



# Parameter Free Piecewise Dynamic Time Warping for time series classification

**Vanel Siyou Fotso** **Engelbert Mephu Nguifo** **Philippe Vaslin**

Laboratoire d'Informatique, de Modelisation et d'Optimisation des Systemes (LIMOS)

Blaise Pascal University, Clermont-Ferrand II, France

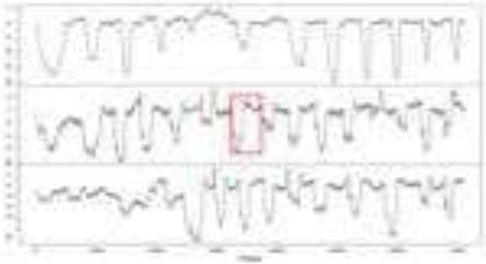
# Context and motivations

Siyou *et al.*, CML@ICML 2015

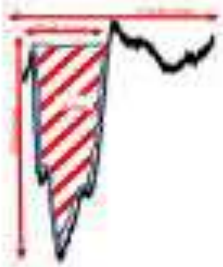
## CONTEXT AND PROBLEM

- The biomechanical analysis of human being movements during their locomotion is performed with various measuring instruments (sensors)

- Those measuring instruments recorded long time series composed of many cycles or patterns, representative of the movements made and effort produced by the subject during his displacement

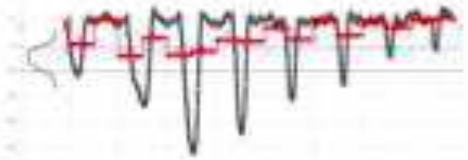


- These cycles are the time series analysis units and have several characteristic properties : minimum, maximum, duration, mean, median, standard deviation, interquartile range, the area below the cycle.

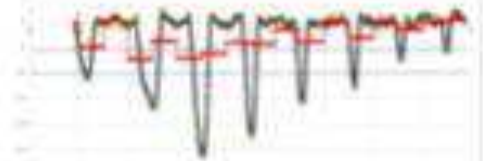


- Piecewise Aggregate Approximation

- Symbolic Aggregate Approximation



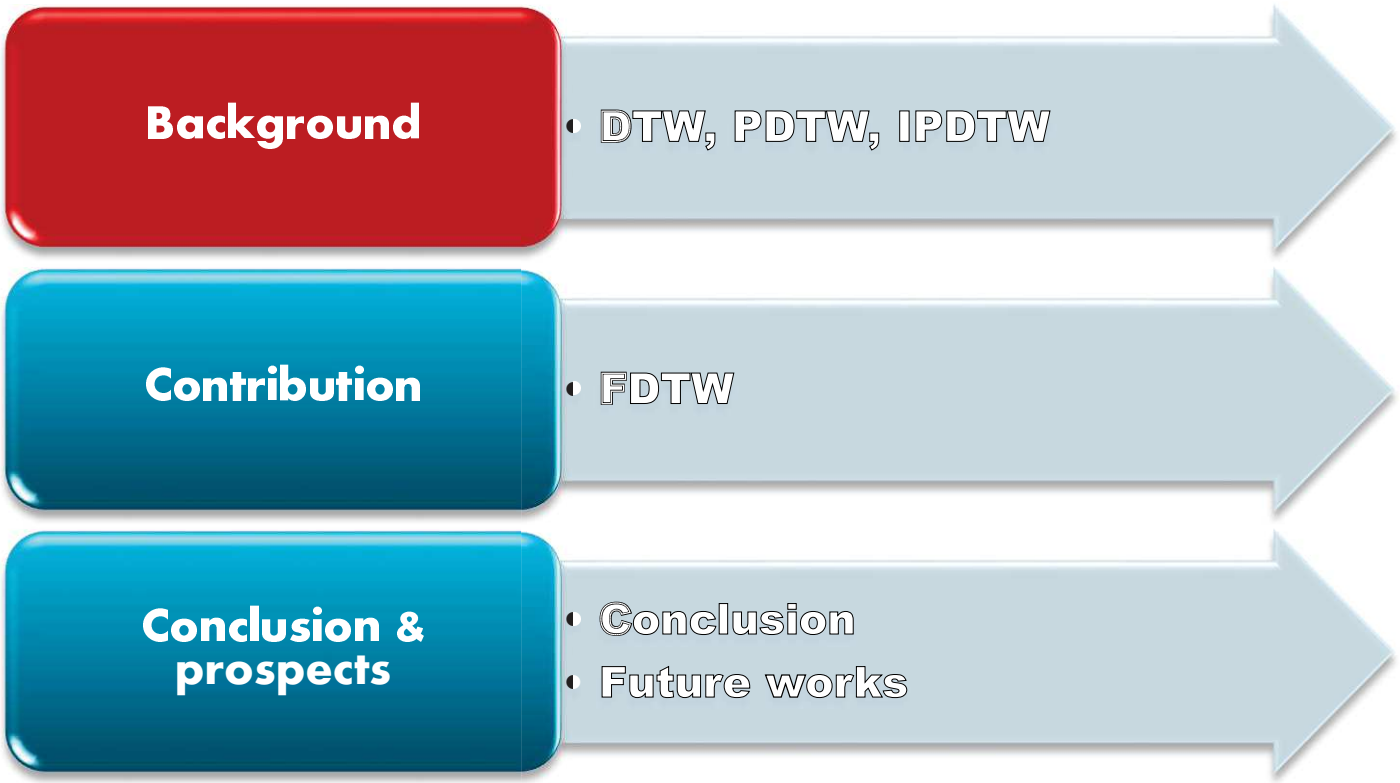
- ESAX



- SAX-TD

- Our goal is to provide a symbolic representation that takes into account several properties for each cycle, but without increasing the number of symbols used for the representation.

