Missing data
DTW-based imputation

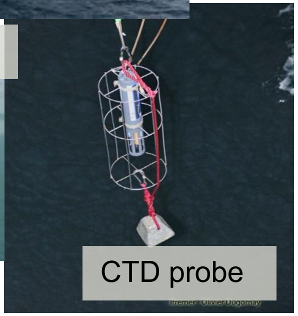
Poisson-caillault Emilie, Phan Thi-Thu-Hong, Lefebvre Alain

ULCO-LISIC, EA 4491, Universite du Littoral Côte d'Opale, FR-62228 Calais

IFREMER, Laboratoire Environnement et Ressources, FR-62321, Boulogne-sur-mer

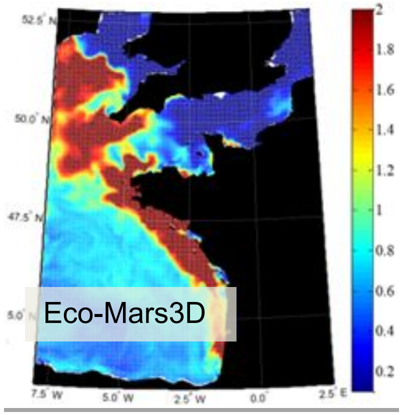
VNUA, Vietnam National University of Agriculture, Hanoï, Vietnam

Observation networks : multi-sources



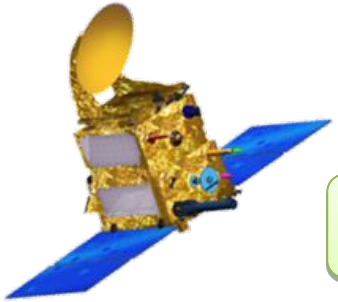
Conventional measurement campaign

Models



Mobile Stations

Satellites



Fixed Stations

First pretreatments to extract OI

- Storage
- Control quality / sensor range
- Temporal/Spatial Alignments
- Na data

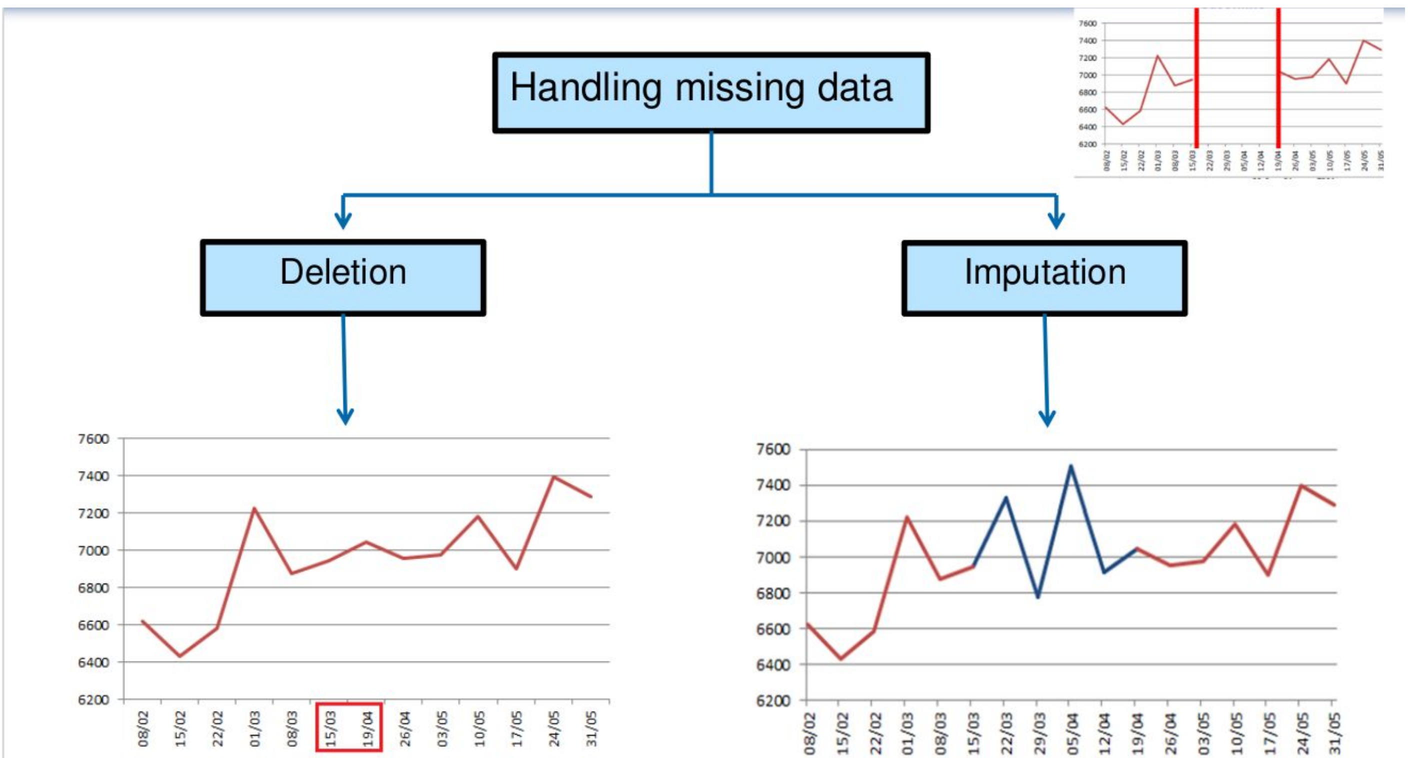
Collecting lots of time series data (big data)

Problems:

- Collected data have missing values due to sensor or communication/transmission problems
- Most of Classification and prediction tools require completed data

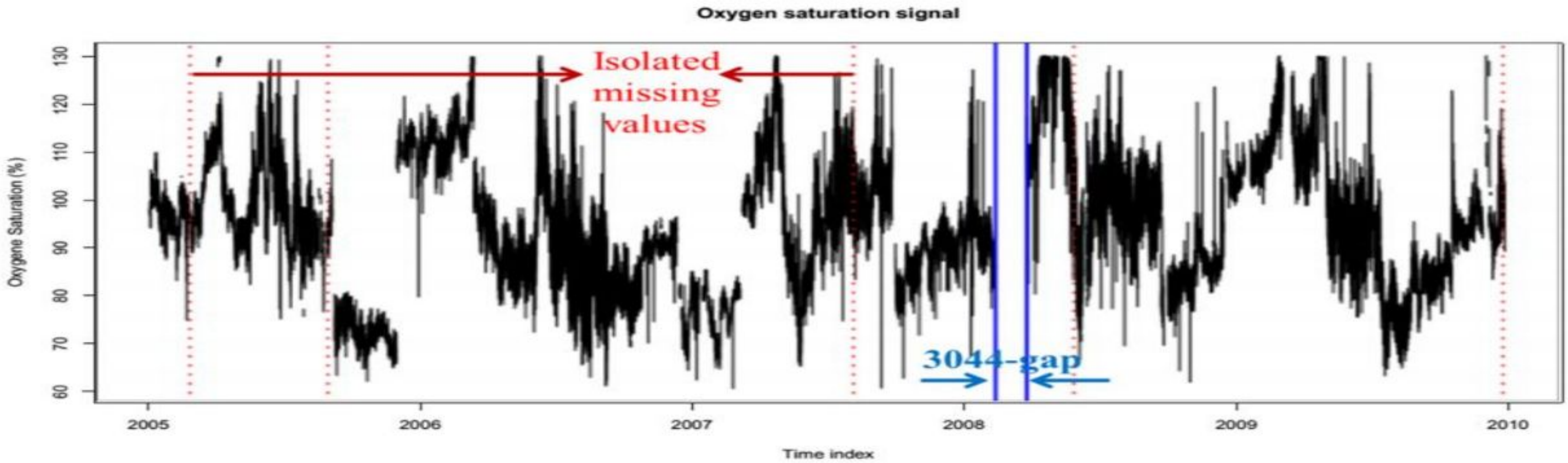
Any available solution?

- Deleting some observations
- Imputation data



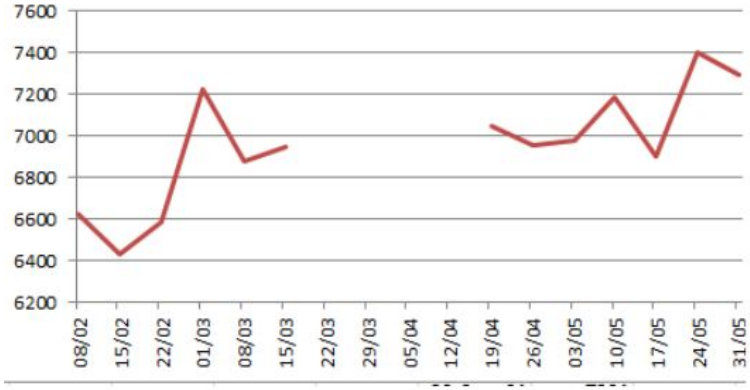
Which methods to deal with missing marine observations ?

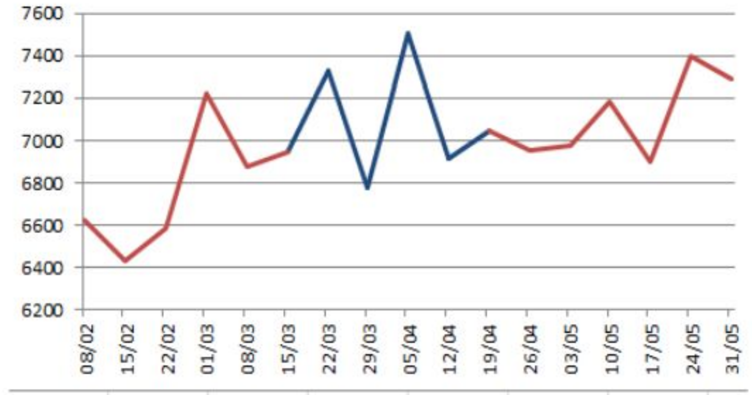
1- Nature of the missing data : MCAR, MAR, NMAR ↔ not so easy to define.
2- Isolated missing values / Gap=missing consecutive values
3- Complex data, low cross-correlation
3- Quick change in the dynamics of the studied process.

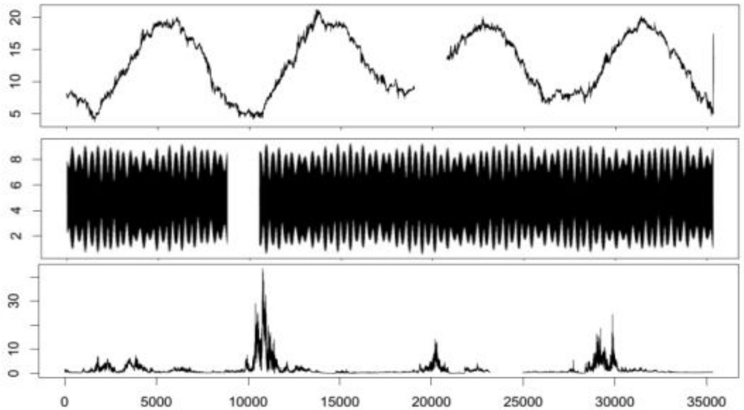


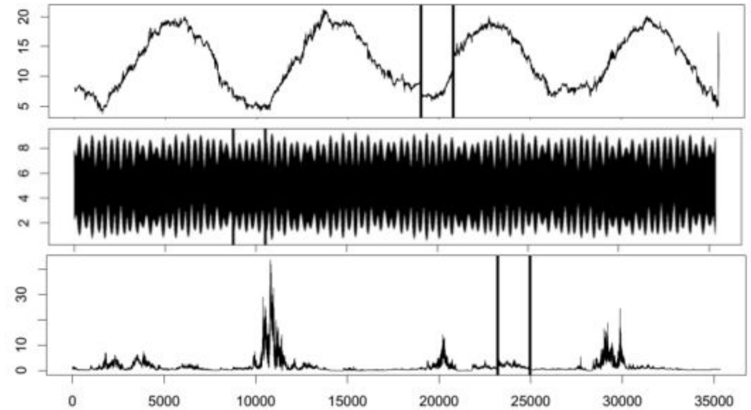
=> Adapted algorithm, **results must be**
- **efficient**: having minimal standard errors
- **reliable/effective**: respecting curve-shape and dynamics process/changes.

