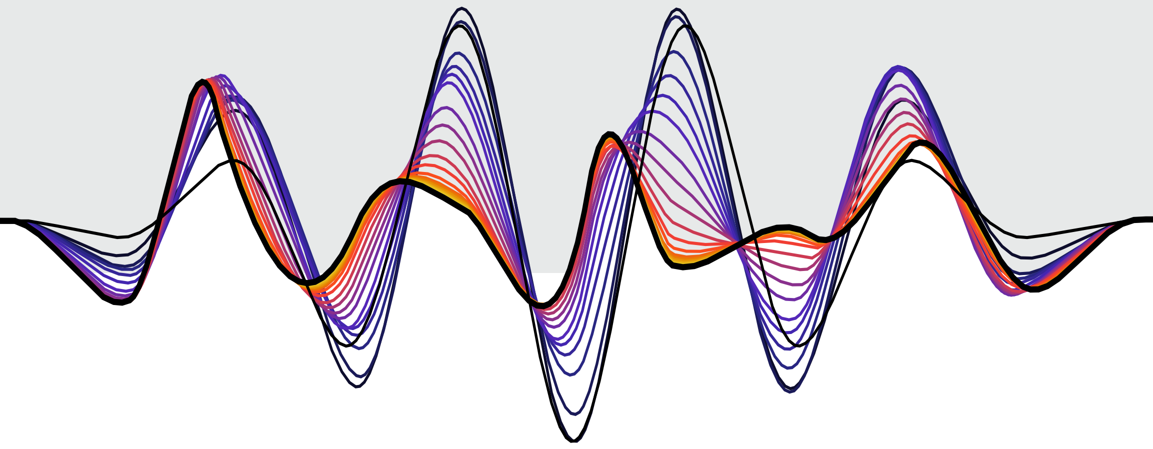
Unsupervised clustering of continuous seismograms with deep learning

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

leonard.seydoux@univ-grenoble-alpes.fr

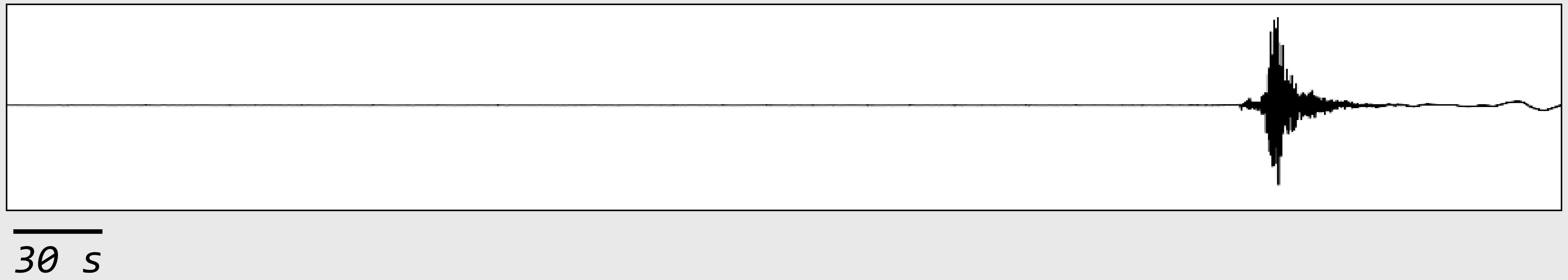
1. ISTerre, Grenoble, France 2. Electrical and Computational Engineering and
3. Computational and Applied Mathematics, Rice University, Houston, TX



Cargese'19 passive imaging workshop

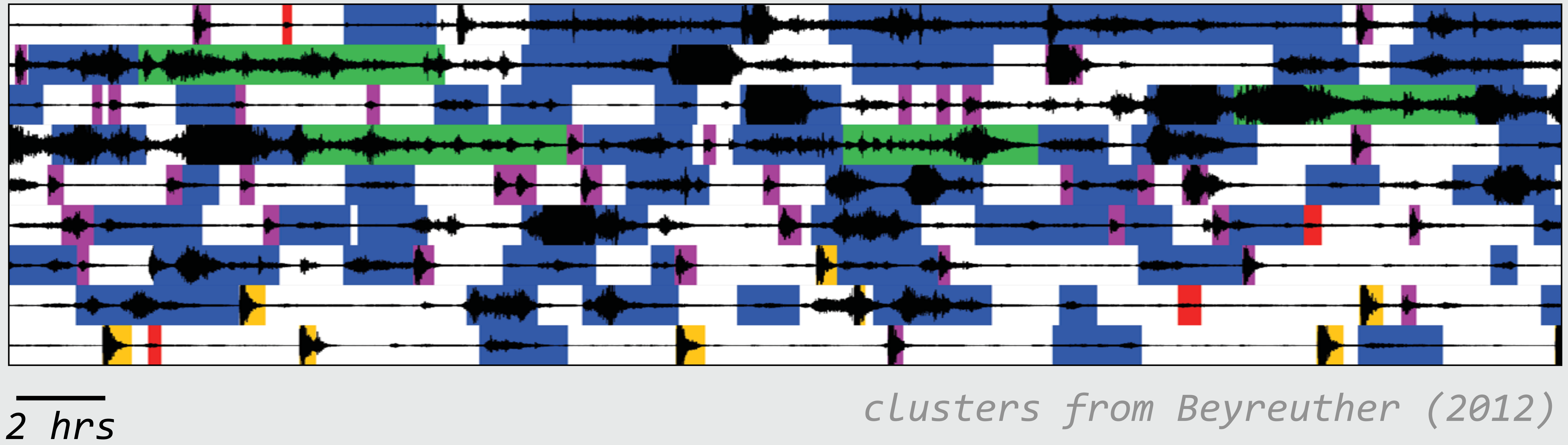
Motivations – blind exploration of seismic data

Highlight any precursory seismic activity before rapid ruptures



Designing single-station detector of non-volcanic tremor

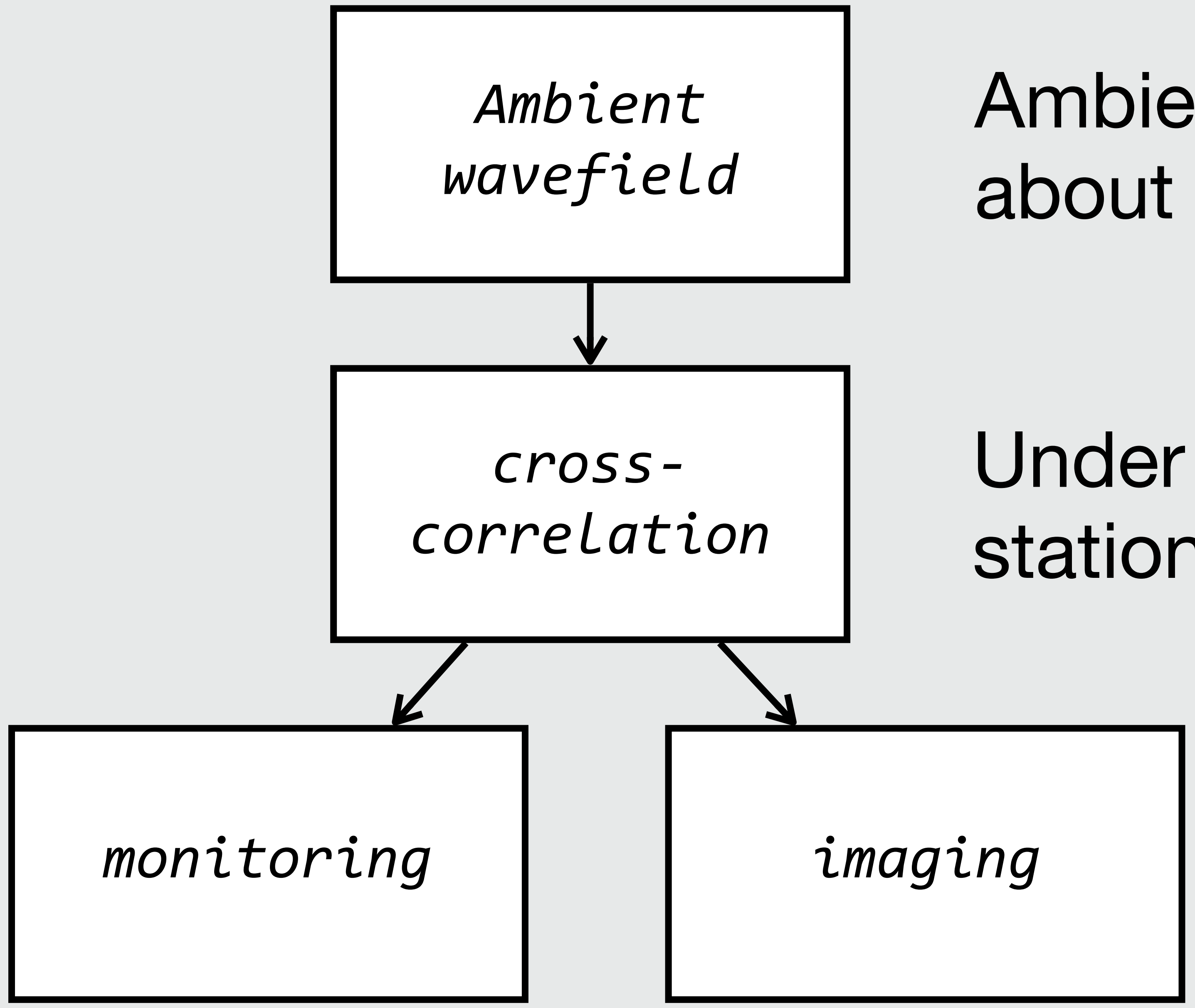
Labeling large datasets before passive experiments
Could we infer new classes of seismic signal?



clusters from Beyreuther (2012)

Seismic signals can have **multiple time scales**

Motivations – clustering ambient wavefield



Ambient seismic wavefield contains information about wave propagation within the Earth

Under assumption of equipartition, source randomness and stationarity, we can recover this information from cross-correlation

Any dominating source that may affect the results?

Have a better a priori on the wavefield content before passive experiment

Class-membership identification – the supervised and the unsupervised way

Supervised (classification)

Learn the non-linear mapping between the data and the labels

model

$$y = f(x)$$

label data

Unsupervised (clustering)

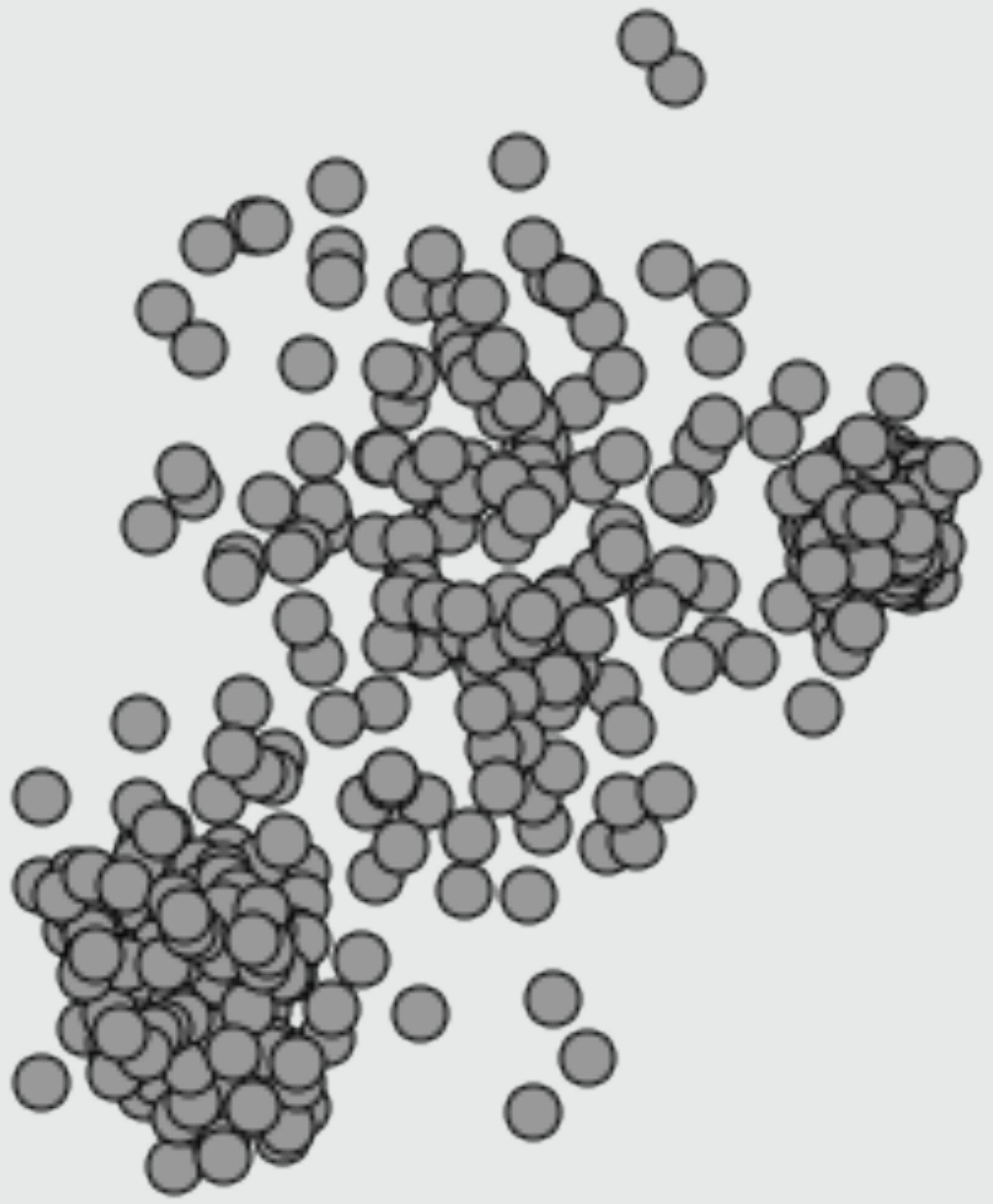
Understand the natural distribution or structure of the input data

model

$$\tilde{x} = f(x)$$

reconstruction of the data data

Cluster analysis – two most common definitions



`data points`

"Segmentation of a heterogeneous population into a number of more homogeneous subgroups (top-bottom)"

