Dissimilarity-based Classification of Multidimensional Signals by Conjoint Elastic Matching

Application to Phytoplanktonic Species Recognition

Émilie Caillault[1, 2] and Pierre-Alexandre Hébert[1, 2] and Guillaume Wacquet[1, 2]

1 University Lille Nord de France, F-59000 Lille, France.

2 ULCO, LASL, 50 rue Ferdinand Buisson, F-62228 CALAIS, France.

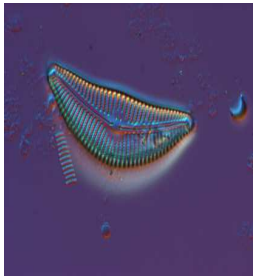
Emilie.Caillault ; Pierre-Alexandre.Hebert@lasl.univ-littoral.fr

Context : BQR PhytoClas project

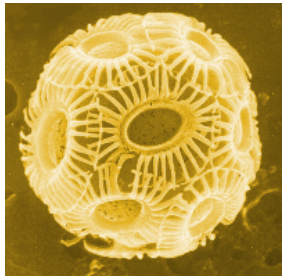
- **Aims**
  - To detect, class and count phytoplanktonic species
- **Funds**
  - Local project for environmental, marine and fishing politics
  - Directive 2000/60/EC of the european parliament and of the council of 23 october 2000 establishing a framework for community action in the field of water policy. Official Journal of the European Communities EN 2000/60/EC (2000)
- **Participants**
  - Denis Hamad, Pierre-Alexandre Hébert, Émilie Caillault Guillaume Wacquet
  - Felipe Artigas, Natacha Guiselin, Xavier Mériaux, Elsa Breton
  - LASL, LOG Wimereux, CNRS, ULCO.

Context : Phytoplankton species recognition

- **What ?**
  - **Autotrophic component** of the plankton community.
  - **Microscopic algae** in suspension in the water, developing according to a seasonal cycle.
  - Length: from 1 to some 100 micrometers.
  - ~5000 known species which 70 toxic ones.

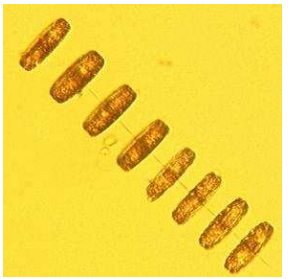


Diatoms

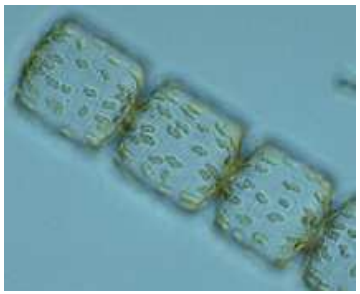


Emiliania huxleyi

- **Why ?**
  - Phytoplankton account for **half of all photosynthetic activity** on Earth.
  - Thus phytoplankton are responsible for much of the oxygen present in the Earth's atmosphere – half of the total amount produced by all plant life.
  - Phytoplankton are a **key food item** in both aquaculture and mariculture.



Thalassiossira rotula



Lauderia annulata

⇒ Important factor in environmental, economic, ecological politics

Directive 2000/60/EC

Context : Phytoplankton species recognition

- **Flow-cytometry**

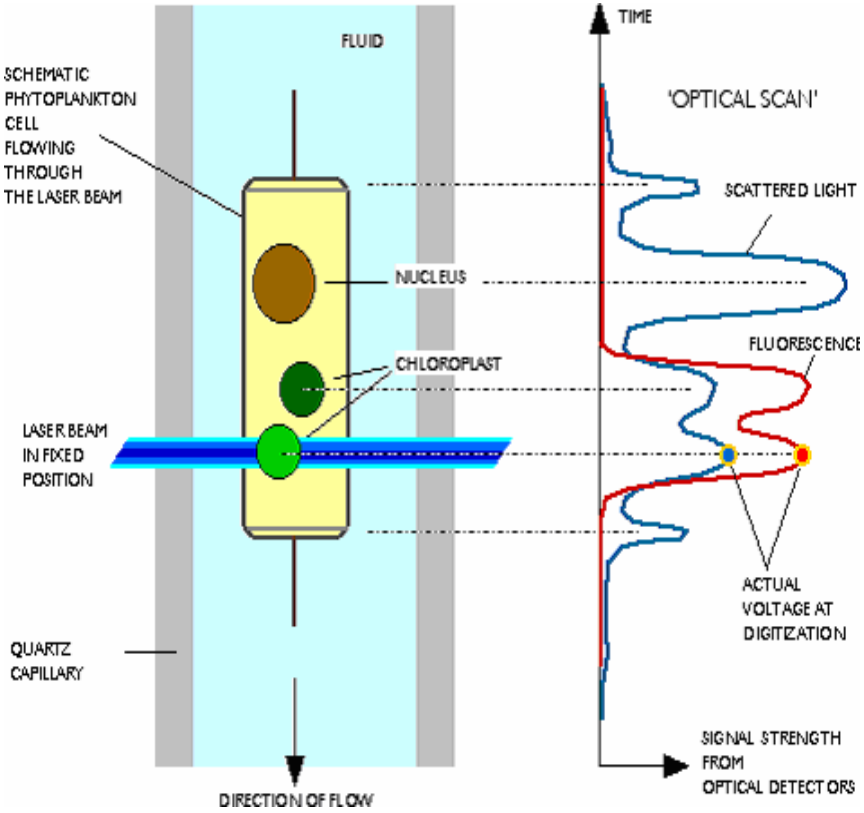
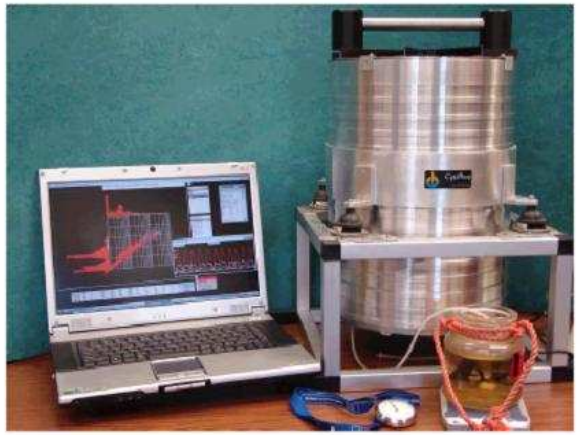


Image © Cytobuoy

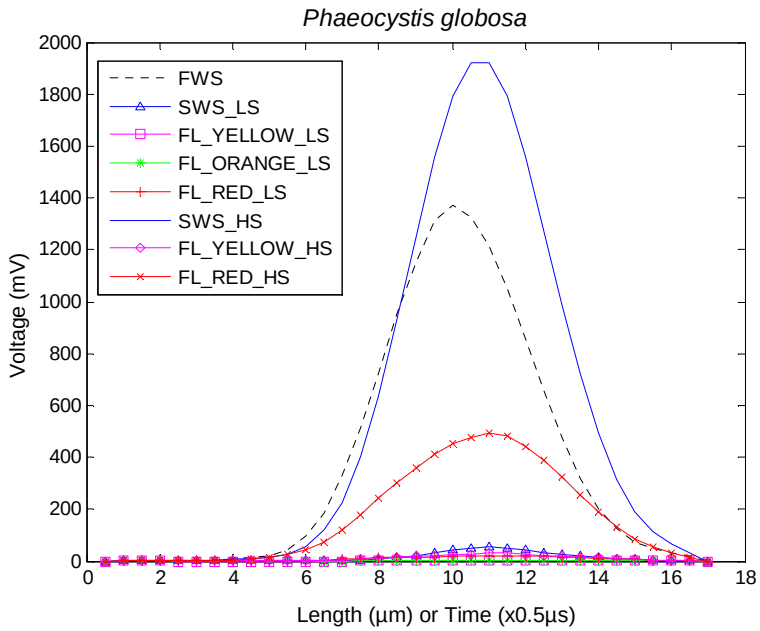


- **Particles drived by a liquid flow at high speed in front of a laser light**
- **Advantages**
  - Temporal signals
  - Information about size, structure and pigments
  - Mobility, easy to use
  - Alive cells
- **Drawbacks**
  - Colonies difficult to count

Context : Phytoplankton species recognition

Cytometric curves

- 8 raw time signals per cell
- identical experimental conditions: same sampling rates, same detection threshold, etc.