# Outline


Background
DTW, PDTW, IPDTW
Contribution
FDTW
Conclusion & prospects
Conclusion
Future works

# Time series classification

- Web site comparison

- PAA + DTW is among the good classifiers

« No free lunch theorem »

# Time series

**Definition**

$X = x_1, \cdots, x_n$is a sequence of numerical values representing the evolution of a specific quantity during the time. $x_n$ is the most recent value.

**Definition**

A segment $X_i$ of length $l$ of the time series $X$ of length $n$ $(l < n)$ is a sequence constituted by $l$ consecutive variables of $X$ starting at the position $i$ and ending at the position $i + l - 1$. We have:
$X_i = x_i, x_{i+1}, \ldots, x_{i+l-1}$

# Time series

**Definition**

The arithmetic average of the data points of a segment $X_i$ of length $l$ is noted $\bar{X}_i$ and is defined by:

$$\bar{X}_i = \frac{1}{l} \sum_{j=0}^{l-1} x_{i+j}$$

# PAA, PDTW

**Definition**

Let $T$ be the set of time series. The Piecewise Aggregate Approximation (PAA) is defined as follows:

$$PAA : T \times \mathbb{N}^{\star} \to T$$

$$(X, N) \mapsto PAA(X, N) = \begin{cases} \bar{X}_1, \cdots, \bar{X}_N \text{ if } N < |X| \\ X \text{ otherwise} \end{cases}$$

**Definition**

Let $N \in \mathbb{N}^{\star}$, $X$ and $Y$ be two time series.

$$PDTW(X, Y, N) = DTW(PAA(X, N), PAA(Y, N)).$$

# DTW

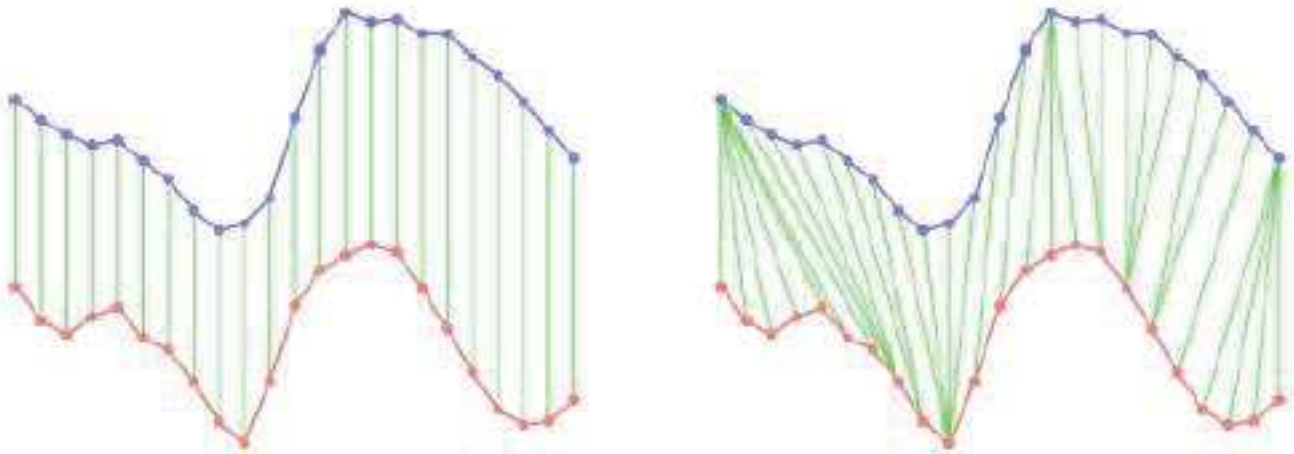


Figure: Euclidean distance (left) - DTW (right)

- The **boundarie** condition: The first (respectively last) point of both time series must be aligned.
- The **monotony** condition: during alignment there is no return to a point which has already been used.
- The **continuity** condition: when aligning all data points are considered

