European Workshop
on
Data Stream Analysis

Caserta, 14-16 mar. 07

Smart Alarming Methods: an overview, highlight on statistical methods

Jean-Paul Valois[1],
Christophe Blondeau[1],
Simplice Dossou-Gbete[2],
Laurent Bordes[2]

1 TOTAL

2 UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

jean-paul.valois@total.com

Background, objective

TOTAL

General Topic, objective of Smart alarming

General framework by G. Cormode yesterday

- **DEFINITION / PURPOSE**
  - to **detect any change (novelty)** before it becomes obvious,
  - to describe it,
  - if needed (or possible) : thus prevent its consequences

- **NEEDS**
  - An automatic “black box” procedure is often the final product

Aim of the present paper

Exhibit the guide lines to correctly design an industrial project

- **A bibliographic investigation**

A large variety of statistical methods
because a lot of diversified concrete situations

The overviews most often take first into account the type of algorithms
*see **Marcou & Singh, 2003, Hodge 2004**.*

→ Which questions to be posed to conceive a smart-alarming work flow ?

- **In some cases :**
  - programming selected methods
  - experimentation on synthetic data sets or real data sets from oil industry

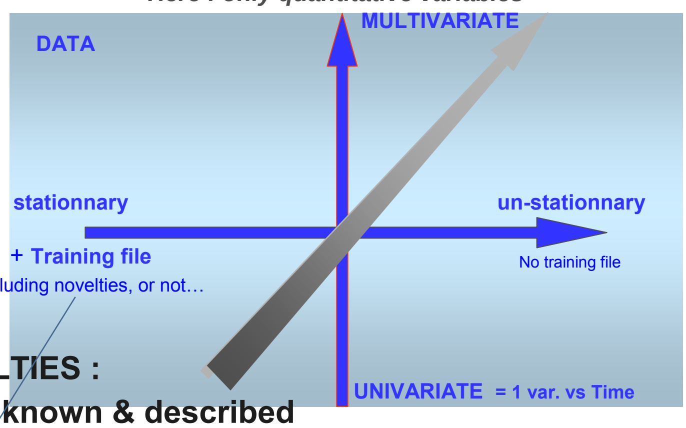
A panorama of the methods

NOVELTIES :
a priori unknown
or a quite new situation
could occur

Constraints on the data

Constraints on the task

Here : only quantitative variables



NOVELTIES :

