Adapted Wavelets for Pattern Recognition

Hector Mesa

UNIVERSITÉ PARIS-SUD 11

Universidad de la Habana

September 25, 2006

Thesis advisors:

Pr. J.-M. Poggi   Pr. H. de Arazoza

Introduction

The fields of applications of wavelets are growing involving, for instance

- Signal Analysis
- Image Processing
- Function Approximation

We use wavelets for

Pattern detection problems.

Introduction

$$\textit{admissibility condition:} \quad \int_{\mathbb{R}} \frac{\left|\widehat{\Psi}(\omega)\right|^2}{|\omega|} d\omega < +\infty$$

For real wavelets, it is sufficient to satisfy

$$\int_{\mathbb{R}} \Psi(x)dx = 0, \quad and \quad x\Psi(x) \in L^1(\mathbb{R})$$

There exist several families of wavelets.
It is possible to create new ones.

Introduction (cont.)

What we do

We give a method to construct
pattern-adapted wavelets
and we use them to solve
some **pattern detection problems**.

We will work only with 1D real wavelets
and compactly supported patterns.

Outline

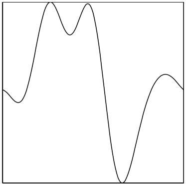
Pattern-detection problems

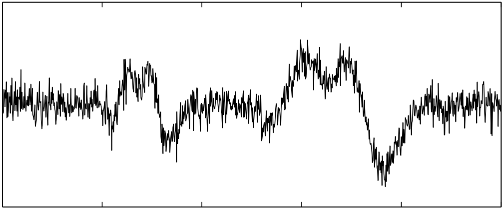
Given:

- A pattern.
- the signal.

Goal:

Detect where the signal is (locally) similar to the motif.





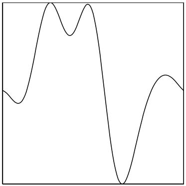
Pattern-detection problems

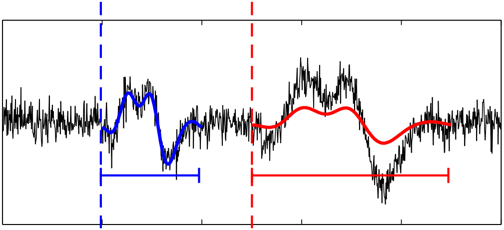
Given:

- A pattern.
- the signal.

Goal:

Detect **where** the signal is (**locally**) similar to the motif.





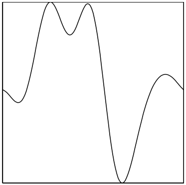
Pattern-detection problems

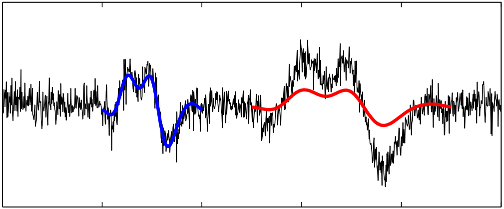
Given:

- A pattern.
- the signal.

Goal:

Detect where the signal is (locally) **similar** to the motif.





Matched Filters

Two methods are commonly used to measure the similarity:

Signal substraction

$$\mathcal{D}(s,f) = \|s-f\|^2$$

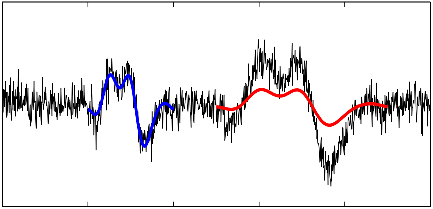
$$\Longleftrightarrow$$

$\|s\|=\|f\|=1$

$\int_{\mathbb{R}} s\cdot dx = \int_{\mathbb{R}} f\cdot dx = 0$

Correlation

$$\cos(s,f) = \frac{\langle s,\ f\rangle}{\|s\|\ \|f\|}$$



Matched Filters

Two methods are commonly used to measure the similarity:

Signal substraction

$$\mathcal{D}(s, f) = \|s - f\|^2$$

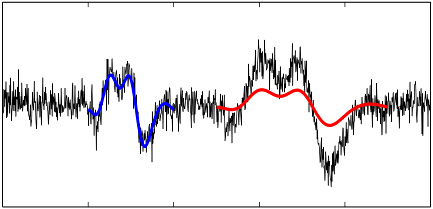
$\Longleftrightarrow$

$\|s\| = \|f\| = 1$

$\int_{\mathbb{R}} s \cdot dx = \int_{\mathbb{R}} f \cdot dx = 0$

Correlation

$$\cos(s, f) = \frac{\langle s, f \rangle}{\|s\| \, \|f\|}$$



Matched Filters

Template matching or pattern matching is
a frequently used technique (Brunelli and Poggio(1997)).

For problems in 1D it is known as Matched Filters.

The cross-correlation between the signal and the matched filter is computed to estimate the local similarity for each time shift.

Why with wavelets?

- A zero-average 1D compactly supported function is close to a wavelet.
- The wavelet transforms allow to decompose any signal in "frequency" bands.
- There exist efficient implementations of the transforms.
- The wavelet analysis also allows to compute the cross-correlation along the scales.

Pattern-Adapted Wavelets.