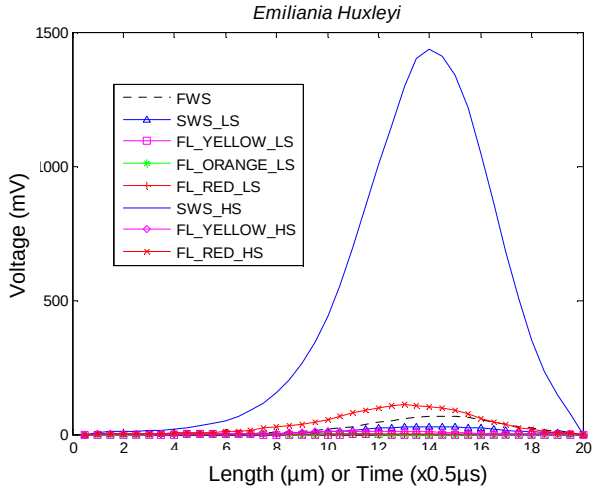


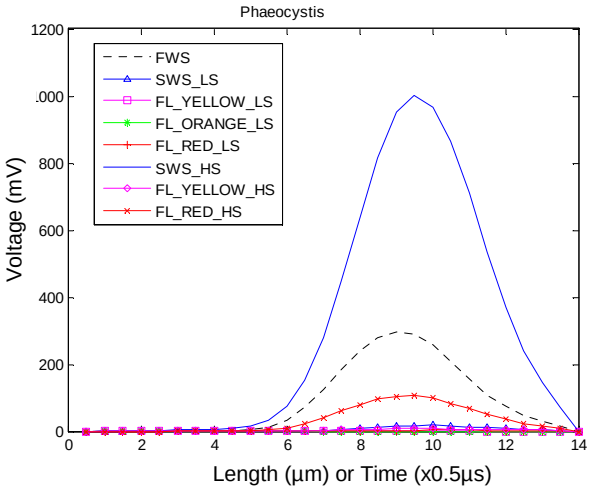
-one signal on forward scatter (FWS), corresponding to the cell external structure;
-two signals on sideward scatter (SWS), corresponding to the internal structure;
-two signals on red fluorescence (FLR) which characterize chlorophyll pigments;
-one signal on orange fluorescence (FLO), ages, specific pigments;
-two signals on yellow fluorescence (FLY), specific pigments.

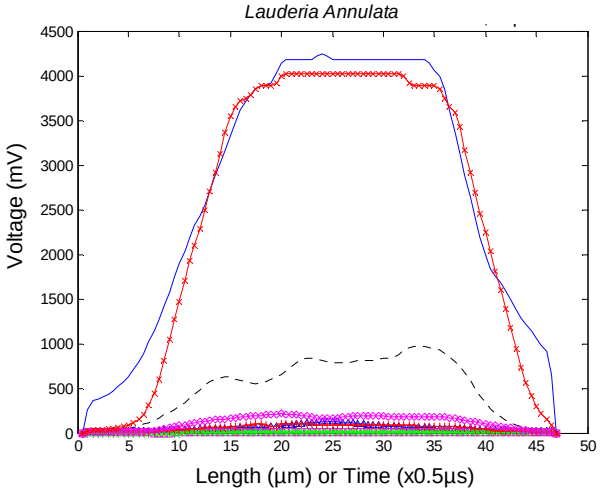
microscopic photo

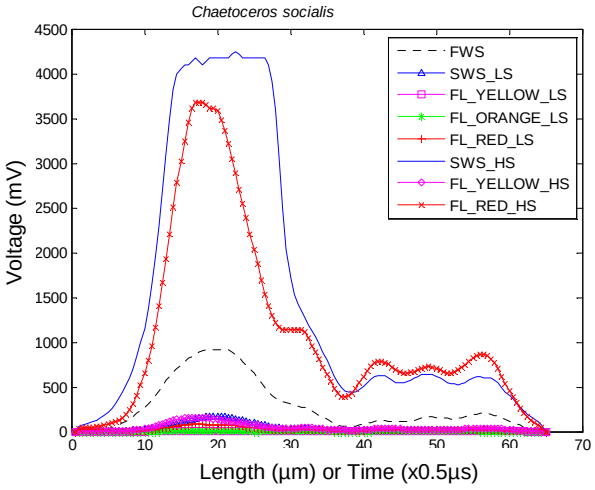
Context : Phytoplankton species recognition

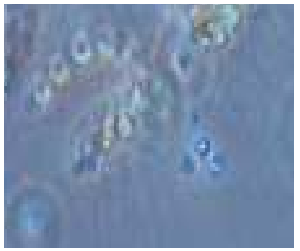
Cytometric curves (=profiles) inter-species





