Pattern Detection in Time Series

Ecker Nick

Table of contents

01

Introduction

- Luxembourgish company founded in 1970
- Global leader in steel industry
- Smart and environmental solutions

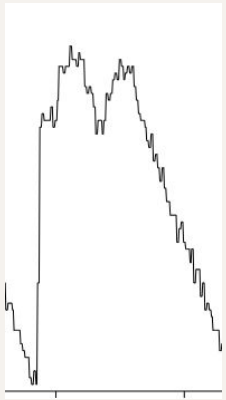
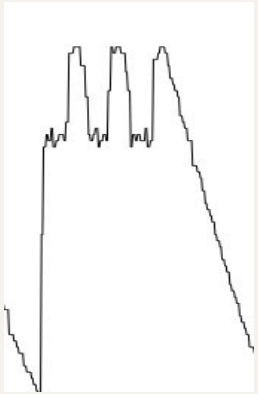
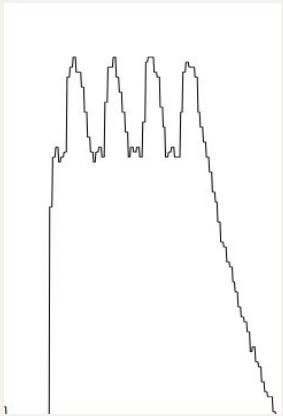
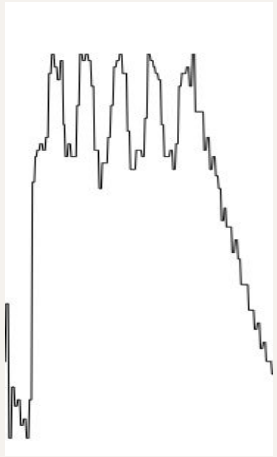
- Software developed by Datathings
- ML models without specialized knowledge
- Regression, classification and pattern detection

02

Data and Objectif

Data Presentation

- 4 time series, each corresponding to a role in a section mill
- The values correspond to the rotation speed of a role at each timestep
- Composed of different types of patterns: from 2 to 5 picks

Objectif: Maximise pattern detections without making wrong detections

03

Problematic

when using AIXPert

Limitations of AIXPert

- Limited number of distance measures:
  - Euclidean Distance
  - Symbolic Aggregate Approximation (SAX)