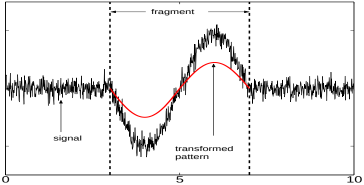
Adapted wavelets for pattern detection

For any compactly supported pattern f, an adapted wavelet ψ_f and a signal S,

The continuous wavelet transform (CWT)

$$W_{\psi_f}S(a,b) = \left\langle \frac{1}{\sqrt{a}}\psi_f(\frac{x-b}{a}),\ S(x) \right\rangle$$

Estimate the local correlation of S and $f(\frac{x-b}{a})$ for any $a > 0$.



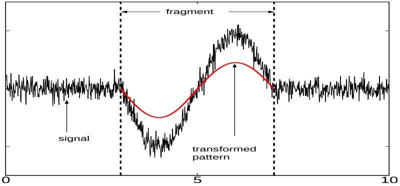
Also, for high correlation:

$$Amplitude \ \propto \ \frac{1}{\sqrt{a}}W_{\psi_f}S(a,b)$$

Our detection problem

Pattern Detection Problem:

Given a finitely supported pattern f and a 1D signal S, to find where the signal is locally similar (in terms of correlation) to a scaled version of the pattern estimating both time-shift and scale factor.



Assumption:

$supp(f) = [0, 1]$

Let us denote

$$\cos(a, b) = \frac{W_{\psi_f} S(a,b)}{\left\| S \mathbb{1}_{[b,a+b]} \right\| \left\| \psi_f \right\|}$$

Three steps procedure

We proved that for any fixed scale $a > 0$,

Local maxima of $W^2_{\psi_f} S(a, b)$ as a function of b
are the ONLY possible values for which
the (squared) similarity of S to $f(\frac{x-b}{a})$ is locally maximum.

alert: Point (a, b), for $a > 0$ such that $W^2_{\psi_f} S(a, b)$ is locally maximum in b.

Selection rules

Two kinds of rules:

- **(P-B)** Pattern-based rules: designed only from the pattern. Allow to check the similarity between the pattern and the fragment of the signal.

Example

$$|\cos(a^*, b^*)| \le \left(1 + \frac{\left|W_{\psi_f} S(a,b) - \Gamma^* \cdot W_{\psi_f} S(a^*, b^*)\right|}{\left|W_{\psi_f} S(a^*, b^*)\right| \sqrt{1 - {\Gamma^*}^2}}\right)^{-1}$$

where $\Gamma^* = \Gamma(\frac{a}{a^*}, \frac{b-b^*}{a^*}) = \left\langle \frac{a}{a^*} \psi_f \left(\frac{x - \frac{b-b^*}{a^*}}{a/a^*}\right),\ \psi_f(x) \right\rangle$

Selection rules

Two kinds of rules:

- **(P-B)** Pattern-based rules: designed only from the pattern. Allow to check the similarity between the pattern and the fragment of the signal.
- **(A-B)** Application-based rules: based on the specifications of the problem. Allow to check the relevance with respect to the problem.

Three steps procedure

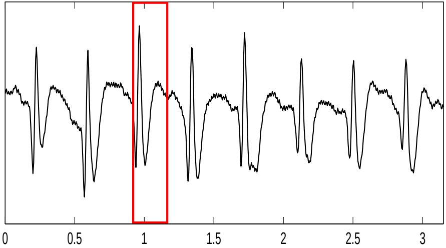
We propose to follow this three steps procedure:

1. Given the motif f to detect, create the pattern-adapted wavelet $\psi_f(x)$.
2. Detect all the alerts on the signal.
3. Discard the false alerts by using selection rules.

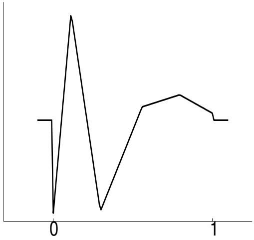
Three steps procedure
An example

Adapting the wavelet.

Figure: An electroencephalogram (EEG)



The pattern

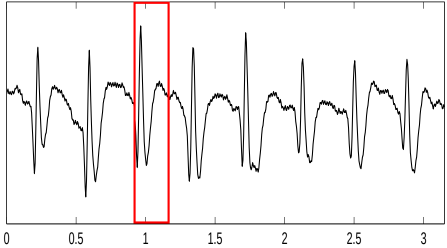


Three steps procedure

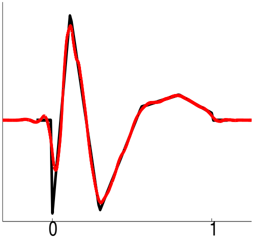
An example

Adapting the wavelet.

Figure: An electroencephalogram (EEG)



The adapted wavelet



Three steps procedure
An example

Detecting all the similarity alerts.

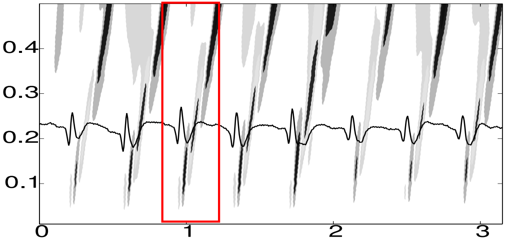
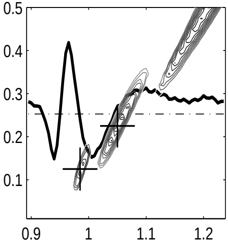


Figure: The signal superimposed to the wavelet energy of an EEG. Black regions means higher wavelet energy



Zooming...