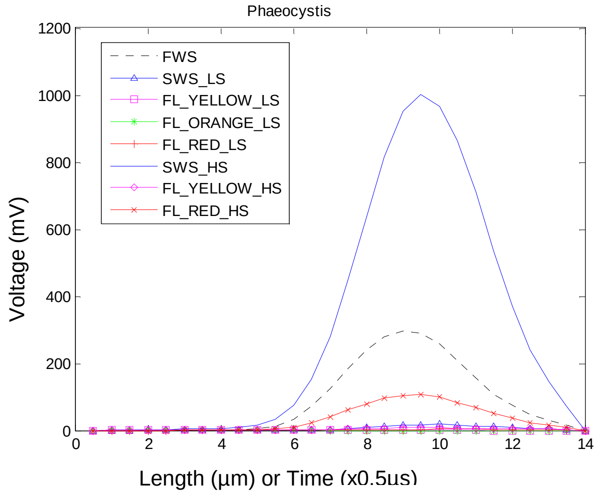


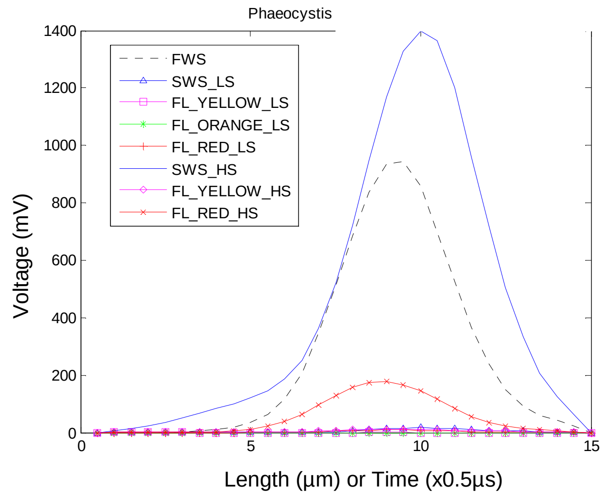


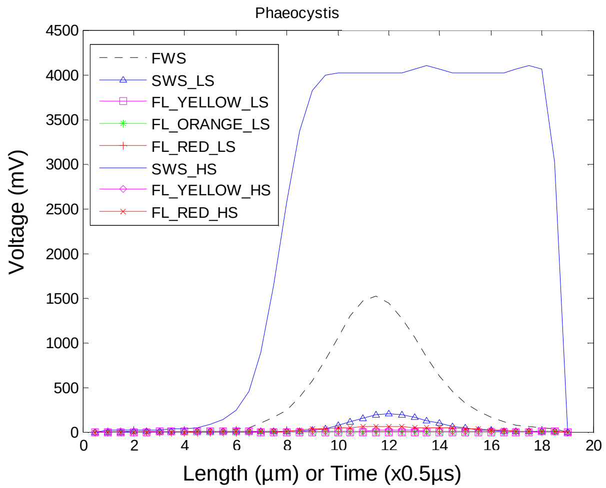


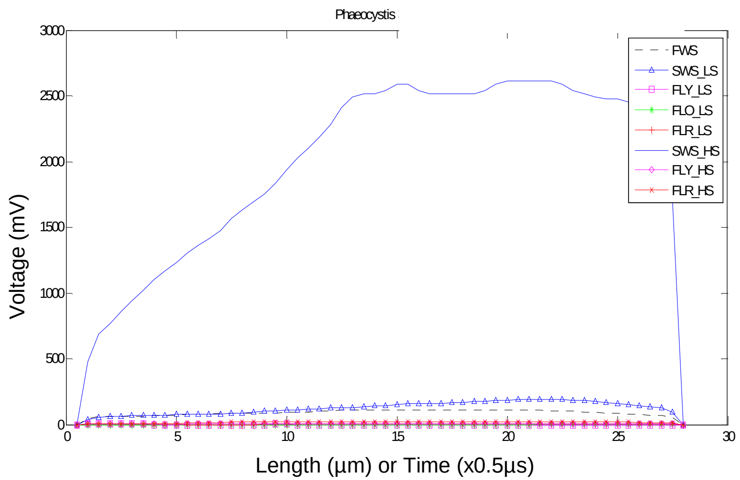
Context : Phytoplankton species recognition

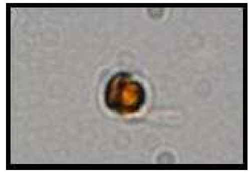
Cytometric curves intra-species

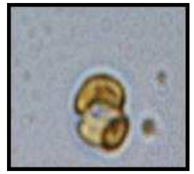


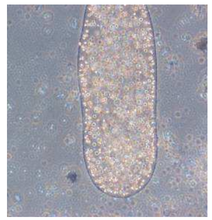








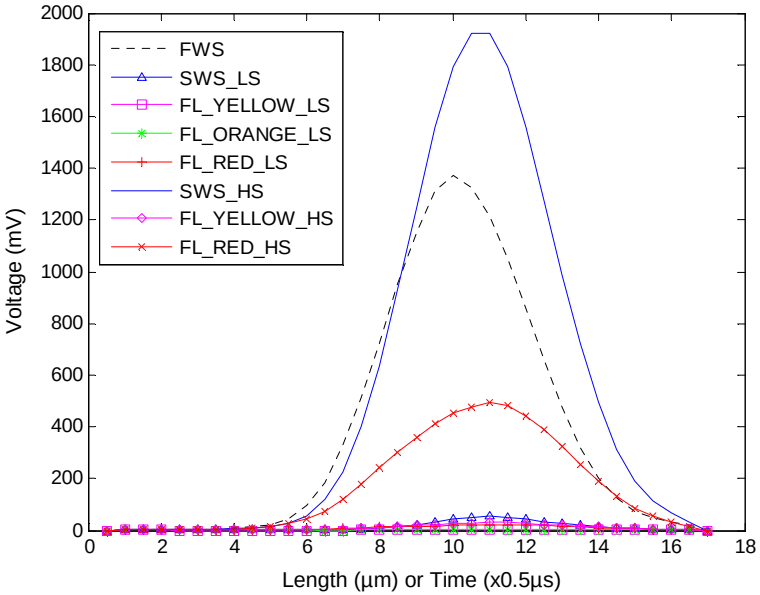




Context: Phytoplankton species recognition

Classification problem

- Cell identification from their cytometric curves



Which one ?

Known profiles:

Chaetoceros socialis
Emiliania Huxleyi
Lauderia annulata
Leptocylindrus minimus
Phaeocystis globosa
Skeletonema costatum
Thalassiosira rotula.

Aims:

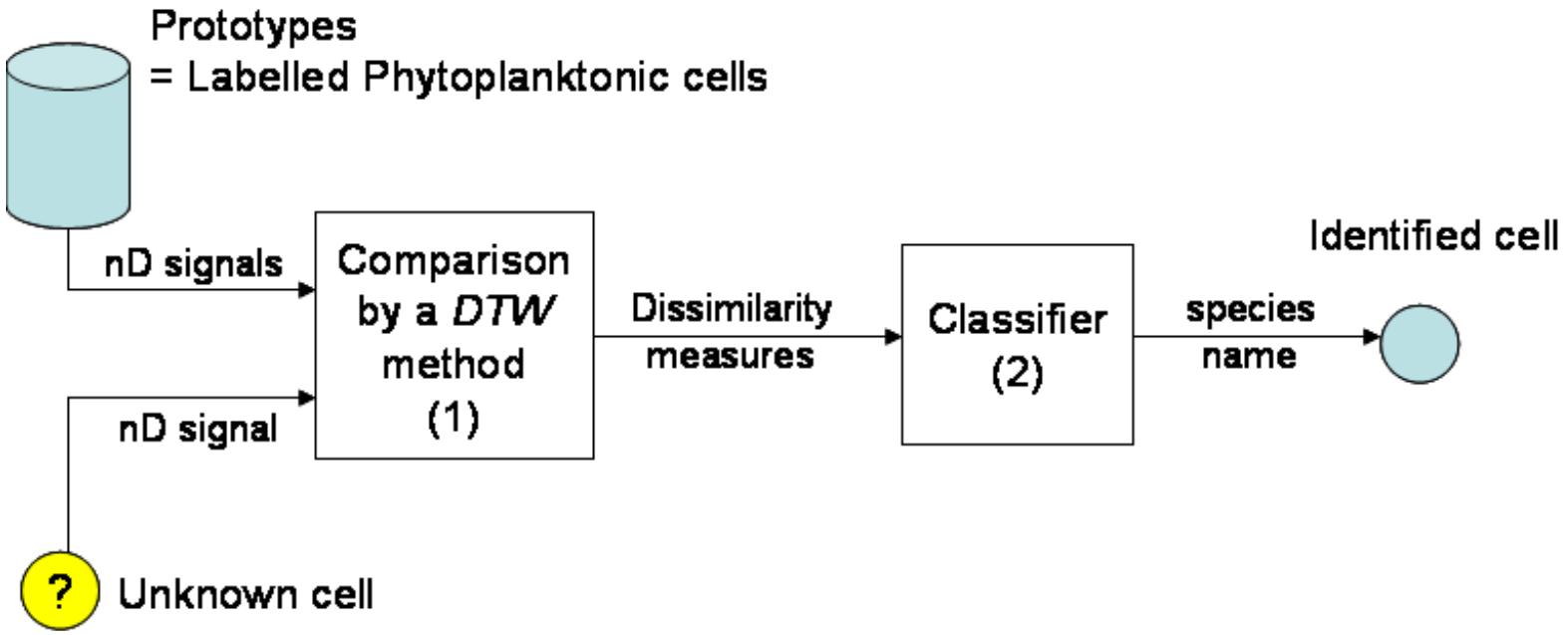
- Make the species recognition process automatic.
- Build a comparison measure between cytometric profiles robust to:
  - the intra-species variability,
  - the sensor sensibility.