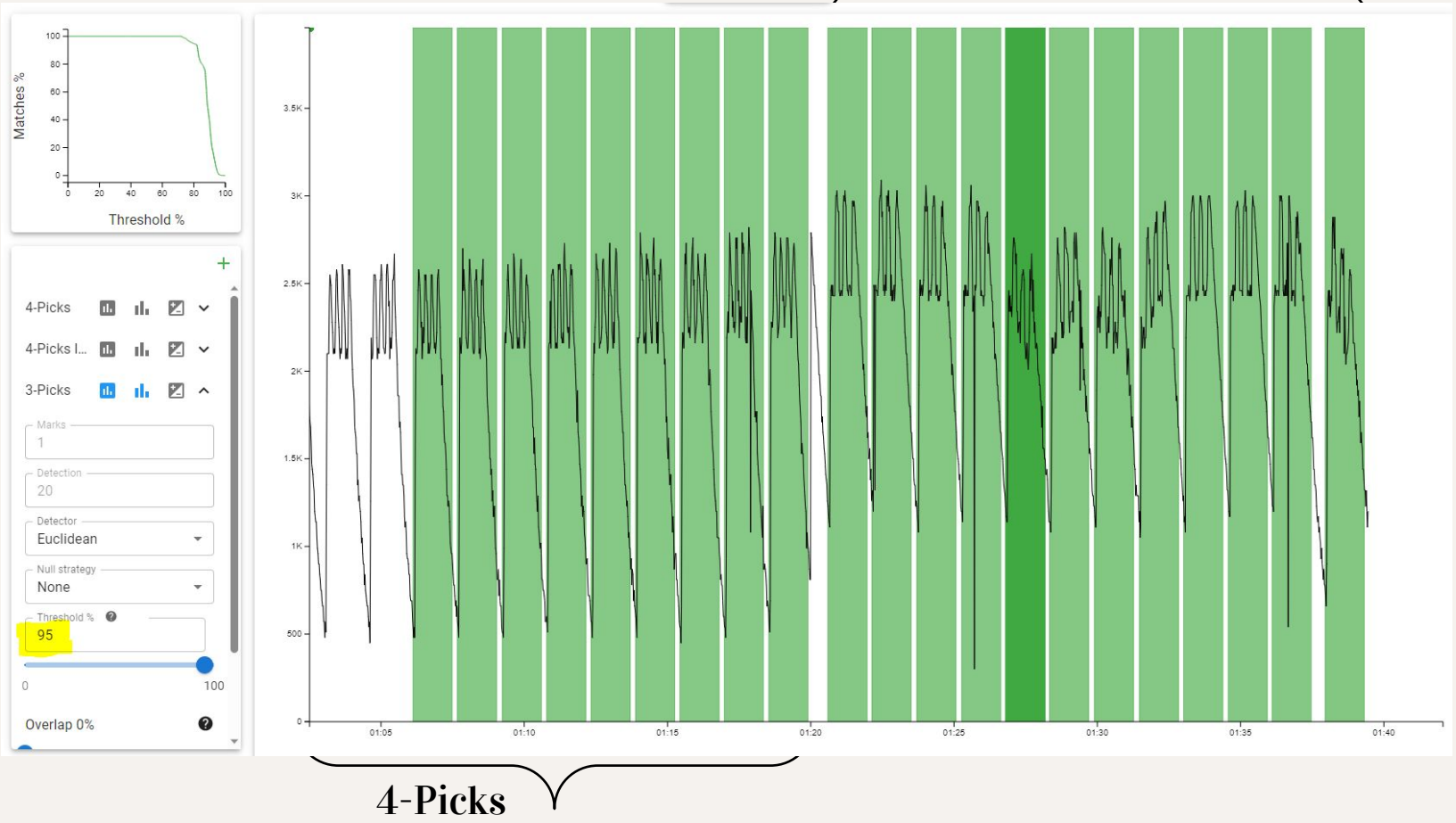
- Threshold too low → Wrong detections

- Threshold too high → Missing detections

- Small changes can have a big impact

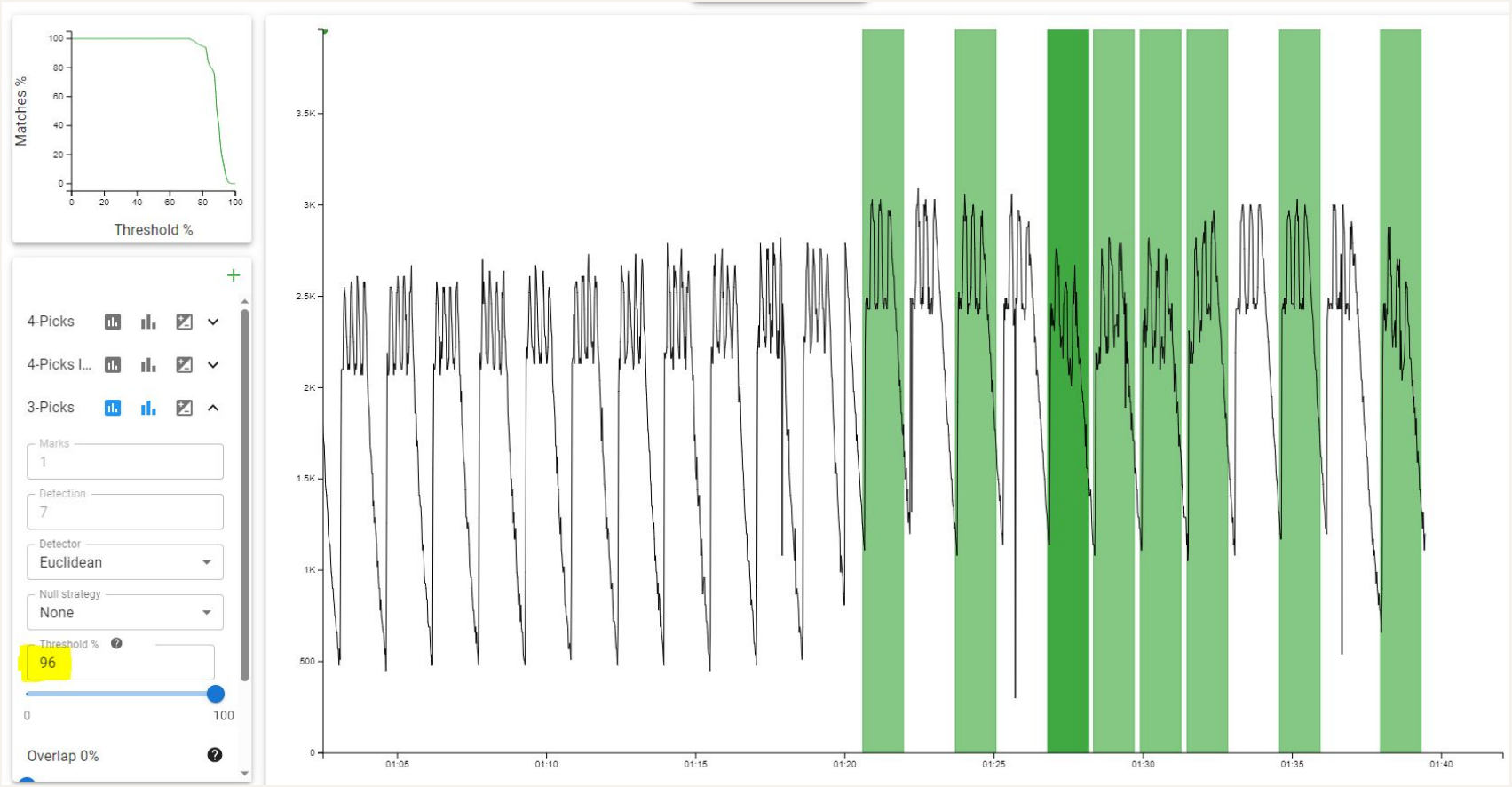
Threshold = 95%

3-Picks



4-Picks

Threshold = 96%



04

New method

Dynamic Time Warping

- Algorithm for measuring similarity between two time series

- Main advantages:
  - Time series can be of different lengths
  - Robust in change of time range

