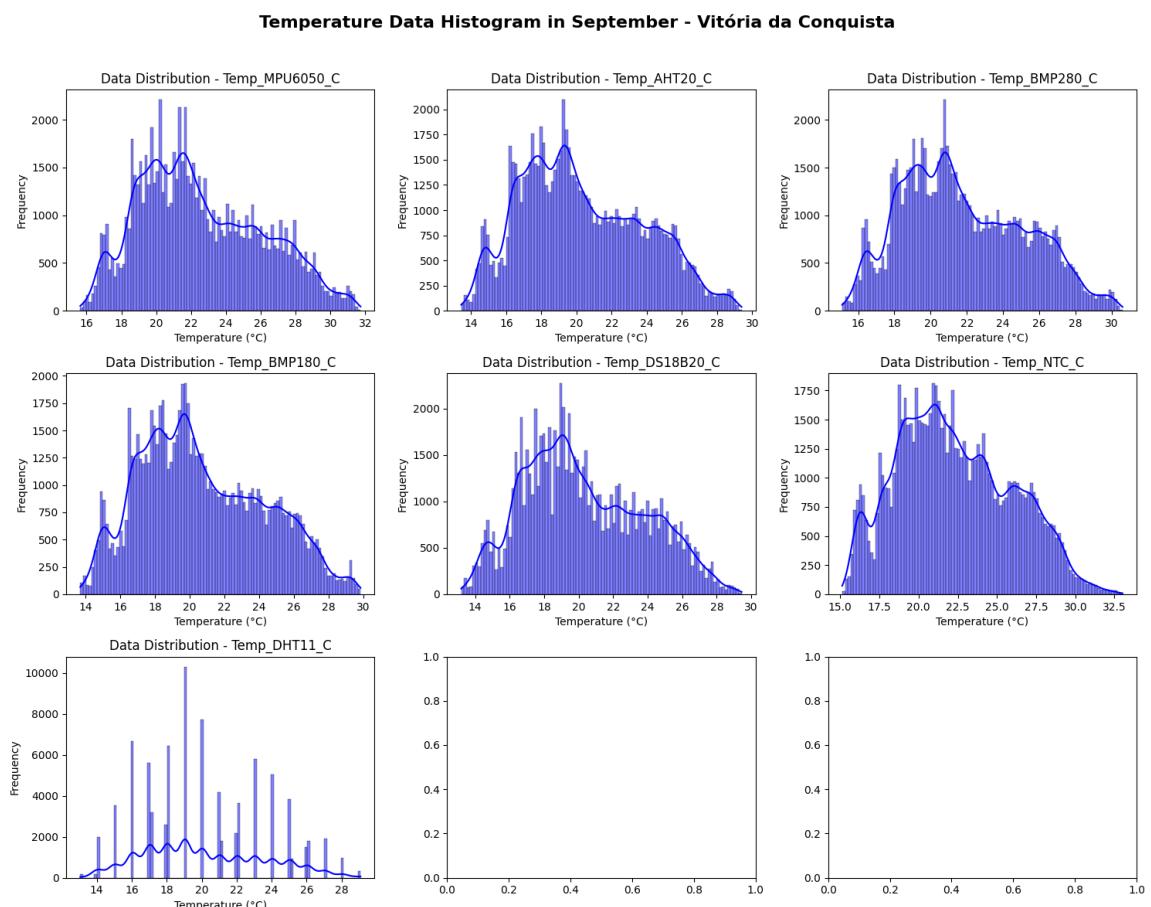


Figura 5: Histograma de distribuições de temperatura (7 sensores)

Observações:

- Distribuições aproximadamente normais com picos entre 20-22°C

- Sensores(BMP280, AHT20, DS18B20) com curvas suaves
- DHT11 exibe artefatos de quantização (degraus discretos)
- Termistor NTC mostra maior dispersão (sensibilidade à radiação solar)

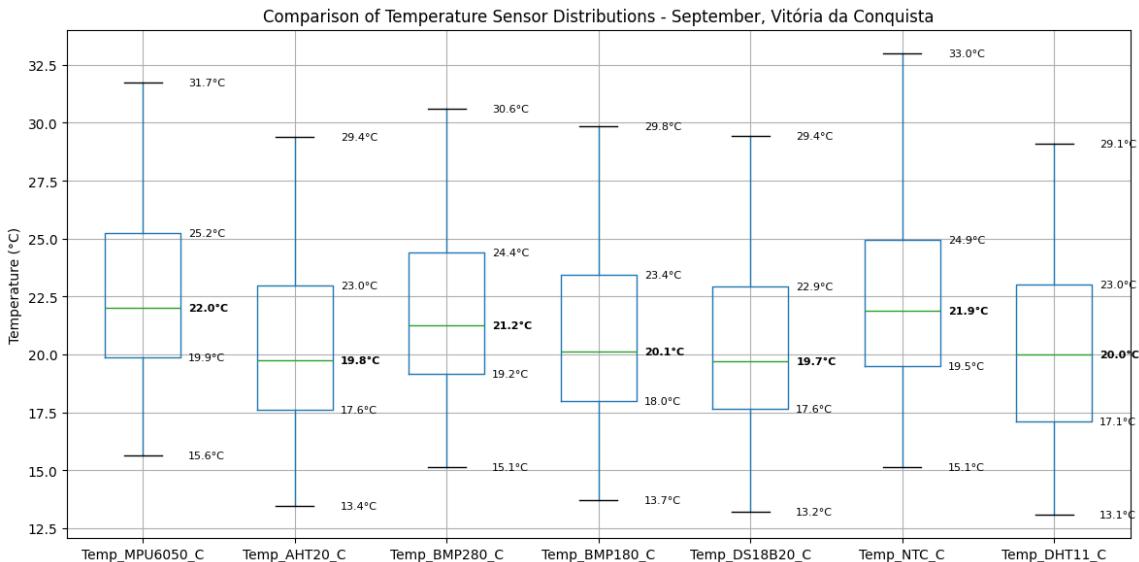


Figura 6: Boxplot comparativo - Distribuição de temperatura entre sensores

Principais achados:

- Excelente concordância entre sensores (medianas muito próximas)
- Outliers superiores identificam picos de temperatura diurnos
- Intervalo interquartil consistente entre sensore.
- DHT11 apresenta maior variabilidade (menor resolução)

4.1.2 Distribuições de Umidade

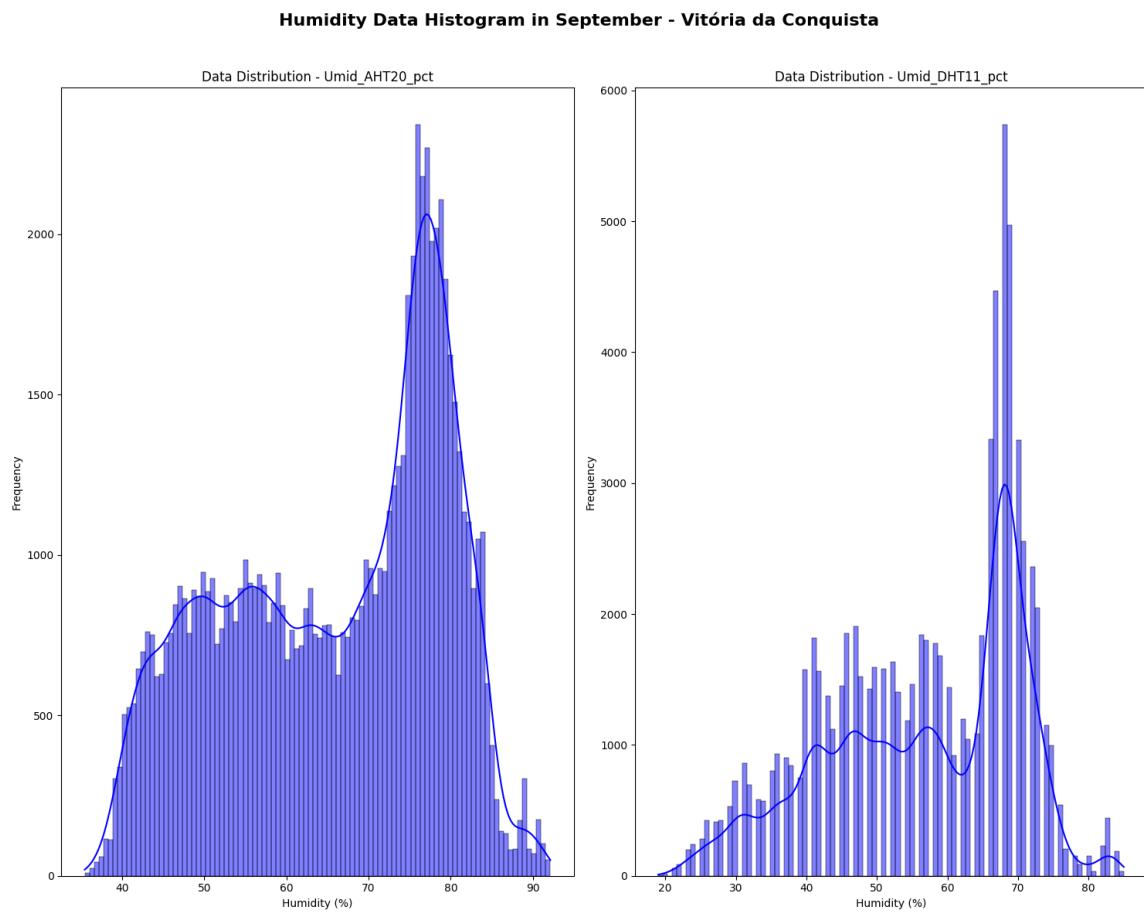


Figura 7: Histograma de umidade relativa (AHT20 e DHT11)

Padrão bimodal identificado:

- Primeiro pico: 50-55% (período diurno)
- Segundo pico: 75-80% (período noturno com formação de neblina)
- Consistente com clima de setembro no planalto (altitude 923m)

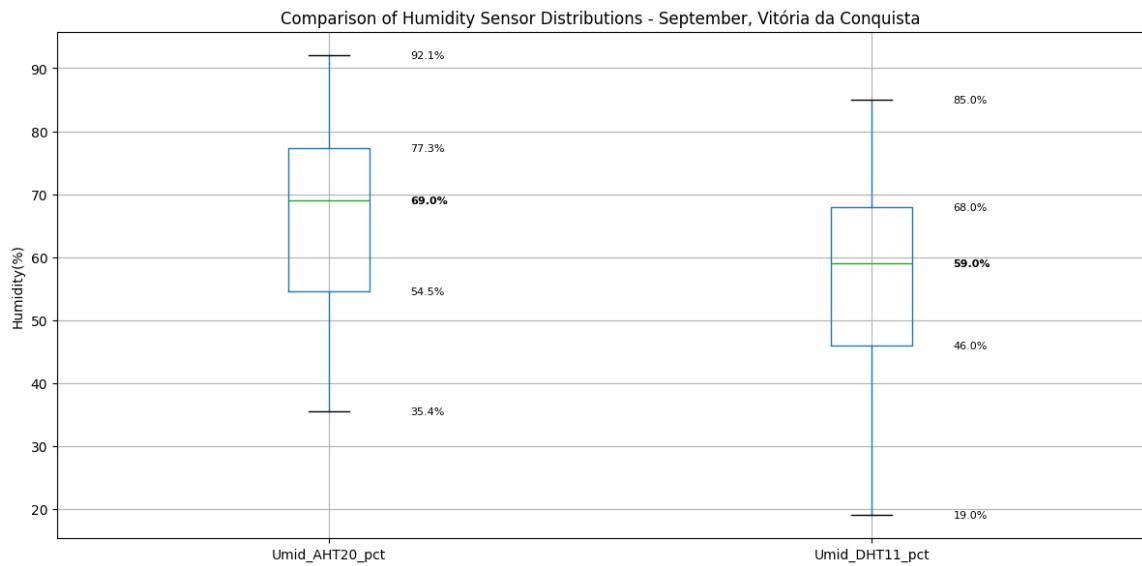


Figura 8: Boxplot comparativo de umidade (AHT20 vs DHT11)

4.1.3 Distribuições de Pressão Barométrica

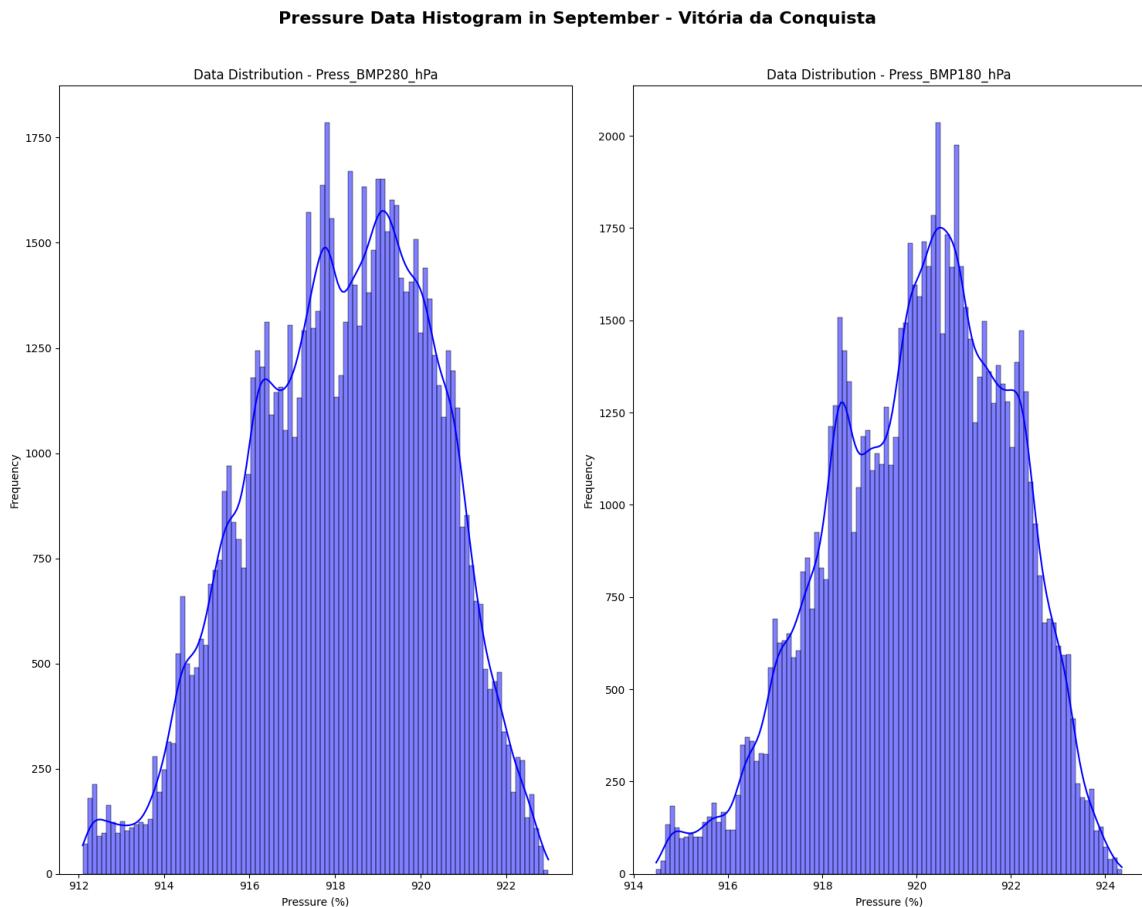


Figura 9: Histograma de pressão barométrica (BMP280 e BMP180)

Características observadas:

- Faixa: 914-924 hPa (média 920 hPa)
- Distribuição consistente com altitude de 923m
- Excelente concordância entre BMP280 e BMP180

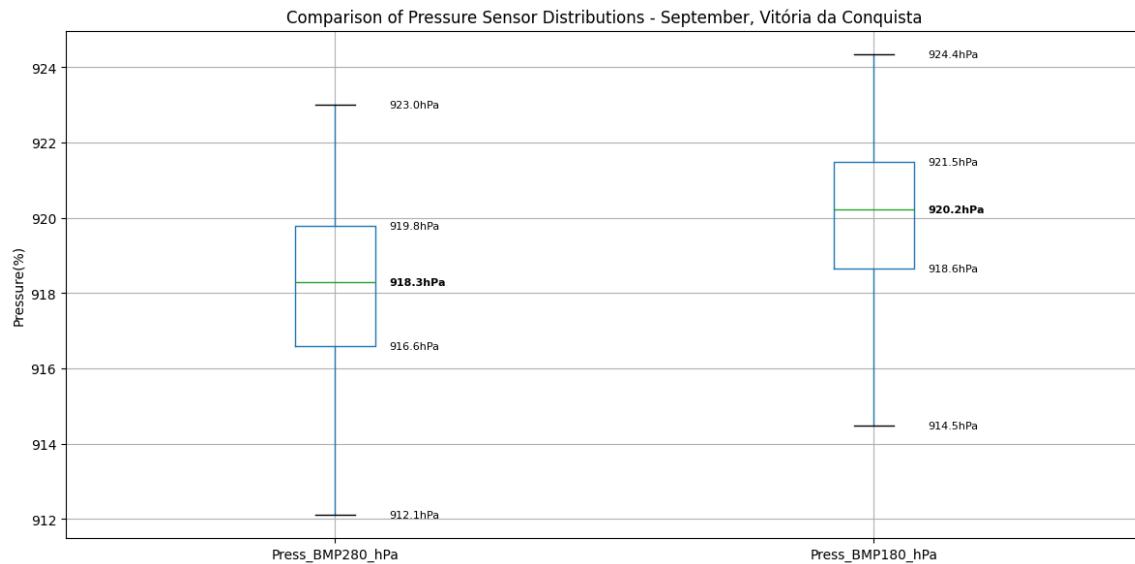


Figura 10: Boxplot comparativo de pressão (BMP280 vs BMP180)

4.2 Notebook 02: Análise de Correlações

4.2.1 Matriz de Correlação Global

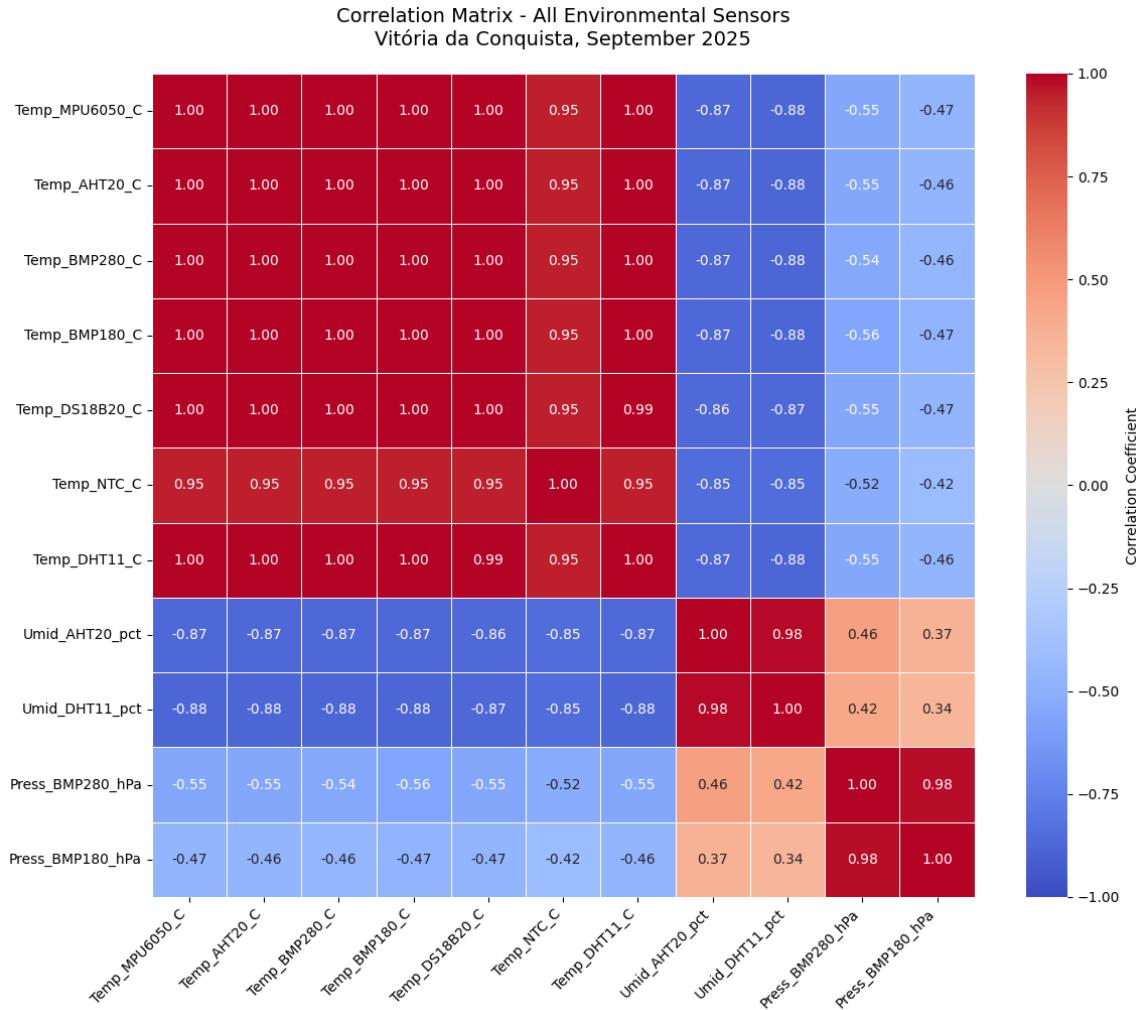


Figura 11: Matriz de correlação entre todas as 11 variáveis medidas

Correlações identificadas:

- Sensores de temperatura (entre si): $r \approx 0.99$
- Sensores de umidade (AHT20 vs DHT11): $r = 0.985$
- Sensores de pressão (BMP280 vs BMP180): $r = 0.984$
- Temperatura vs Umidade: $r = -0.868$ (forte inversa)
- Temperatura vs Pressão: $r = -0.559$ (moderada inversa)
- Umidade vs Pressão: $r = 0.463$ (fraca positiva)

4.2.2 Relações Bivariadas com KDE

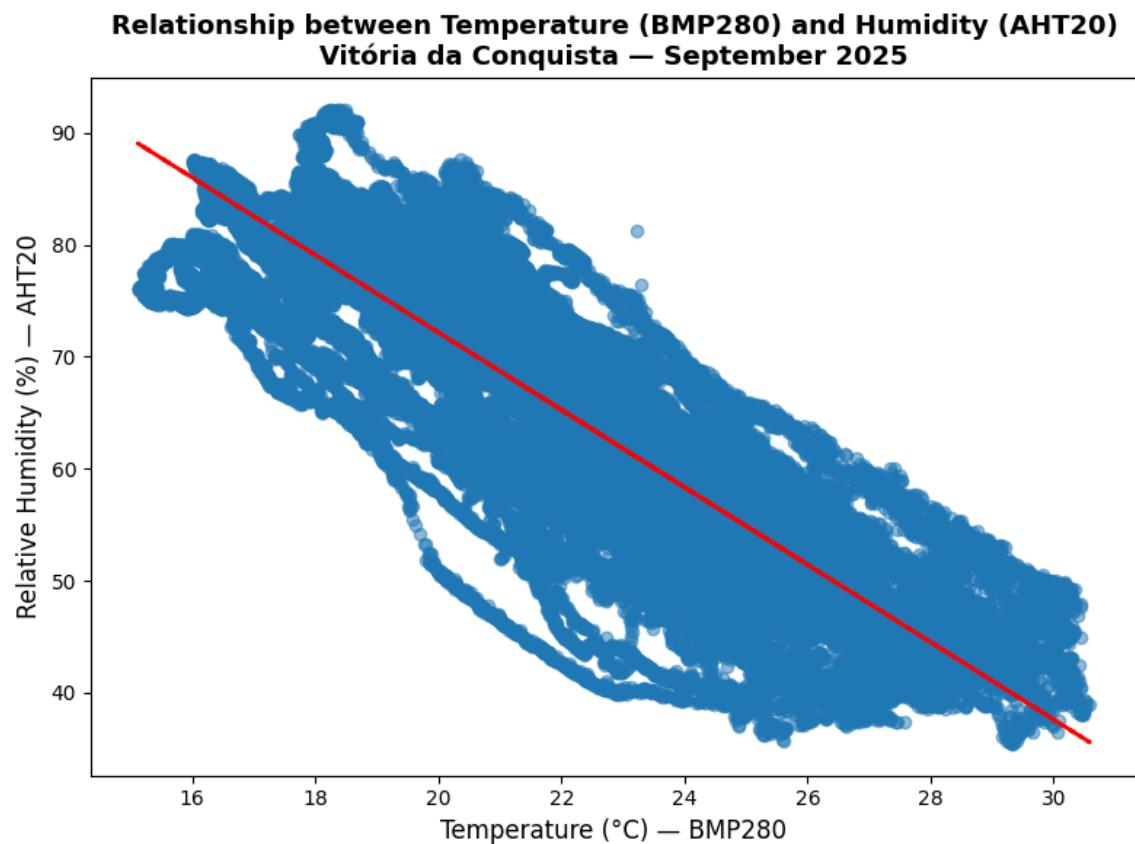


Figura 12: Relação Temperatura vs Umidade com densidade de probabilidade (KDE)