2 methods:

- Features
- Comparisons

Method: DTW-dissimilarity based classification

- Scheme of the classification system for Phytoplanctonic cells



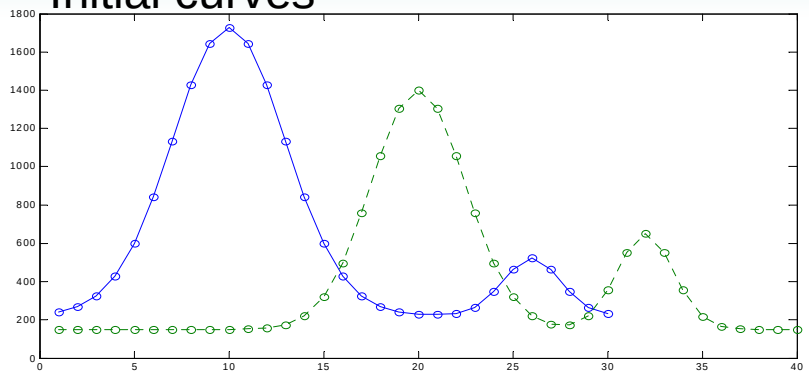
- 7 species trained
- Each cell characterized by an 8D signal

Method: DTW-dissimilarity based classification

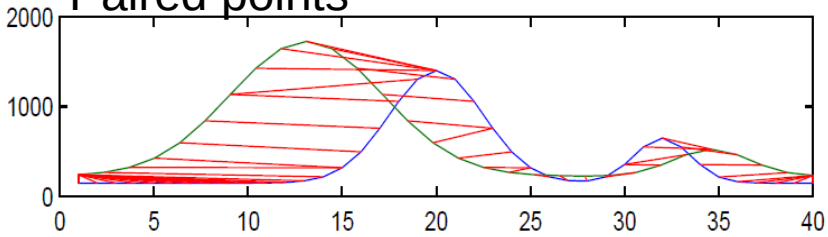
DTW: Dynamic Time Warping

- Sakoe and Chiba's work – 1978
  - Elastic maching between two curves.
  - Look for the best matching between points corresponding to the best time distorsions, minimizing a global distance cost.
  - Cost defined as a sum of intensity gaps between paired points.
  - Constraints:
    - Time continuity,
    - Extrem points paired,
    - Time monotonicity.

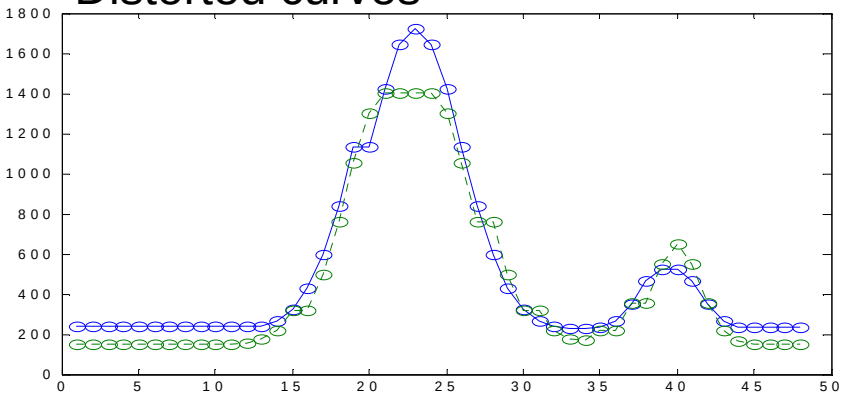
Initial curves



Paired points



Distorted curves



Method: DTW-dissimilarity based classification

DTW: Dynamic Time Warping

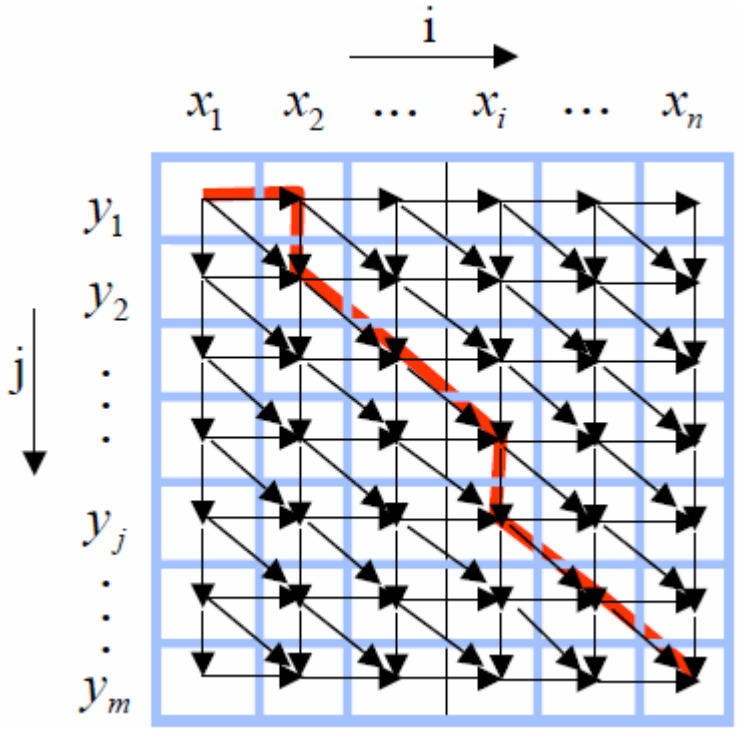
- Sakoe and Chiba's algorithm

$$
\begin{aligned}
X &= \{(x_i), i = 1, \ldots, n\} \\
Y &= \{(y_j), j = 1, \ldots, m\} \\
P &= \{(i_k, j_k), k = 1, \ldots, n_k\} \\
W &= \{(w(k), k = 1, k = 1, \ldots, n_k\}
\end{aligned}
$$

Signals to be compared

Path of paired points

Weighting of the pairs



Recursive partial cost for matching points *i* and *j*

$$
\begin{aligned}
Dist(i,j) = \min\{ & Dist(i-1,j) + d(x_i, y_j), \\
& Dist(i,j-1) + d(x_i, y_j), \\
& Dist(i-1,j-1) + 2.d(x_i, y_j)\}.
\end{aligned}
$$