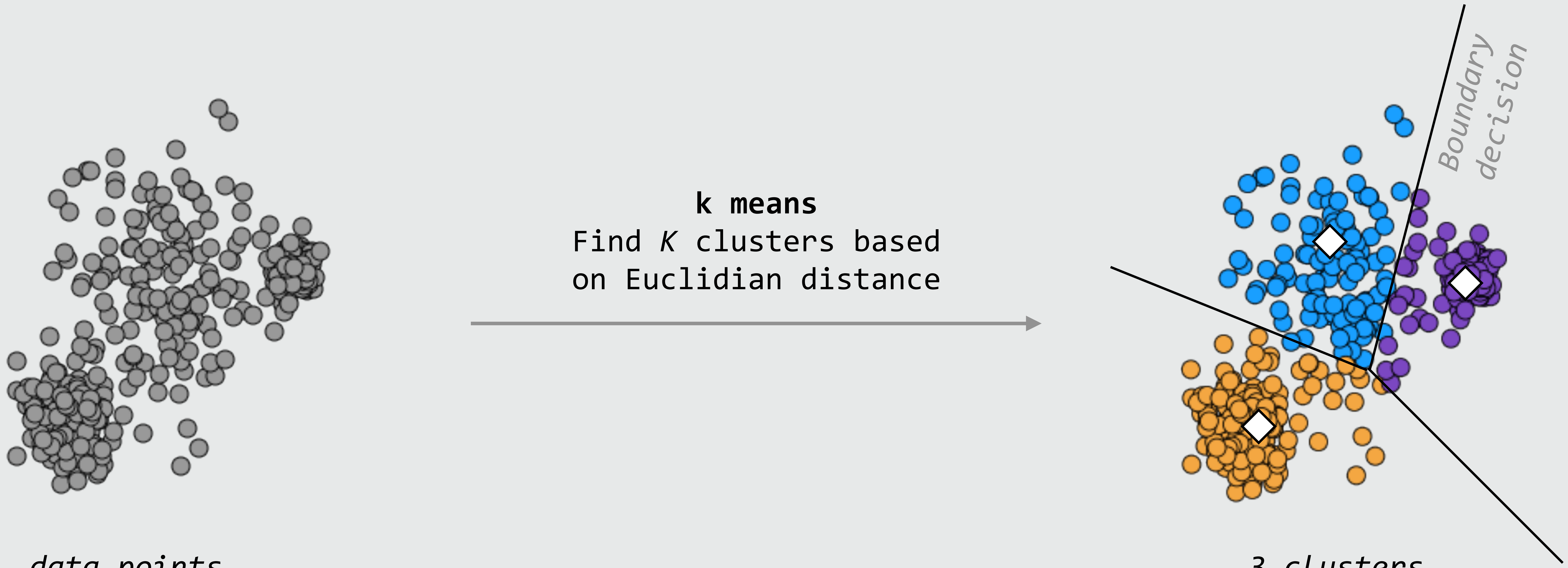
"Finding groups in a set of points by some natural criterion of similarity (bottom-top)"

Depending on the definition used, the techniques and results may vary

Clustering is an exploratory task, *every result make sense*

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

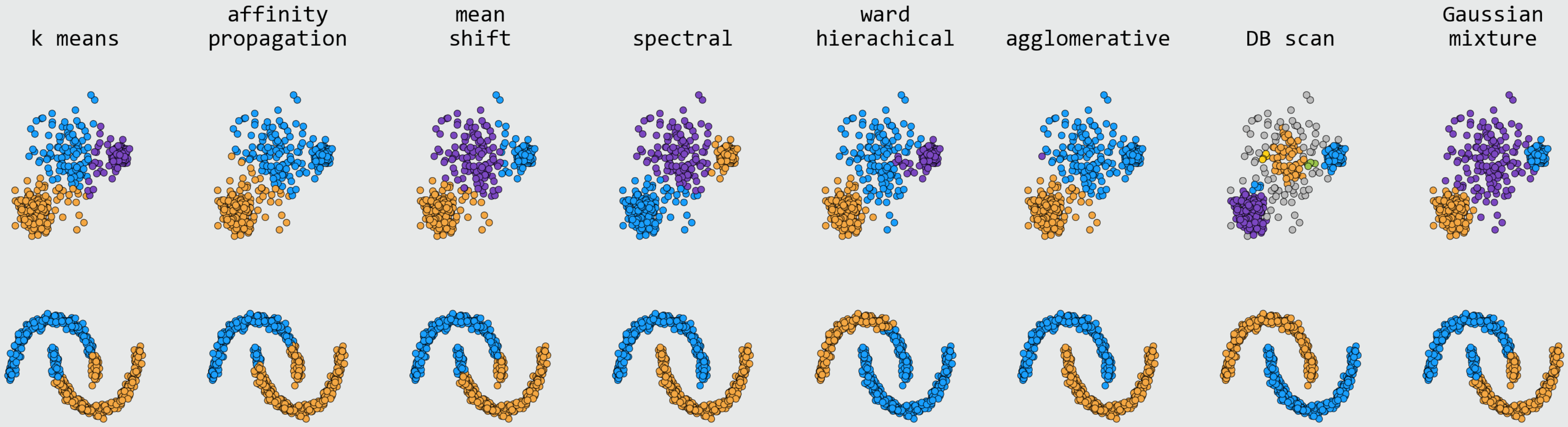
Cluster analysis – example of similarity-based clustering



data points *3 clusters*

Which algorithm is best suited for your dataset?

Cluster analysis – pick up the right one!

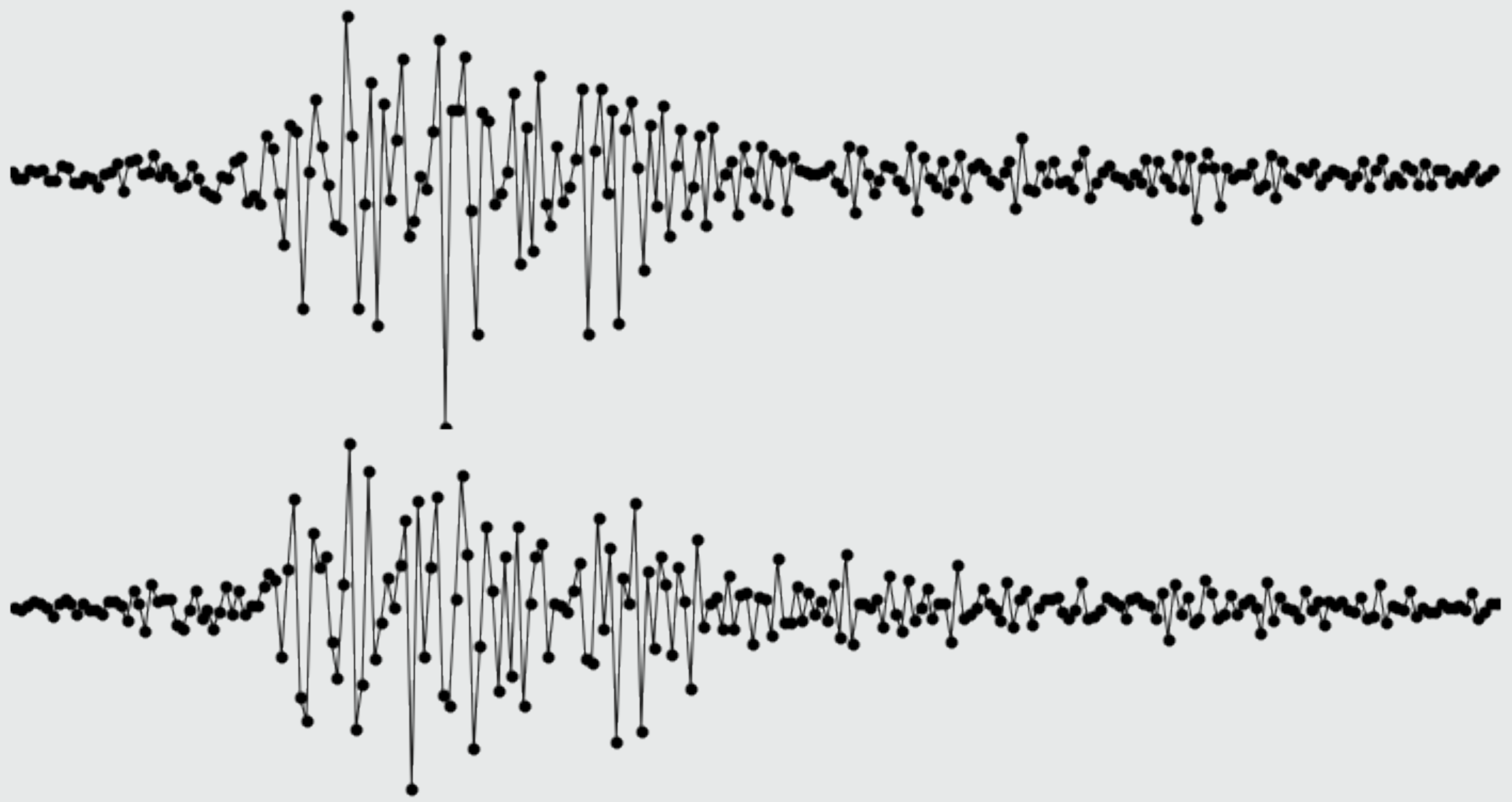


scikit-learn.org

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

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
$x \in \mathbb{R}^N$

→

Features
$x' \in \mathbb{R}^F$

→

Clustering
$c \in \mathbb{N}^C$

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***?

Feature extraction techniques

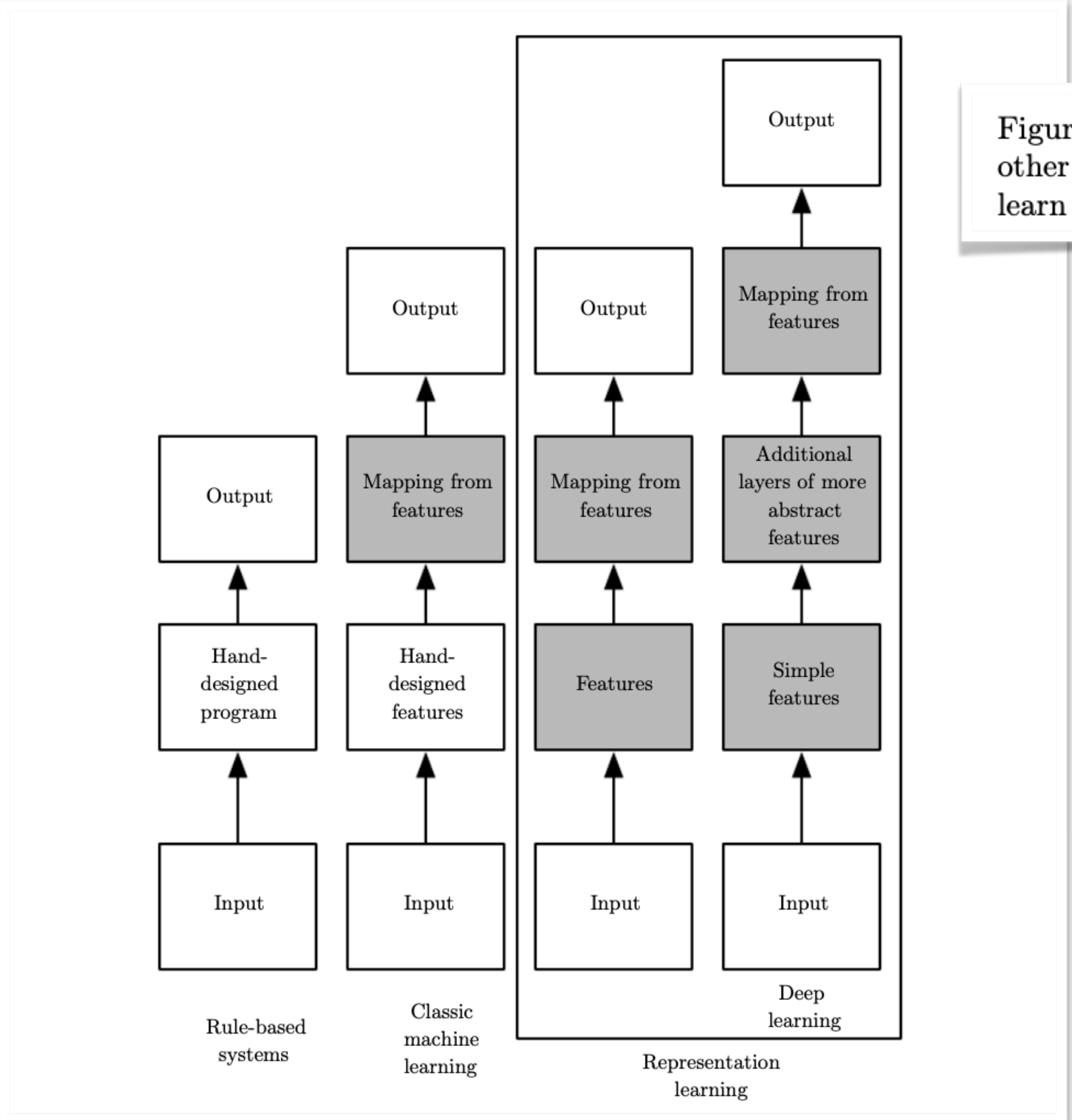


Figure 1.5: Flowcharts showing how the different parts of an AI system relate to each other within different AI disciplines. Shaded boxes indicate components that are able to learn from data.