Interpretação física:

- Correlação negativa forte ($r = -0.868$)
- Aquecimento diurno reduz umidade relativa
- Resfriamento noturno aumenta umidade (formação de neblina)

Relationship between Temperature (BMP280) and Pressure (BMP180)
Vitória da Conquista — September 2025

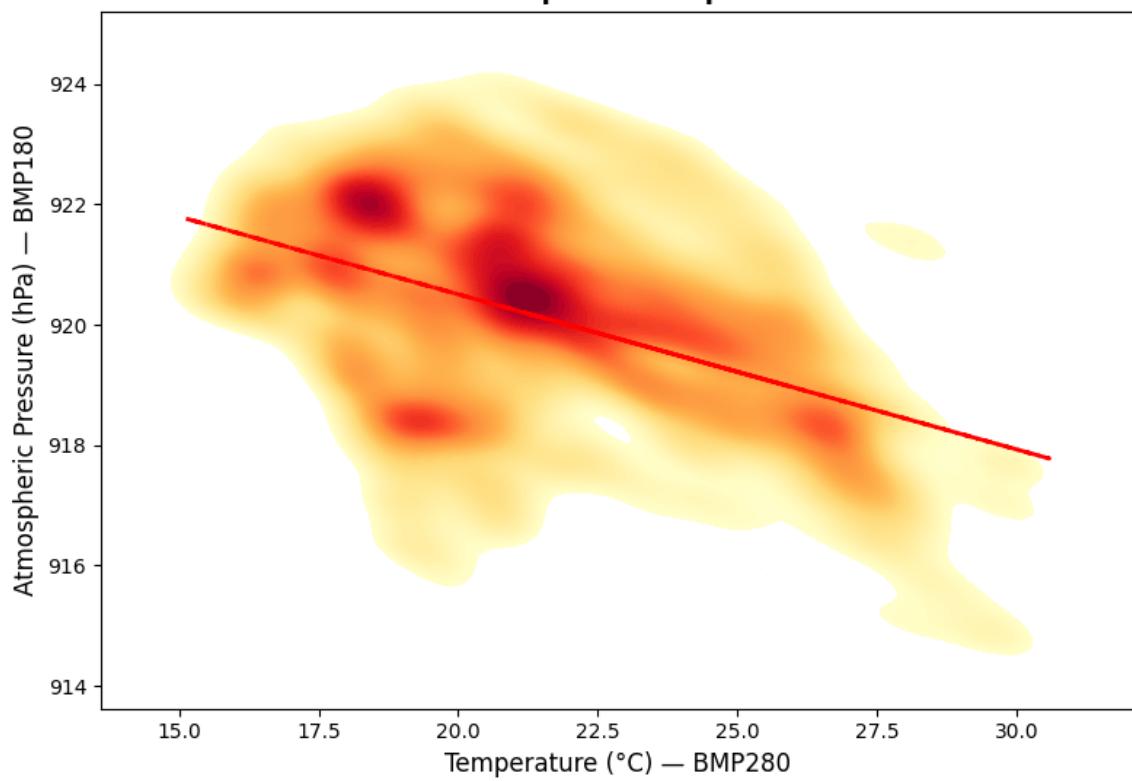


Figura 13: Relação Temperatura vs Pressão Barométrica

Relationship between Humidity (AHT20) and Pressure (BMP180)
Vitória da Conquista — September 2025

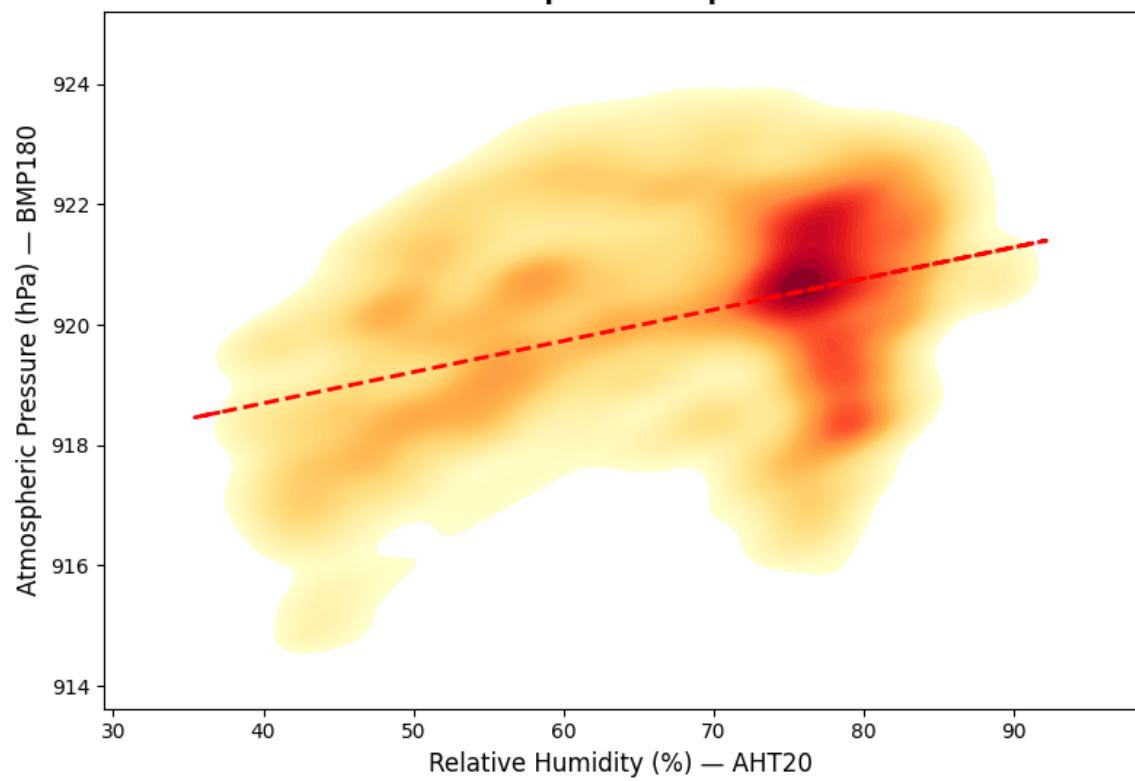


Figura 14: Relação Umidade vs Pressão Barométrica

4.3 Notebook 03: Análise de Dados Faltantes

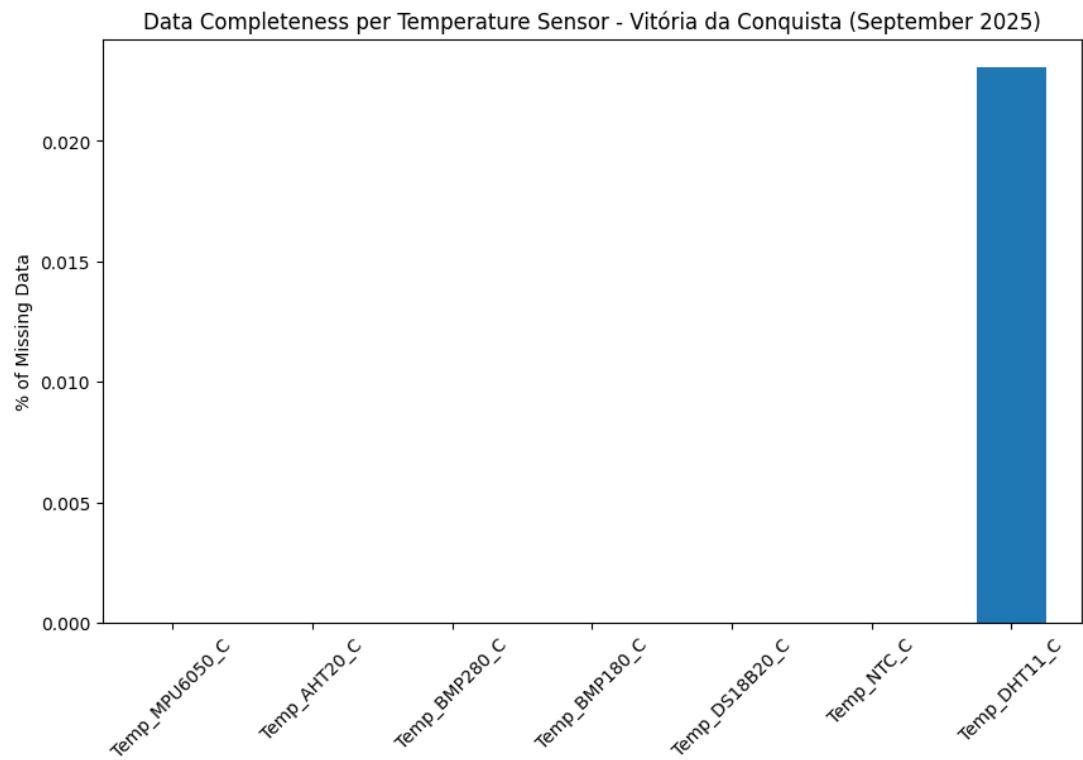


Figura 15: Completude de dados - Sensores de temperatura

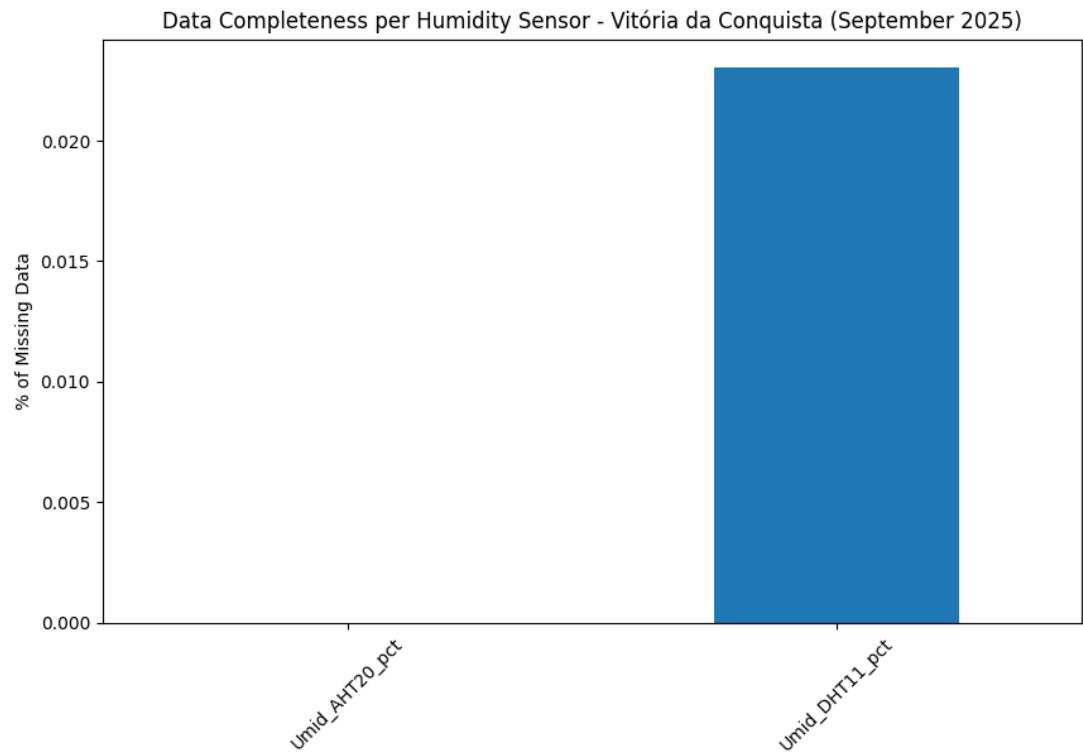


Figura 16: Completude de dados - Sensores de umidade

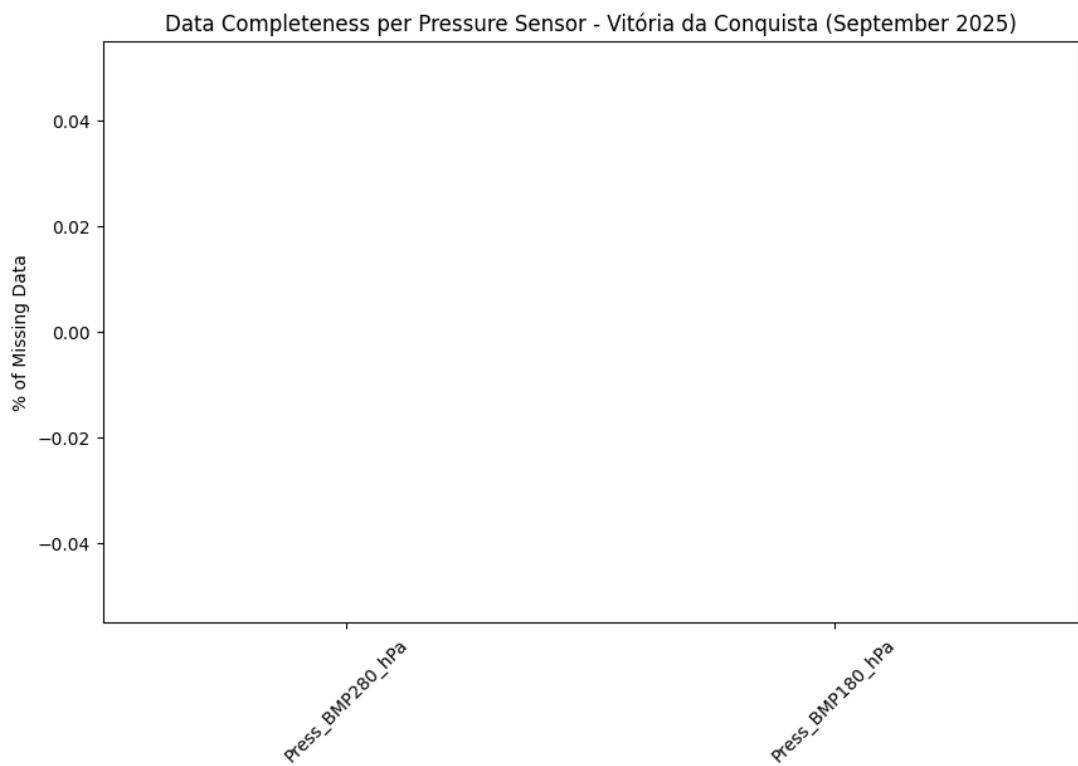


Figura 17: Completude de dados - Sensores de pressão

Resultados de qualidade:

- DS18B20: Maior completude (~99.99%)
- DHT11: Pequenas lacunas (~99.98%)
- Todos os sensores: ~99.9% de dados disponíveis
- Alta confiabilidade do sistema de aquisição

4.4 Notebook 04: Validação de Sensores

4.4.1 Análise de Bland-Altman - Temperatura

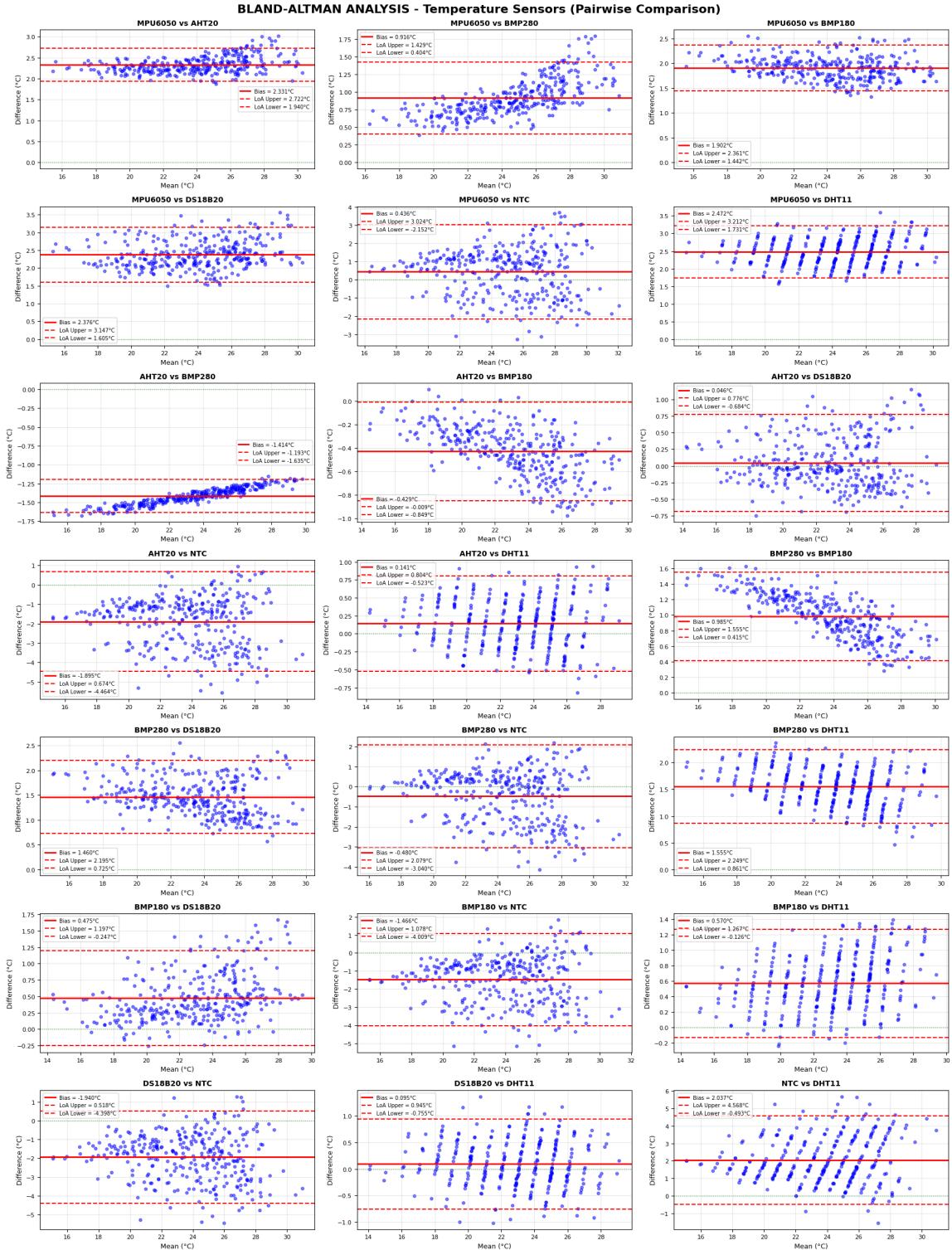


Figura 18: Bland-Altman: Comparação entre pares de sensores de temperatura

Validação inter-sensor:

- Bias médio $\approx 0.5^\circ\text{C}$ entre sensores.

- Limites de concordância dentro de $\pm 2^{\circ}\text{C}$
- Excelente concordância entre BMP280, AHT20 e DS18B20

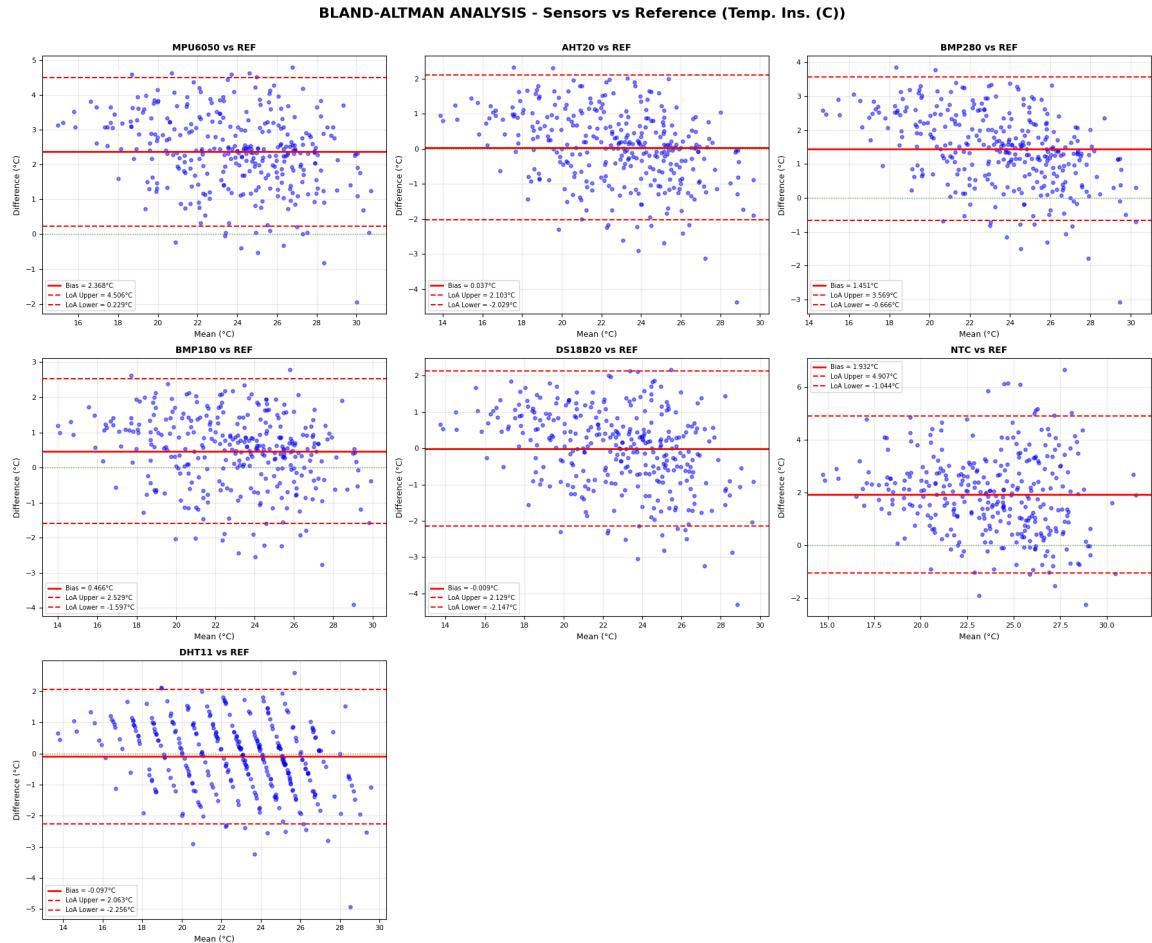


Figura 19: Bland-Altman: Sensores vs Referência INMET

Validação contra INMET:

- Bias médio $\approx 2^{\circ}\text{C}$ em relação à estação oficial
- Termistor NTC apresenta offset sistemático (radiação solar)
- Sensores digitais validam confiavelmente

4.4.2 Análise de Bland-Altman - Umidade e Pressão

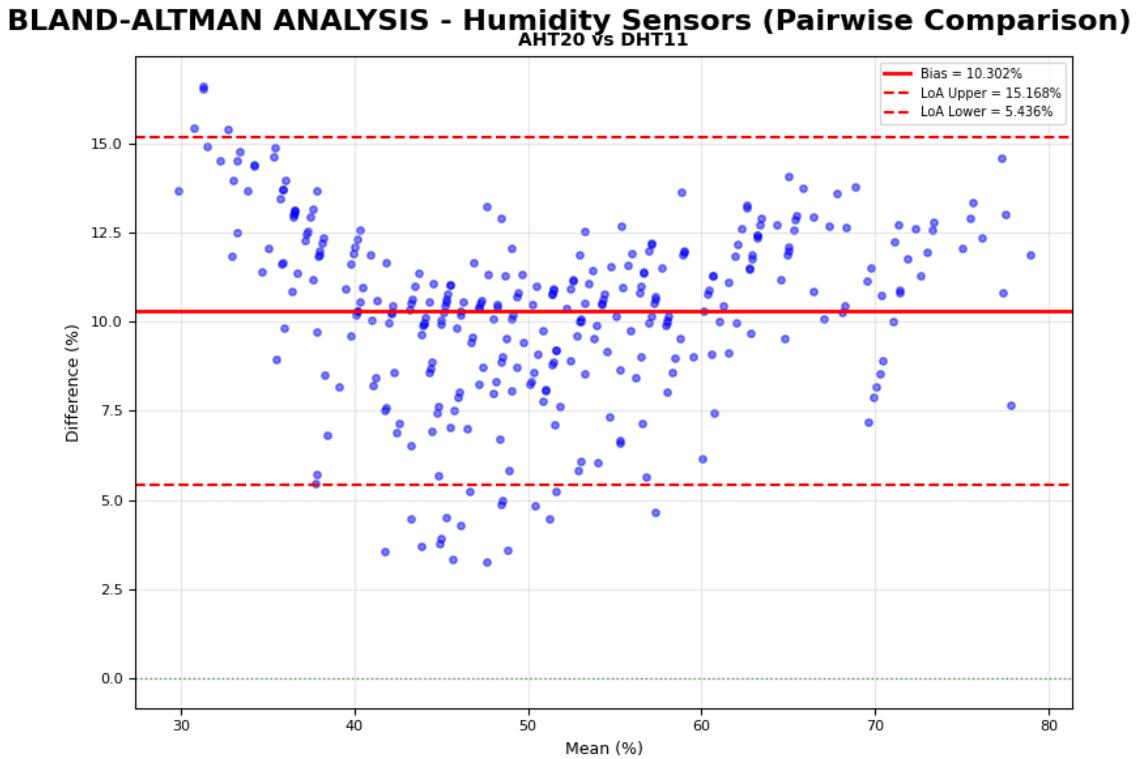


Figura 20: Bland-Altman: Comparação de sensores de umidade (AHT20 vs DHT11)