a priori known & described

= metadata,
technical knowledge
previous expert diagnostic, etc

* or the variables can be taken separately
=> no correlation between them

Univariate* Stationnary Data

Unknown novelty

Even in this simplest case,
You can be faced with different situations

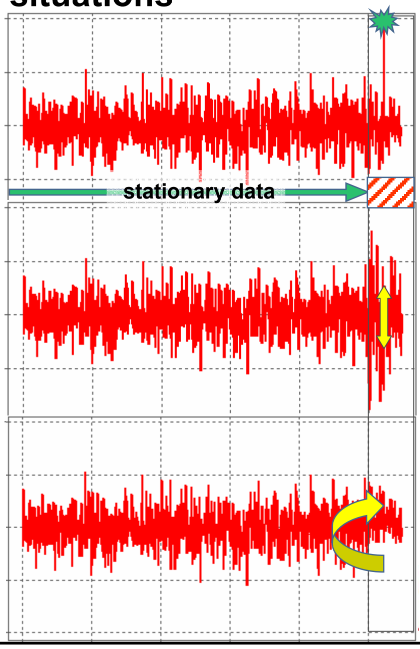
(1) An exceptional frost

(2) A repetition of both

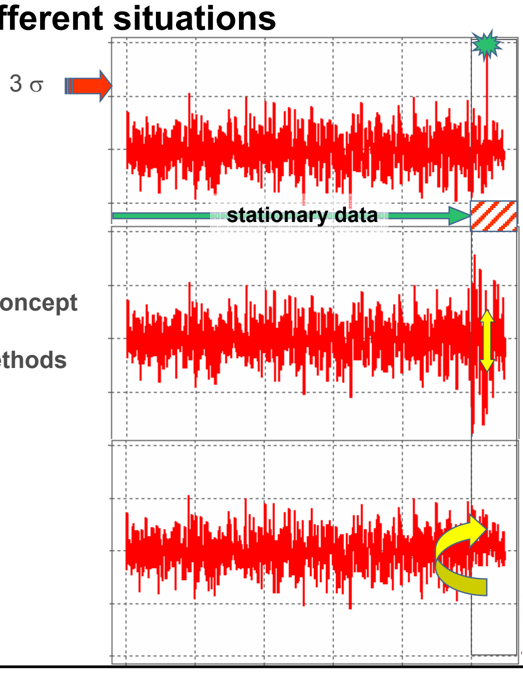
hotter sommers

colder winters

(3) A lack of very cold winters

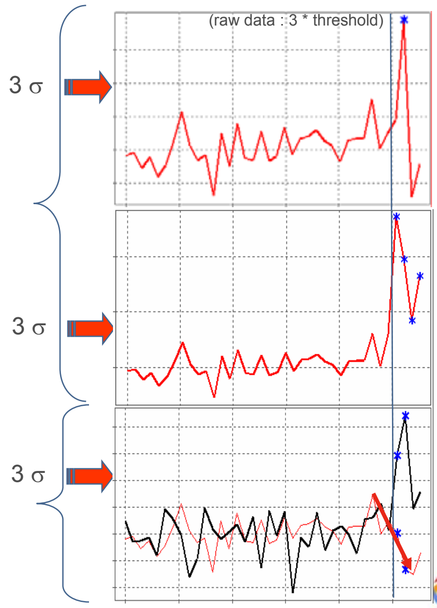


Even in this simplest case,
You can be faced with different situations



Threshold (ex: 3 σ) is the basic concept used as indicator in a lot of more sophisticated methods

Basic threshold method should sometimes be adapted

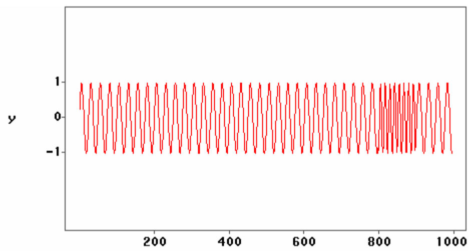


σ in a sliding window

1+(m/σ) in a mobile window

σ not suitable

A data preprocessing can be needed



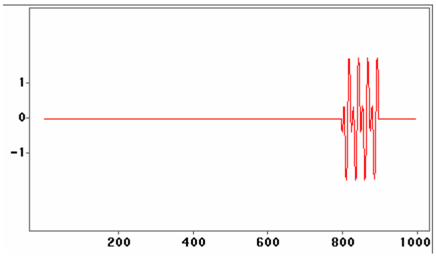
Several methods :

•difference with the base signal

•Fourier transform,

•wavelett

•…

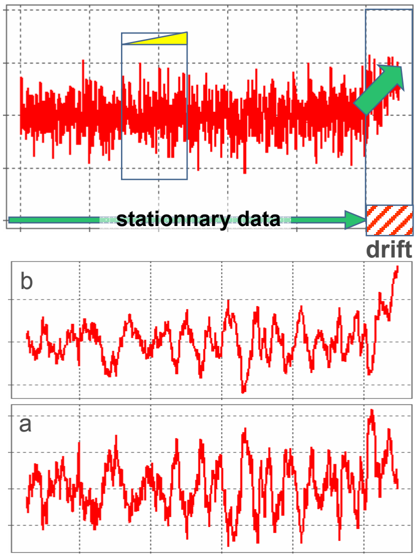


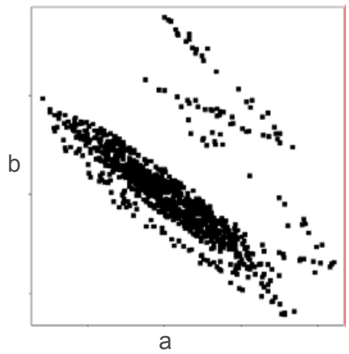
When the novelty is a drift

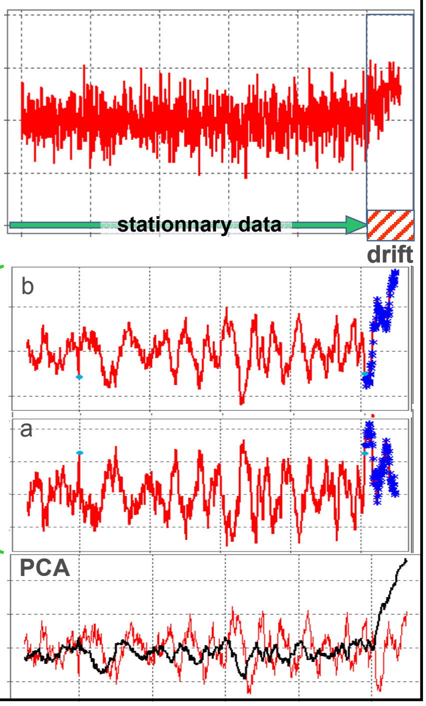
Linear Regression in a sliding window

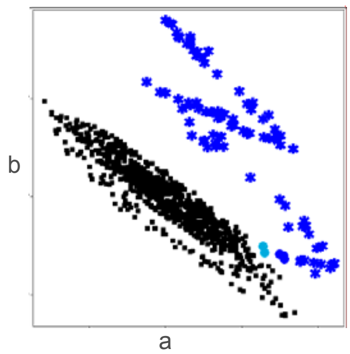
Model : y = a time + b + weigth

last value is **active** in the model
a, b could change in case of novelty









* or the variables can be taken separately
=> no correlation between them

Univariate* novelty

Stationnary Data

Unknown novelty

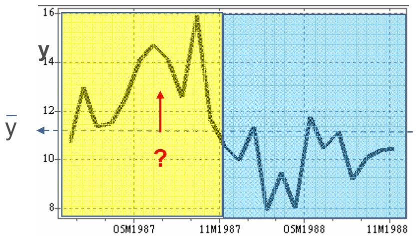
Change point analysis

PAGE (1955,1957),
HINKLEY (1971), PETTITT (1980), → Cusum
HINKLEY SCHECHTMAN (1987) → Boostrap

