**A presentation for the 2008 Young Research Day,
Institute of Statistics, UCL.**

Using unbalanced Haar wavelets for classification of time series.

Catherine Timmermans
PhD student.

A project in collaboration with Véronique Delouille
Solar Influences Data analysis Center, Royal Observatory of Belgium
& Rainer von Sachs
Institute of Statistics, Catholic University of Louvain, Belgium

Using unbalanced Haar wavelets for classification of time series.

- ***Time series :***

Expansion in canonical basis, basic concepts, representation and motivation

- ***Unbalanced Haar wavelets:***

Data reduction, basis pursuit and representation.

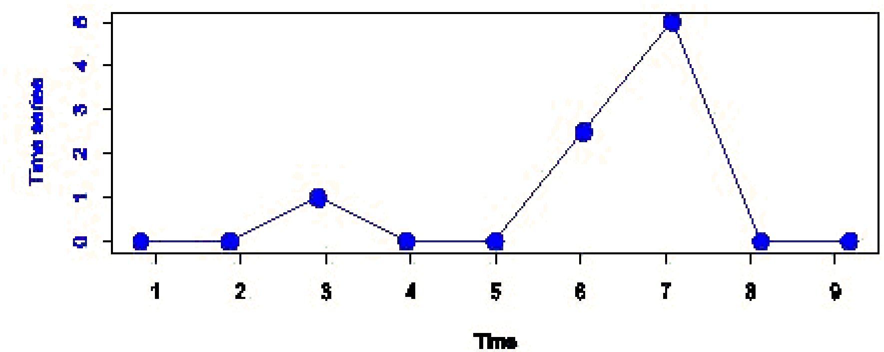
- ***Classification:***

Distances, simulations and perspectives…

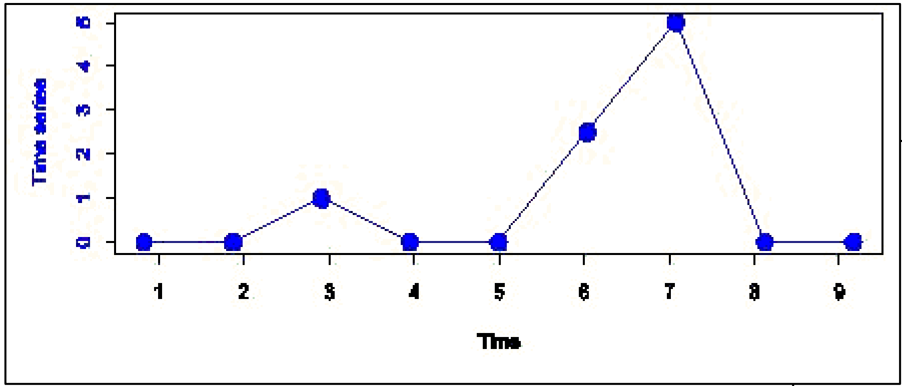
C. Timmermans

Describing a time series in the canonical basis.

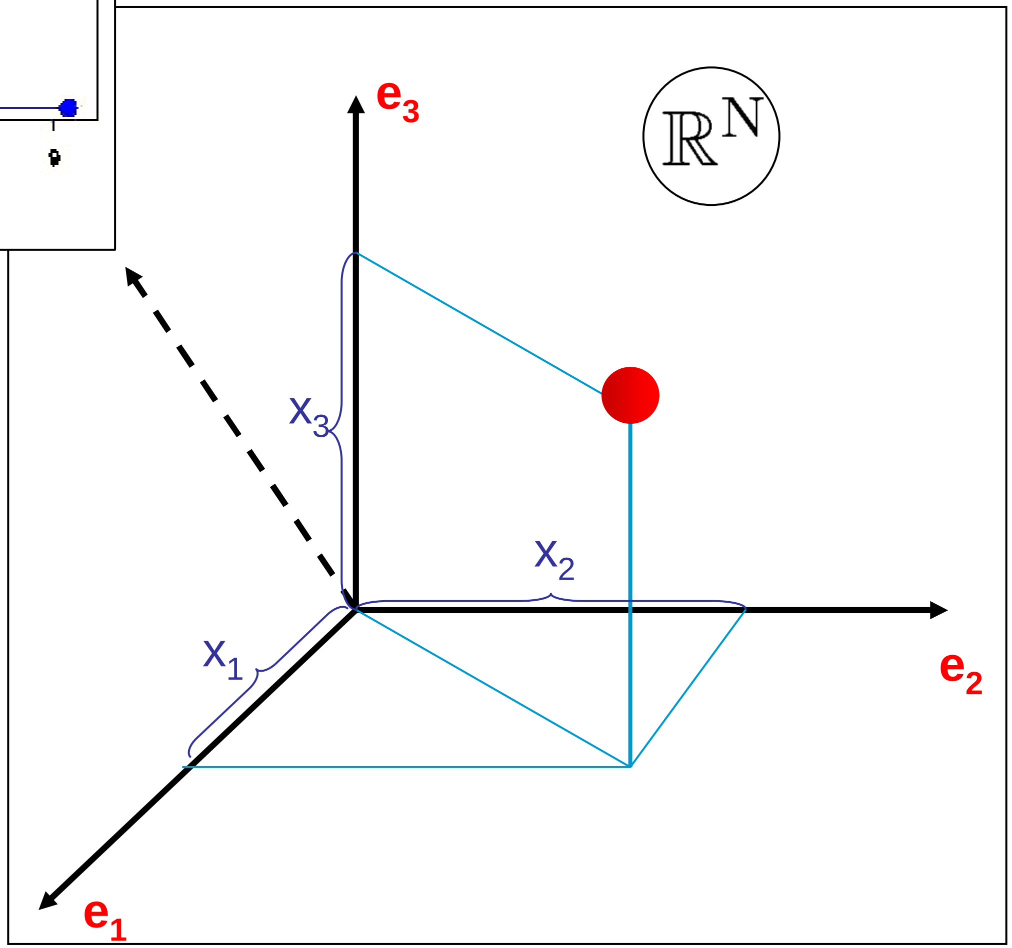
Here is a series.



Describing a time series in the canonical basis.
A series of N observations can be seen as a point in R^N.



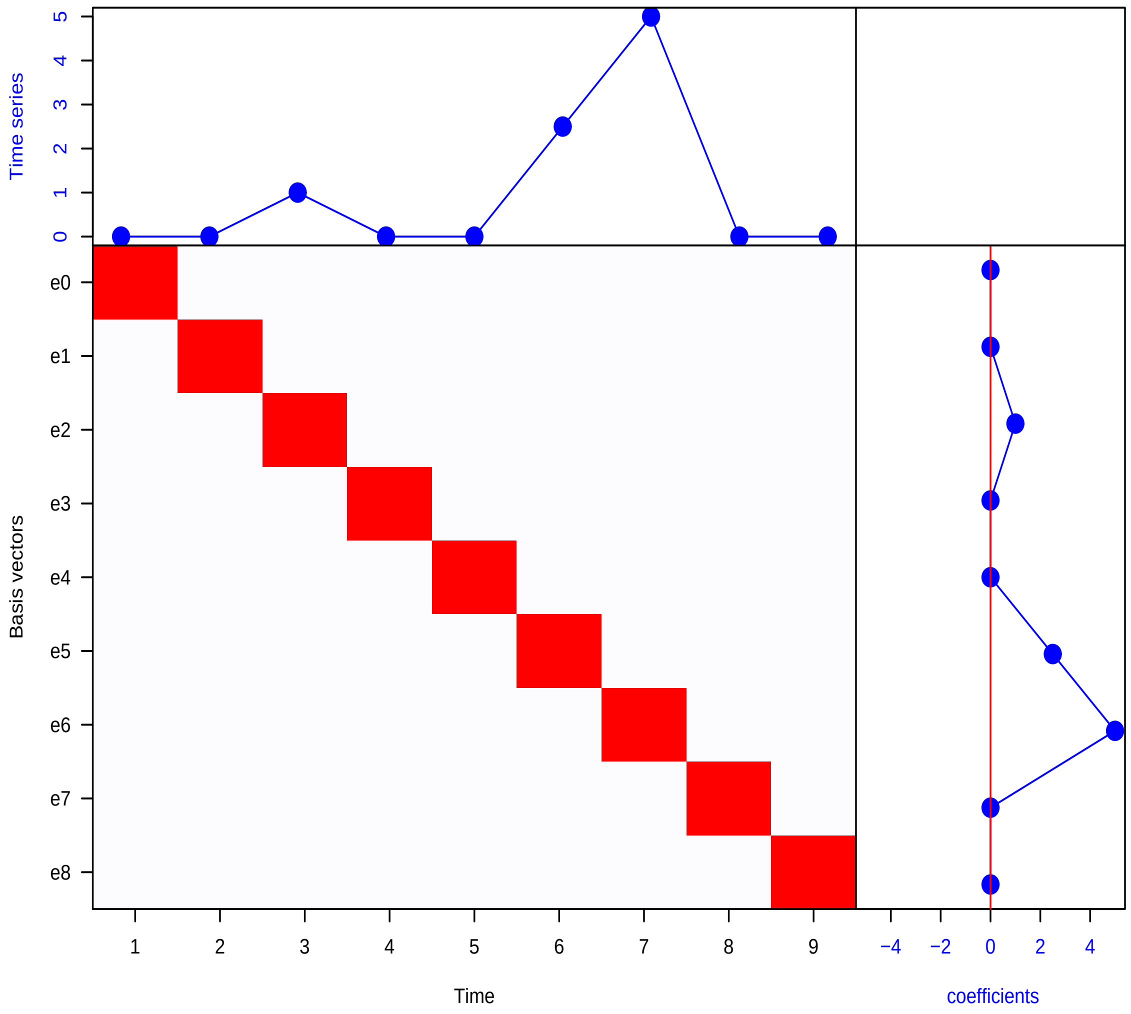
$$ {}_{e}\underline{\boldsymbol{x}} = \sum_{i=1}^{N} \langle\, \underline{\boldsymbol{x}} \,|\, \underline{\boldsymbol{e_i}} \,\rangle \cdot \underline{\boldsymbol{e_i}} = \sum_{i=1}^{N} x_i \underline{\boldsymbol{e_i}} $$