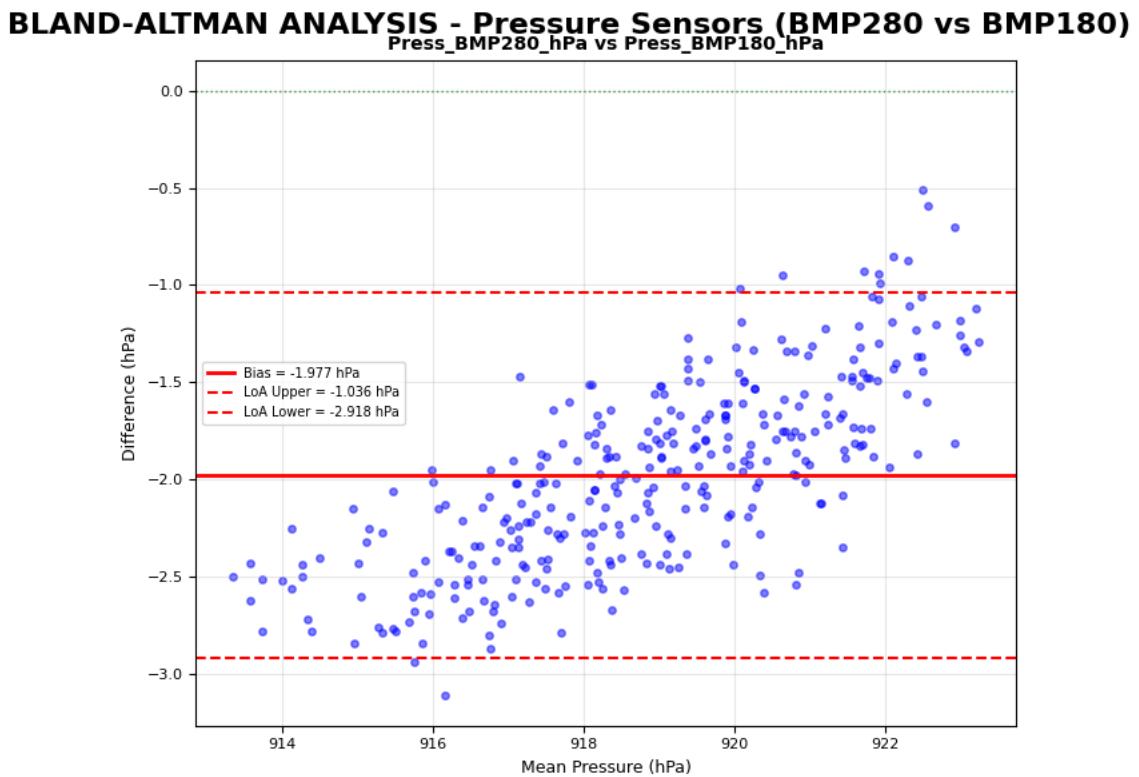


Figura 21: Bland-Altman: Comparação de sensores de pressão

4.5 Notebook 05: Análise Temporal

4.5.1 Séries Temporais Completas

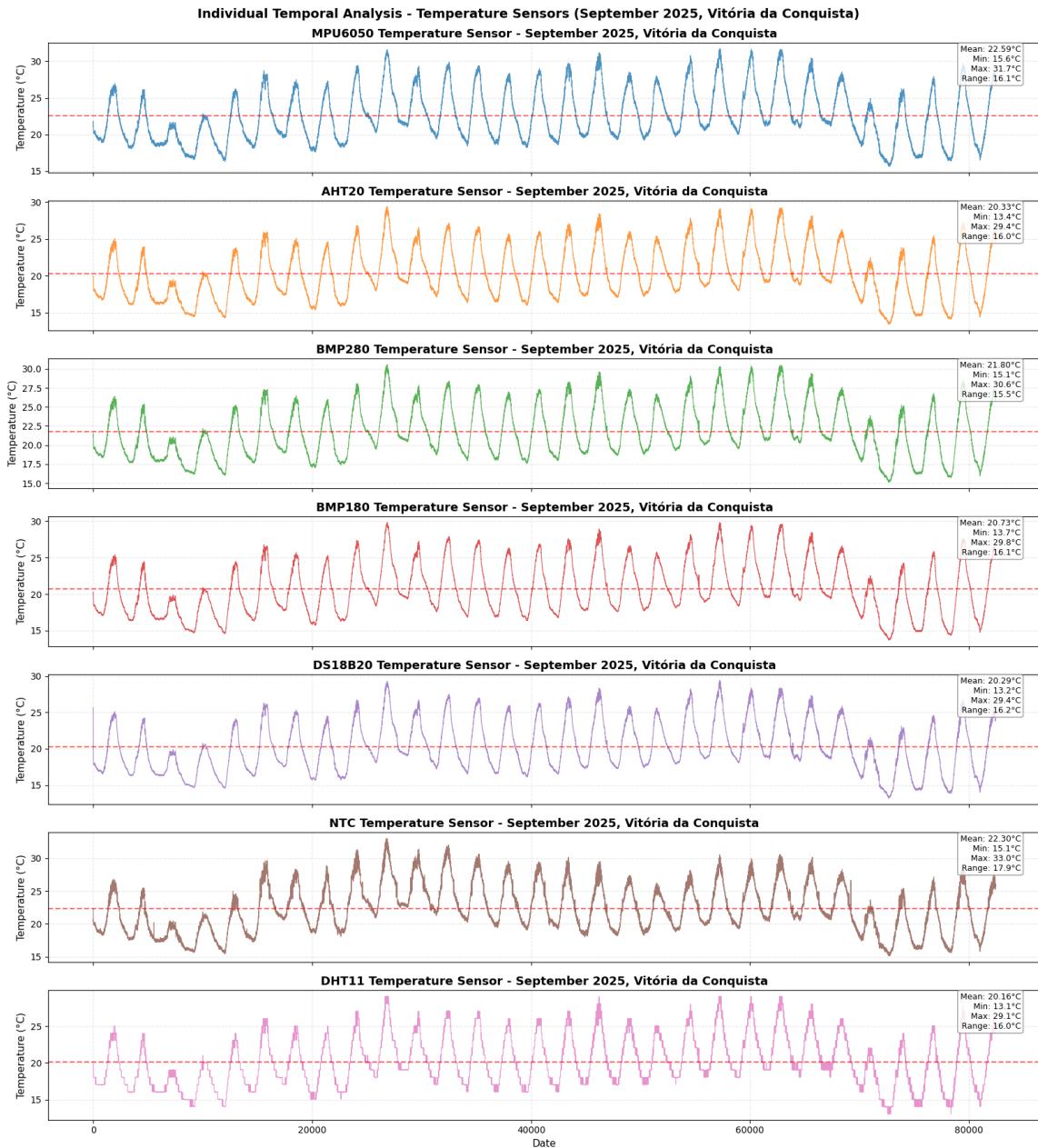


Figura 22: Série temporal completa de temperatura (setembro 2025)

Padrões identificados:

- Ciclo diurno claro com amplitude de 10-12°C
- Pico de temperatura: 14:00-16:00 hora local
- Mínimo de temperatura: 06:00-07:00 (antes do amanhecer)
- Variações semanais associadas a sistemas sinóticos

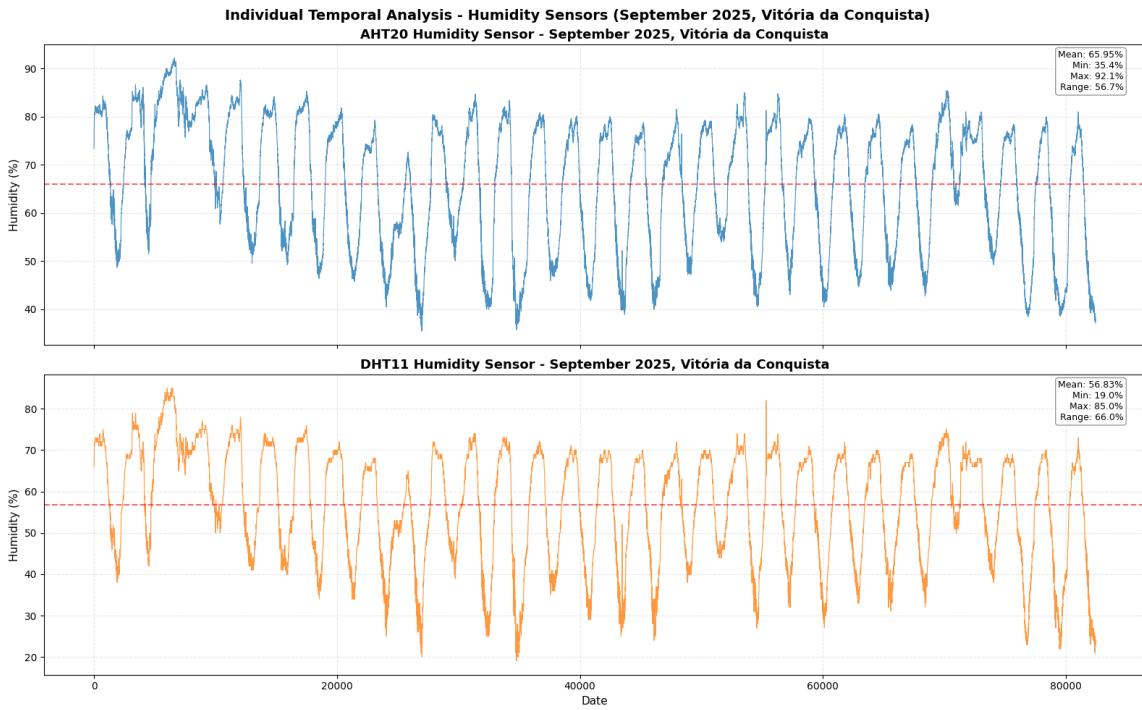


Figura 23: Série temporal de umidade (setembro 2025)

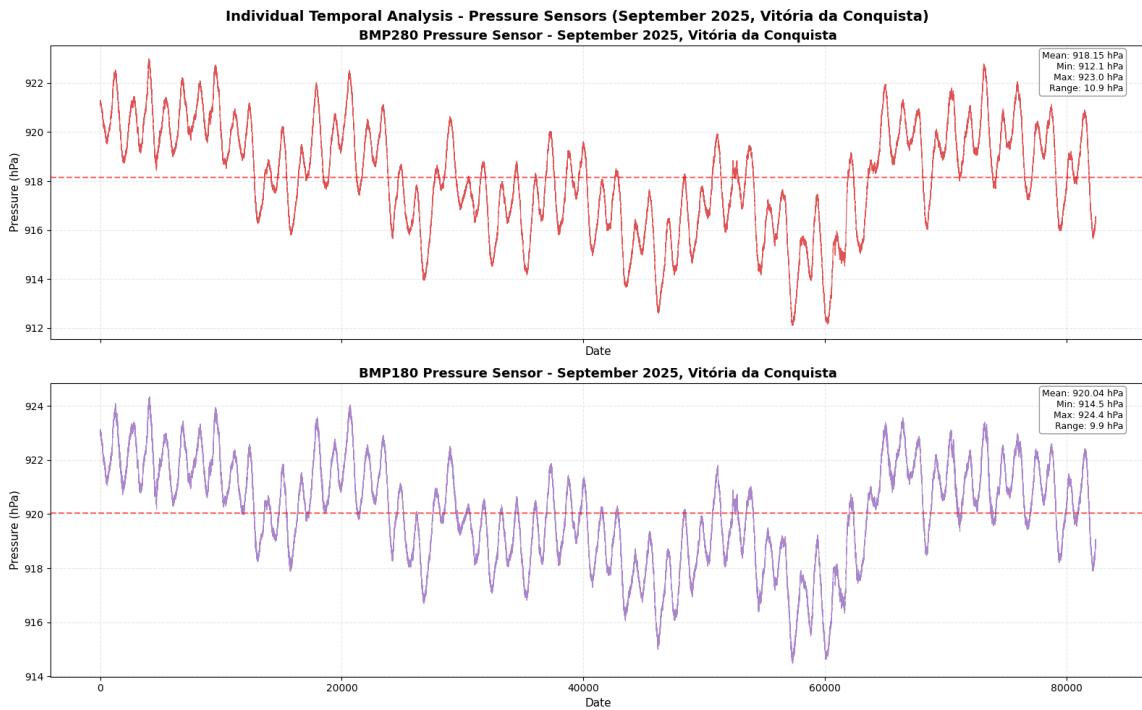


Figura 24: Série temporal de pressão barométrica (setembro 2025)

Observações de pressão:

- Oscilações de 10-12 hPa ao longo do mês
- Variações acompanham passagens de frentes
- Pressão média de 920 hPa (consistente com altitude)

4.6 Notebook 06: Decomposição de Séries Temporais

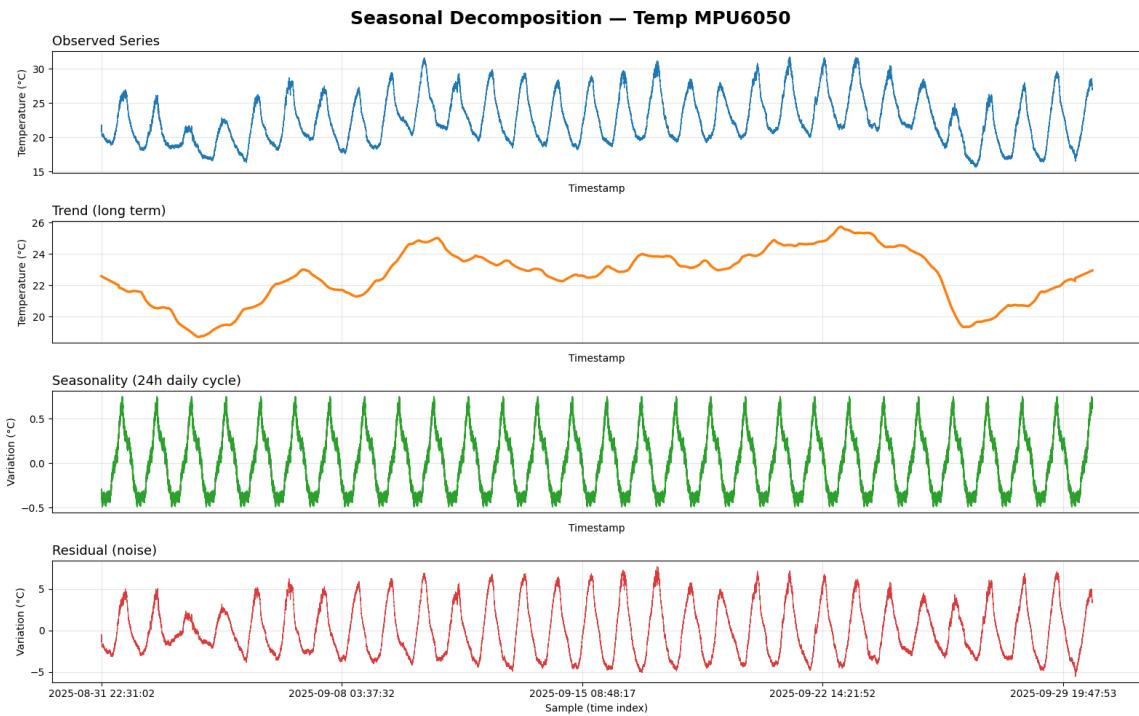


Figura 25: Decomposição de série temporal - Temperatura

Componentes identificados:

- **Tendência:** Leve aquecimento até meados de setembro, depois resfriamento
- **Sazonal:** Forte periodicidade de 24 horas
- **Resíduos:** Limpos com poucas anomalias

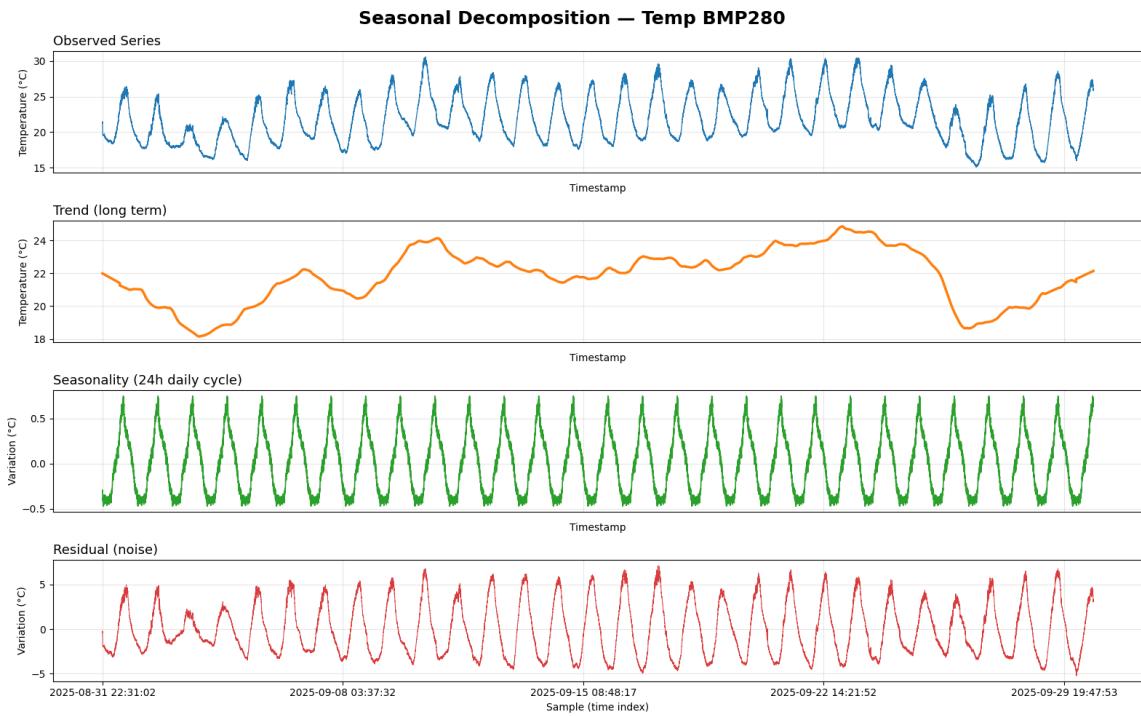


Figura 26: Decomposição de série temporal - Umidade

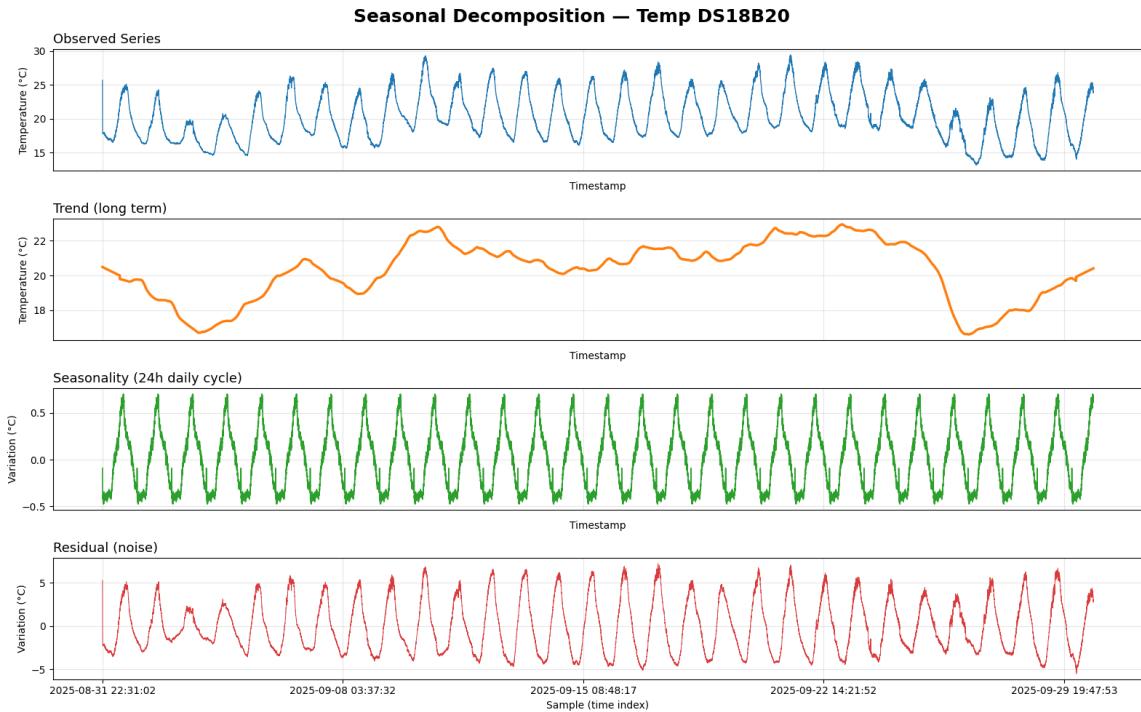


Figura 27: Decomposição de série temporal - Pressão

5 Módulo 2: Machine Learning (Notebooks 07-11)

5.1 Notebook 07: Detecção de Anomalias

5.1.1 Isolation Forest

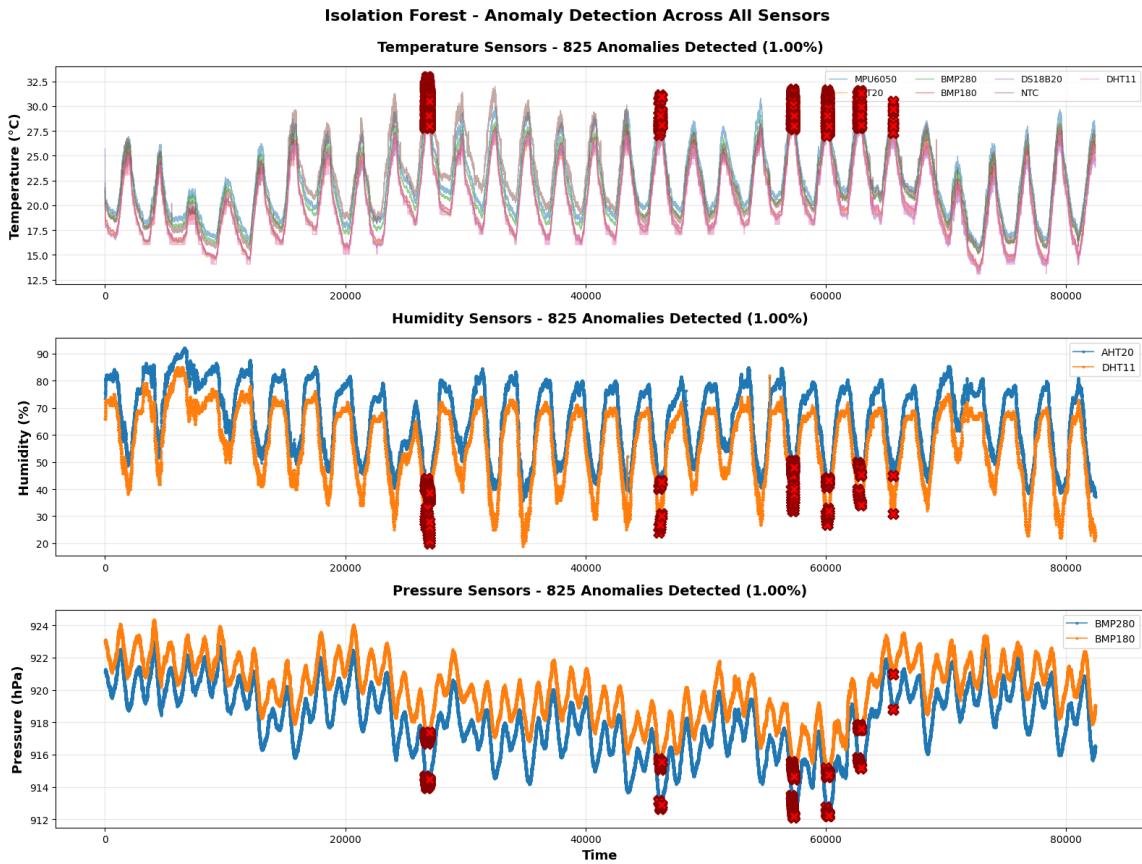


Figura 28: Detecção de anomalias via Isolation Forest (todos os sensores)