Three steps procedure
An example

Discarding false alerts.

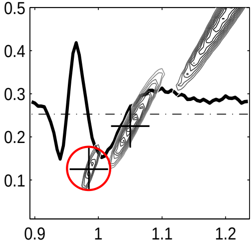


Figure: Zoom of the wavelet energy graphic of the signal

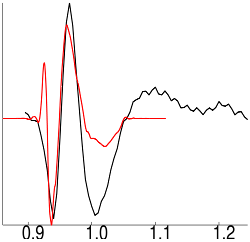


Figure: A false alert

Three steps procedure
An example

Discarding false alerts.

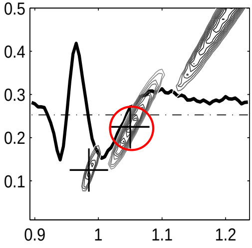


Figure: Zoom of the wavelet energy graphic of the signal

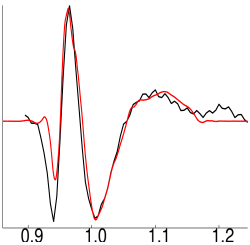


Figure: A true alert

Outline

1. A methodology for pattern detection problems
   - The problems
   - Why with wavelets?
   - Three steps procedure
2. Adapted wavelets construction
   - Our approach
   - Properties
3. Real life applications
   - Spike detection
   - Evoked Potentials analysis
   - Defect detection in overhead lines in trains

Adapting wavelets

There exists many methods to construct pattern-adapted wavelets:

- Solving the (bi)orthogonal relations.
  Chapa and Rao (2000).
- Projection based methods.
  Unser and Aldroubi (1993), Abry and Aldroubi (1995).
- Lifting scheme.
  Sweldens (1996).

Lifting

It is based on the definition of Multiresolution Analysis (MRA). Let φ and ψ be a scaling function and a wavelet. Let f be the pattern.

The new primal wavelet ψ_l is computed by:

$$\psi_l(x) = \psi(x) + \sum_i l_i \varphi(x-i)$$

where $\sum_i l_i = 0$.
The adapted wavelet would be ψ_f such that

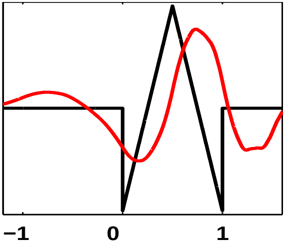
$$\|\psi_f - f\| = \min_{\sum_i l_i = 0} \|\psi_l - f\|$$

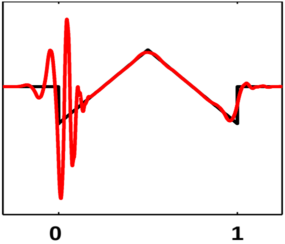
Our approach

- To approximate a dilated version of f:

$$f_\rho(x) = \frac{1}{\sqrt{\rho}} f(x/\rho), \text{ for some } \rho > 0.$$

It takes profit of the good approximation properties of the scaling functions





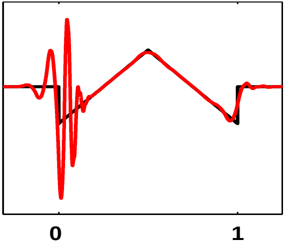
Our approach

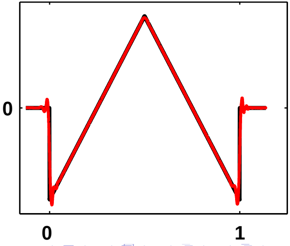
- To approximate a dilated version of f:

$$f_\rho(x) = \frac{1}{\sqrt{\rho}} f(x/\rho), \text{ for some } \rho > 0.$$

- To reduce the influence of the original wavelet.

$$\psi_f(x) = c\psi(x - k) + \sum_i l_i \varphi(x - i)$$





Our approach

- To approximate a dilated version of f:

$$f_\rho(x) = \frac{1}{\sqrt{\rho}} f(x/\rho), \text{ for some } \rho > 0.$$

- To reduce the influence of the original wavelet.

$$\psi_f(x) = c\psi(x - k) + \sum_i l_i \varphi(x - i)$$

For orthogonal MRAs:

$$\begin{aligned} k^* &= \arg\max_k \langle f,\ \psi(x-k)\rangle \\ c^* &= \langle f,\ \psi(x-k^*)\rangle \\ l^*[n] &= \langle f,\ \varphi(x-n)\rangle \end{aligned}$$

The new filters

The associated filters $(u,\ v^N,\ \mathring{u}^N,\ \mathring{v}^N)$ are computed with

$$\begin{aligned}
v^N &= c \cdot \delta_{2k} \star v \ + \ [l]_{\uparrow_2} \star u \\
\mathring{u}^N &= u \ - \ \frac{1}{c}\delta_{2k} \star [l]^{\vee}_{\uparrow_2} \star v \\
\mathring{v}^N &= \frac{1}{c}\delta_{2k} \star \mathring{v} \ .
\end{aligned}$$

Operators:

$$\begin{aligned}
\delta_k[j] &= \delta(k-j) && \text{(Kronecker delta)} \\
u \star v[n] &= \textstyle\sum_j u[j]v[n-j] && \text{(Convolution)} \\
[u]_{\uparrow_2}[n] &= u[2n] && \text{(Upsampling)} \\
u^{\vee}[n] &= u[-n] && \text{(Reflection)}
\end{aligned}$$

Properties

Pros

- The filters are finite, their size are relatively small, and they are easy to compute.

