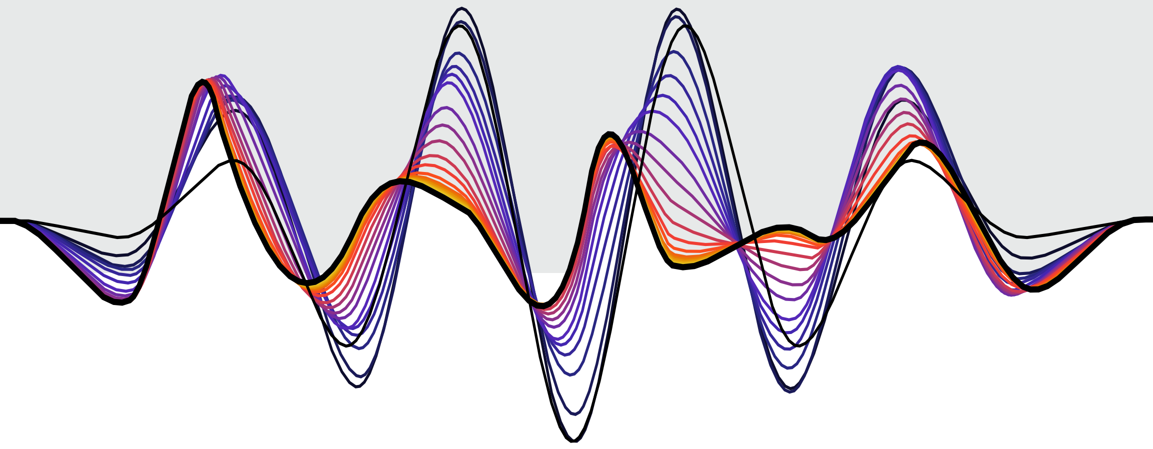
Identification de familles de signaux par apprentissage profond dans les données sismiques continues

Léonard Seydoux[1], Randall Balestriero[2], Piero Poli[1], Maarten de Hoop[3], Michel Campillo[1] and Richard Baraniuk[2]

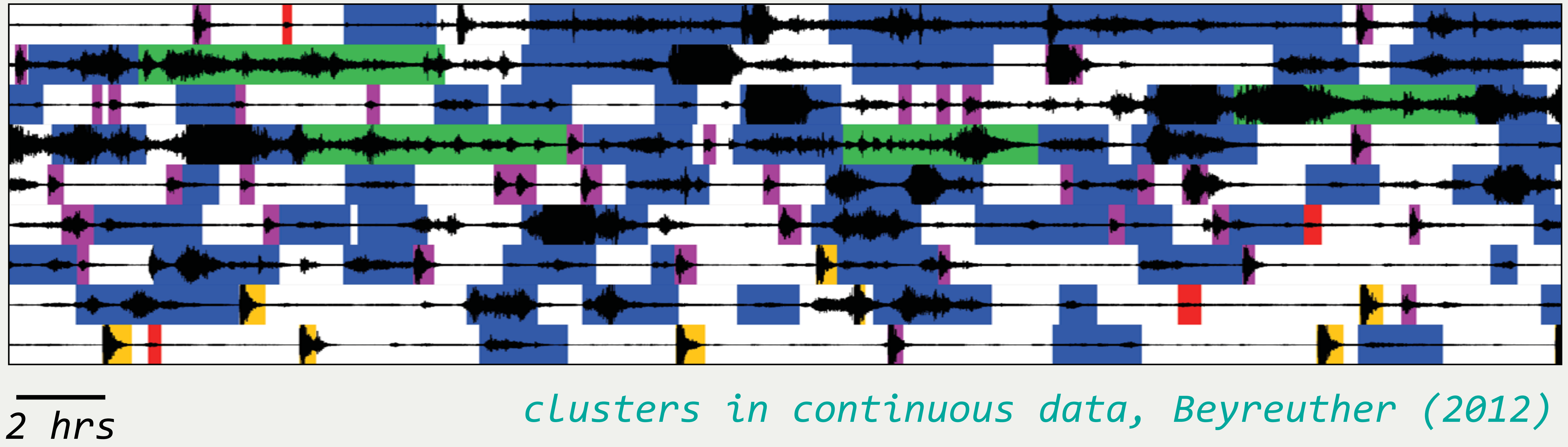
1. ISTerre, Grenoble, France 2. Electrical and Computational Engineering and
3. Computational and Applied Mathematics, Rice University, Houston, TX



Rencontres Scientifiques et Techniques **Résif** 2019 à Biarritz

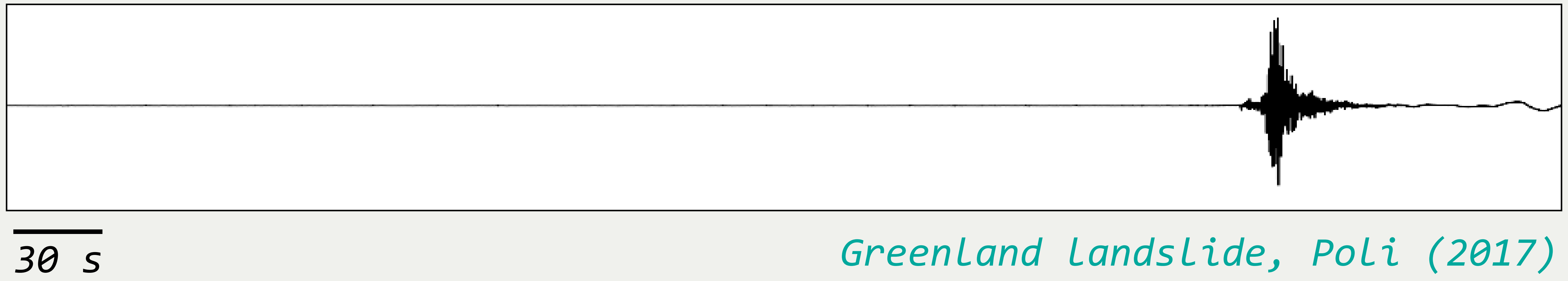
Motivations – détection & identification de signaux dans données continues

Identifier automatiquement de grandes bases de données



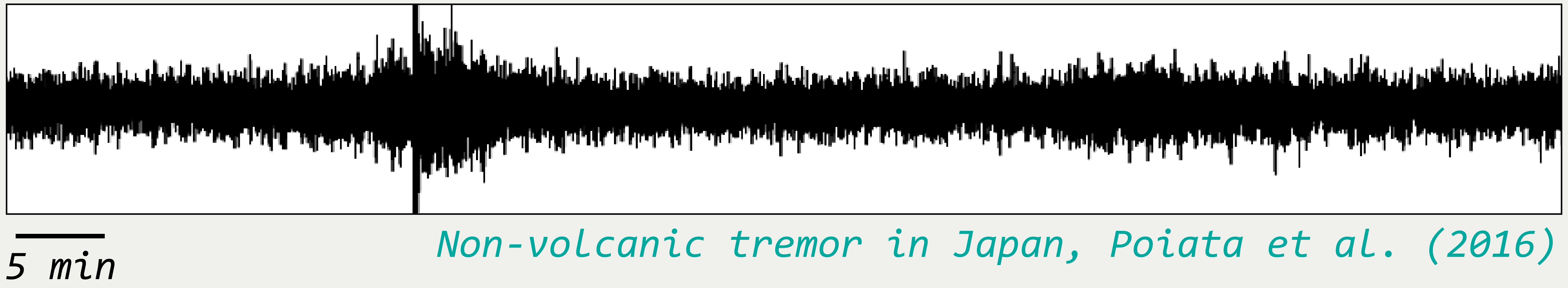
clusters in continuous data, Beyreuther (2012)

Détecter une activité sismique de faible amplitude



Greenland landslide, Poli (2017)

Détecter des signaux émergents



Non-volcanic tremor in Japan, Poiata et al. (2016)

Mettre en évidence des **nouvelles classes** de signaux sismiques

Identification de familles – méthode supervisée et non-supervisée

Supervisée (classification)

Régression entre les labels et la donnée

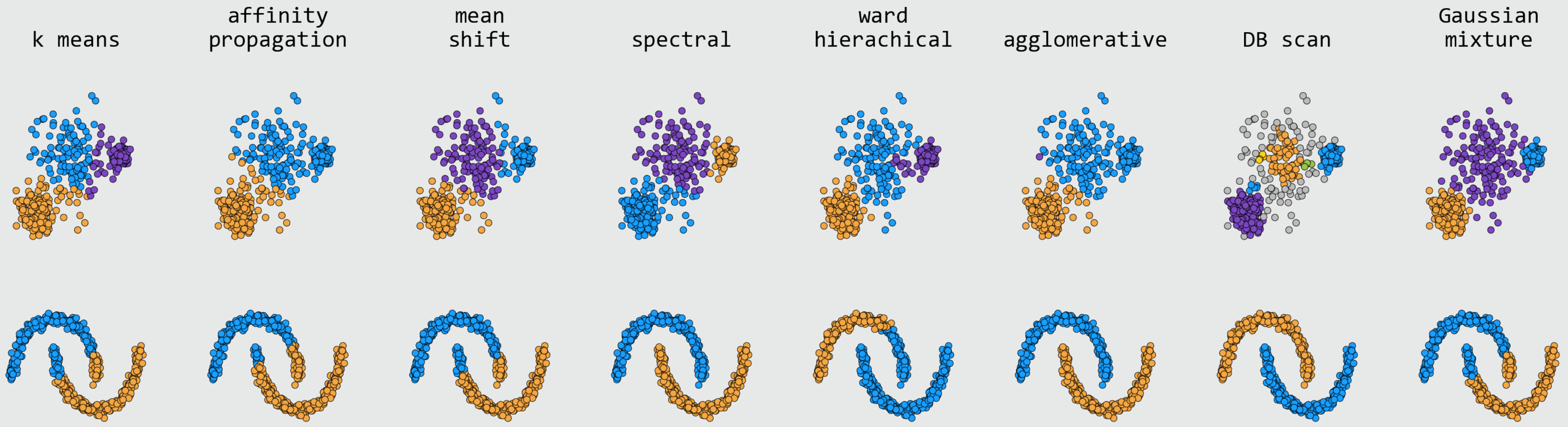
$$\underbrace{y}_{\text{label}} = \overbrace{f}^{\text{modèle}}(\underbrace{x}_{\text{donnée}})$$

Non-supervisée (clustering)

Modélisation de la distribution des données

$$\underbrace{\tilde{x}}_{\text{approximation de la donnée}} = \overbrace{f}^{\text{modèle}}(\underbrace{x}_{\text{donnée}})$$

Identification de familles – un grand nombre de définitions, autant de solutions



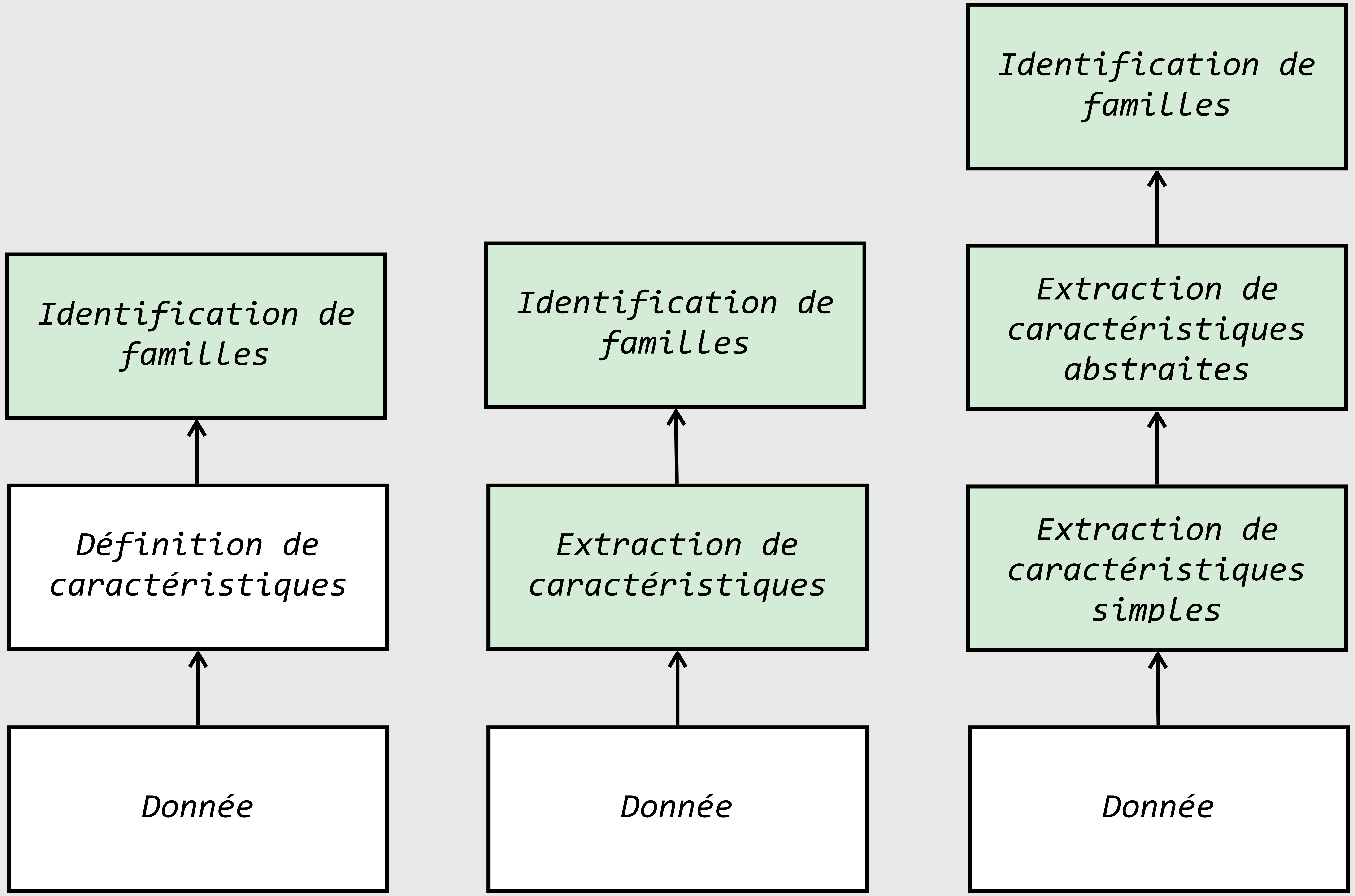
source: [scikit-learn.org]

Aldenderfer & Bashfield (1984), Duda & Hart (1973), Estivill-Castro (2002)

C’est une tâche d’exploration, **tout résultat a du sens**

Cas particulier – identification de familles de formes d’onde

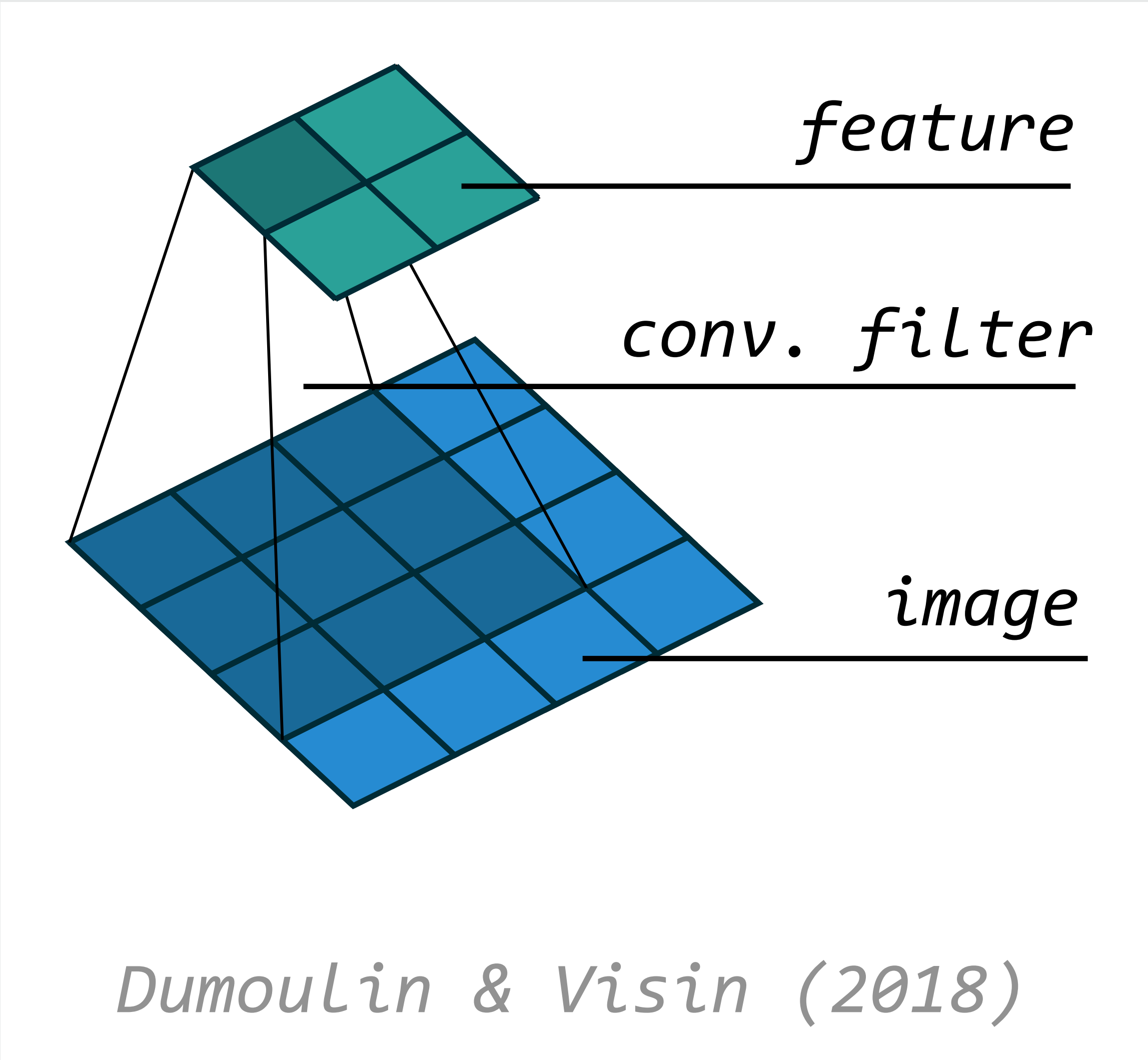
modifié de Goodfellow et al. (2016)



Apprentissage classique | Apprentissage de représentation | Apprentissage profond

Concept des réseaux de neurones convolutifs

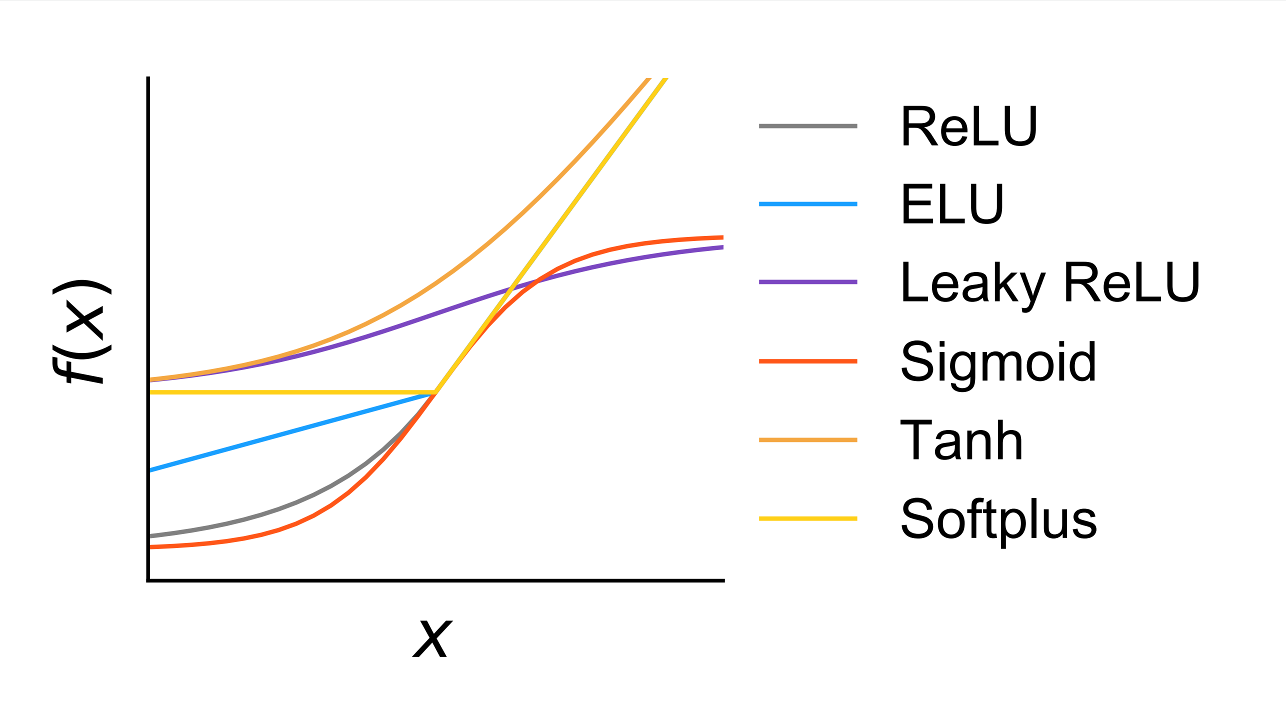
Extraction de features

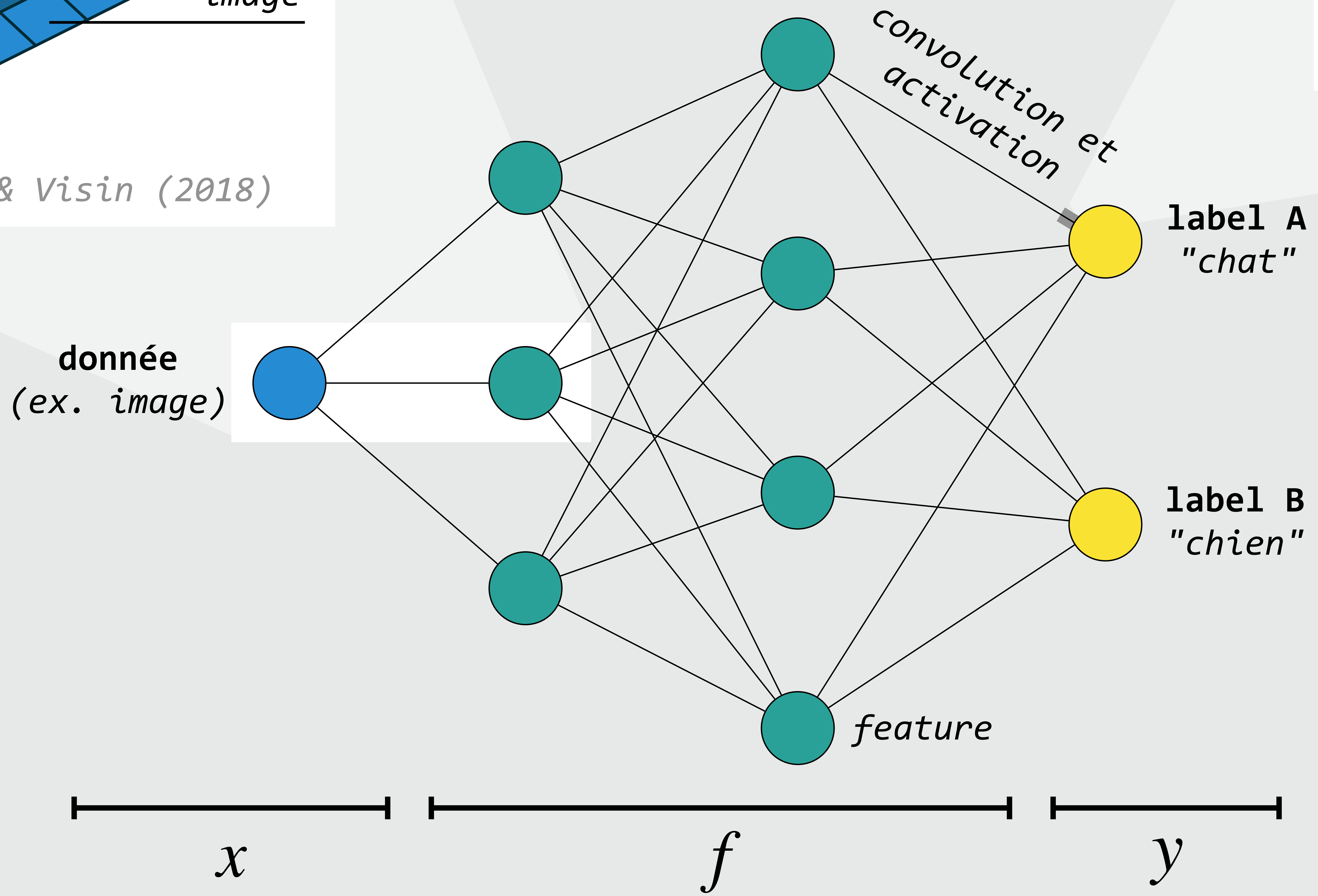


"chat"

$$f(x) = y$$

Activation non linéaire

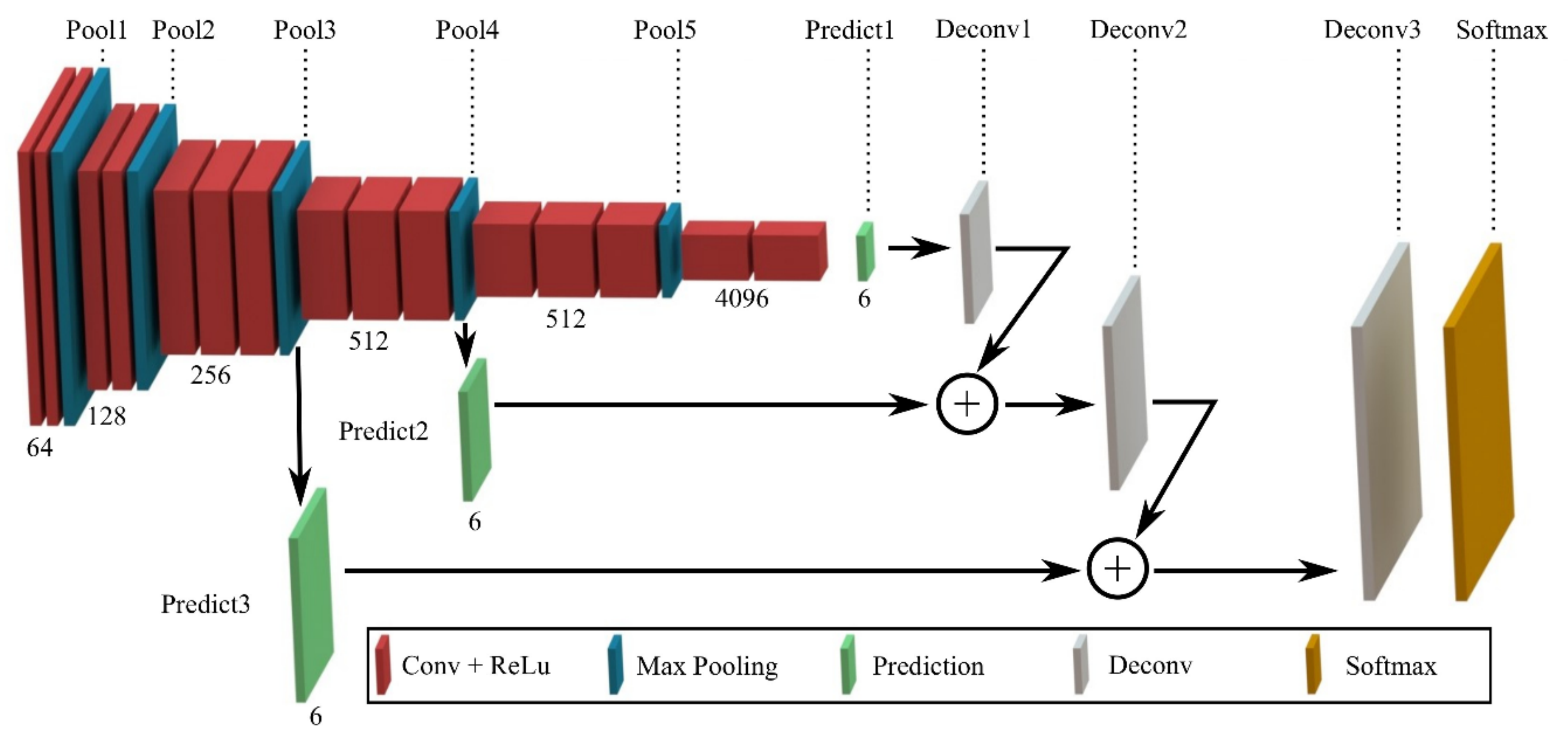




Quelle **architecture**?