C. Timmermans

Describing a time series in the canonical basis.

Canonical basis expansion.



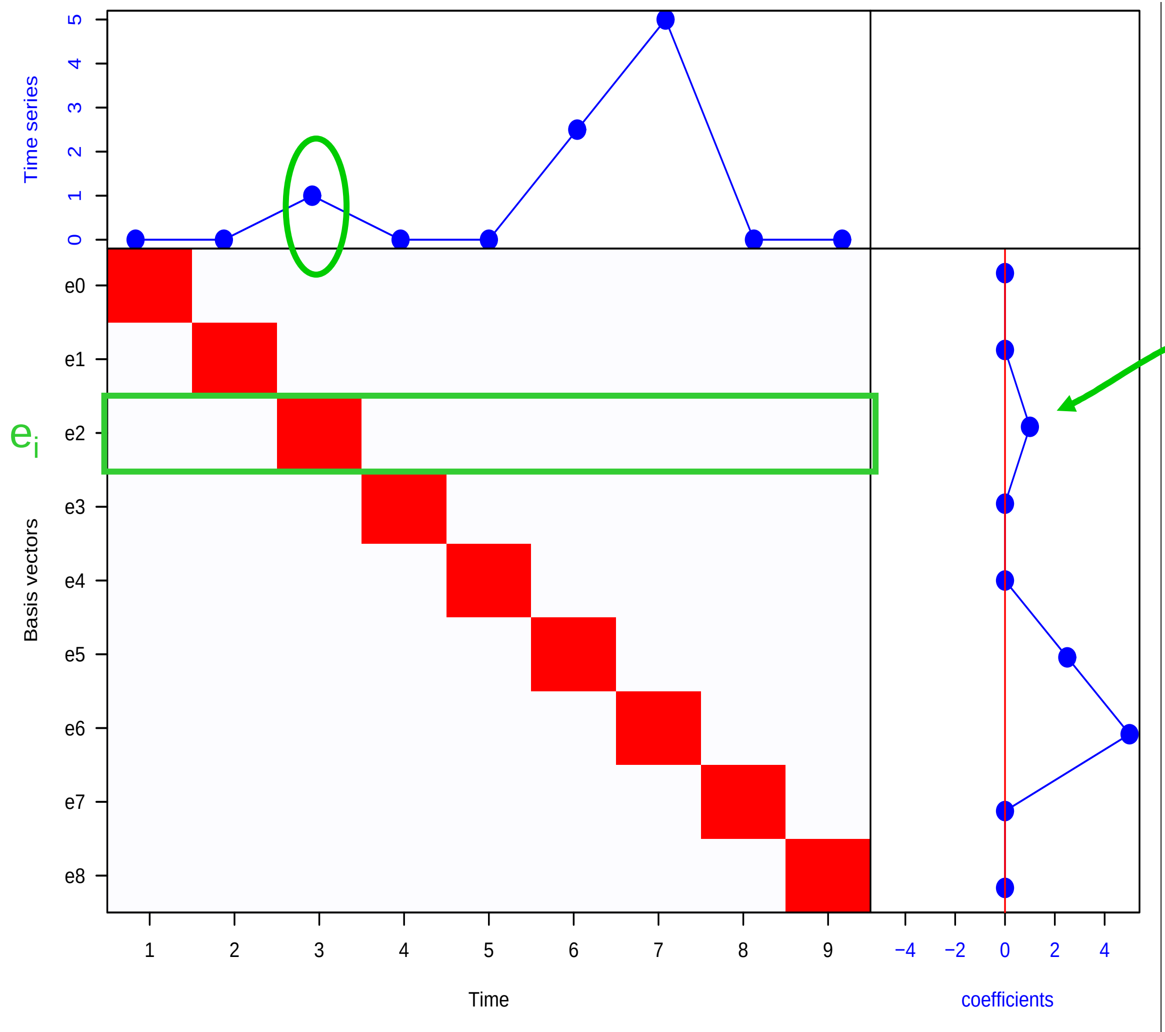
$$\underline{\boldsymbol{x}} = \sum_{i=1}^{N} x_i \underline{\boldsymbol{e_i}}$$

■ = 1; □ = 0.

Time series

C. Timmermans

Describing a time series in the canonical basis.
Canonical basis is an “instant-focusing” basis.



■ = 1; □ = 0.

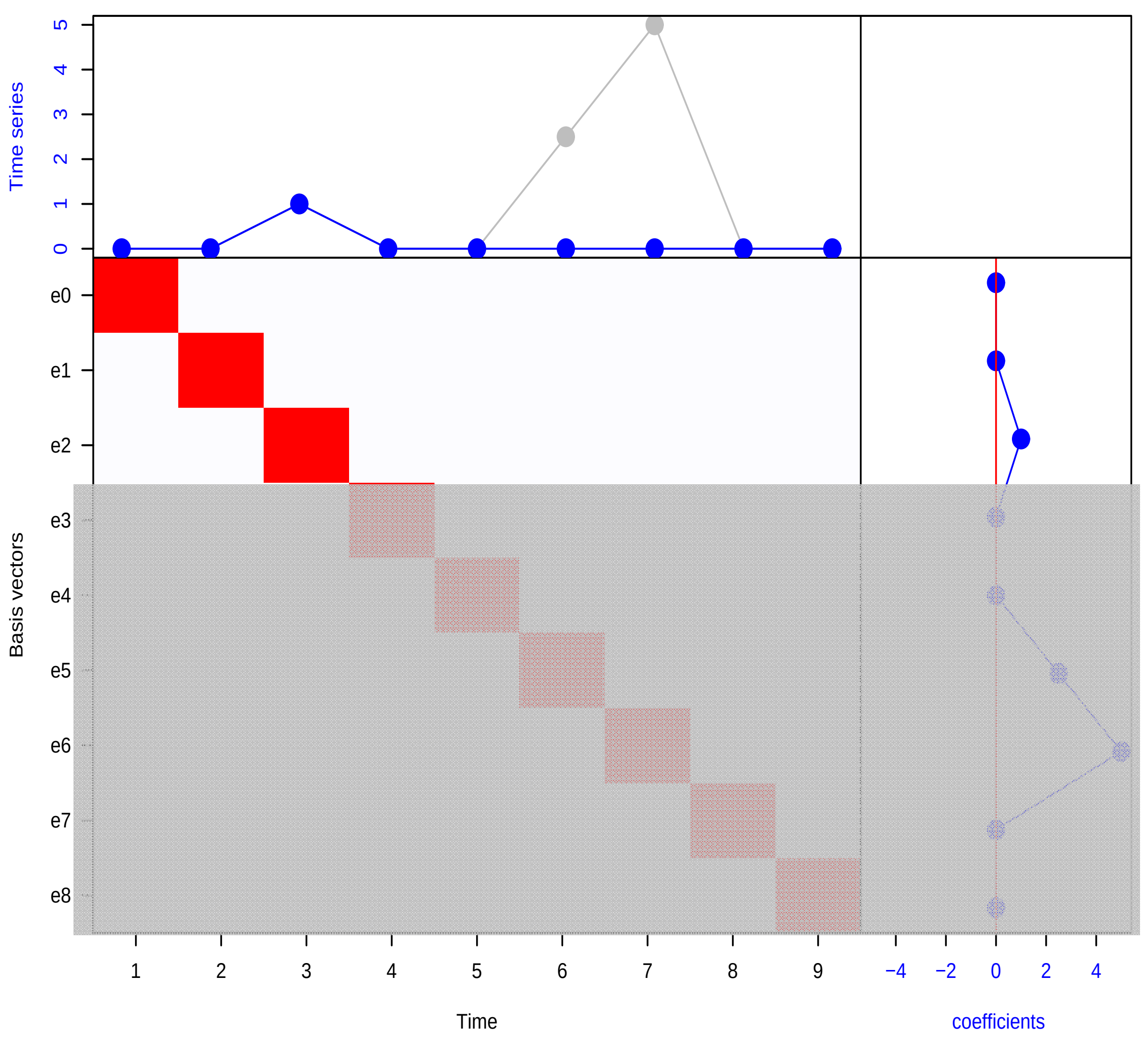
$$\underline{\boldsymbol{x}} = \sum_{i=1}^{N} x_i \underline{\boldsymbol{e_i}}$$

$$x_i = \langle\, \underline{\boldsymbol{x}} \mid \underline{\boldsymbol{e_i}} \,\rangle$$

The ordering
of the vectors
reflects the temporal order
of the measurements.