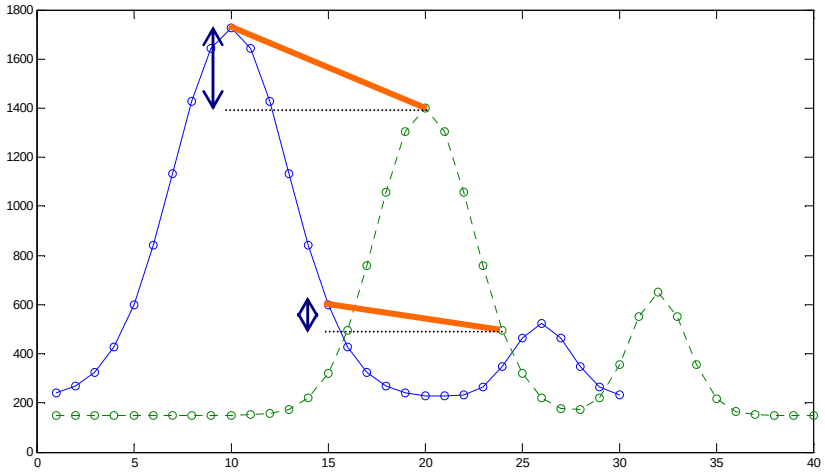
Global cost to be minimized

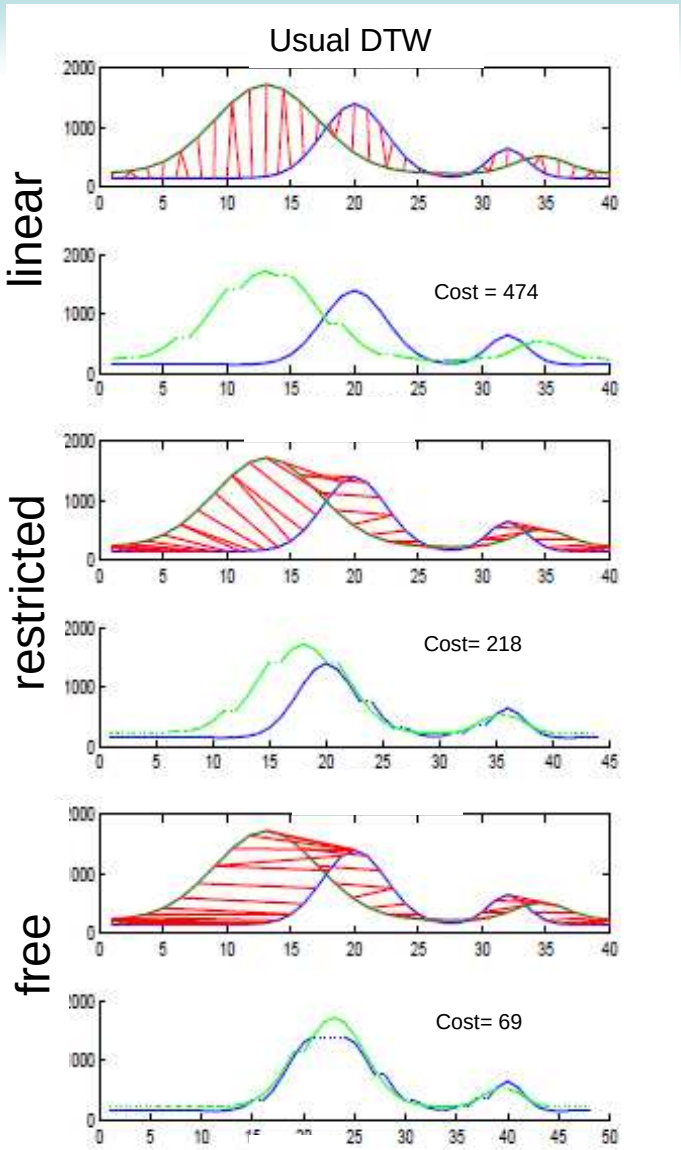
$$
\begin{aligned}
C(X,Y,P,W) &= \frac{\sum_{k=1}^{n_k} d(x_{i_k}, y_{j_k}).w(k)}{\sum_{k=1}^{n_k} w(k)} \\
&= \frac{Dist(n,m)}{\sum_{k=1}^{n_k} w(k)}.
\end{aligned}
$$

Method: DTW-dissimilarity based classification

DTW : Dynamic Time Warping

- Original method: Sakoe et Chiba
  - Distance measure





Method: DTW-dissimilarity based classification

DTW: Dynamic Time Warping

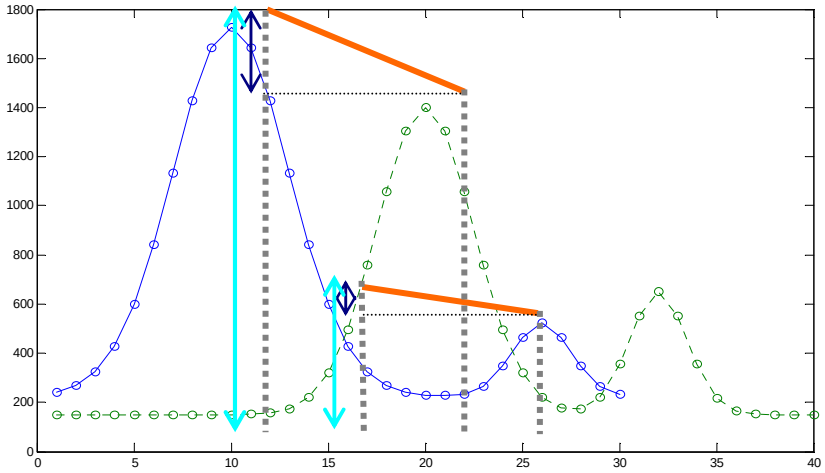
- Proposed « dissimilarity measure »
  - for positive signals (or same sign)

$$s(x_{i_k}, y_{j_k}) = \frac{d(x_{i_k}, y_{j_k})}{\max\{d(x_{i_k}, 0), d(y_{j_k}, 0)\}}$$

Distance measure

Normalisation

  - extension to nD signals



Cost	linear	restricted	free
DTW-distance	474,2	218,9	69,9
DTW-dissimilarity	0,52	0,31	0,17

DTW-dissimilarity: a qualitative measure

Application: Phytoplankton species recognition

Bench test

- **Comparison of two possible approaches**
  - Features-based approach: 32 extracted features.
  - Dissimilarity-based approach: 175 = 7 species * 25 prototypes
    - Original/usual DTW
    - Conjoint DTW
    - Variants: linear DTW, restricted-neighbourhood DTW, free DTW.

- **Used classifiers**
  - k-NN: 1-nearest neighbour.
  - MLP: multi layer perceptron
    - Sigmoïd transfert function
    - DTW-MLP : 175/91/7 neurones (input/hidden layer/output)
    - Features-MLP : 32/19/7 neurons.
  - SVM: support vector machine
    - Polynomial kernel (support vectors).

- **Cross validation**
  - 4 sets of 25 profiles * 7 species.

Application: Phytoplankton species recognition

Classification: dissimilarity-based vs DTW-distance

– 1-NN classifier

Training folds	Fold 1	Fold 2	Fold 3	Fold 4	**Mean**	Std
Classical distance-based DTW						
linear	93.3	90.8	94.2	92.1	**92.6**	1.4
10%-restricted	94.8	92.5	94.8	93.7	**94.0**	1.0
20%-restricted	96.3	92.9	94.6	93.1	**94.2**	1.5
no-restricted	96.1	90.2	93.5	91.8	**92.9**	2.5
Proposed dissimilarity-based DTW						
linear	97.7	94.8	95.0	96.1	**95.9**	1.3
10%-restricted	97.9	94.6	96.0	96.1	**96.1**	1.3
20%-restricted	98.2	95.4	96.1	97.1	**96.7**	1.2
no-restricted	97.3	95.6	96.0	96.9	**96.4**	0.8

Application: Phytoplankton species recognition

Classification: dissimilarity vs feature

- Features-based classification

Training fold	Fold 1	Fold 2	Fold 3	Fold 4	Mean	Std
1-NN	93.7	90.2	93.7	92.5	**92.5**	1.7
MLP	96.9	94.8	96	94.8	**95.6**	1.1
SVM1	90	87.4	91	92.5	**90.2**	2.2
SVM2	95	91.2	90	93.9	**92.5**	2.4