Goodfellow et al. (2016)

We can learn the features from the data according to a given task (representation learning)

Neural networks

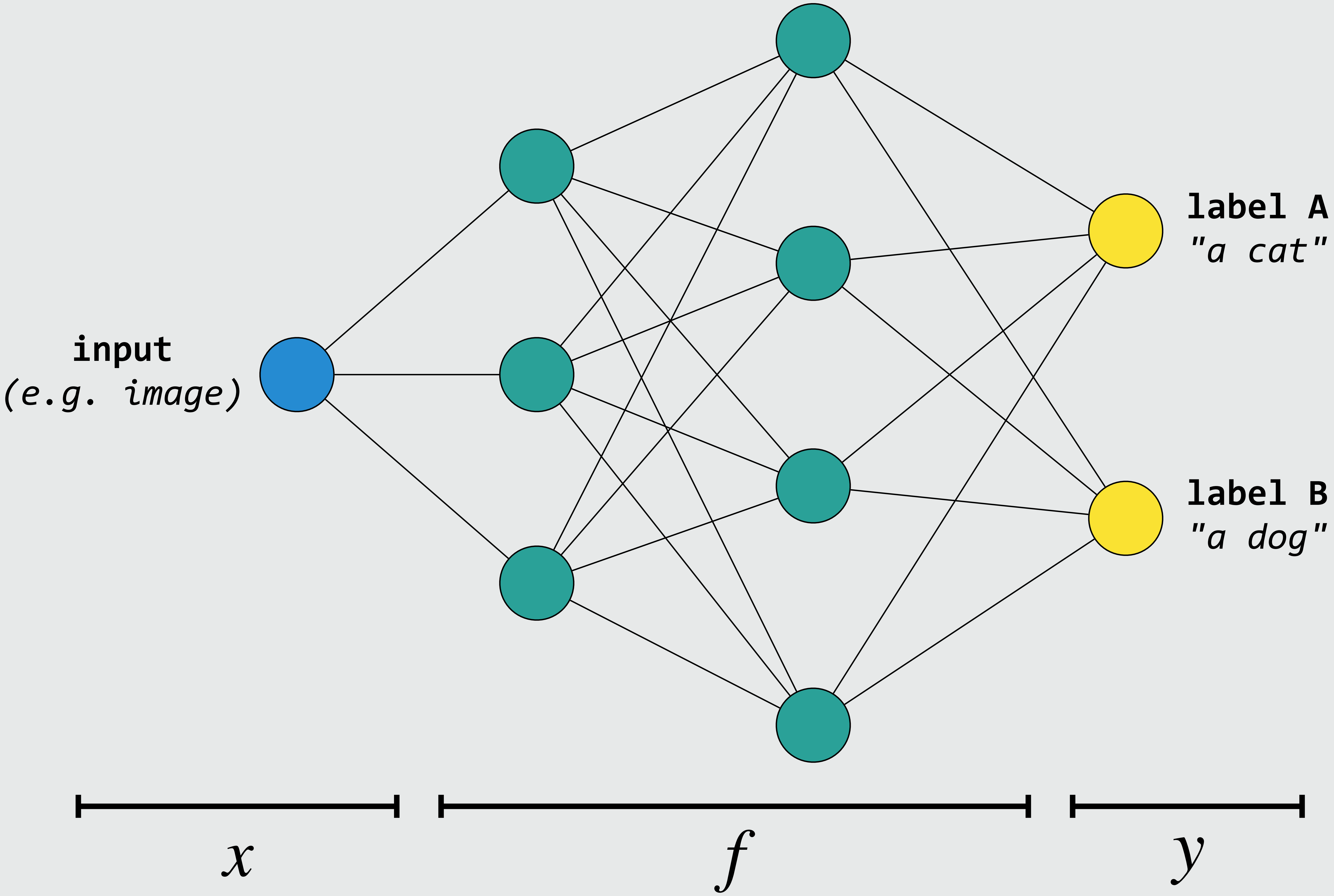
"a cat"

$$f(x) = y$$

Neural networks

$$f(x) = y$$

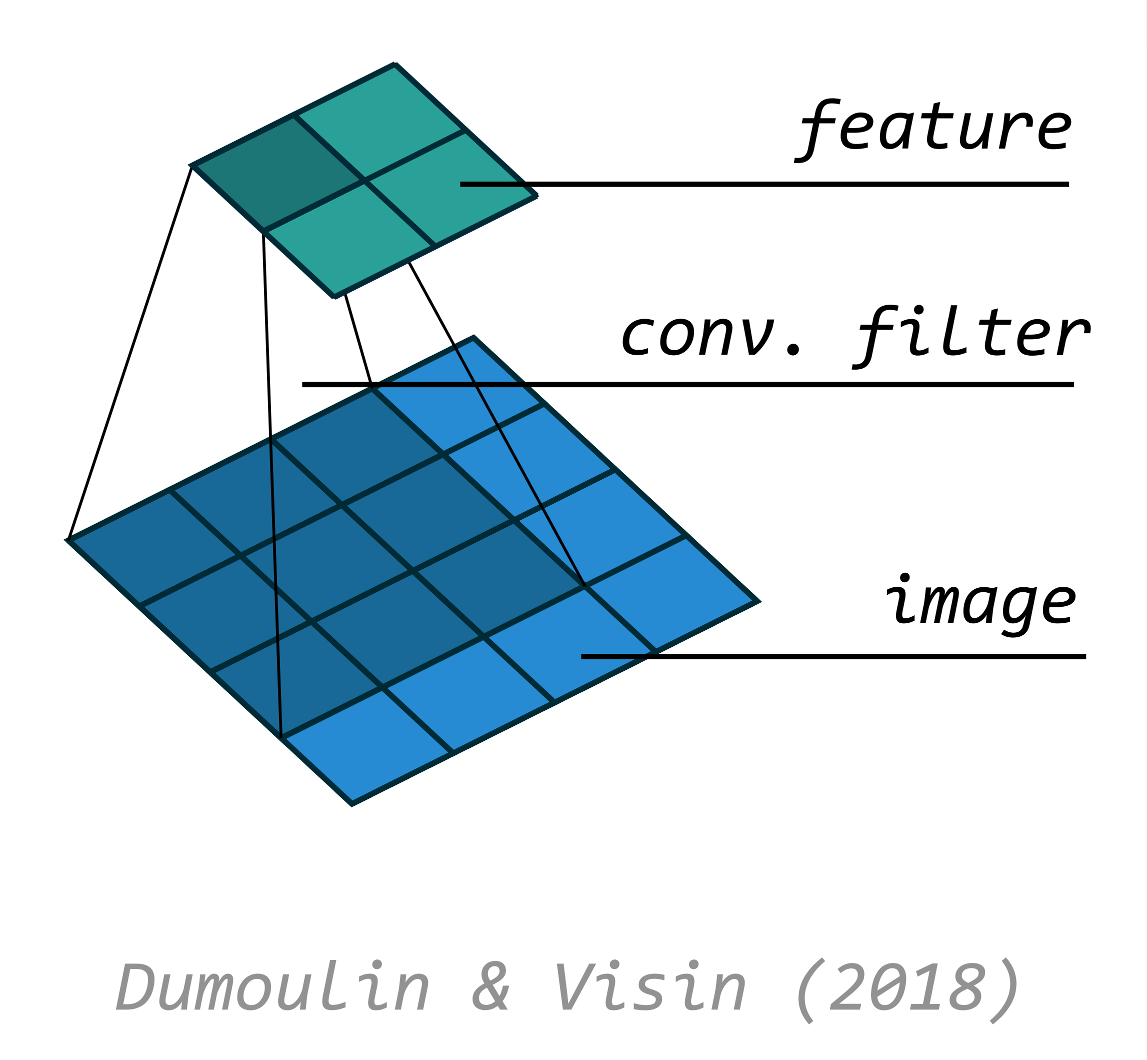
"a cat"