Projecting $\underline{\boldsymbol{x}}$
on the ith basis vector
allow us to focus
on the value of the series
at time t_i.

Describing a time series in the canonical basis.
Canonical basis has no “sum-up potential”.



= 1; = 0.

$$\mathrm{e}\underline{\boldsymbol{x}}^* = \sum_{i=1}^{N^*<N} x_i \underline{\boldsymbol{e_i}}$$

When looking only to a few axis, you have no idea of the global shape of the series.

No “sum-up” potential for the canonical basis expansion.

C. Timmermans

What should be a meaningful basis for summarization?
A motivation for unbalanced Haar wavelets.

Our goal is to define
a new orthonormal basis
such that

the first basis vectors carry
the main part of the information
describing the time sequence,

and subsequent basis vectors carry information
of decreasing importance.

C. Timmermans

What should be a meaningful basis for summarization?
A motivation for unbalanced Haar wavelets.

What do we consider as

- important features?

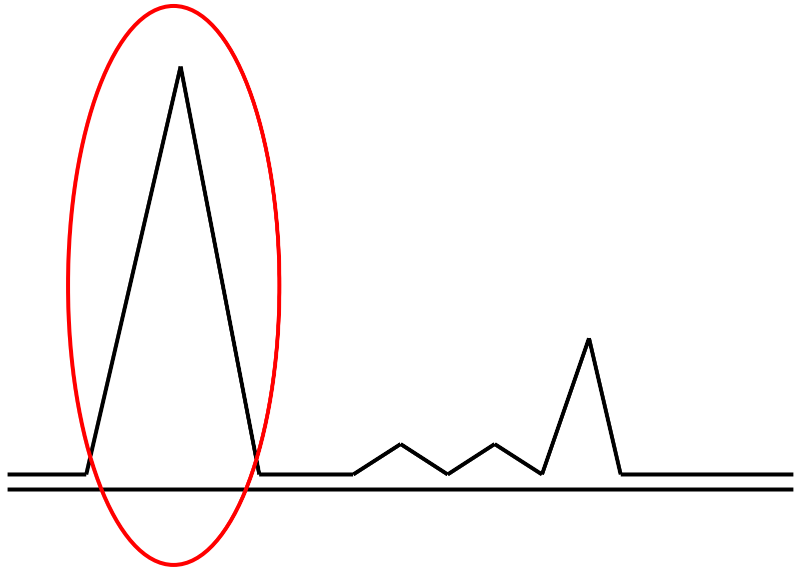
"high" or "extended" level changes.

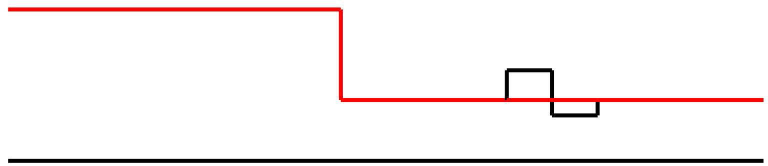
- secondary features?

"small" or "narrow" level changes.

- not a feature?

equal consecutive data.





What should be a meaningful basis for summarization?
A motivation for unbalanced Haar wavelets.

What do we consider as

➢ important features?

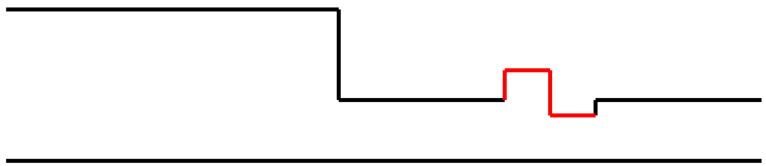
"high" or "extended" discontinuities.

➢ secondary features?

"small" or "narrow" discontinuities.

➢ not a feature?

equal consecutive data.

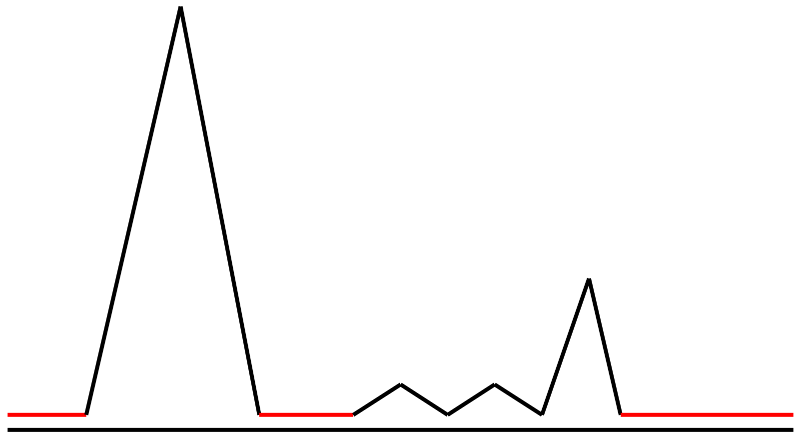


What should be a meaningful basis for summarization?
A motivation for unbalanced Haar wavelets.

What do we consider as

- important features?

“high” or “extended” discontinuities.

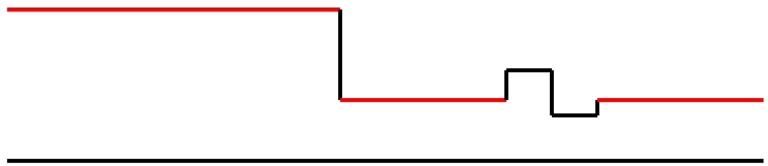


- secondary features?

“small” or “narrow” discontinuities.

- not a feature?

equal consecutive data.



C. Timmermans

Using unbalanced Haar wavelets for classification of time series.

➢ *Unbalanced Haar wavelets:*

Data reduction, basis definition, wavelets and representation.

Reference: Fryzlewicz,P., 2007, *Unbalanced Haar technique for nonparametric function estimation,* Journal of the American Statistical Association (Theory and Methods), 102, 1318—1327.

➢ *Classification:*

Distances, simulations and perspectives…

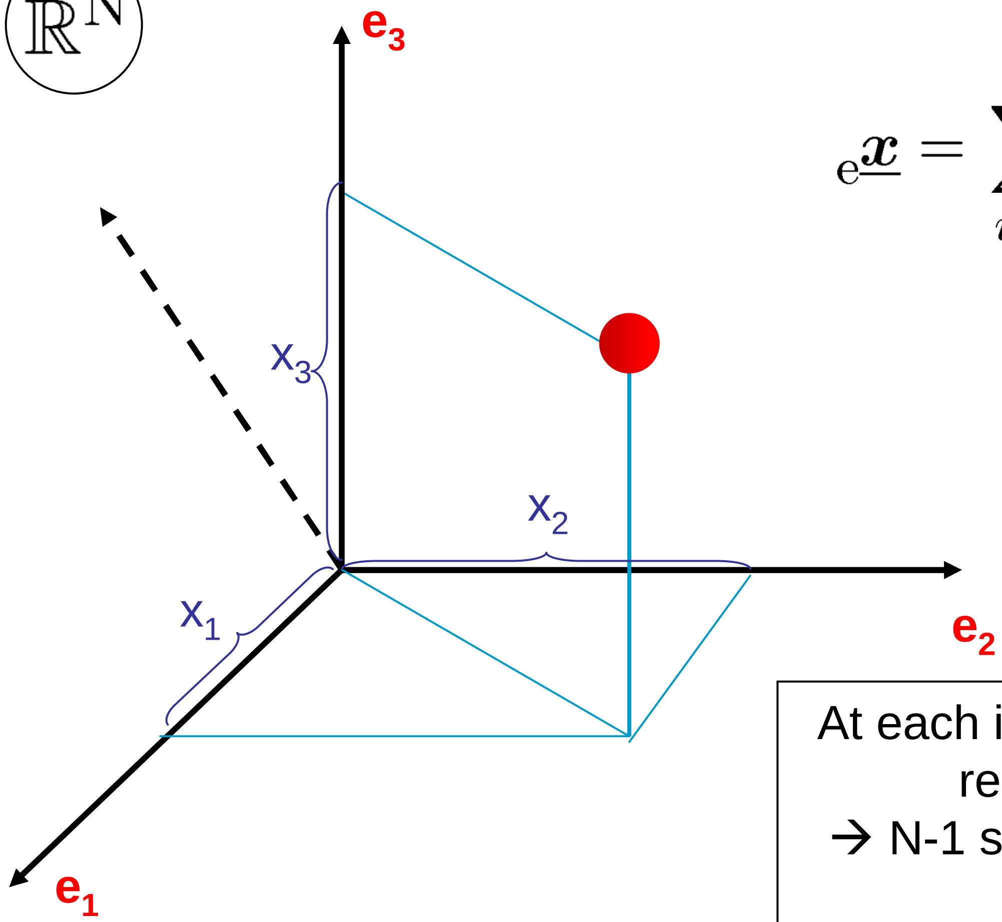
C. Timmermans

Bottom-up unbalanced Haar wavelets basis.