Exemple d'un grand champion de la classification d'images

Deep convolutional VGG16



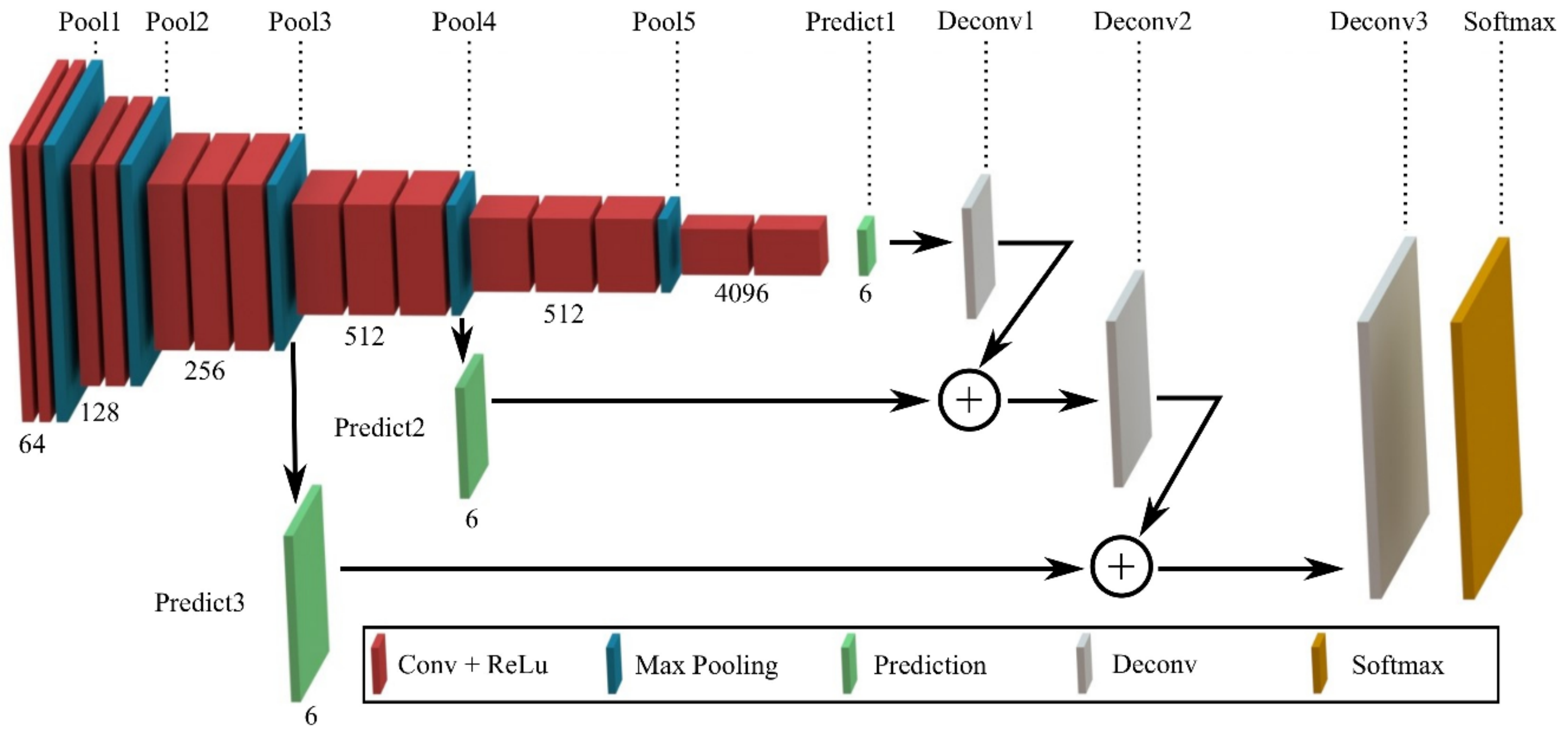
Simonyan & Zisserman (2015)

Du détail à l'abstrait

Cette architecture est le fruit **d'essais empiriques** inspirés par la nature

Exemple d’un grand champion de la classification d’images

Deep convolutional VGG16



Simonyan & Zisserman (2015)

Du détail à l’abstrait

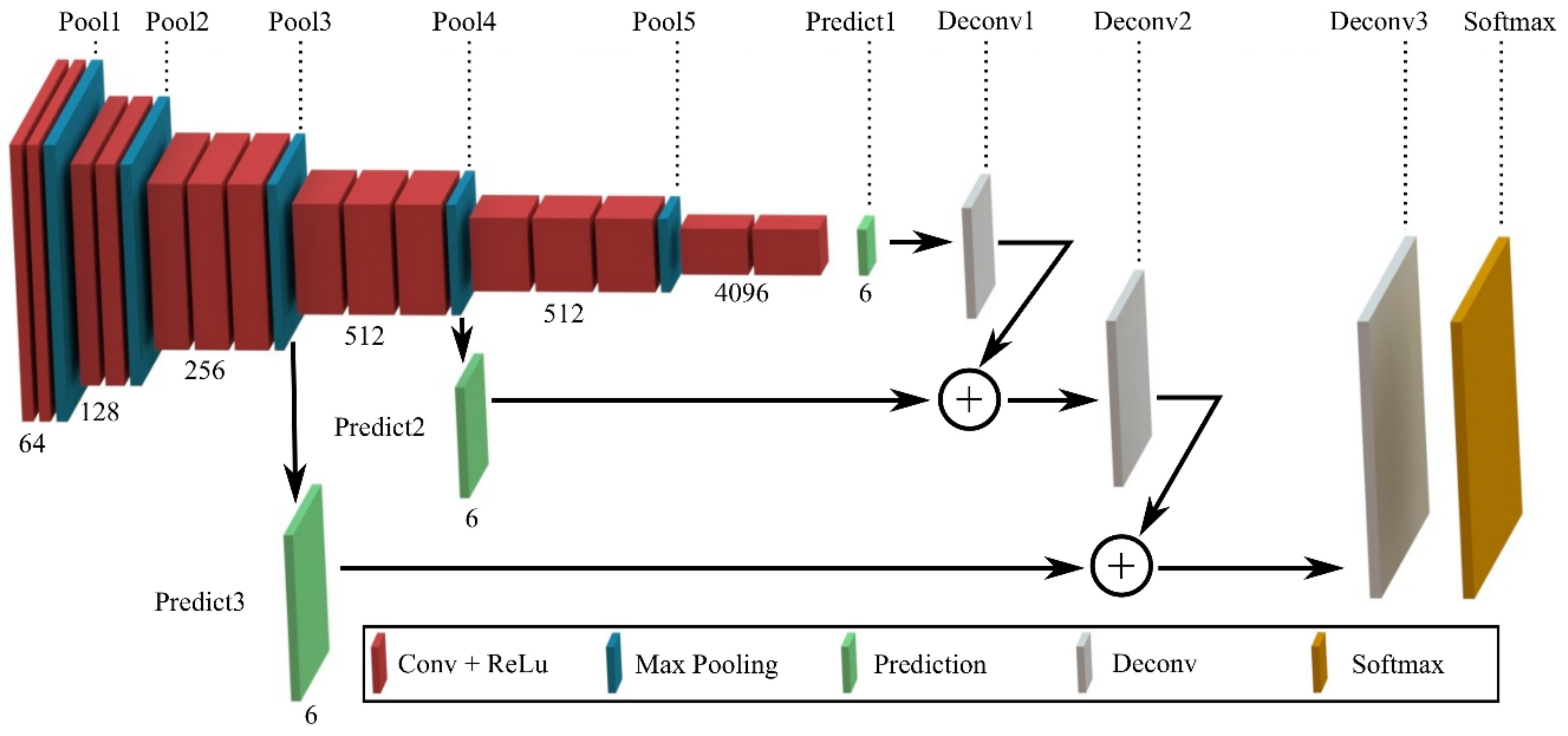
Cette architecture est le fruit **d’essais empiriques** inspirés par la nature

Exemples de filtres appris (première couche)

Les filtres de VGG16 sont sensibles à l’orientation

Exemple d’un grand champion de la classification d’images

Deep convolutional VGG16



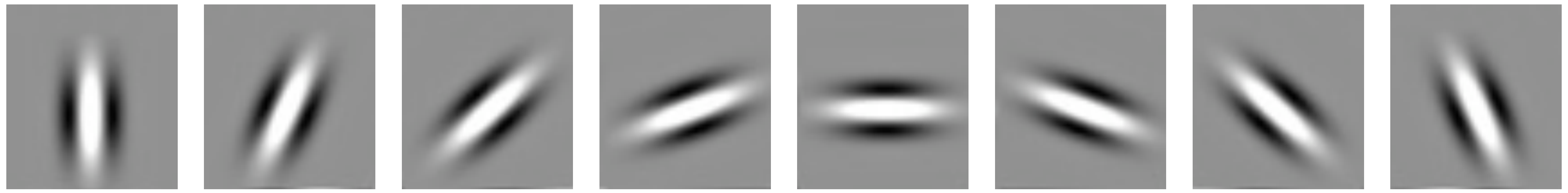
Simonyan & Zisserman (2015)

Du détail à l’abstrait

Cette architecture est le fruit **d’essais empiriques** inspirés par la nature

Exemples de filtres appris (première couche)

Ondelettes de Gabor en 2D

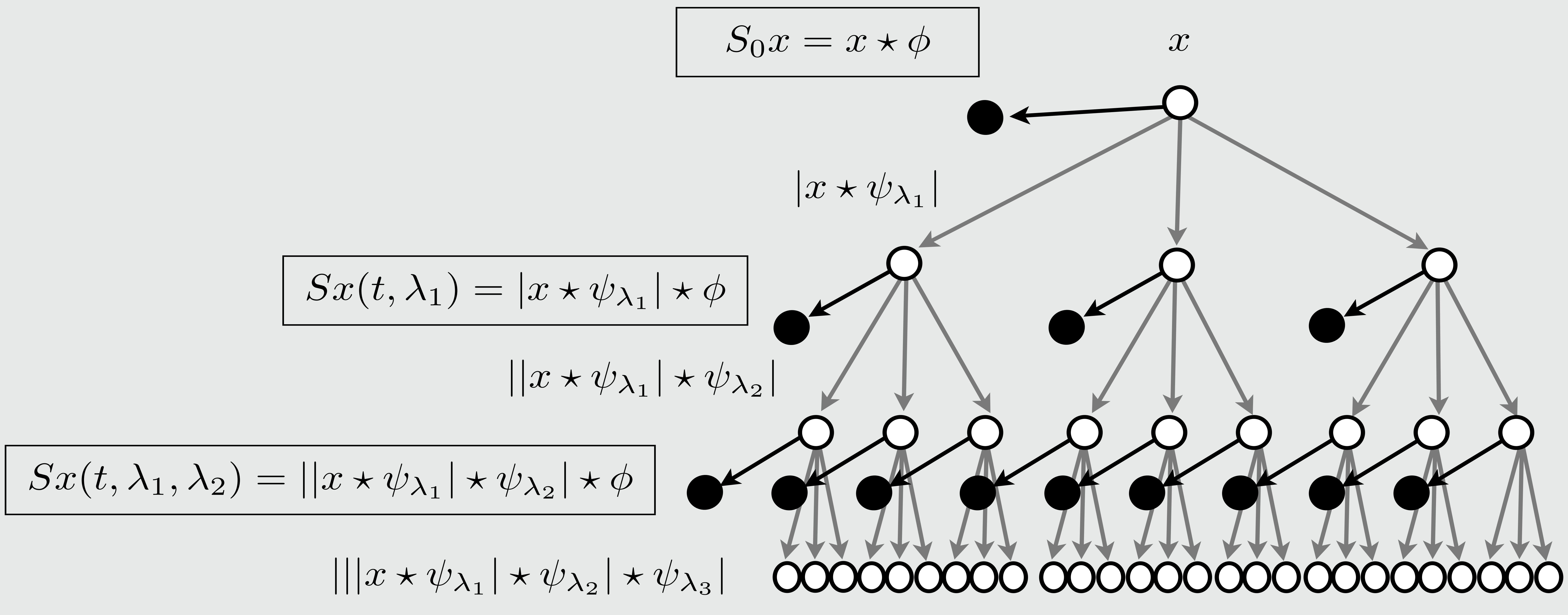


Les filtres de VGG16 sont sensibles à l’orientation

On peut les remplacer par des ondelettes

Réseau de neurones convolutif à ondelettes

Réseau diffusif *(scattering network)*

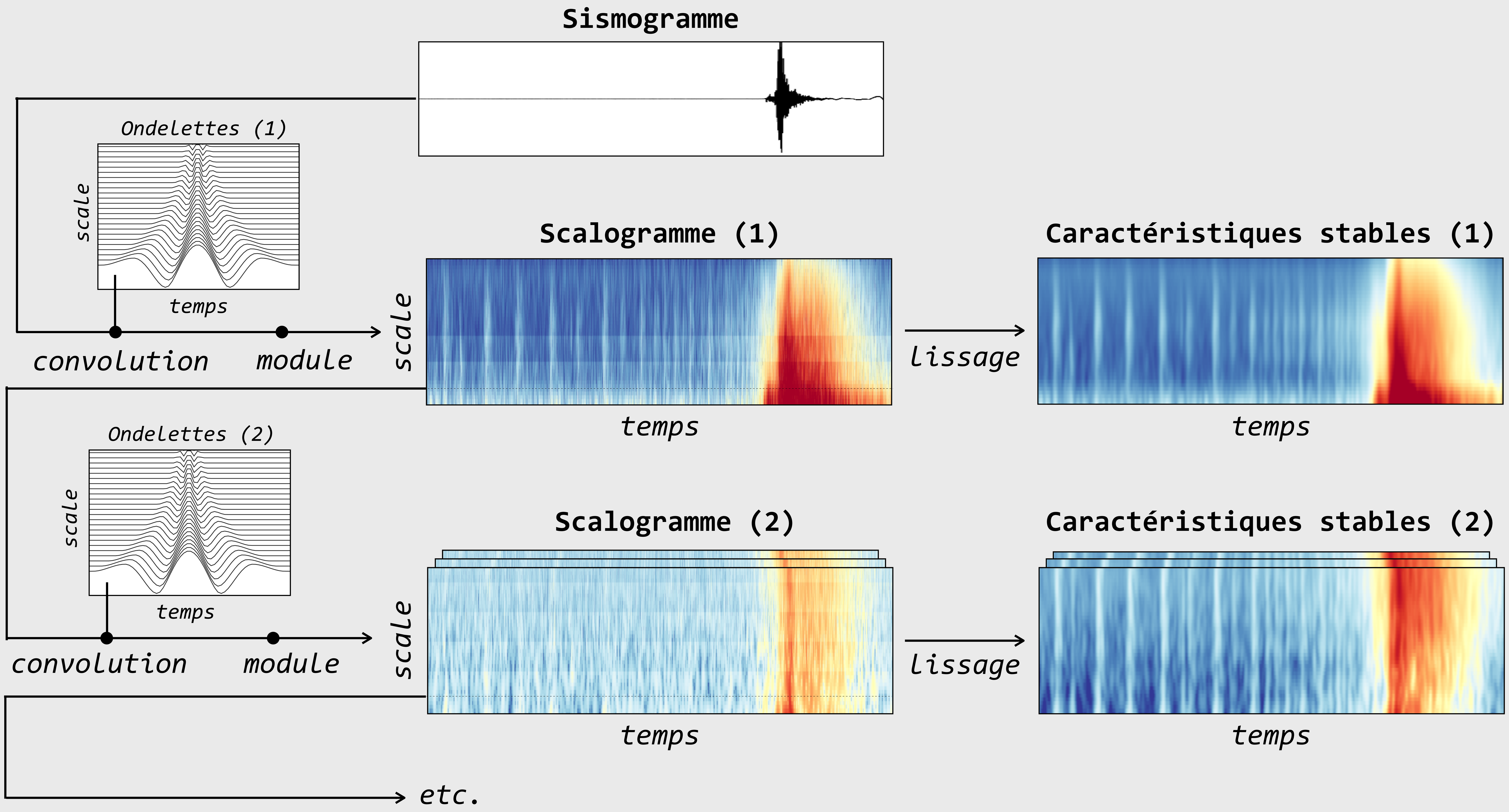


Andén & Mallat (2014)

- Filtres analytiques
- Pas d'apprentissage
- Propriétés explicites
- Architecture intuitive

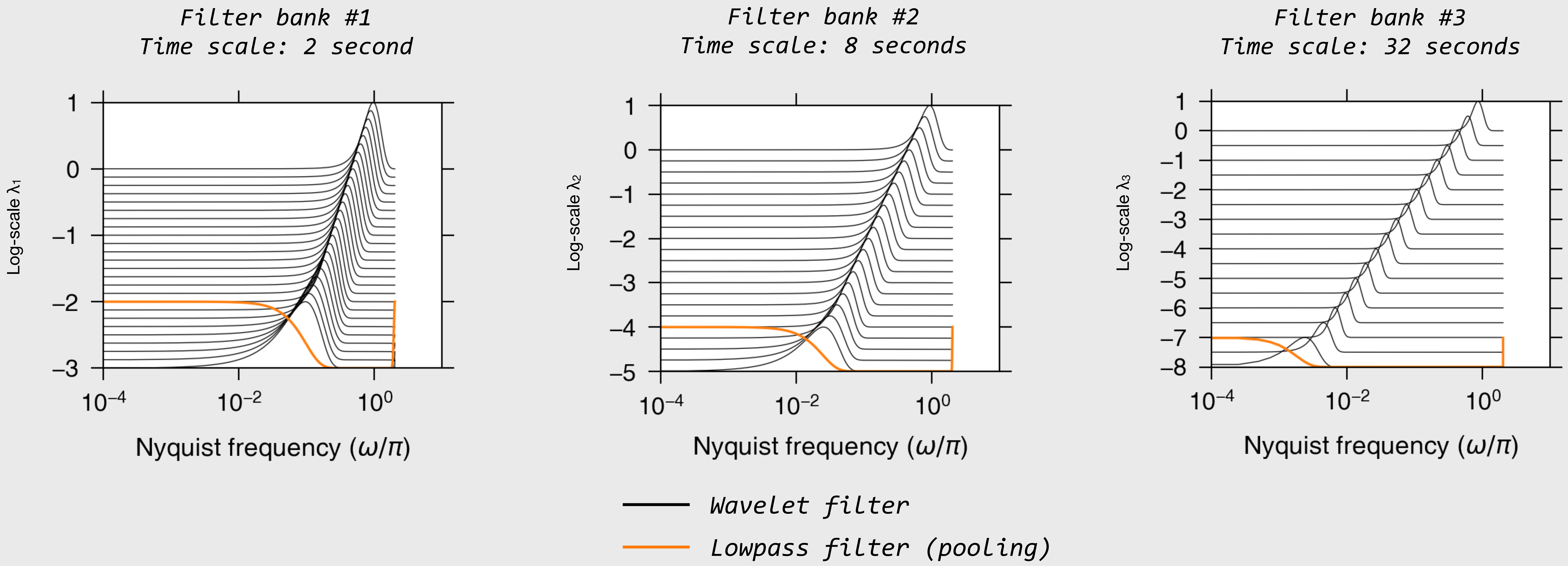
Excellent résultats sur la classification de signaux **audio** (Andén 2014), **électrocardiongrammes** & **chants d'oiseaux** (Balestriero 2017)

Représentation du signal à travel un réseau diffusif

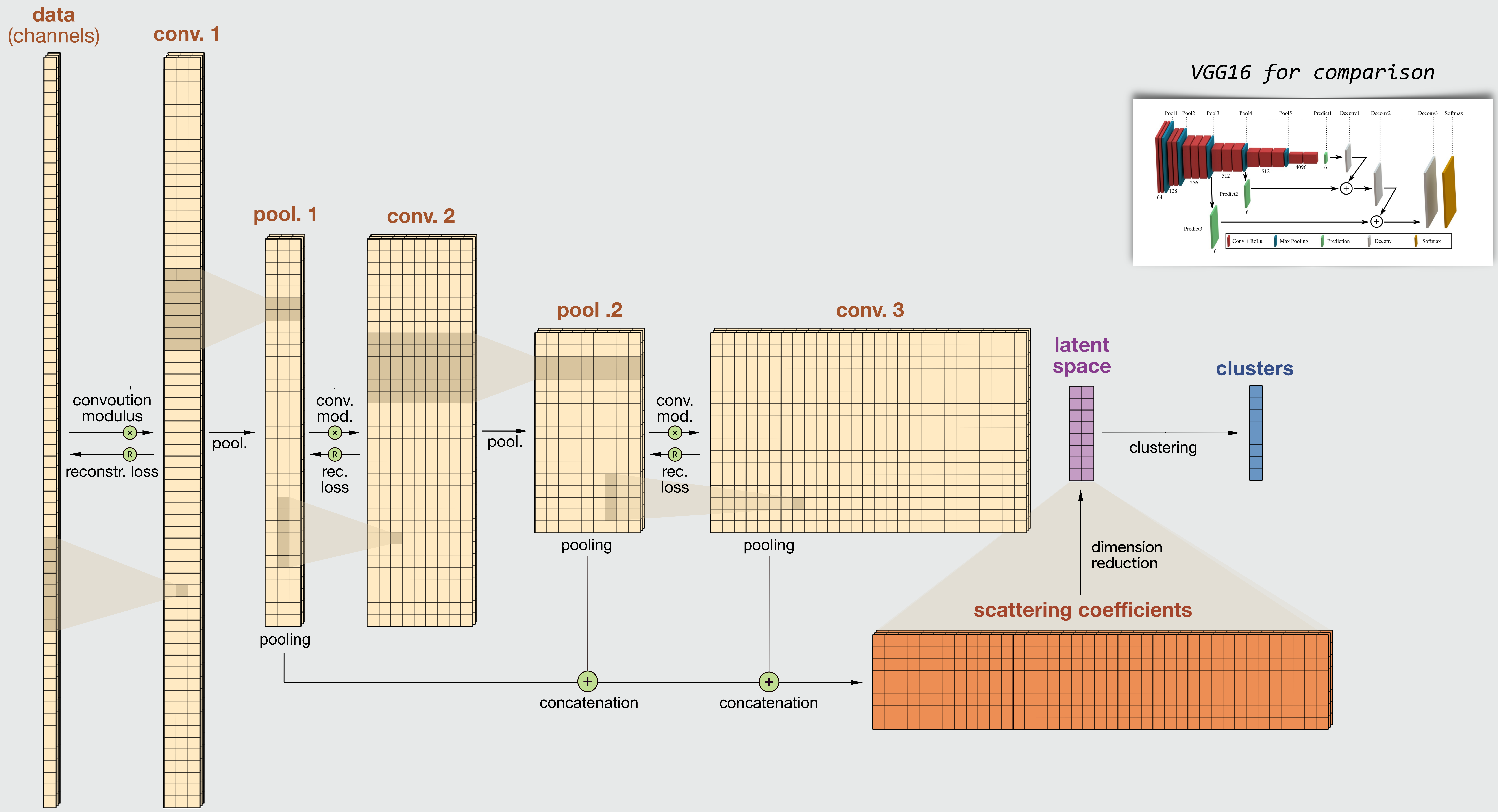


Extraction de **caractéristiques stables** à
différentes échelles de temps et de fréquence