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

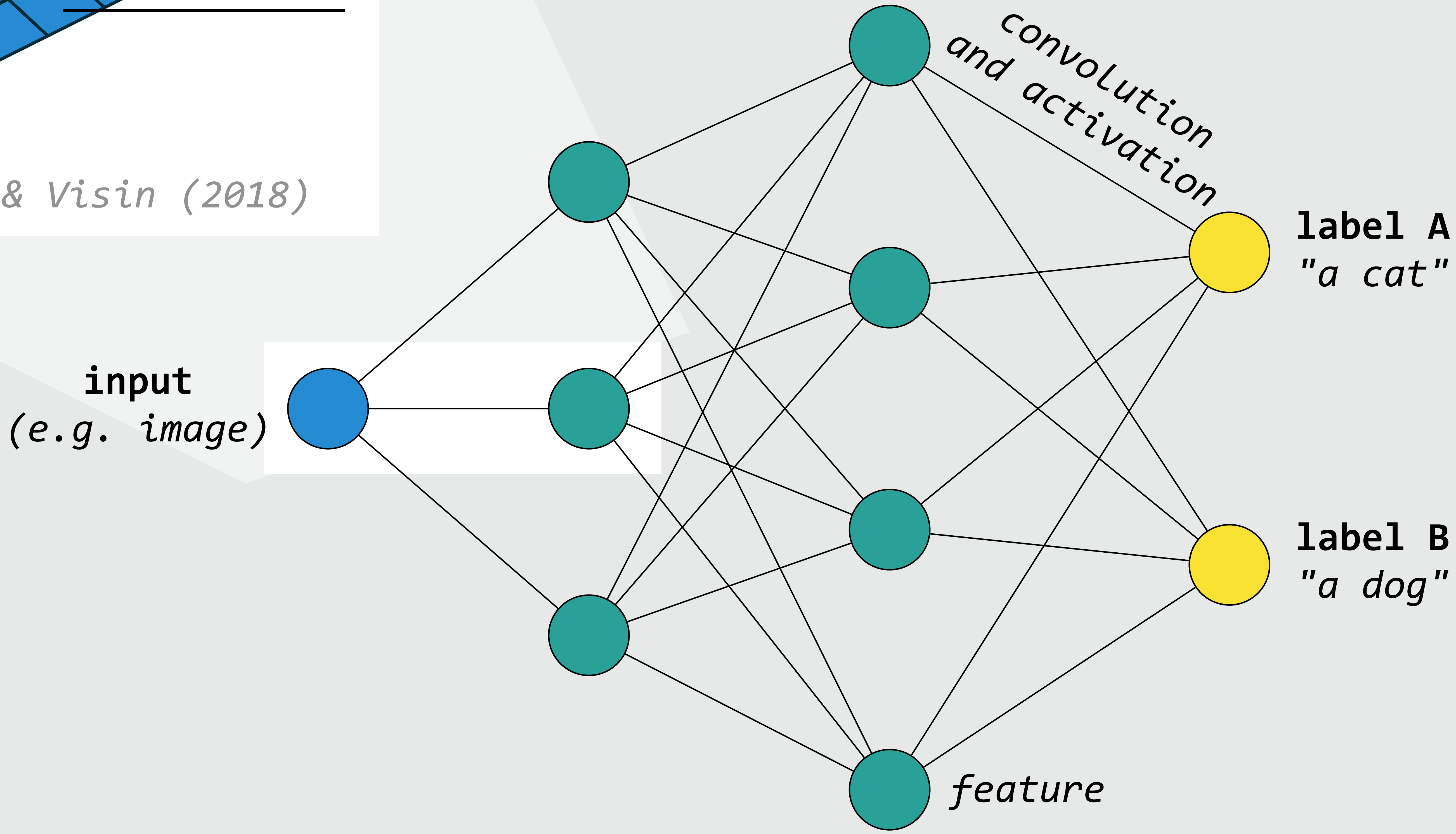
Neural networks

Learn to recognize patterns



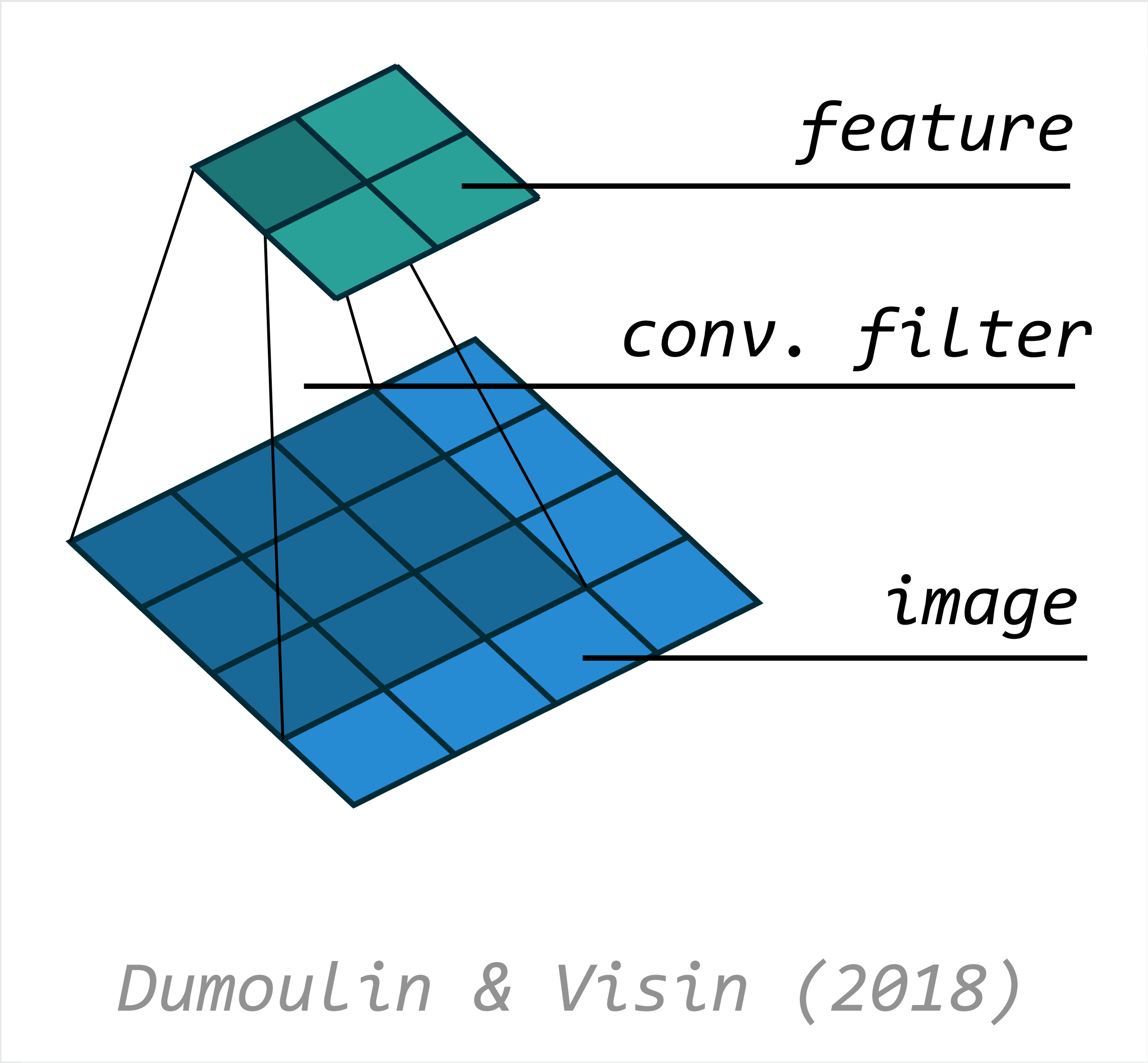
"a cat"

$$f(x) = y$$



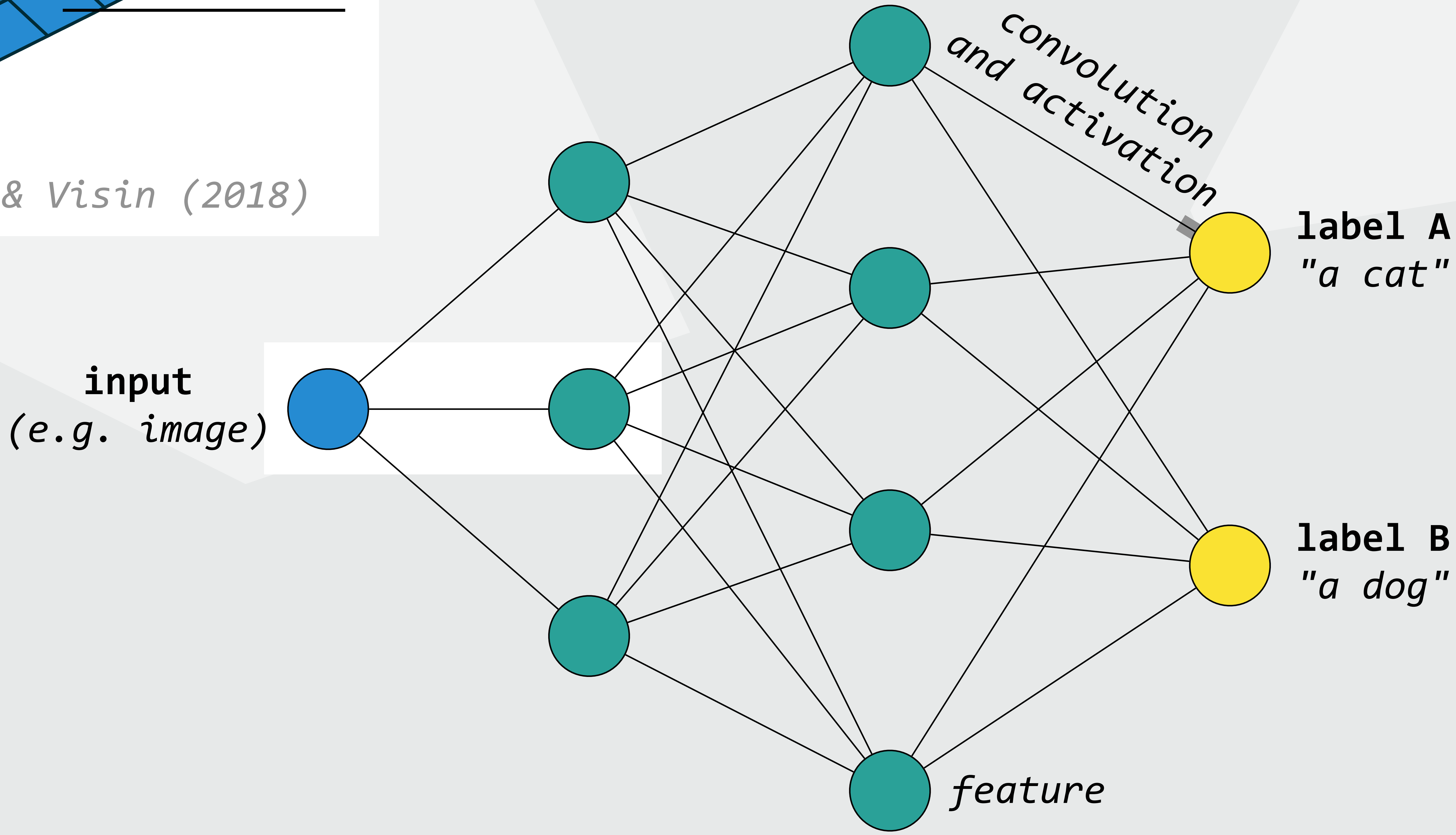
Neural networks

Learn to recognize patterns

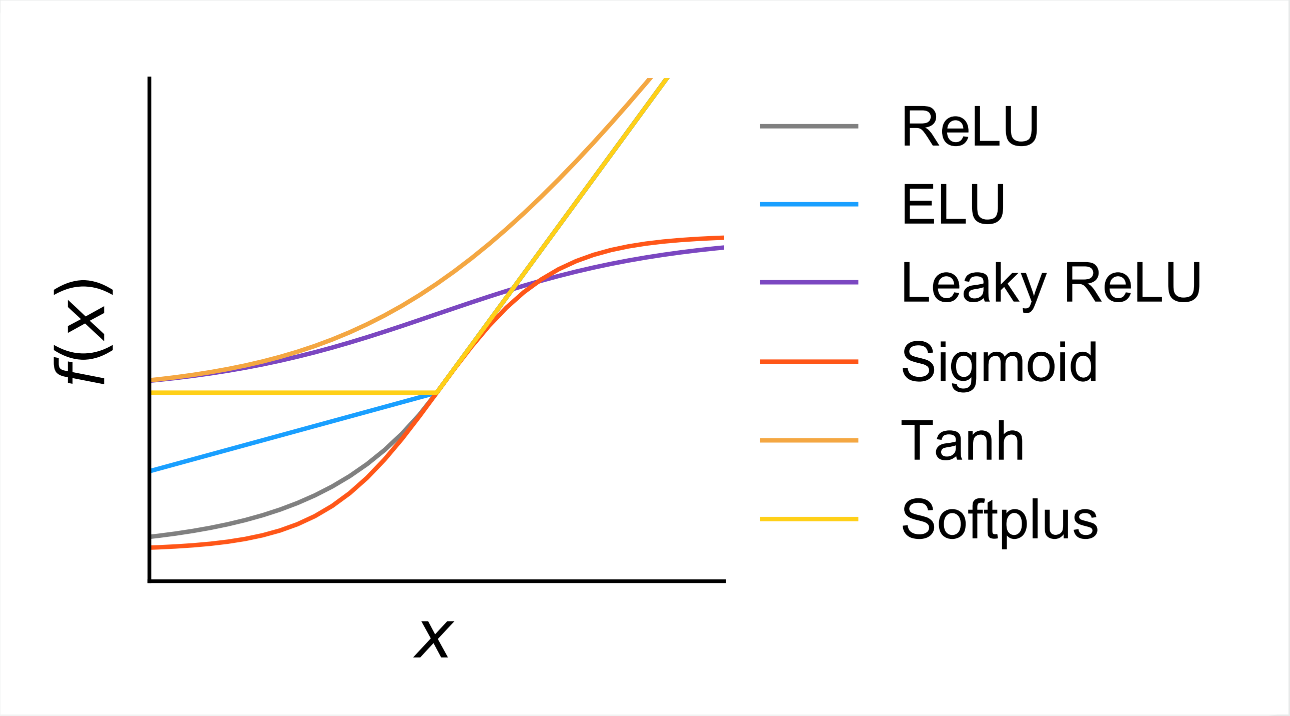


"a cat"