A data reduction algorithm as a first step to basis construction.

→ Step 0: Series in the canonical basis $\{e_i\}$ (i=1..N) of R^N

$\mathbb{R}^N$



$$e\underline{\boldsymbol{x}} = \sum_{i=1}^{N} x_i \underline{\boldsymbol{e_i}} = \sum_{i=1}^{N} \langle\, \underline{\boldsymbol{x}} \,|\, \underline{\boldsymbol{e_i}} \,\rangle \,.\, \underline{\boldsymbol{e_i}}$$

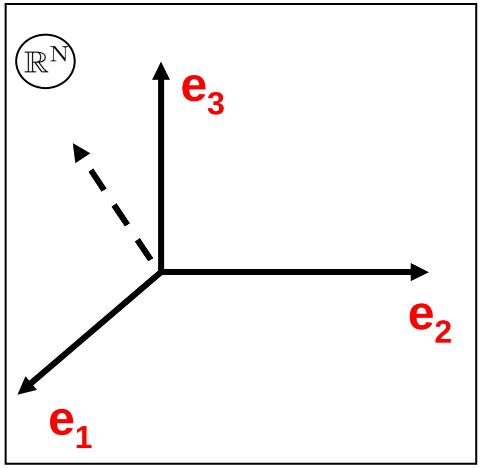
At each iteration of the algorithm, we will reduce the dimension by 1.
→ N-1 steps for a reduction to a unique scalar value.

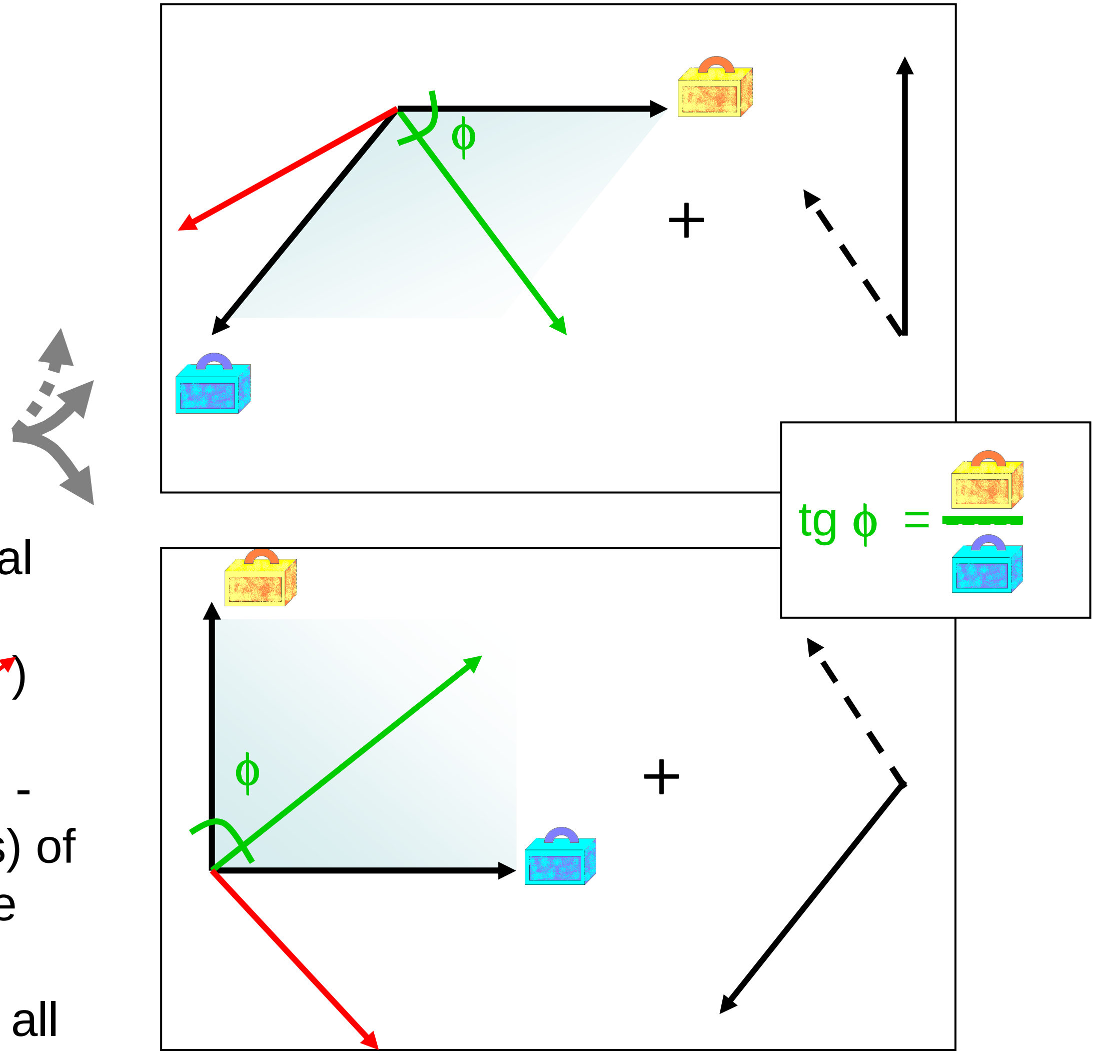
C. Timmermans

A data reduction algorithm as a first step to basis construction.

We want to summarize 1 “well chosen” plan by 1 “well chosen” axis.

→ Step 1: Define optimal summary-axis in each plan $\{e_i, e_{i+1}\}$.





.

In each plan, there is 1 optimal summary-axis: ( ), and 1 optimal detail-axis: ( )

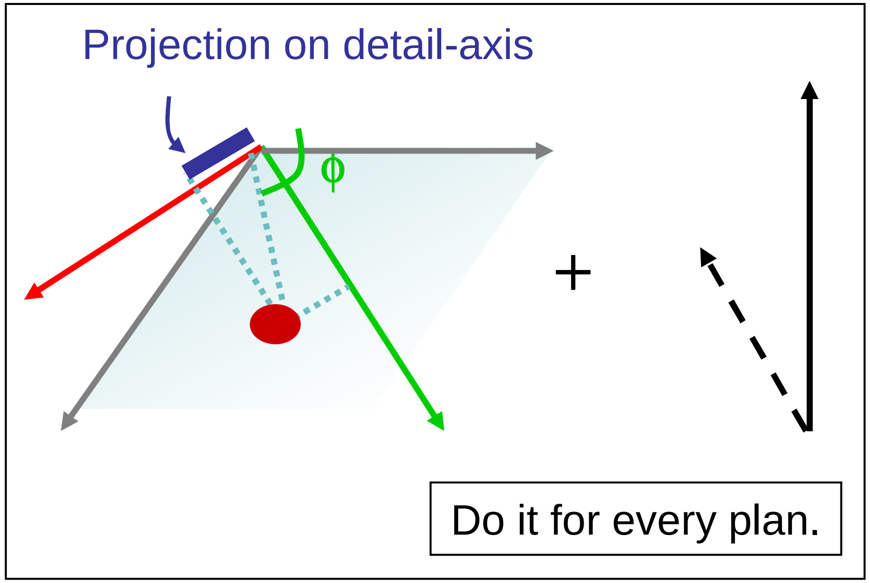
It depends on the importance - weights given “a priori” - (‘s) of the information carried by the original axes.
First, this importance is = 1 for all axes.

C. Timmermans

A data reduction algorithm as a first step to basis construction.

We want to summarize 1 “well chosen” plan by 1 “well chosen” axis.

→ Step 2: Select the plan you will actually summarized.



You should lose as less information as possible when summarizing 1 plan by 1 axis.