- Main disadvantage:
  - Complexity of O(n*m)
  - Euclidean Distance: O(n)

DTW Algorithm

	0	1	1	2	2	3	5
0							
1							
2							
3							
5							
5							
5							
6							

	0	1	1	2	2	3	5
0	0	inf	inf	inf	inf	inf	inf
1	inf						
2	inf						
3	inf						
5	inf						
5	inf						
5	inf						
6	inf						

$$M[i, j] = |A[i] - B[j]| + \min(M[i-1, j-1], M[i-1, j], M[i, j-1])$$

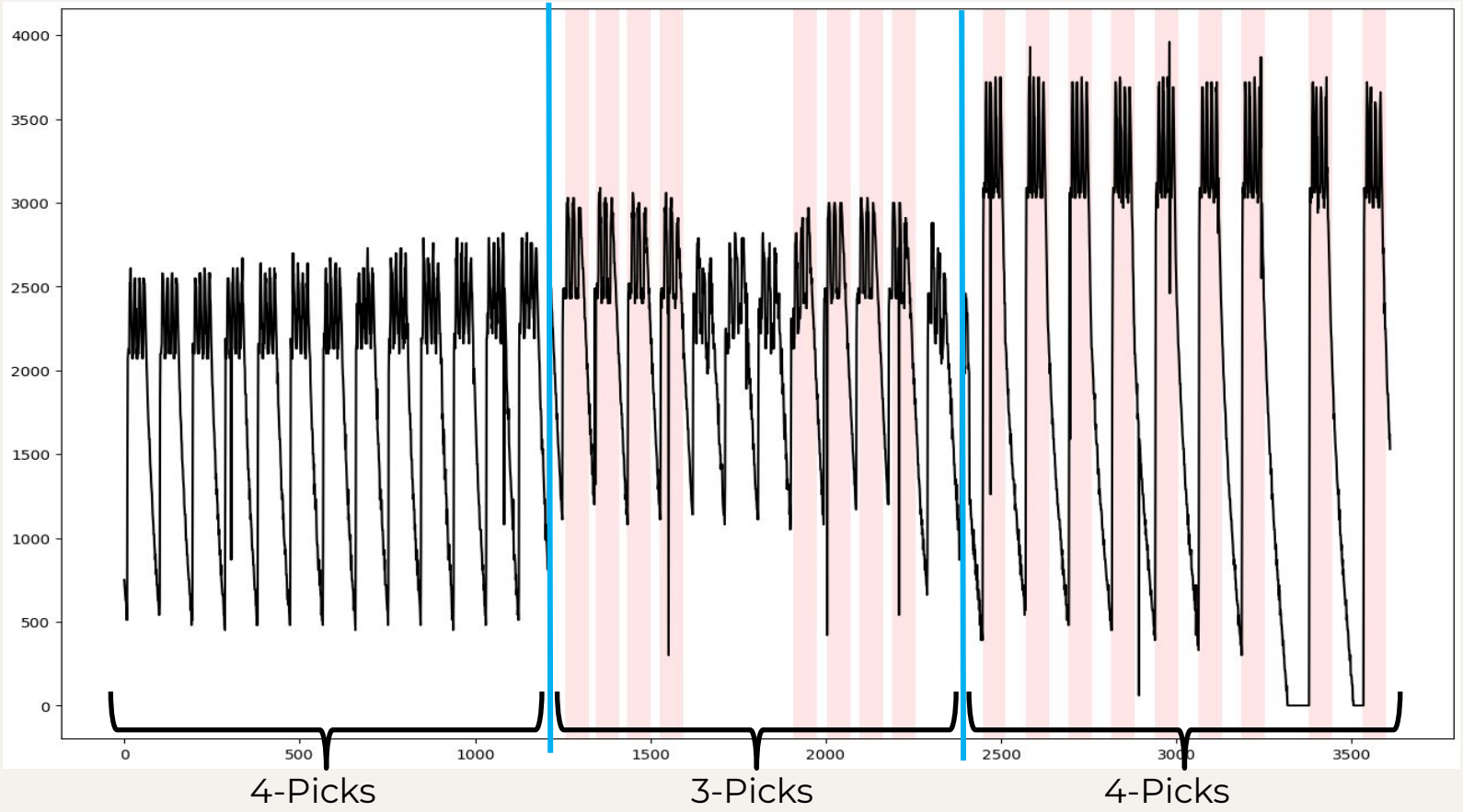
	0	1	1	2	2	3	5
0	0	inf	inf	inf	inf	inf	inf
1	inf	0	0	1	2	4	8
2	inf	1	1	0	0	1	4
3	inf	3	3	1	1	0	2
5	inf	7	7	4	4	2	0
5	inf	11	11	7	7	4	0
5	inf	15	15	10	10	6	0
6	inf	20	20	14	14	9	1