Which methods to deal with missing marine observations









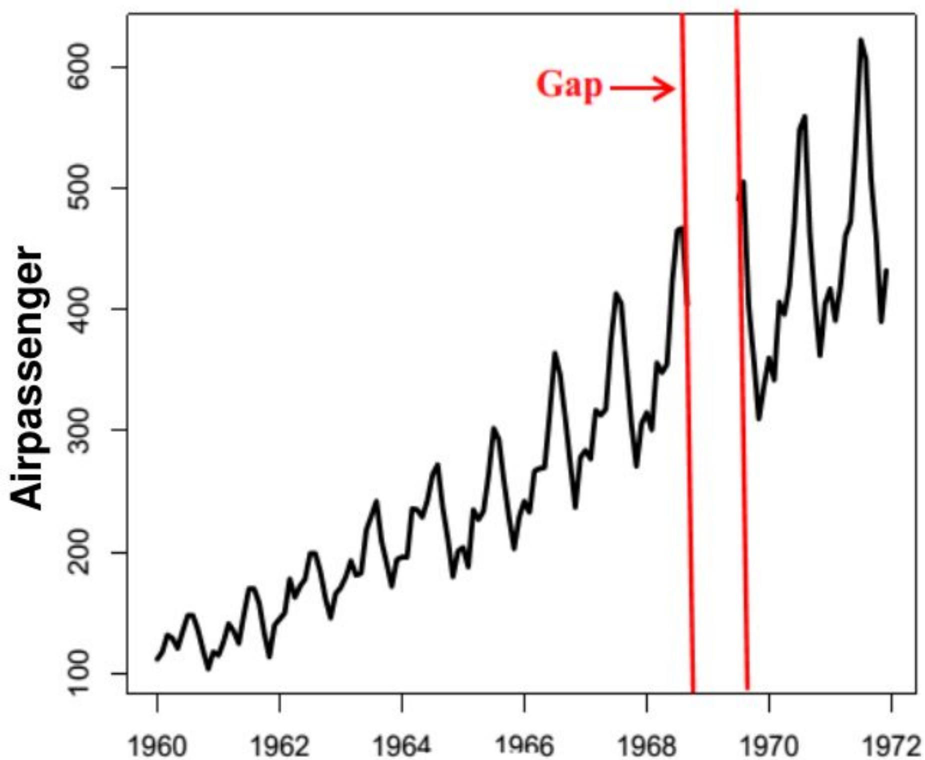
=> Adapted algorithm, **results must be**

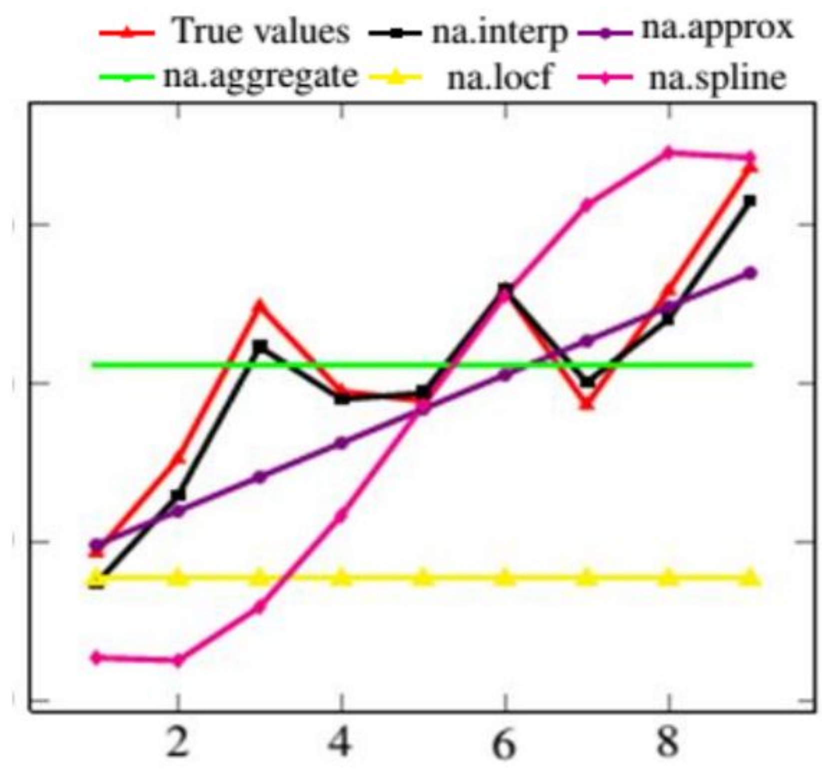
- **efficient**: having minimal standard errors
- **reliable/effective**: respecting curve-shape and dynamics process/changes.

Case of Isolated missing values in **univariate times series**

-> Usual methods : local averages, interpolation, regression.

But for gap ?





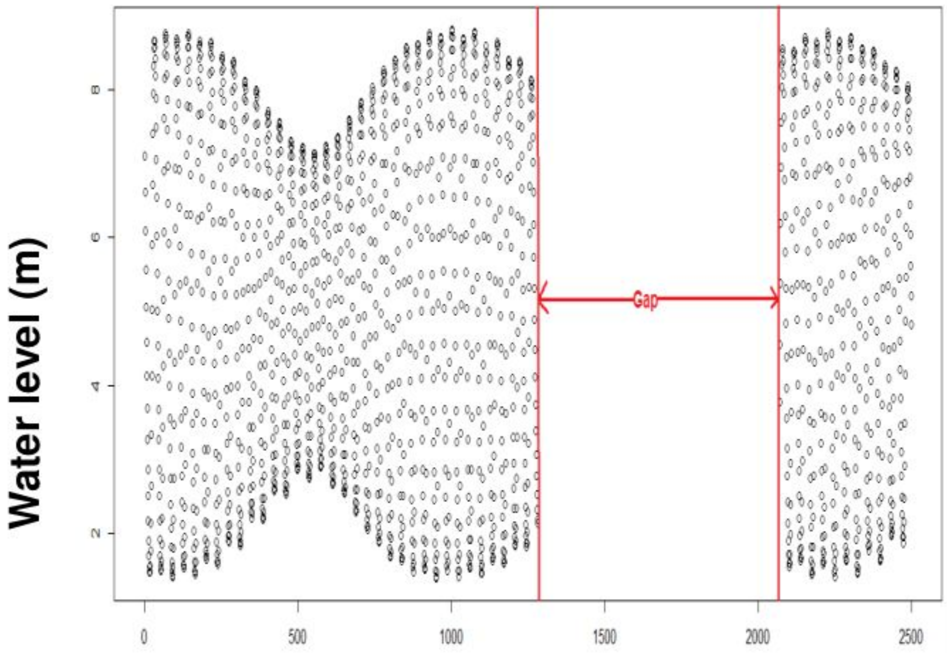
Kaggle dataset. **Univariate** time series.

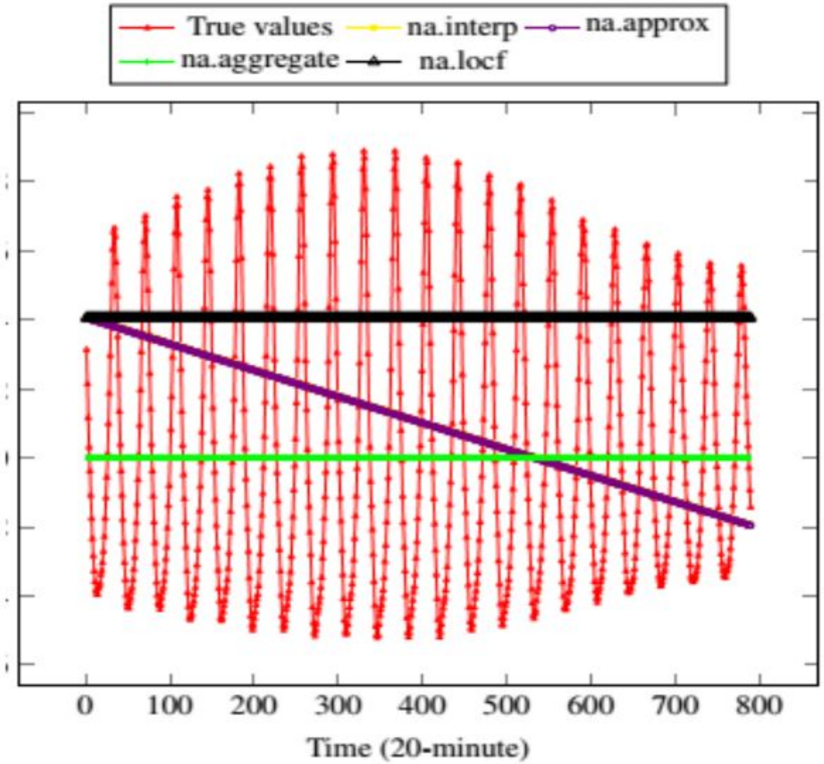
na.interp only respects the **pattern shape**

Solution for Isolated missing values in **univariate times series**

-> Usual methods : local averages, interpolation, regression.

And large gap ?

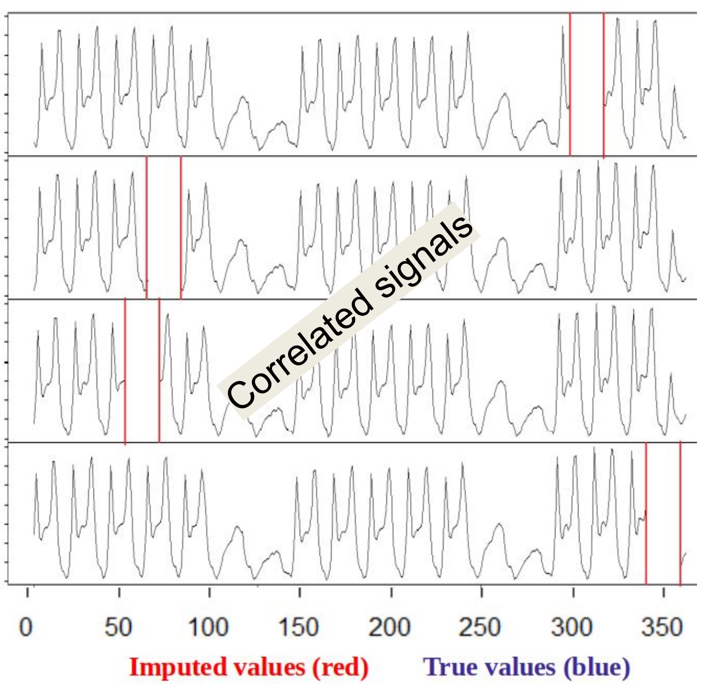


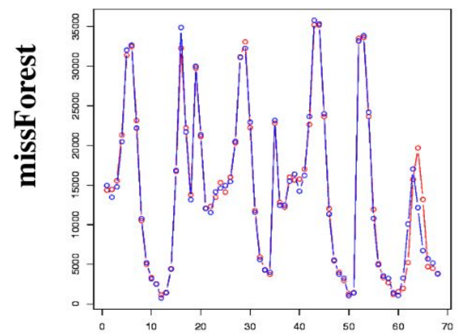


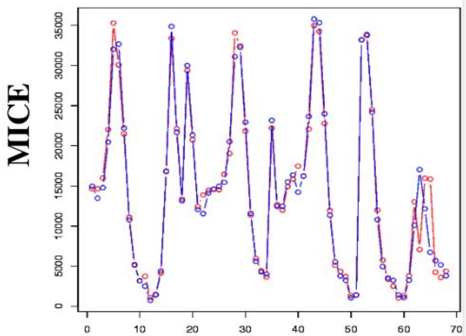
Efficient but no respect for the **pattern shape**.

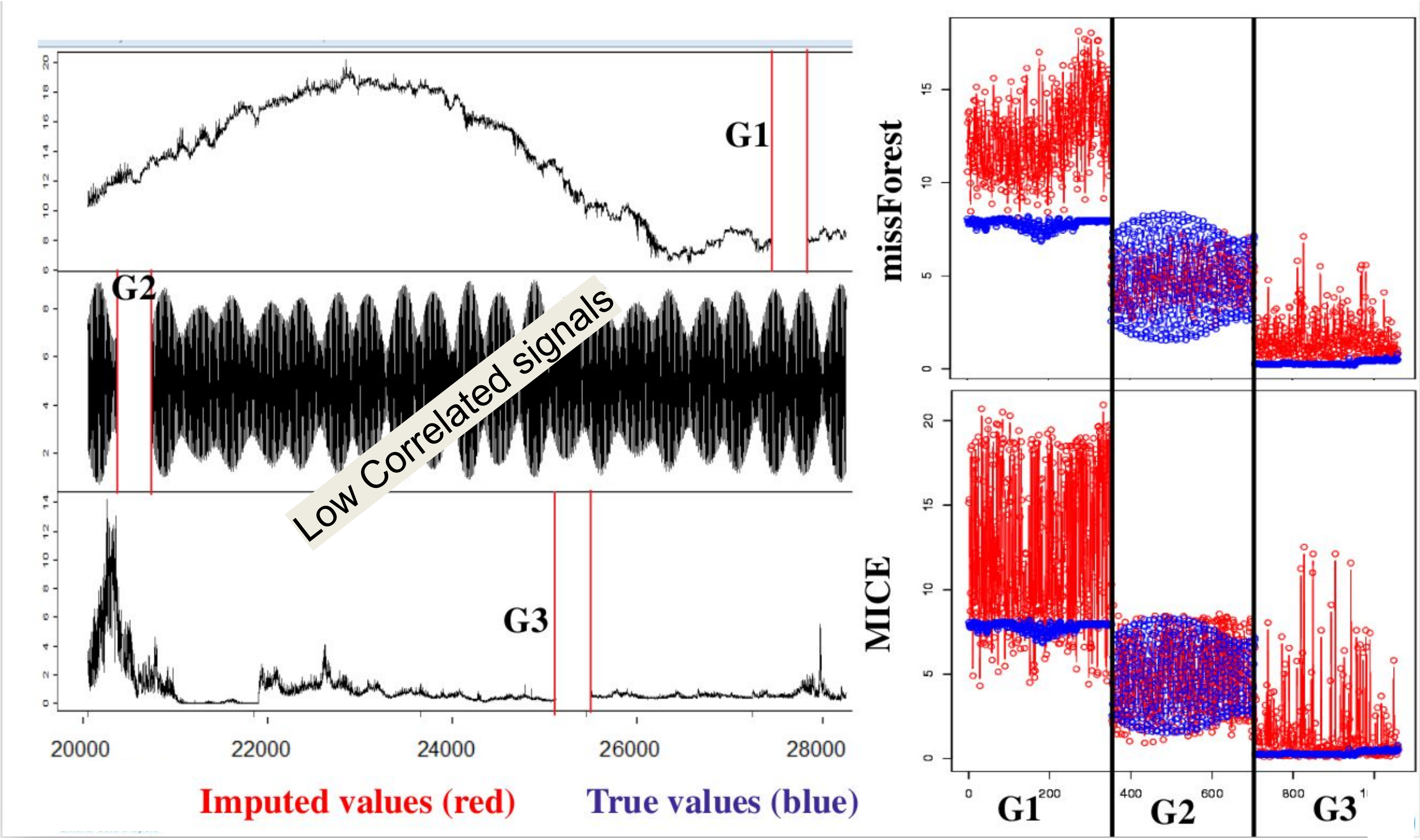
Solution for **multivariate times series**