$$f(x) = y$$

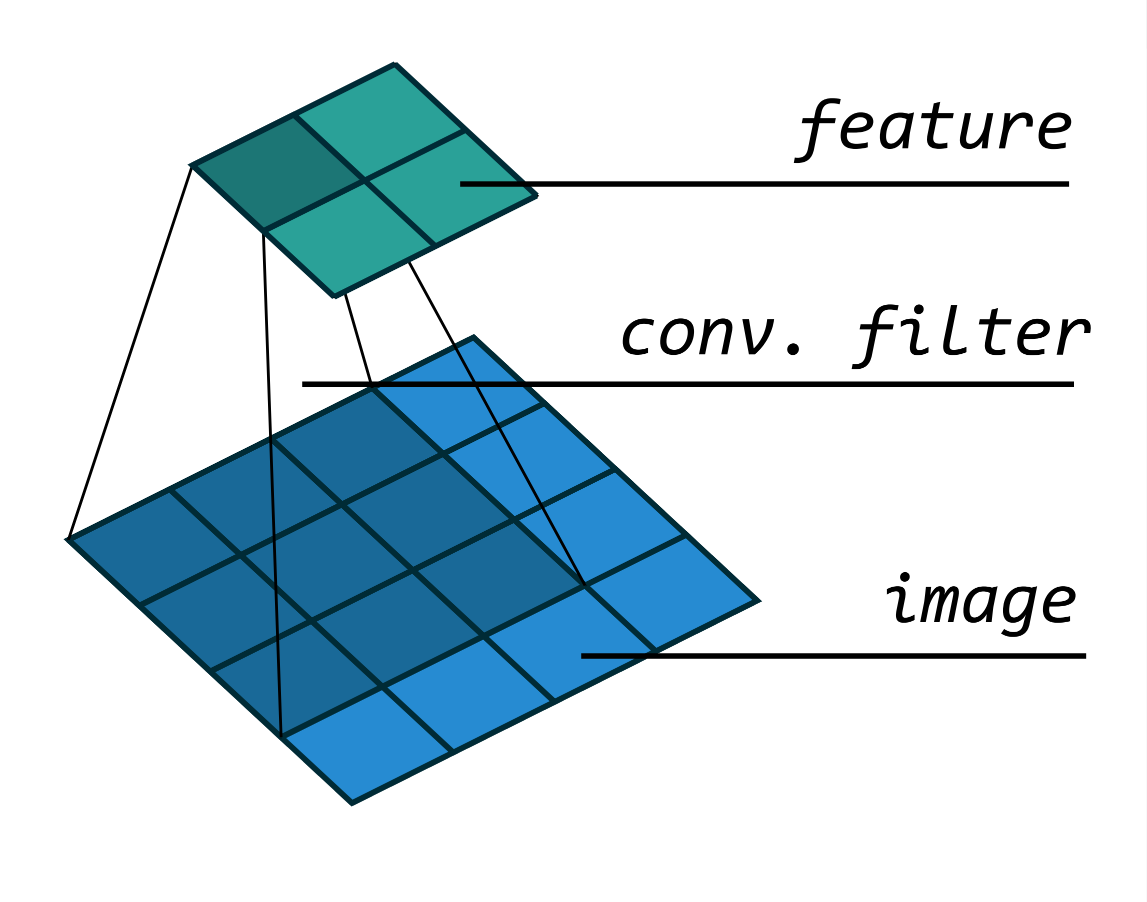


Non linear activation function



Neural networks

Learn to recognize patterns

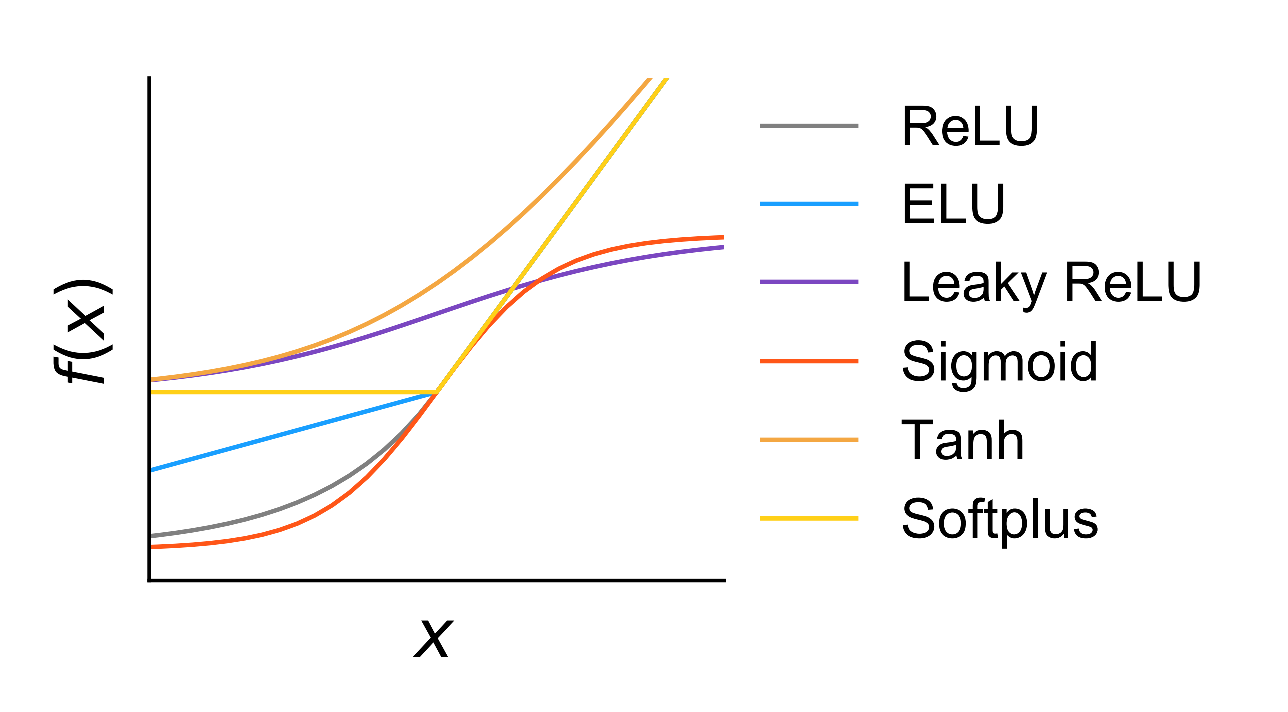


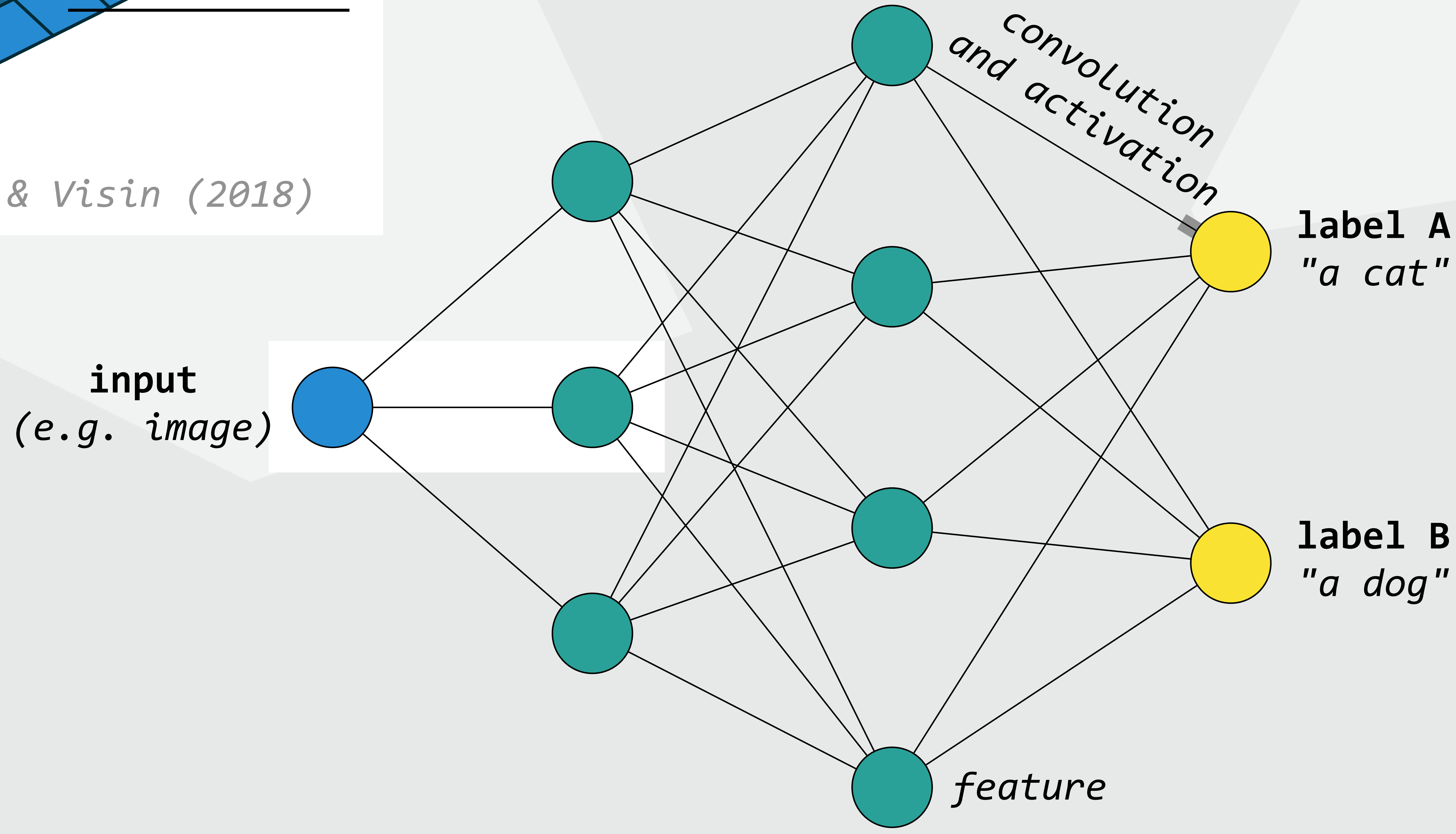
Dumoulin & Visin (2018)

$$f(x) = y$$

"a cat"

Non linear activation function

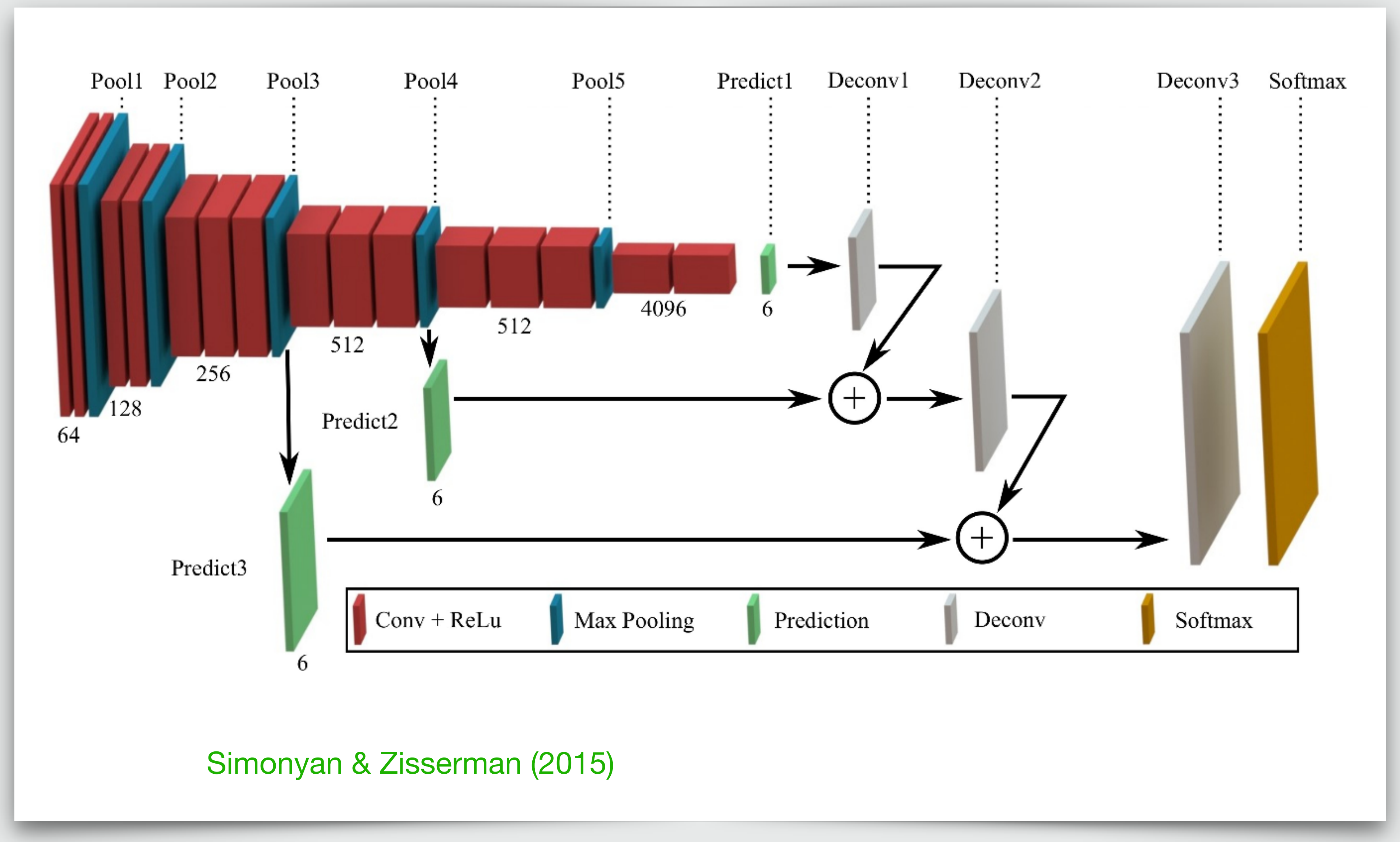




How neural networks should be **designed**?
How many layers?
What kind of filters?
Which activations?

Example of deep convolutional network for image classification

Deep convolutional VGG16

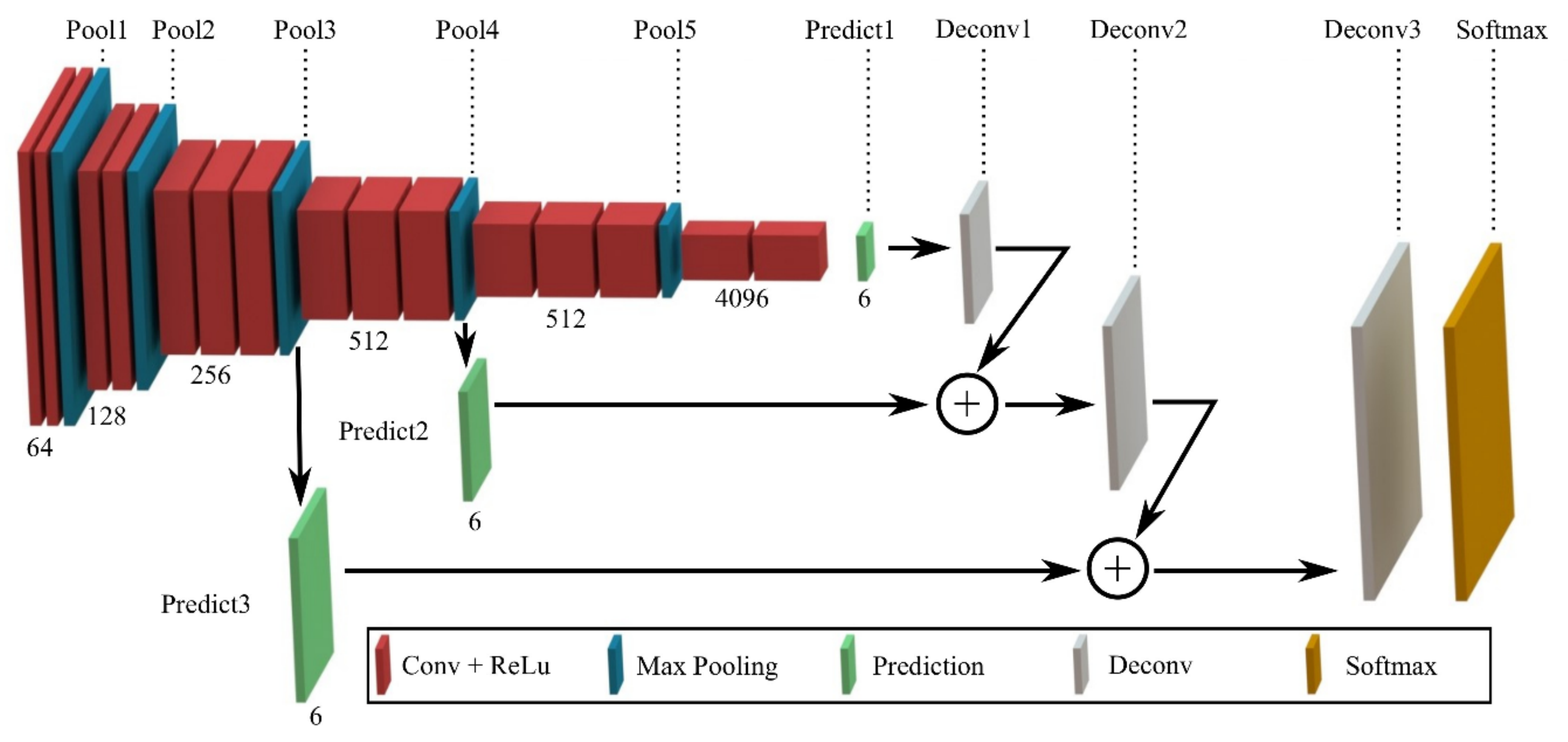


From details to abstraction

Architectures obtained empirically
with insights from signal processing
and inspired from nature