Architecture du réseau diffusif

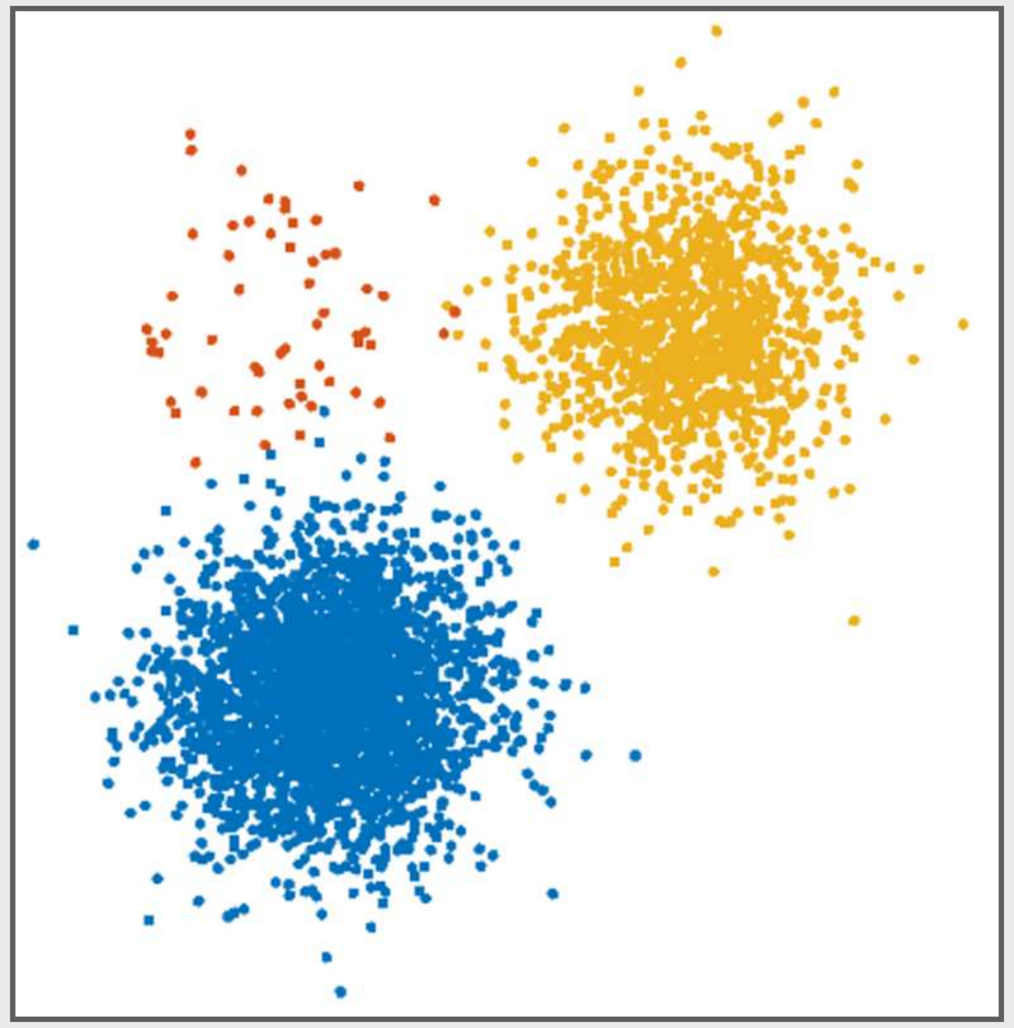


Échelle de temps



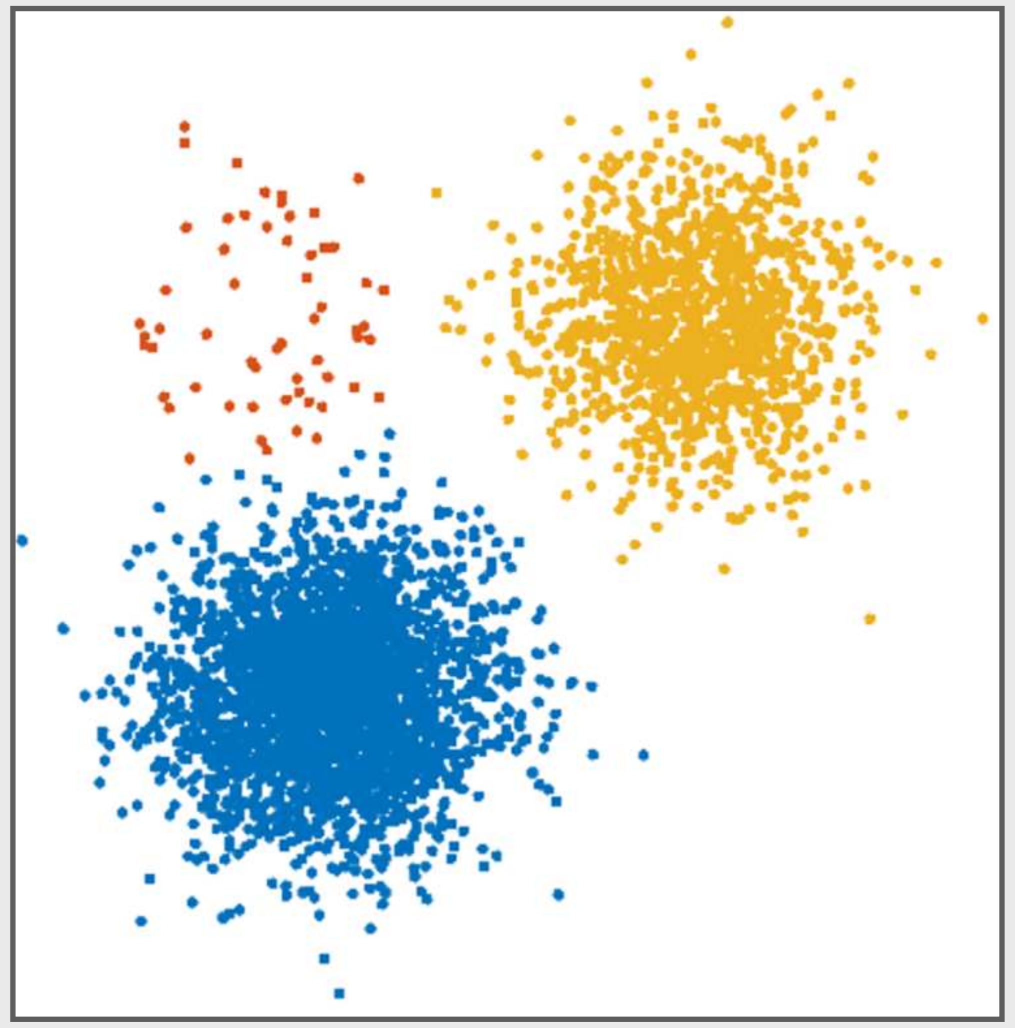
data
(channels)
conv. 1
pool. 1
conv. 2
pool .2
conv. 3
latent
space
clusters
convoution
modulus
reconstr. loss
pool.
conv.
mod.
rec.
loss
pool.
conv.
mod.
rec.
loss
clustering
pooling
pooling
pooling
dimension
reduction
scattering coefficients
concatenation
concatenation
VGG16 for comparison
Pool1
Pool2
Pool3
Pool4
Pool5
Predict1
Deconv1
Deconv2
Deconv3
Softmax
64
128
256
512
512
4096
6
Predict2
6
Predict3
6
Conv + ReLu
Max Pooling
Prediction
Deconv
Softmax

Identification de familles avec un modèle de mélange gaussien



(a) *generated synthetic data from 3 normal processes with unbalanced covariance and population size*

(b) *K-means*



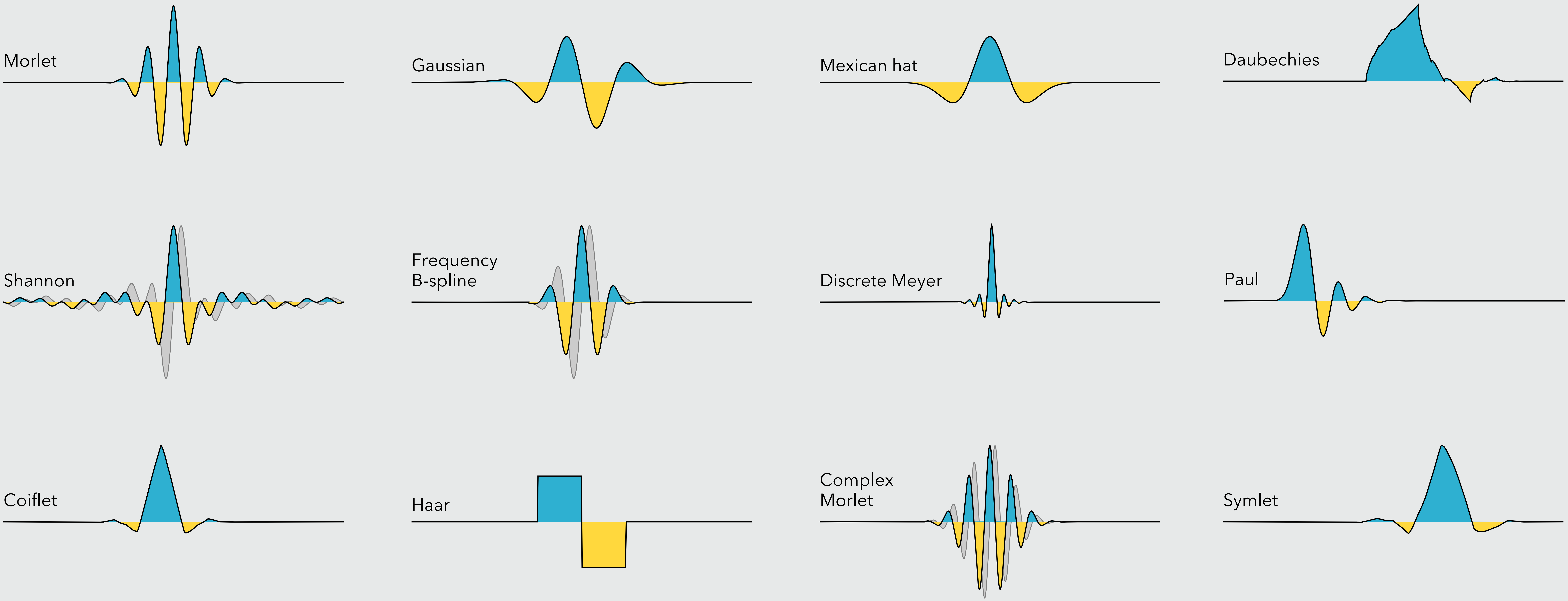
(c) *GMM, a soft probabilistic version of K-mean*

modified from Raykov et al. PONE (2016)

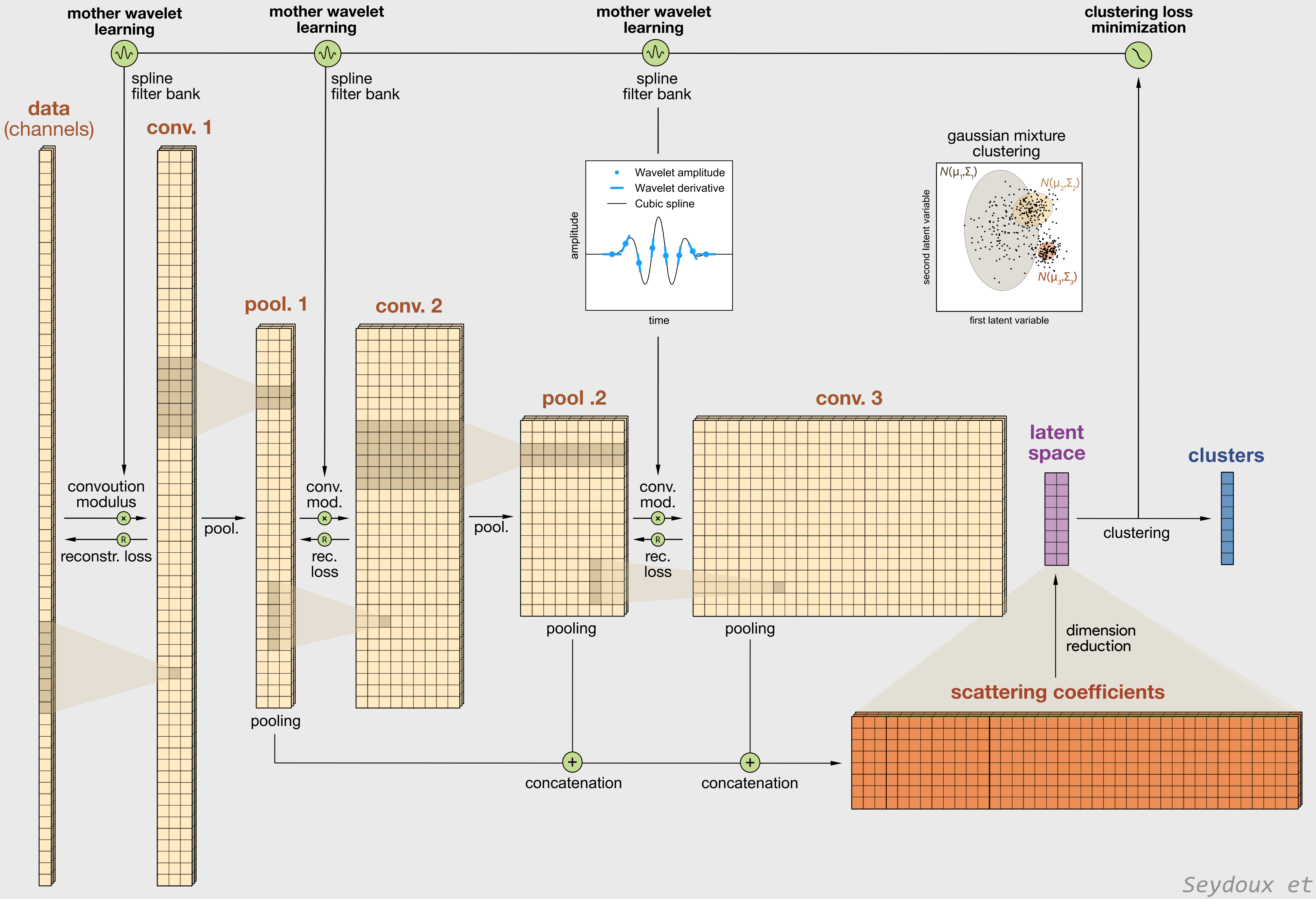
$$x \sim \prod_{k=1}^{K} \mathcal{N}(\mu_k, \Sigma_k)\mathbf{1}_{\{t=k\}}$$

GMM peut identifier des familles avec des **populations déséquilibrées**

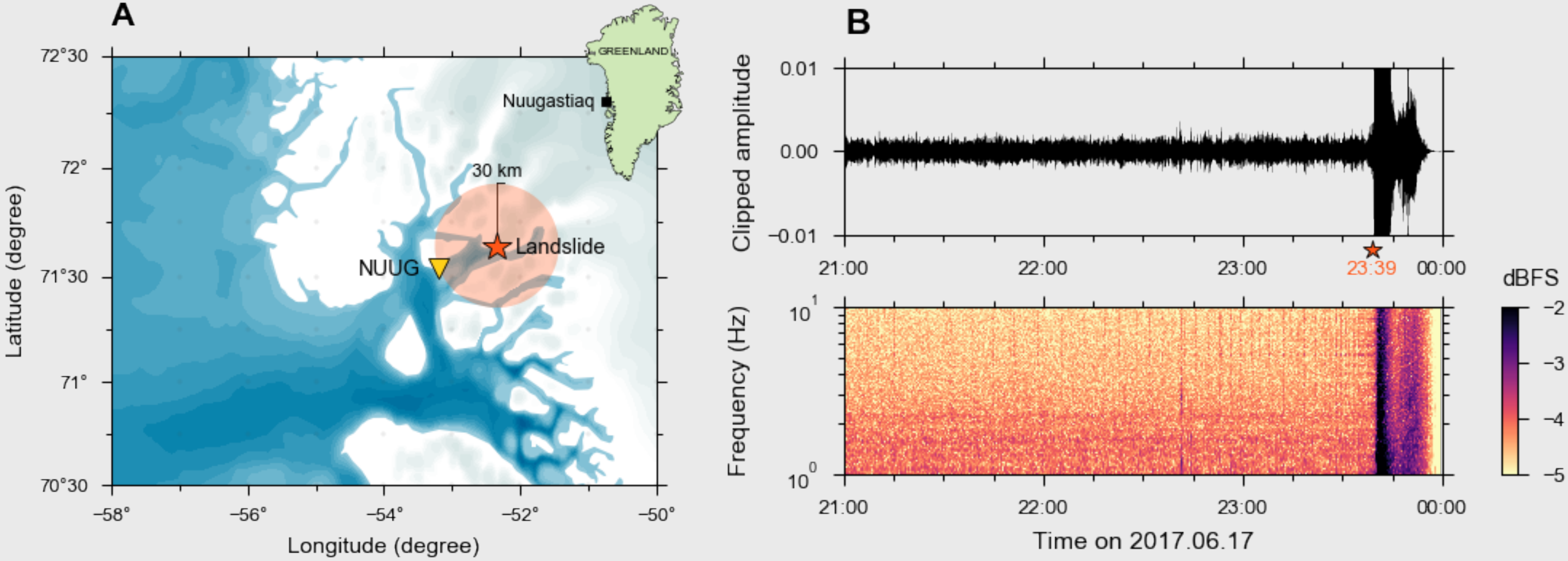
Quelle ondelette?
Toutes ces ondelettes ont des propriétés particulières



Nous pouvons aussi **apprendre l'ondelette** qui résout une tâche donnée au mieux


mother wavelet learning
mother wavelet learning
mother wavelet learning
clustering loss minimization
spline filter bank
spline filter bank
spline filter bank
data (channels)
conv. 1
pool. 1
conv. 2
pool .2
conv. 3
Wavelet amplitude
Wavelet derivative
Cubic spline
amplitude
time
gaussian mixture clustering
second latent variable
first latent variable
latent space
clusters
convoution modulus
reconstr. loss
pool.
conv. mod.
rec. loss
pool.
conv. mod.
rec. loss
clustering
pooling
pooling
pooling
dimension reduction
scattering coefficients
concatenation
concatenation

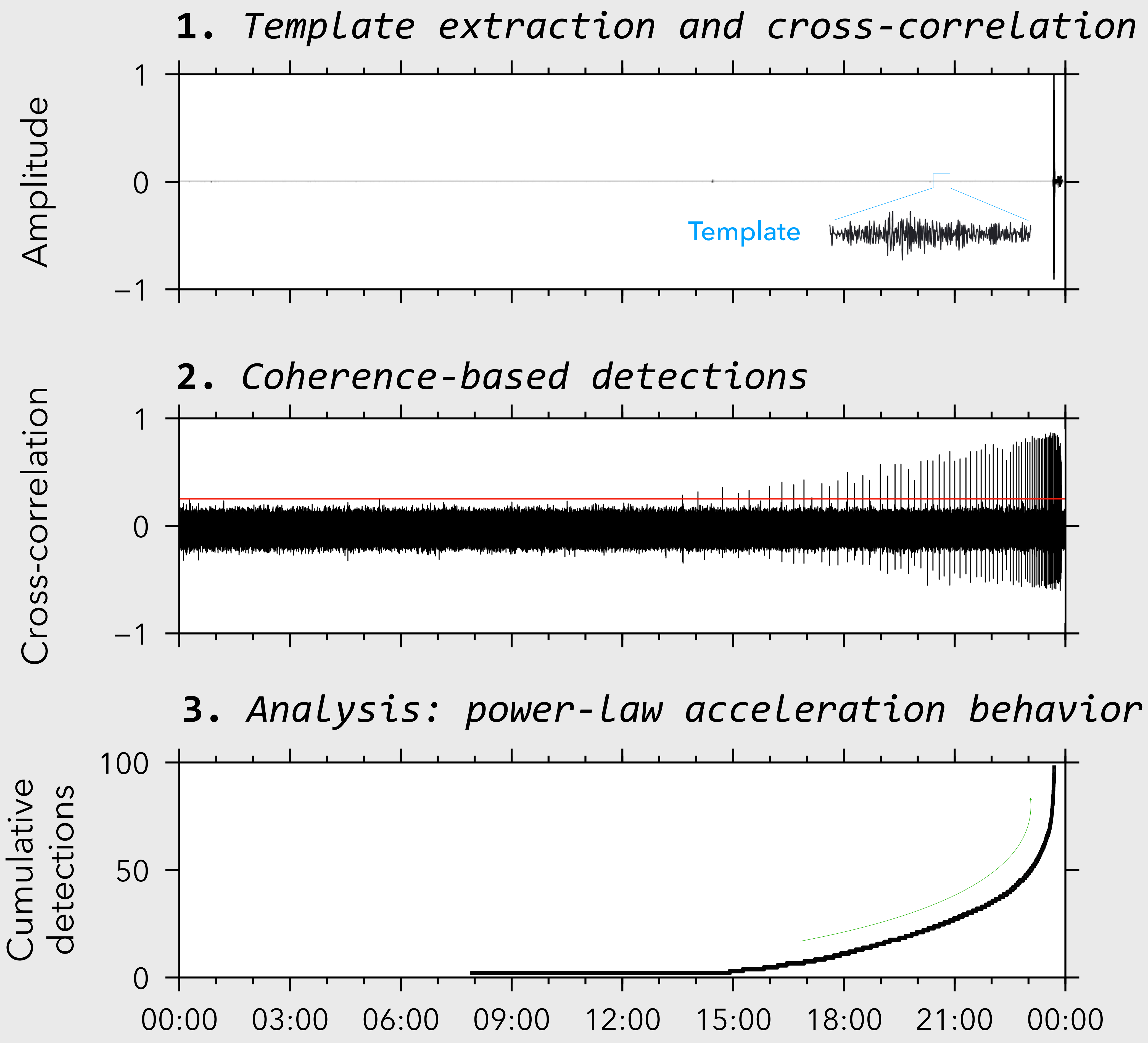
Glissement de terrain de Nuugaastiaq (2017) **– faible précurseur sismique?**



Seydoux et al. (rev.)

Glissement de terrain de Nuugaastiaq (2017) – faible précurseur sismique?

Mise en évidence par *template matching*



modified from Poli (2017)

Les avantages

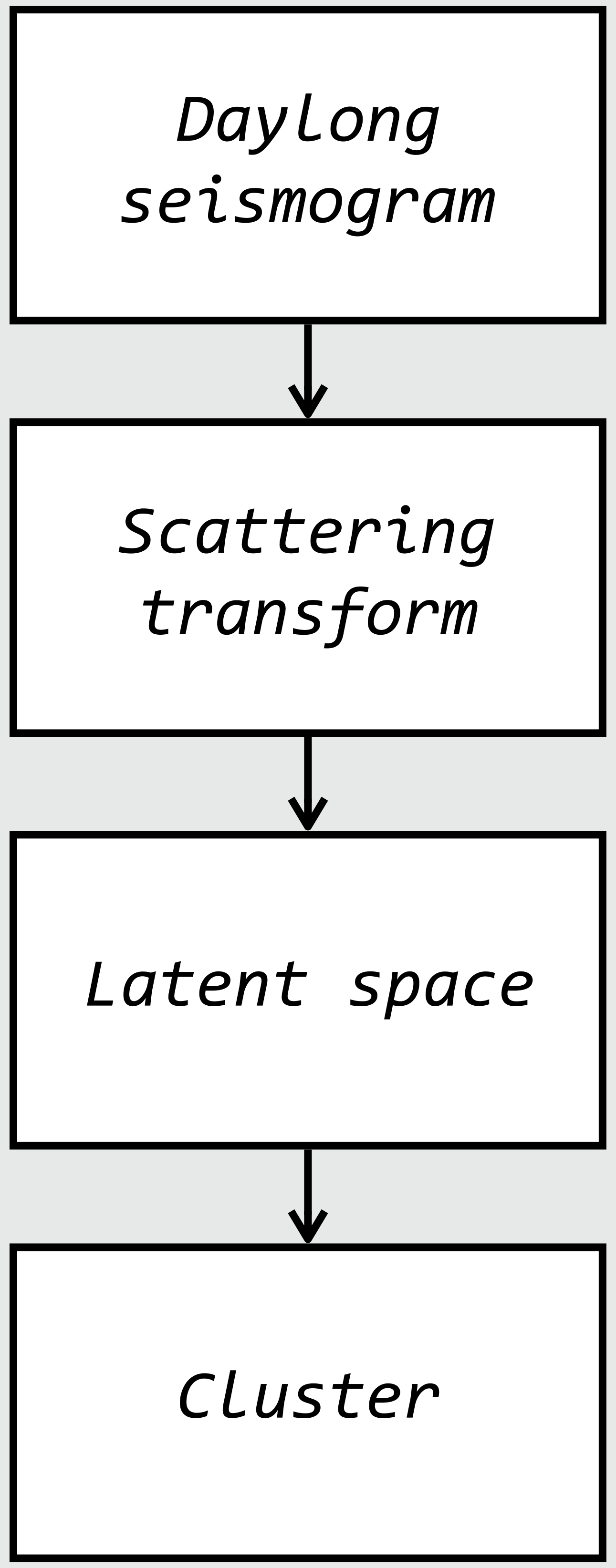
- Robuste au bruit
- Rapide à calculer

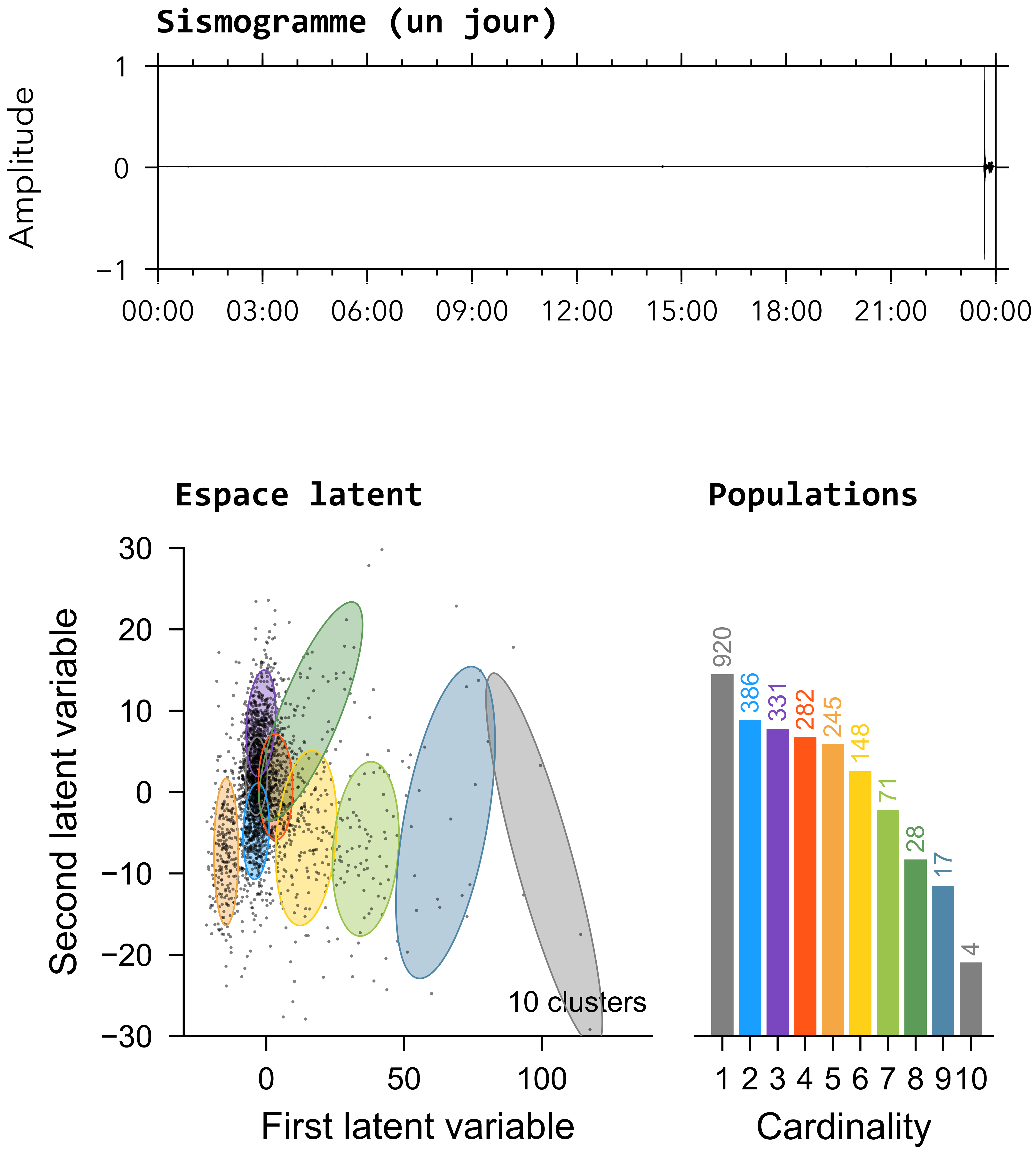
Les inconvénients

- Sensible à la définition du template
- Sensibles à plusieurs paramètres (fenêtre, fréquence)
- Limité à des signaux connus (classification à deux classes)

Peut-on retrouver ces résultats à l’aveugle?