Properties
Pros

- The filters are finite, their size are relatively small, and they are easy to compute.
- Continuity and differentiability properties of the constructed wavelet are ensured by the original scaling function's properties.

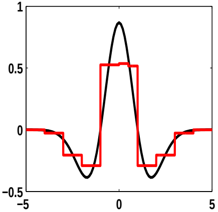


Figure: with the Haar basis

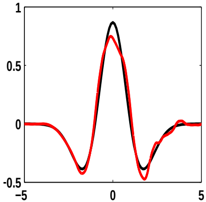


Figure: with the $Db5$ basis

Properties
Pros

- The filters are finite, their size are relatively small, and they are easy to compute.
- Continuity and differentiability properties of the constructed wavelet are ensured by the original scaling function's properties.
- The new wavelet may be as close (in terms of the MSE) to the pattern as desired (choosing ρ).

Mexican hat

$\psi_f(x) = c\psi(x-k) - \sum_i l_i \varphi(x-i)$

$k = 0$, $c = 0.01$

Wavelet(ρ)	MSE
Haar(1)	0.431079
Haar(8)	0.057565
Haar(32)	0.017394
$Db5$(1)	0.169692
$Db5$(8)	0.010001
$Db5$(32)	0.010000

Properties
Pros

- The filters are finite, their size are relatively small, and they are easy to compute.
- Continuity and differentiability properties of the constructed wavelet are ensured by the original scaling function's properties.
- The new wavelet may be as close (in terms of the MSE) to the pattern as desired (choosing ρ).
- This construction method is proved to be stable for small variations in the pattern.

$$\begin{array}{c} f_\varepsilon(x) = f(x) + \varepsilon g(x), \\ \|g\| = 1 \end{array} \Longrightarrow \|\psi_{f_\varepsilon} - \psi_f\|^2 \leq \varepsilon^2 C$$

Properties

Pros

- The filters are finite, their size are relatively small, and they are easy to compute.
- Continuity and differentiability properties of the constructed wavelet are ensured by the original scaling function's properties.
- The new wavelet may be as close (in terms of the MSE) to the pattern as desired (choosing ρ).
- This construction method is proved to be stable for small variations in the pattern.
- It is possible to obtain an arbitrary number of vanishing moments (linear constraints on l).

Properties

Pros

- The filters are finite, their size are relatively small, and they are easy to compute.
- Continuity and differentiability properties of the constructed wavelet are ensured by the original scaling function's properties.
- The new wavelet may be as close (in terms of the MSE) to the pattern as desired (choosing ρ).
- This construction method is proved to be stable for small variations in the pattern.
- It is possible to obtain an arbitrary number of vanishing moments (linear constraints on l).
- Easy to extend to other domains:
  - 2D separable wavelets,
  - Multiwavelets.

Properties

Cons

- The dual filters may have large coefficients.

$$\begin{aligned} \mathring{u}^N &= u - \frac{1}{c}\delta_{2k} \star [l]^{\vee}_{\uparrow_2} \star v \\ \mathring{v}^N &= \frac{1}{c}\delta_{2k} \star \mathring{v}\ . \end{aligned}$$

Properties

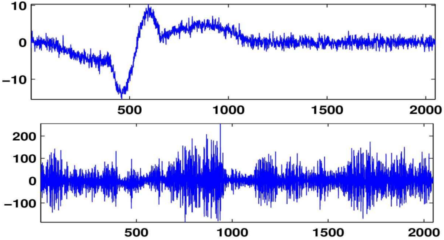
Cons

- The dual filters may have large coefficients.
- The dual scaling function and wavelet do not exist in L^2 in general.

Properties

Cons

- The dual filters may have large coefficients.
- The dual scaling function and wavelet do not exist in L^2 in general.
- The reconstruction step of the fast algorithm becomes unstable when the coefficients are modified.



with the true wavelets coefficients

with modified coefficients

Outline

Spike Detection

Electroencephalogram (EEG) is an important clinical tool for diagnosing, monitoring and managing neurological disorders related to epilepsy.

Spikes correspond to tiny epileptic discharges which last for a fraction of a second.
Automatic spike detection helps to make quantitative descriptions of spike density, topology and morphology what could help to determine patient syndrome and surgical outcome (Scott (2002)).

The Problem

Three characteristics of spikes given by Gloor (1975):

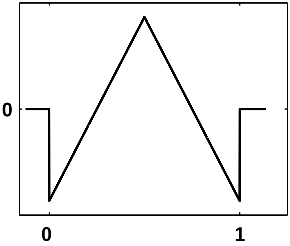
- **Shape**: restricted triangular transients having an amplitude of, at least, twice of the preceding 5s of background activity.
- **Duration**: less than $200ms$.
- **Electric field**: defined by involvement of a second adjacent electrode.

The Detection Procedure

Constructing the pattern-adapted wavelet

We take a triangular function as a very simple pattern.