TAYLOR (2000), → synthesis + multilevel approach

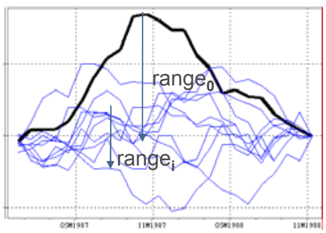
WEISHAUPT et al, 1991, Rig Computer system improves Safety for deep HP/HT wells by kicks detection and well monitoring, SPE 23053

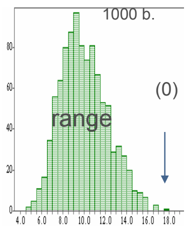
$$CUSUM_{(k)} = CUSUM_{(k-1)} + (y_{(k)} - \bar{y})$$



Data by Taylor (2000)

Time





Bootstraps i (1..n)

DIAGNOSTIC

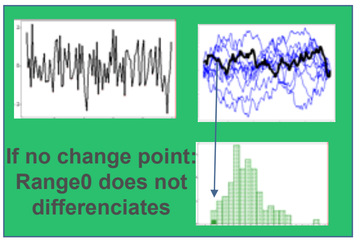
In case of 2 trends, range (0) is higher

Proba. test

Yes No

LOCALISATION

MULTILEVEL ANALYSIS



DISCUSSION (Taylor)

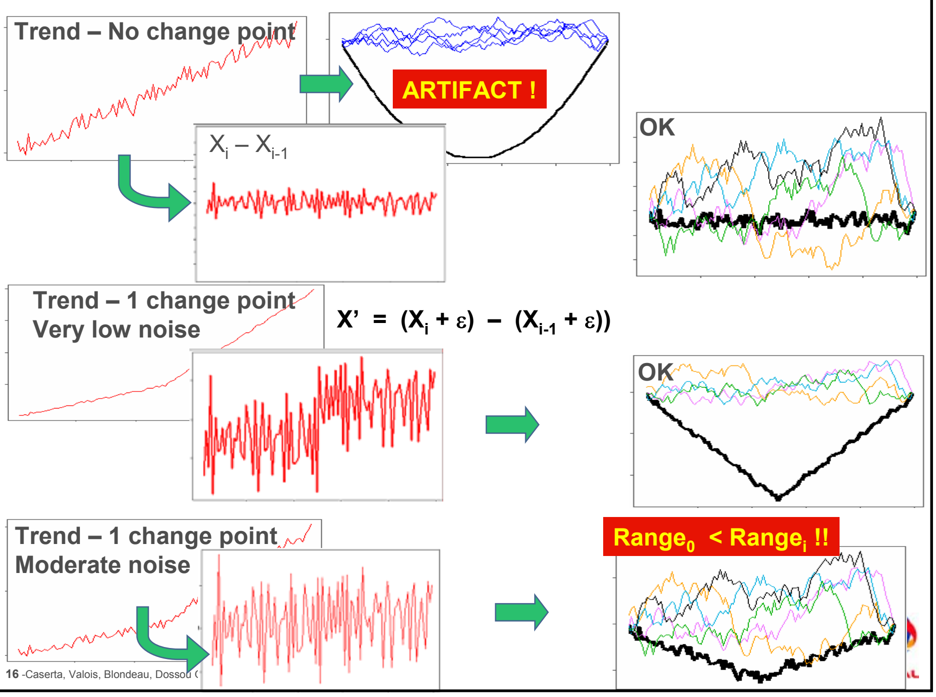
- Reasonnably robust. If needed, use the ranks.
- More flexible than control chart.
- Can detect minor shifts of the mean
- Not suitable for detecting isolated high values.

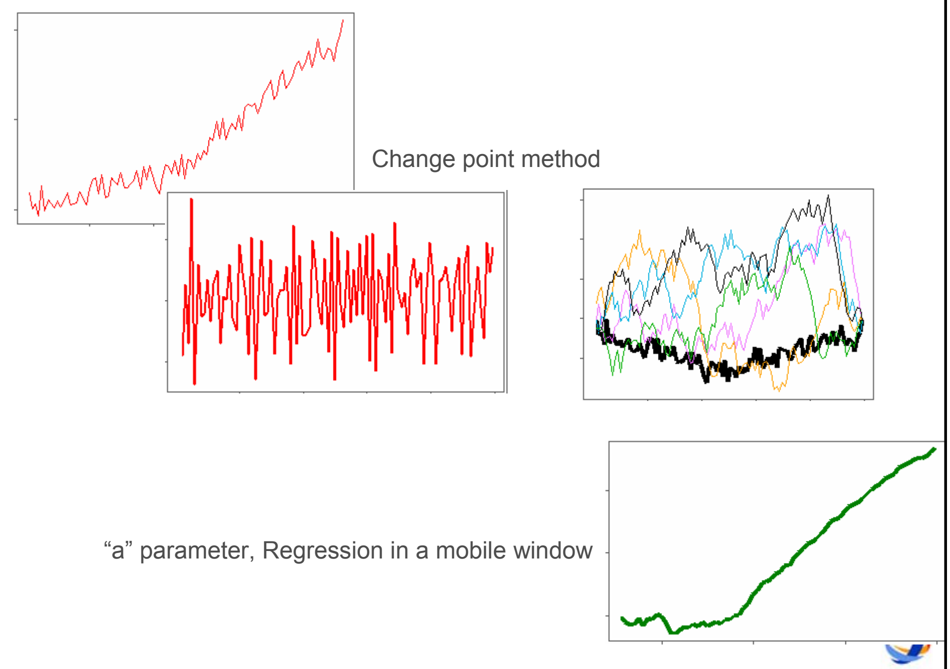
- COMMENT
- Is not a really on line method, rather an a posteriori one