05.1

Results

Single Pattern

Application of DTW



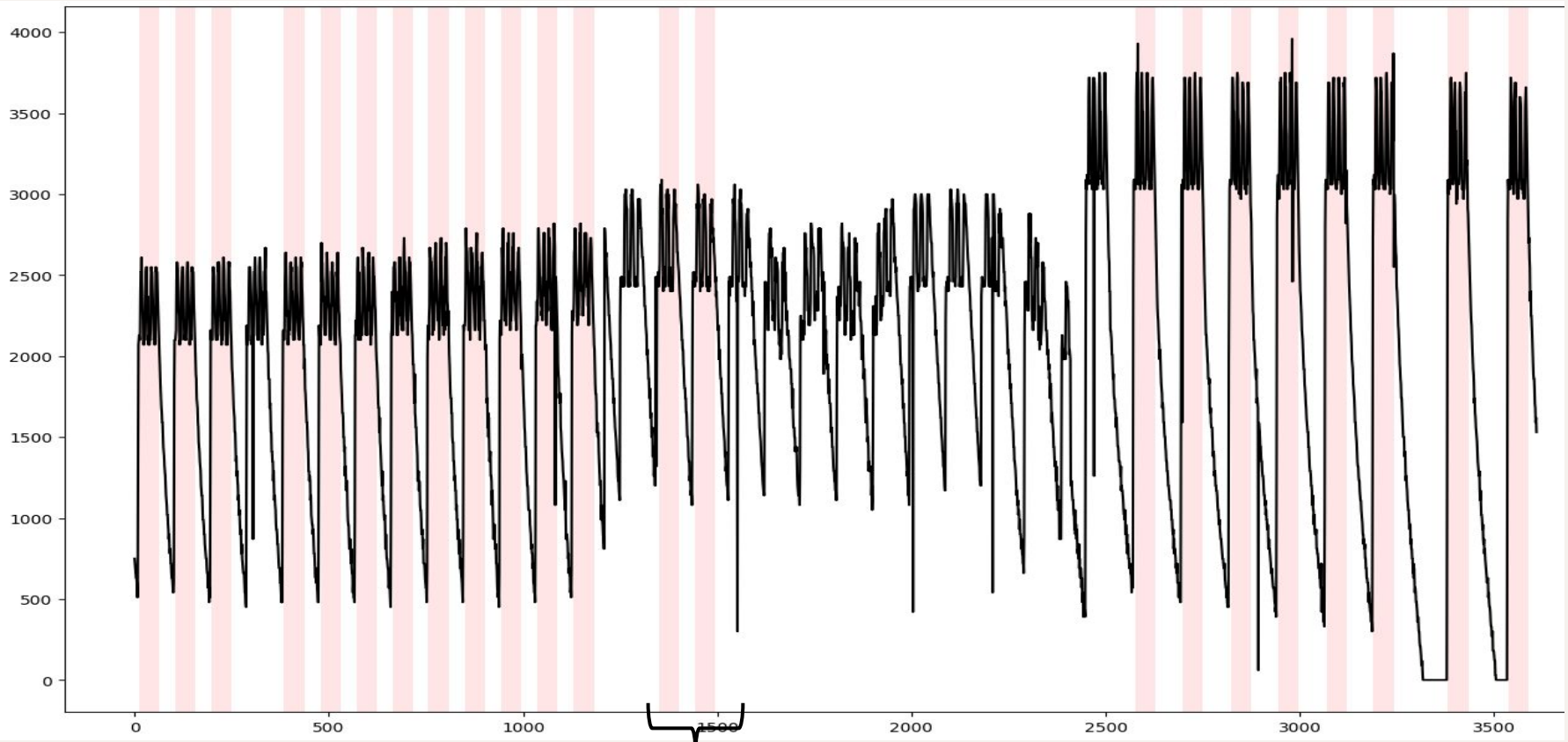
Problem: Amplitude of a signal has too much impact in making detections

Solution: comparing patterns locally

1. Adjusting by mean: $Z = X - \text{mean}(X)$
2. Normalization: $Z = X - \min(X) / \max(X) - \min(X)$
3. Standardization: $Z = X - \text{mean}(X) / \text{std}(X)$