Example of deep convolutional network for image classification

Deep convolutional VGG16



Simonyan & Zisserman (2015)

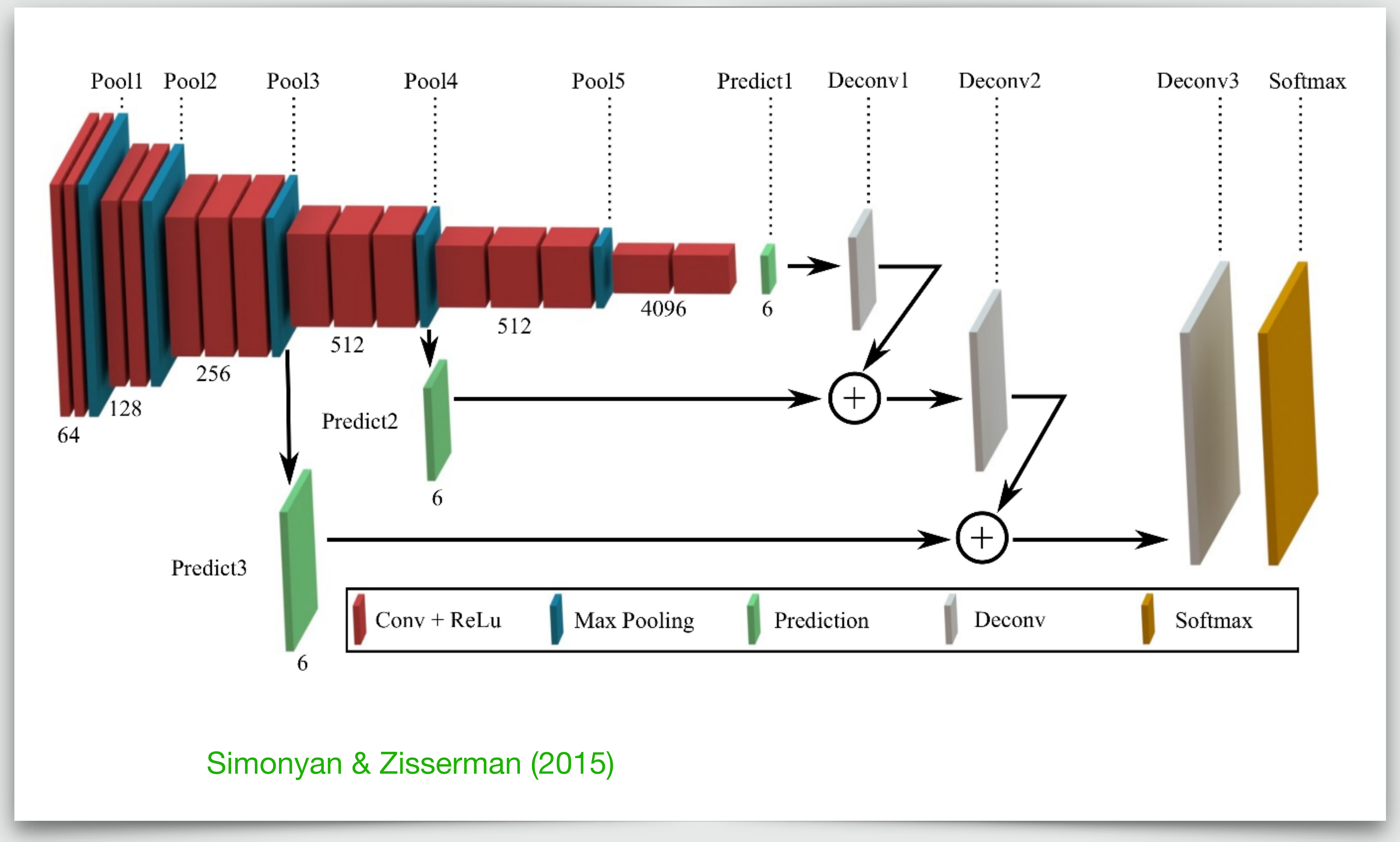
From details to abstraction

Architectures obtained empirically with insights from signal processing and inspired from nature

Filters typically learned with VGG16 (first layer)

Example of deep convolutional network for image classification

Deep convolutional VGG16



From details to abstraction

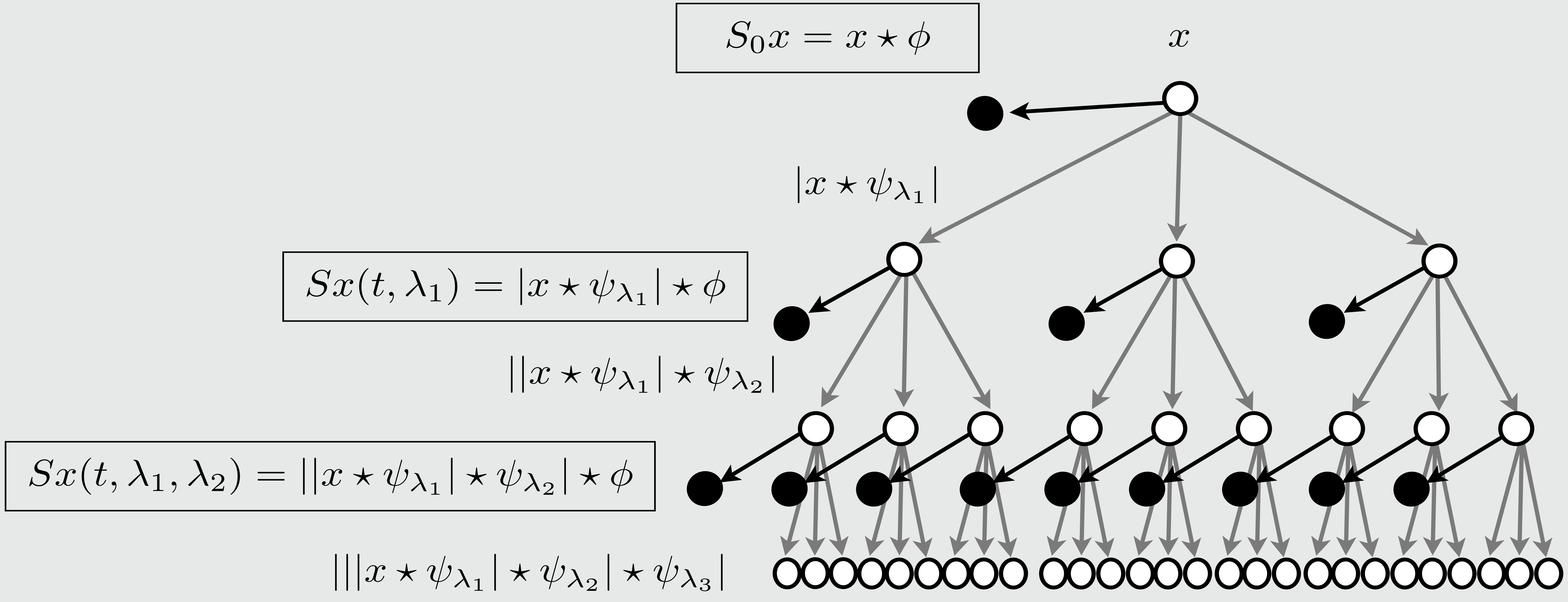
Architectures obtained empirically
with insights from signal processing
and inspired from nature

Filters typically learned with VGG16 (first layer)

Sample two-dimensional Gabor wavelet filters

Convolutional filters are sensitive to geometries.
They could be replaced with wavelets filters.

The scattering network — a convolutional network with wavelet filters



Andén & Mallat (2014)

- Analytical wavelet filters
- No learning involved
- Explicit properties
- Straightforward design

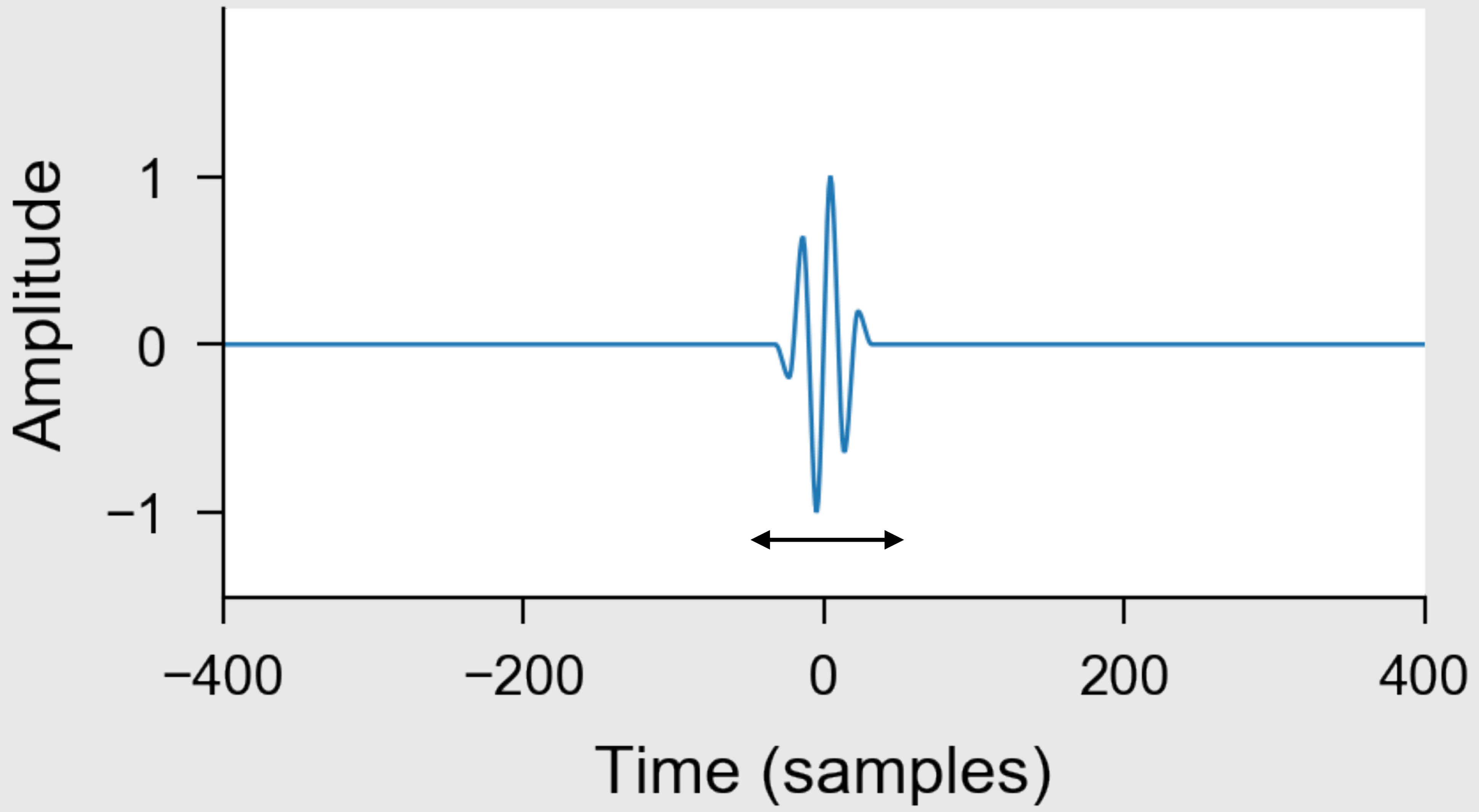
State-of-the-art performances on one dimension signal analysis