Identification de signaux dans un jour de donnée

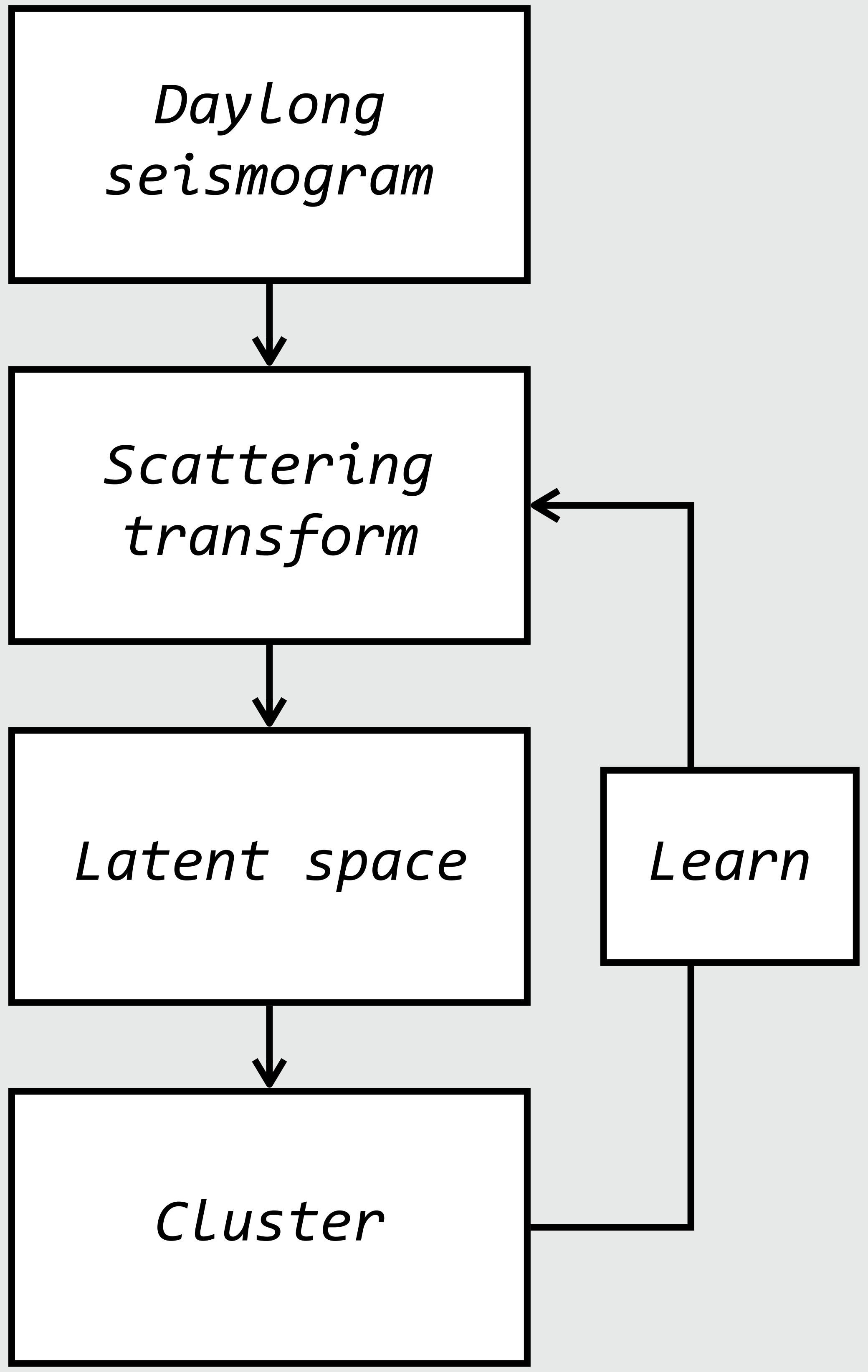


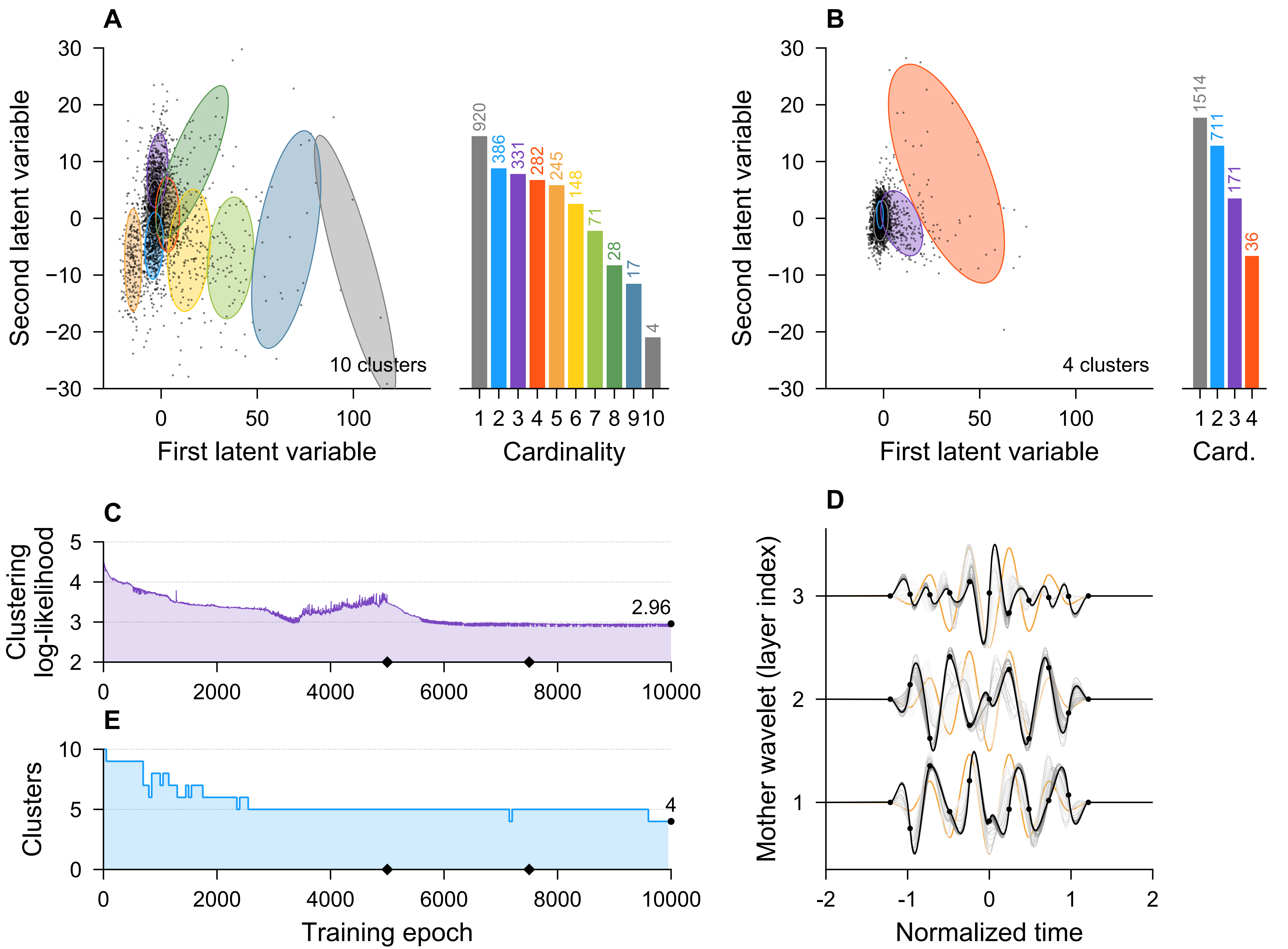


Début – beaucoup de clusters identifiés, données éparpillées

Seydoux et al. (rev.)

Identification de signaux dans un jour de donnée

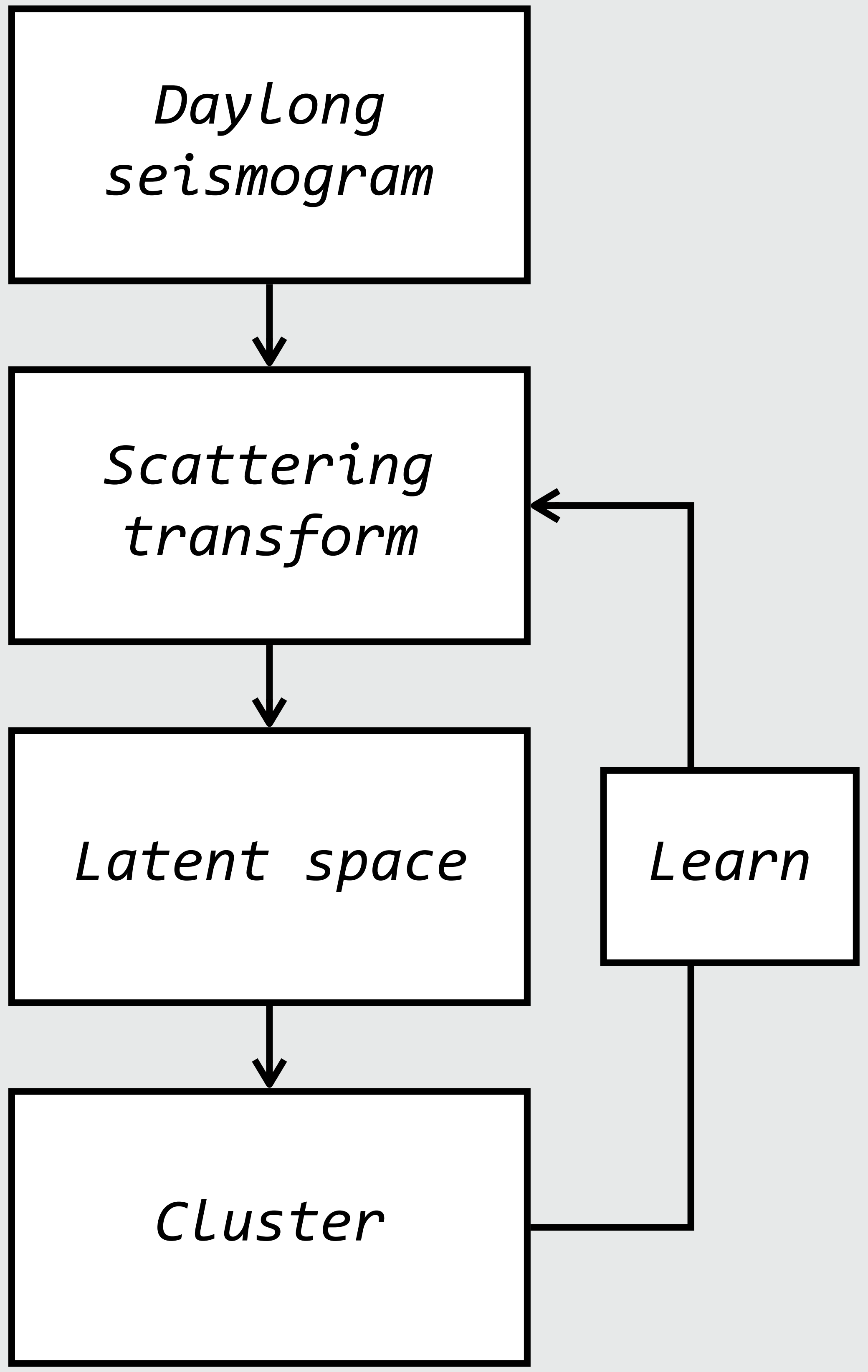


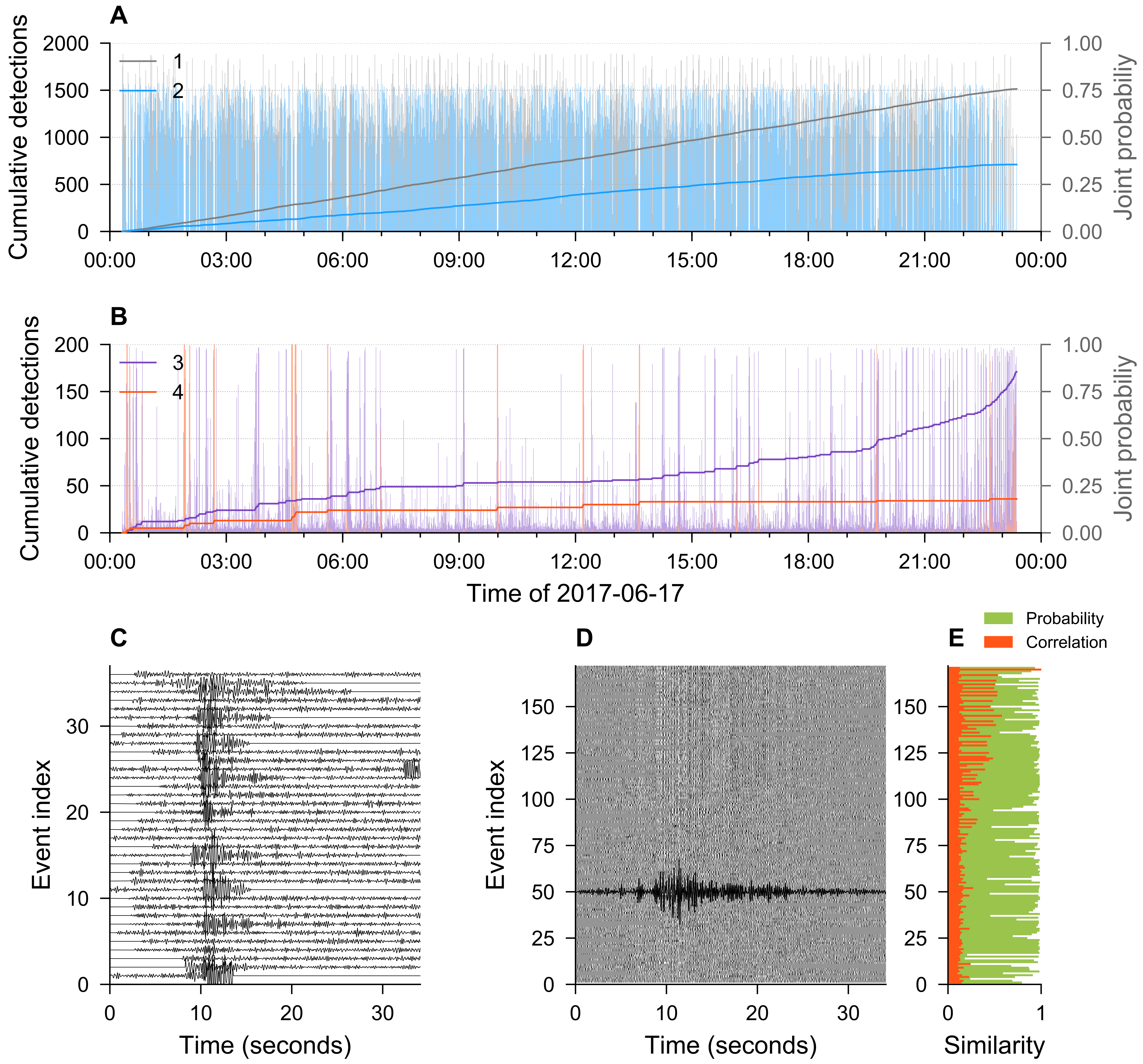


Seydoux et al. (rev.)

Après entrainement, certains points se concentrent, d'autres non

Identification de signaux dans un jour de donnée

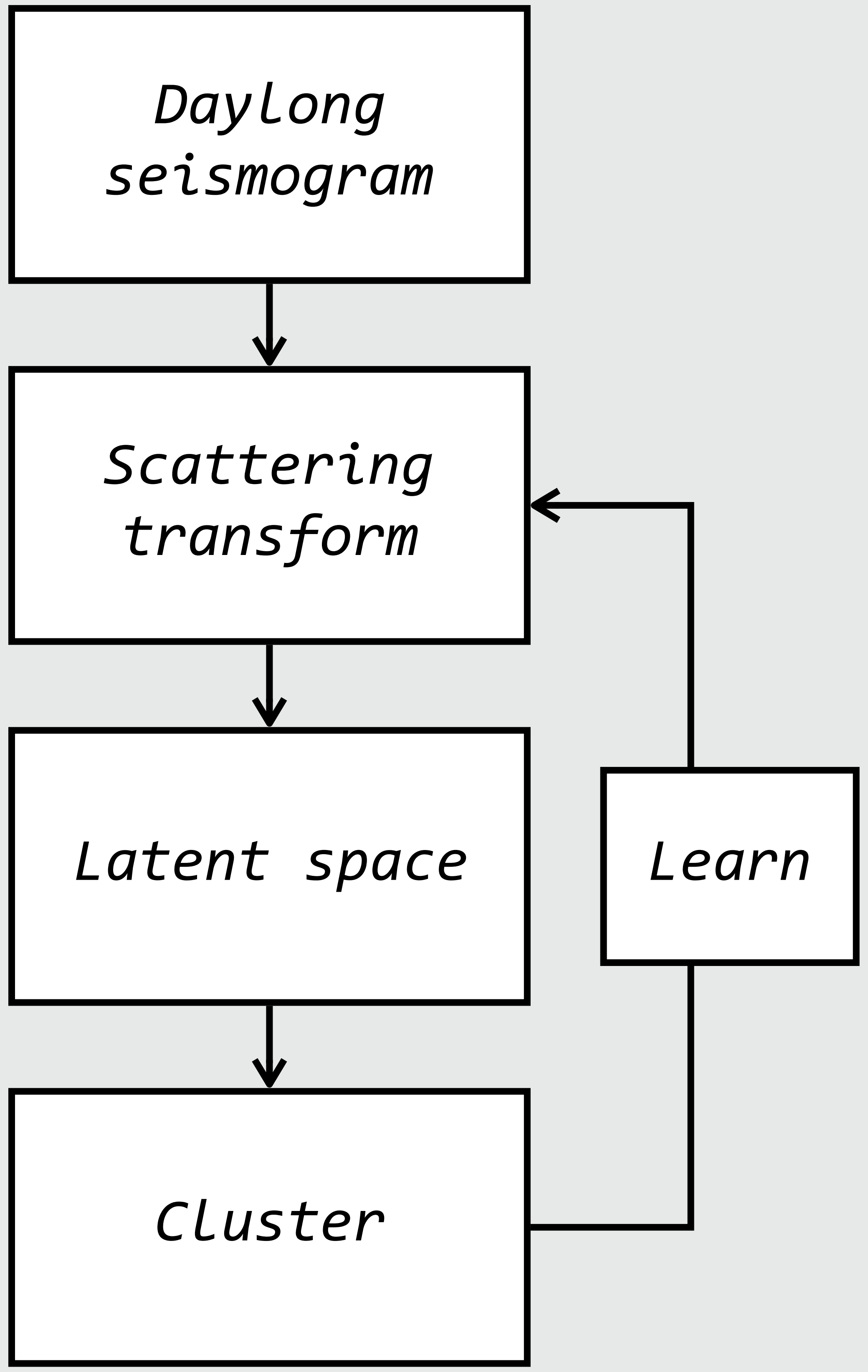


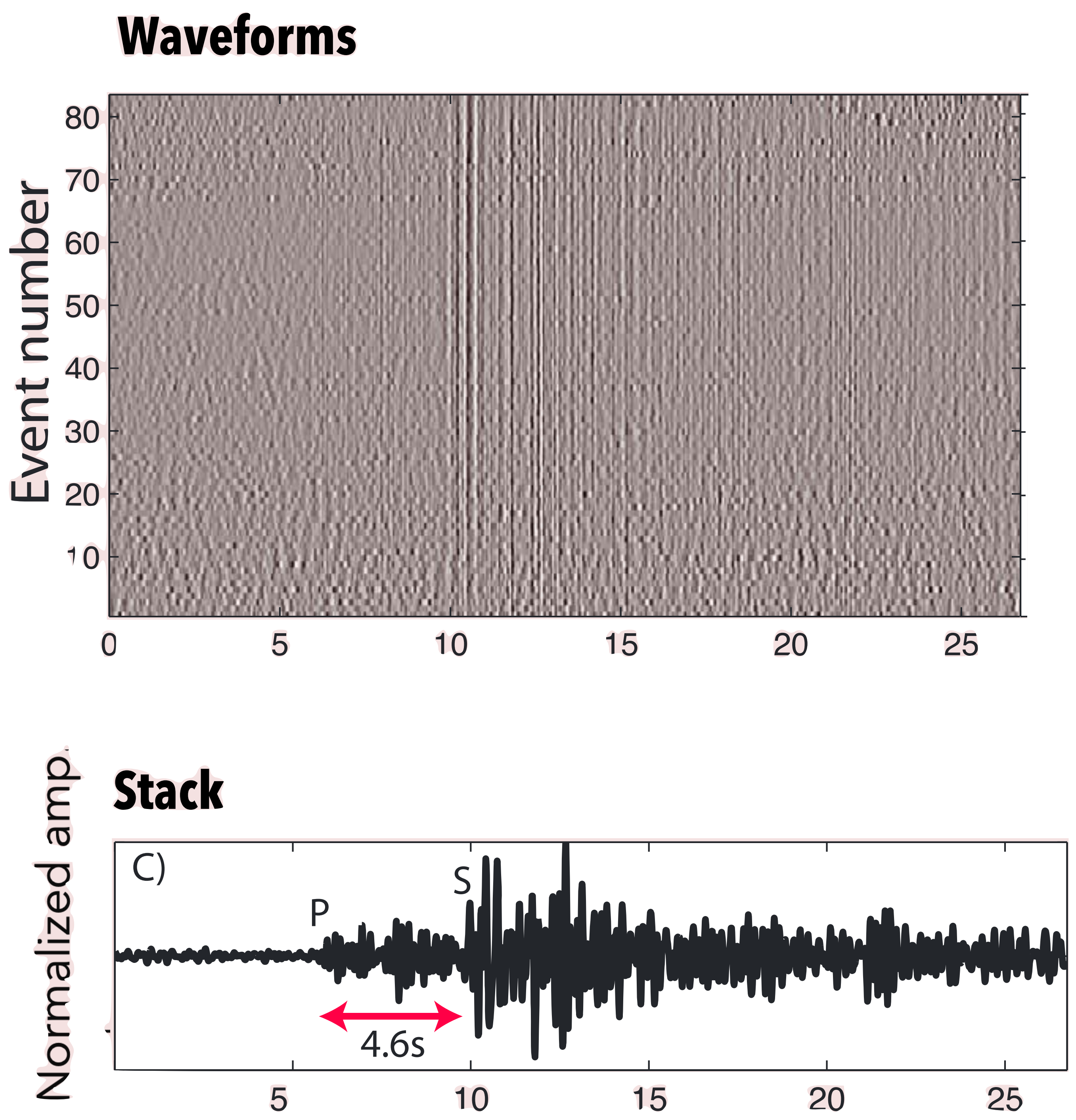


Seydoux et al. (rev.)