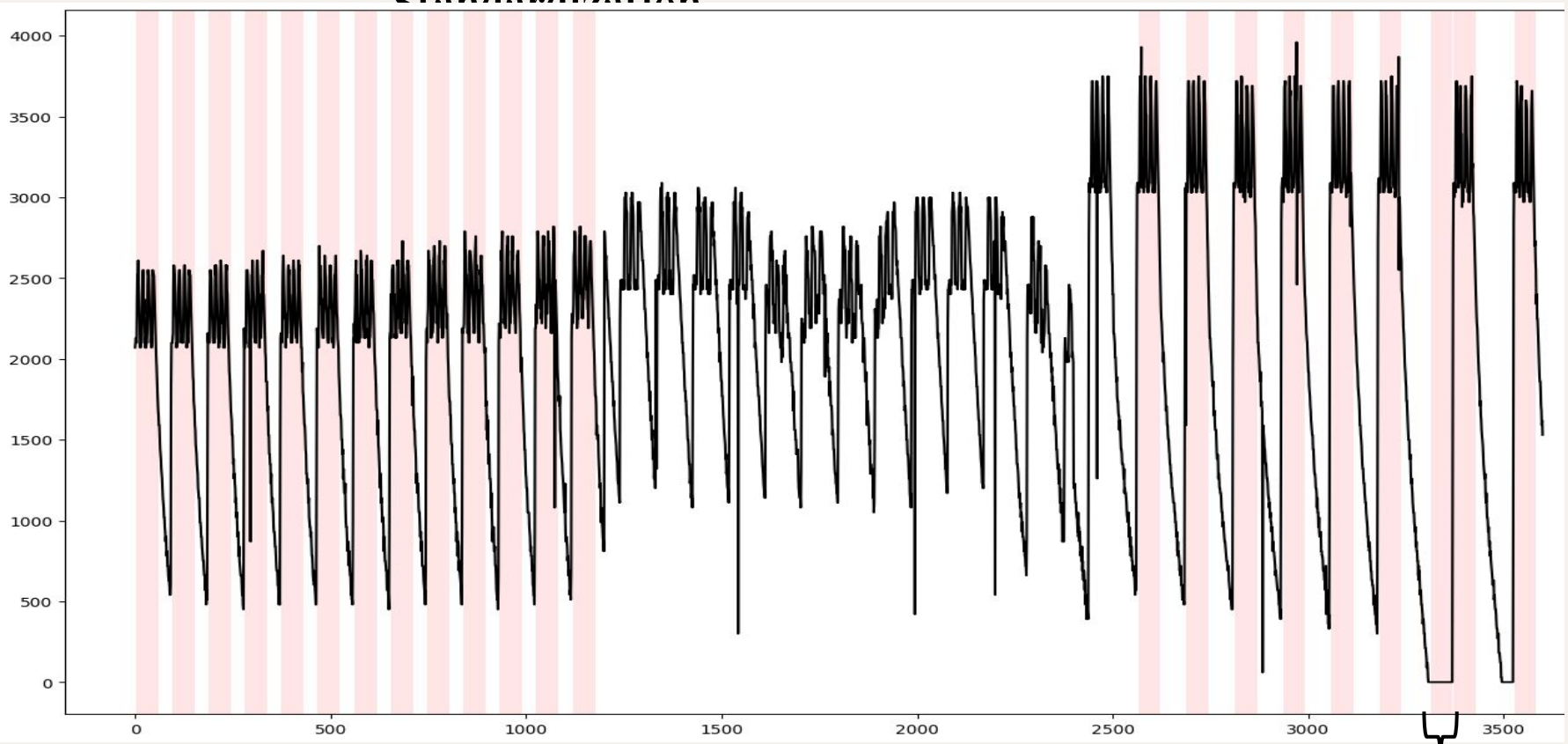
→ Best results with standardization

DTW with Standardization

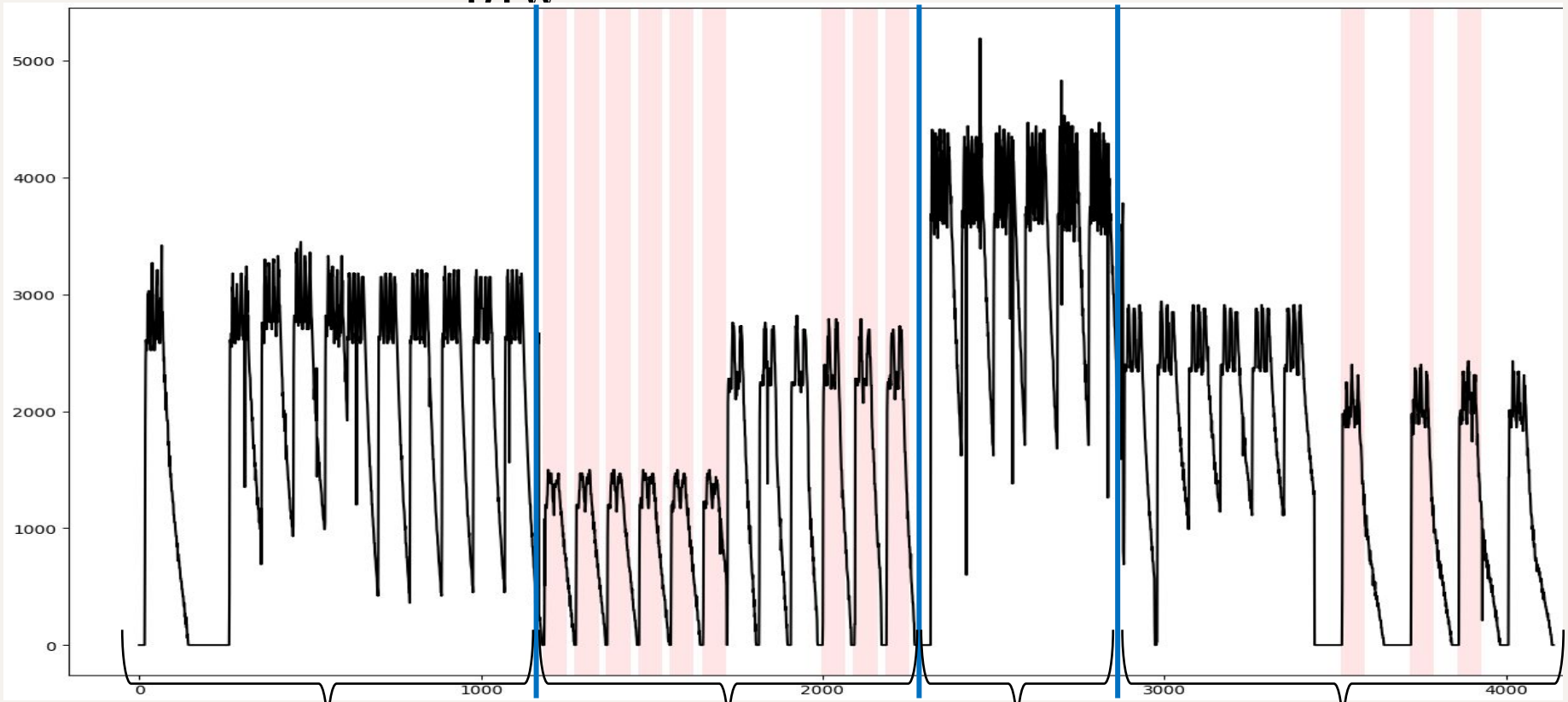