- MI- **Multiple Imputation**: predicts imputed value by finding an observation (from available values) with the closest predictive mean to that variable containing missing values for each observation.
- **MICE - Multivariate Imputation via Chained Equations**: estimate imputed values based on the conditional distribution (on all of other variables) for each variable containing missing values.
- **na.approx** (zoo package): uses a generic function with interpolated values to estimate each missing data.
- **missForest**: for each variable missForest builds a random forest model on the observed part. Then this model is used to predict missing values in the variable. (look at the code : fill by means !)
- **k-nearest neighbors**



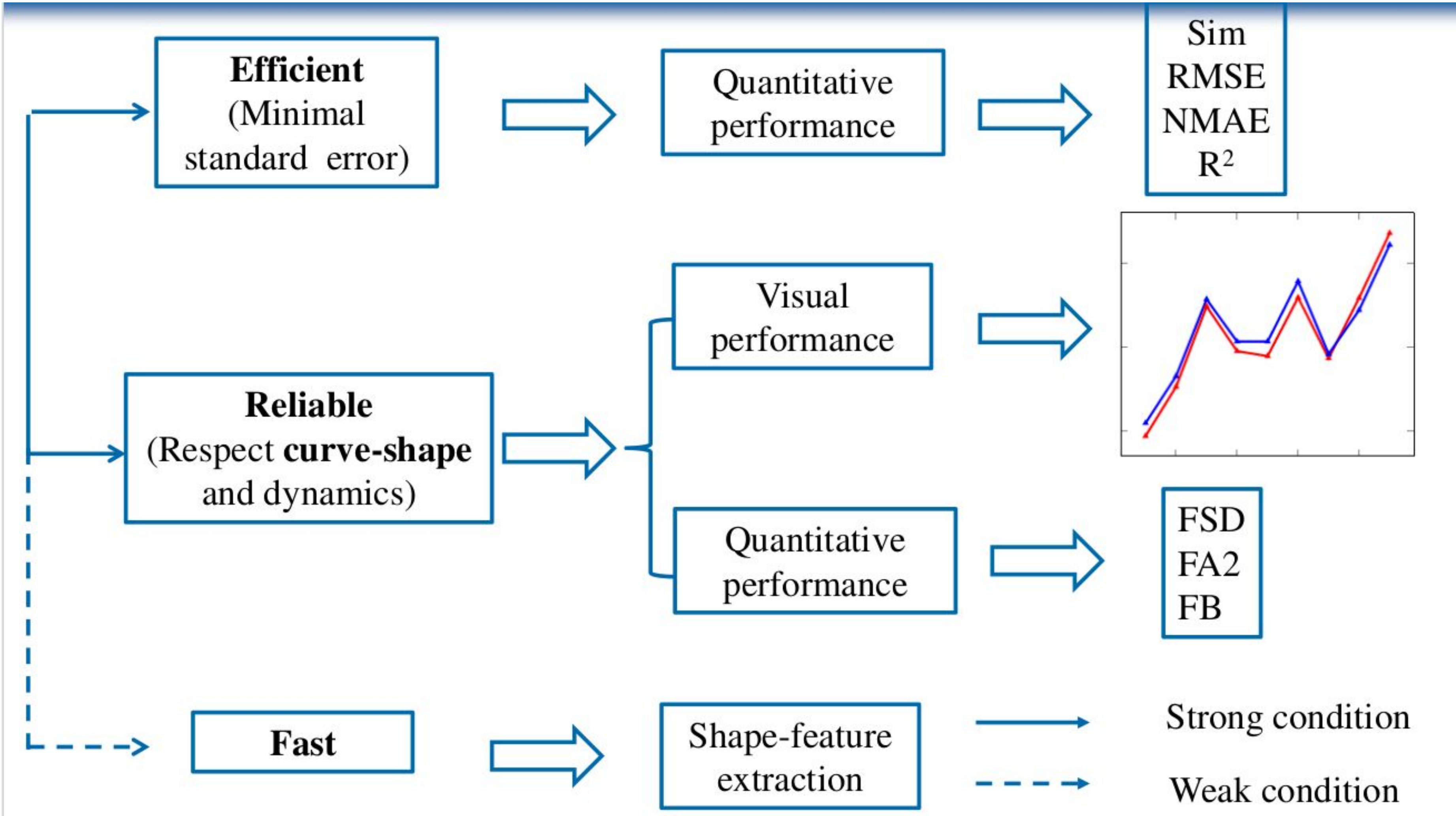






-> No efficiency / no shape respect for large gap

Goal and Trade off.



Idea

Proposed approach

Focus on Marel Carnot Dataset

MAREL Carnot Dataset

Dataset
MAREL Carnot

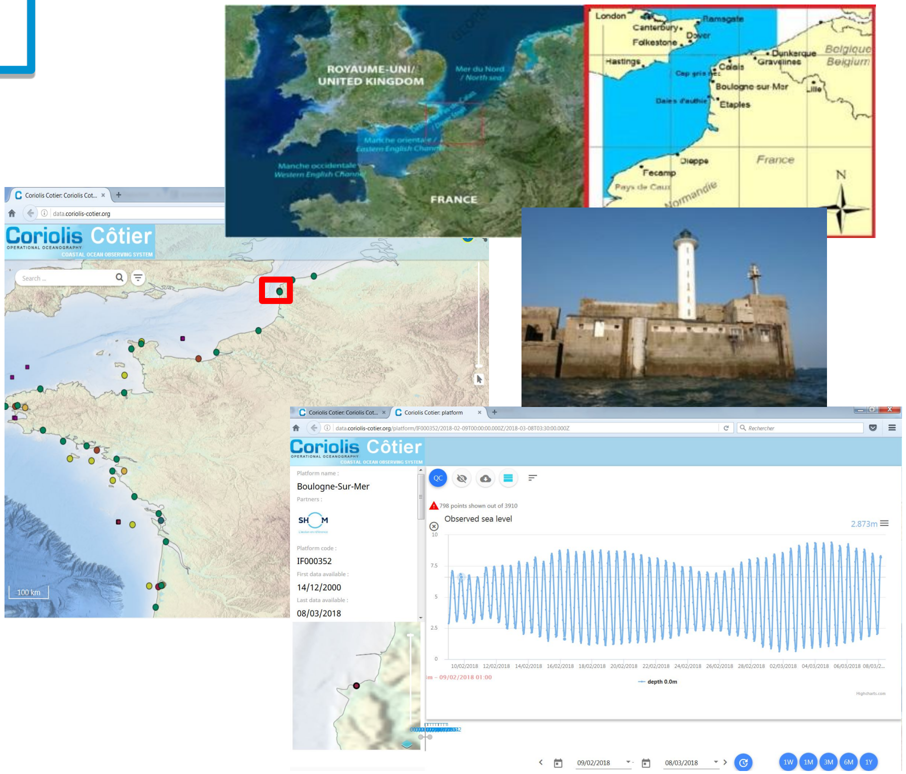
High frequency instrumented station (HF)
20 minutes sampling

Tests based on 4 years (2005-2009)

Raw Data
*Size : 92 968 * 9*
#NA: 320 401 : 38 %

9 Essential Ocean Variables :

- **Temperature**
- **Salinity**
- **Dissolved Oxygen**
- **Nitrate**
- **Phosphate**
- **Silicate**
- **Turbidity**
- **PAR** (Photosynthetically Active Radiation)
- **Fluorescence**



http://data.coriolis-cotier.org/

Completion Protocol

Completion Protocol:

T-gap: size T of consecutive missing values

completion size criterion: smallHole and acceptedHole

Completion steps:

1- Isolated points (T=1)-gap: completion by local average (t-1 and/or t+1)

2- small T-gap: completion by weighting local average of a mobile centered windows [t-T; t +T]
T<**smallHole**

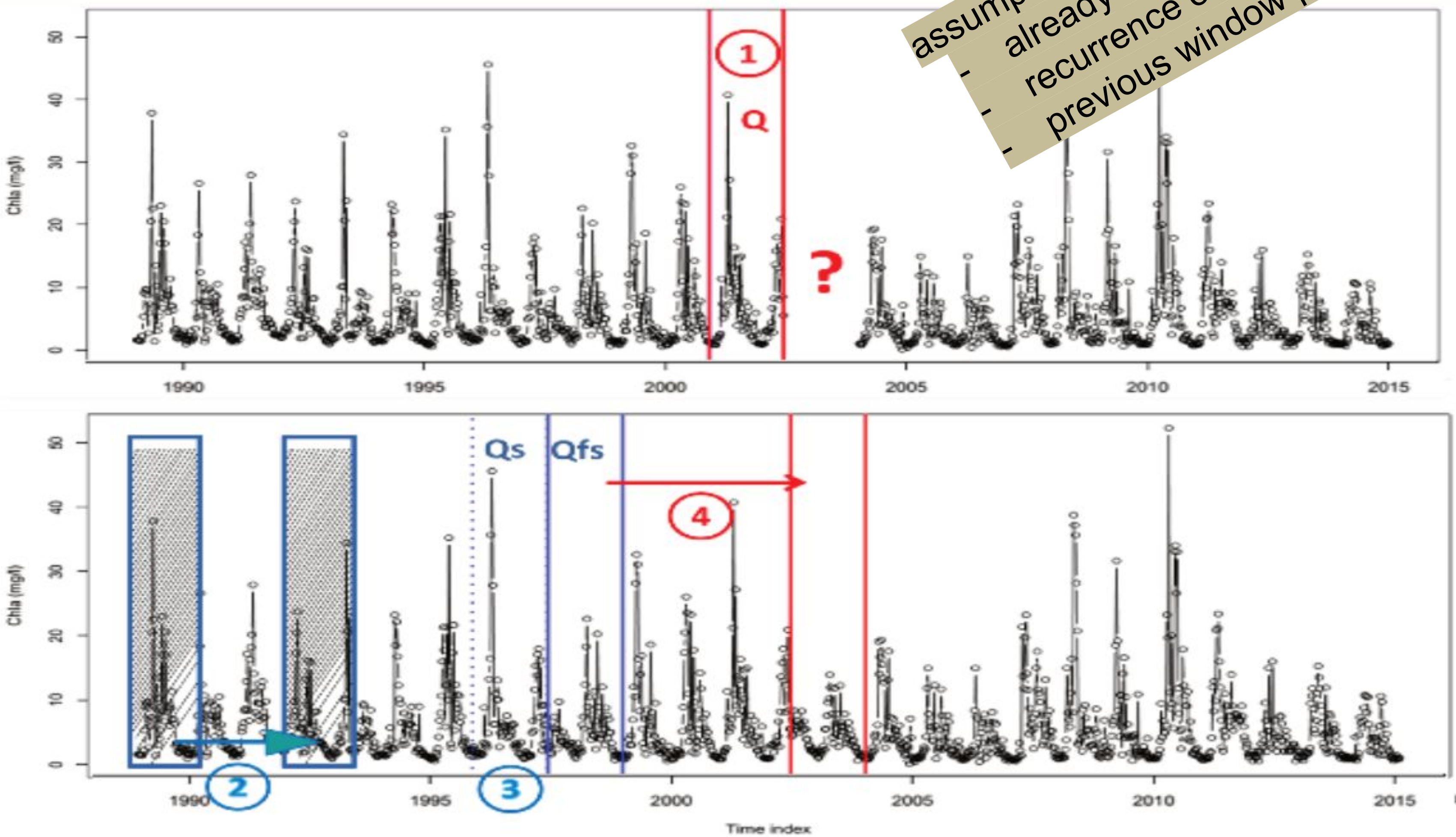
3- large T-gap: filling by existing recurrent pattern with DTW-matching
 (**smallHole** < T ≤ **acceptedHole**)

smallHole , acceptedHole depends on
- dynamics change
- subject of interest.

Proposed approaches for large T>>1-gap:

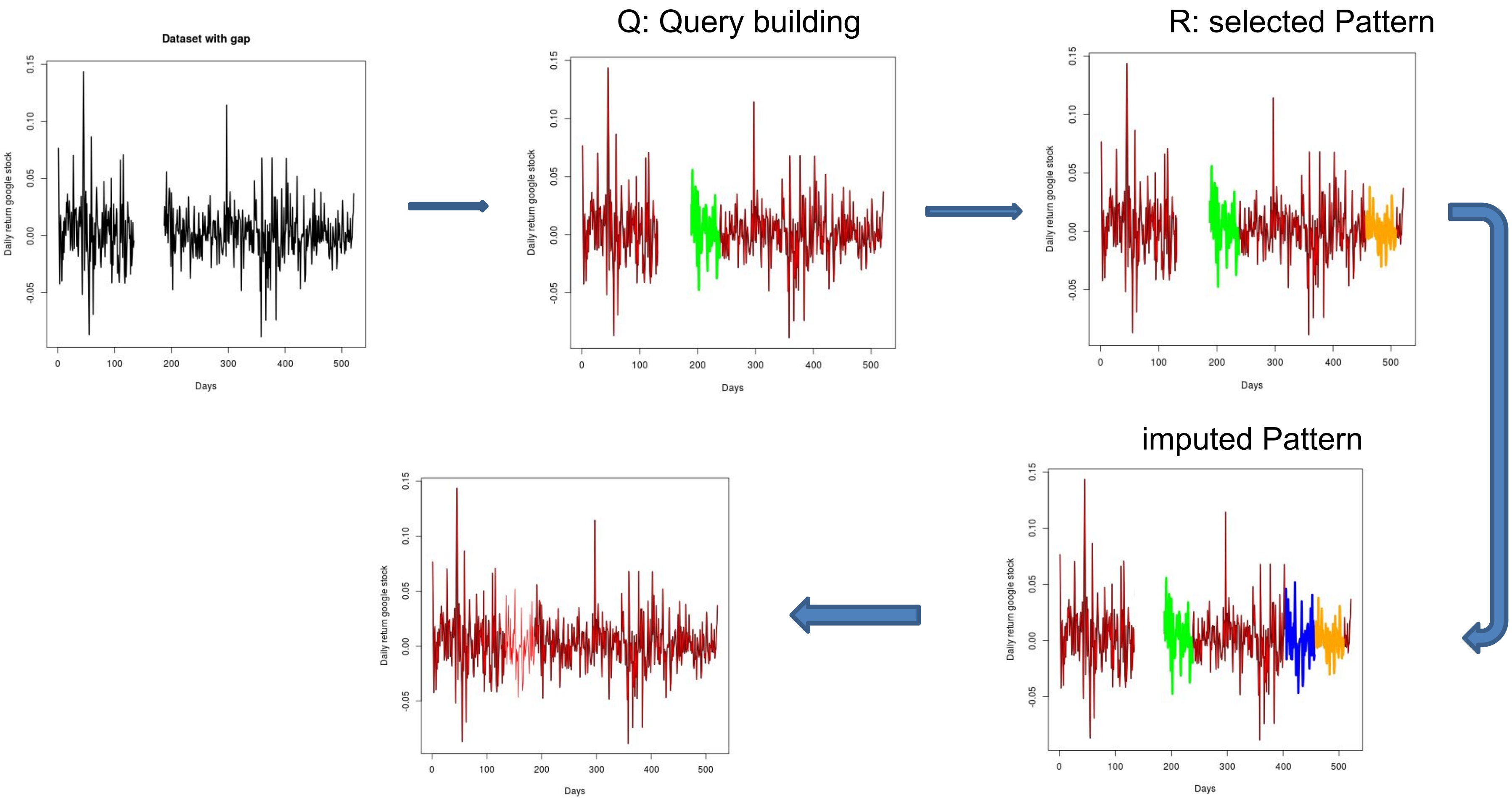
- DTWBI : univariate iseries - imputation based on DTW criterion
- DTWUMI : extension of DTWBI for multivariate series
- FSMUMI : Fuzzy Logic Theory
- eDTWBI : add trend conservation contraints + future window

assumption :
- already existing information
- recurrence of phenomena
- previous window pressure