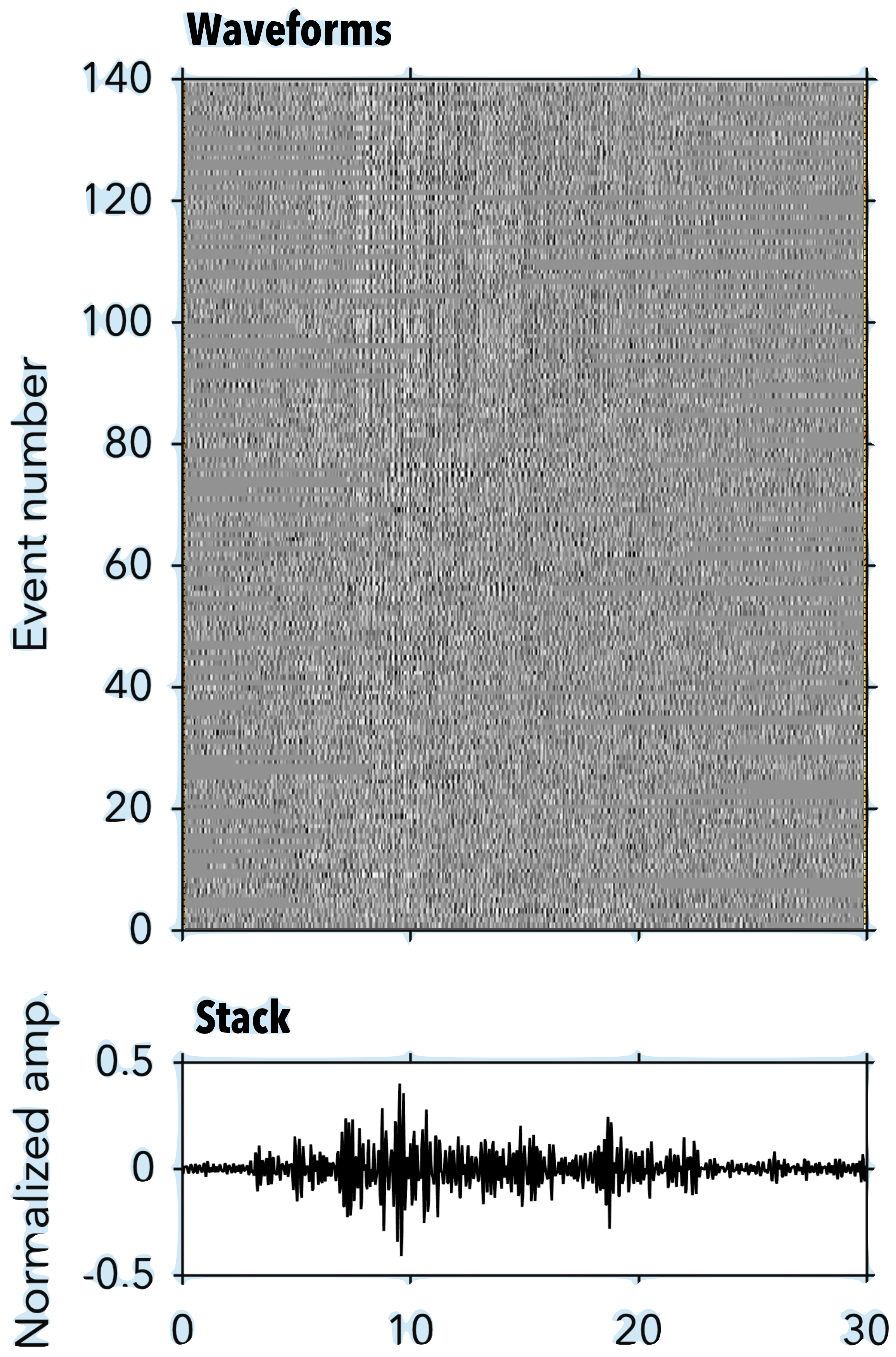
Détection cumulatives de chaque familles & formes d’ondes

vs. template matching





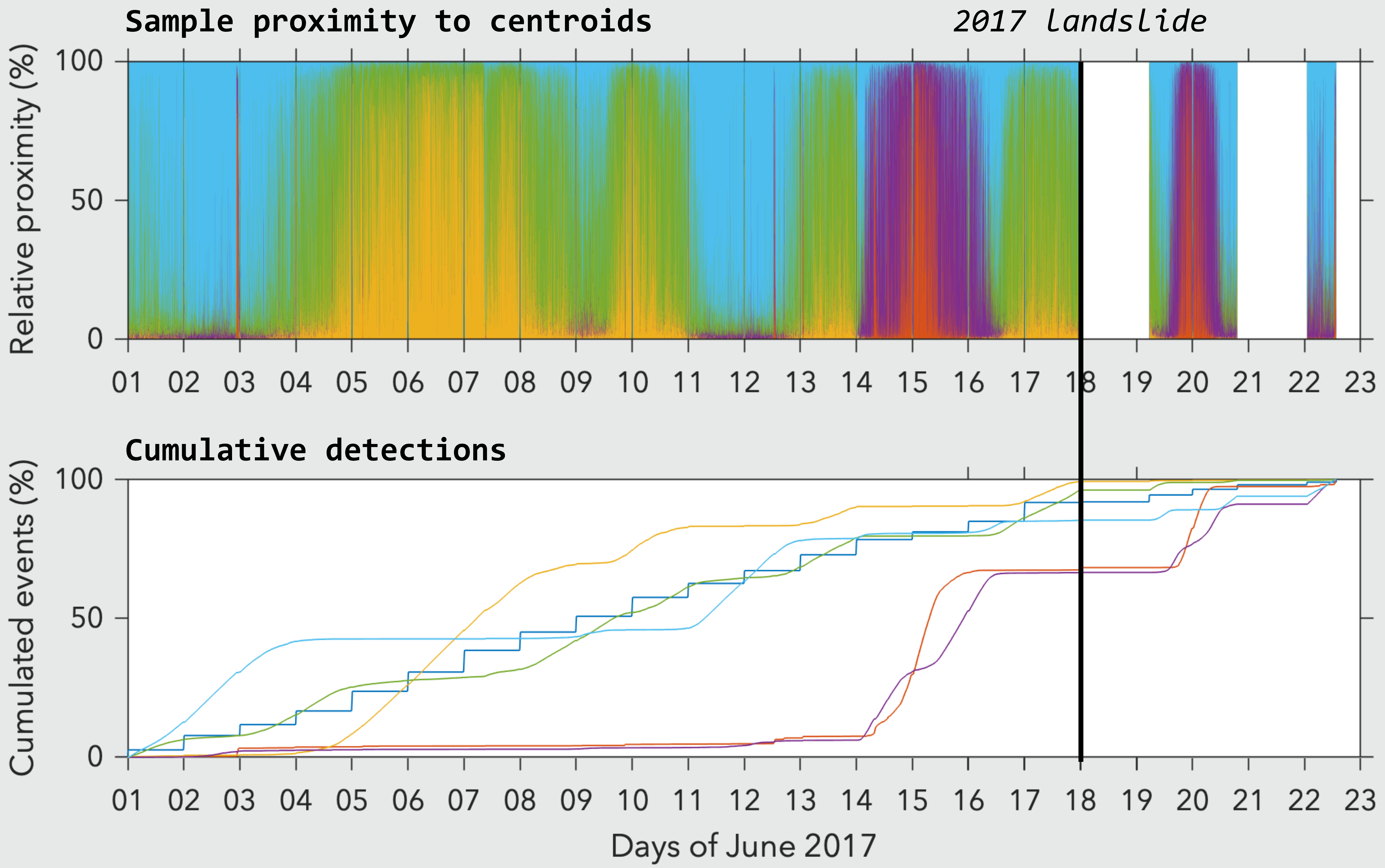
Poli (2017)

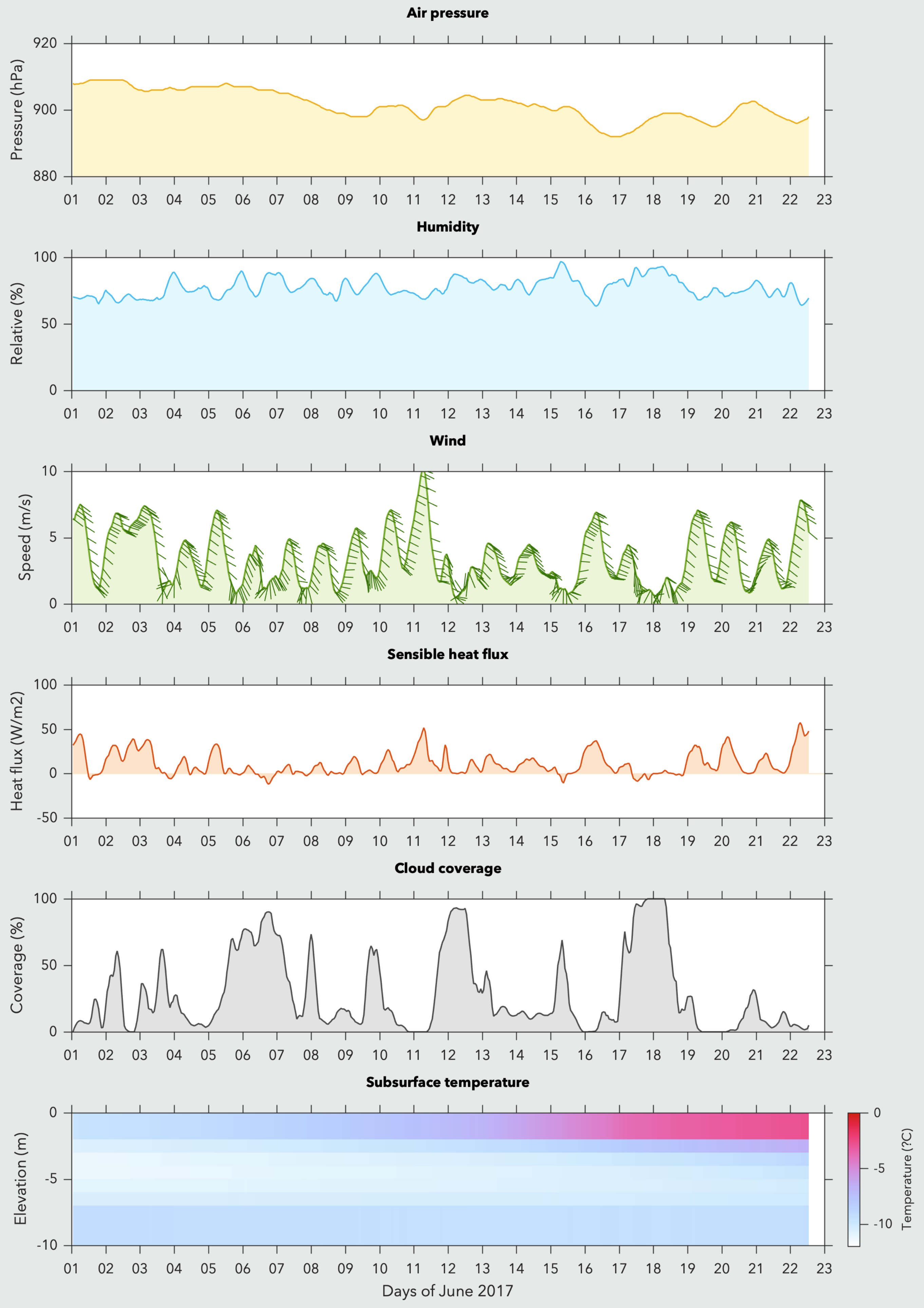


Seydoux et al. (soon)

Identification statistique du template trouvé précédemment

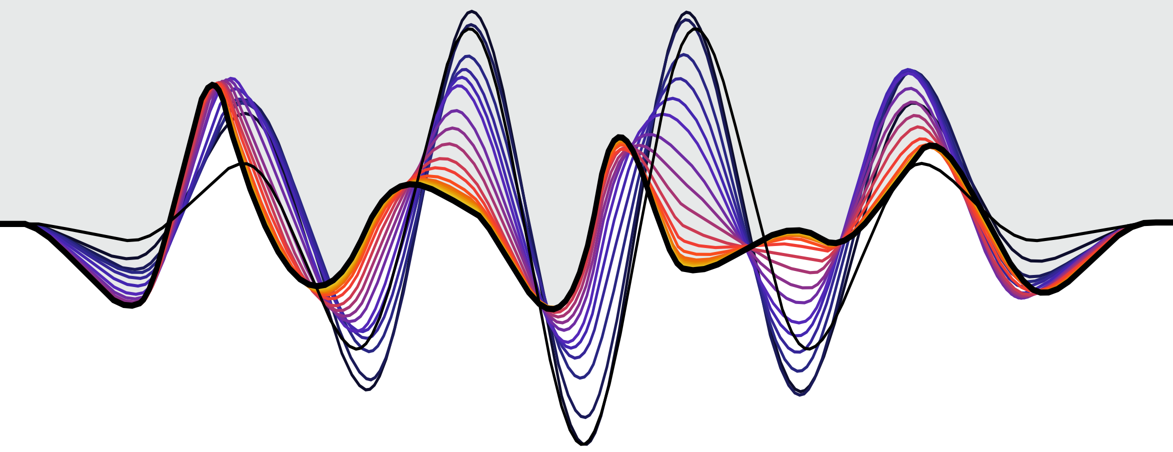
Discussion – classification du bruit à de plus grandes échelles





On ne voit plus la (rare) sismicité, mais des structures relativement stables dans le bruit.

Conclusions



Scattering network is used as a **stable multiple time-scale** representation of the seismic data

PCA and GMM are used to **cluster** the seismic data in a two-dimensional space

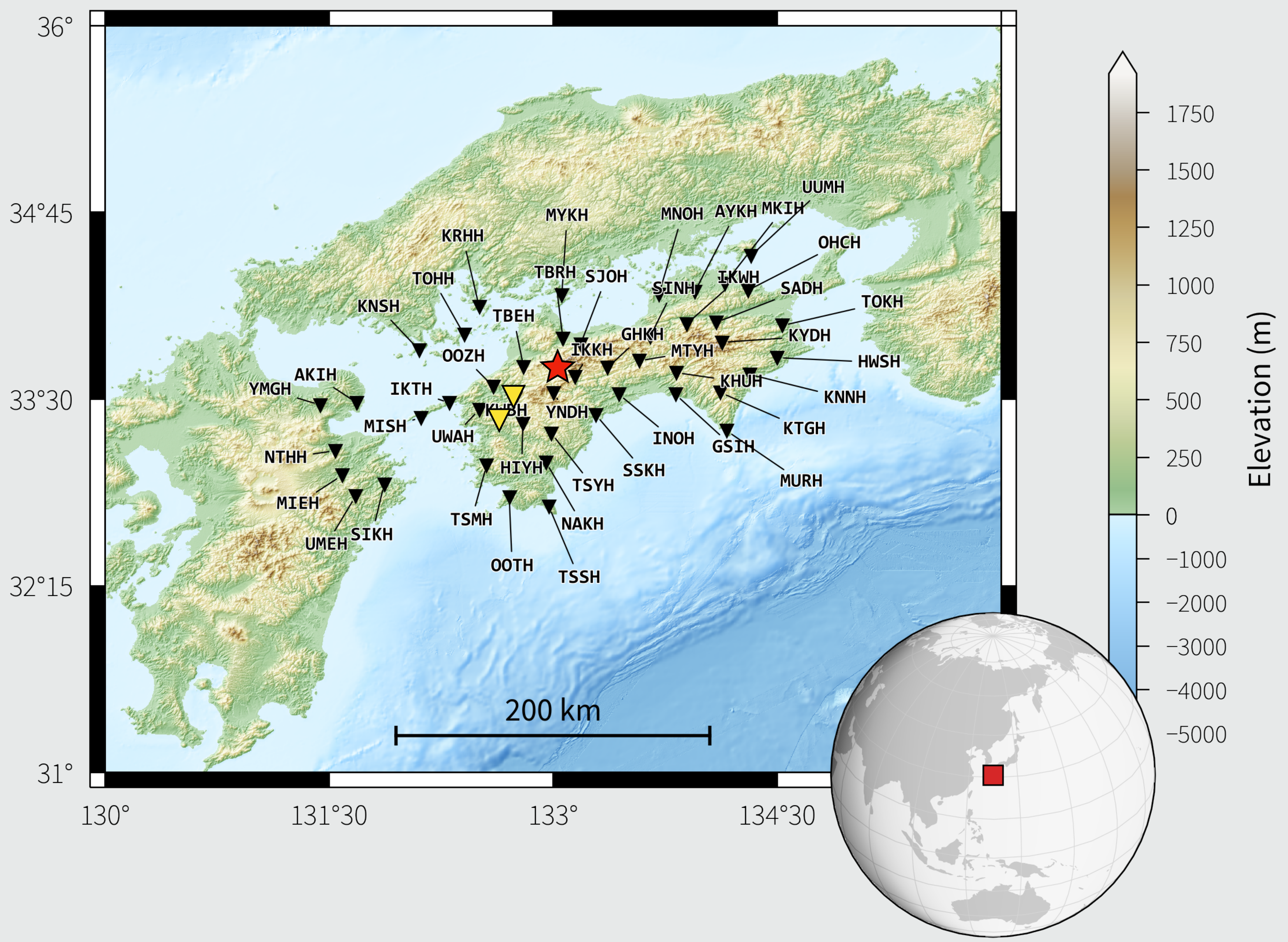
We **learn the wavelet** that minimizes the clustering loss (representation learning)

We were able to **blindly recover the precursory repeater** preceding the main landslide rupture

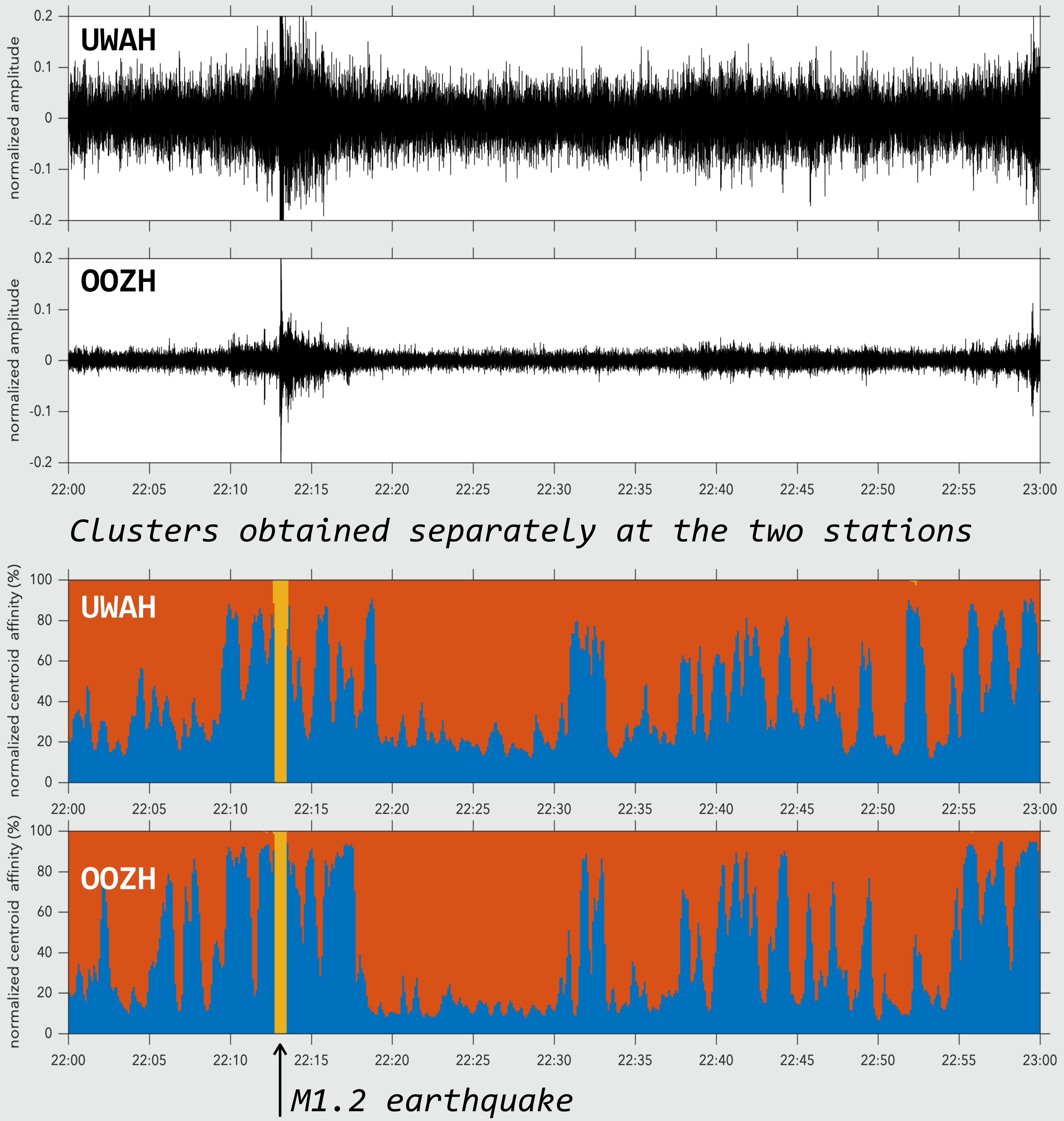
Annexes

Discussion – towards single-station detection of non-volcanic tremors

Japan hi-net

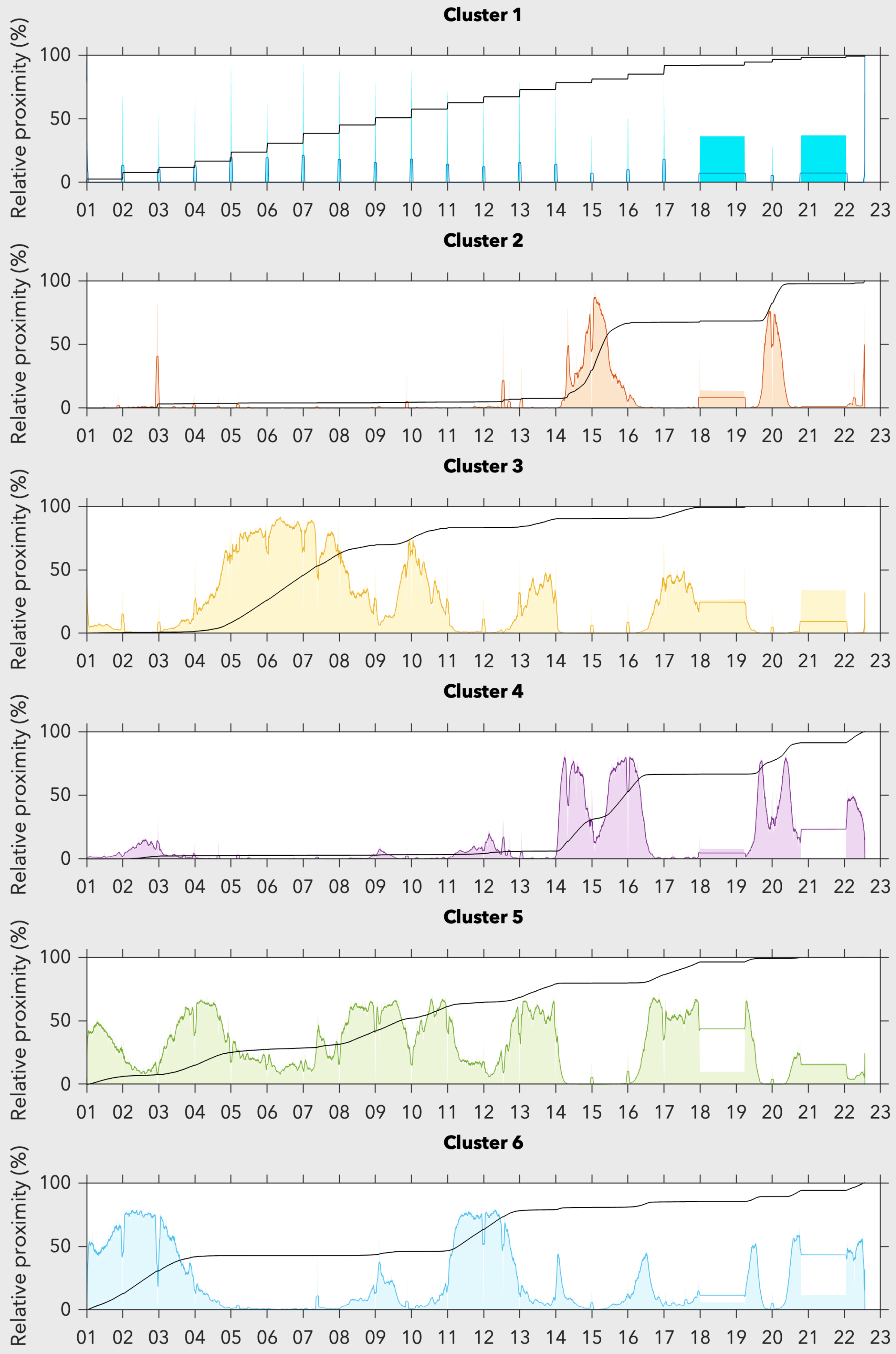


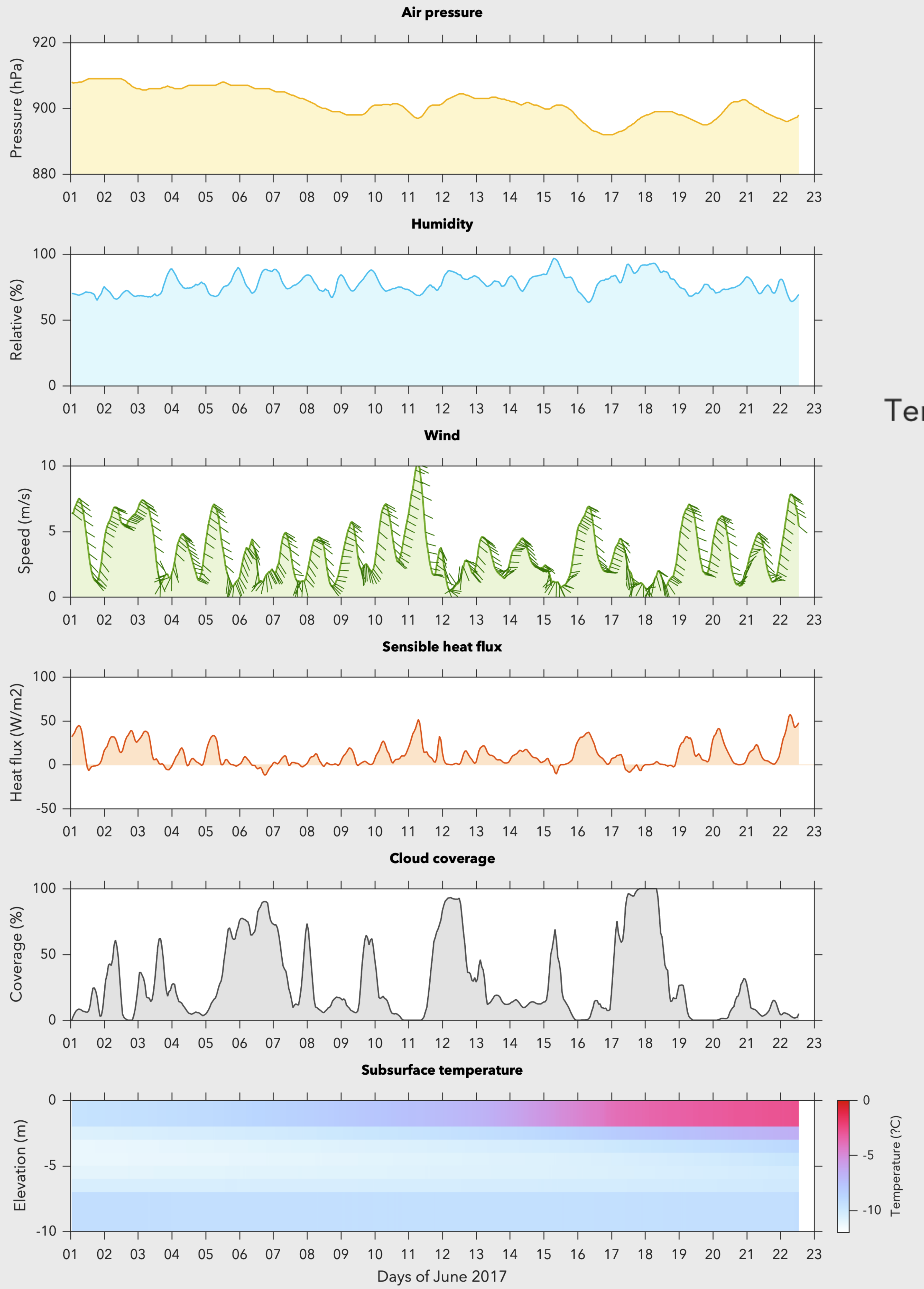
Broadband records at two stations located 50 km apart