2 wrong detections

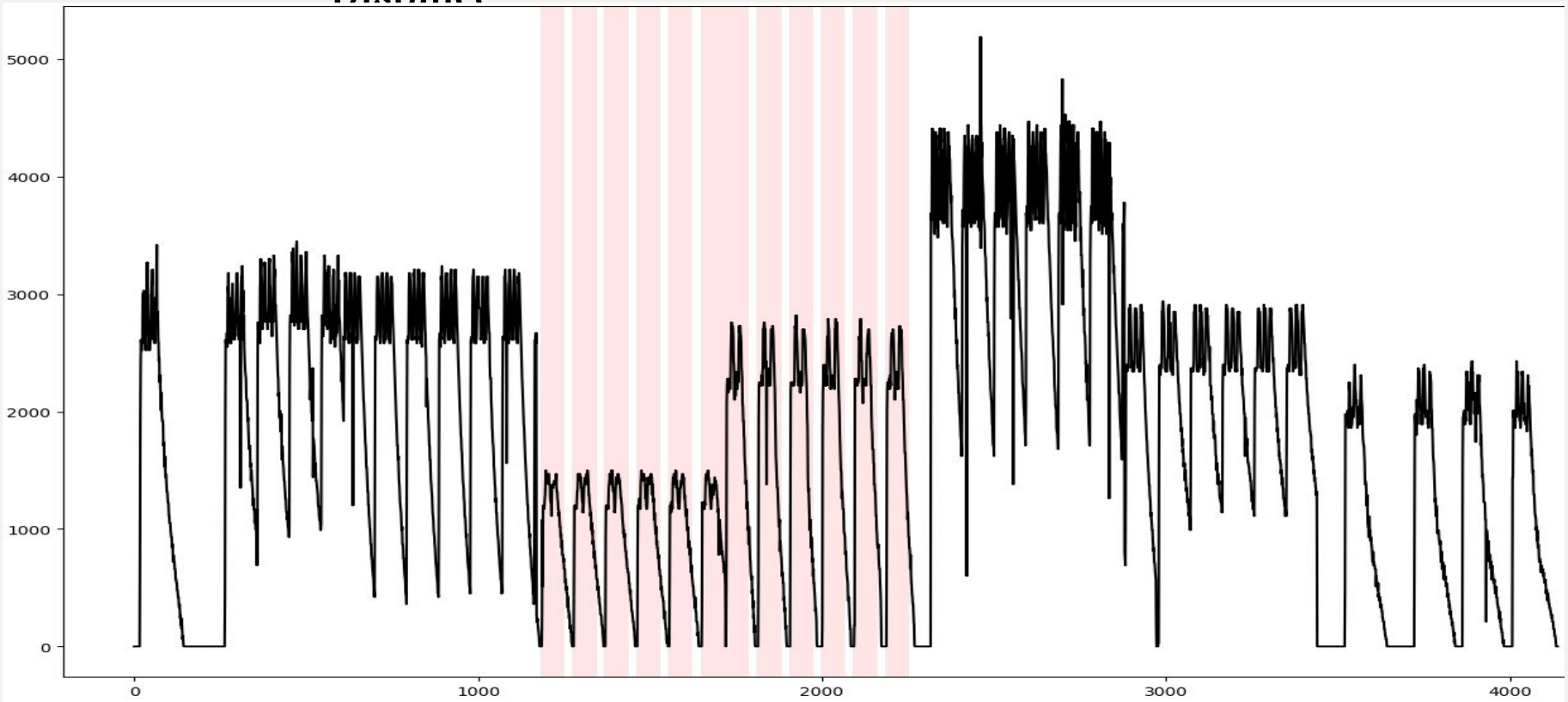
Euclidean Distance with Standardization



Detecting a 2-Picks Pattern - DTW