# DTW

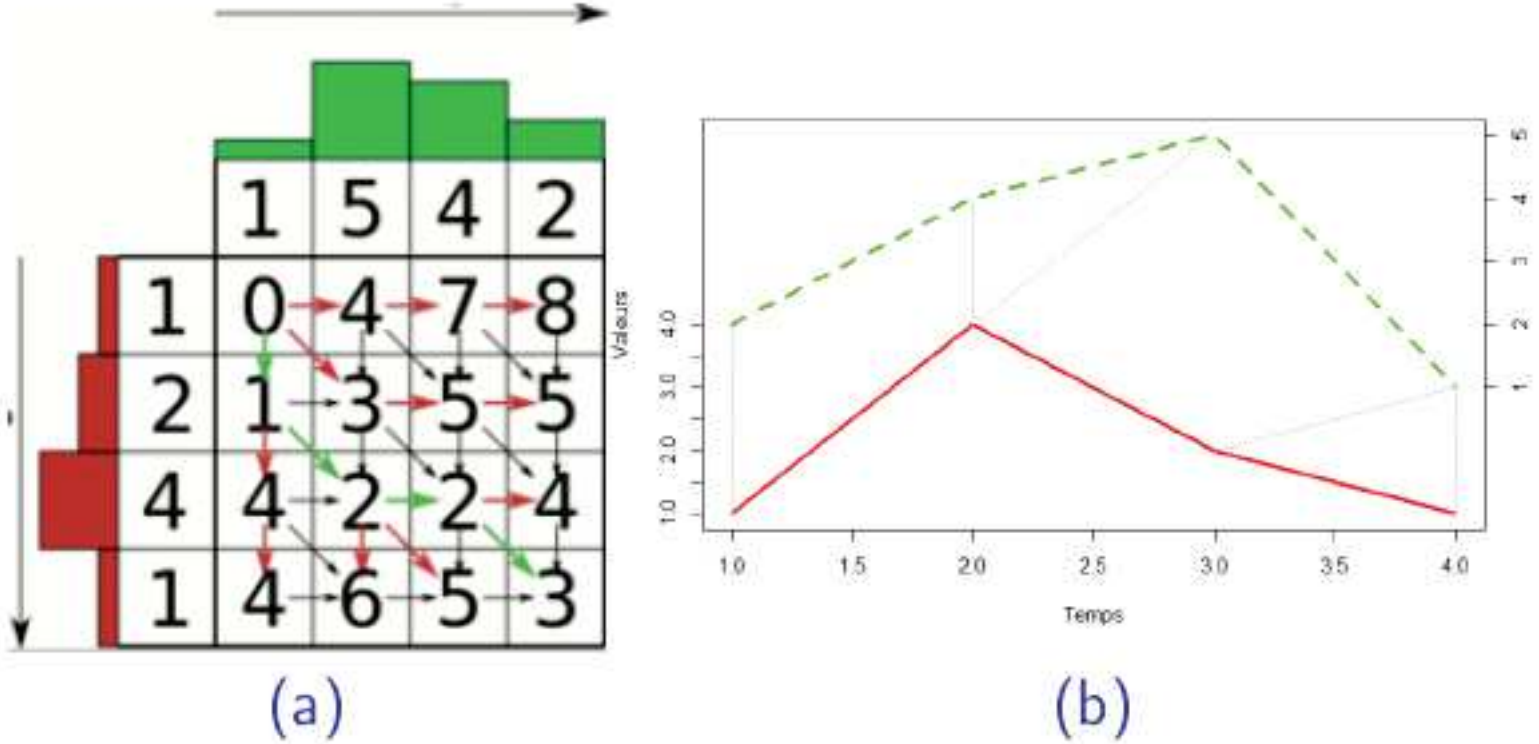
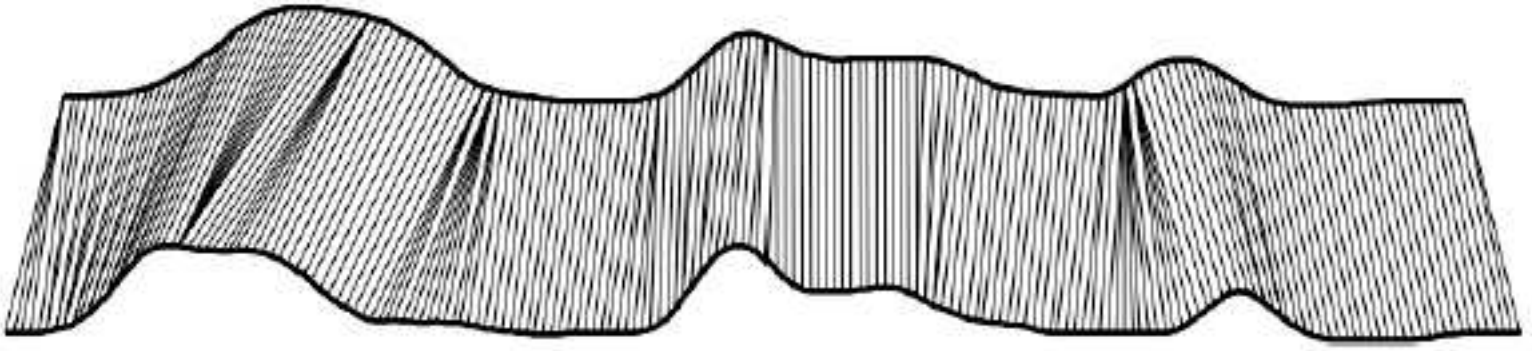
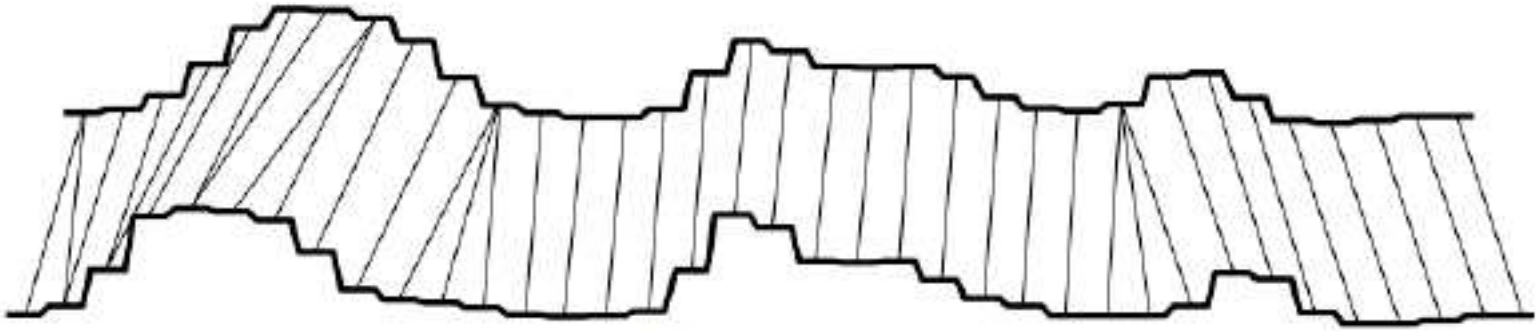