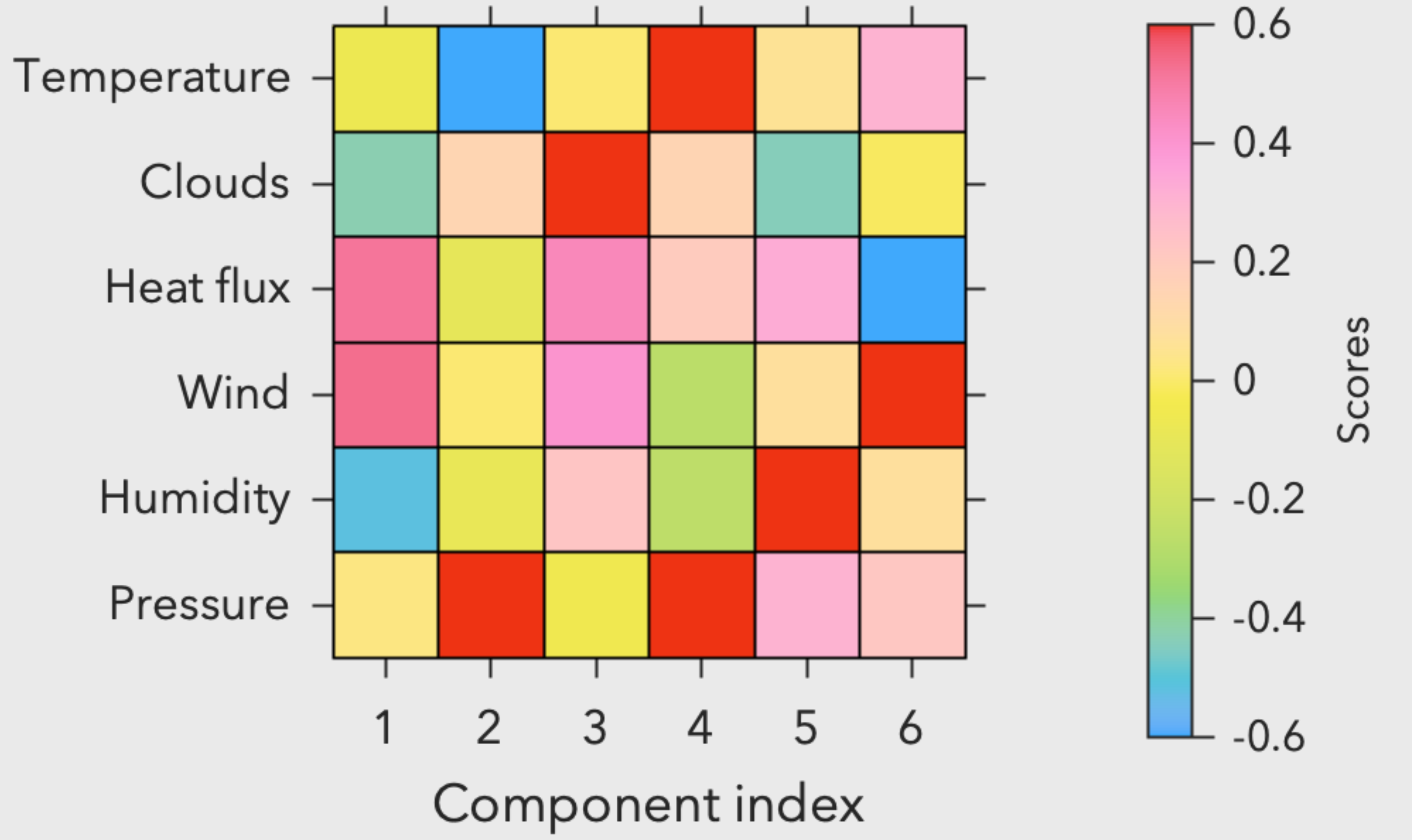


Two continuous records independently analyzed lead to the same clusters

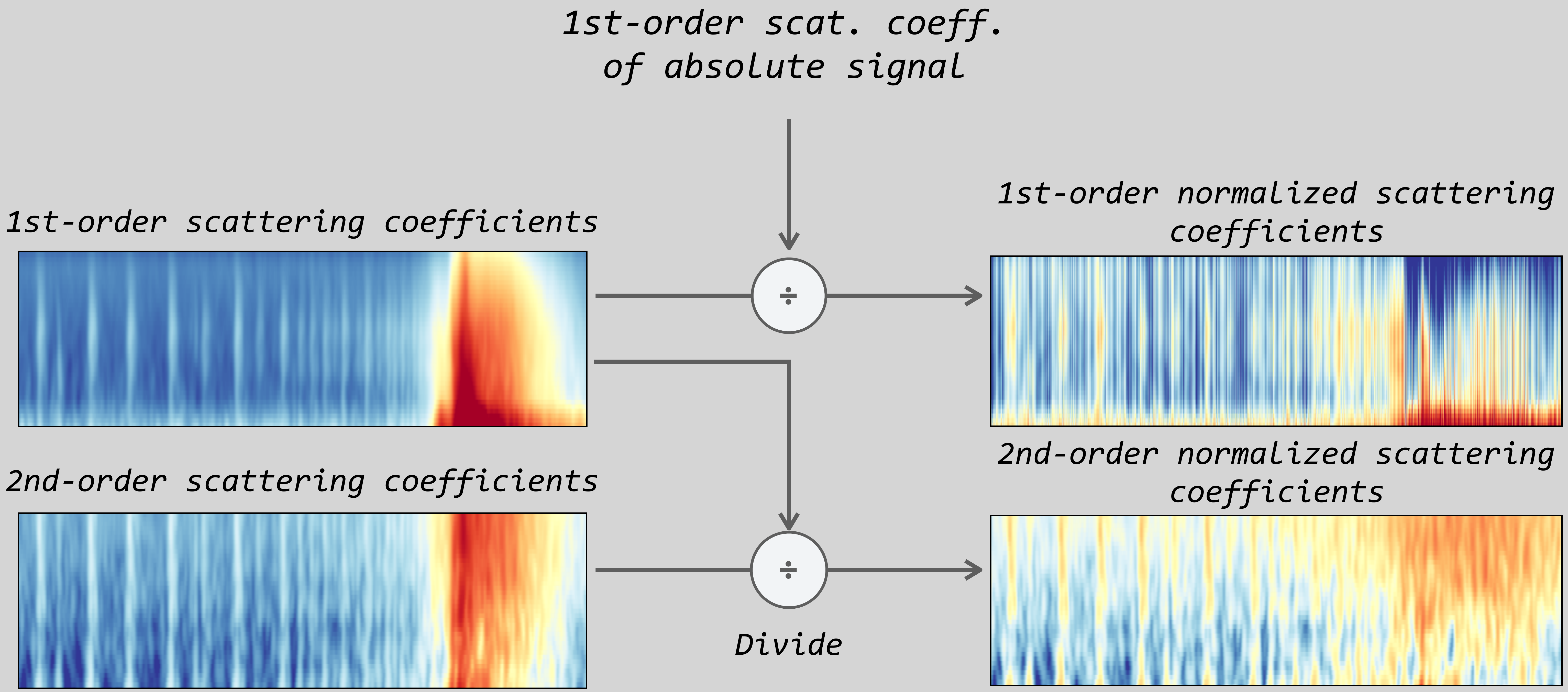
Discussion – clusters versus meteorological data







Appendix – parental normalization of the scattering coefficients



Several order of magnitude of amplitude difference between signals in the seismic data.
We normalize the amplitude w.r.t. the parent scattering coefficients.

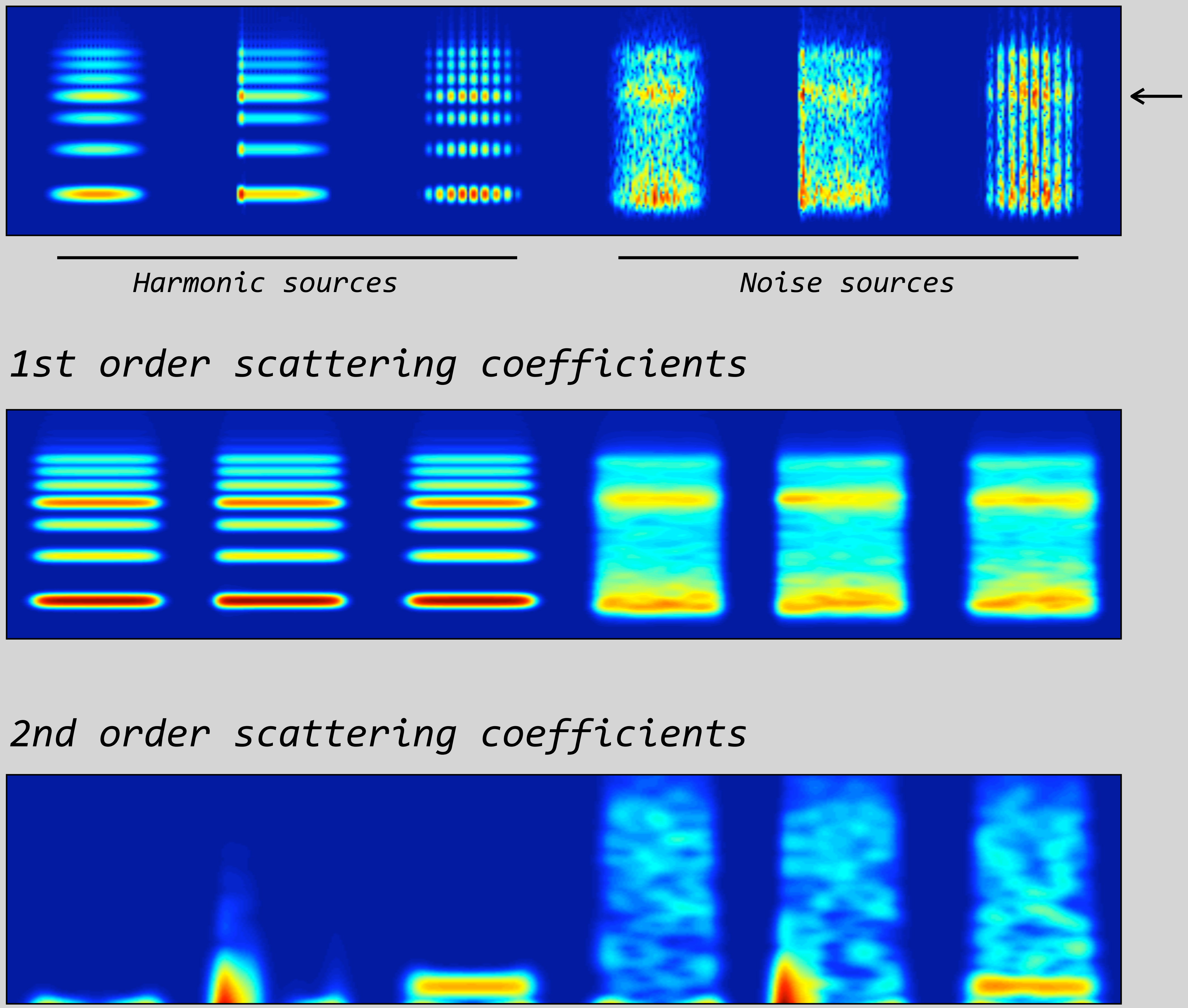
Siffre et al. (2013)

Toy example: **a two-layer scattering network**

Scalogram

1st order scattering coefficients

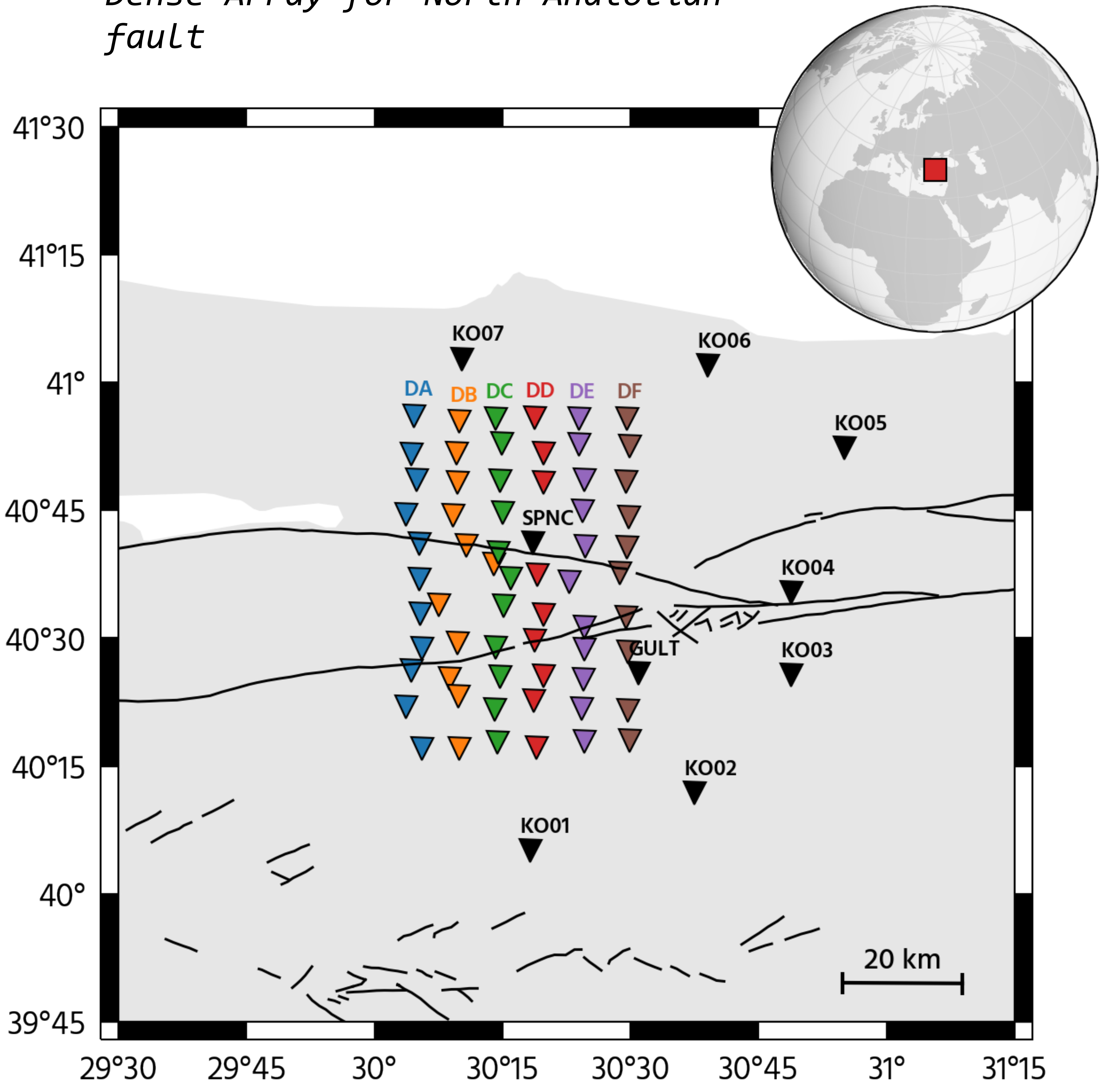
2nd order scattering coefficients



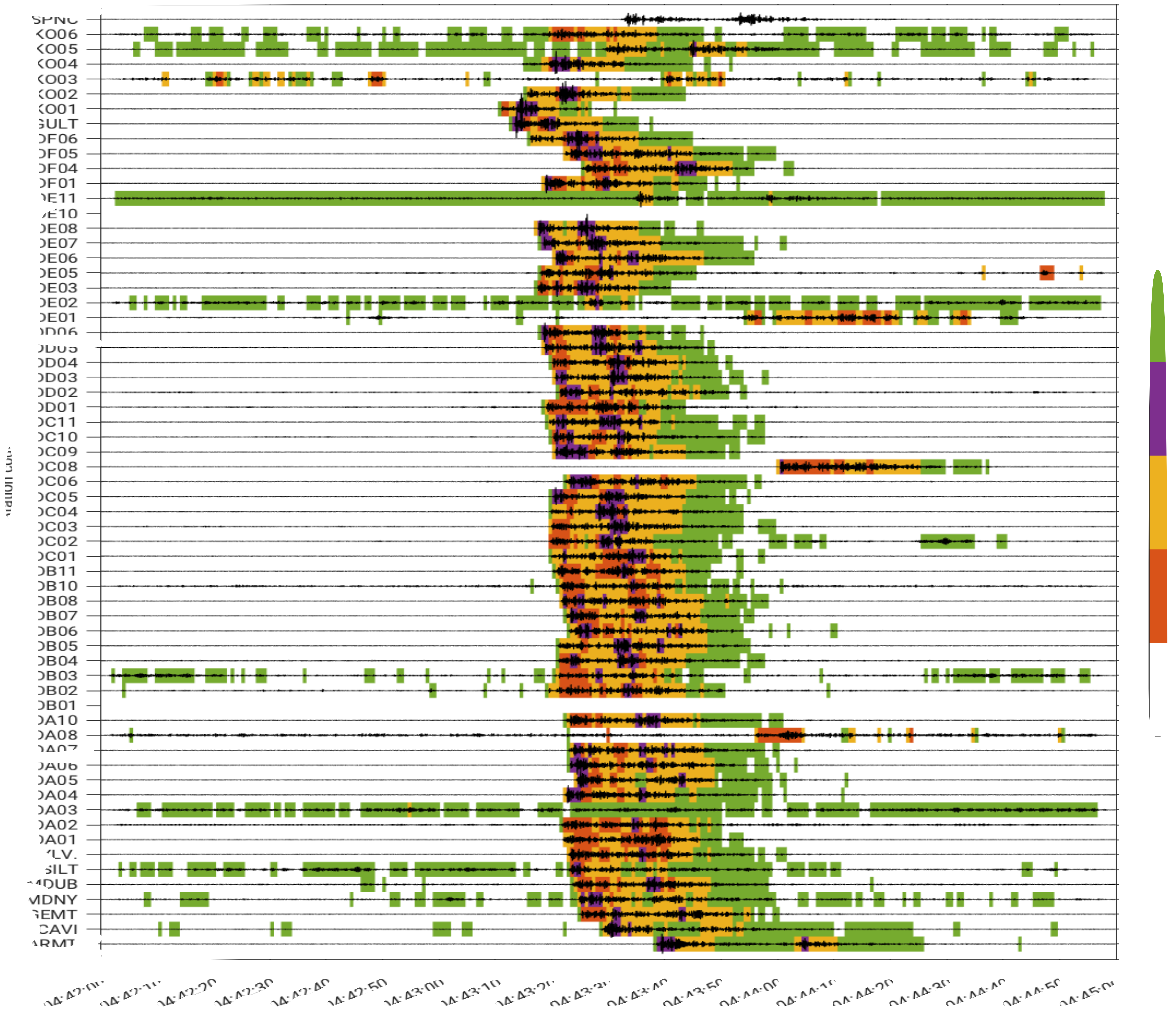
Anden & Mallat IEEE (2014)

Ongoing work – differentiate between seismic phases

Dense Array for North Anatolian fault

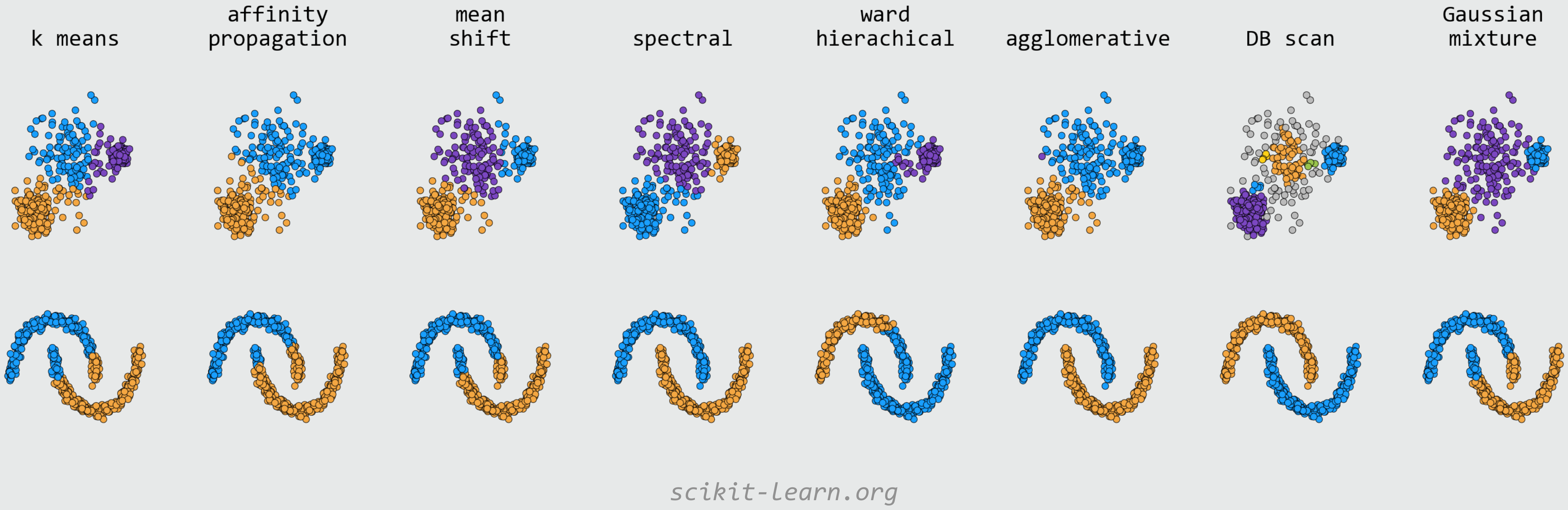


Analysis of a M1.6 earthquake



C

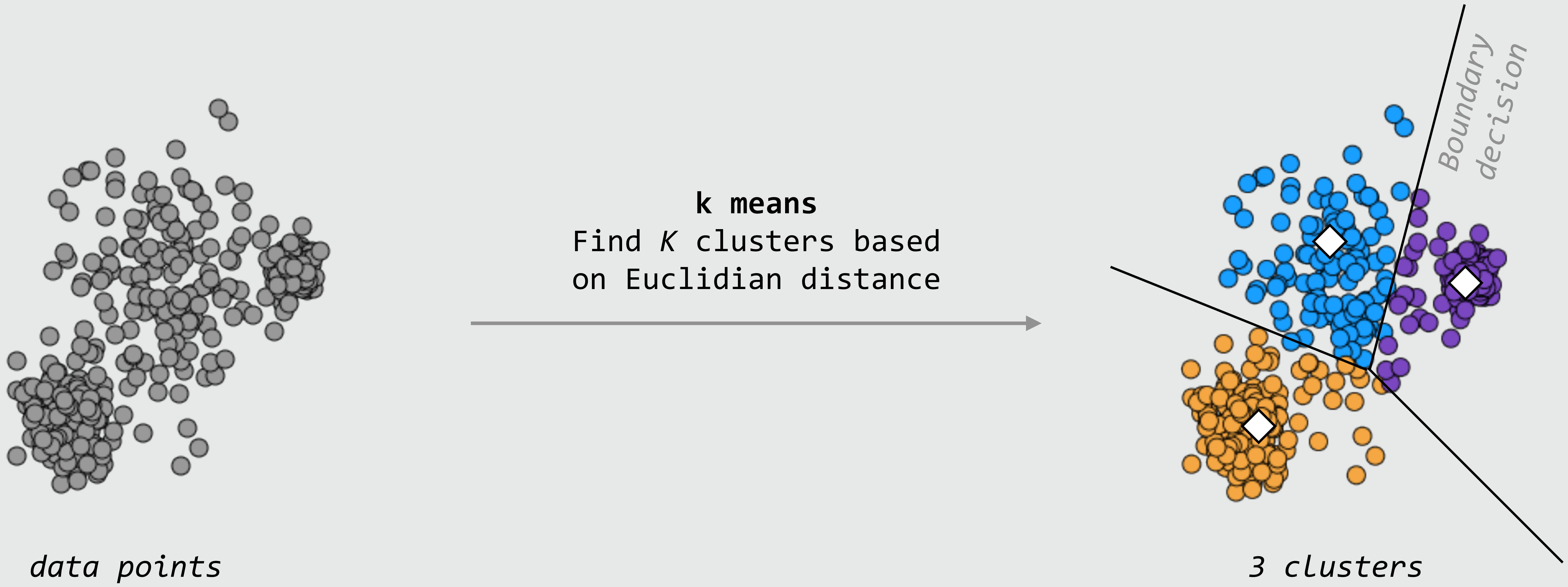
Cluster analysis – pick up the right one!