- Dissimilarity-based classification (10% restricted neighbourhood)

Training fold	Fold 1	Fold 2	Fold 3	Fold 4	Mean	Std
1-NN	98.2	95.4	96.1	97.1	**96.7**	1.3
MLP	98.2	97.3	97.3	96.7	**97.3**	0.7
SVM1	98.8	95.6	95.6	96.1	**96.5**	1.6
SVM2	92.3	93.5	93.3	92.9	**93**	0.6

Conclusions

DTW-dissimilarity based classifier/measure between nD signals

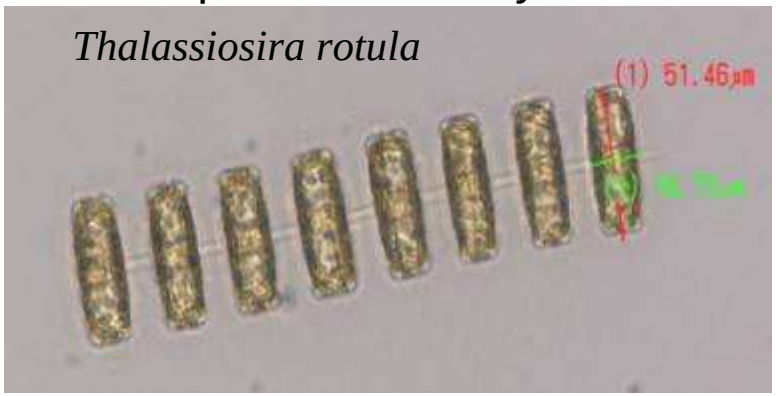
- Proposition of a dissimilarity measure, issued from Sakoe and Chiba's elastic matching method
  - Intuitive measure, easy to understand,
  - Extension to nD signals.

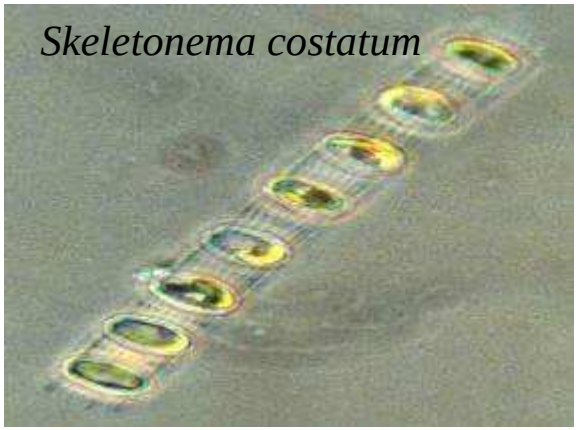
- Possible application to every nD signals
  - Handwritten characters recognition,
  - Scriptor identification,
  - Signatures verification (ICDAR 2009)
  - ...

- Promising results obtained for natural and culture marine samples of Eastern Channel
  - (identification + counting of 3 species: *Phaeocystis globosa, Pseudonitzia seriata, Chaetoceros socialis*).