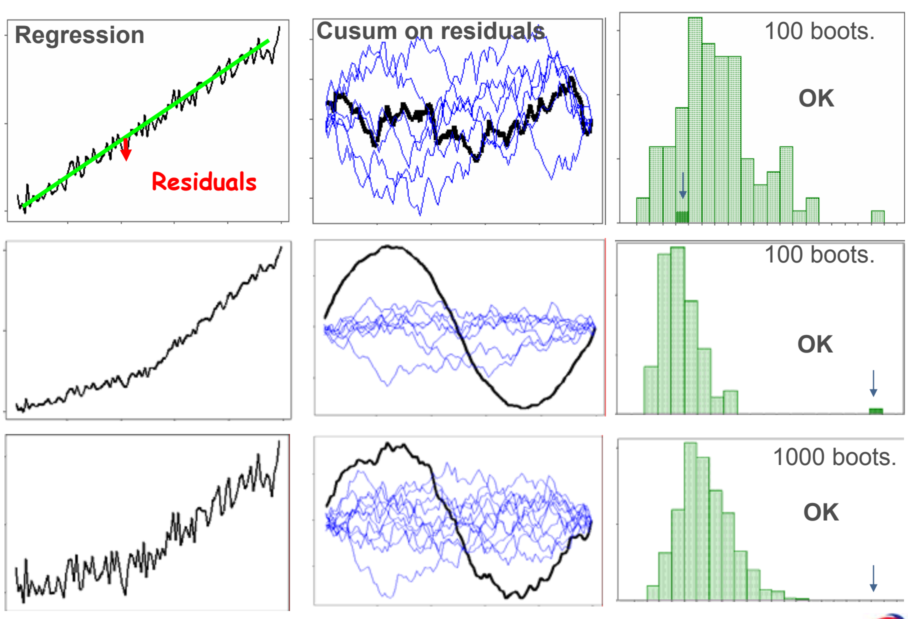
EXTENSION

Using **$X' = (X_i - X_{i-1})$** instead of (X_i), as suggested by Taylor to diagnose evolutions

Could then the CHANGE POINT method handle **no-stationnary data ?**




Change point method
“a” parameter, Regression in a mobile window


Regression
Residuals
Cusum on residuals
100 boots.
OK
100 boots.
OK
1000 boots.
OK

Change point method

is adapted to detect <u>shift of the mean</u>, *OK with Taylor*

is only suitable for <u>stationnary data</u> : *not clearly mentionned by Taylor*