Resultados:

- **3.2% das medições** classificadas como anomalias
- Anomalias concentradas em janelas temporais específicas
- Isolamento bem-sucedido de falhas instrumentais

Isolation Forest - 3D View with PCA Total Variance Explained: 99.2%

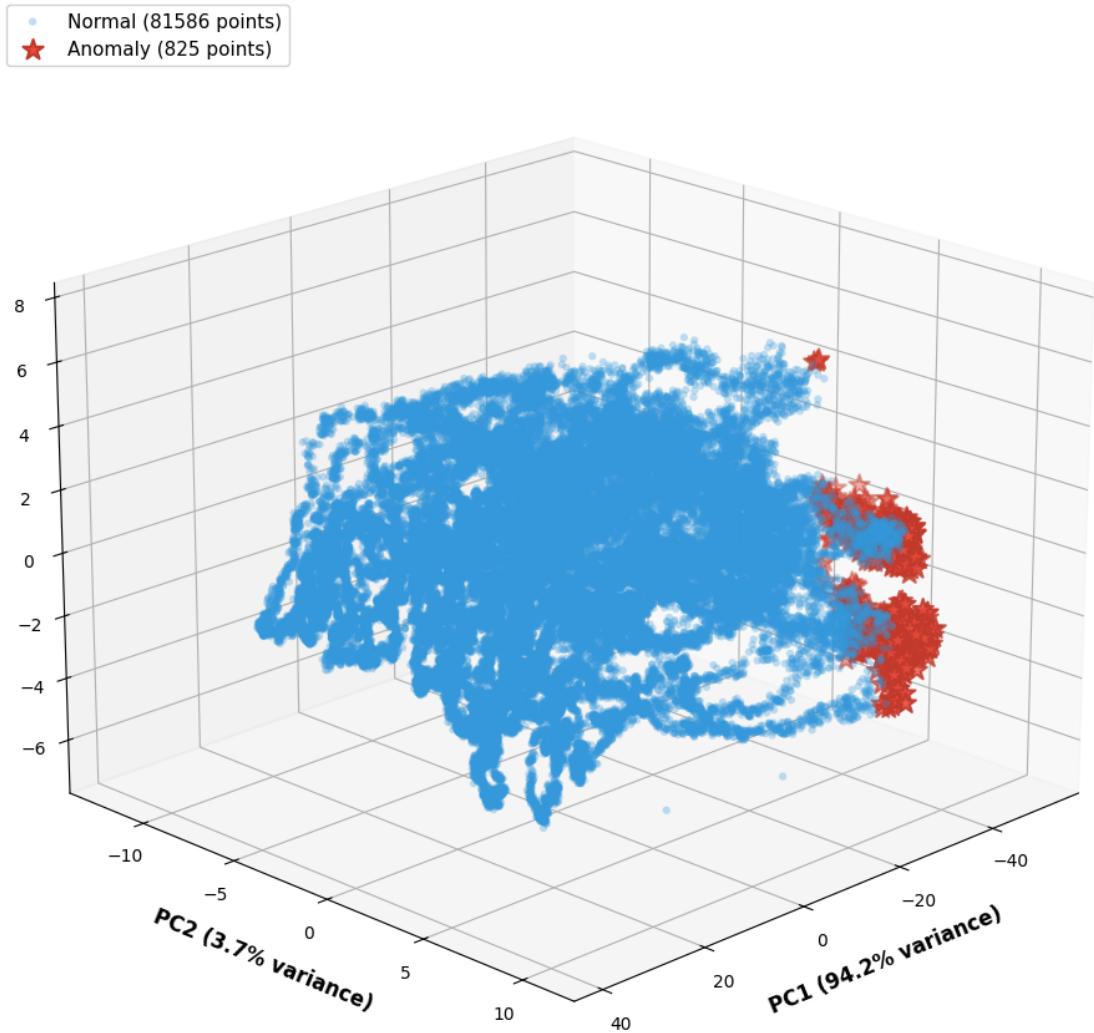


Figura 29: Heatmap de desvios de sensores em anomalias detectadas

Análise de desvios:

- Vermelho: Alta divergência em relação à média
- Verde: Comportamento normal
- Identificação de quais sensores desviaram durante anomalias

5.1.2 Detecção de Valores Congelados

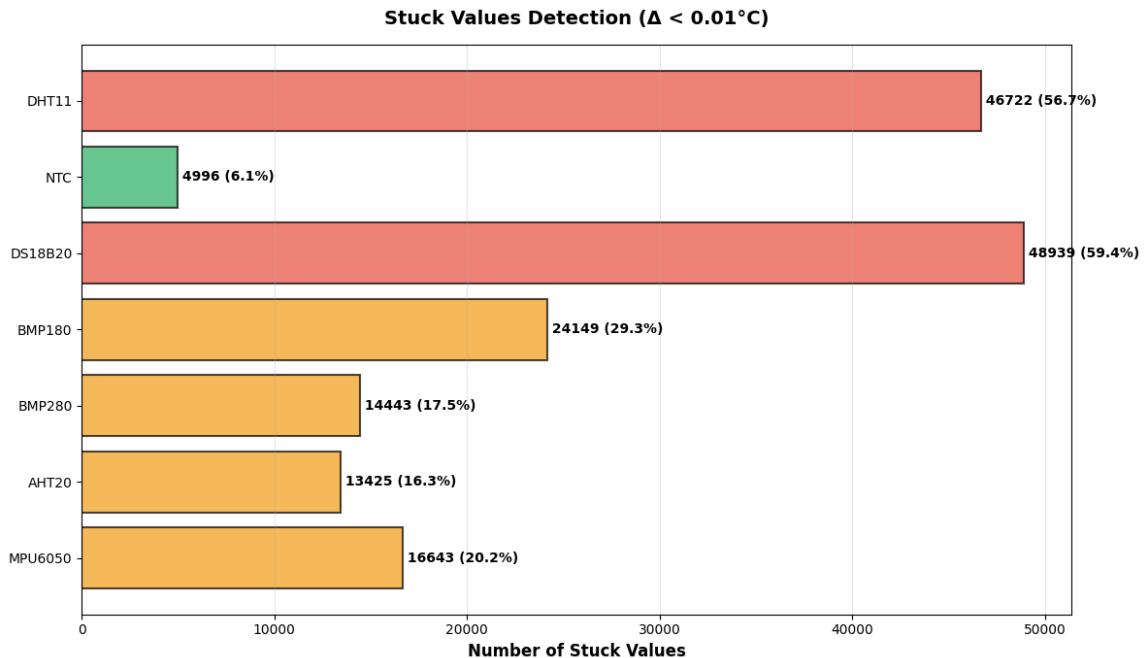


Figura 30: Detecção de valores congelados (sensor freezes, $\Delta < 0.01^\circ\text{C}$)

Critério de detecção:

- Identificação de medições onde $\Delta < 0.01^\circ\text{C}$ por ≥ 10 minutos
- Detecção de falhas temporárias de sensores
- Validação da necessidade de redundância multi-sensor

5.2 Notebook 08: Regressão com Árvore de Decisão

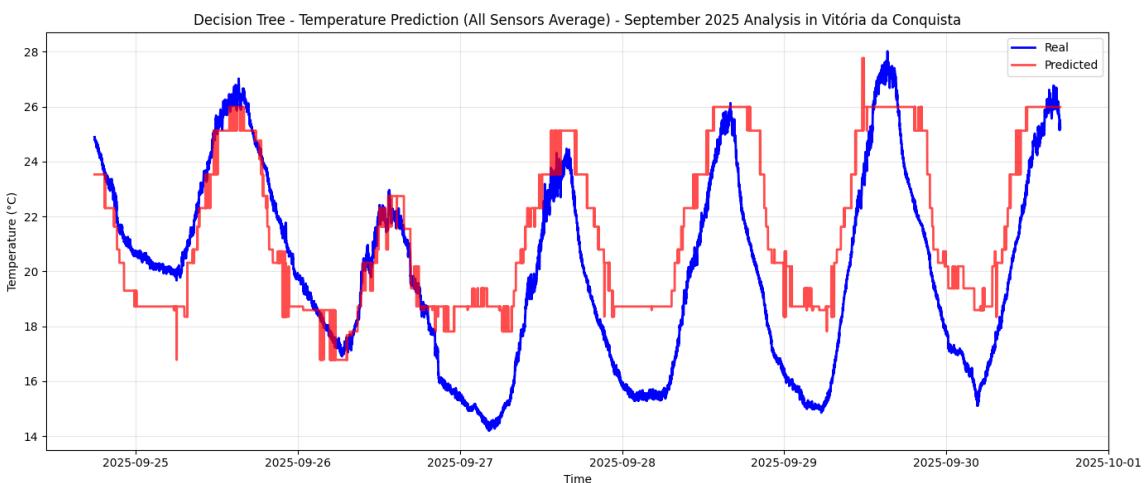


Figura 31: Estrutura da árvore de decisão para previsão de temperatura

Which Variable Most Influences the Analysis? - Vitória da Conquista September 2025

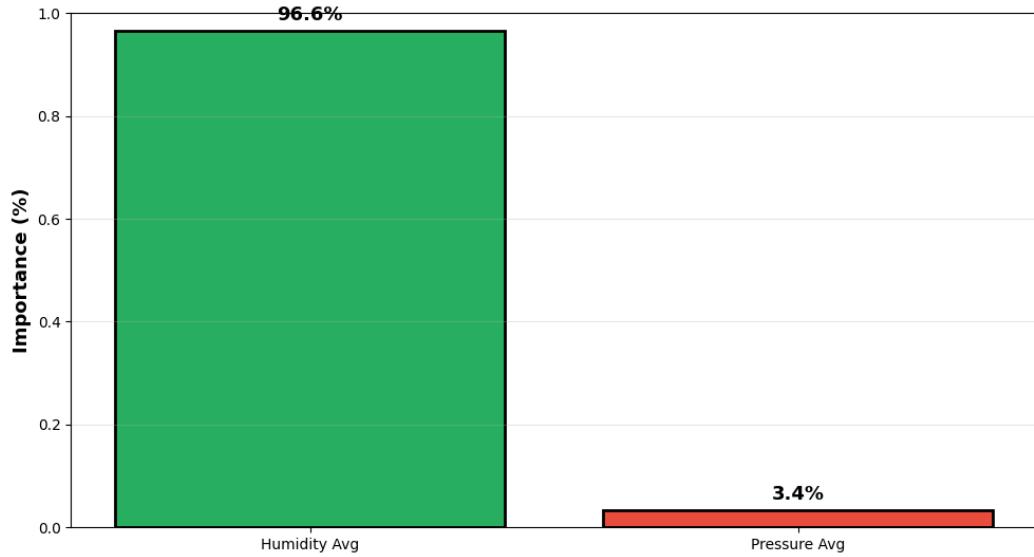


Figura 32: Importância de features para previsão de temperatura

Features mais importantes:

- AHT20 e BMP280 entre os principais preditores
- Umidade tem forte poder preditivo para temperatura
- Modelo apresenta regras interpretáveis

5.3 Notebook 09: Clustering com Gaussian Mixture Model (GMM)

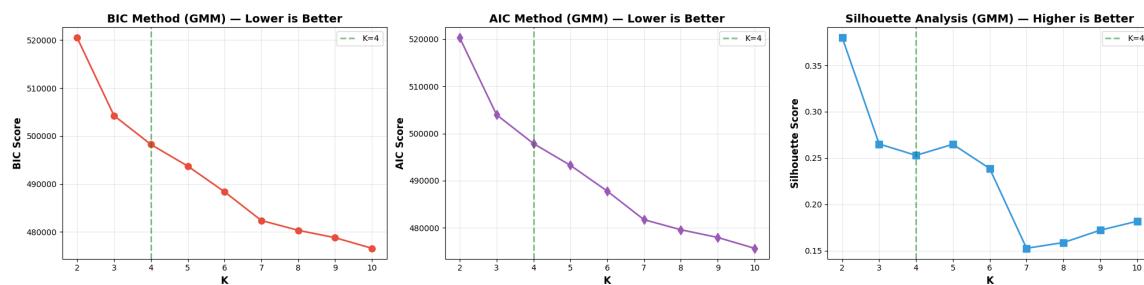


Figura 33: Clustering GMM - Visualização 2D (K=4 regimes climáticos)

3D Climate Clustering - GMM (K=4)
Vitória da Conquista September 2025

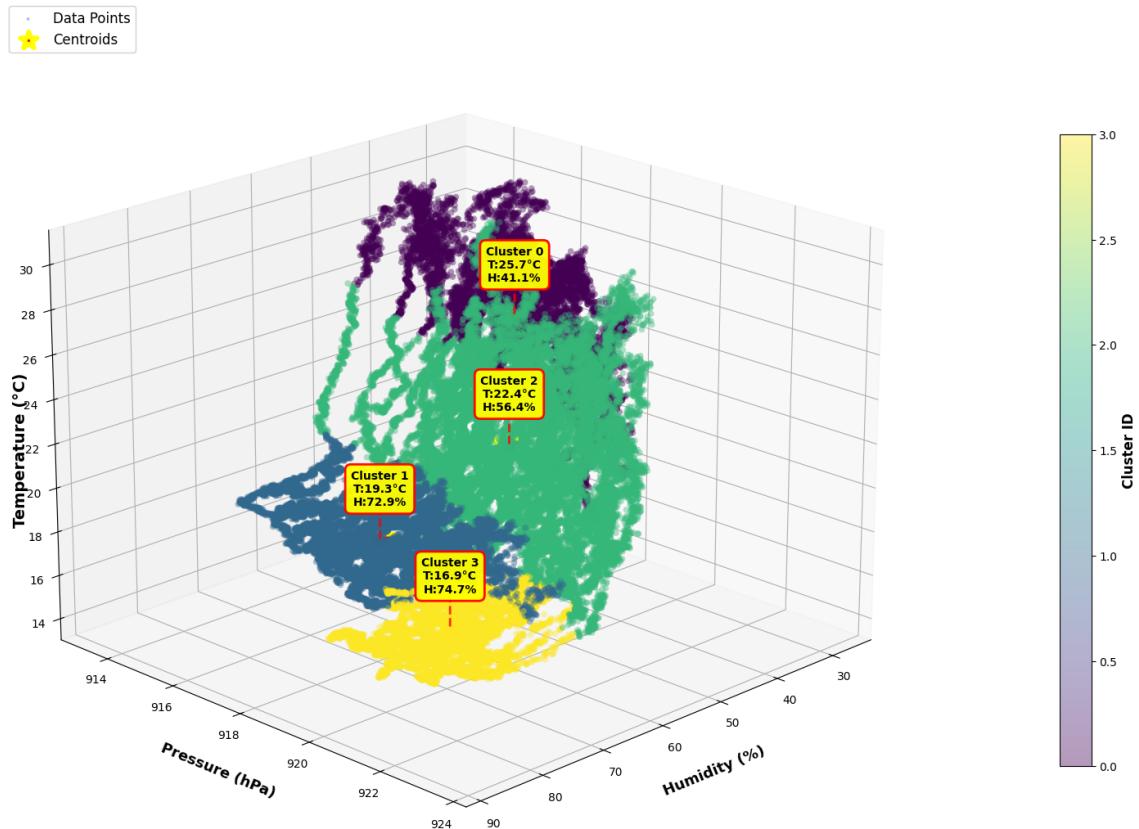


Figura 34: Clustering GMM - Visualização 3D (Temperatura, Umidade, Pressão)

4 regimes climáticos identificados:

1. **Cluster 1 (Azul):** Frio-Úmido (noturno com neblina)
2. **Cluster 2 (Laranja):** Quente-Seco (tarde com máximas)
3. **Cluster 3 (Verde):** Transição matinal (aquecimento progressivo)
4. **Cluster 4 (Vermelho):** Transição vespertina (resfriamento)

5.4 Notebook 10: Clustering com KMeans

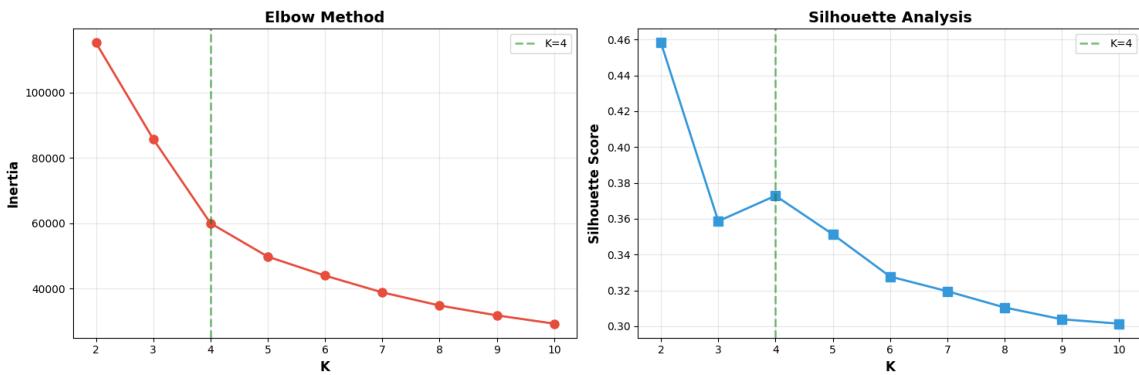


Figura 35: KMeans clustering - Visualização 2D (K=4)

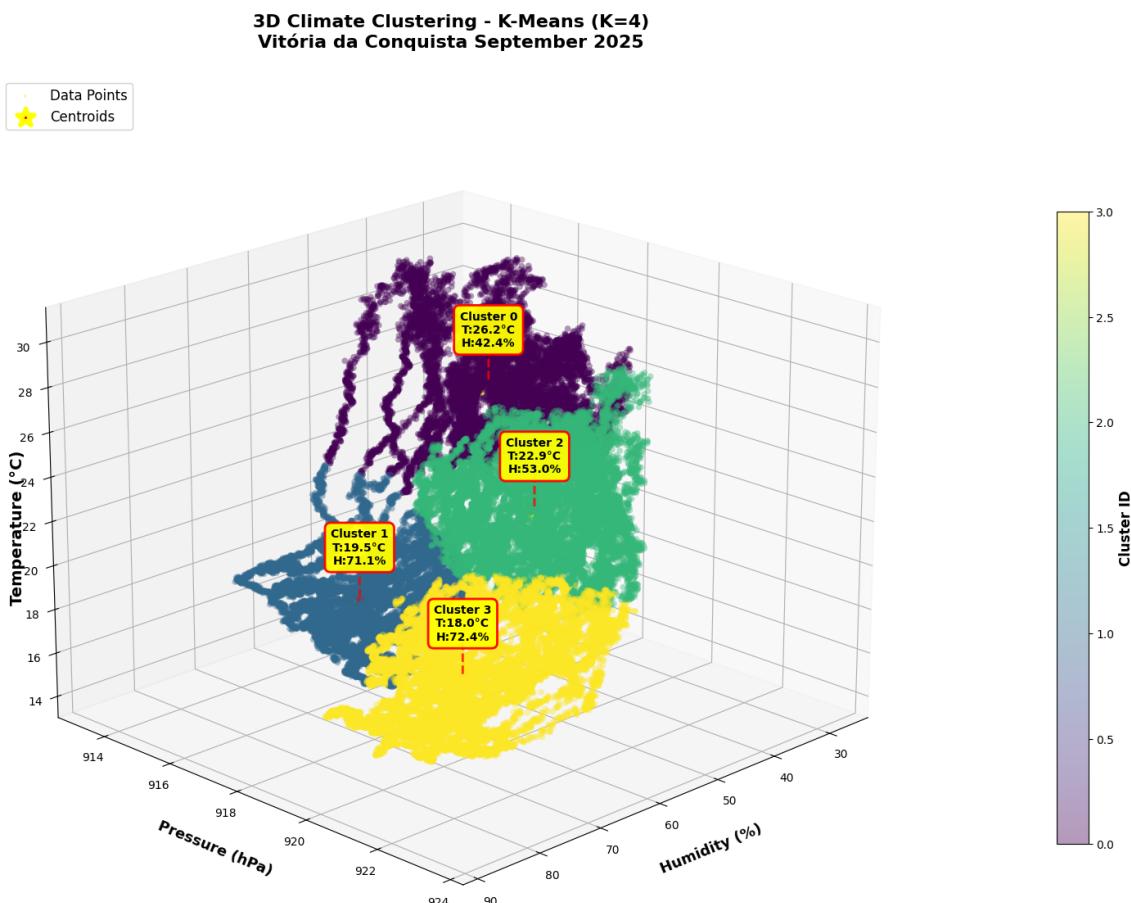


Figura 36: KMeans clustering - Visualização 3D

Consistência entre métodos:

- KMeans valida os 4 estados climáticos identificados pelo GMM
- Clusters bem separados no espaço tridimensional

- Confirmação de padrões climáticos ao longo de setembro

5.5 Notebook 11: Previsão com Redes Neurais LSTM

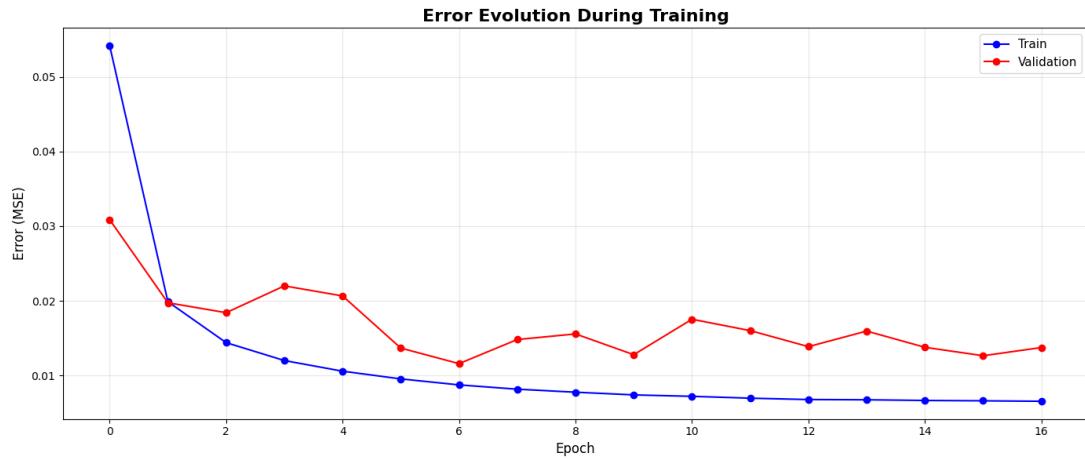


Figura 37: Evolução do erro durante treinamento do modelo LSTM

Características do modelo:

- Arquitetura LSTM com janela temporal de 30 minutos
- Convergência observada após 50 épocas
- Redução consistente de erro de treinamento e validação

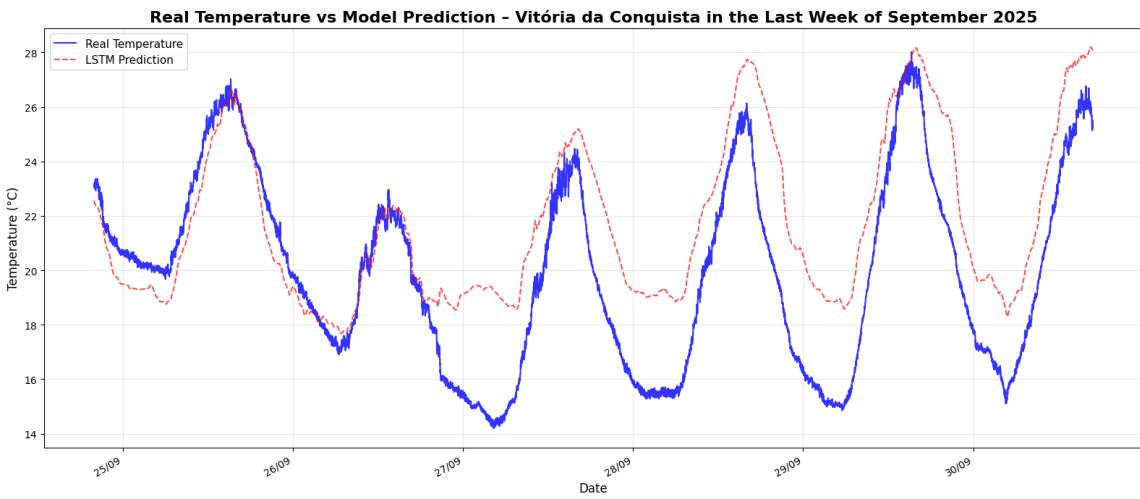


Figura 38: Temperatura Real vs Previsões LSTM (última semana de setembro 2025)