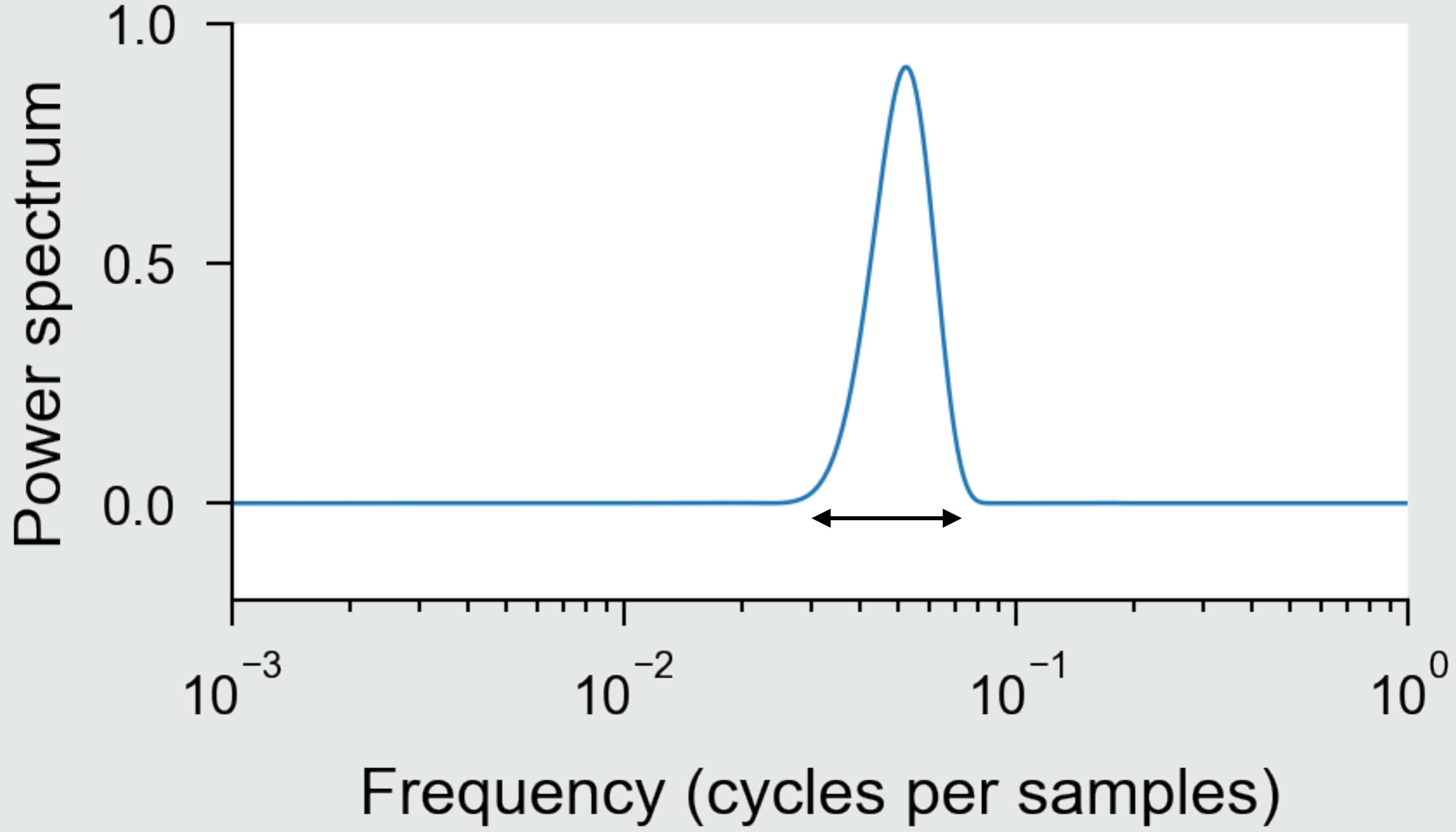
audio classification (Andén 2014), electrocardiograms & birds (Balestriero 2017)

Wavelet transform

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

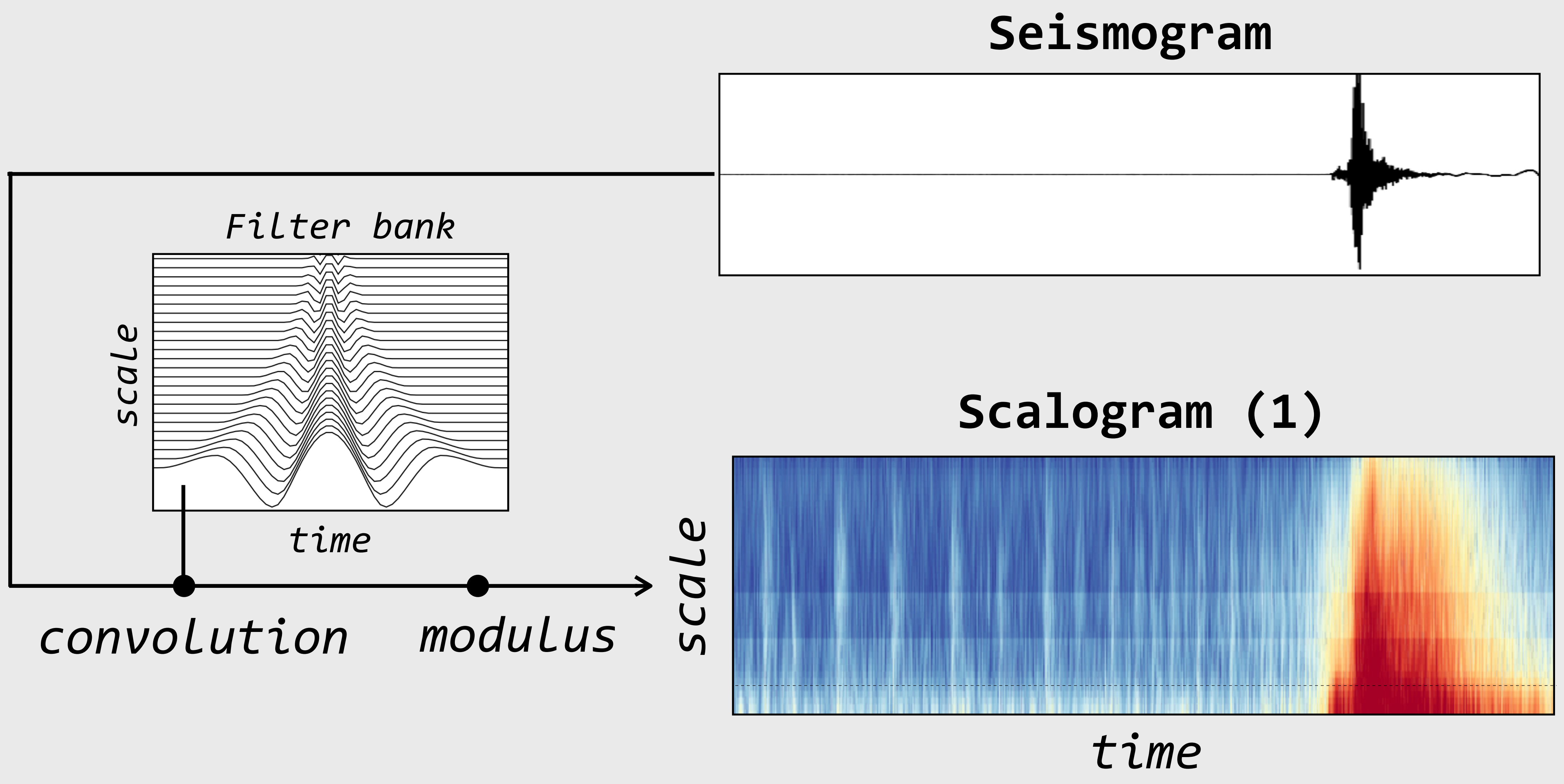
$$Wx(\lambda, t) = (\psi_\lambda \otimes x)(t)$$





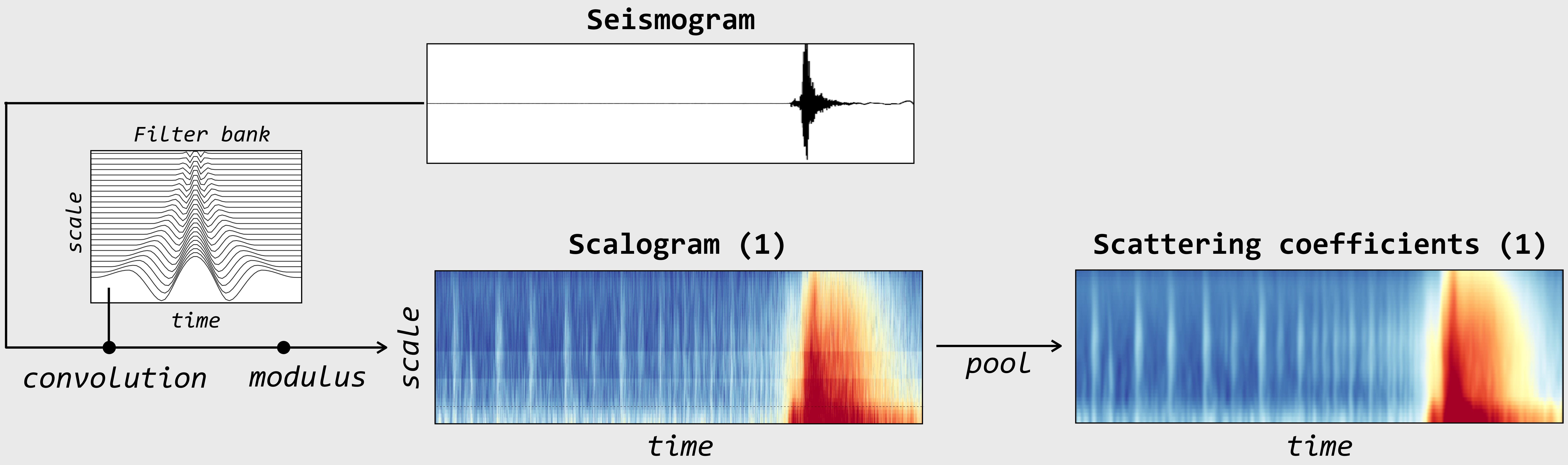
Wavelets are localized in time and frequency