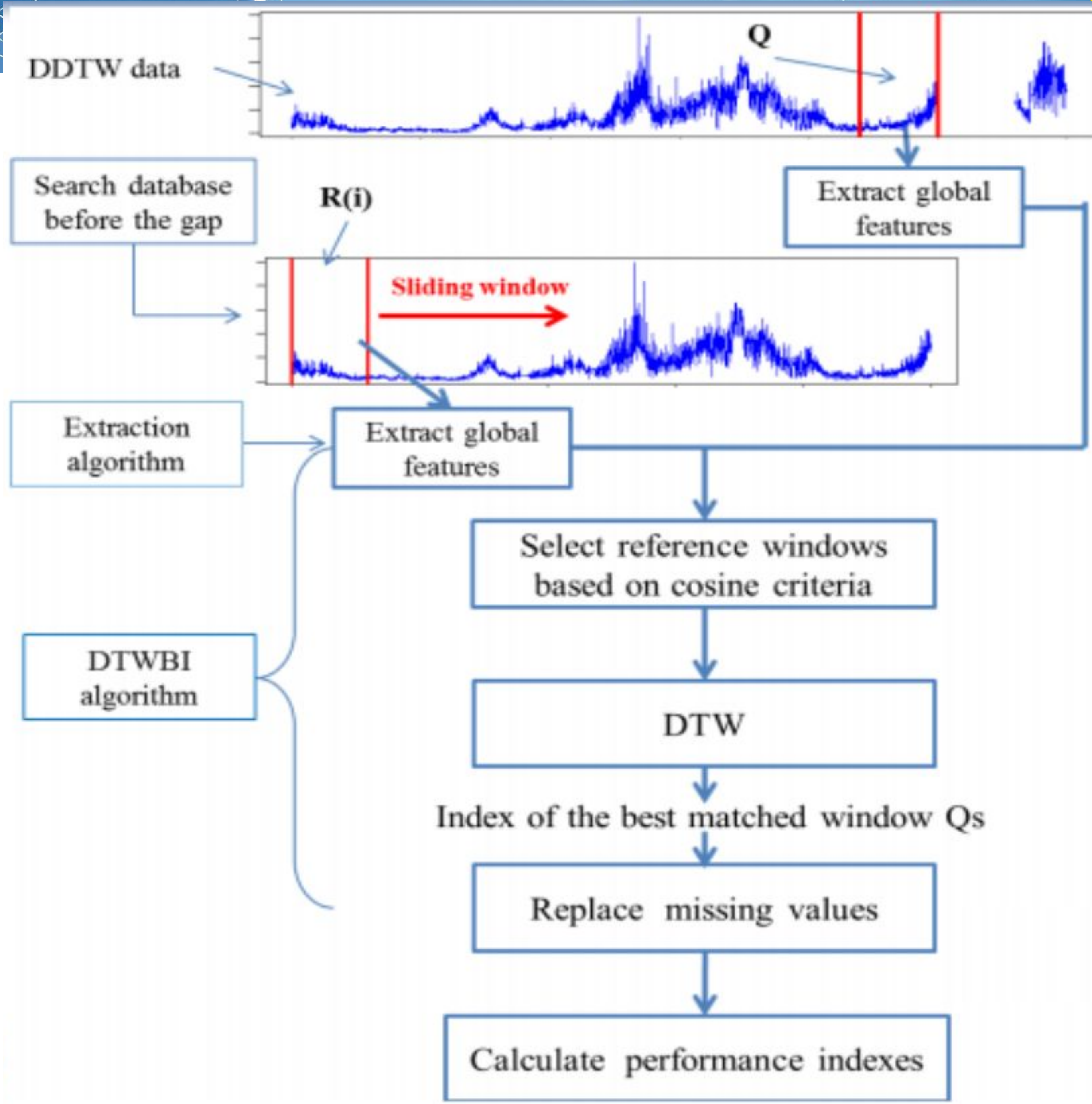


Illustration/Vocabulary:

Completion of one large gap by “DTW-similar” recurring pattern



Univariate signal

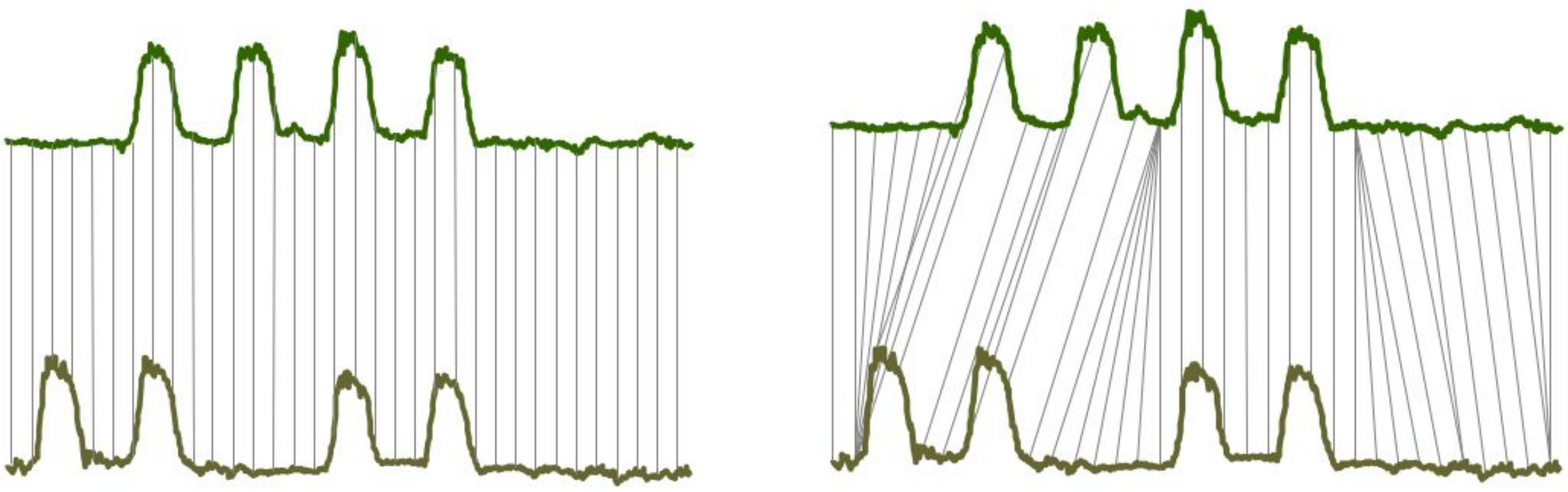


DDTW data
Q
Search database before the gap
R(i)
Extract global features
Sliding window
Extraction algorithm
Extract global features
Select reference windows based on cosine criteria
DTWBI algorithm
DTW
Index of the best matched window Qs
Replace missing values
Calculate performance indexes

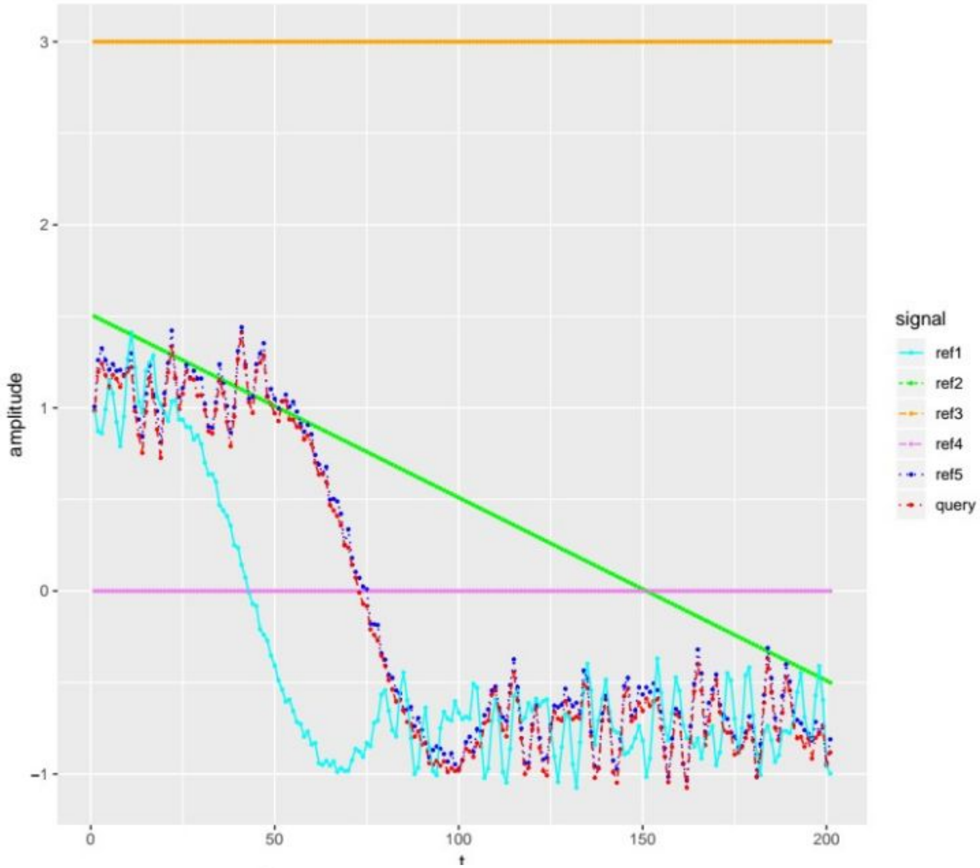
Similarity problem

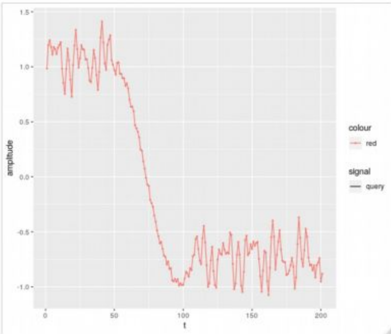
- Given: two sequences $X= \{x_1,x_2,...,x_n\}$ and $Y= \{y_1,y_2,...,y_m\}$
- How to define and compute Sim(X, Y) according to acceptable time variability?

Euclidean distance vs. **DTW**



Signal comparison - **Euclidean distance**
Metrics ?