Figure: Alignment example with DTW

# DTW, PDTW



(a) DTW



(b) PDTW

Figure: Keogh et Al, KDD 2000

# PDTW

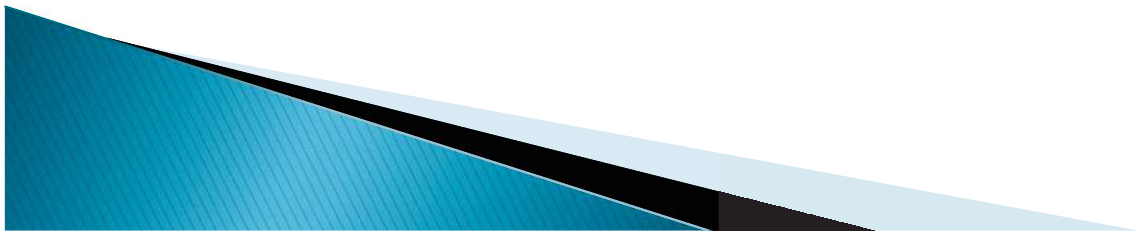
- An optimisation problem :
  - Finding the number of segments
- Objective function :
  - Accuracy
  - …
- Possible solutions
  - Brute Force approach
  - IDDTW (Shu et al., SDM 2002)

# IDDTW

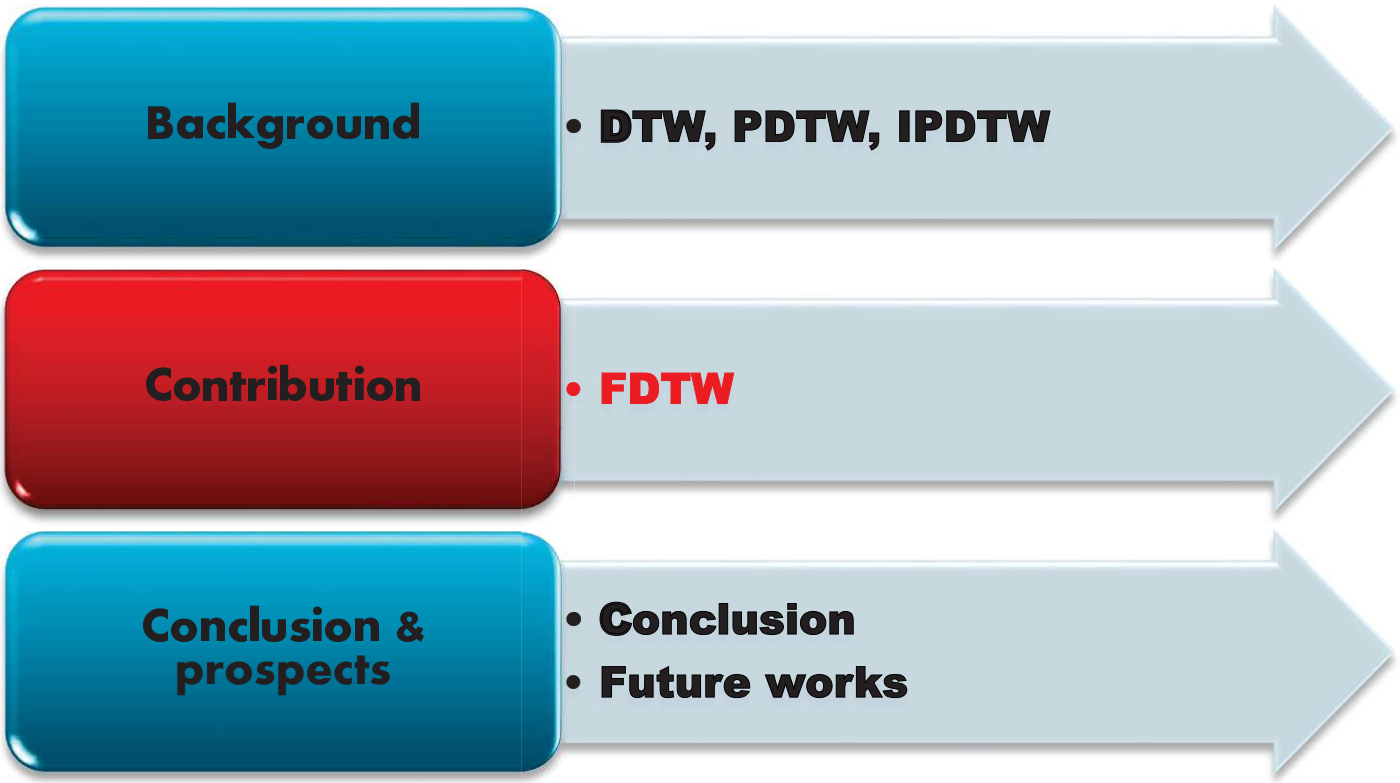
Shu et al., SDM 2002

Considers the number of segments which are powers of two :