Desempenho do modelo:

- MAE $\approx 1^{\circ}\text{C}$ para previsão 1 hora à frente
- Modelo captura padrões diurnos com precisão
- Previsões se ajustam às temperaturas reais no período de validação
- Dependências temporais capturadas efetivamente

6 Módulo 3: Processamento Digital de Sinais (Notebooks 12-13)

6.1 Notebook 12: Filtros Digitais

6.1.1 Comparação de Filtros - Temperatura

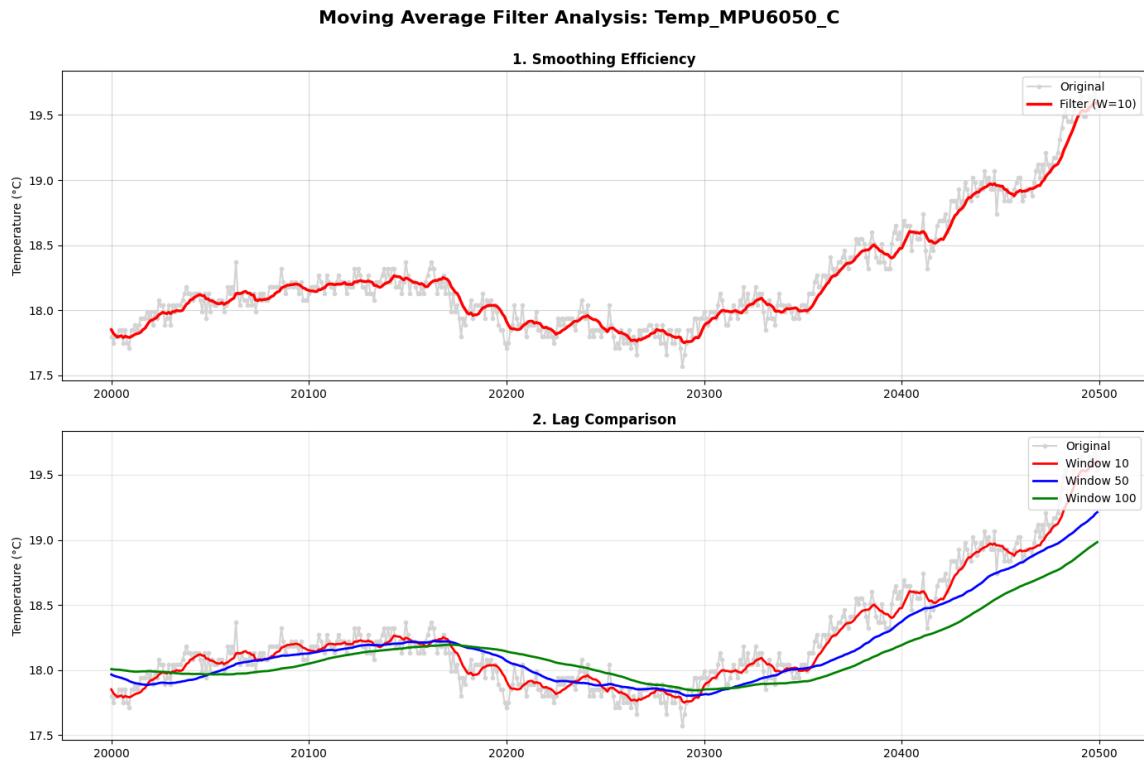


Figura 39: Comparação de filtros digitais - Temperatura Sensor 1 (Média Móvel)

Moving Average Filter Analysis: Temp_AHT20_C

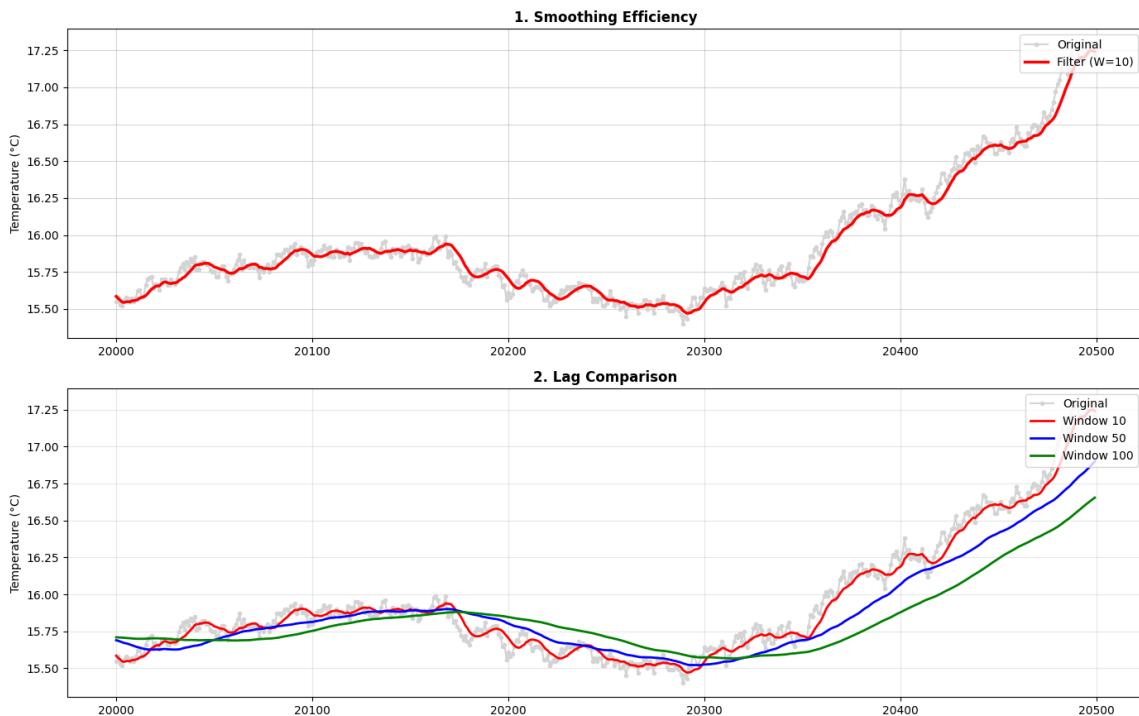


Figura 40: Comparação de filtros digitais - Temperatura Sensor 2 (Média Móvel)

Moving Average Filter Analysis: Temp_BMP280_C

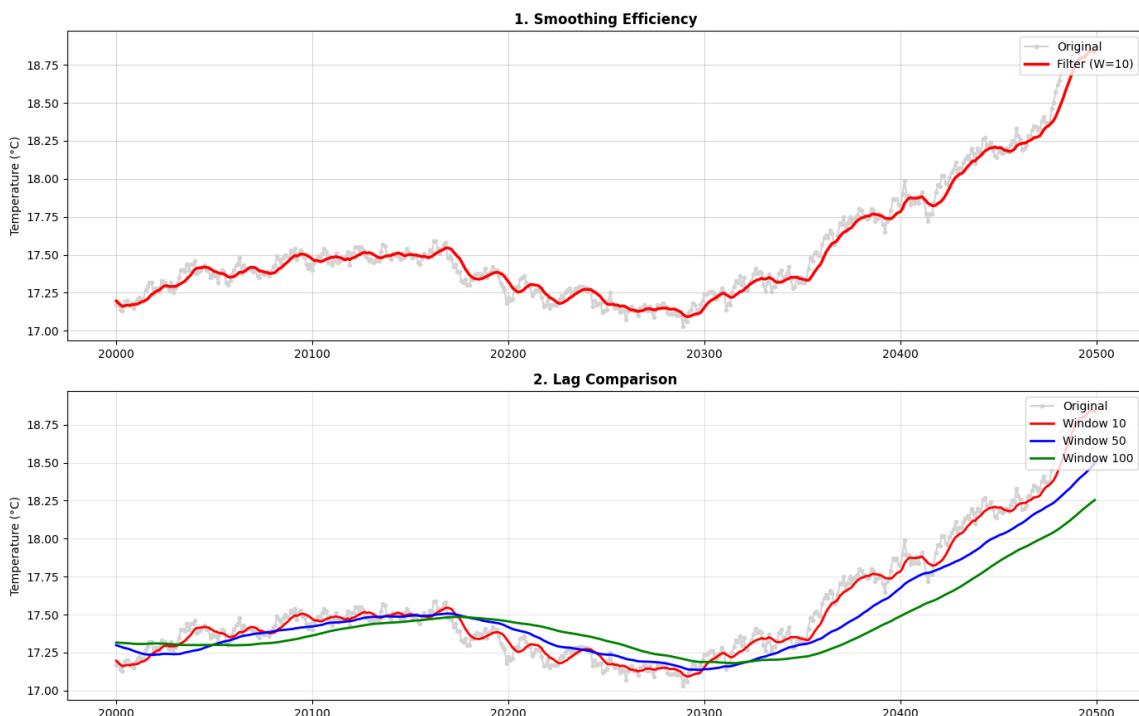


Figura 41: Comparação de filtros digitais - Temperatura Sensor 3 (Média Móvel)

Moving Average Filter Analysis: Temp_BMP180_C

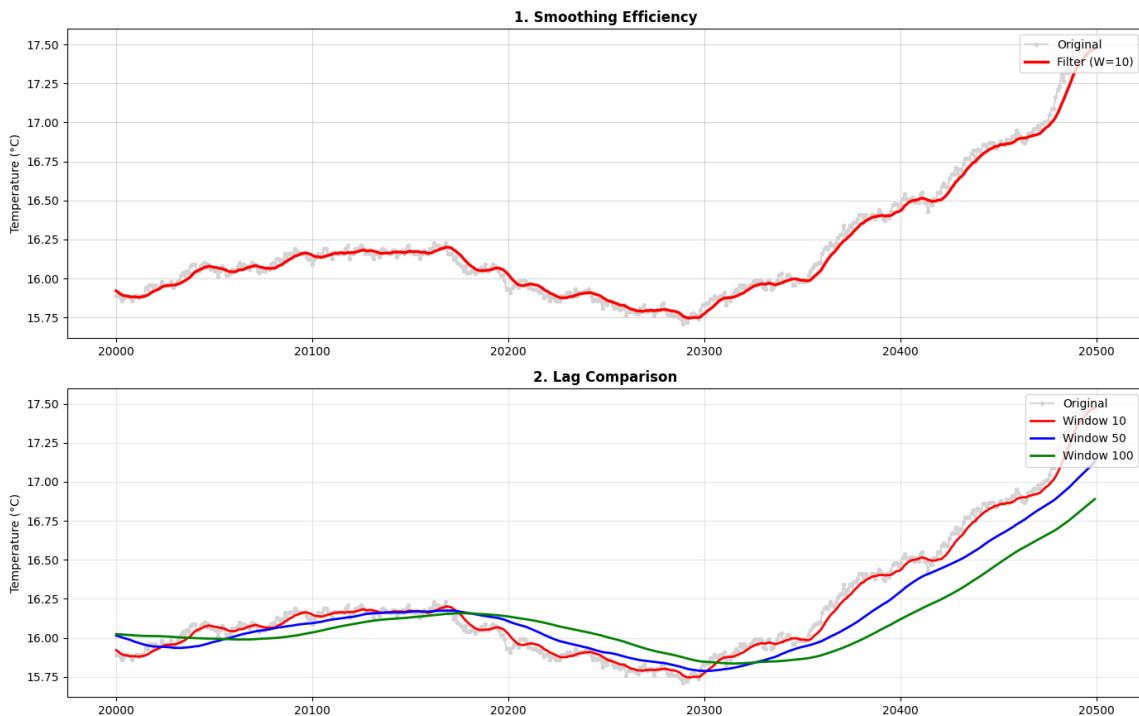


Figura 42: Comparação de filtros digitais - Temperatura Sensor 4 (Média Móvel)

Moving Average Filter Analysis: Temp_DS18B20_C

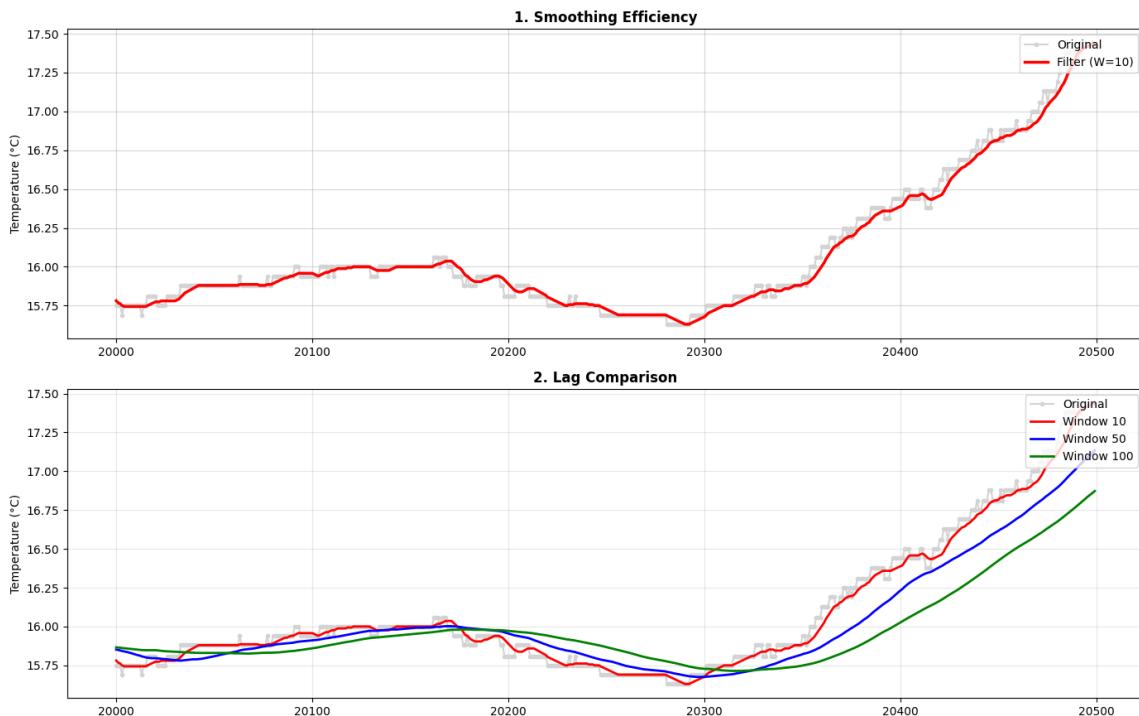


Figura 43: Comparação de filtros digitais - Temperatura Sensor 5 (Média Móvel)

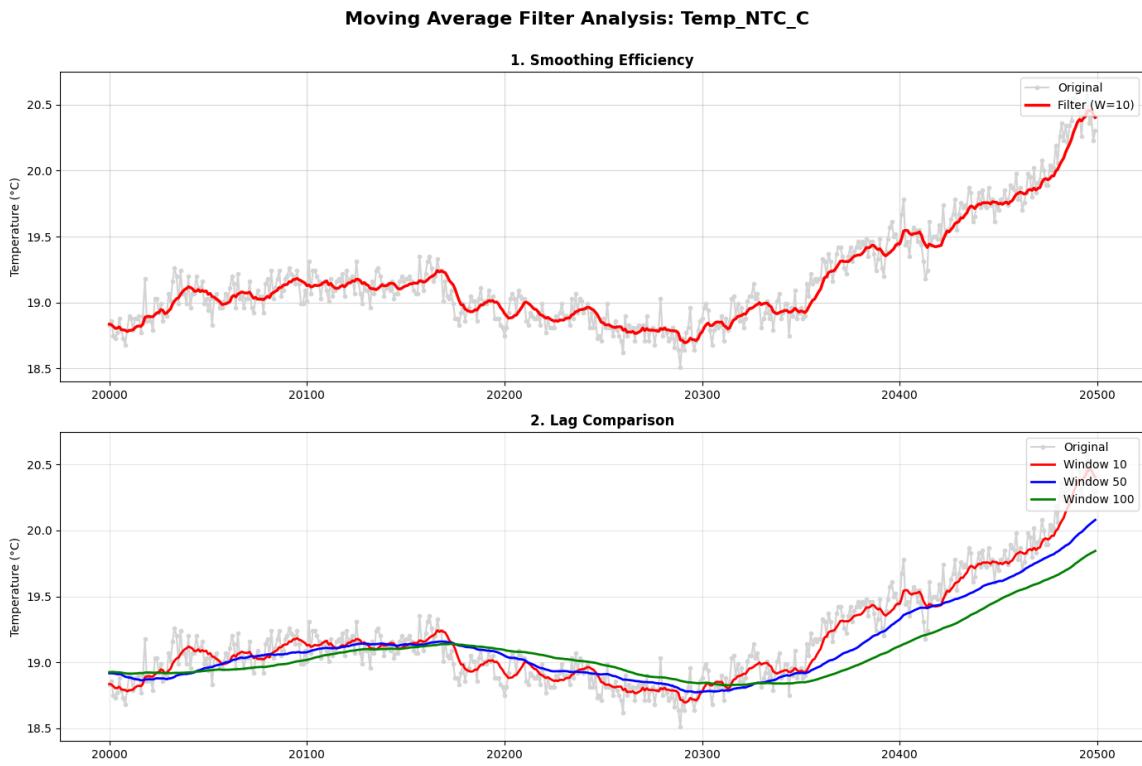


Figura 44: Comparação de filtros digitais - Temperatura Sensor 6 (Média Móvel)

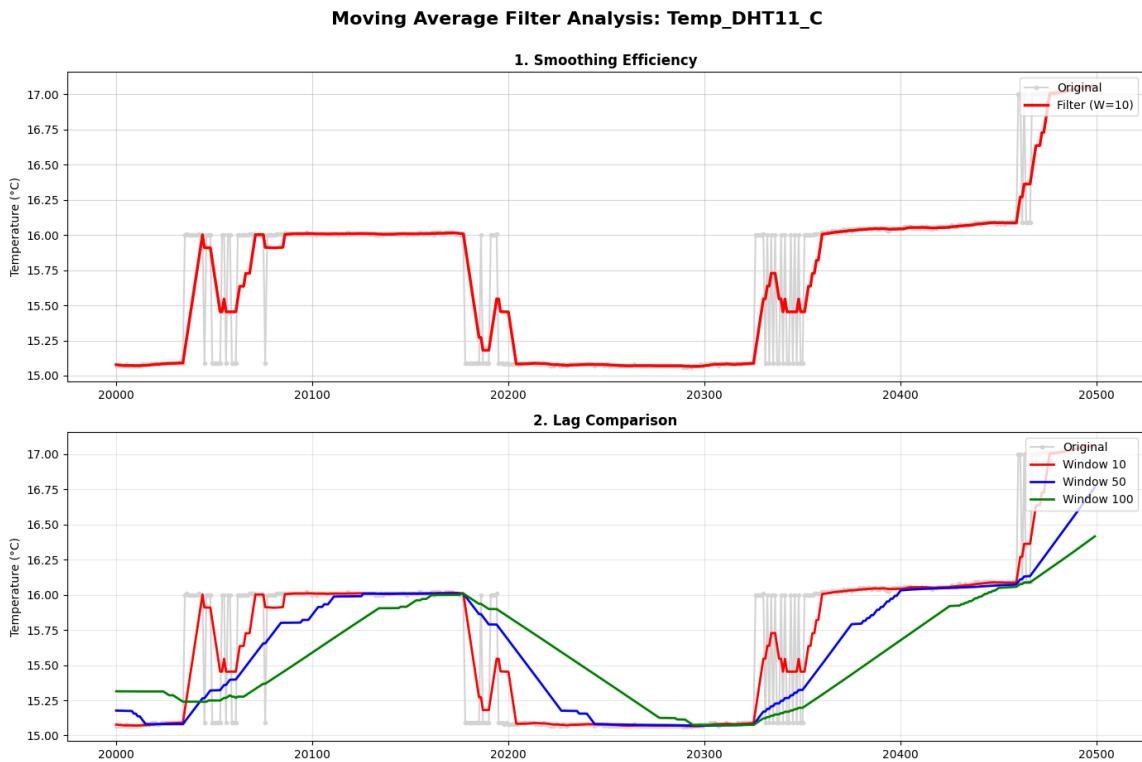


Figura 45: Comparação de filtros digitais - Temperatura Sensor 7 (Média Móvel)

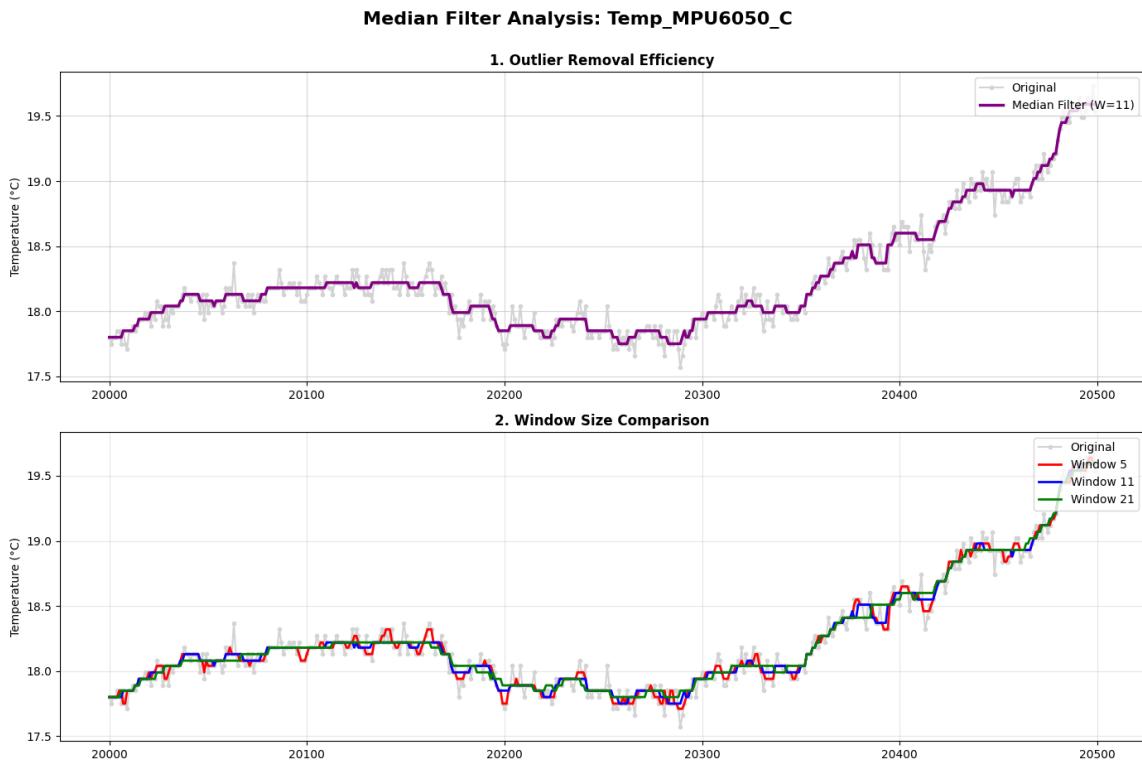


Figura 46: Comparação de filtros digitais - Temperatura Sensor 1 (Mediana)

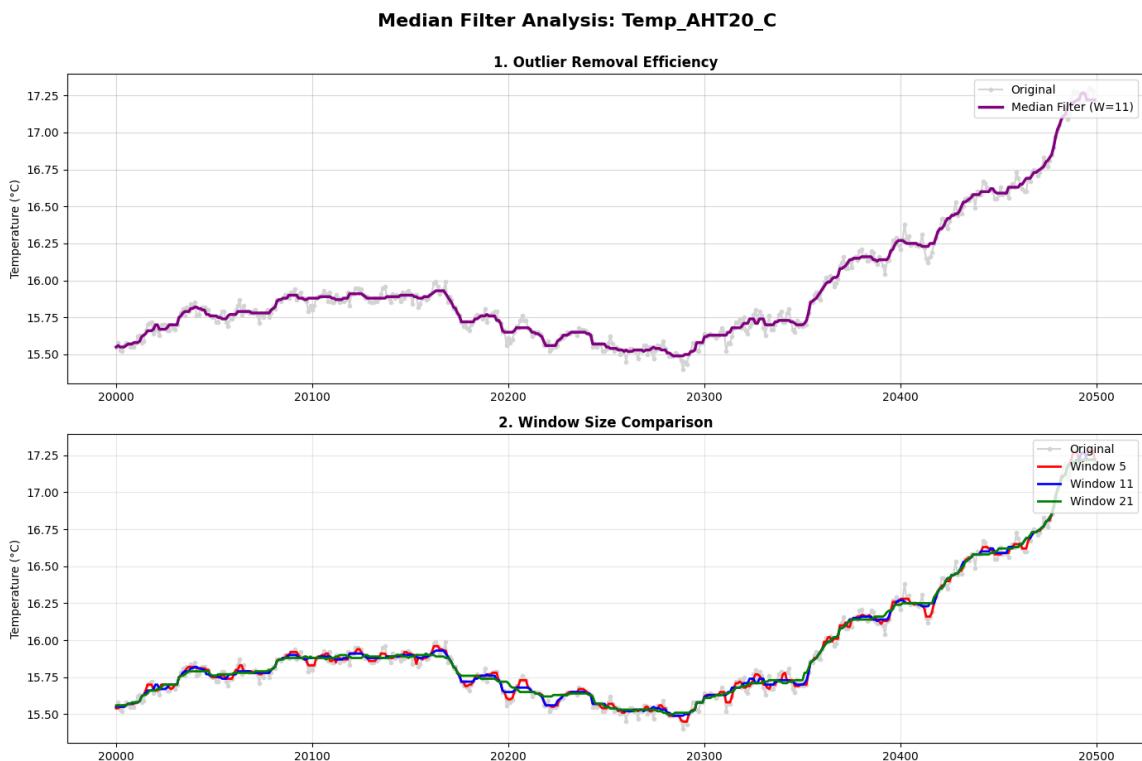


Figura 47: Comparação de filtros digitais - Temperatura Sensor 2 (Mediana)