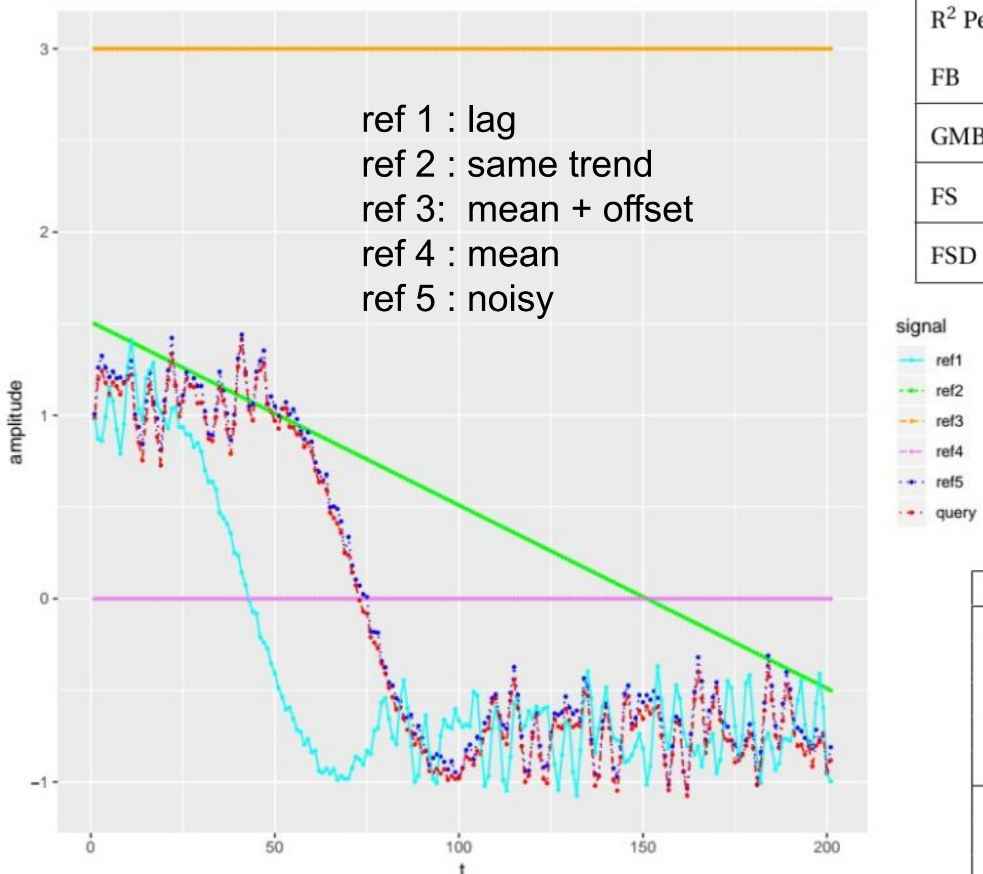


ref 1 : lag
ref 2 : same trend
ref 3: mean + offset
ref 4 : mean
ref 5 : noisy

	query	ref1	ref2	ref3	ref4	ref5
distMax	0.00	1.64	1.53	4.07	1.41	0.10
ME	0.00	0.27	-0.63	-3.13	-0.13	-0.05

Signal comparison - Euclidean distance Metrics ?

ref 1 : lag
ref 2 : same trend
ref 3: mean + offset
ref 4 : mean
ref 5 : noisy



GMV=VG	$\exp(\overline{(\ln(q)-\ln(r))^2})$	min.	$\in \mathbb{R}^+$	$+\ 0.75 \leq \text{VG} \leq 1.25$
FA2	$\frac{1}{T}\text{card}(0.5 \leq q_i/r_i \leq 2)$	max.	$\in [0,1]$	$+\ \text{FA2} > 0.8$
R^2 Pearson	$(\frac{cov(q,r)}{\sigma_q \times \sigma_r})^2$	max.	$\in [0,1]$	$++\ R^2 \geq 0.9$ p-value
FB	$2 \times \frac{\overline{q}-\overline{r}}{\overline{q}+\overline{r}}$	min.	$\in [-2,2]$	$++\ \lvert\text{FB}\rvert \leq 0.3$
GMB=BG	$\exp(\overline{\ln(q)}-\overline{\ln(r)})$	min.	$\in \mathbb{R}^+$	$+\ 0.75 \leq \text{BG} \leq 1.25$
FS	$2 \times \frac{\lvert(\sigma_q)^2-(\sigma_r)^2\rvert}{(\sigma_q)^2+(\sigma_r)^2}$	min.	$\in [0,2]$	$++\ \text{FS} \leq 0.5$
FSD	$2 \times \lvert\frac{\sigma_q-\sigma_r}{\sigma_q+\sigma_r}\rvert$	min.	$\in [0,2]$	$++\ \text{FSD} \leq 0.5$

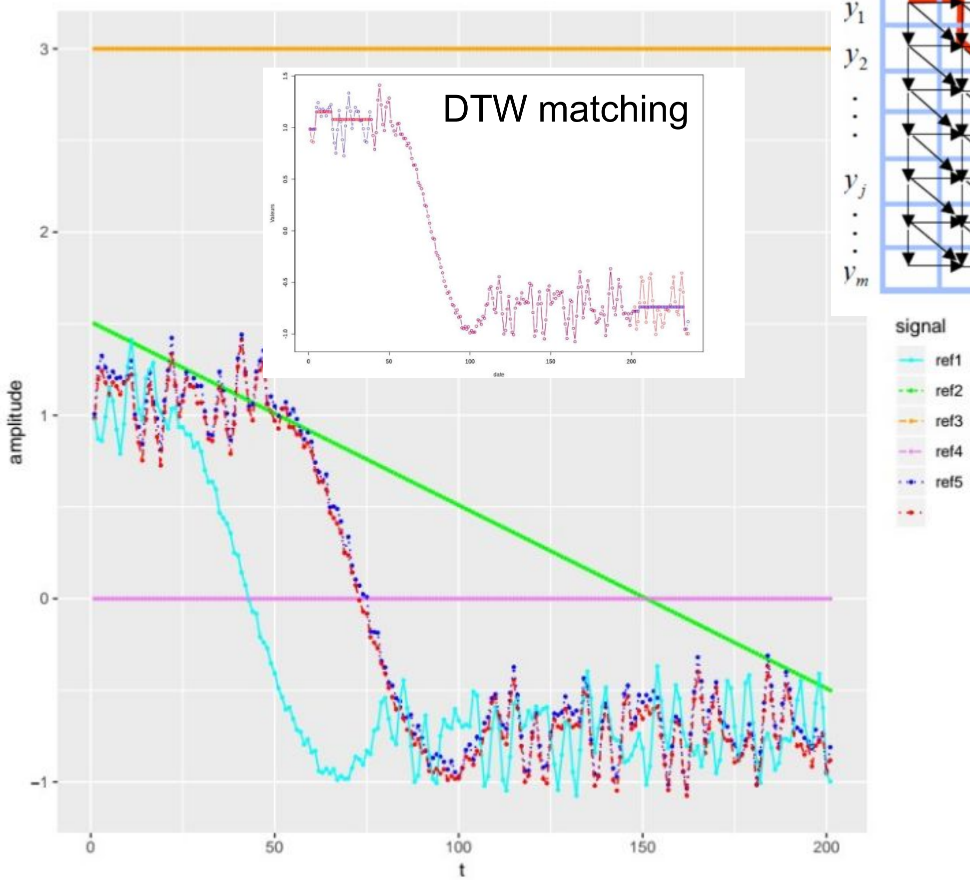
bounded scores

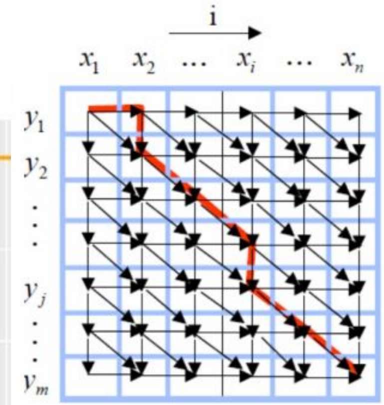
Bold results = criterion is satisfied

	query	ref1	ref2	ref3	ref4	ref5
distMax	0.00	1.64	1.53	4.07	1.41	0.10
ME	0.00	0.27	-0.63	-3.13	-0.13	-0.05
VG	1.00	14.71	12.71	137.54	-	**1.02**
FA2	1.00	0.78	0.41	0.00	0.00	**0.99**
R2.cor	1.00	0.55	0.72	0.00	0.00	**1.00**
Bias	0.00	0.27	-0.63	-3.13	-0.13	-0.05
FB	0.00	-1.04	-3.33	-2.17	2.00	0.55
BG	1.00	1.33	**0.78**	0.18	-	**0.98**
FS	0.00	**0.41**	0.69	2.00	2.00	**-0.00**
FSD	0.00	**0.21**	**0.36**	2.00	2.00	**0.00**

-> **no signal satisfies all criteria**

Signal comparison - Elastic matching
Metrics ?





$$X = \{(x_i), i = 1, \ldots, n\}$$
$$Y = \{(y_j), j = 1, \ldots, m\}$$
$$P = \{(i_k, j_k), k = 1, \ldots, n_k\}$$

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

Elastic matching
Deformation tolerance
DTW / Derivative DTW / Features DTW

	VG	FA2	R2.cor	FB	BG	FS	FSD
ref1-DTW	1.04	1.00	0.99	0.04	1.01	-0.01	0.00
ref1-DDTW	1.08	1.00	0.98	-0.05	0.97	0.14	0.07
ref1-AFBTDTW	1.06	1.00	0.99	-0.03	1.02	0.05	0.03
ref2-DTW	g.	0.96	0.96	-0.47	89.37	0.23	0.11
ref2-DDTW	83.98	0.41	0.33	-8.43	1.57	-0.50	0.25
ref2-AFBTDTW	3.63	0.42	0.92	-2.84	1.35	0.05	0.03
ref3-DTW	12.08	0.00	0.00	-1.30	0.39	2.00	2.00
ref3-DDTW	g.	0.00	0.00	-2.87	0.09	2.00	2.00
ref3-AFBTDTW	g.	0.00	0.00	-2.70	0.14	2.00	2.00
ref4-DTW	g.	0.00	0.00	2.00	g.	2.00	2.00
ref4-DDTW	g.	0.00	0.00	2.00	g.	2.00	2.00
ref4-AFBTDTW	g.	0.00	0.00	2.00	g.	2.00	2.00
ref5-DTW	1.04	0.99	1.00	0.19	1.07	-0.00	0.00
ref5-DDTW	1.05	0.94	0.99	0.68	0.99	-0.05	0.03
ref5-AFBTDTW	1.02	0.99	1.00	0.52	0.98	-0.01	0.00

Experiments - Univariate case

N0	dataset name	N0 of instants	Trend (Y/N)	Seasonal (Y/N)	Frequency
1	Air passenger	144	Y	Y	Monthly
2	Beersales	192	Y	Y	Monthly
3	Google	521	N	N	Daily
4	SP	168	Y	Y	Quarterly
5	CO2 concentrations	160	Y	Y	Monthly
6	Mackey-Glass chaotic	1201	N	N	
7	Phu Lien temperature	648	N	Y	Monthly
8	Water level	131472	N	Y	20 minutes

Imputation performance indicator

Accuracy indexes

- Similarity: the similar percentage between the imputed value (Y) and the respective true values (X)
- NMAE: The Normalized Mean Absolute Error between Y and X
- RMSE: The Root Mean Square Error average squared difference Y and X

Shape indexes

- FSD: Fraction of Standard Deviation between Y and X

Maximum cross-correlation between query and selected windows

-> recurrent patterns

Gap size	dataset							
	#1	#2	#3	#4	#5	#6	#7	#8
6%	0.88	0.92	0.58	0.78	0.99	1	0.91	1
7.50%	0.91	0.91	0.55	0.74	0.99	0.99	0.91	1
10%	0.94	0.87	0.5	0.67	0.98	0.99	0.91	1
12.50%	0.95	0.89	0.44	0.65	0.98	0.99	0.9	1
15%	0.95	0.85	0.4	0.65	0.98	0.99	0.9	1

#1-Airpassenger, #2-Beersales, #3-Google, #4-SP, #5-Co2 concentrations
#6-Mackey-Glass chaotic, #7-Phu Lien temperature, #8-water level

Experiments - Univariate case

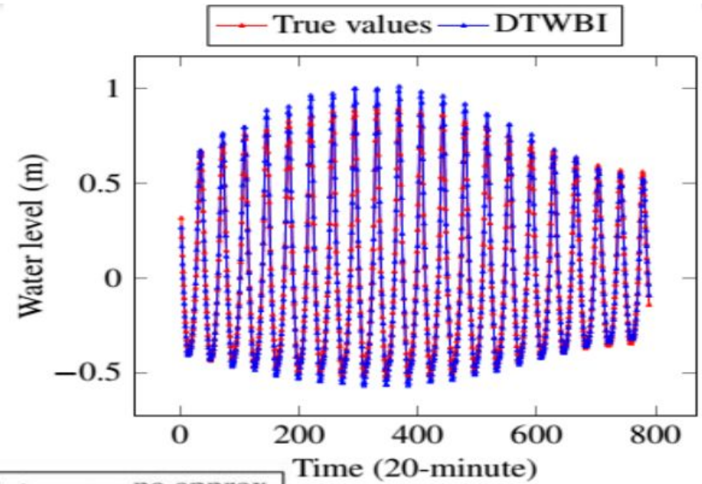
Phu Temperature
15% ~ 97 jours

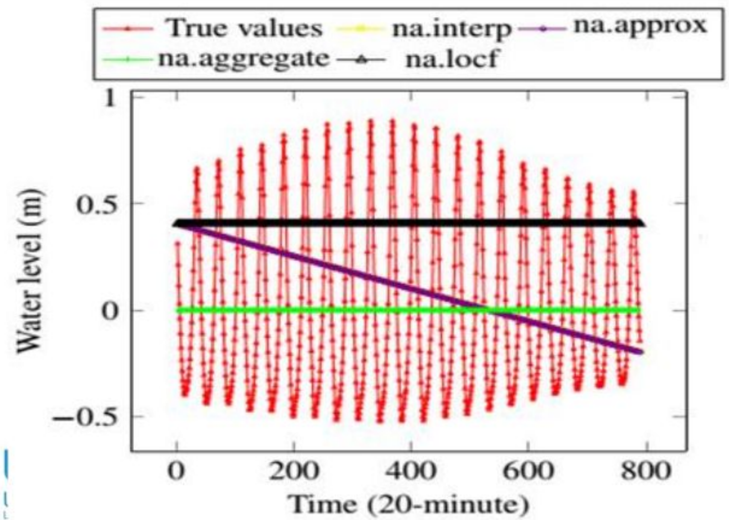
Marel dataset
Gap size/10
1.5% ~ 28 days