$$1, 2, 4, \ldots, 2^k$$



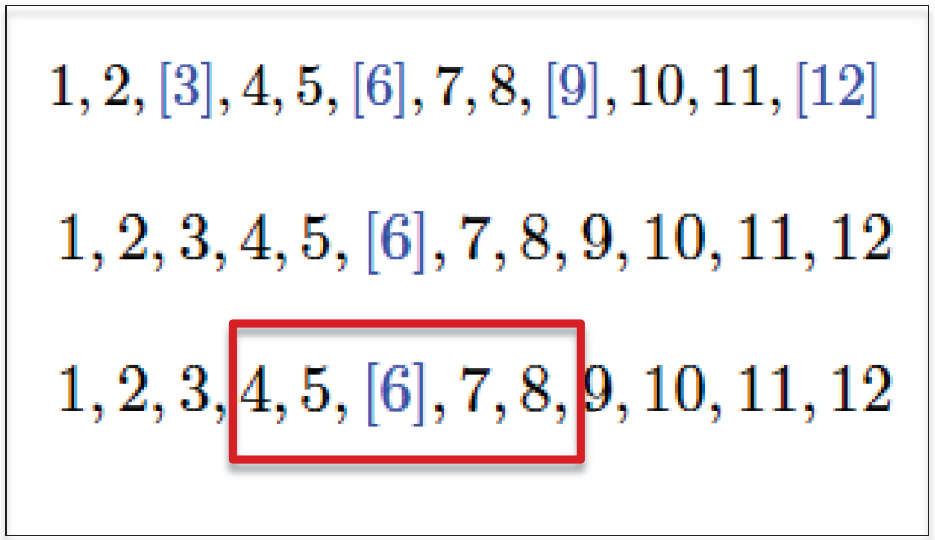
# Outline


Background
• DTW, PDTW, IPDTW
Contribution
• FDTW
Conclusion & prospects
• Conclusion
• Future works

# FDTW

- Idea :
  - Express the number of trials as a function of the time series length, n
  - Look at the minimum number of trials with equal width, Nc
  - Nc = square root (2n)

- Example :
  - n = 12
  - Nc = 4 ➔ {3, 6, 9, 12}
  - Min accuracy with 6
  - Explore [4,8]



# FDTW

**Fact**

*In short, in the worst case, we test the* $N_c$ *first candidates to find the best one. Then, we test* $\frac{2n}{N_c}$ *other candidates to find the local minimum. We finally perform* $nb(N_c) = N_c + \frac{2n}{N_c}$ *tests. The minimal number of tests is done when the number of candidates* $N_c = \sqrt{2n}$.

# FDTW

Lemma :

*For a given a dataset* $d_i$ $FDTW(d_i) \leq 1NNDTW(d_i)$ . *The quality of the alignment of our heuristic is better than that of DTW.*

Proof

$1NNDTW(d_i) = 1NNPDTW(d_i, n)$. $1NNDTW(d_i)$is then one of the candidate considered by the heurisitic *FDTW*. Since *FDTW*returns the minimal classification error from all candidates, the classification error of $1NNDTW$is always greater than or equal to *FDTW*. □

# FDTW

Proposition :

*For a given dataset $d_i$ that has $c_i$ classes, $c_i \in \mathbb{N}^*$,*

$$acc_{DTW} \geq \frac{1}{c_i} \implies \frac{1}{c_i} \times acc_{max} \leq acc_{FDTW} \leq acc_{max}$$