Median Filter Analysis: Temp_BMP280_C

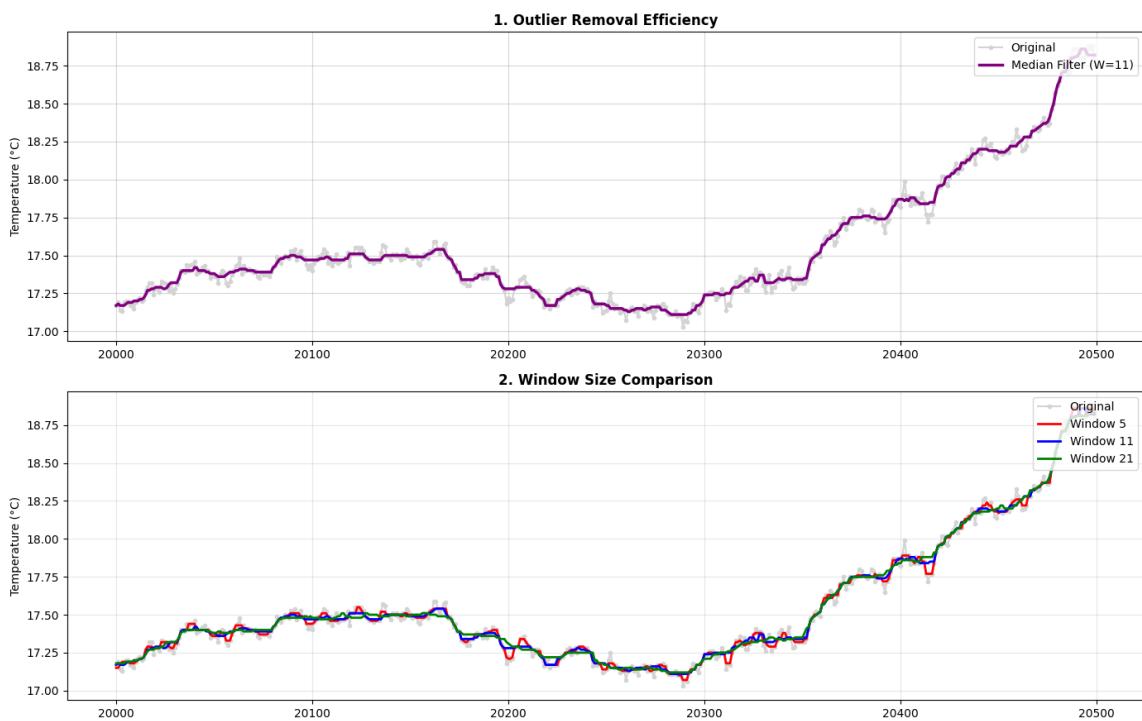


Figura 48: Comparação de filtros digitais - Temperatura Sensor 3 (Mediana)

Median Filter Analysis: Temp_BMP180_C

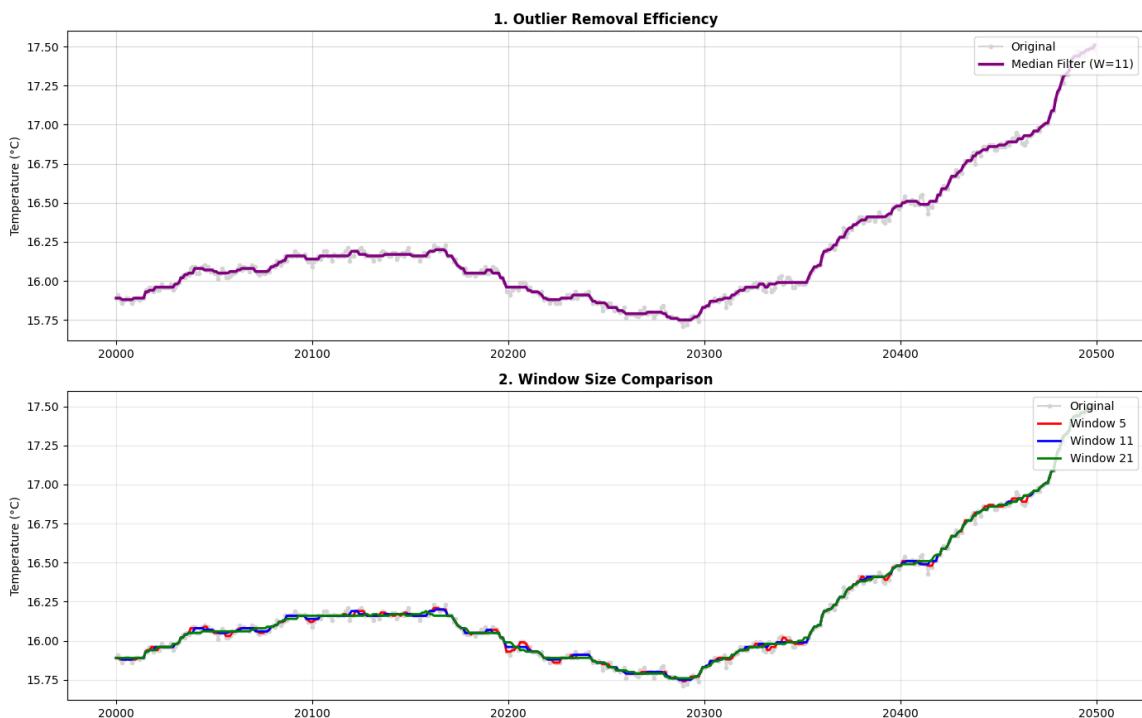


Figura 49: Comparação de filtros digitais - Temperatura Sensor 4 (Mediana)

Median Filter Analysis: Temp_DS18B20_C

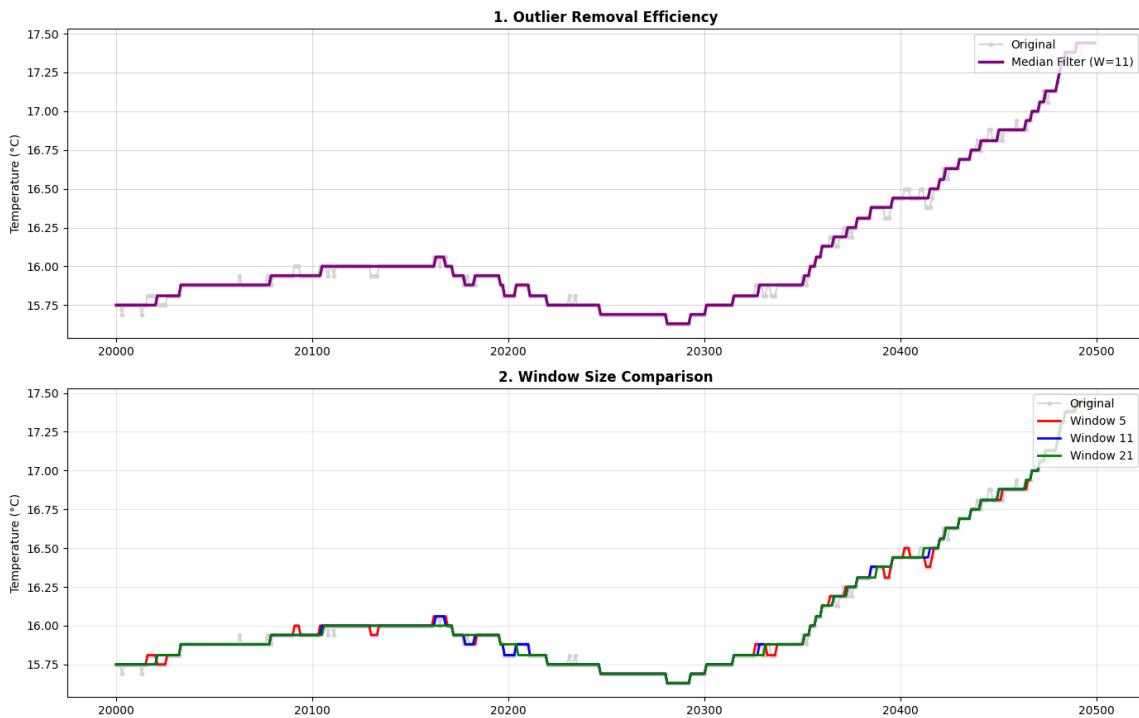


Figura 50: Comparação de filtros digitais - Temperatura Sensor 5 (Mediana)

Median Filter Analysis: Temp_NTC_C

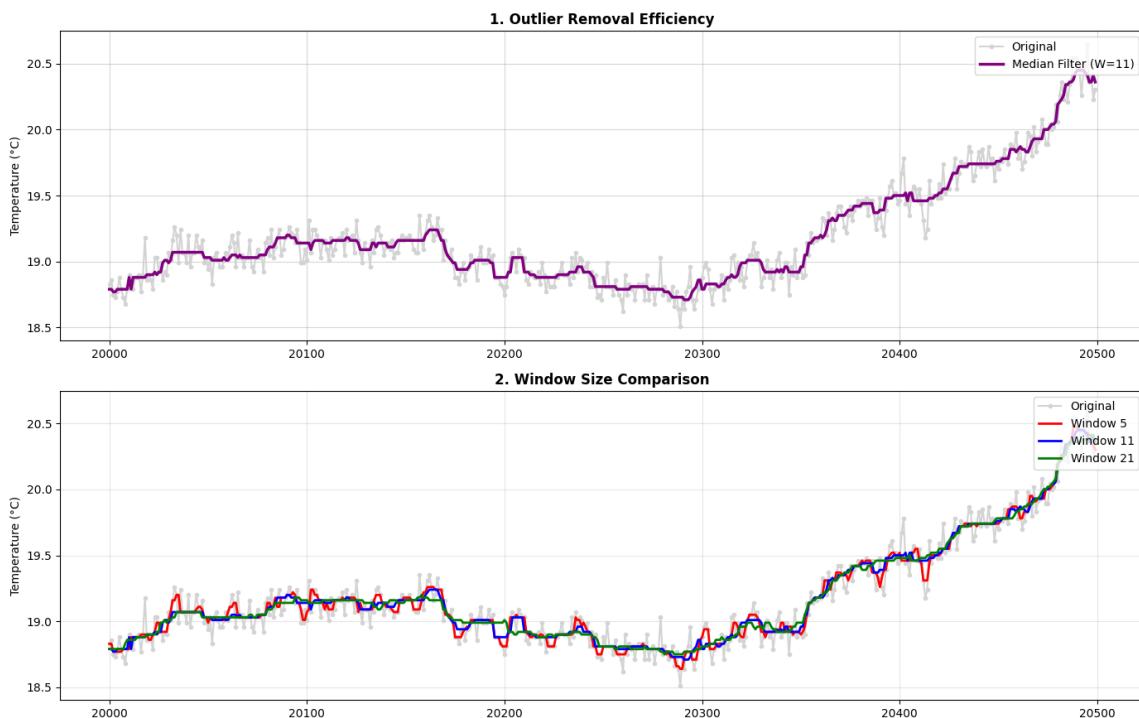


Figura 51: Comparação de filtros digitais - Temperatura Sensor 6 (Mediana)

Median Filter Analysis: Temp_DHT11_C

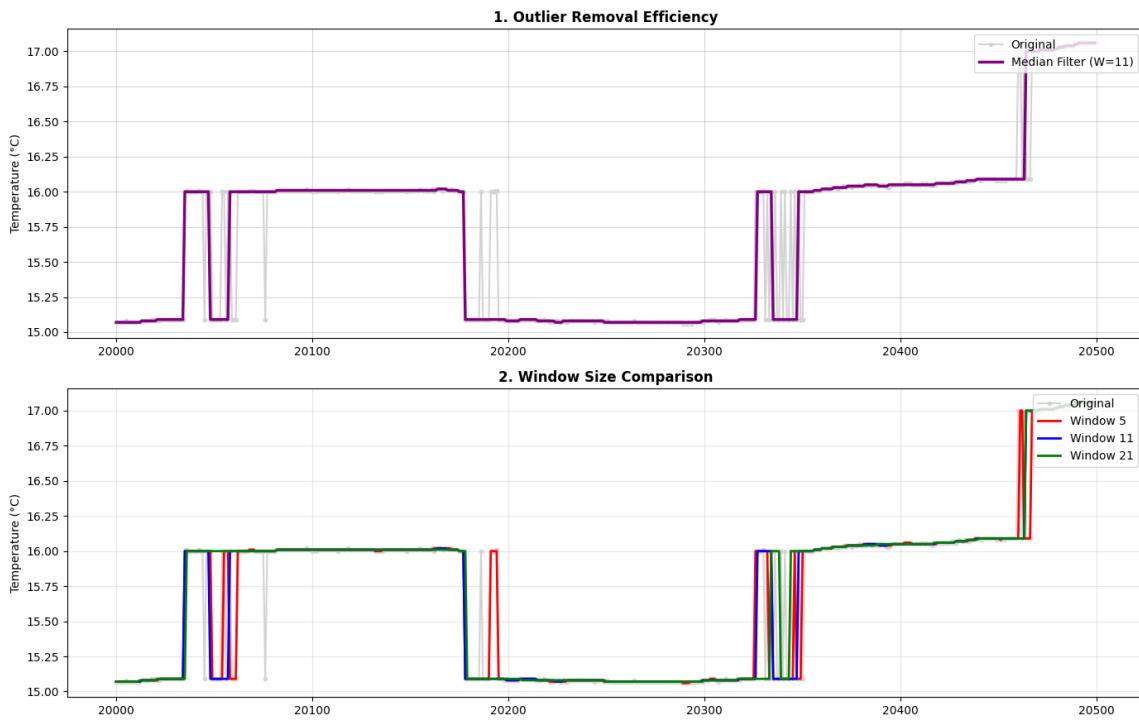


Figura 52: Comparação de filtros digitais - Temperatura Sensor 7 (Mediana)

EWMA Filter Analysis: Temp_MP6050_C

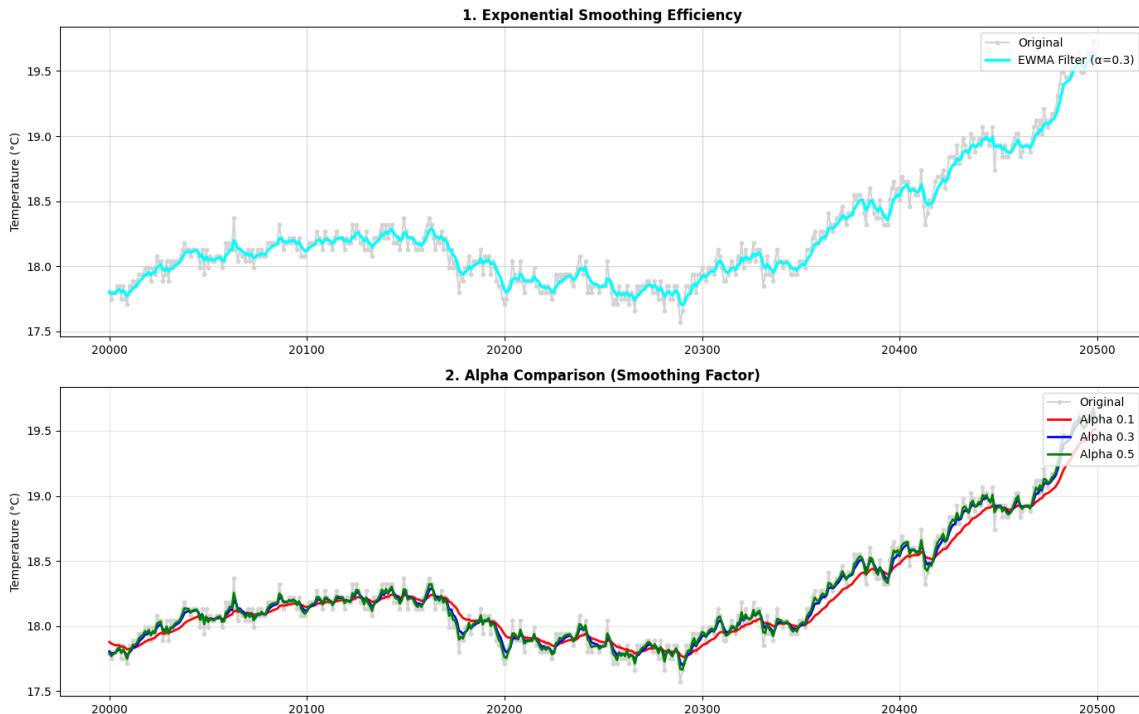


Figura 53: Comparação de filtros digitais - Temperatura Sensor 1 (EWMA)

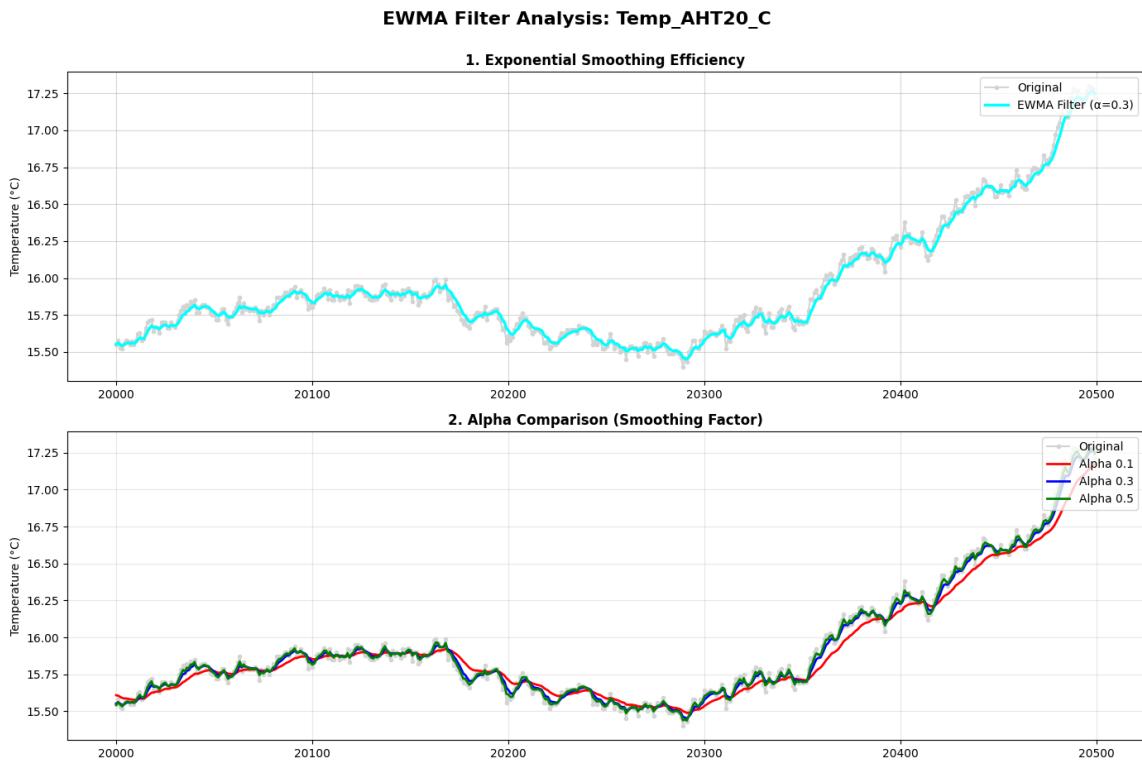


Figura 54: Comparação de filtros digitais - Temperatura Sensor 2 (EWMA)

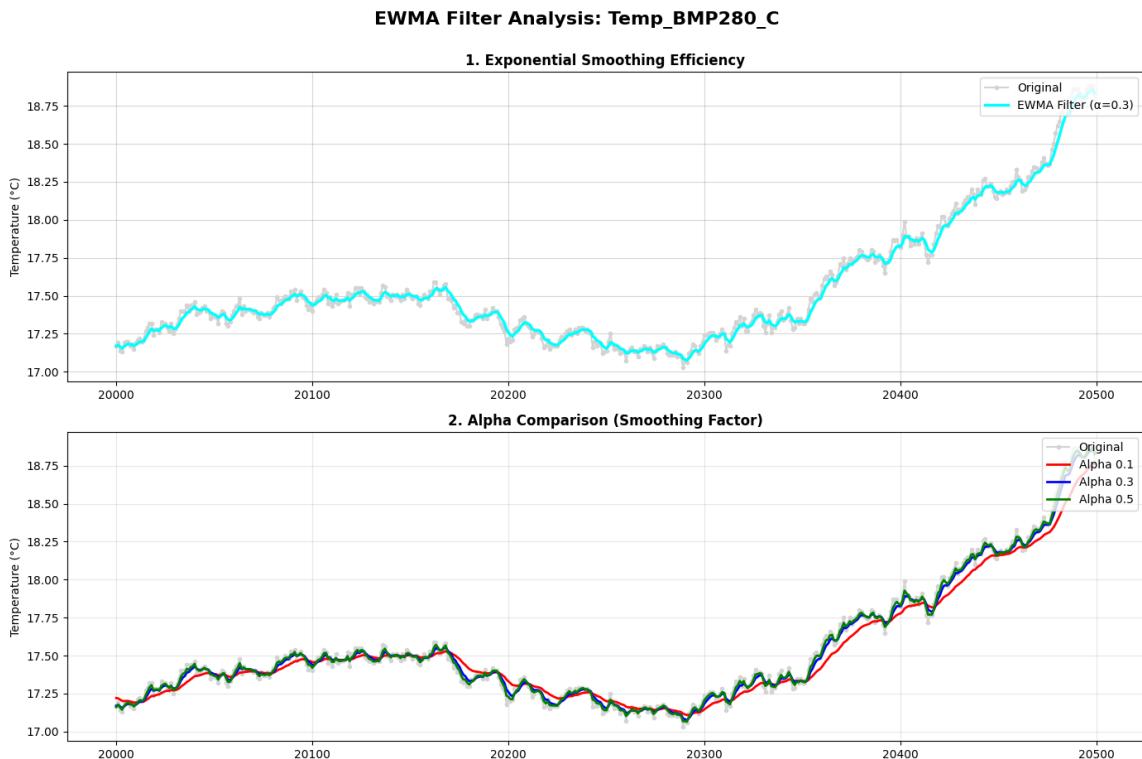


Figura 55: Comparação de filtros digitais - Temperatura Sensor 3 (EWMA)

EWMA Filter Analysis: Temp_BMP180_C

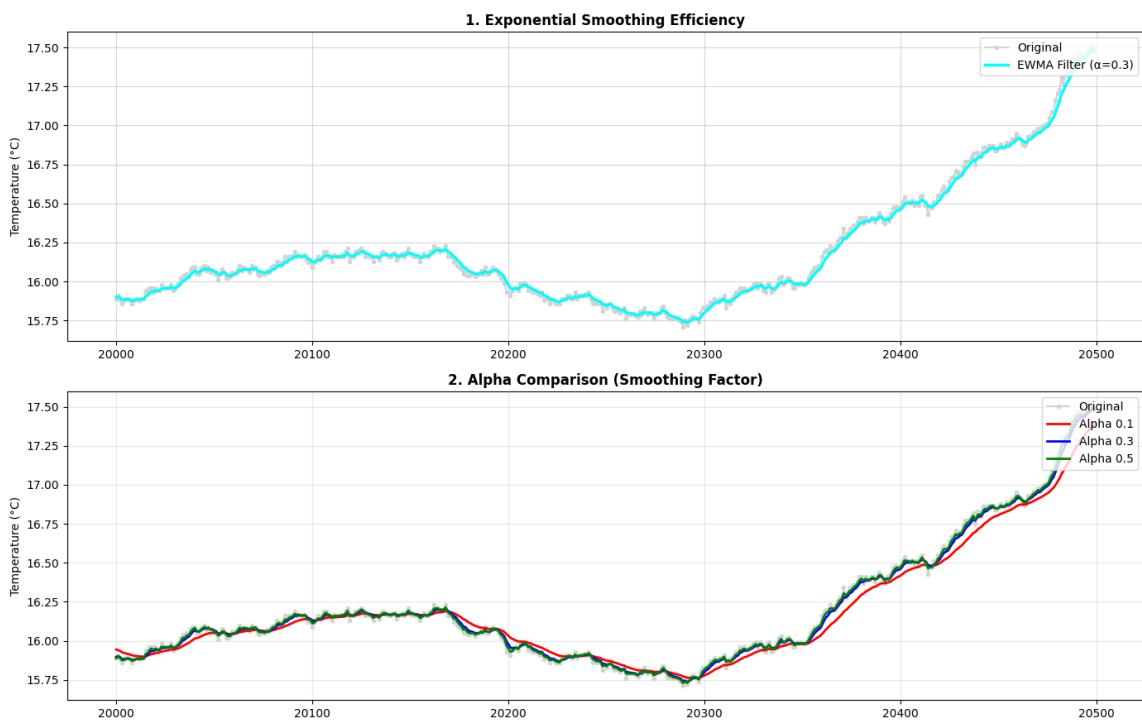


Figura 56: Comparação de filtros digitais - Temperatura Sensor 4 (EWMA)