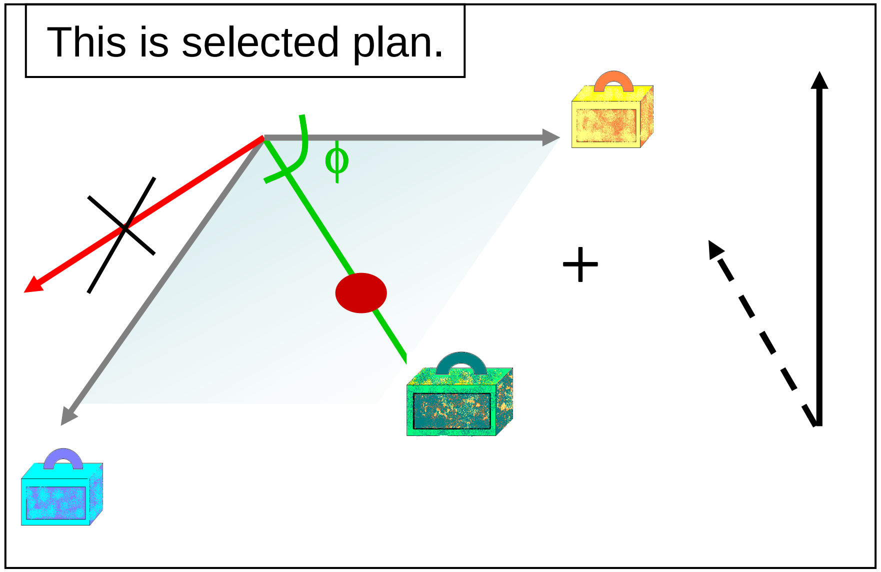
→ Compare projections of the data on all the possible detail-axis ( ) and select the plan were this projection is minimum.

A data reduction algorithm as a first step to basis construction.

We want to summarize 1 “well chosen” plan by 1 “well chosen” axis.

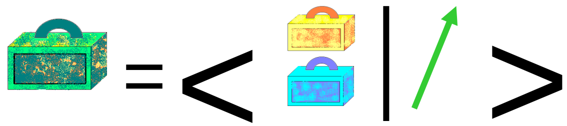
C. Timmermans

→ Step 3: Reduce data, adapt weights.



- In the selected plan, suppress the detail- axis. Keep only summary-axis –and the projection of the data on it.

- Compute weight for the summary-axis.



- Others axes remain unchanged.

→Step 4: Back to Step 1 - until reduced series is a scalar (or until you’re satisfied of your data reduction).

From data reduction to basis definition.

Our new basis vectors are the “by-products” of our data reduction.

Unbalanced Haar wavelets

C. Timmermans

Series in the canonical basis $\{e_i\}$ (i=1..N) of R^N

Iterate and go from R^N to R.

→ **Step 1:** **Define optimal summary-axis in each plan $\{e_i, e_i+1\}$.**

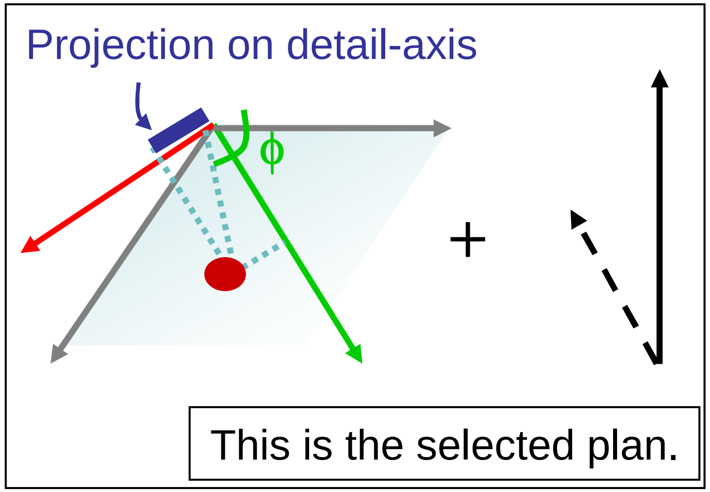
→ **Step 2:** **Select the plan you will actually summarized.**

→ **Step 3:** **Reduce data, adapt weights.**

→ **Step 4:** **Back to Step 1. Until reduced series is a scalar.**

→ **Final step:** **a scalar.**

Store selected detail-vector and the projection of the data on it. This shall be a new basis vector and its associate coefficient.

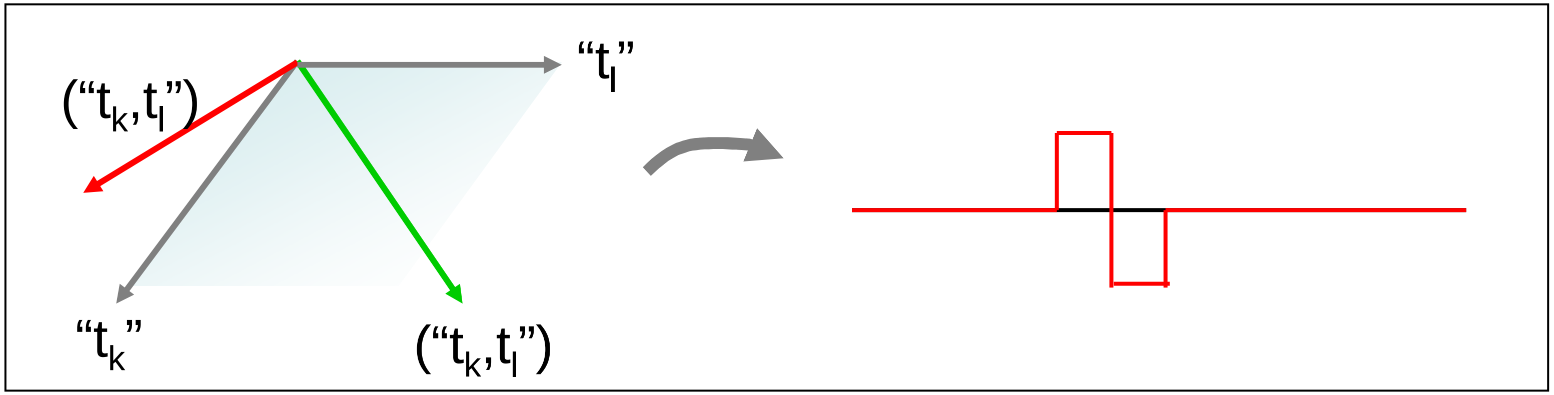


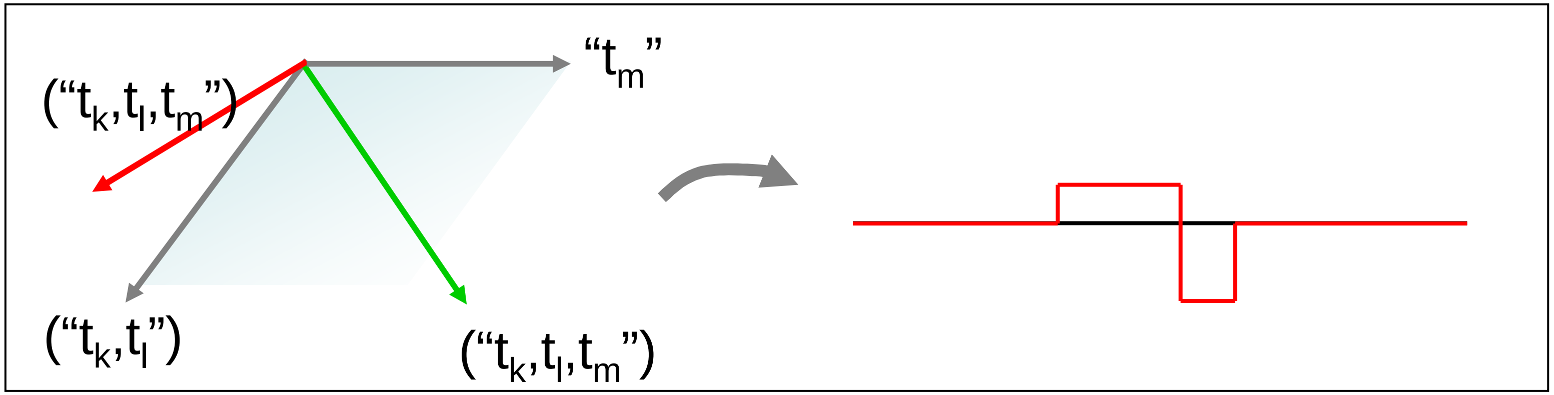
Iterate and define basis vectors, starting from the last one.

First basis vector is a constant and its associate coefficient is the remaining scalar.

Series in its unbalanced Haar wavelet basis of R^N.

Unbalanced Haar wavelet basis.
Interpretation, example and the wavelets.