- Boni

Different species in colony in the LOG database



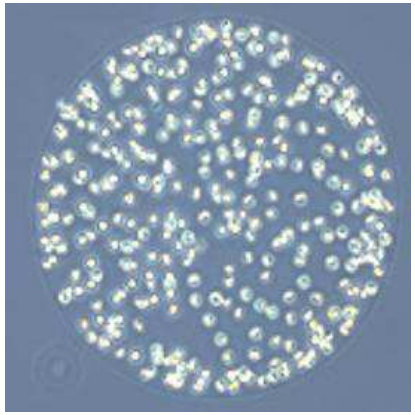


Lauderia annulata

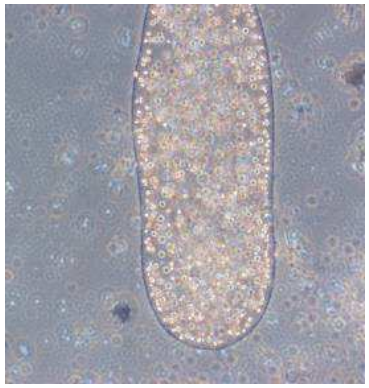
Leptocylindrus minimus

Different life cycle of *Phaeocystis globosa colonies*

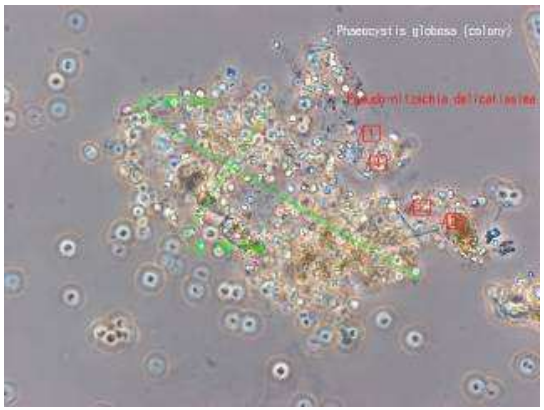
1)Young colony 50-100 µm



2) Colony 100-500 µm

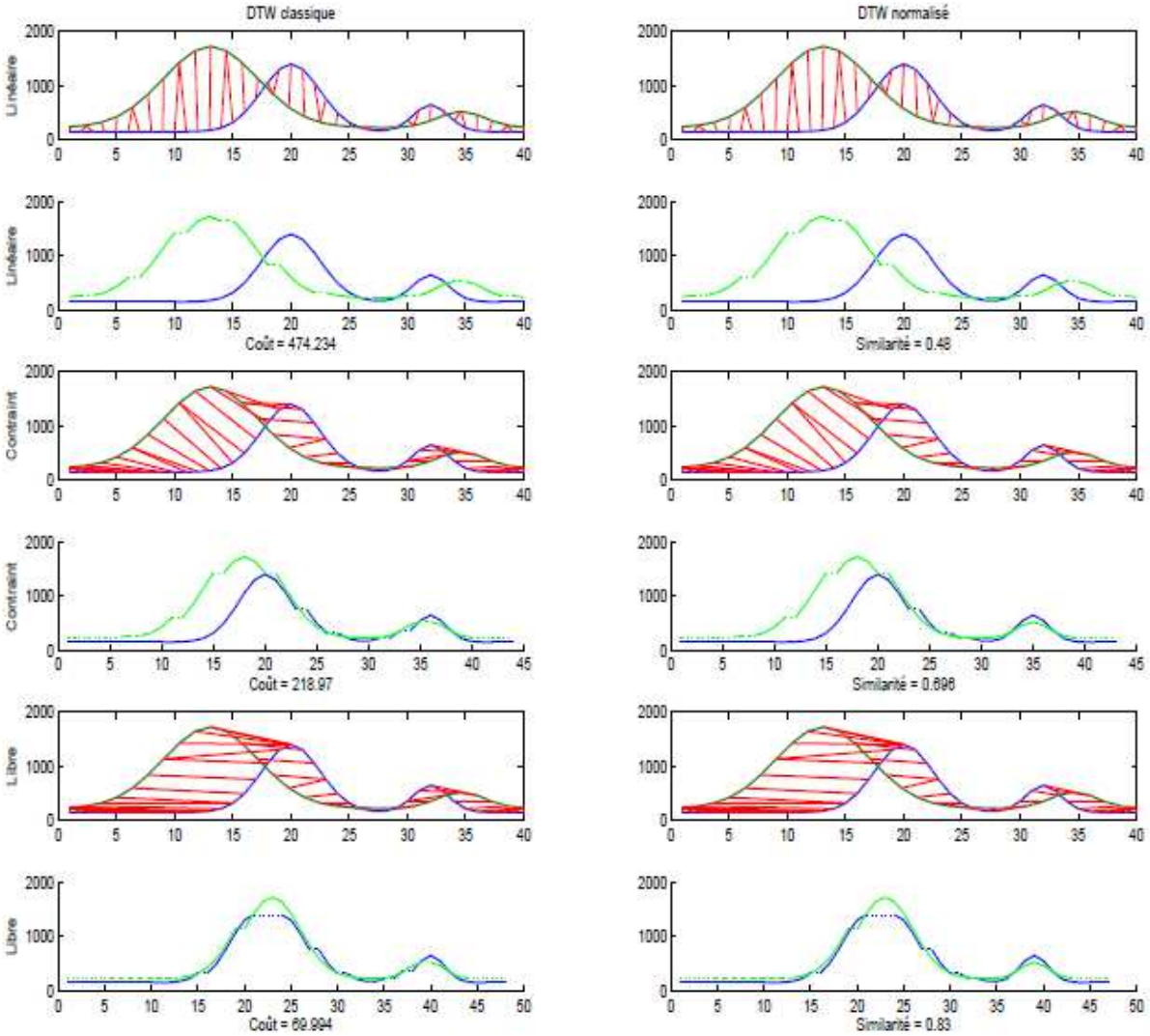


3) Mature colony 500->1000 µm



4) Old colony

Classification DTW-dissimilarity



Classification DTW-dissimilarity

DTW : Dynamic Time Warping

- Conjoint elastic matching of nD signaux
  - Positive Signals (same sign)

$$\bar{X} = \{(\bar{x}_i), i = 1, \ldots, n_x\}$$
$$\bar{Y} = \{(\bar{y}_j), j = 1, \ldots, n_y\}$$

Signals to compare

$$\bar{x}_i = \{(x_{ic}), c = 1, \ldots, n_c\}$$

Series of intensity of n_c curves characterising the signal

Conjoint DTW-distance

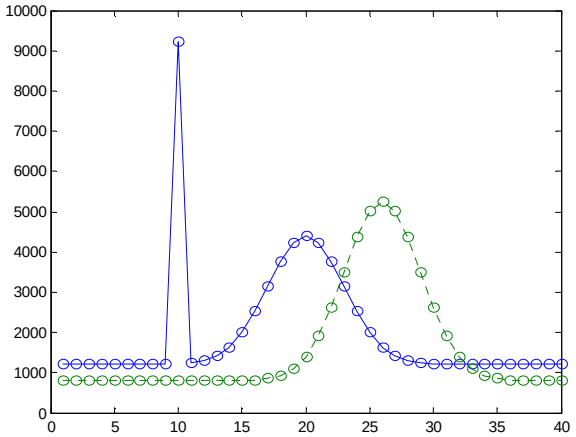
$$d(\bar{x}_i, \bar{y}_j) = \sum_{c=1}^{n_c} |x_{ic} - y_{ic}| = \sum_{c=1}^{n_c} d(x_{ic}, y_{ic}).$$

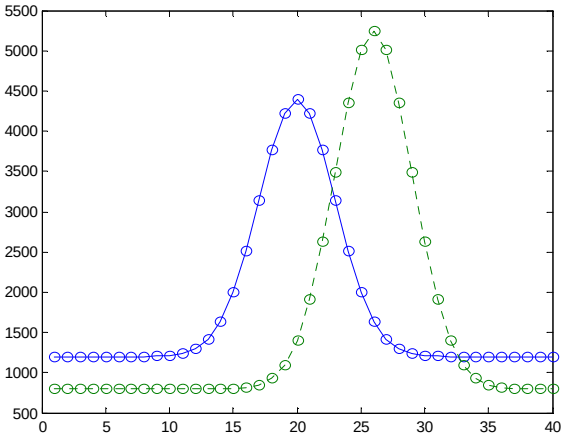
Conjoint DTW-dissimilarity

$$s(\bar{x}_i, \bar{y}_j) = \frac{1}{n_c} \sum_{c=1}^{n_c} s(x_{ic}, y_{ic}).$$

Classification DTW-dissimilarity

DTW-dissimilarity vs DTW-distance : noise sensitivity



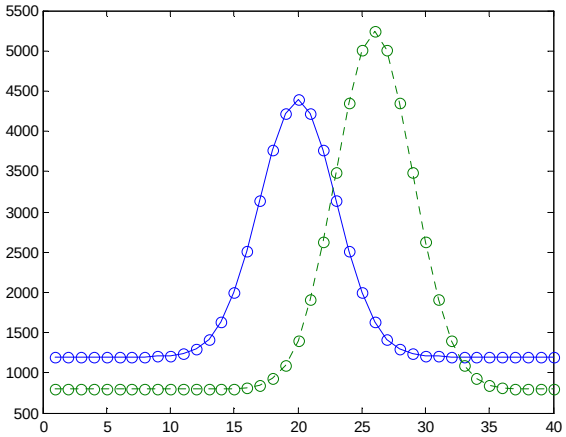


	linear	restricted	free
DTW distance	1368	516	332
DTW dissimilarity	0,45	0,25	0,17

	linear	restricted	free
DTW distance	1168	416	234
DTW dissimilarity	0,43	0,24	0,16

Classification DTW-dissimilarity

DTW-dissimilarity vs DTW-distance: low intensity signals



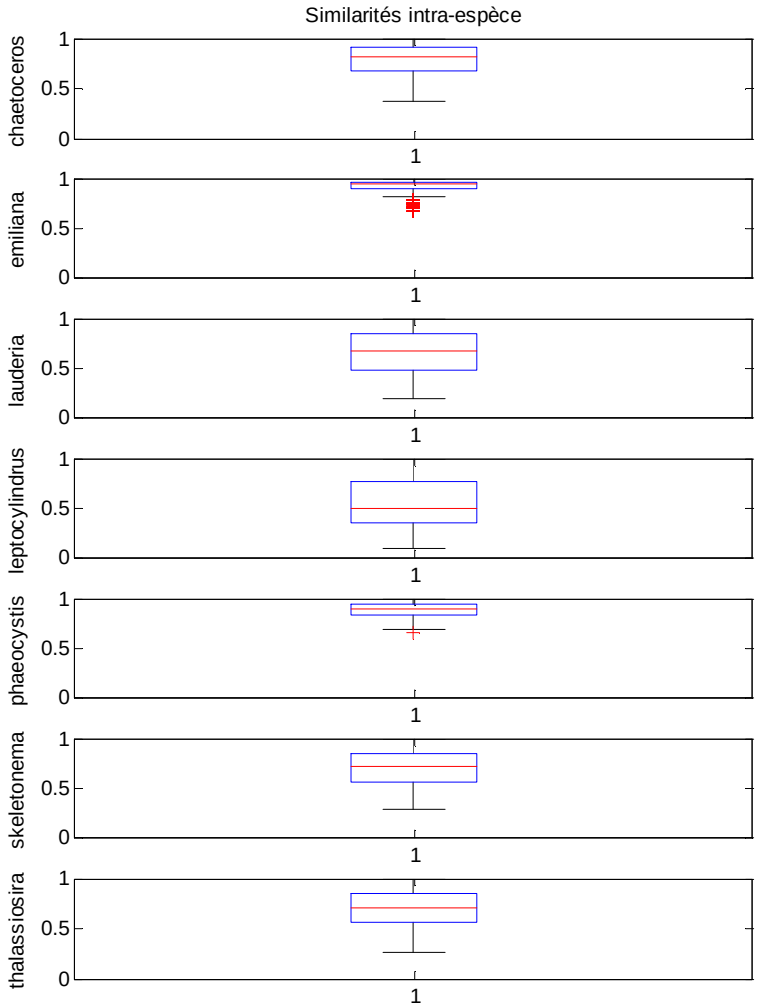
Same figure
Intensity : coefficient x1/8000
(max value from 5500 to 0.8).