Diversity of definition leads to variety of algorithms

We need **data experts** to have a priori on the data in order to select the right algorithm

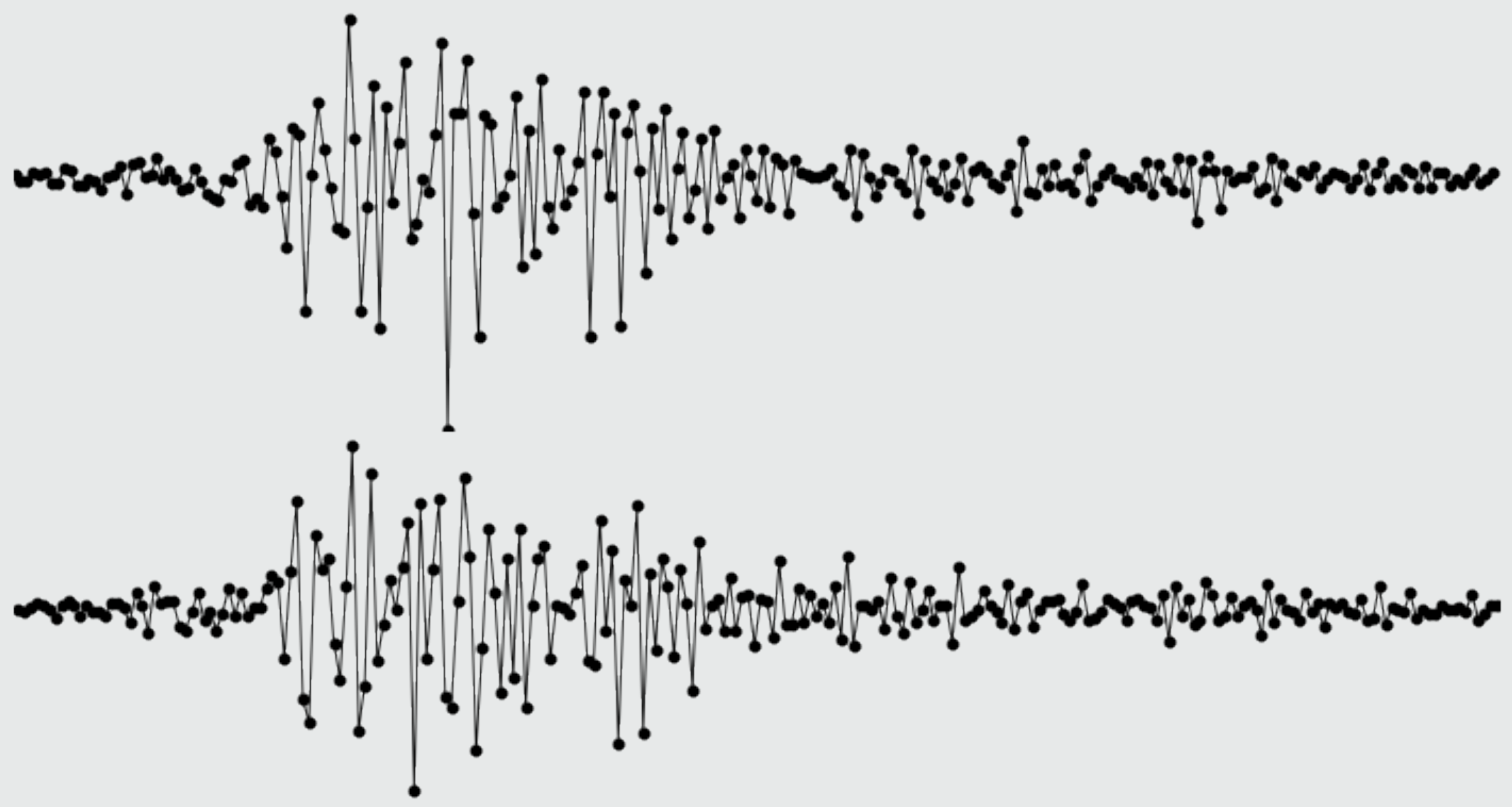
Cluster analysis – example of similarity-based clustering



Which algorithm is best suited for your dataset?

Waveform clustering

How can we consider waveform data?



N-points waveform
correlation: 32% !

A waveform is a point in a *N* dimensional space

Time-domain representation is highly unstable (sensitive to translation in time, amplitude, frequency, etc.)

We need to extract **features** that have some properties of invariance

Waveform clustering

General workflow

Waveform		Features		Clustering
$x \in \mathbb{R}^N$	→	$x' \in \mathbb{R}^F$	→	$c \in \mathbb{N}^C$

Dependent to:
Translation in time
Deformation (scattering)
Frequency content
Not suited for clustering

Which features have invariance properties?

Find groups based on similarity in the feature similarity

How do we select the right ***features***?

Neural networks

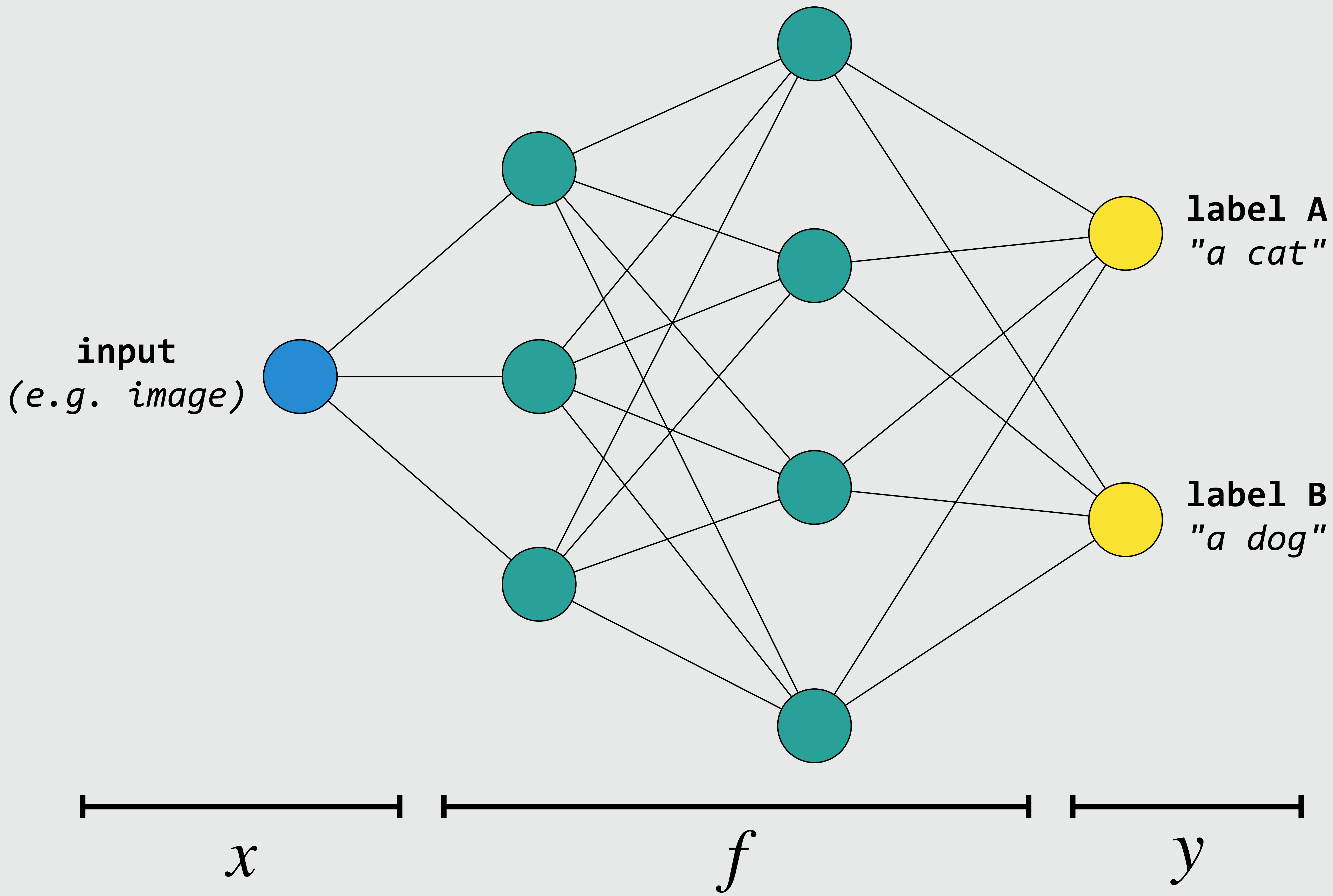
$$f(x) = y$$

"a cat"

Neural networks

$$f(x) = y$$

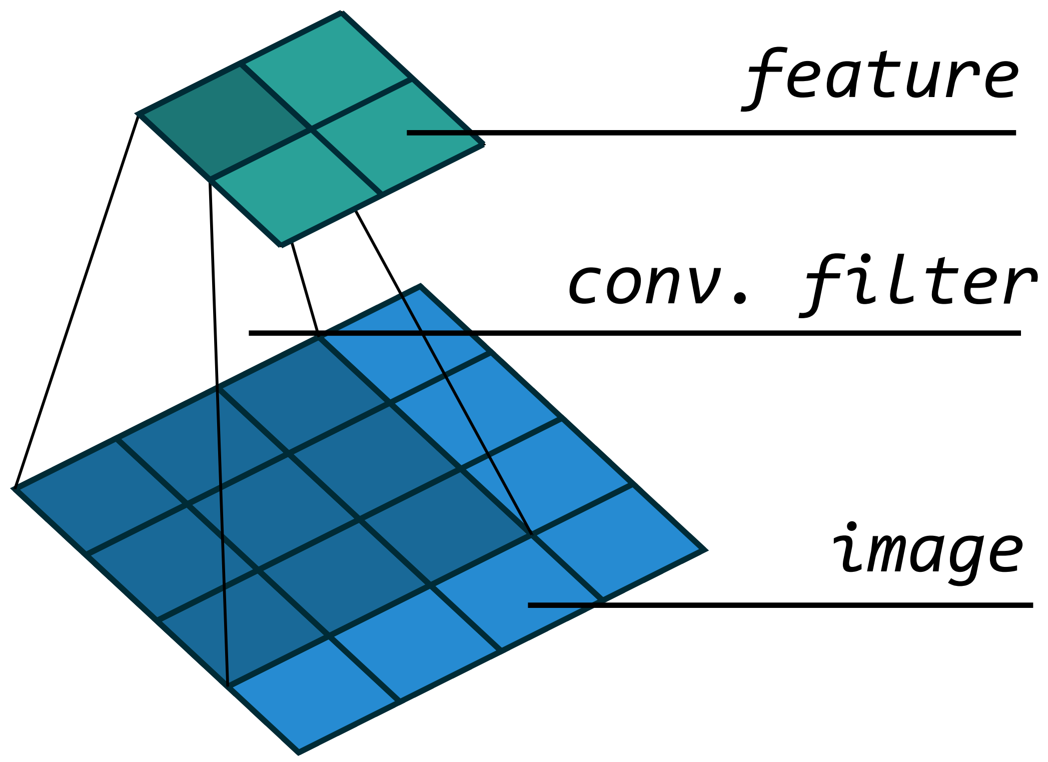
"a cat"



Neural networks can **approximate highly non-linear** functions

Neural networks

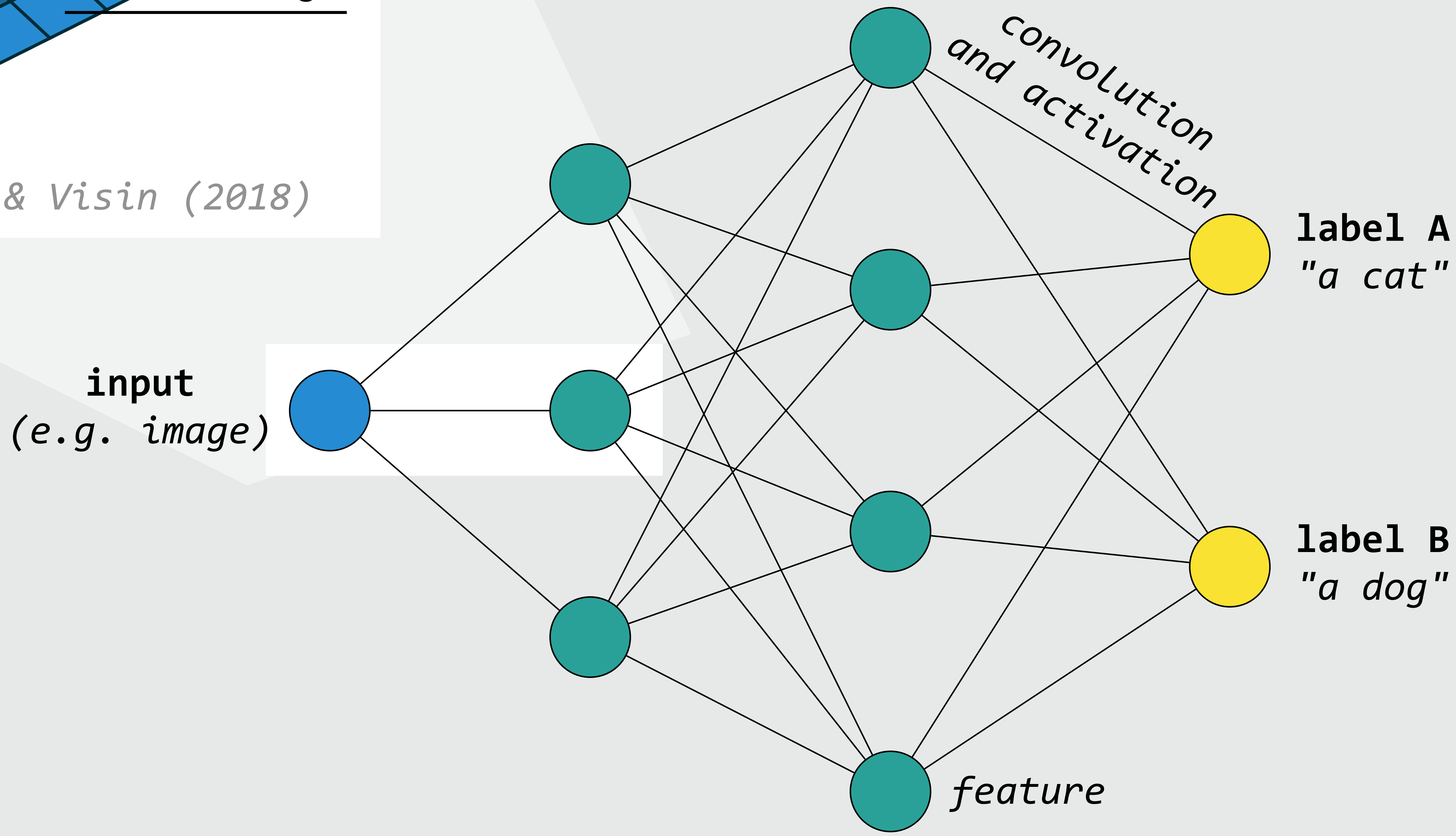
Learn to recognize patterns



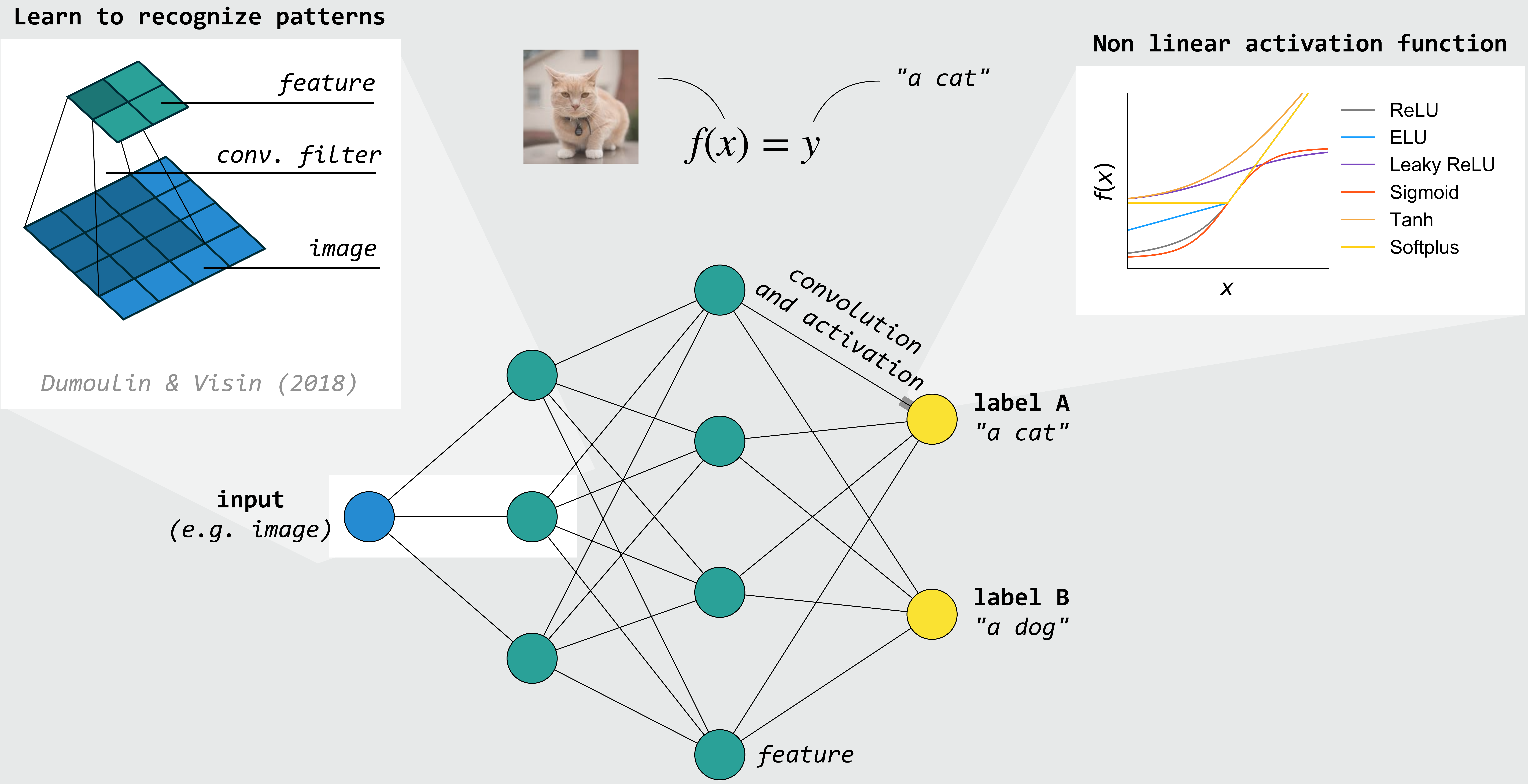
Dumoulin & Visin (2018)

$$f(x) = y$$

"a cat"



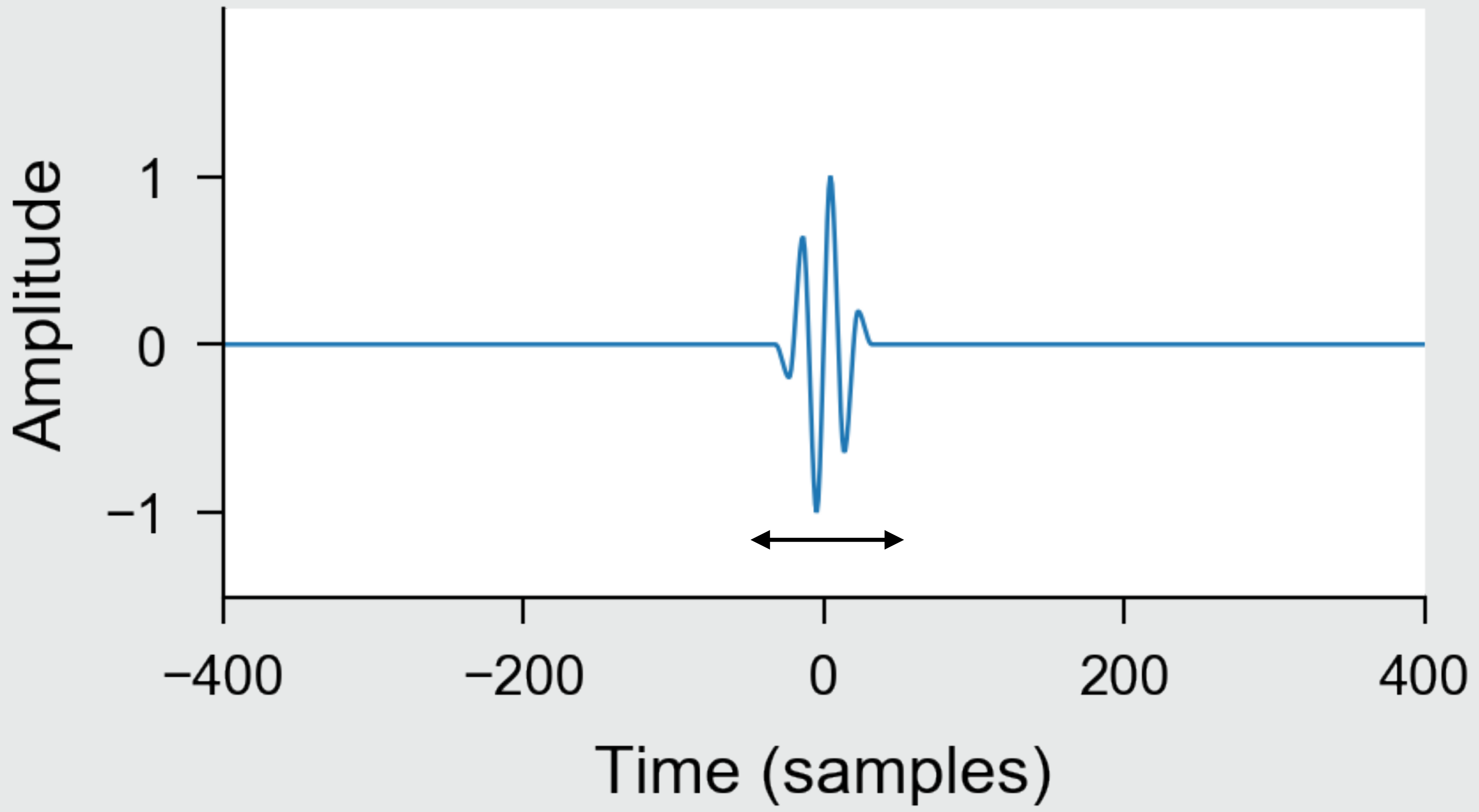
Neural networks

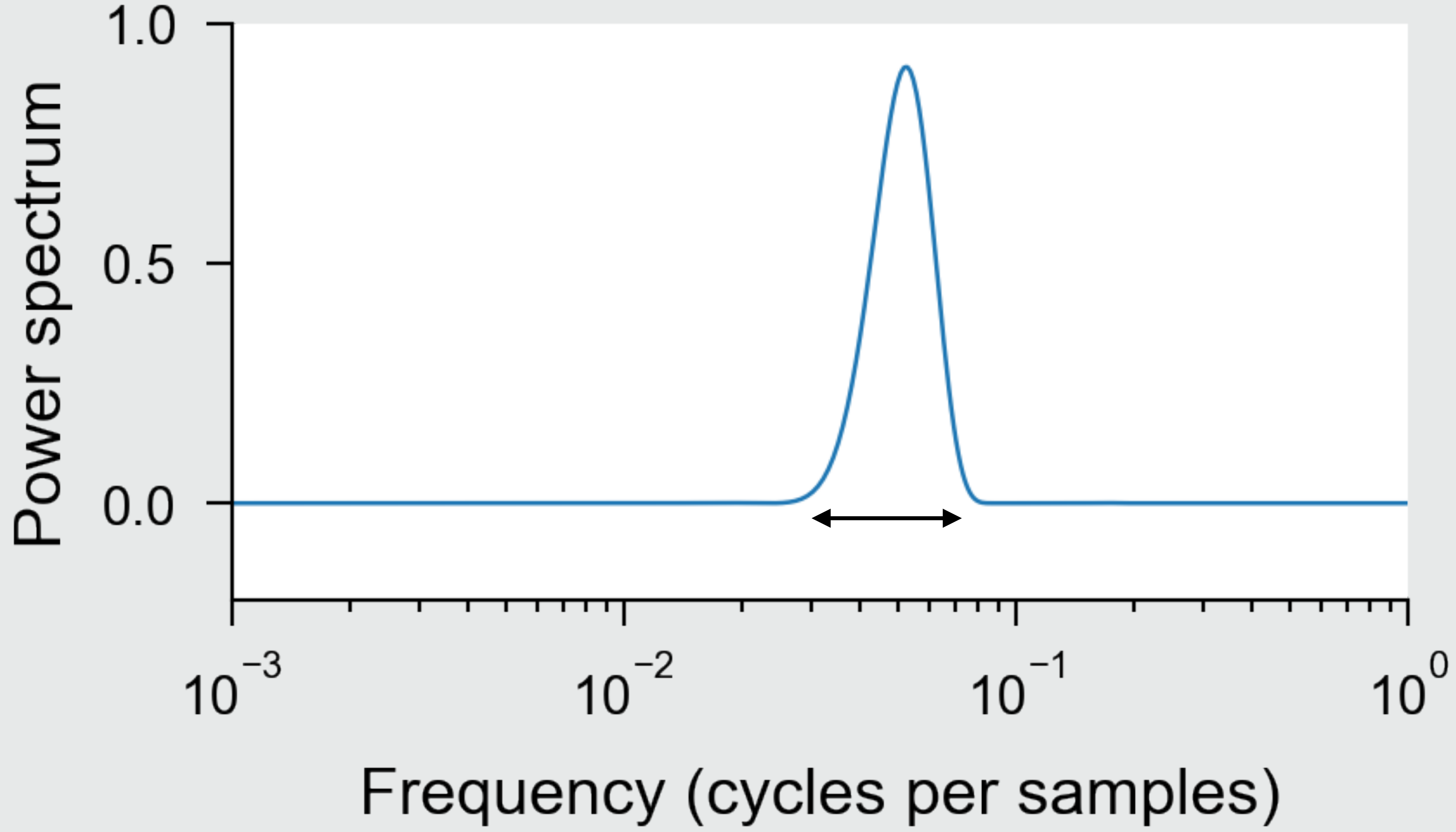


Wavelet transform

Explore the **time and frequency** content of a one-dimensional signal with convolution with different wavelets localized in time and frequency