Detecting a 2-Picks Pattern - Euclidean Distance



DTW: 9/12 good detections

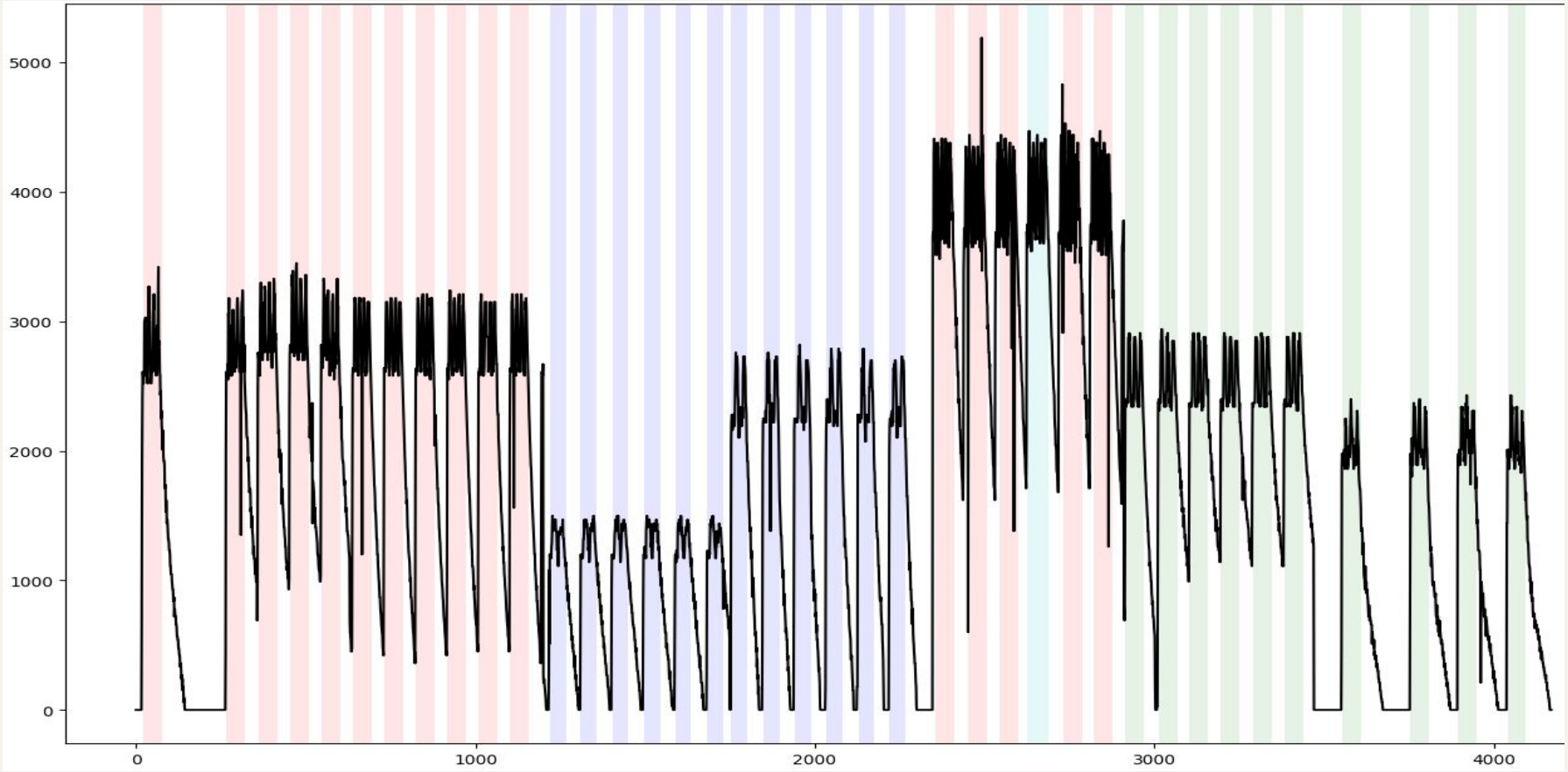
Euclidean Distance: 12/12 good detections