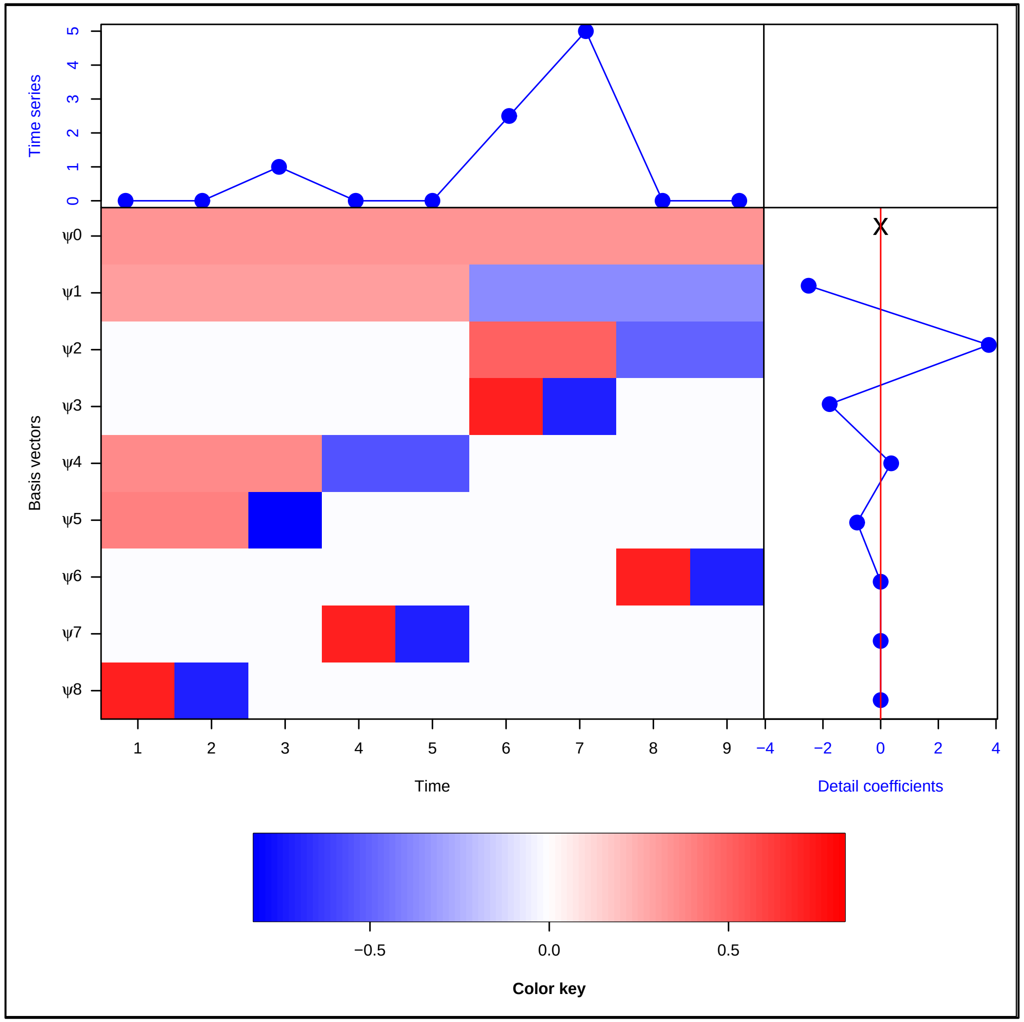
→Looks like Haar wavelets. One vector encode a difference between 2 adjacent values (or groups of values) .

But non-zero part is not symmetrical. This is the **"unbalanced" property**. It allows us to overcome the "dyadic restriction" of traditional wavelets.

Our basis is truly adaptative – and remains orthonormal.

C. Timmermans

Unbalanced Haar wavelet basis.
Interpretation, example and the wavelets.



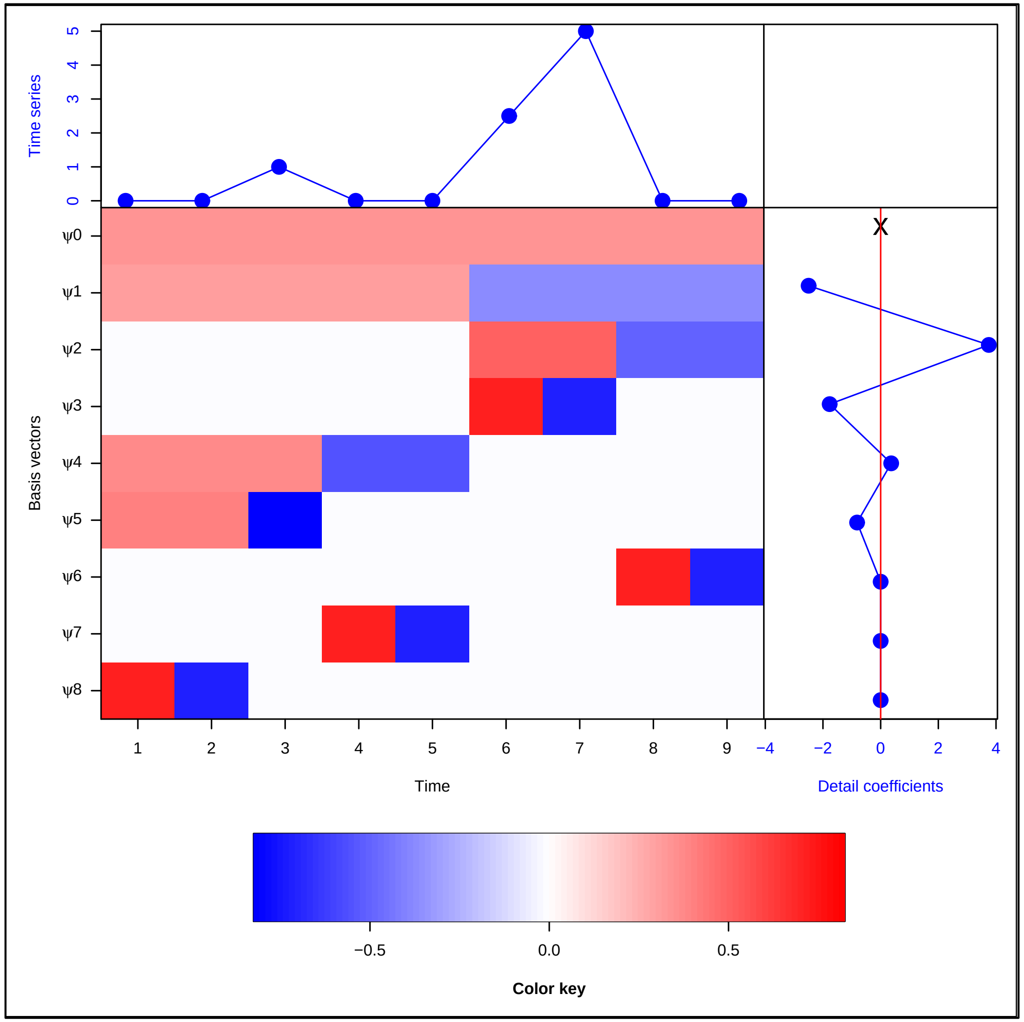
Support mean value.

Points to the large peak.

Points to the small peak.

No features.

Unbalanced Haar wavelet basis.
Interpretation, example and the wavelets.



This basis is individually suited to a particular time series.

We cannot expect our basis to be convenient to expand another time series.

Comparing series expanded in their associated unbalanced Haar wavelet basis is not straightforward.