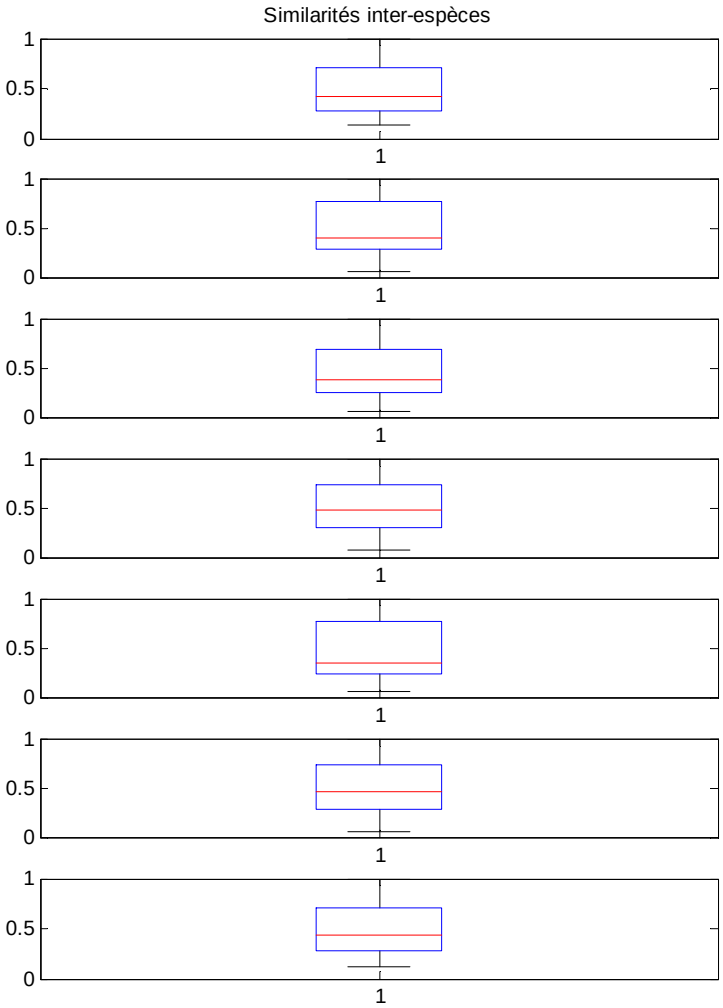
	linear	restricted	free
DTW distance	1168	416	234
DTW Dissimilarity	0,57	0,76	0,84

	linear	restricted	free
DTW distance	0,15	0,05	0,02
DTW dissimilarity	0,56	0,76	0,84

Application: Phytoplancton classification

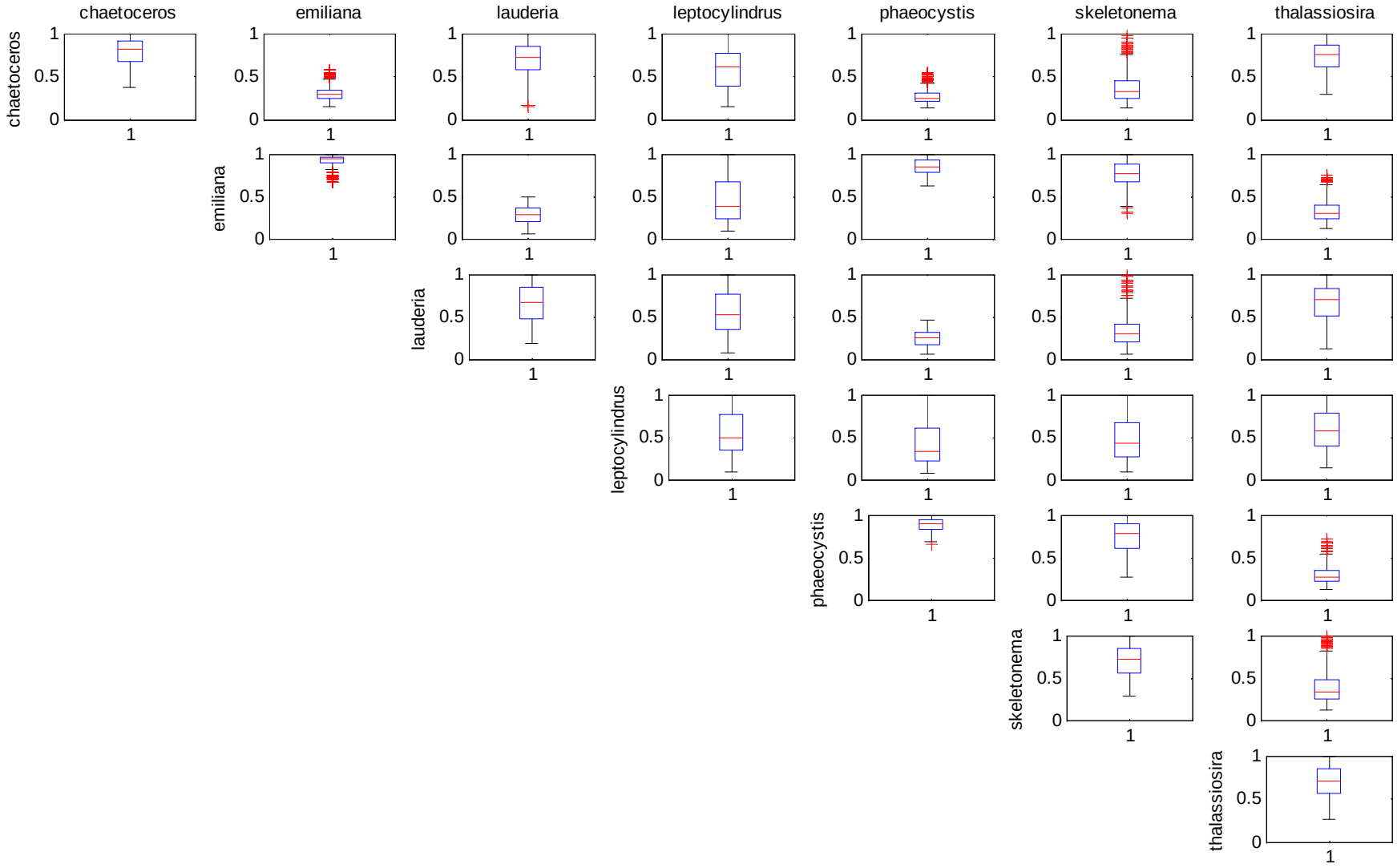
Similarity inter and intra-species





Application: Phytoplancton classification

Similarity inter and intra-species



Application: Phytoplancton classification

Impact of signal intensity

– Cross partition of classifiers based on shape and intensity

Classifieurs		$SimDTW$ *ajusted*		$SimDTW$		
		Rec	No -rec.	Rec	No -rec.	Total
$SimAmp$	Rec.	$85,3$	$8,4$	$93,0$	$0,7$	$93,7$
	No -rec.	$3,6$	$2,7$	$5,2$	$1,1$	$6,3$
	Total	$88,9$	$11,1$	$98,2$	$1,8$	100