EWMA Filter Analysis: Temp_DS18B20_C

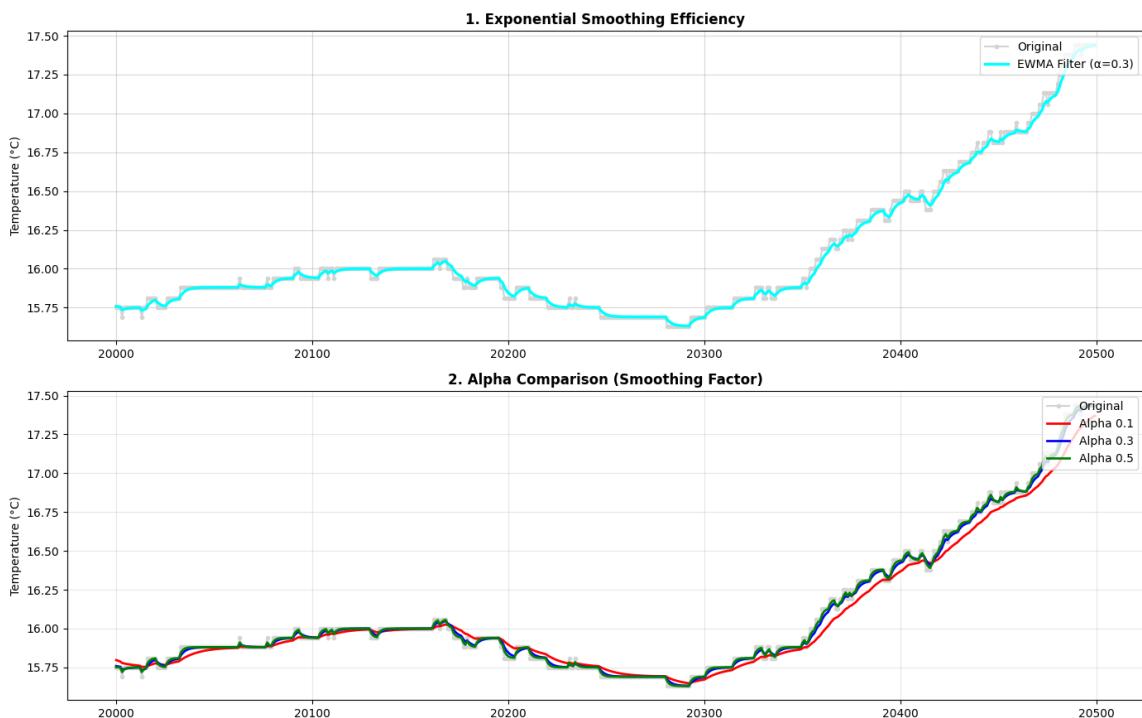


Figura 57: Comparação de filtros digitais - Temperatura Sensor 5 (EWMA)

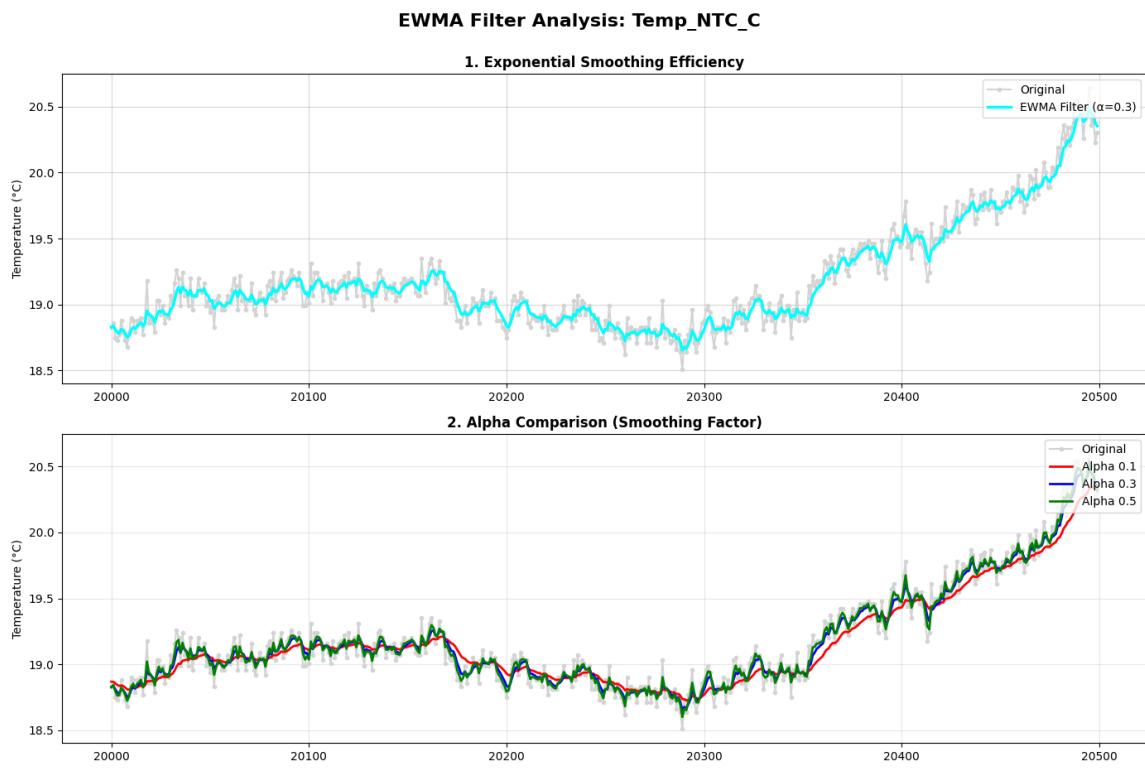


Figura 58: Comparação de filtros digitais - Temperatura Sensor 6 (EWMA)

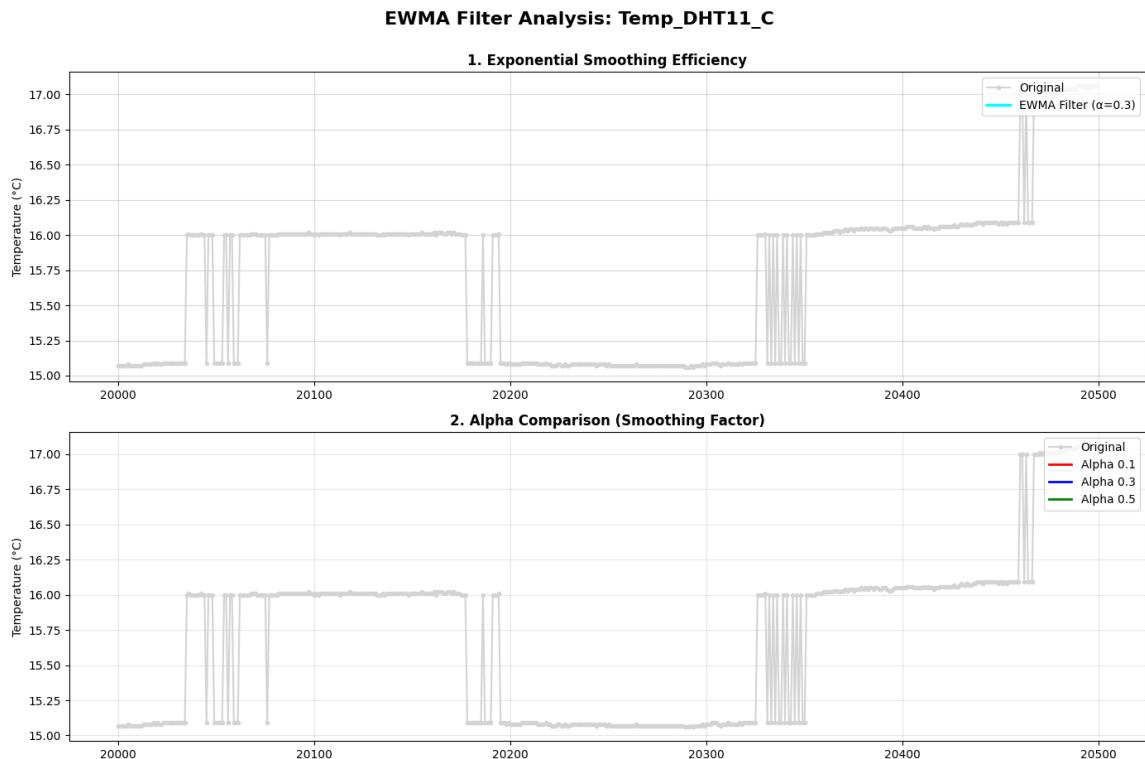


Figura 59: Comparação de filtros digitais - Temperatura Sensor 7 (EWMA)

6.1.2 Comparação de Filtros - Umidade

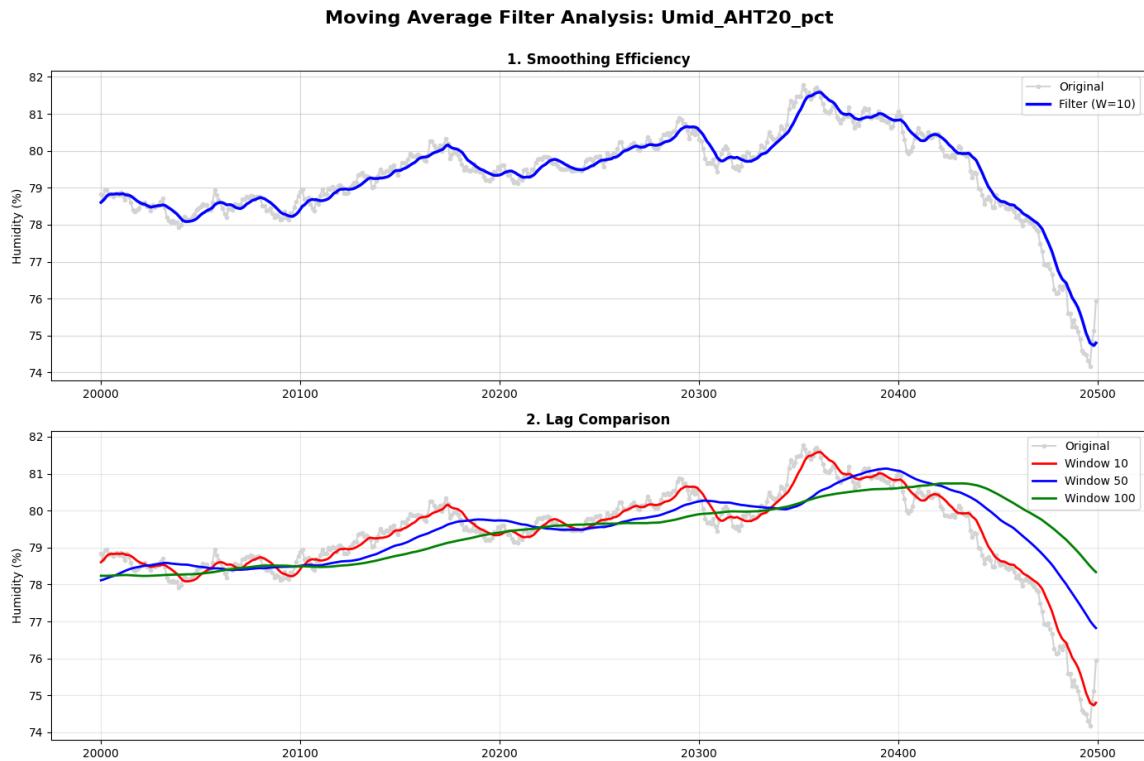


Figura 60: Comparação de filtros digitais - Umidade Sensor 1 (Média Móvel)

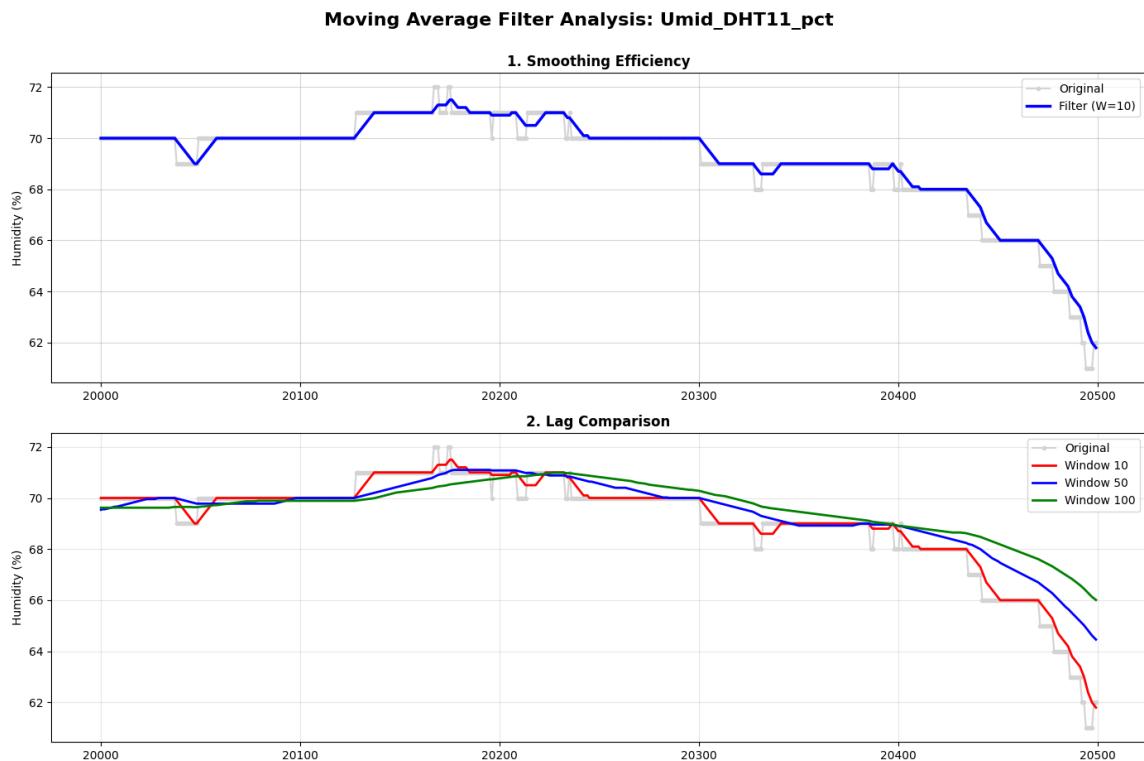


Figura 61: Comparação de filtros digitais - Umidade Sensor 2 (Média Móvel)

Median Filter Analysis: Umid_AHT20_pct

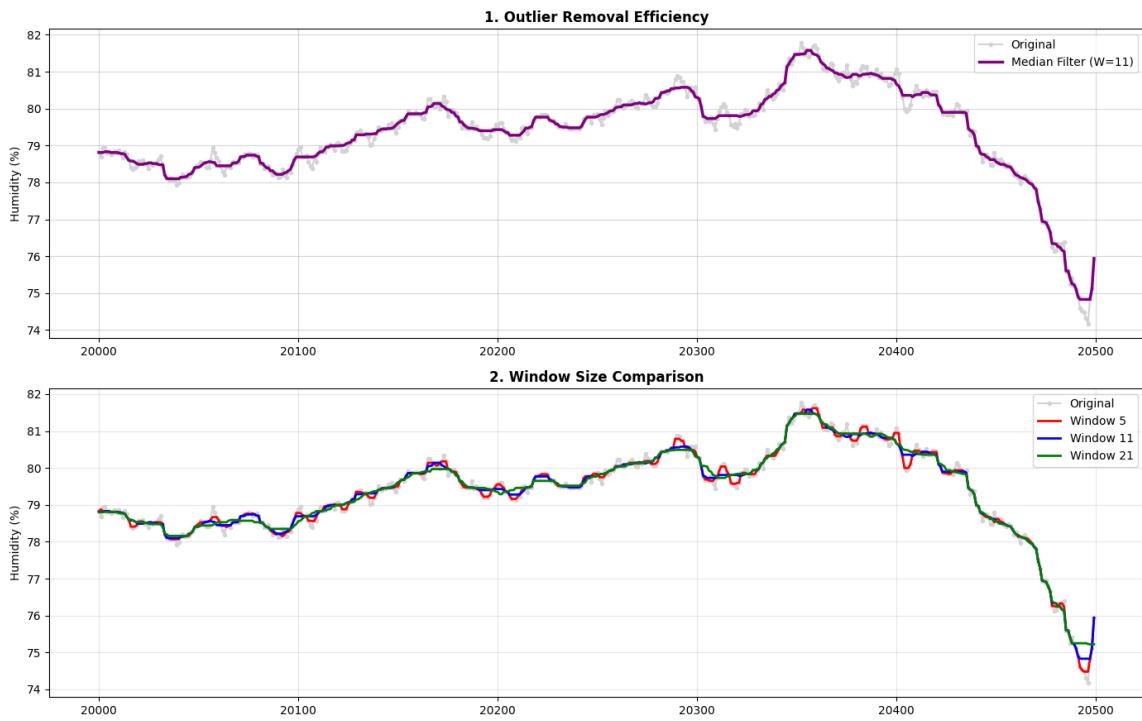


Figura 62: Comparação de filtros digitais - Umidade Sensor 1 (Mediana)

Median Filter Analysis: Umid_DHT11_pct

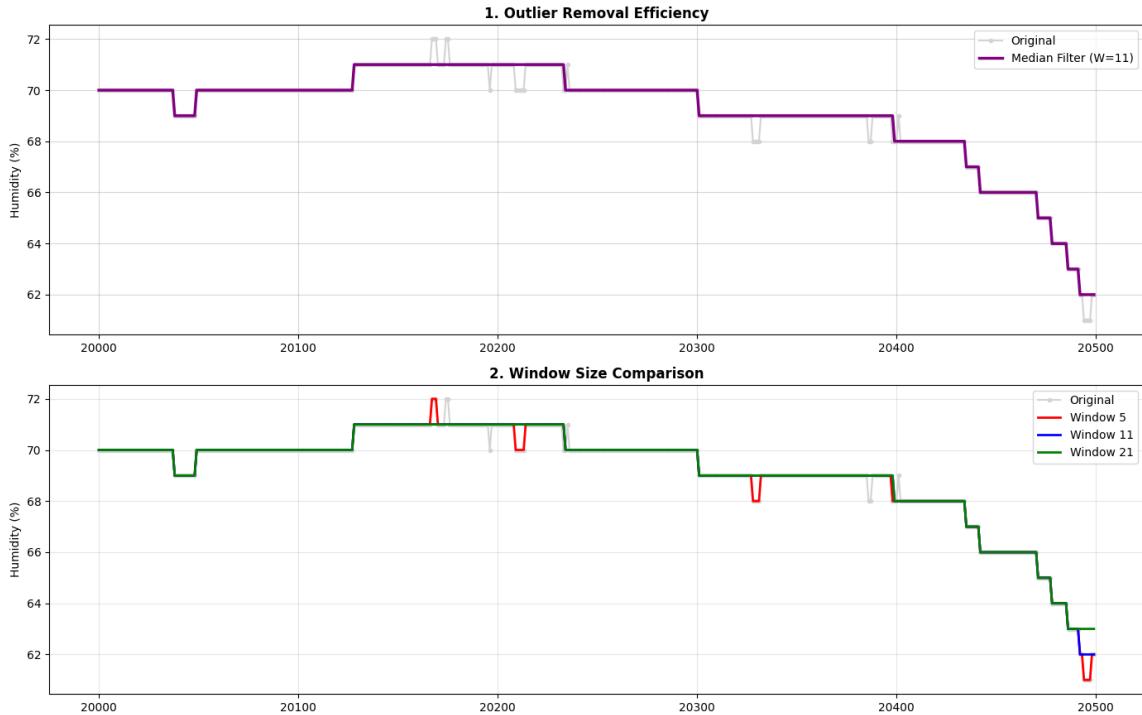


Figura 63: Comparação de filtros digitais - Umidade Sensor 2 (Mediana)

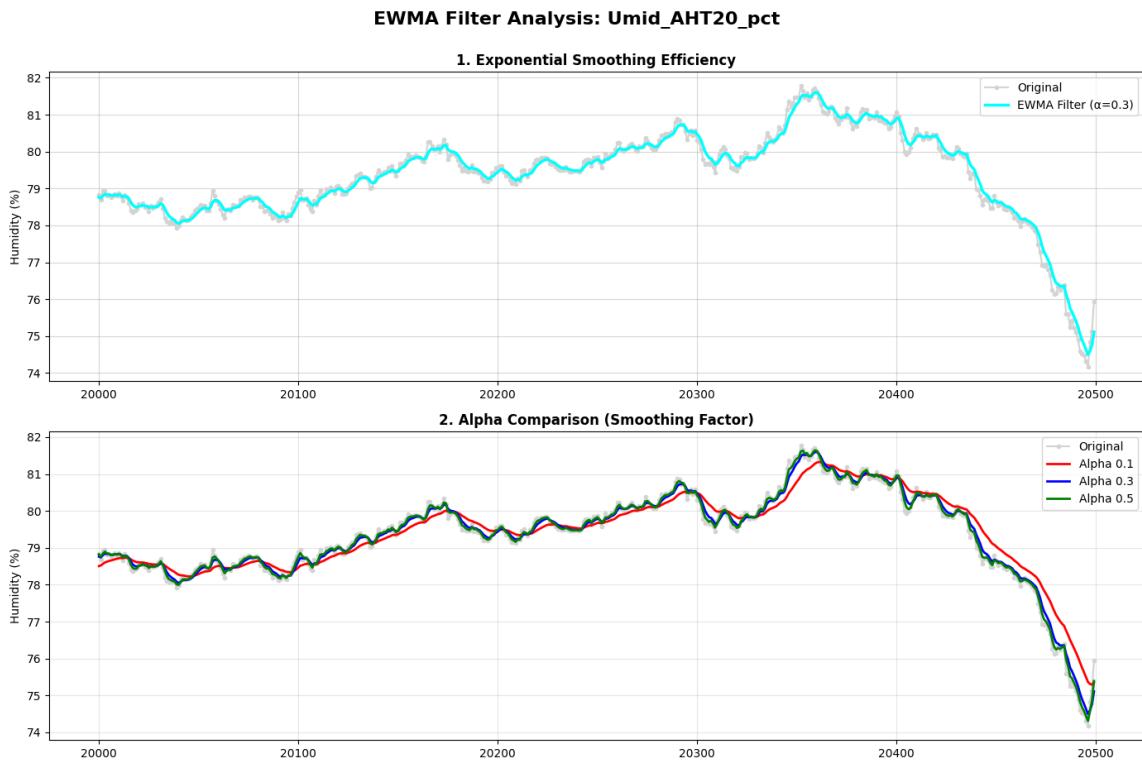


Figura 64: Comparação de filtros digitais - Umidade Sensor 1 (EWMA)