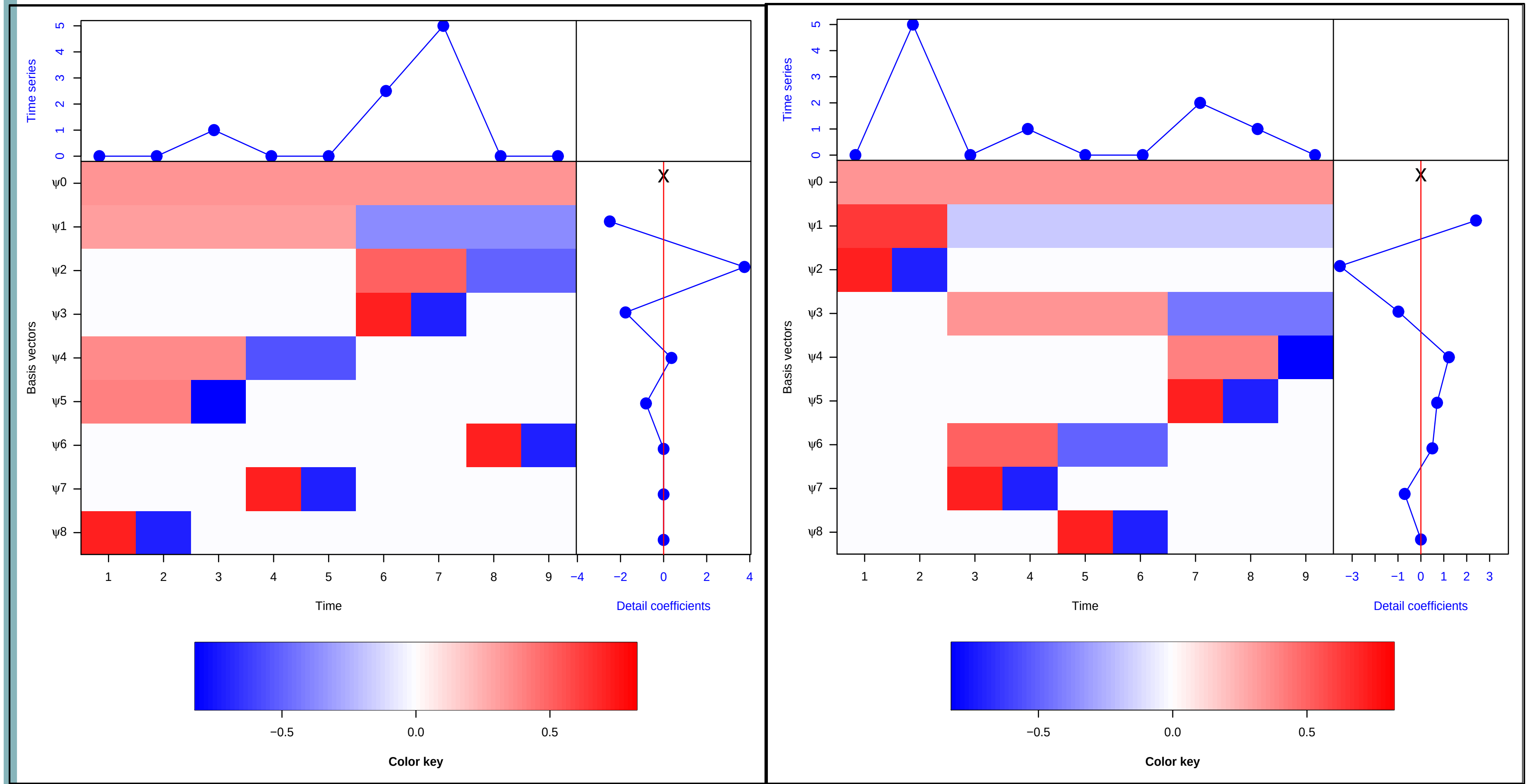
Using unbalanced Haar wavelets for classification of time series.

➢ *Classification:*

Distances, simulations and perspectives…

C. Timmermans

Defining a distance.
Ingredients.



<u>Notation</u>

- k: rank of the vector (from 0 to N-1)
- i^k: split index at rank k.
- d^k: detail coefficient at rank k.

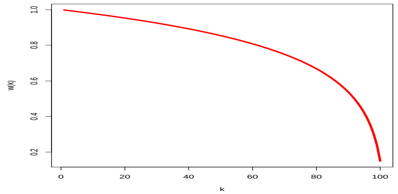
Defining a distance.
Ingredients.

At each given rank k:

- Differences in splits indexes $(i_1^k - i_2^k)$ should increase the distance.
- Differences in details coefficients $(d_1^k - d_2^k)$ should increase the distance.

Importance of differences in i's or d's should decrease with rank k.

→ Weights: $w(k) = \dfrac{\log(N+1-k)}{\log N + 1}$



$$d_1(S_1, S_2) = \sum_{k=1}^{N} w(k)\{n(k) f_1(i_1^k - i_2^k) + f_2(d_1^k - d_2^k)\}$$

$$f_1(.), f_2(.) = \mathrm{abs}(.) \quad \text{or} \quad (.)^2$$

$$n(k) \quad = \begin{cases} 0 & \text{if} \quad d_1^k . d_2^k = 0 \\ 1 & \text{else.} \end{cases}$$

Defining a distance.
Ingredients.

A cross-term i's / d's can be included:

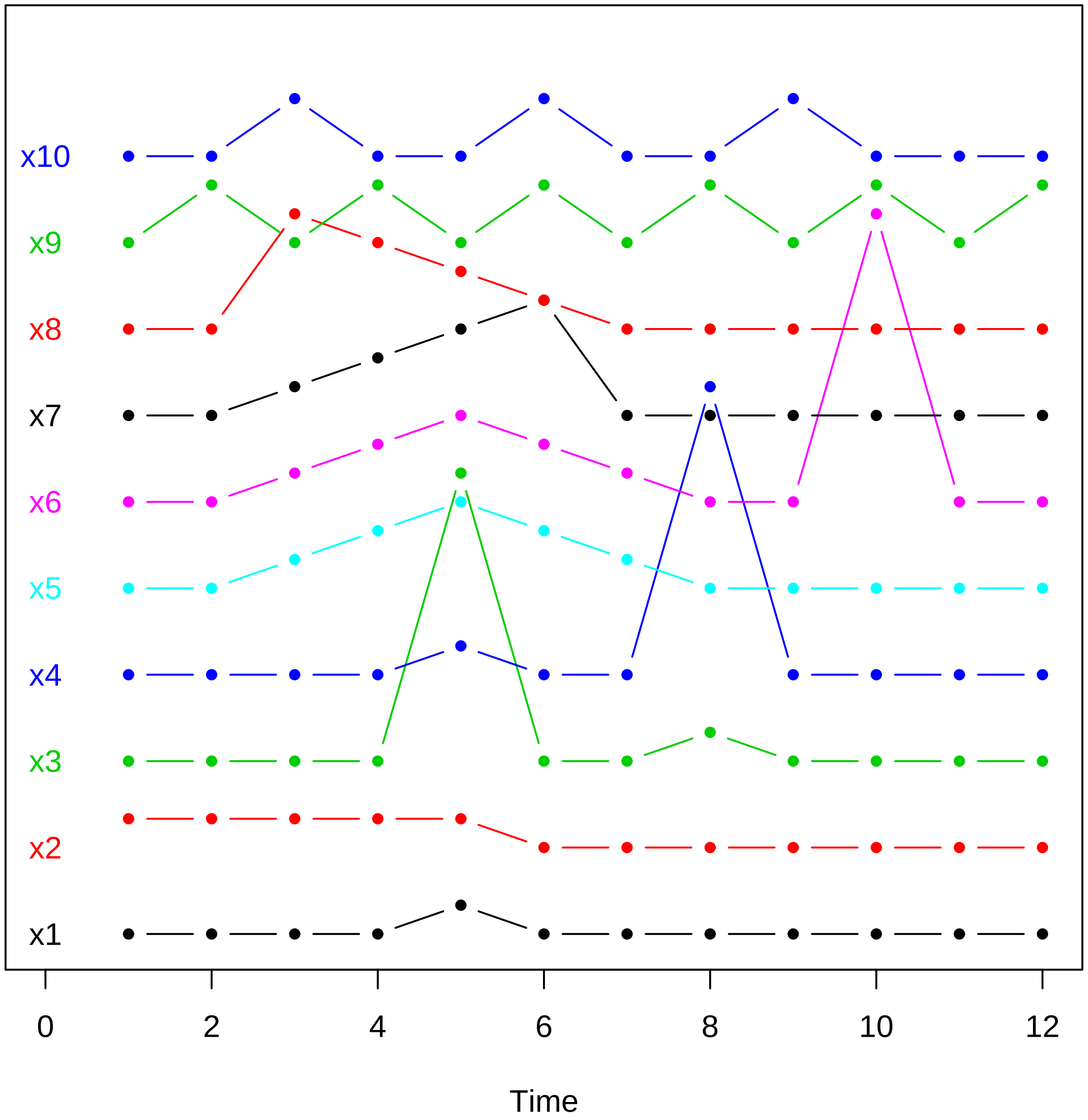
$$d_2(S_1,S_2) = \sum_{k=1}^{N} w(k)\{f_1(i_1^k - i_2^k).f_2(d_1^k - d_2^k) + n(k)f_1(i_1^k - i_2^k) + f_2(d_1^k - d_2^k)\}$$

$$f_1(.), f_2(.) = \mathrm{abs}(.) \quad \text{or} \quad (.)^2$$

$$n(k) = \begin{cases} 0 & \text{if} \quad d_1^k.d_2^k = 0 \\ 1 & \text{else.} \end{cases}$$

<u>Rmq</u> : sum starts from k=1 -> this is a semi-metric.