# FDTW

**Algorithm 1** FDTW(training_set, test_set, n, nb_rep=log(n))

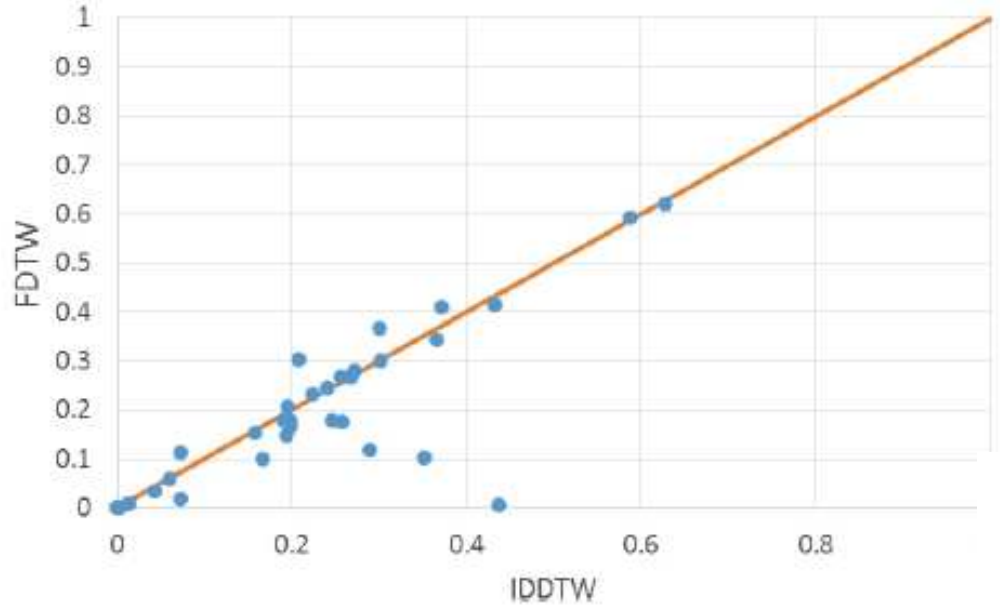
```
# Look for a good value of the number of segments
N
# using the training set
for (i in 0 : (nb_rep − 1)) do
  tab_N ← 1 : (n − i)
  l ← floor(n/sqrt(2 ∗ n))
  tab_N_candidats ← seq(from = n, to = 1, by =
  −l)

  # Parallel execution of 1NNPDTW
  mat_r ← 1NNPDTW(training_set,
  tab_N_candidats)

  # Mark candidates already used to not reuse
  for (i in tab_candidats) do
    tab_N[i] ← −1
  end for

  # Search for the best candidate with the minimal
  error
  min ← minimun(mat_r)

  # look for the local minimun near of the best
  candidate
  result[[(i + 1)]] ← localMinimun(min.N_min,
  min.error_min, training_set, tab_N)
end for

# The best local minimal error
m ← minimun(result)
return m
```