In case of no-stationnary data

+ noisy data, better use residuals of regression

TOTAL

* or the variables can be taken separately
=> no correlation between them

Univariate* Stationnary Data

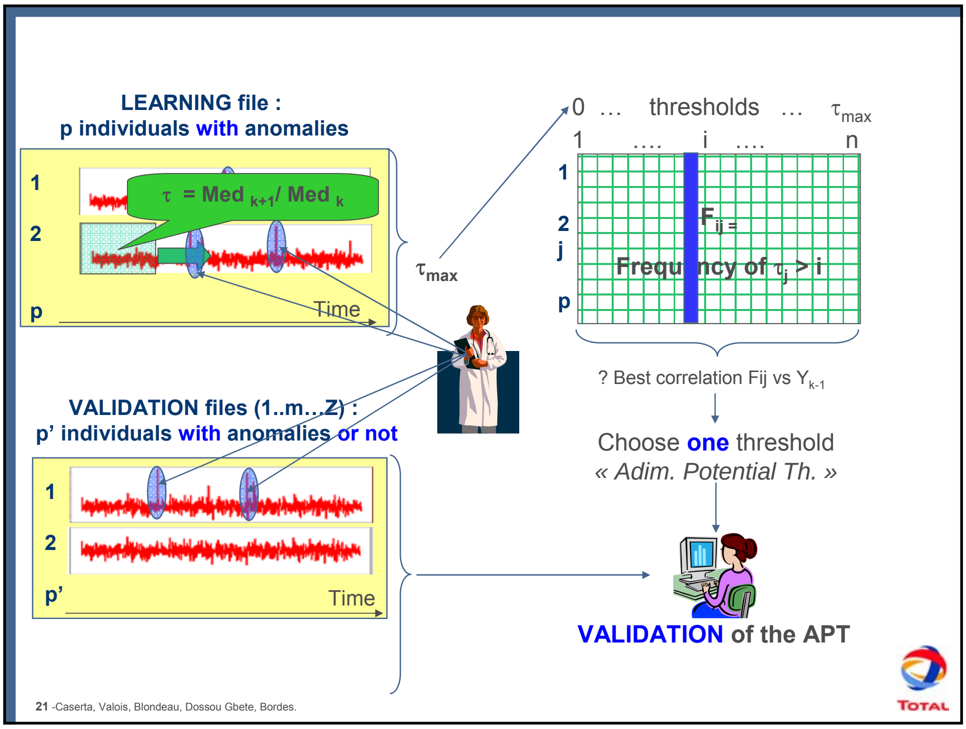
Known novelty

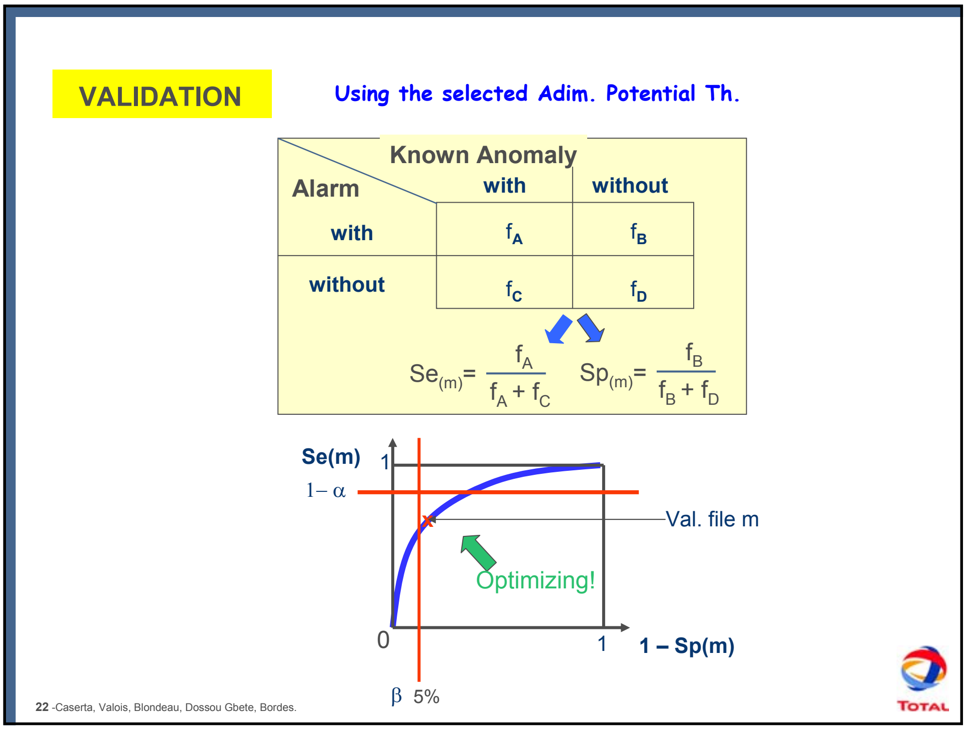
Median method

See :

SEAGULL and SANDERSON, 1998, Anesthesia alarms in surgical context,
Proceeding of the human factors and ergonomics society 42nd annual meeting

TOTAL


LEARNING file :
p individuals with anomalies
τ = Med $_{k+1}$/ Med $_k$
1
2
p
Time
τ_{max}
0 … thresholds … τ_{max}
1 …. i …. n
1
2
j
p
F_{ij} =
Frequency of τ_j > i
? Best correlation Fij vs Y_{k-1}
Choose one threshold
« Adim. Potential Th. »
VALIDATION files (1..m…Z) :
p' individuals with anomalies or not
1
2
p'
Time
VALIDATION of the APT
21 -Caserta, Valois, Blondeau, Dossou Gbete, Bordes.
TOTAL


VALIDATION
Using the selected Adim. Potential Th.
Known Anomaly
Alarm
with
without
with
f_A
f_B
without
f_C
f_D
$Se_{(m)} = \frac{f_A}{f_A + f_C}$
$Sp_{(m)} = \frac{f_B}{f_B + f_D}$
Se(m)
1
$1-\alpha$
Val. file m
Optimizing!
0
1
1 – Sp(m)
β 5%
22 -Caserta, Valois, Blondeau, Dossou Gbete, Bordes.
TOTAL