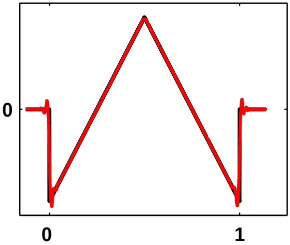
the motif



The Detection Procedure
Constructing the pattern-adapted wavelet

We take a triangular function as a very simple pattern.

the motif



the adapted wavelet

The Detection Procedure
Detecting the alerts

- Detection of local maxima in b of the wavelet energy $W^2_{\psi_f} S(a,b)$ for every scale $a > 0$.
- We chose the optimum scale a^* for any b^*, as those scales where
$$\frac{amplitude}{scale} = \left| \frac{1}{\sqrt{a^3}} W_{\psi_f}(a,b) \right|$$
is locally maximum as a two-variables function.

Pattern-based selection rules:

- Thresholds on the similarity:
$$\cos(\psi_f(\frac{x-b^*}{a^*}), S) = \frac{W_{\psi_f} S(a,b)}{\left\| S \mathbb{1}_{[b,a+b]} \right\|}$$
- Thresholds on upper bounds of $\cos(\psi_f(\frac{x-b}{a}), S)$

The Detection Procedure

Detecting the alerts

- Detection of local maxima in b of the wavelet energy $W^2_{\psi_f} S(a,b)$ for every scale $a > 0$.
- We chose the optimum scale a^* for any b^*, as those scales where

$$\frac{amplitude}{scale} = \left| \frac{1}{\sqrt{a^3}} W_{\psi_f}(a,b) \right|$$

is locally maximum as a two-variables function.

Pattern-based selection rules:

- Thresholds on the similarity:

$$\cos(\psi_f(\frac{x-b^*}{a^*}), S) = \frac{W_{\psi_f} S(a,b)}{\left\| S \mathbb{1}_{[b,a+b]} \right\|}$$

- Thresholds on upper bounds of $\cos(\psi_f(\frac{x-b}{a}), S)$

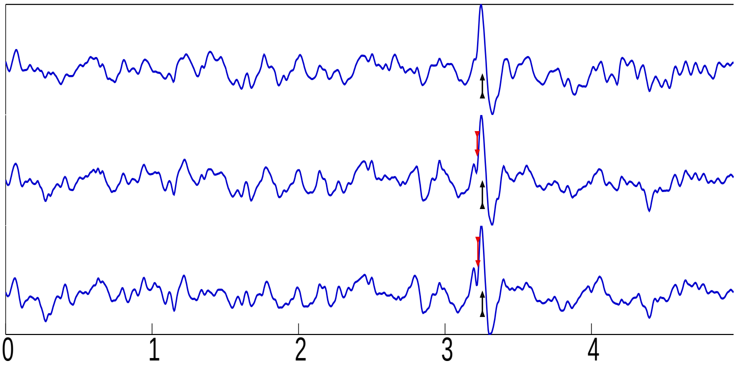
The Detection Procedure

Problem-dependent rules

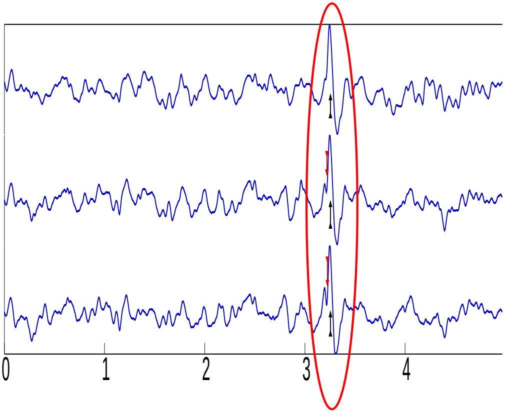
Taken from the characteristics of spikes:

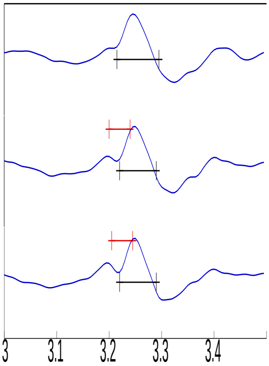
- **Duration:** $a \leq 200ms$. Discontinuities will cause false alerts for small a, so we keep only $a \geq 10ms$.
- **Amplitude:** $\frac{1}{a}W^2(a,b) > 4\sigma_{bg}^2$ where $\sigma_{bg}^2(b)$ is the average background amplitude computed as the average of the local maxima of $\frac{1}{a}W^2(a,b)$ for $5s$ of signal before the alert-time b.
- **Not a random event:** $W^2(a,b) > \tau\sigma(b) + \mu(b)$, i.e. the wavelet energy must be larger than a multiple of the standard deviation σ of the background amplitudes plus its mean μ
- **Must cause a field:** At the end, it is checked if there exist a field (a similarity-alert) in an adjacent channel.