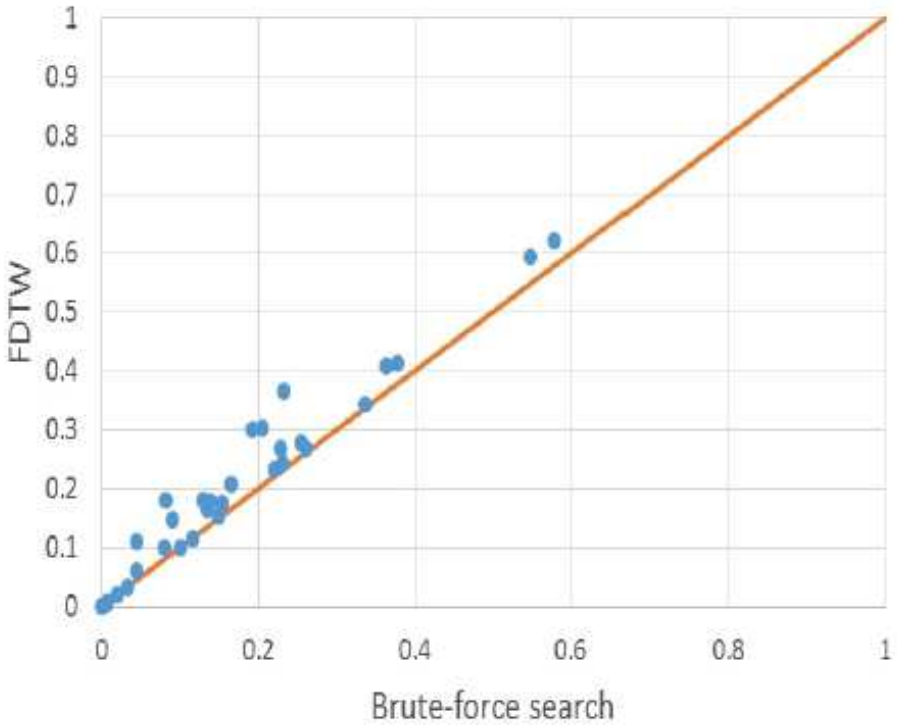
# Experiments & results

## Evaluation methodology

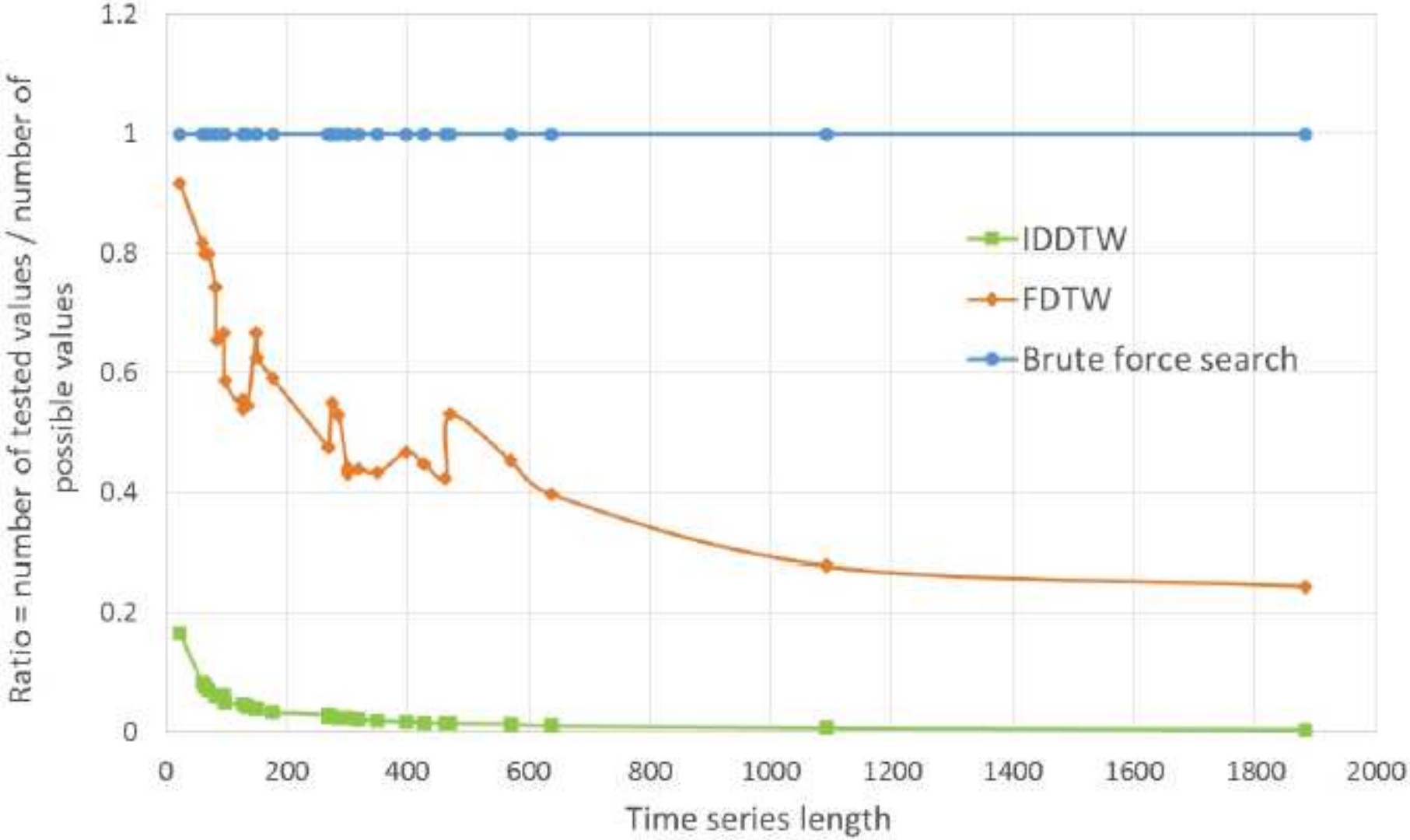
- **Interestingness**
  - Classification accuracy
  - Number of trials
- **Runtime**