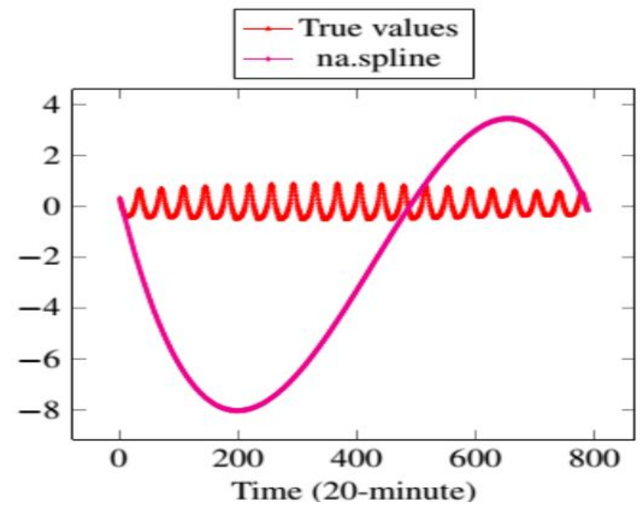
Gap size	Method	CO2 concentrations				Mackey-Glass Chaotic				Phu Lien temperature				Water level			
		Sim	NMAE	RMSE	FSD	Sim	NMAE	RMSE	FSD	Sim	NMAE	RMSE	FSD	Sim	NMAE	RMSE	FSD
6%	DTWBI	**0.93**	**0.001**	**0.3**	**0.04**	**0.95**	**0.005**	**0.01**	**0.03**	**0.88**	**0.06**	**1.7**	**0.08**	**0.95**	**0.009**	**0.1**	**0.05**
	na.interp	0.75	0.055	1.6	1.5	0.79	0.031	0.04	0.81	0.8	0.142	3.1	0.63	0.81	0.042	0.5	1.05
	na.locf	0.73	0.059	1.7	2	0.77	0.036	0.05	2	0.77	0.173	3.8	2	0.8	0.043	0.4	2
	na.approx	0.75	0.055	1.6	1.5	0.79	0.031	0.04	0.81	0.8	0.142	3.1	0.63	0.81	0.042	0.5	1.05
	na.aggregate	0.45	0.185	4.7	2	0.82	0.025	0.03	2	0.83	0.114	2.4	2	0.83	0.035	0.4	2
	na.spline	0.75	0.057	1.6	0.75	0.65	0.072	0.09	0.38	0.61	0.413	8.5	0.52	0.3	0.654	6.6	1.61
15%	DTWBI	**0.94**	**0.001**	**0.3**	**0.04**	**0.92**	**0.01**	**0.02**	**0.01**	**0.882**	**0.066**	**1.8**	**0.05**	**0.96**	**0.007**	**0.1**	**0.04**
	na.interp	0.76	0.053	1.6	1.46	0.81	0.03	0.04	0.99	0.81	0.145	3.2	1	0.81	0.044	0.5	1.6
	na.locf	0.77	0.052	1.6	2	0.79	0.037	0.05	2	0.79	0.175	3.8	2	0.81	0.043	0.5	2
	na.approx	0.76	0.053	1.6	1.46	0.81	0.03	0.04	0.99	0.81	0.145	3.2	1	0.81	0.044	0.5	1.6
	na.aggregate	0.43	0.202	5.1	2	0.84	0.025	0.03	2	0.84	0.117	2.5	2	0.83	0.036	0.4	2
	na.spline	0.69	0.085	2.5	0.58	0.57	0.129	0.16	0.73	0.44	1.268	26.3	1.27	0.21	1.185	11.8	1.83

Gap size	Method	Airpassenger				Beersales				Google				SP			
		1-Sim	NMAE	RMSE	FSD	1-Sim	NMAE	RMSE	FSD	1- Sim	NMAE	RMSE	FSD	1- Sim	NMAE	RMSE	FSD
15%	DTWBI	**0.1**	**0.02**	**12.8**	**0.36**	0.16	**0.054**	1	**0.1**	0.15	0.13	0.031	**0.29**	0.19	0.029	40.7	0.59
	na.interp	0.14	0.025	15.6	0.35	**0.11**	0.069	**0.7**	0.17	0.14	0.11	0.031	0.99	0.21	0.033	43.6	**0.49**
	na.locf	0.21	0.047	28.2	2	0.18	0.126	1.2	2	0.16	0.13	0.034	2	0.19	0.028	36.3	2
	na.prox	0.2	0.043	26.5	1.17	0.17	0.117	1.1	1.42	0.14	0.11	0.031	0.99	0.19	0.032	41	1
	na.aggregate	0.17	0.035	22.1	2	0.16	0.11	1.1	2	**0.11**	**0.08**	**0.023**	2	**0.18**	**0.025**	**32**	2
	na.spline	0.45	0.175	106.1	0.95	0.51	0.731	6.3	0.88	0.66	12.34	2.928	1.6	0.39	0.136	162.5	0.68

Experiments - Univariate case / comparison with random selection



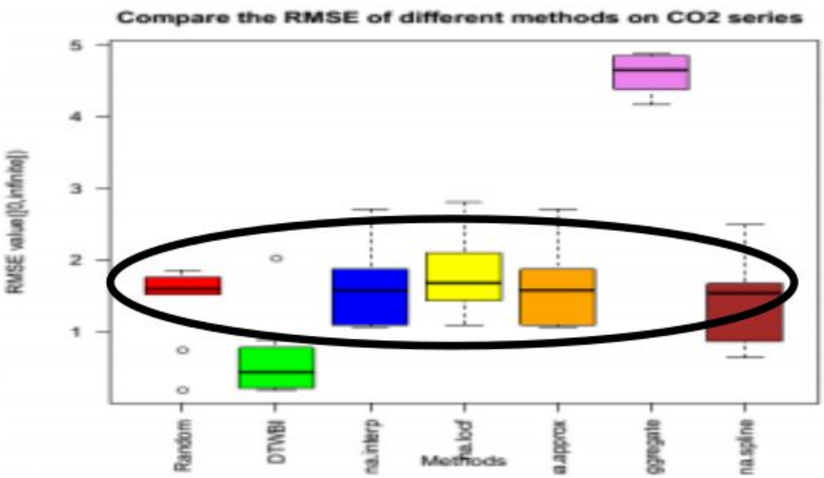


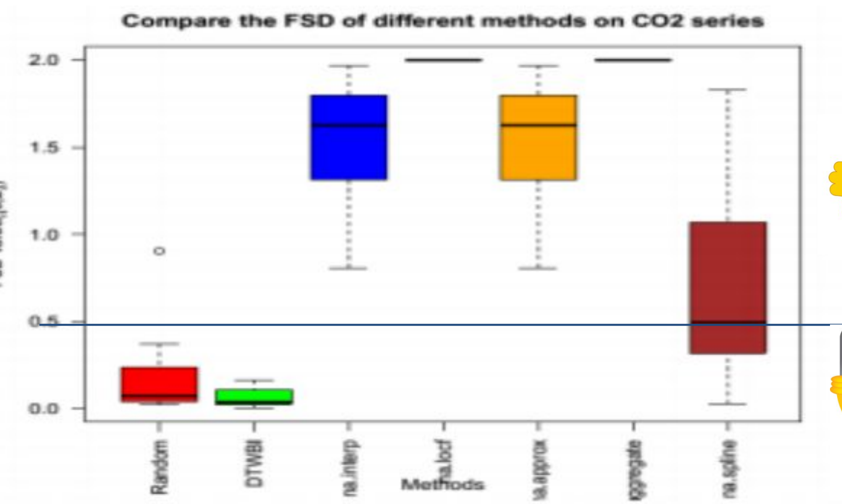


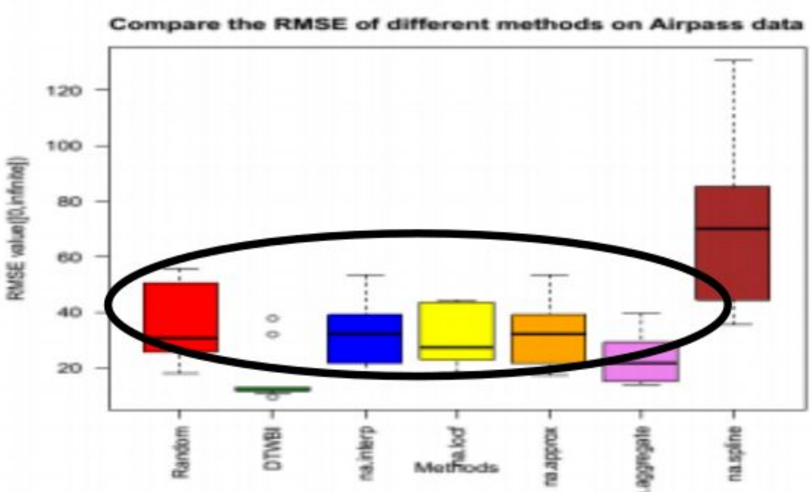
Experiments - Univariate case / comparison with random selection

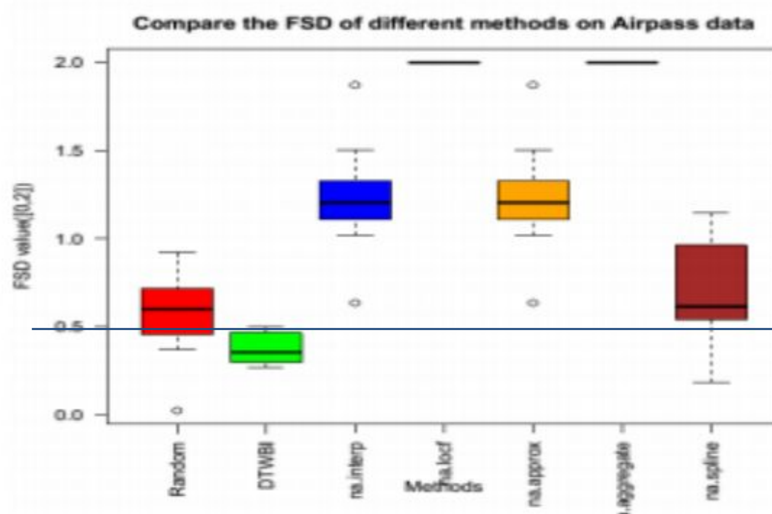
Efficiency - RMSE ?

Shape respect - FSD ?





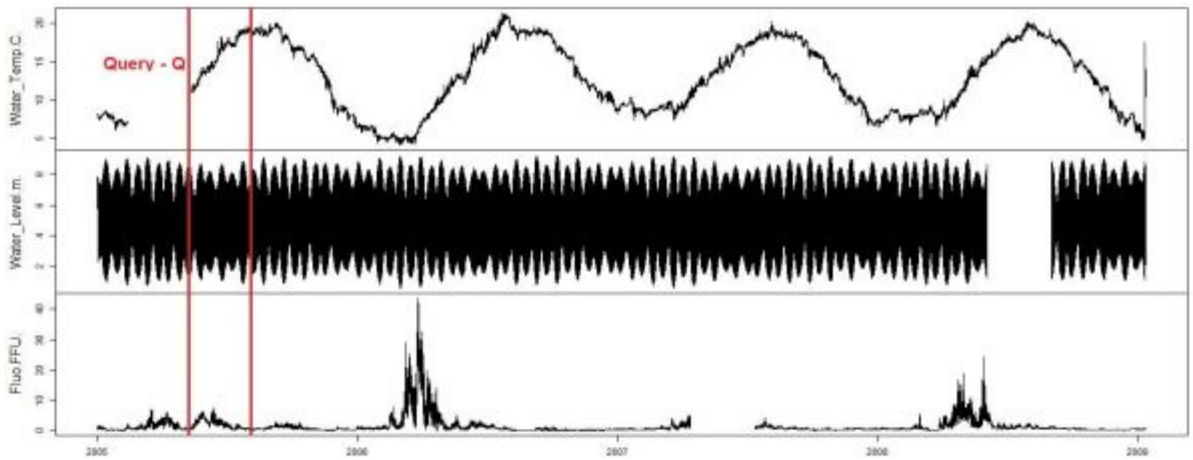


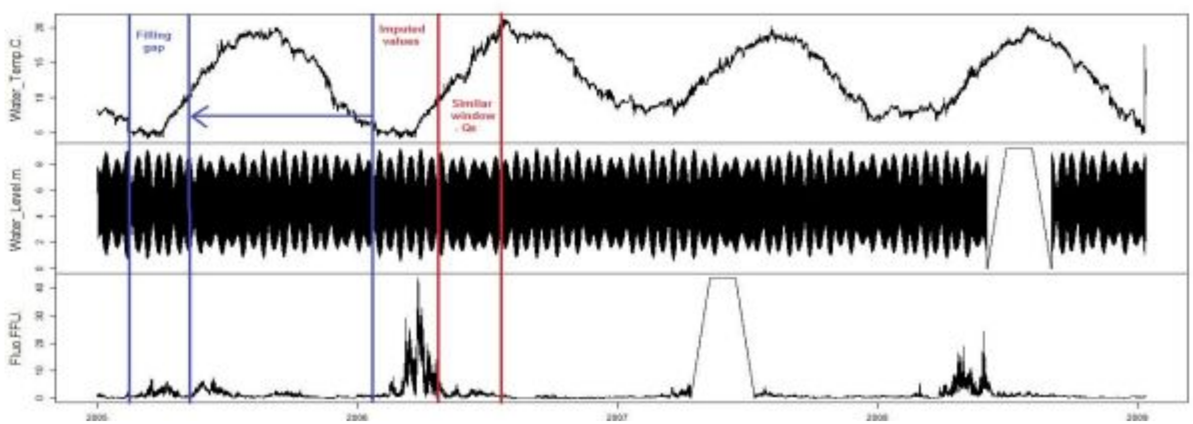


Multivariate dataset

- one-by-one independant filling
- **multivariate similarity**

Extended approach to **multivariate signals - DTWUMI**





Trapezoid filling of uncomplete signals during the sliding research.