MULTIvariate Stationnary Data

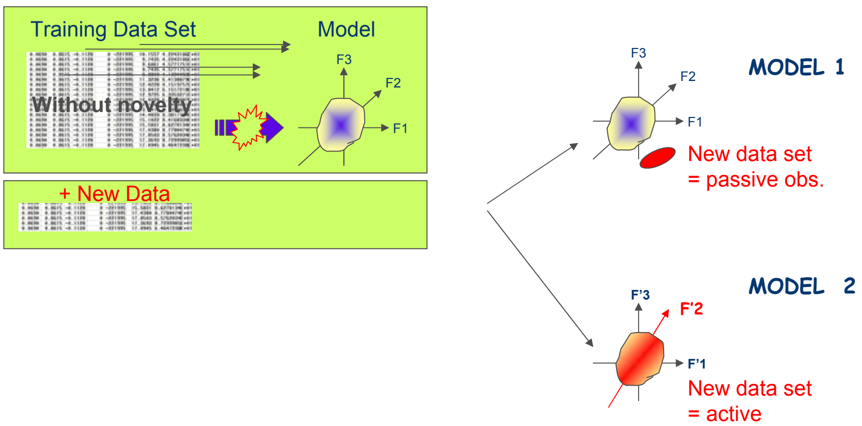
Unknown novelty

P.C.A.

Air survey

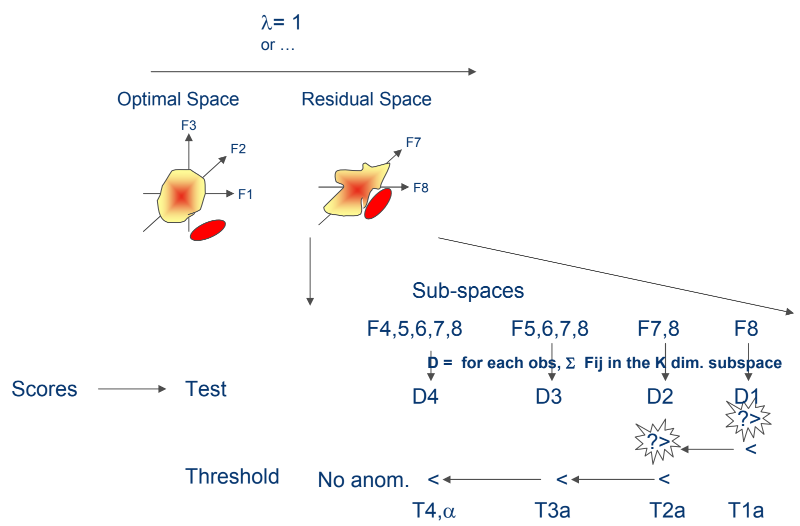
HARKAT, 2003, Détection de défauts par analyse en composantes principales, Doctorat de l'Institut Polytechnique de Lorraine, laboratoire CRAN

23 -

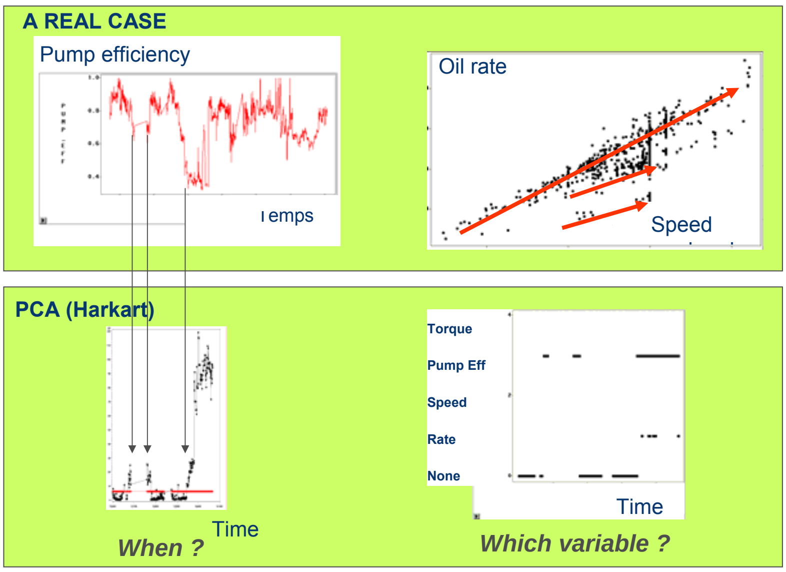


24 -Caserta, Valois, Blondeau, Dossou Gbete, Bordes.

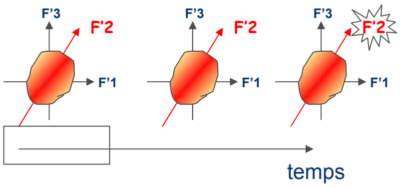
MODEL 1 New data taken as passive ones



T = f (alpha, K[th] Eig. Val.)



MODEL 2 ACP using a sliding window



? Stability of the correlations (variables, factors) ?