# Experiments & results : Accuracy

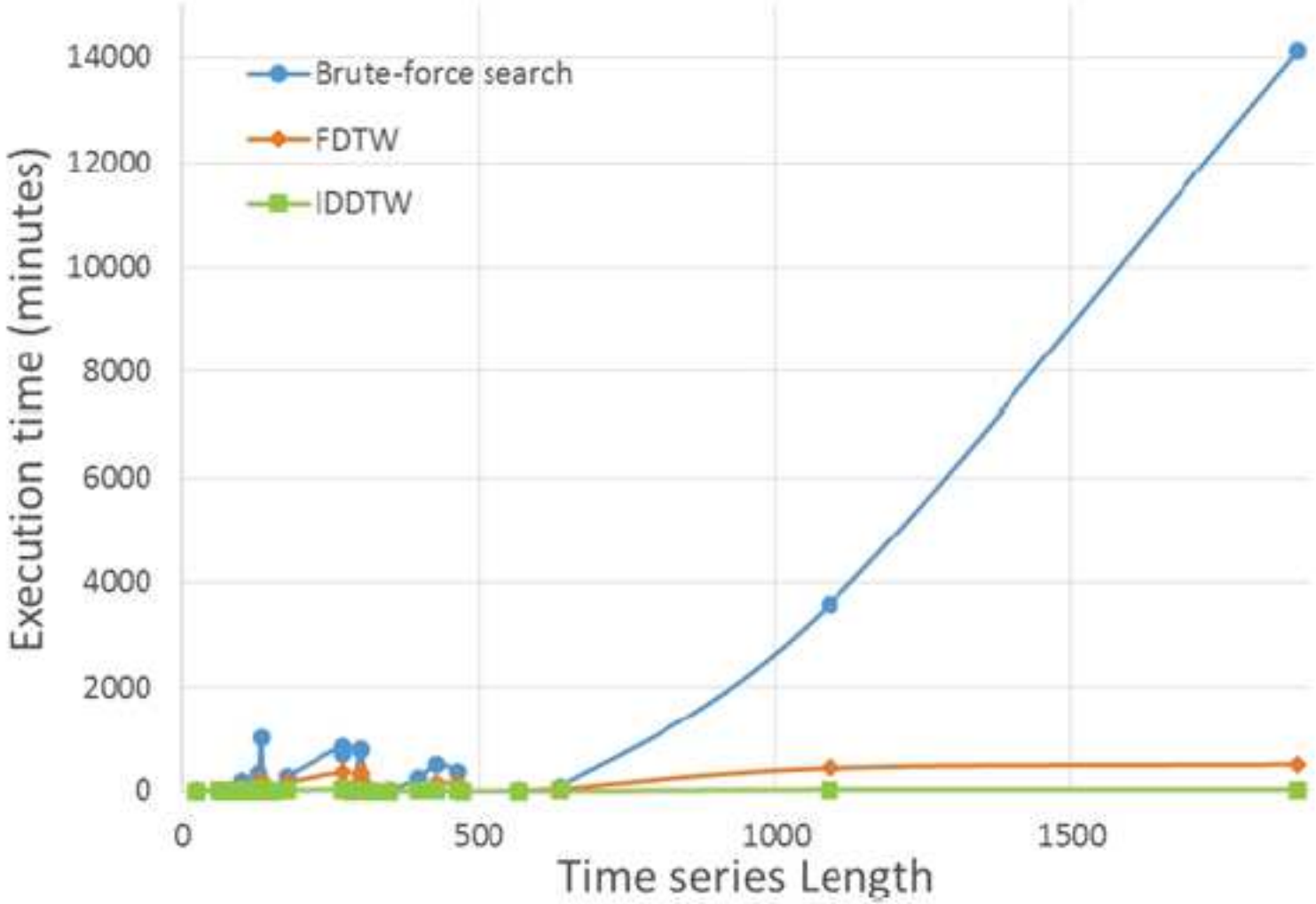


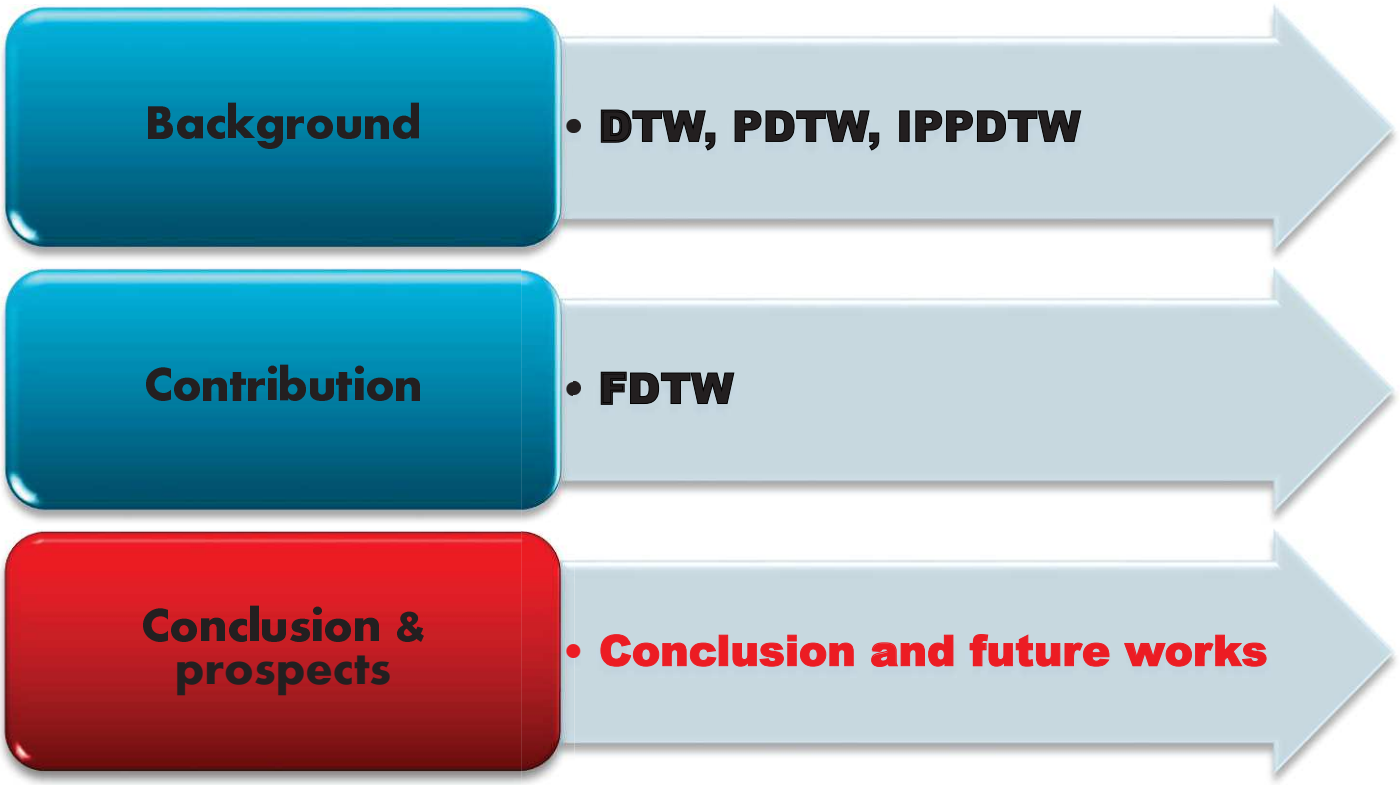
# Experiments & results : #trials ratio



# Experiments & results : Runtime



# Outline


Background
• DTW, PDTW, IPPDTW
Contribution
• FDTW
Conclusion & prospects
• Conclusion and future works

# Conclusion

- Choice of segments length for Time Series Analysis
- The proposed approach:
  - Allows selecting a minimal number of trials
  - Searches for a local minimum
  - Can be extensible

# Future works

- Extension to global minimum
- Uncertainty

# Thanks