Some results



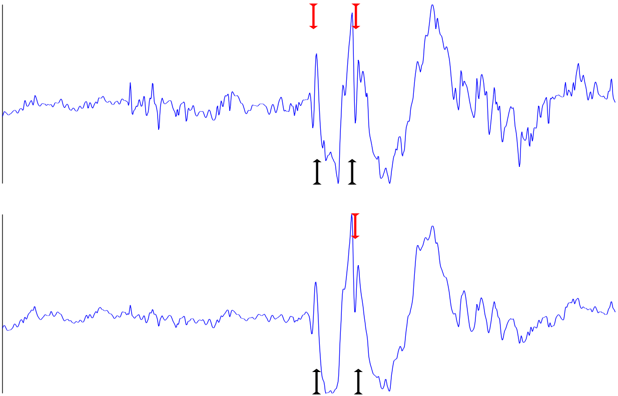
Some results





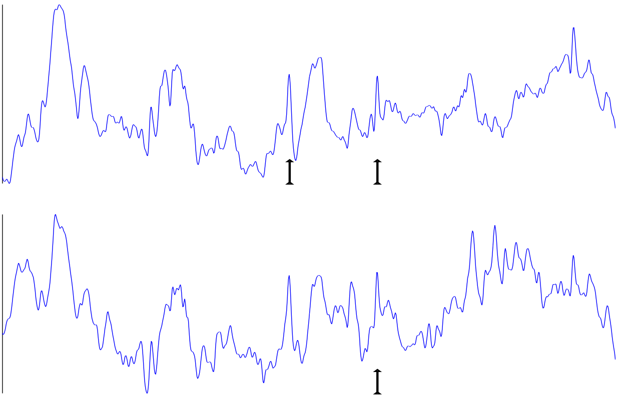
Results (cont.)

Other examples



Results (cont.)

Other examples



Outline

Evoked Potential

Each stimulus cause an event related potential in the EEG.

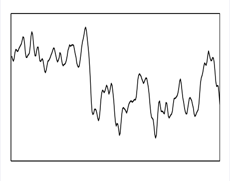
Stimulus (visual, auditive, etc.)

Evoked Potential

Each stimulus cause an event related potential in the EEG.

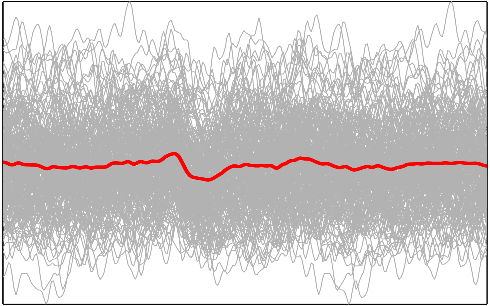
Stimulus (visual, auditive, etc.)

Reaction in the EEG (Evoked Potential)



Evoked Potential

Too many trials are needed to obtain the averaged evoked potential (EP).



Evoked Potential

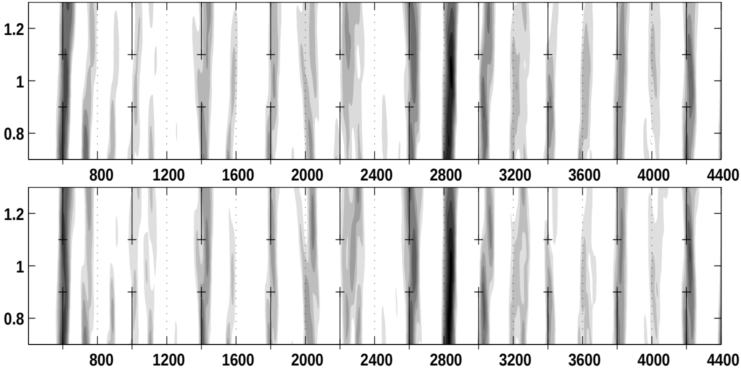
Reducing the number of trials

Table: Some descriptive statistics of the correlation between subset-adapted and whole set-adapted wavelets (217 trials).

size	mean	std	min	5%	10%	25%	50%
5%	0.6240	0.1440	0.0532	0.3563	0.4221	0.5309	0.6404
10%	0.7632	0.1018	0.3460	0.5691	0.6189	0.7027	0.7781
15%	0.8281	0.0754	0.4772	0.6785	0.7195	0.7815	0.8387
20%	0.8722	0.0593	0.5838	0.7543	0.7848	0.8380	0.8819
25%	0.9010	0.0512	0.5096	0.8026	0.8281	0.8703	0.9104
30%	0.9223	0.0403	0.6862	0.8389	0.8612	0.9012	0.9262

Detection

The CWT with the whole set-adapted wavelet does not change significatively when using a subset-adapted wavelet of 30% trials.



Trial Rejection

Some trials are not good for the analysis.

The adapted wavelets may be used to detect bad trials even when it is difficult to find the potential in the trial.

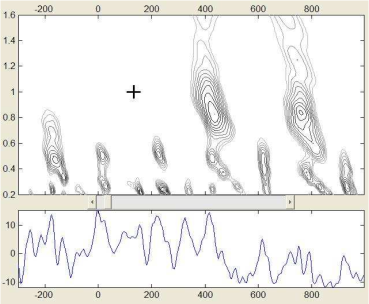
Possible rules:

A-B.1 There must exists an alert close to $(1,\ b_0)$, where b_0 is the expected latency of the potential.

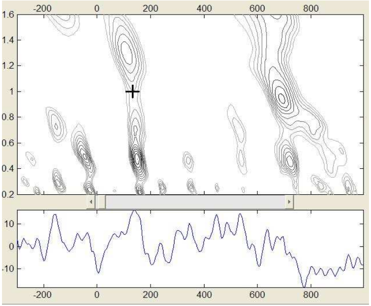
A-B.2 The amplitude can be compared with that of the average potential.
$$\frac{1}{\sqrt{a}} W_{\psi_f} S(a^*, b^*) \geq \tau_m.$$

Trial Rejection

Example of rejection.