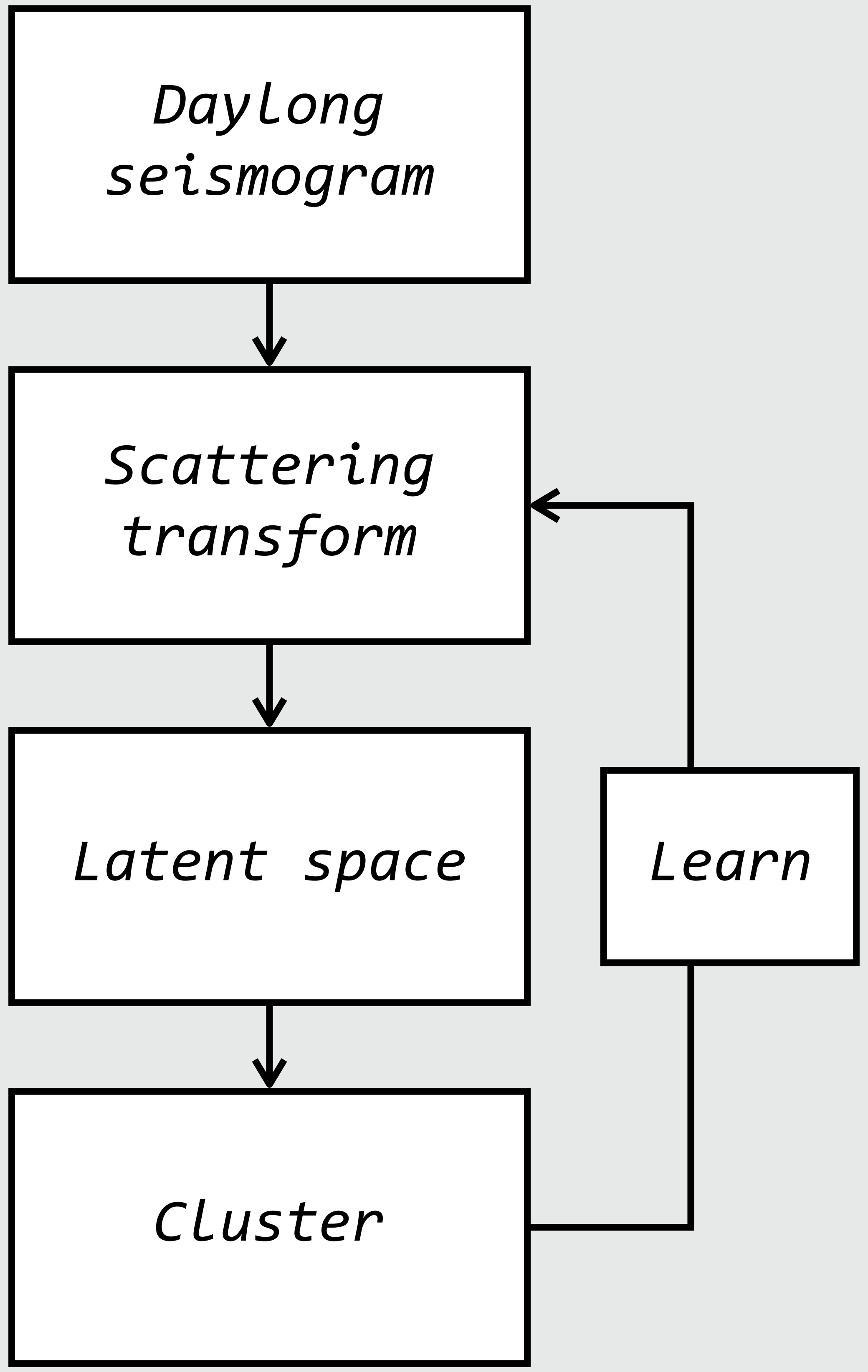
$$Wx(\lambda, t) = (\psi_\lambda \otimes x)(t)$$



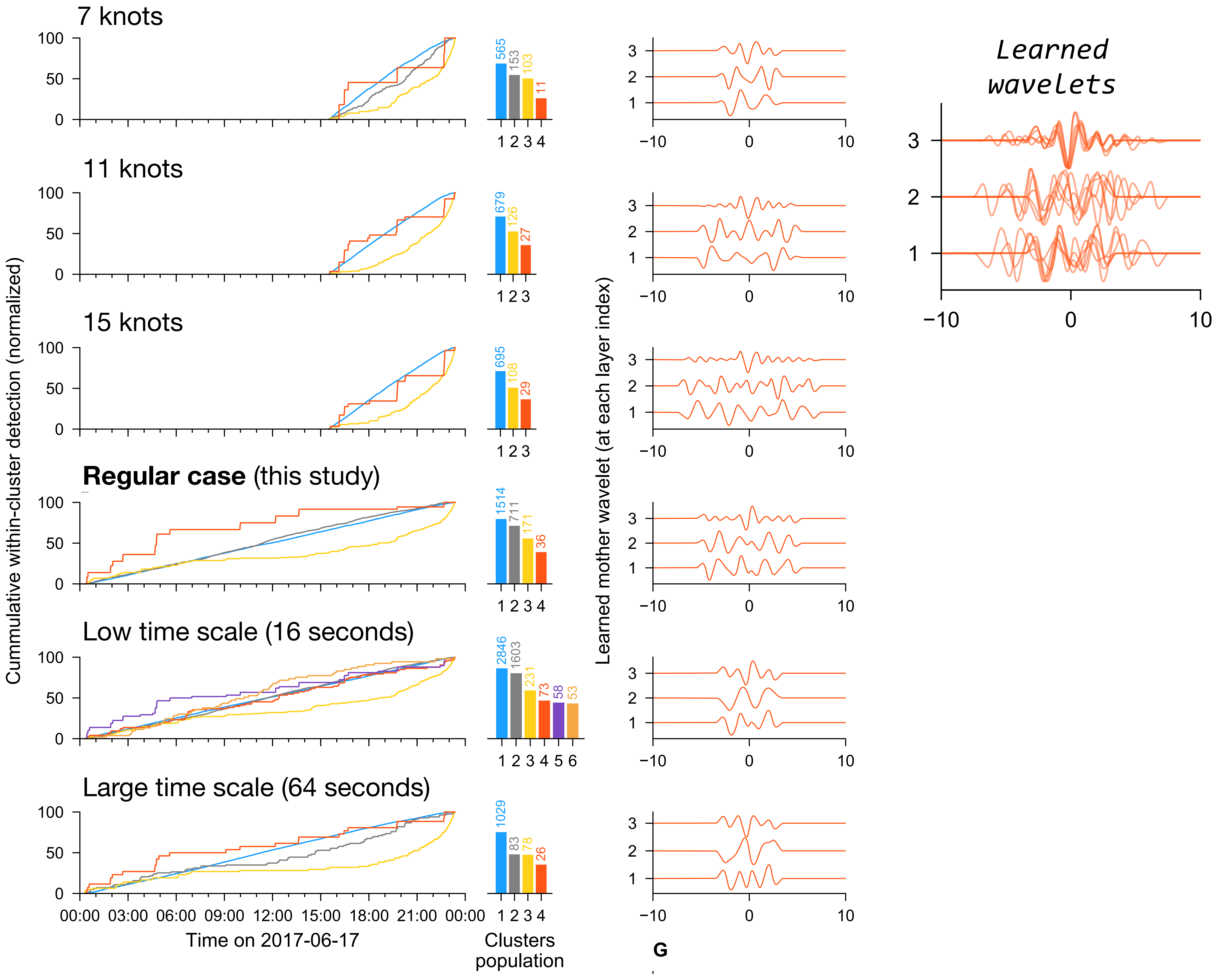


Wavelets are localized in time and frequency

Robustness to scattering network design

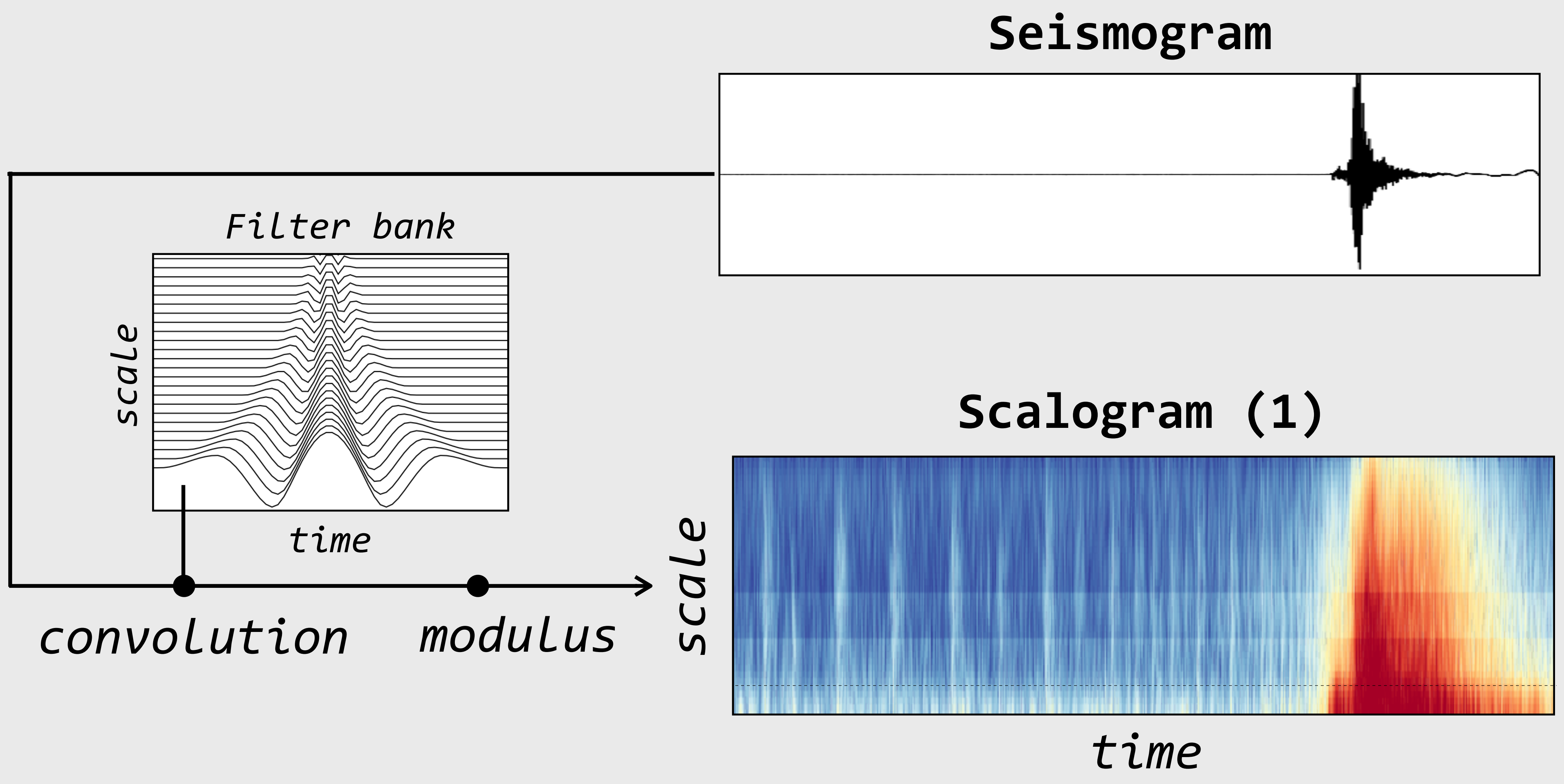


Seydoux et al. (rev.)



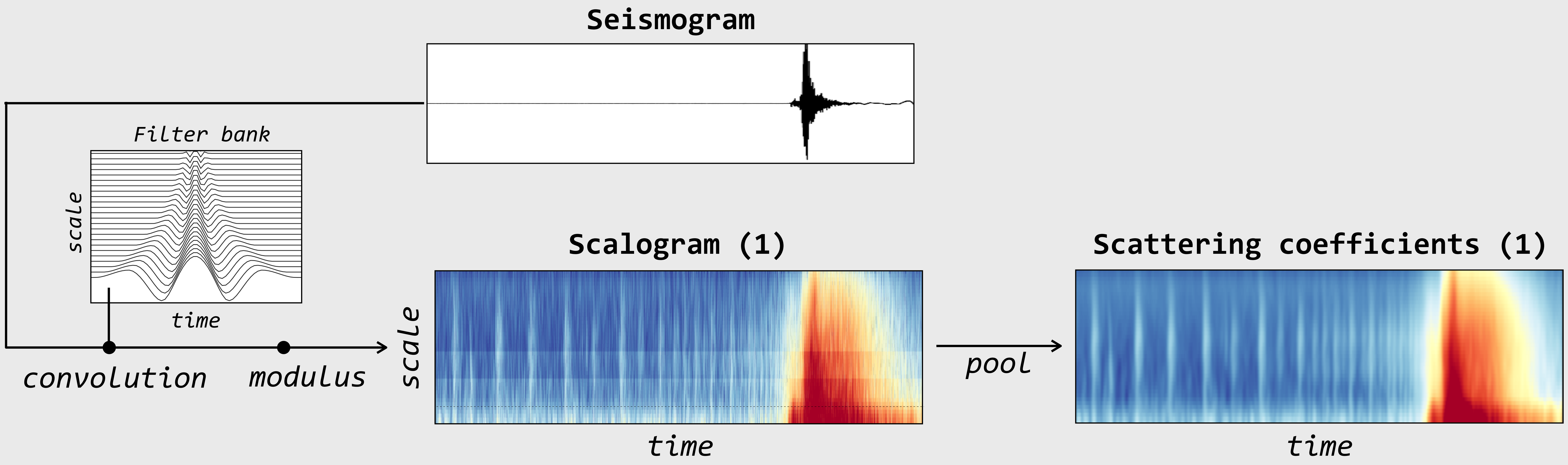
Different parameters always recover the precursory pattern

Idea of a scattering network



The first layer is a time-frequency representation of the waveform

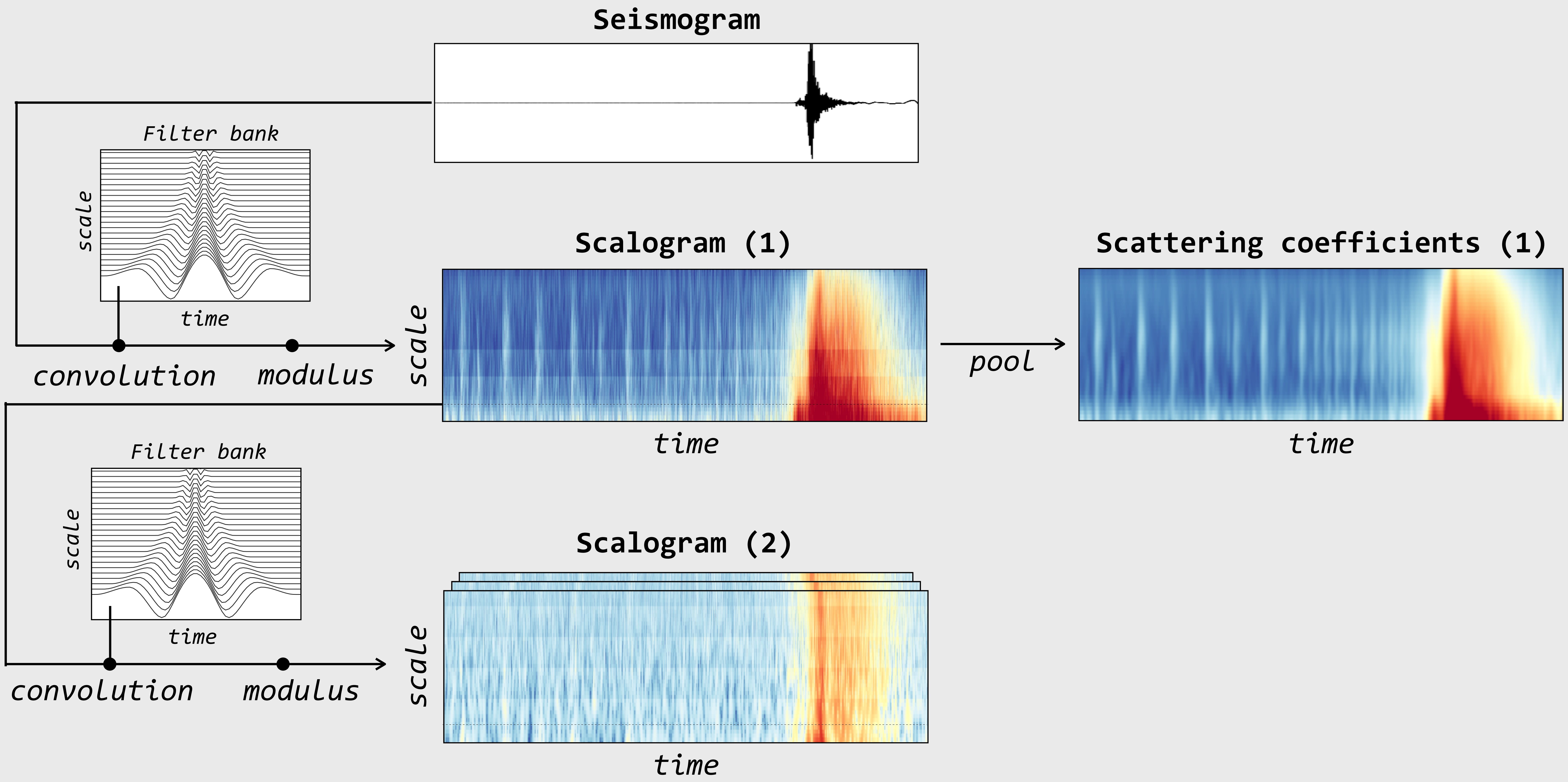
Idea of a scattering network



The first layer is a time-frequency representation of the waveform

The first-order scattering coefficients provide a **locally stable signal description** at small time scales.

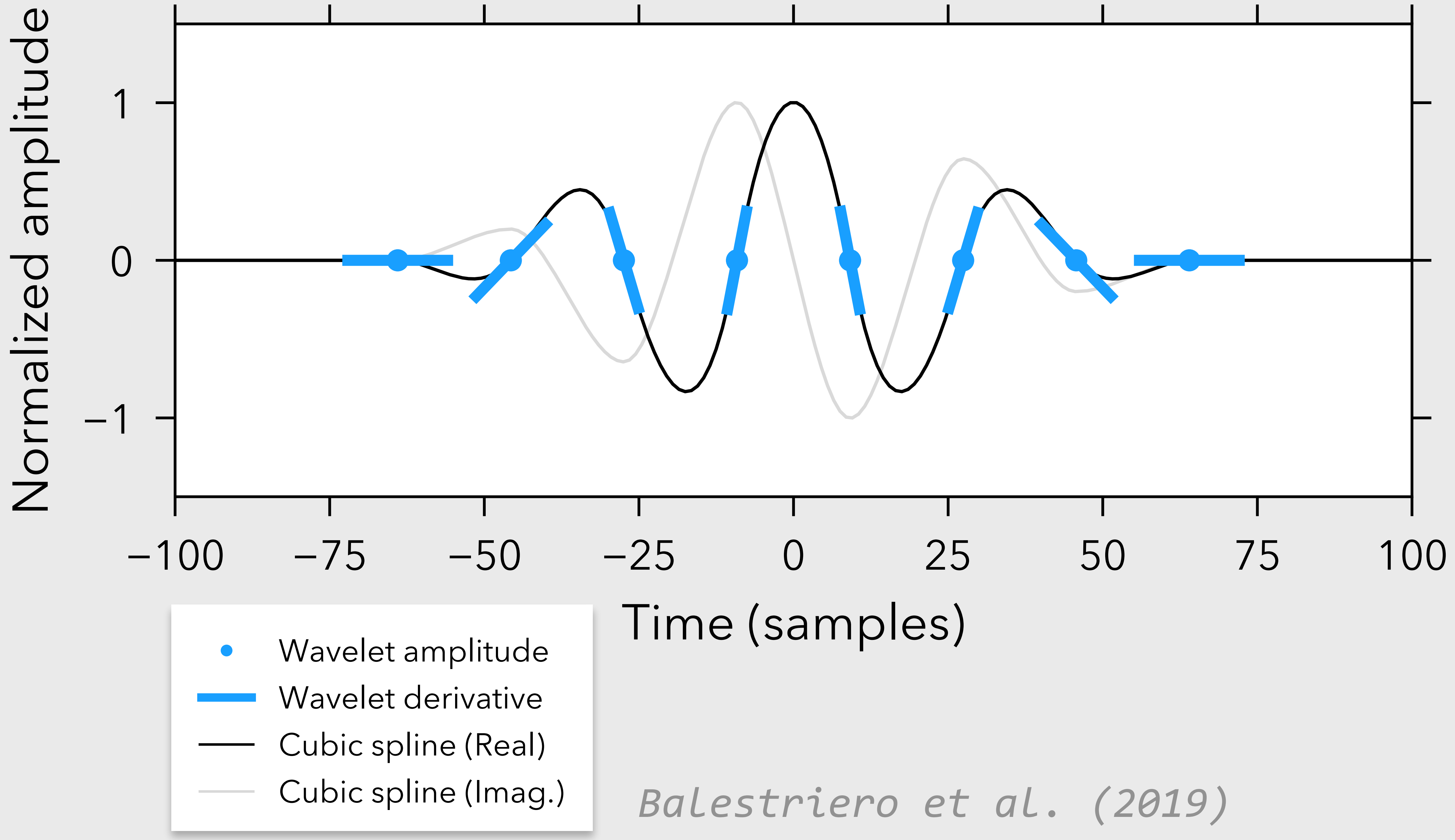
Idea of a scattering network



Larger time scales are analyzed at second order

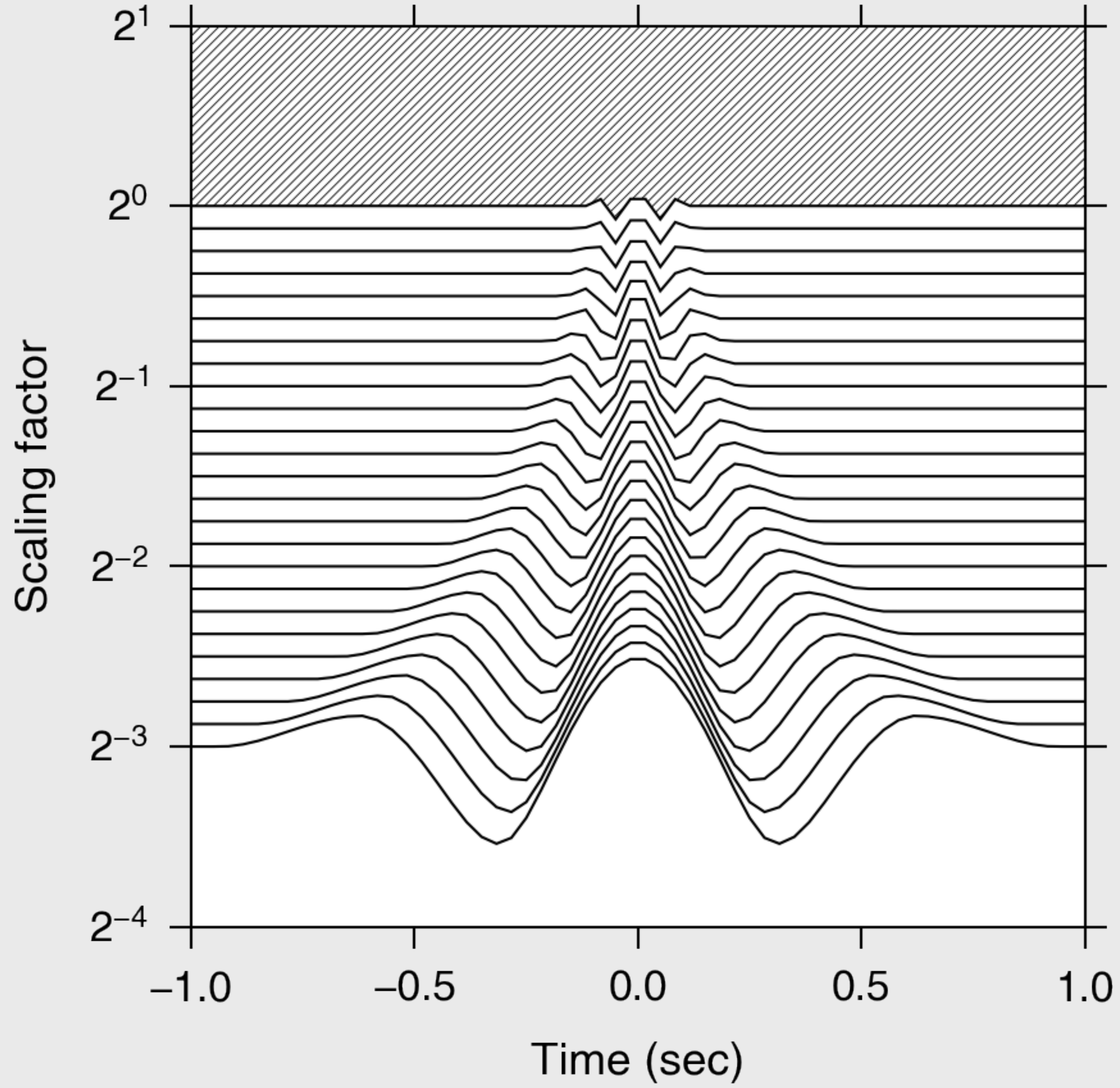
Learnable wavelets from Hermite cubic spline interpolation

1. Amplitude and derivative learned at knots
2. Full wavelet interpolated with cubic splines



Balestriero et al. (2019)

3. Filter bank obtained from dilation of the mother wavelet



We can learn the wavelets given any task (e.g. clustering, classification, …).
Only a few coefficients are learned compared with classical convolutional nets