1. Extract shape-features for Q and Reference

2. Select n References satisfying cosine criterion

3. Compute DTW cost of Q and each Reference

4. Select Qs having the the minimum DTW cost

5. Replace the missing values by previous window of Qs on the incomplete signal

Experiments **multivariate signals - DTWUMI**

N0	Dataset name	N0 of instants	Correlation (No/Low/High)	Frequency
1	NNGC	1,745	High	Hourly
2	Simulation	32,000	No	
3	MAREL-Carnot	35,344	Low	Hourly

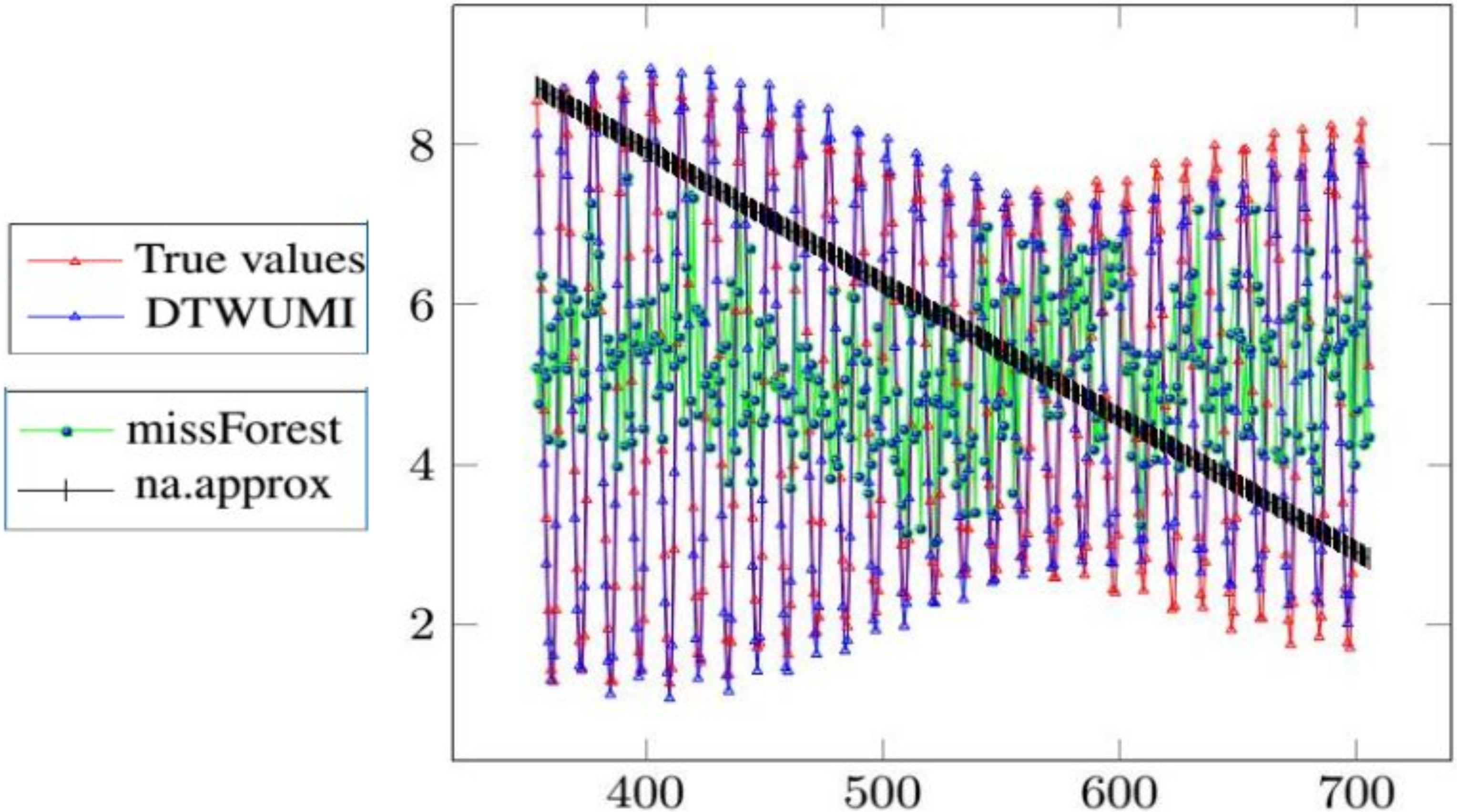
Average imputation performance indexes of various imputation algorithms

Gap size	Method	NNGC dataset: Accuracy indices 1-Sim	NNGC: Accuracy $1-R^2$	NNGC: Accuracy RMSE	NNGC: Shape indices FSD	NNGC: Shape FB	NNGC: Shape 1-FA2	Simulated dataset: Accuracy indices 1-Sim	Simulated: Accuracy $1-R^2$	Simulated: Accuracy RMSE	Simulated: Shape indices FSD	Simulated: Shape FB	Simulated: Shape 1-FA2	Marel CARNOT dataset: Accuracy indices 1-Sim	Marel: Accuracy $1-R^2$	Marel: Accuracy RMSE	Marel: Shape indices FSD	Marel: Shape FB	Marel: Shape 1-FA2
1%	na.approx	0.2	0.99	11786	0.41	0.19	0.52	0.126	0.994	1.99	0.52	**1.86**	0.81	0.068	0.15	1.62	**0.07**	**0.03**	0.21
	MI	0.1	0.32	5774	0.02	0.01	0.26	0.14	0.999	2.22	0.12	1.89	0.79	0.19	0.44	4.48	0.42	0.24	0.48
	MICE	0.03	0.06	2382	0.03	0.01	0.05	0.14	0.997	2.23	0.13	2.39	0.79	0.16	0.46	4.51	0.37	0.2	0.39
	missForest	**0.02**	**0.02**	**1286**	**0.01**	**0.01**	**0.01**	0.11	0.996	1.69	0.89	5.49	0.85	0.15	0.26	3.2	0.35	0.18	0.32
	DTWUBI	0.12	0.51	7591	0.03	0.12	0.3	**0.085**	**0.51**	**1.22**	**0.01**	5.86	**0.58**	**0.056**	**0.04**	**1.02**	0.11	0.05	**0.15**
3%	na.approx	0.18	0.99	11329	0.66	0.29	0.55	0.11	0.998	1.88	0.69	2.08	0.81	0.08	0.17	1.8	0.09	0.07	0.19
	MI	0.1	0.29	5317	0.04	0.02	0.24	0.13	1	2.27	0.03	2.63	0.8	0.21	0.49	4.53	0.41	0.33	0.47
	MICE	0.03	0.11	3112	0.02	0.02	0.05	0.13	1	2.27	0.03	2.63	0.8	0.19	0.53	5.17	0.49	0.36	0.41
	missForest	**0.02**	**0.02**	**1375**	**0.02**	**0.01**	**0.01**	0.1	1	1.71	0.91	2.49	0.85	0.18	0.37	4.09	0.39	0.37	0.36
	DTWUBI	0.05	0.19	4219	0.05	0.08	0.08	**0.064**	**0.45**	**1.16**	**0.01**	**1.72**	**0.54**	**0.056**	**0.06**	**1.07**	**0.09**	**0.02**	**0.12**
10%	na.approx	0.18	1	11419	0.62	0.25	0.56	0.11	1	2.01	0.46	**2.02**	0.79	0.083	0.23	3.09	0.15	0.16	0.27
	MI	0.1	0.35	5892	**0.008**	0.02	0.27	0.12	1	2.24	0.02	2.18	0.79	0.13	0.43	4.35	0.16	0.14	0.46
	MICE	0.04	0.13	3435	0.01	0.01	0.06	0.12	1	2.25	0.02	16.56	0.79	0.12	0.5	4.78	0.21	0.18	0.41
	missForest	**0.02**	**0.05**	**1990**	0.02	**0.004**	**0.03**	0.09	1	1.7	0.91	1.35	0.86	0.1	0.29	3.47	0.25	0.15	0.3
	DTWUBI	0.05	0.21	4402	0.02	0.04	0.08	**0.064**	**0.47**	**1.18**	**0**	4.49	**0.56**	**0.065**	**0.2**	**2.58**	**0.12**	**0.13**	**0.2**

boundary : 1-FA2<0.2 |FB|<0.2 FSD<0.5

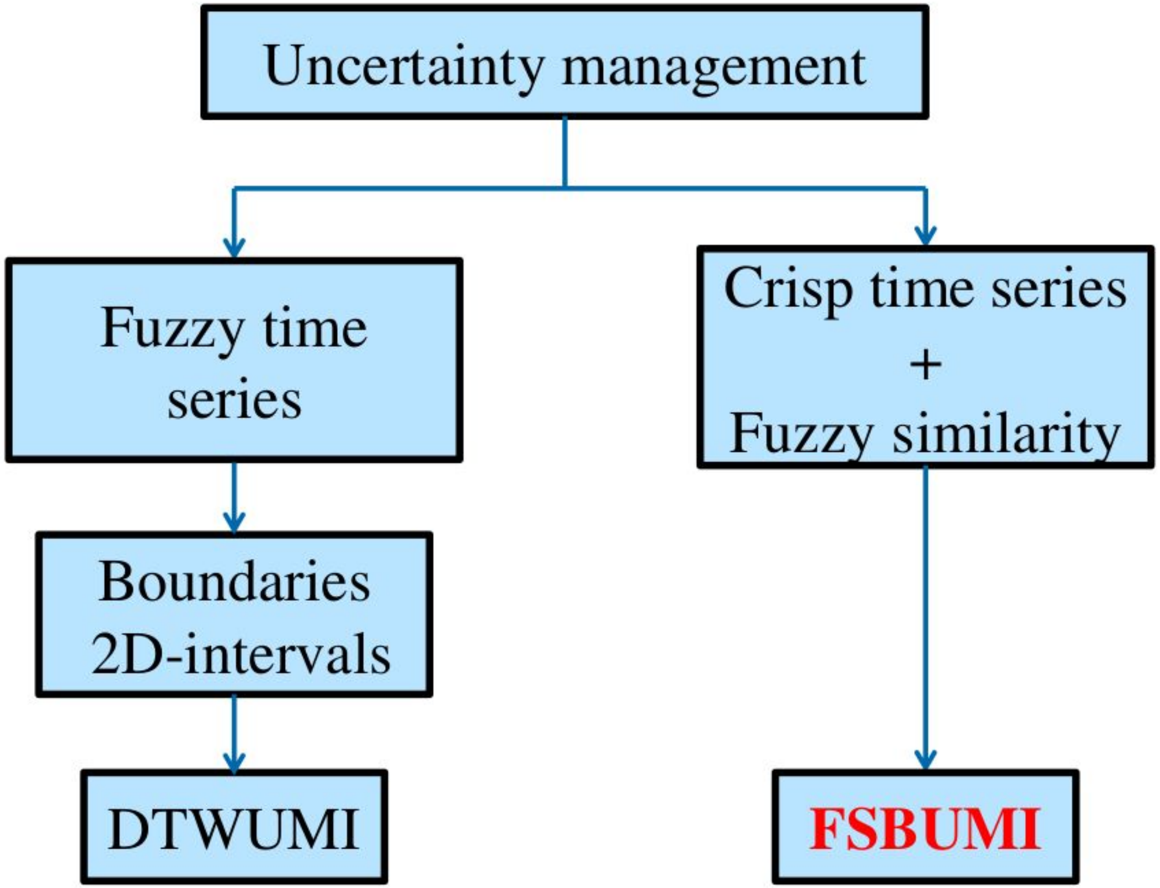
Experiments **multivariate signals - DTWUMI**

N0	Dataset name	N0 of instants	Correlation (No/Low/High)	Frequency
1	NNGC	1,745	High	Hourly
2	Simulation	32,000	No	
3	MAREL-Carnot	35,344	Low	Hourly

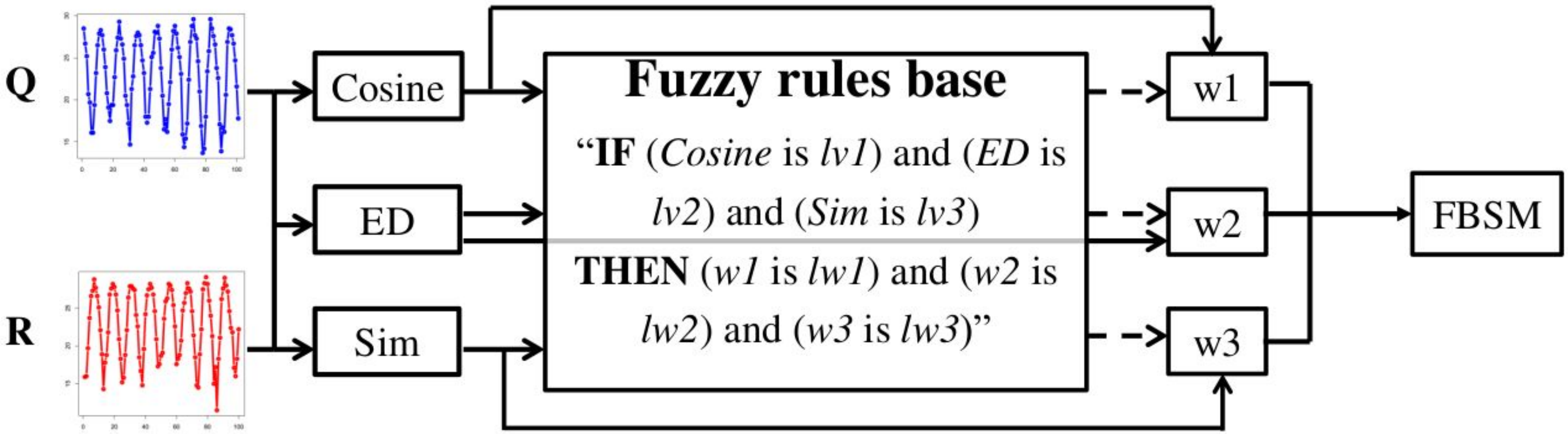


Visual comparison of imputed values of different imputation methods with true values on Marel Carnot dataset with the gap size of 353 on the 2nd signal.