Bad trial



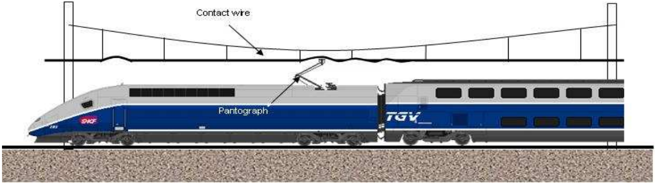
Good trial

Outline

Defect detection on catenary lines

The TGV

Defects may cause the pantograph to lose contact.



Automatic defect detection on the catenary from contact force measures.

Distinguish defects from singularities (road bridges, sections overlap).

Defect detection on catenary lines
The TGV

Defects may cause the pantograph to lose contact.

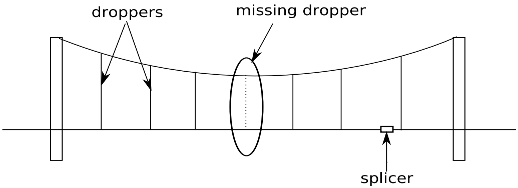
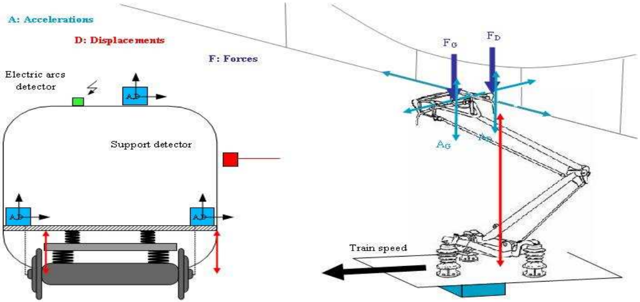


Figure: Schema of a catenary with two defects

Automatic defect detection on the catenary from contact force measures.
Distinguish defects from singularities (road bridges, sections overlap).

The problem

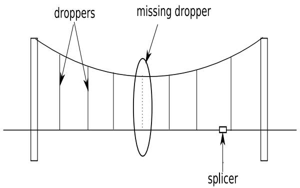
The data: contact force signal

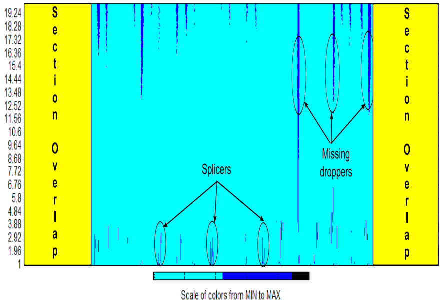


- Real data
- Simulated data

Two points of view

- Frequencies.
  Search large values of the wavelet coefficients on some frequency bands, different for each defect (Massat *et al.* (2006)).



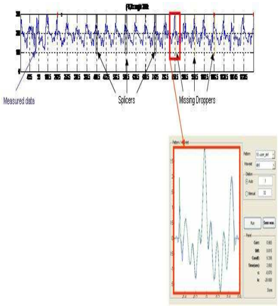


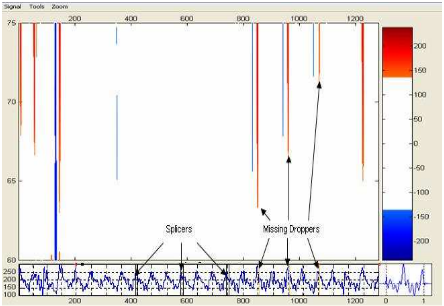
Two points of view

- Frequencies.
  Search large values of the wavelet coefficients on some frequency bands, different for each defect (Massat *et al.* (2006)).
- Shapes.
  Search into the signal some known patterns. Those patterns are chosen according to the signature of each defect.

Missing droppers
The choice of the pattern

With a real signal

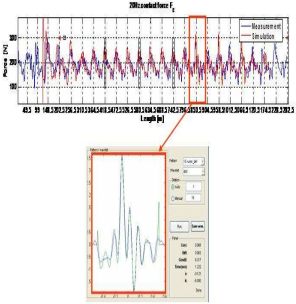


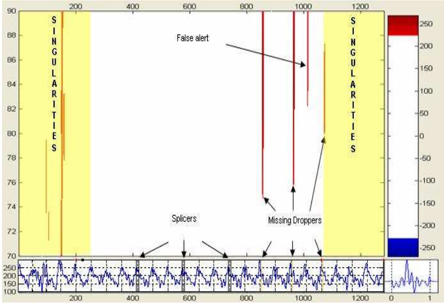


Missing droppers

The choice of the pattern

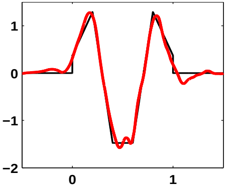
With a simulated signal





Missing droppers

The adapted wavelet



We used $\rho = 8$ and $Db5$ as initial MRA.

Missing droppers

Selection rules.

(P-B.1) Thresholds on the estimates of the similarity.

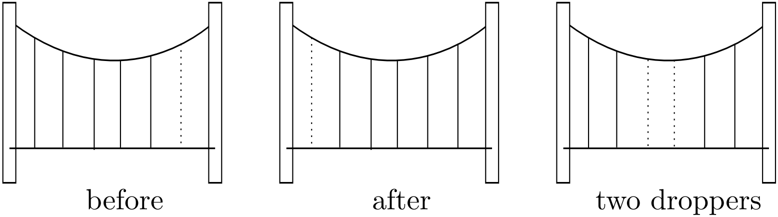
(A-B.1) Select only the alerts at some scales (duration of the signature).