No stat. Test => empirical threshold

PCA by Harkart

- Needs stationnary data set + a training set without novelty
- Detection with a short delay
- Can detect simoultaneous novelties in several variables

MULTIvariate Stationnary Data

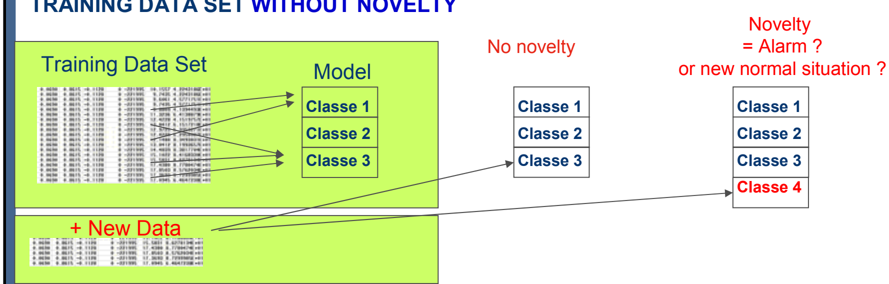
Novelty known or not

Clustering

TOTAL

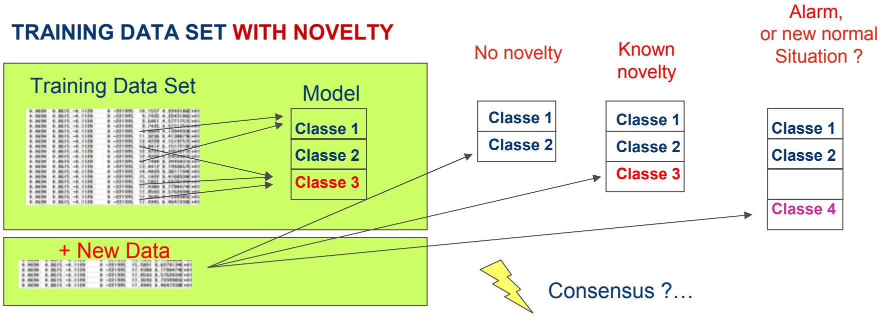
CLUSTERING

TRAINING DATA SET WITHOUT NOVELTY



TOTAL

CLASSIFICATION



Works well,
Nevertheless does not describe the novelty (which variable, …)

MULTIvariate Stationnary Data

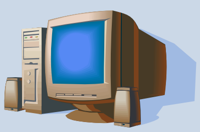
Novelty known(*)

(*) parameters can be ajdusted in the learning

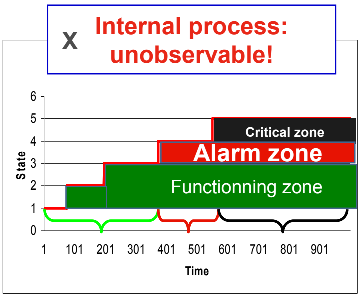
Hidden Markov Model (HMM)

(Viterbi algorithm, 1967)

Andrew J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory 13(2):260?269, April 1967.



Exemple (1) internal process



In this example : Y_n is $\mathcal{N}(m_{X_n}, 0.4)$ distributed

where $(m_1, m_2, m_3, m_4, m_5) = (1, 1.5, 2, 2.5, 3)$

Φ is finite and identified with $E = \{1,2,\ldots,k\}$ via the one-to-one map X, from Φ to E.

Write $X_n = X(\varphi_n)$ the sequence of *internal states*.

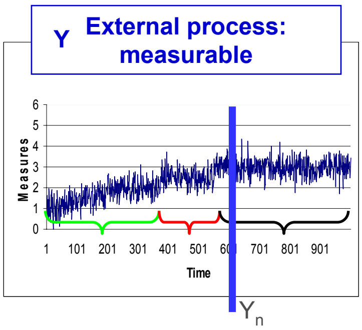
We assume that (X_n) is an homogeneous Markov chain with known para-meters and E as state space.

Here : synthetic case :
+ random function → external process

Exemple (1) external process

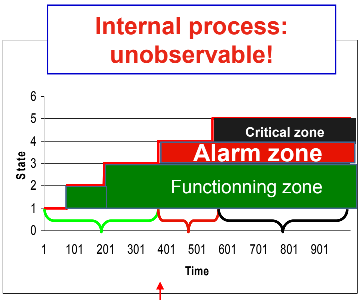
Write Y_n the *external state* of the system measured at time t_n.

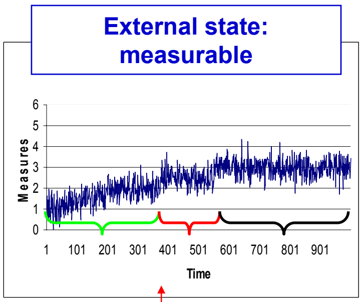
- Conditional on (X_n), the Y_n's are independent;
- Y_n depends on (X_n), only thought X_n.



uni or multivariate

Exemple (1) of internal/external processes





We should alarm at this stage

Internal process

Y signal

Viterbi result (analysis)

?

PROBLEM: given the external process (blue),

how to detect the entry of the internal process (red) in the alarm zone?