Dealing with uncertainty - speed trade-off



FSBMI scheme. New similarity computation to select best similar windows

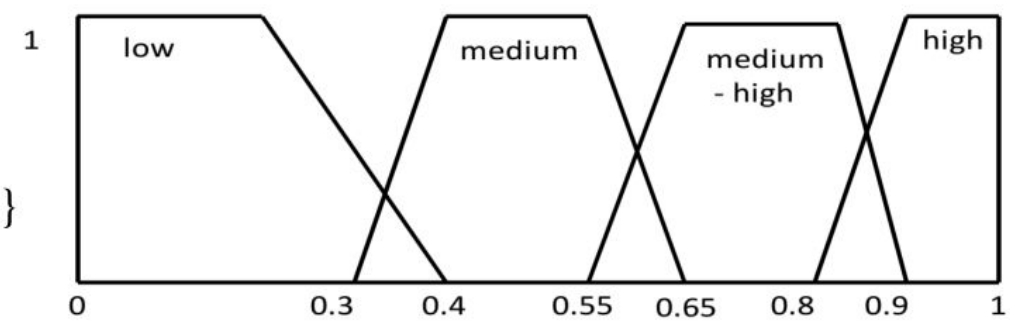


Computing scheme of the new similarity measure

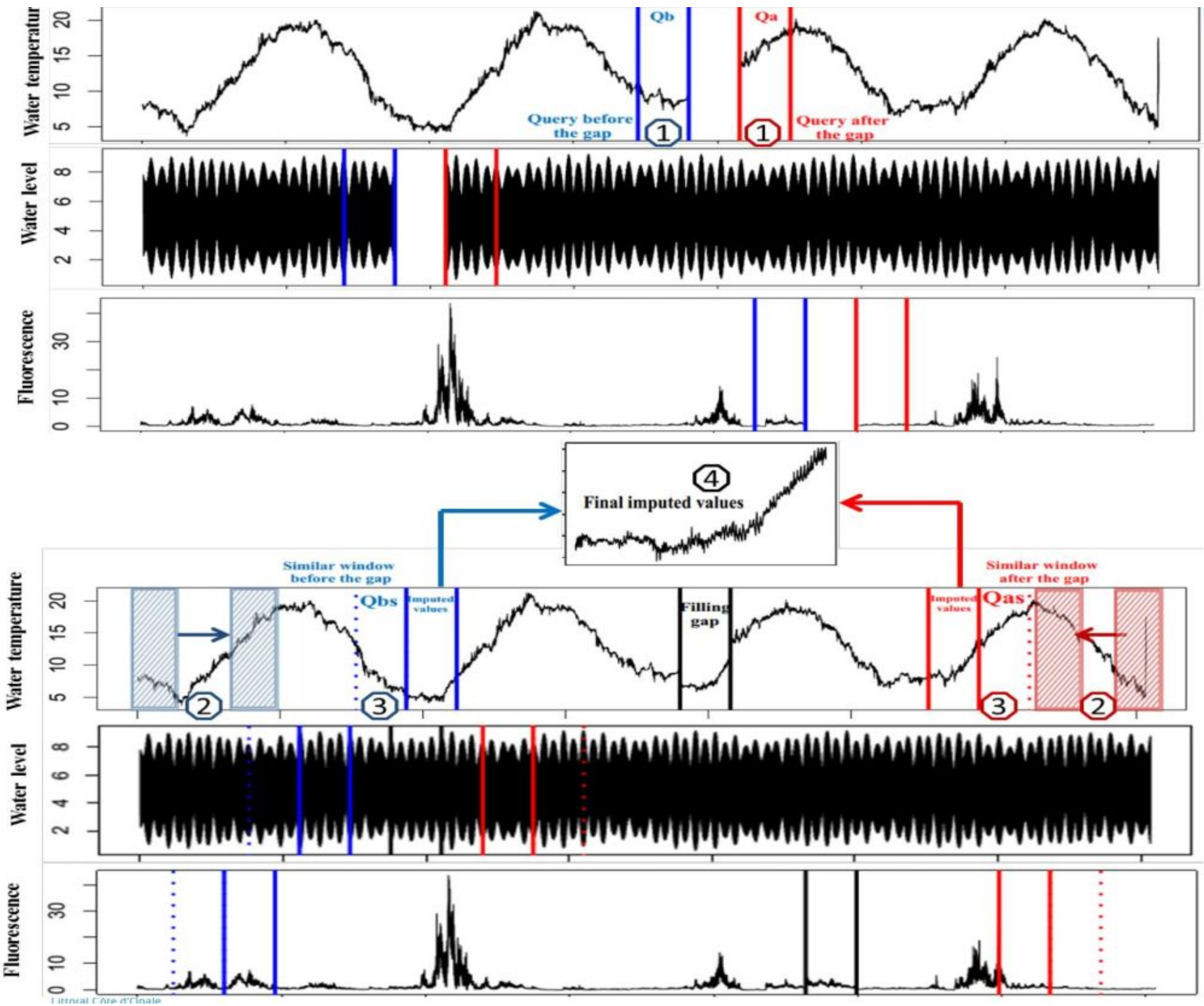
lv: linguistic variable

lw: linguistic weight

lv, *lw* ϵ {low, medium, medium-high, high}



Membership function of fuzzy similarity values


Water temperatur
Qb
Qa
Query before the gap
1
Query after the gap
Water level
Fluorescence
4
Final imputed values
Similar window before the gap
Qbs
Imputed values
Filling gap
Imputed values
Qas
Similar window after the gap
Water temperature
2
3
Water level
Fluorescence

Simulated dataset (5 signals- 32,000 points)

Gap size	Method	Accuracy indices			Shape indices		
		1-Sim	1-R^2	RMSE	FSD	FB	1-FA2
7.5%	FSMUMI	**0.049**	**0.071**	**0.027**	0.069	**0.505**	**0.184**
	Amelia	0.197	0.998	0.147	0.045	1.305	0.792
	FcM	0.158	0.809	0.104	1.813	1.866	0.991
	MI	0.2	0.992	0.15	**0.038**	1.645	0.797
	MICE	0.205	0.988	0.15	0.057	10.744	0.799
	missForest	0.188	0.97	0.136	0.284	4.396	0.812
	na.approx	0.192	0.971	0.142	0.669	2.163	0.712
	DTWUMI	0.133	0.653	0.908	0.064	1.113	0.571
10%	FSMUMI	**0.061**	**0.181**	**0.043**	0.114	**0.511**	**0.26**
	Amelia	0.202	0.999	0.147	0.034	4.062	0.788
	FcM	0.164	0.872	0.104	1.837	2.201	0.992
	MI	0.21	0.997	0.155	0.12	2.954	0.785
	MICE	0.209	0.996	0.15	0.055	3.994	0.779
	missForest	0.194	0.97	0.135	0.308	3.024	0.811
	na.approx	0.183	0.997	0.129	0.372	1.455	0.719
	DTWUMI	0.155	0.782	0.893	**0.026**	1.182	0.626

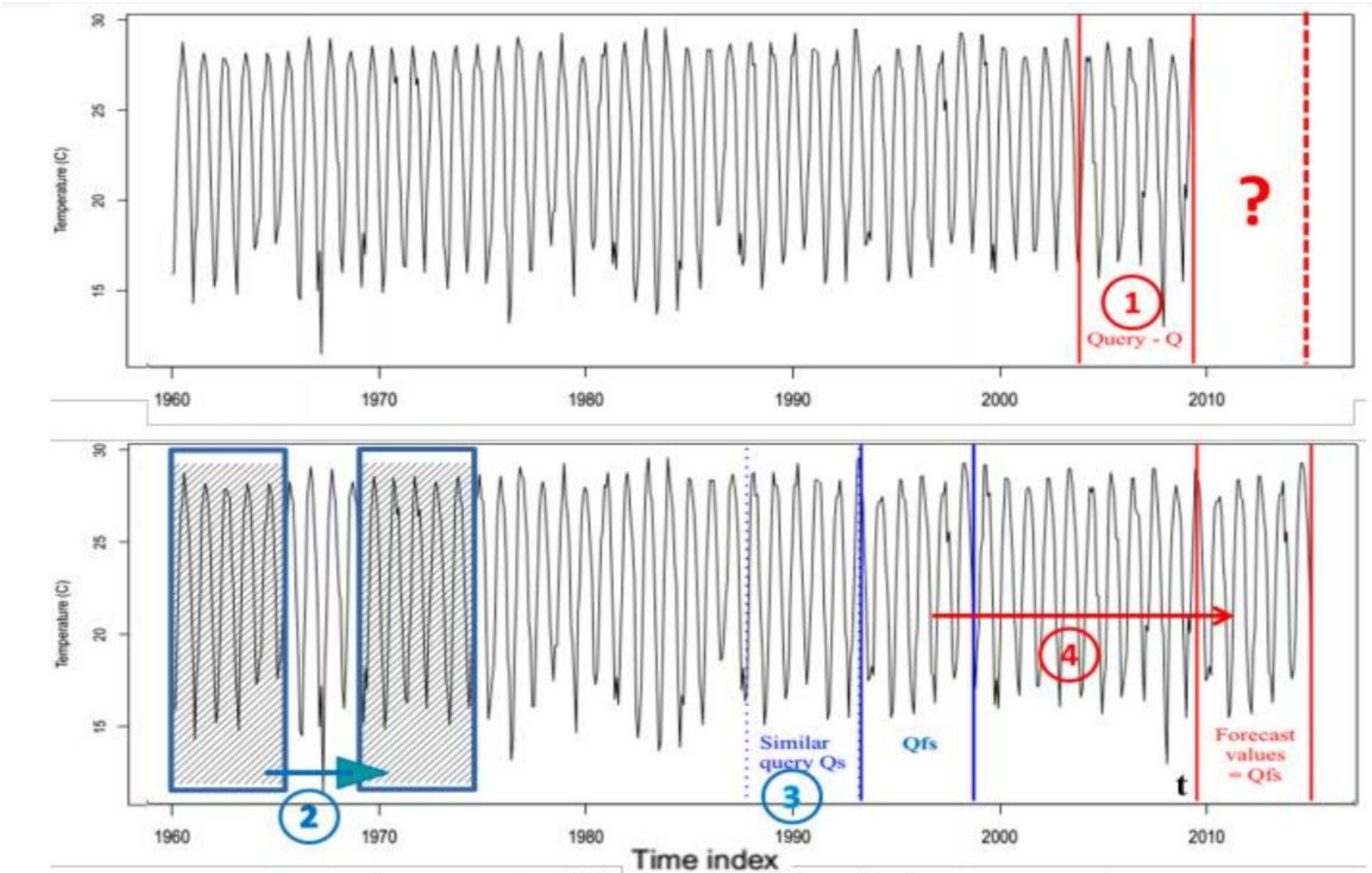
Marel Carnot hourly dataset
(3 signals- 35,334 points)

Gap size	Accuracy indices			Shape indices		
	1-Sim	1-R^2	RMSE	FSD	FB	1-FA2
4%	**0.059**	**0.058**	**1.466**	0.094	**0.101**	0.183
	0.171	0.44	4.389	0.287	0.2	0.456
	0.126	0.152	2.779	0.285	0.203	0.727
	0.166	0.41	4.234	0.277	0.204	0.444
	0.15	0.379	4.15	0.268	0.19	0.411
	0.129	0.234	3.134	0.23	0.187	0.303
	0.077	0.13	2.006	**0.068**	0.135	0.268
	0.07	0.105	1.77	0.15	0.12	**0.138**
10%	**0.053**	**0.098**	**1.642**	**0.083**	**0.055**	**0.191**
	0.14	0.3	4.294	0.24	0.142	0.442
	0.1	0.098	3.68	0.136	0.101	0.303
	0.14	0.112	4.294	0.24	0.142	0.442
	0.12	0.42	4.066	0.152	0.077	0.383
	0.097	0.461	3.049	0.104	0.117	0.255
	0.071	0.529	1.873	0.098	0.094	0.253
	0.081	0.381	3.293	0.119	0.124	0.224

boundary : 1-FA2<0.2 |FB|<0.2 FSD<0.5

Other application: Near Future Prediction

Direct application : DTWBI for Forecasting Univariate TS



1-Query building, 2-Sliding window comparison,
3-Window selection, 4-Forecasting values

Documentation

Related Papers.

Univariate time series imputation

• DTWBI - Dynamic time warping-based imputation for univariate time series data. Pattern Recognit. Lett. 139: 139-147 (2020) https://doi.org/10.1016/j.patrec.2017.08.019

• eDTWBI - eDTWBI: Effective Imputation Method for Univariate Time Series Advanced Computational Methods for Knowledge Engineering. ICCSAMA 2019. Advances in Intelligent Systems and Computing, vol 1121. Springer, https://doi.org/10.1007/978-3-030-38364-0_11

• Comparison of DTW variants for the imputation (Ocean'2017), Which DTW Method Applied to Marine Univariate Time Series Imputation https://doi.org/10.1109/OCEANSE.2017.8084598

Uncorrelated multivariate time series imputation

• DTWUMI (MLSP'2017), DTW-Approach for uncorrelated multivariate time series imputation. Machine Learning Signal Processing. https://doi.org/10.1109/MLSP.2017.816816

• FSMUMI (ACISC journal - 2018), A New Fuzzy Logic-Based Similarity Measure Applied to Large Gap Imputation for Uncorrelated Multivariate Time Series. https://doi.org/10.1155/2018/9095683

Applications to classification and forecasting

• Shape-feature extraction algorithm (ICCE'2016), **Comparative study on supervised learning methods for identifying phytoplankton species.** 10.1109/CCE.2016.7562650

• Forecasting meteorological univariate time series (EUSIPCO'2018) **Comparative Study on Univariate Forecasting Methods for Meteorological Time Series. https://doi.org/10.23919/EUSIPCO.2018.8553576**

CRAN pakages
https://cran.r-project.org/web/packages/DTWBI/index.html
https://cran.r-project.org/web/packages/DTWUMI/index.html
https://cran.r-project.org/web/packages/FSMUMI/index.html

Documentation: http://mawenzi.univ-littoral.fr/

Support: emilie.poisson@univ-littoral.fr

Other study:

- GAIN vs DTW-completion
- Adaptive system according signal characteristics (in progress)
- GUI interface (-> 2022)