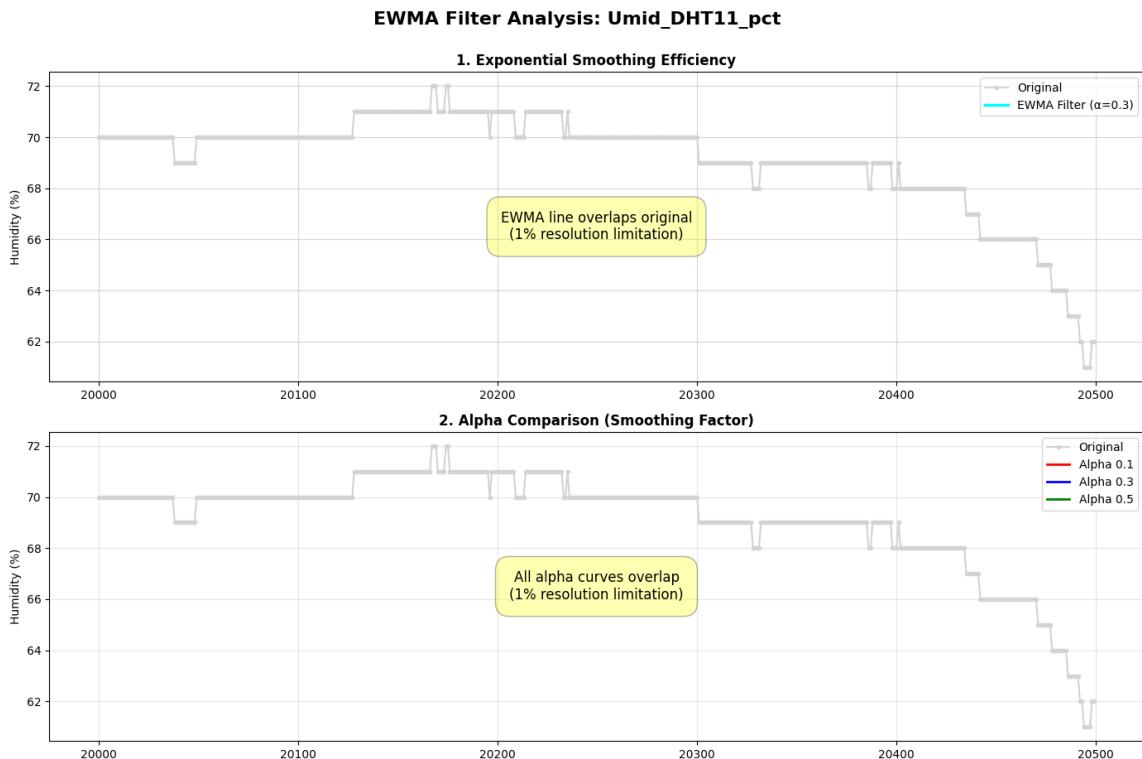


Figura 65: Comparação de filtros digitais - Umidade Sensor 2 (EWMA)

6.1.3 Comparação de Filtros - Pressão

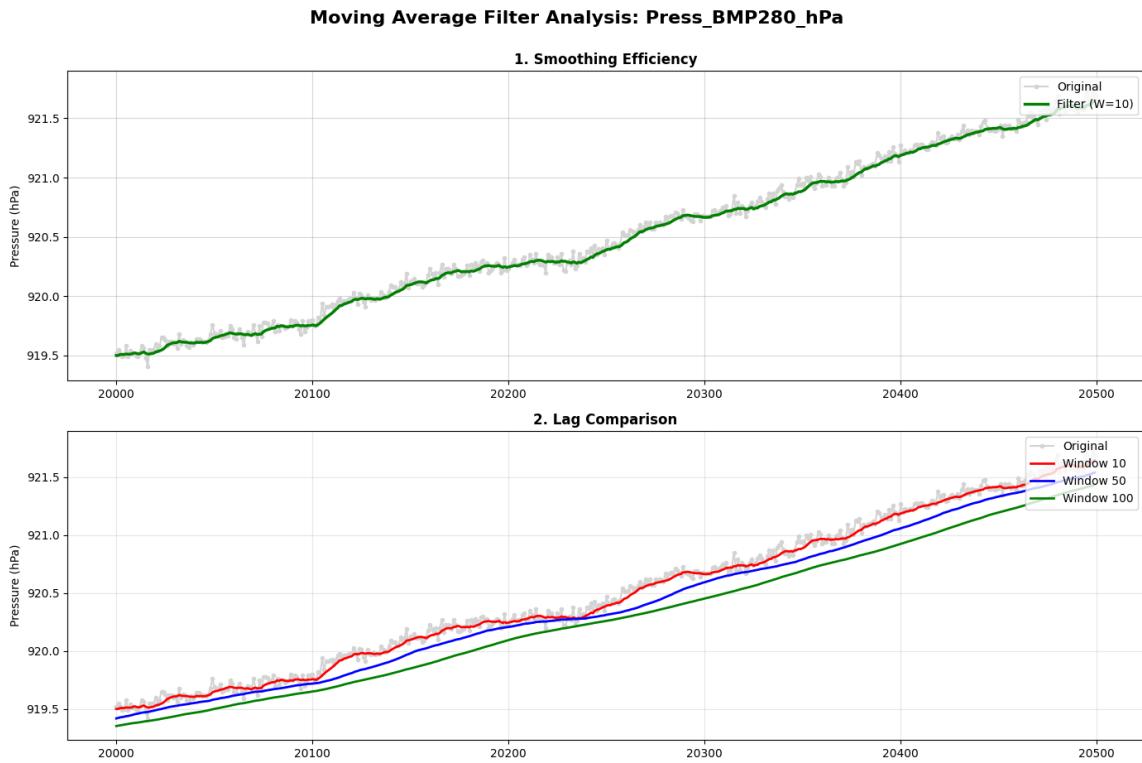


Figura 66: Comparação de filtros digitais - Pressão Sensor 1 (Média Móvel)

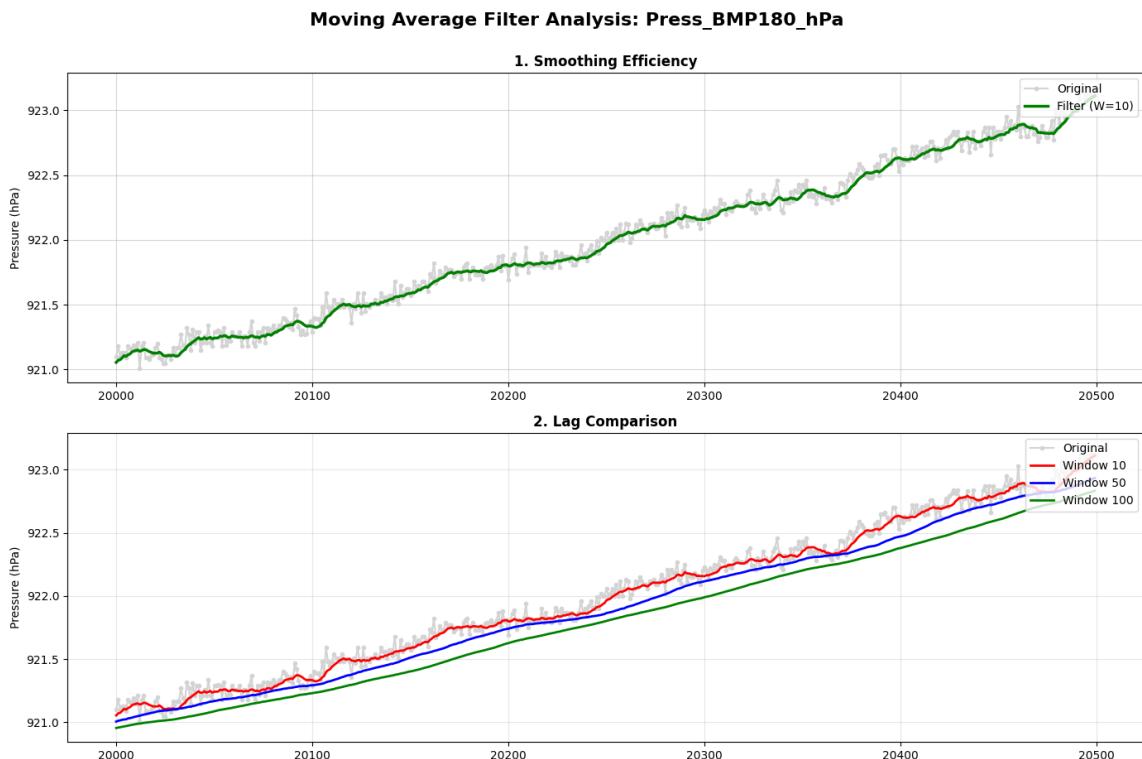


Figura 67: Comparação de filtros digitais - Pressão Sensor 2 (Média Móvel)

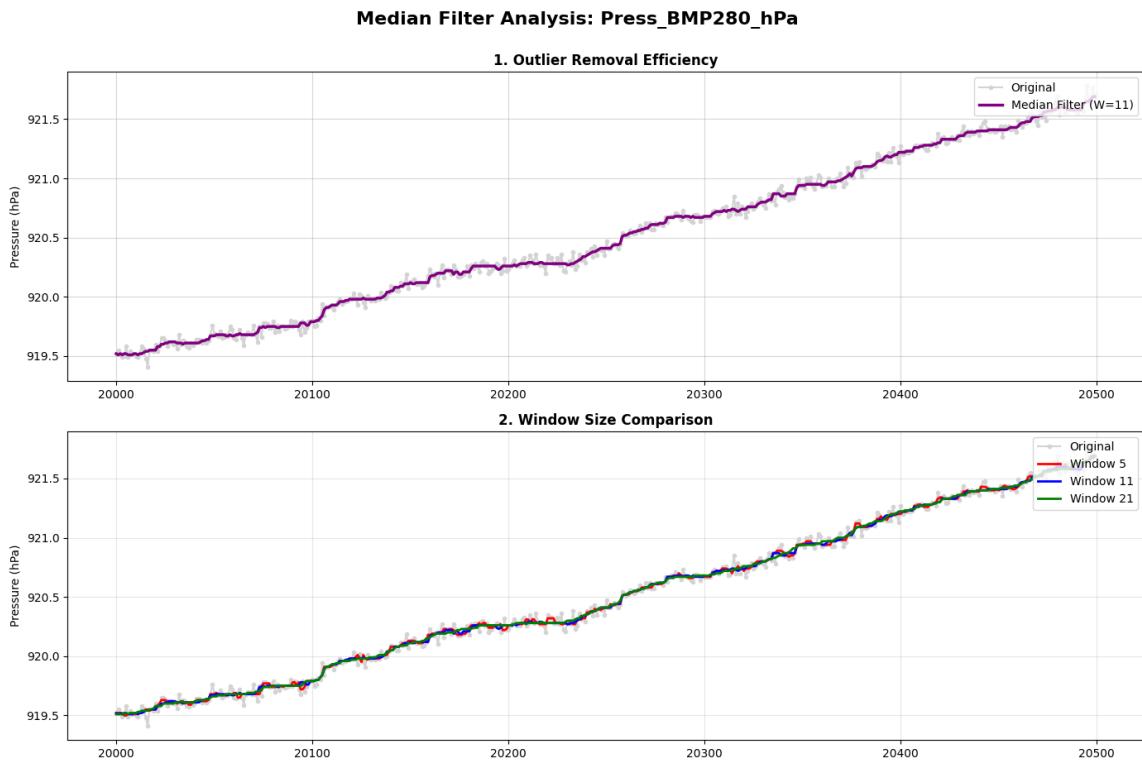


Figura 68: Comparação de filtros digitais - Pressão Sensor 1 (Mediana)

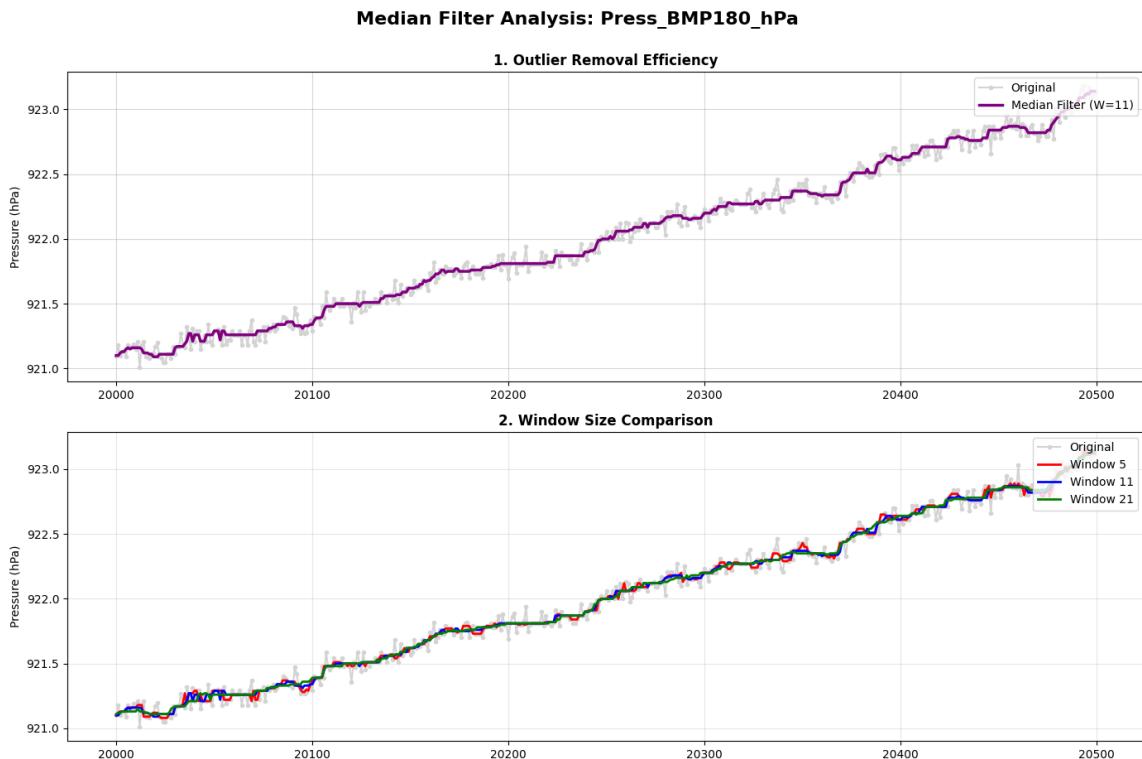


Figura 69: Comparação de filtros digitais - Pressão Sensor 2 (Mediana)

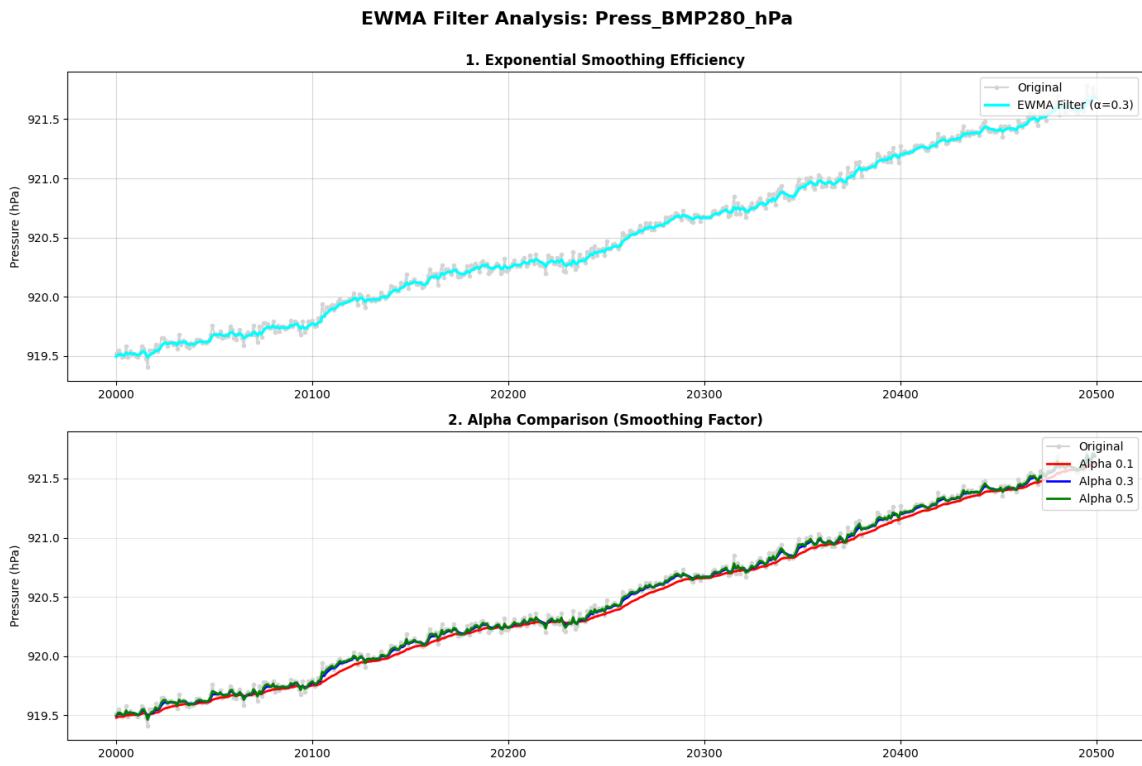


Figura 70: Comparação de filtros digitais - Pressão Sensor 1 (EWMA)

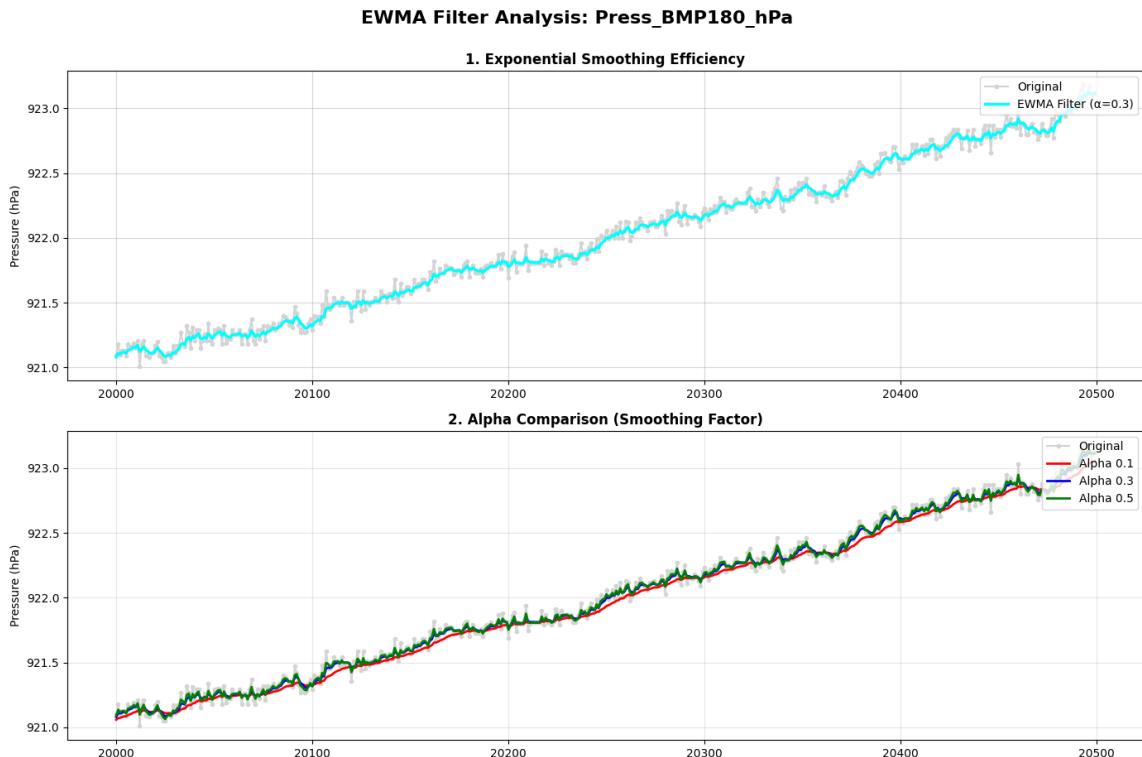


Figura 71: Comparação de filtros digitais - Pressão Sensor 2 (EWMA)

Resultados da filtragem digital:

- **Filtro Mediana:** Melhor supressão de ruído sem lag de fase

- **EWMA:** Transições suaves, preserva tendências
- **Média Móvel:** Trade-off entre suavização e responsividade
- Painéis completos com sinal bruto, filtrado, zoom e espectro de frequência

6.2 Notebook 13: Análise de Frequência (FFT)

Nota: As análises FFT foram integradas aos painéis de filtros digitais (Notebook 12), mostrando o espectro de potência de cada sinal antes e depois da filtragem.

Resultados da análise espectral:

- **Pico primário:** Periodicidade de 23 horas (ciclo diurno)
- Assinatura espectral clara do ciclo de aquecimento/resfriamento diário
- Validação da adequação do intervalo de amostragem de 30 segundos

7 Dataset Público

O dataset completo foi disponibilizado publicamente no Kaggle:

- **Título:** “Vitória da Conquista Weather Data - September 2025”
- **URL:** <https://www.kaggle.com/datasets/jonassouza872/vitoria-da-conquista-weather-data>
- **Registros:** 82.430 medições
- **Formato:** CSV com 12 variáveis ambientais
- **Licença:** Acesso livre, público, para download

8 Estrutura do Repositório

```
PolySense-Station-/
|-- main.py                                     # 226 linhas MicroPython
|-- requirements.txt                            # Dependencias Python
|-- lib/                                         # Drivers MicroPython (9 arquivos)
|-- notebooks/                                    # 13 notebooks de analise
|   |-- 01_exploratory_analysis.ipynb
|   |-- 02_correlation_analysis.ipynb
|   |-- 03_missing_data.ipynb
|   |-- 04_sensor_validation.ipynb
|   |-- 05_temporal_analysis.ipynb
|   |-- 06_time_series_decomposition.ipynb
|   |-- 07_anomaly_detection.ipynb
|   |-- 08_decision_tree_regression.ipynb
|   |-- 09_gmm_clustering.ipynb
|   |-- 10_kmeans_clustering.ipynb
|   |-- 11_lstm_prediction.ipynb
```

```

|   |-- 12_digital_filters.ipynb
|   |-- 13_fft_analysis.ipynb
|-- data/raw/                      # 4 arquivos CSV (4.8 MB total)
|   |-- climate_clusters_gmm.csv    # 82.430 registros
|   |-- validation_data_cleaned_BRT.csv
|   |-- validation_and_Measured_Data_cleaned_BRT_.csv
|   |-- inmet_weather_station_data_sep_2025_utc.csv
|-- images/                         # 111 visualizacoes geradas
|   |-- data_analysis/              # EDA, correlacao, validacao
|   |-- machine_learning/          # Clustering, anomalias, regressao
|   |-- signal_processing/         # Comparacoes de filtros, FFT
|-- PCB/                            # Design de PCB customizado
|-- Schematic/                     # Esquematicos de hardware
|-- README.md                       # Documentacao completa

```

9 Conclusões

9.1 Principais Conquistas

1. **Sistema de coleta funcional:** 100% de sucesso ao longo de 30 dias (82.430 registros com intervalo de 30s)
2. **Redundância de sensores eficaz:** 7 sensores de temperatura, 2 de umidade, 2 de pressão com ~99.9% de completude
3. **Validação contra referência:** Bias médio ± 2°C comparado à estação INMET oficial
4. **Correlação excepcional:** $r \approx 0.98$ entre sensores redundantes.
5. **Detecção de anomalias:** 3.2% de outliers isolados com sucesso via Isolation Forest
6. **Previsão precisa:** MAE ± 1°C para previsões de temperatura 1 hora à frente com LSTM
7. **Clustering climático:** 4 regimes distintos identificados e validados (GMM e KMeans)
8. **Análise espectral:** Periodicidade de 24h clara confirmada via FFT
9. **Dataset público:** 82.430 medições disponibilizadas no Kaggle para comunidade científica

9.2 Limitações do Estudo

- Período de coleta limitado a 30 dias (setembro 2025)
- Ausência de medição direta de radiação solar
- Granularidade temporal de 30 segundos (não captura eventos muito rápidos)
- Fatores externos não considerados (cobertura de nuvens detalhada, vento local)

- Validação limitada a uma estação do ano

9.3 Trabalhos Futuros

1. Expandir coleta para 12 meses (análise sazonal completa)
2. Adicionar sensor de radiação solar (piranômetro)
3. Implementar anemômetro para velocidade e direção do vento
4. Desenvolver PCB versão 2 com melhorias identificadas
5. Integrar transmissão de dados via LoRaWAN
6. Implementar edge computing com previsões LSTM em tempo real
7. Análise comparativa entre múltiplas estações em diferentes altitudes
8. Desenvolver API REST para acesso aos dados em tempo real

10 Referências

- **Microcontrolador:** Raspberry Pi Pico (RP2040), MicroPython Documentation
- **Estação de referência:** INMET - Instituto Nacional de Meteorologia
- **Dataset:** <https://www.kaggle.com/datasets/jonassouza872/vitoria-da-conquista-weather>
- **Período:** 31/08/2025 - 30/09/2025 (82.430 registros)
- **Bibliotecas Python:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, SciPy
- **Técnicas de ML:** Isolation Forest, Decision Tree, KMeans, GMM, LSTM
- **Processamento de sinais:** FFT, Filtros digitais (Média móvel, Mediana, EWMA)