The Viterbi algorithm as a solution

Y_n may be seen as a random function of X_n

- Assume that at time t_n we know for all x in E:

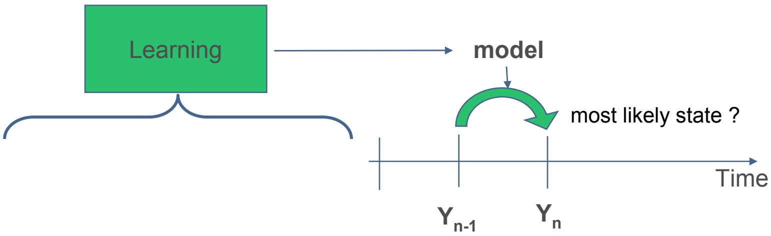
$$V_n(x) = \max_{x_1,\ldots,x_{n-1}} P(X_1=x_1,\ldots,X_{n-1}=x_{n-1},X_n=x \mid Y_1,\ldots,Y_n)$$

- Update $t_n \rightarrow t_{n+1}$ (induction):

$$V_{n+1}(x) = \underbrace{f(Y_{n+1} \mid \varphi_x)}_{\text{known density of } Y_{n+1} \text{ given } X_{n+1} = x} \times \max_{u \in E} V_n(u) \times \underbrace{p(u,x)}_{\text{known transition probability}}$$

- **At time t_{n+1} the most likely state is $x_{n+1} = \arg\max_{x \in E} V_{n+1}(x)$,**

then send an alarm if x_{n+1} belongs to the alarm set Φ_2 .



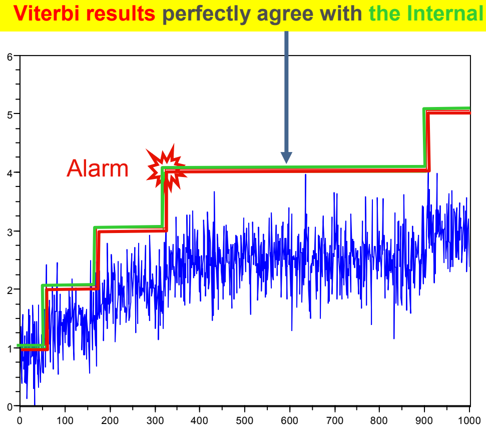
Yn depends only on Yn-1

(via the probabities given in the matrix model)

→ suitable in the Data Stream context

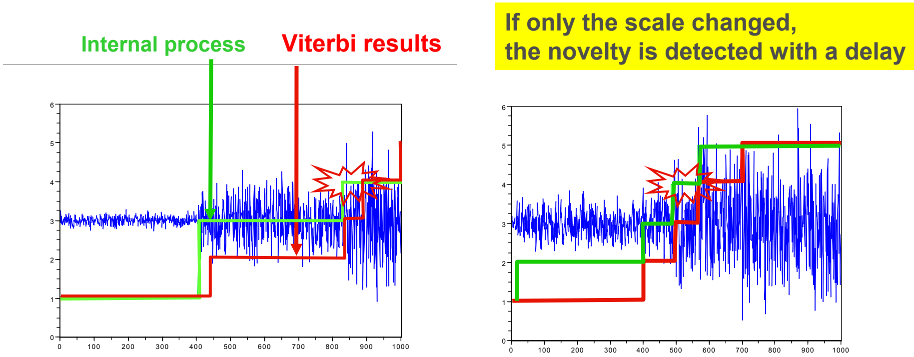
Exemple (1): shift of the mean

Viterbi results perfectly agree with the Internal process



$$P = \begin{pmatrix} 0.995 & 0.005 & 0 & 0 & 0 \\ 0 & 0.995 & 0.005 & 0 & 0 \\ 0 & 0 & 0.995 & 0.005 & 0 \\ 0 & 0 & 0 & 0.995 & 0.005 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Example (2): change in scale



Same transition matric P as for example 1,

Yn is N(3, σXn) distributed

where (σ1, σ2, σ3, σ4, σ5) = (0.1, 0.3, 0.5, 0.8, 1).

HMM, discussion

- Values of the model parameters not need to be perfectly known
  can be estimated from learning data via (e.g.) maximum likelihood techniques.
- HMM can be seen as dynamic mixture models,
  - it includes the iid case (all the lines of P are equal ⇒ Y1,Y2,… are iid) for which the Viterbi algorithm is useless
- For the iid case we have just to estimate at tn: P(φn ∈ Φ2 | Yn) (classification method based on mixture model by EM-type algorithms: Celeux, Diebolt, Govaert,…).

NB : Some LAN (Local Asymptotic Normality) methods available to detect small changes in parameters (Basseville, Beneveniste, Tromp,…).

CONCLUSION

41 -

=> SELECT THE METHOD SUITABLE FOR YOUR WORKFLOW :

- Data → Stationnary or not ?
  Uni or multivariate ?
  Characteristics of variables, etc
- Task → Novelty previously described or not ?

= Some basic methods work very well in case of basic problem

= More sophiscated methods are needed :
to obtain a better signal/noise ratio,
or in case of more complex (e.g. multivariate) situations

Our paper does not examplify two important points :

1/ DATA PREPROCESSING CAN BE USEFUL :
= avoid false alarm and no-detections

2/ GRAPHIC TOOLS FOR DISPLAYING THE ALARM
(Cleveland, Tukey… Heiberger & Holland 2004…)

42 -Caserta, Valois, Blondeau, Dossou Gbete, Bordes.