05.2

Results

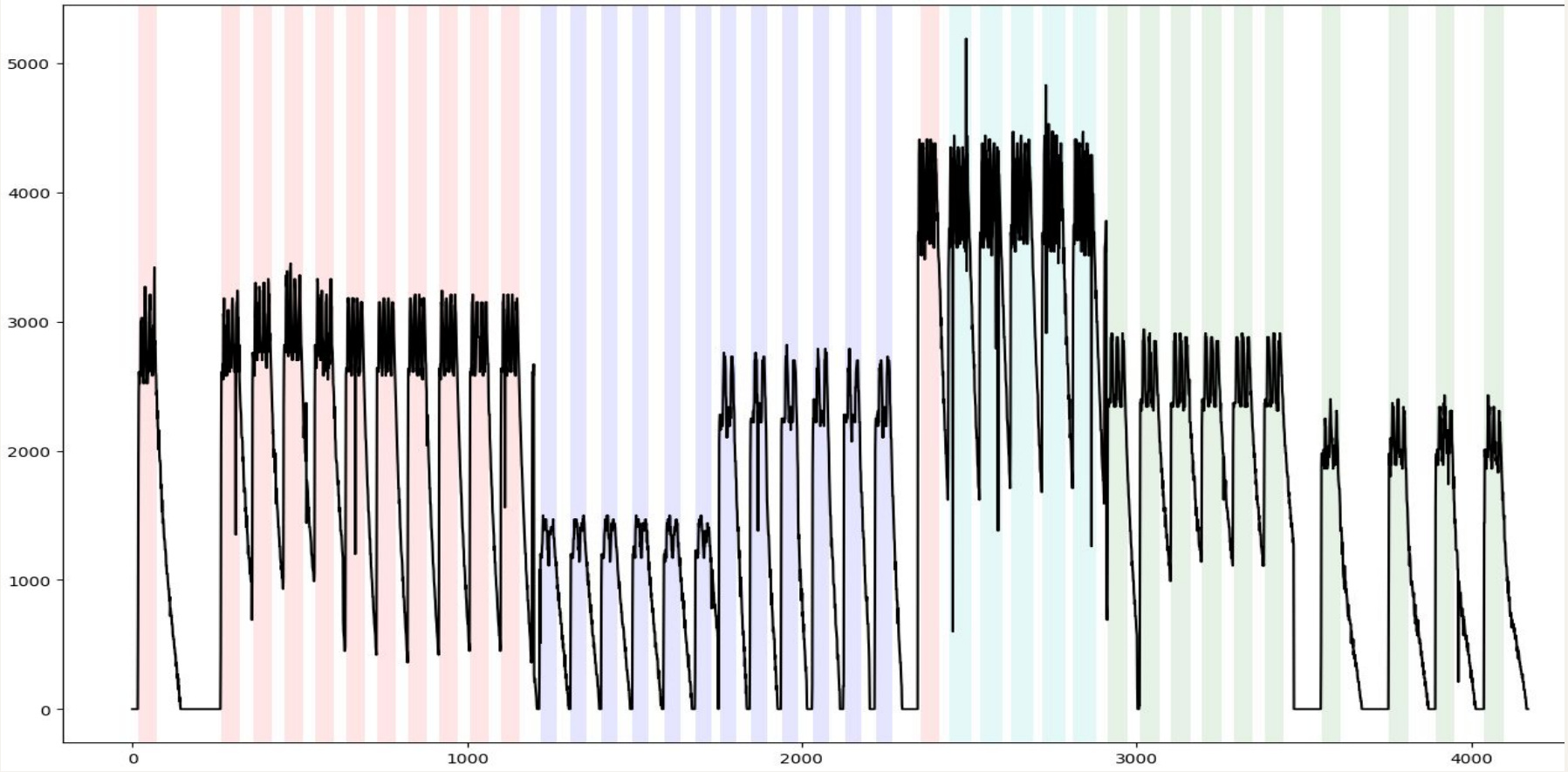
Multiple Patterns

Multiple Patterns - DTW



None of the 5-Picks patterns have been detected

Multiple Patterns - Euclidean Distance



Only 1 wrong detection

Analysis

- Global detections:
  - In most cases, same results for DTW and ED
  - Neither DTW nor ED very good when big differences in amplitude

- Local detections using standardization:
  - ED equal or better than DTW in nearly every test
  - Big difference in computation time:
    - 1s vs. 3min for single pattern
    - 10s vs. 15min for multiple patterns

06

Conclusion

Conclusion

- Euclidean Distance with Standardization → Best results
- Low computation time

→ Should definitely be implemented in AIXPert

- In most cases, DTW worse than Euclidean Distance
- High computation Time

→ Not helpful in this testcase

→ Tests on another testcase may provide new insights

Thank you!