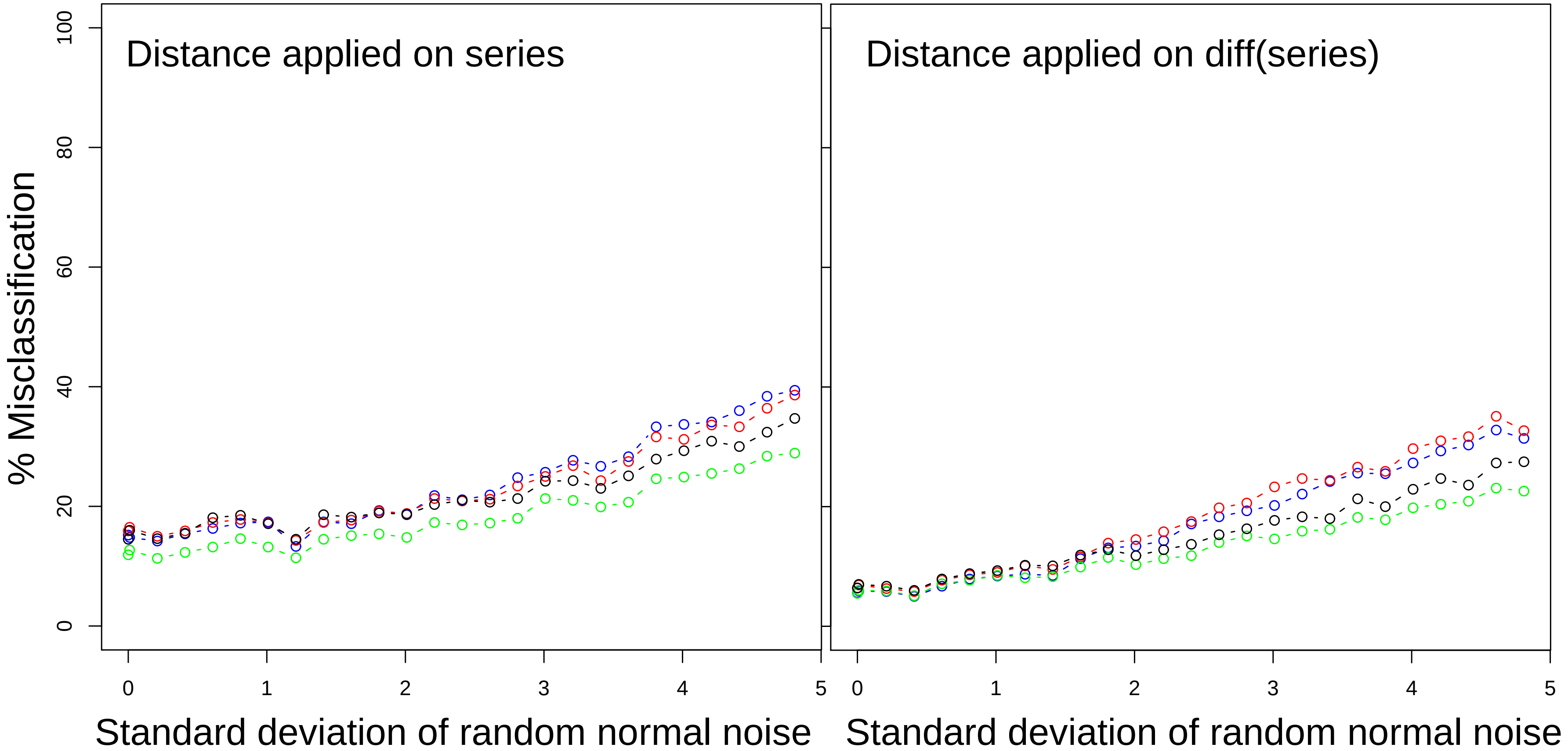
Defining a distance.
Test on simulated data.



- 10 typical series.
- Simulation of 100 series for each familly by adding a gaussian random noise.
- Supervised classification of the 1000 simulated series.
- Count percentage of misclassification.

C. Timmermans

Defining a distance.
Test on simulated data.



Blue: f_1 = abs(.), f_2 =abs(.), cross-term.
Red: f_1 = abs(.), f_2 =$(.)^2$, cross-term.
Green: f_1 = abs(.), f_2 =abs(.),no cross-term.
Black: f_1 = abs(.), f_2 =$(.)^2$, no cross-term.

Maximum noise tested has standard deviation = 5% of the highest peak, 50% of the smallest one.

Defining a distance.
Test on simulated data.

Let's include some more difficulties…

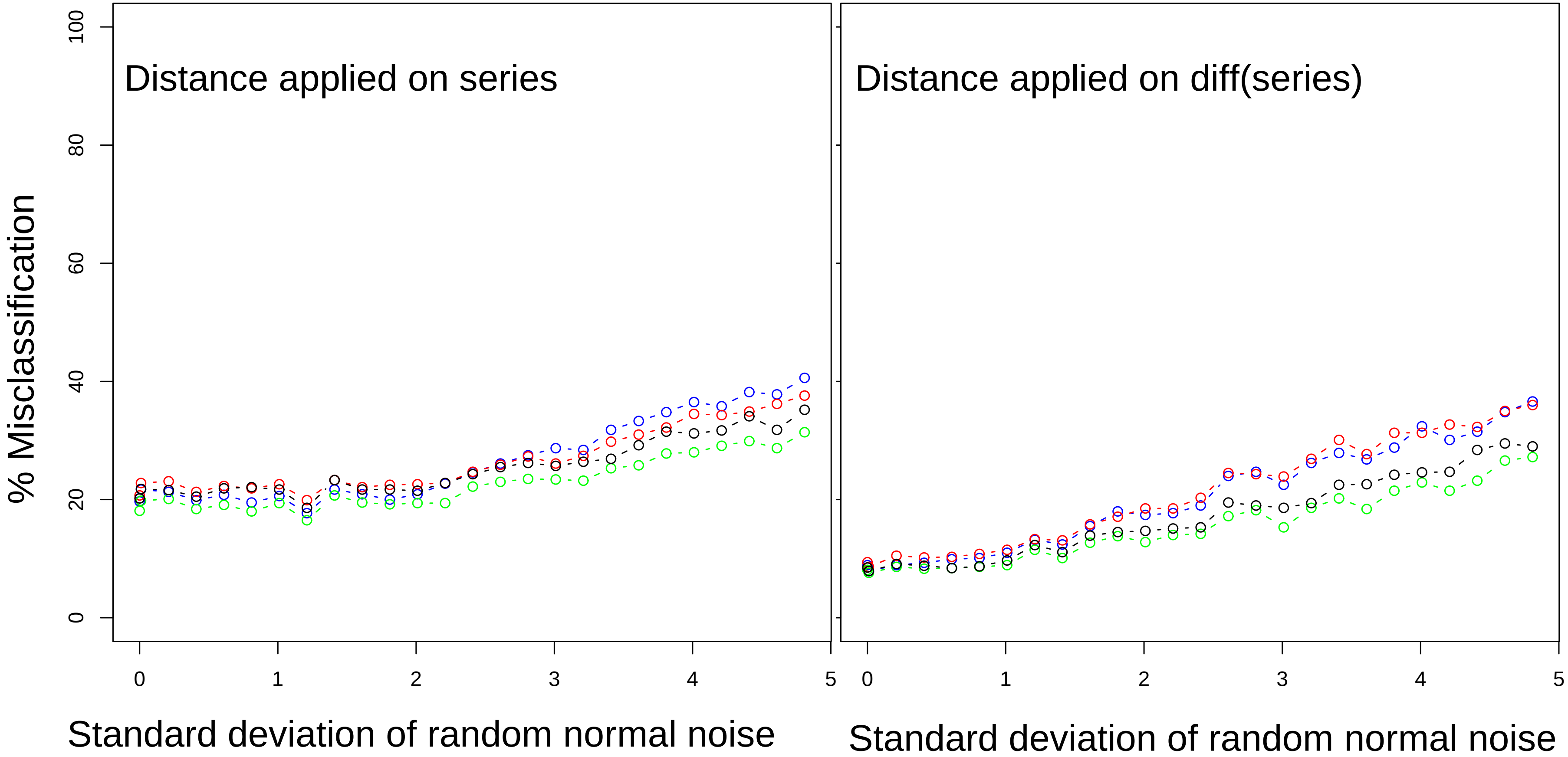
For each of the 10 families, generate:

- 20 series with gaussian random noise : $Y(t) = X(t) + Noise$
- 10 "emphasized" series: $Y(t) = X(t) * 1.5 + Noise.$
- 10 "right-shifted" series: $\begin{cases} Y(t) = X(t-1) + Noise \\ Y(1)=X(1) + Noise \end{cases}$
- 10 "left-shifted" series: $\begin{cases} Y(t) = Y(t+1)+Noise \\ Y(N) = X(N)+Noise \end{cases}$

$\longrightarrow$

- Supervised classification of the 1000 simulated series.
- Count percentage of misclassification.

Defining a distance.
Test on simulated data.



Blue: f_1 = abs(.), f_2 =abs(.), cross-term.
Red: f_1 = abs(.), f_2 =(.)², cross-term.
Green: f_1 = abs(.), f_2 =abs(.),no cross-term.
Black: f_1 = abs(.), f_2 =(.)², no cross-term.

Maximum noise tested has standard deviation = 5% of the highest peak, 50% of the smallest one.

A presentation for the 2008 Young Research Day,
Institute of Statistics, UCL.

Using unbalanced Haar wavelets for classification of time series.

- ***Time series :***

Expansion in canonical basis, basic concepts, motivation and notations.

- ***Unbalanced Haar wavelets:***

Data reduction, basis pursuit and representation.

- ***Classification:***

Distances, simulations and perspectives…

C. Timmermans