(A-B.2) Thresholds on the wavelet coefficients.

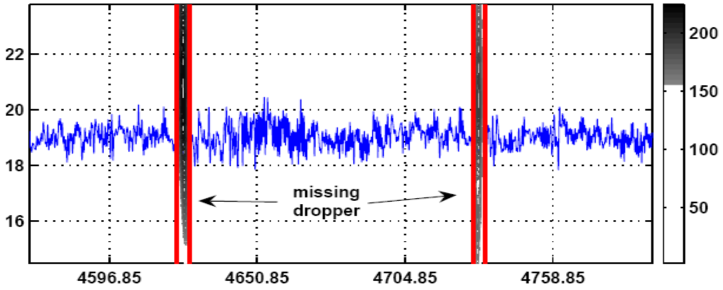
(A-B.3) Time of occurrence (to ignore those sections with known singularities).

Missing droppers

Three cases:



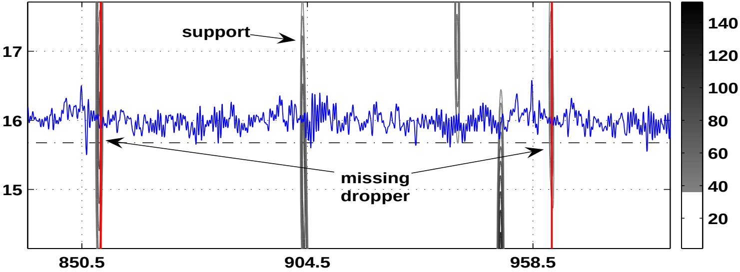
Missing droppers
Results



Two missing droppers

Missing droppers

Results



Missing dropper after the support.

Conclusions

- A method to construct pattern-adapted wavelets.

  Analysis of the extensions of the construction method.

- A Three steps procedure to solve pattern detection problems.

  - the pattern.
  - selection rules

- Three applications.

  Two papers are in preparation.

References

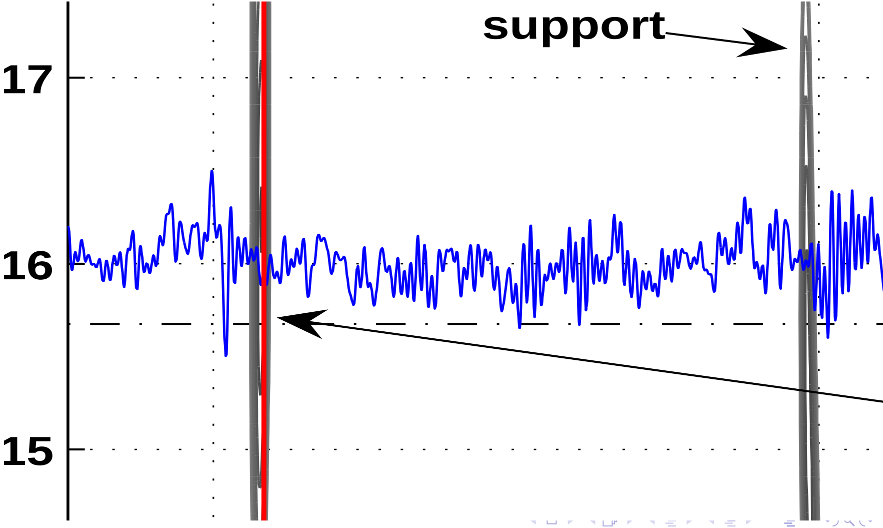
P. Abry.
Ondelettes et turbulences: Multirésolutions, algorithmes de décomposition, invariance d'échelles et signaux de pression.
Diderot Editeur, Paris, 1997.

S. Mallat.
A wavelet tour of signal processing.
Academic Press, 1998.

J. P. Marques de Sá.
Pattern Recognition.
Springer-Verlag, 2001.

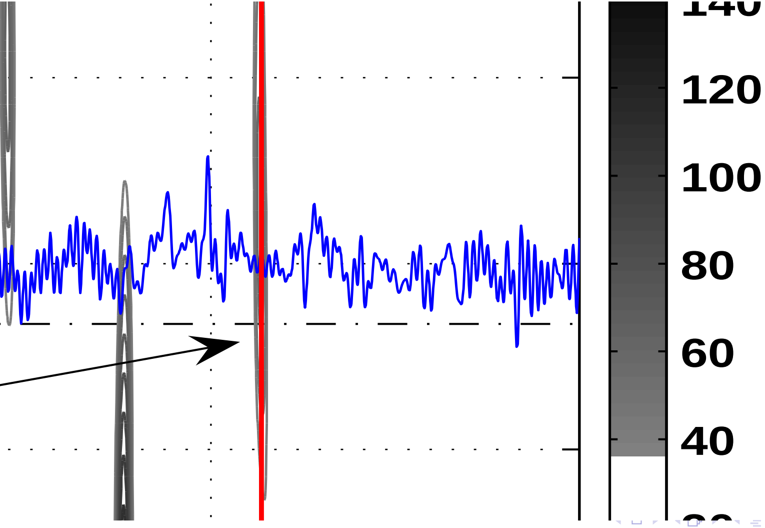
Missing droppers

Results



Missing droppers

Results



Missing droppers

Results