Idea of a scattering network



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

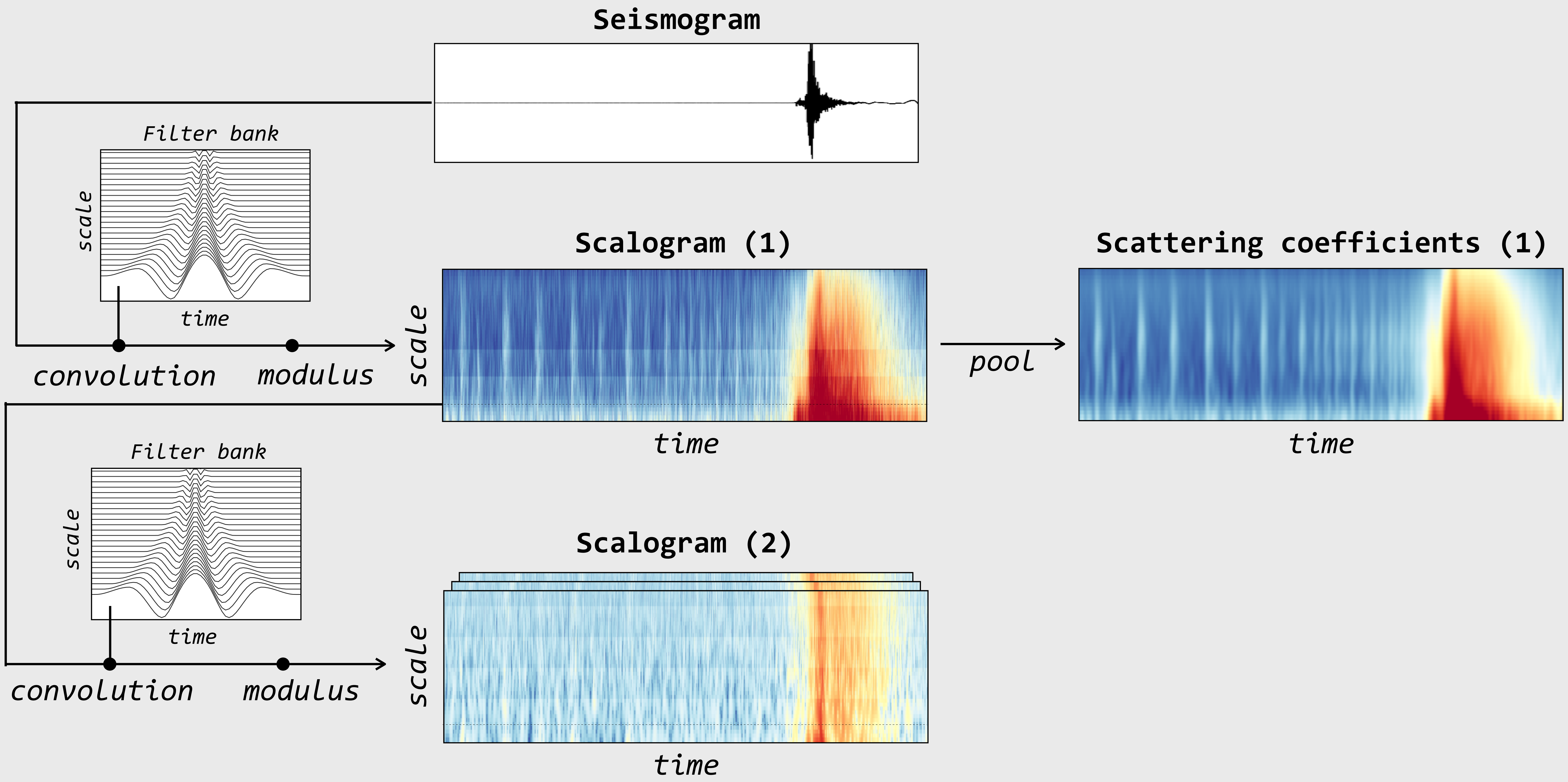
Idea of a scattering network



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

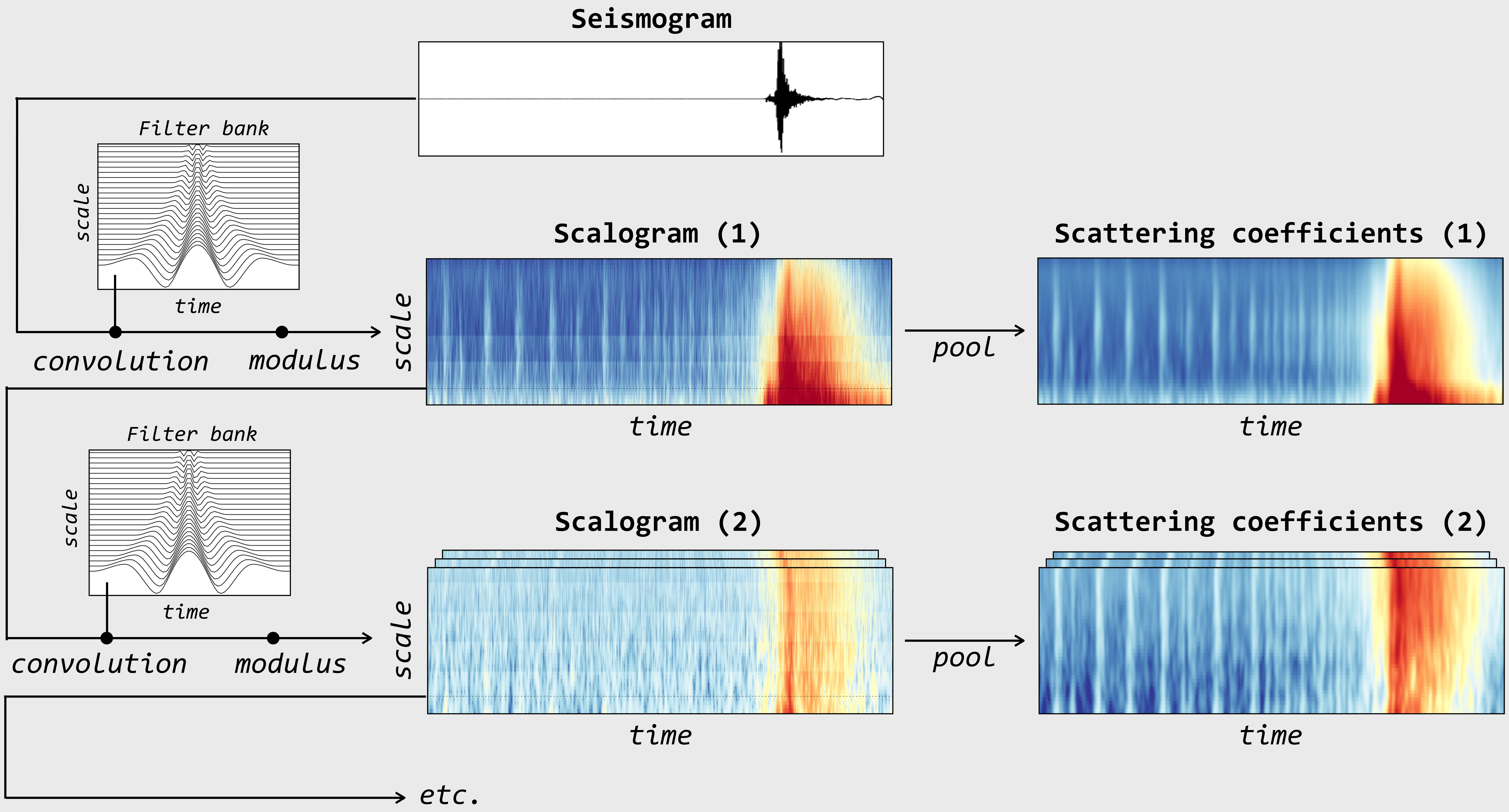
The first-order scattering coefficients provide a **locally stable signal description** at small time scales.

Idea of a scattering network



Larger time scales are analyzed at second order

Idea of a scattering network

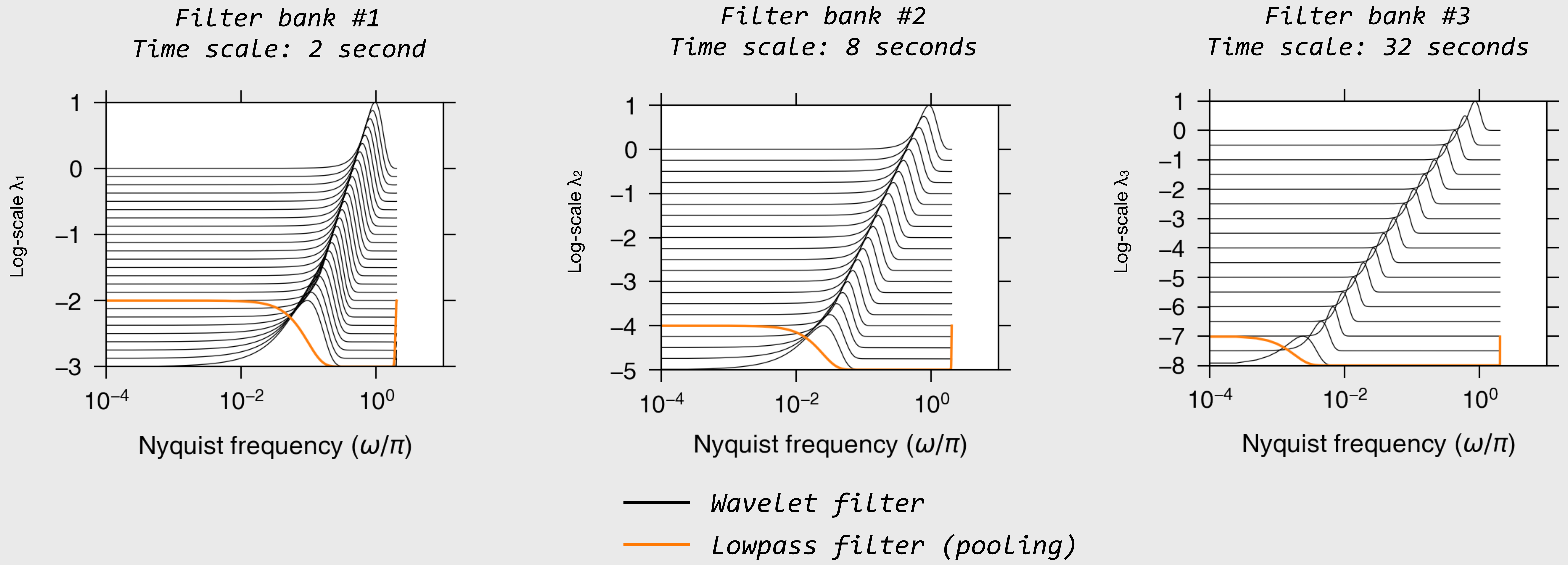


At each order, features are made **locally stable to signal deformations**
The signal structure is scattered across **multiple time scales**

Scattering network design is straightforward

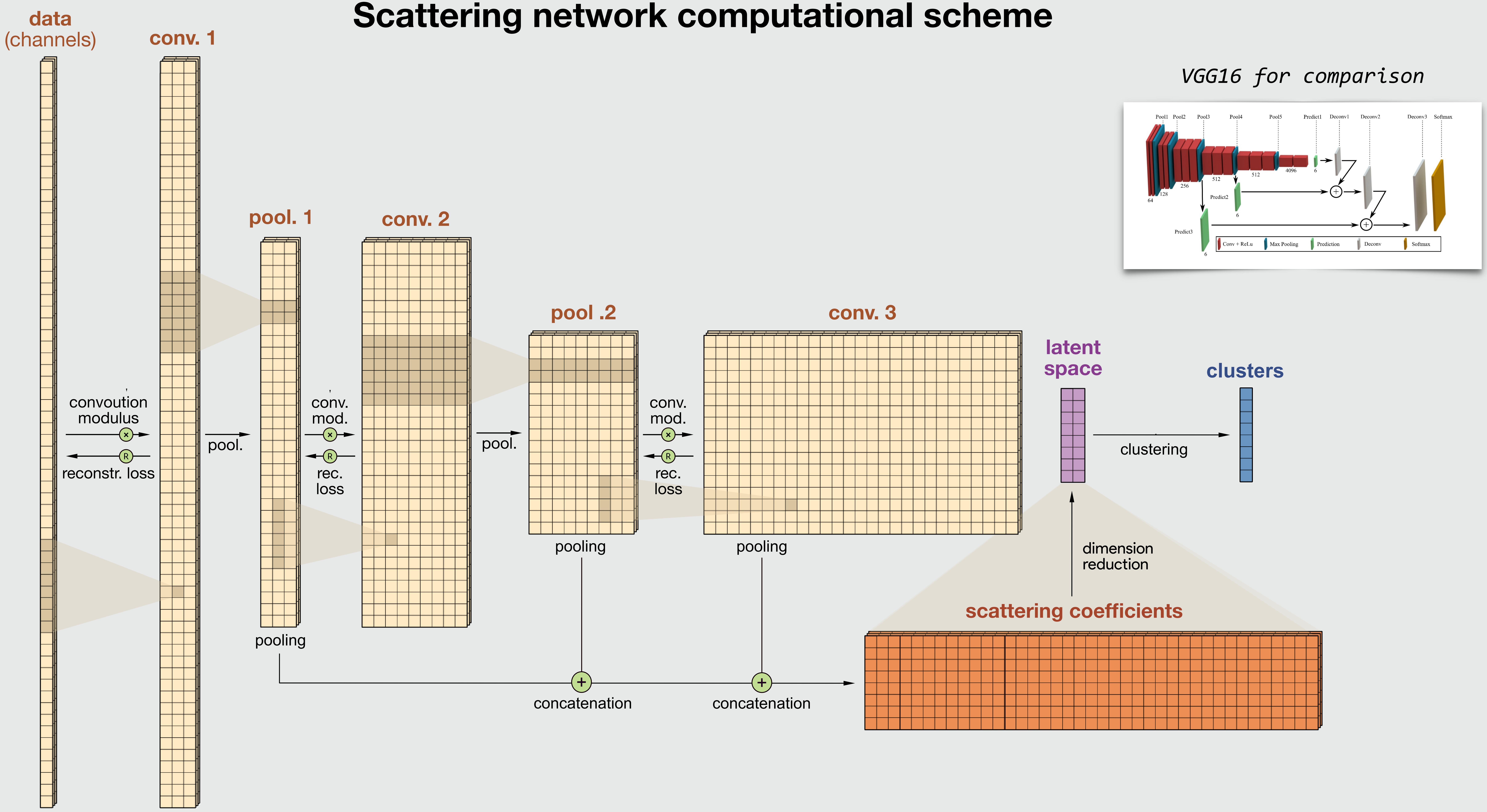
From small to large time scales

From dense to sparse representation



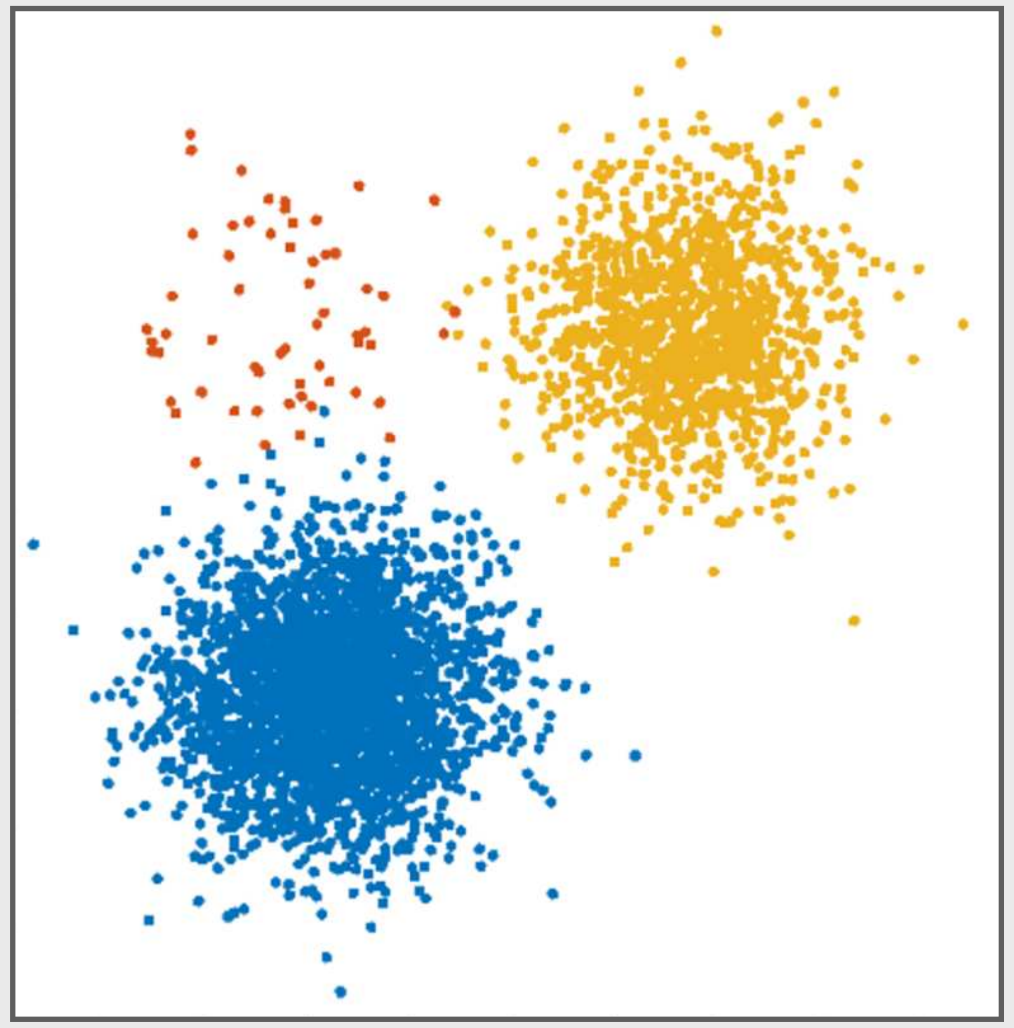
Scattering network is a deep (convolutional) neural network with straightforward architecture and insights from the underlying physics

Scattering network computational scheme



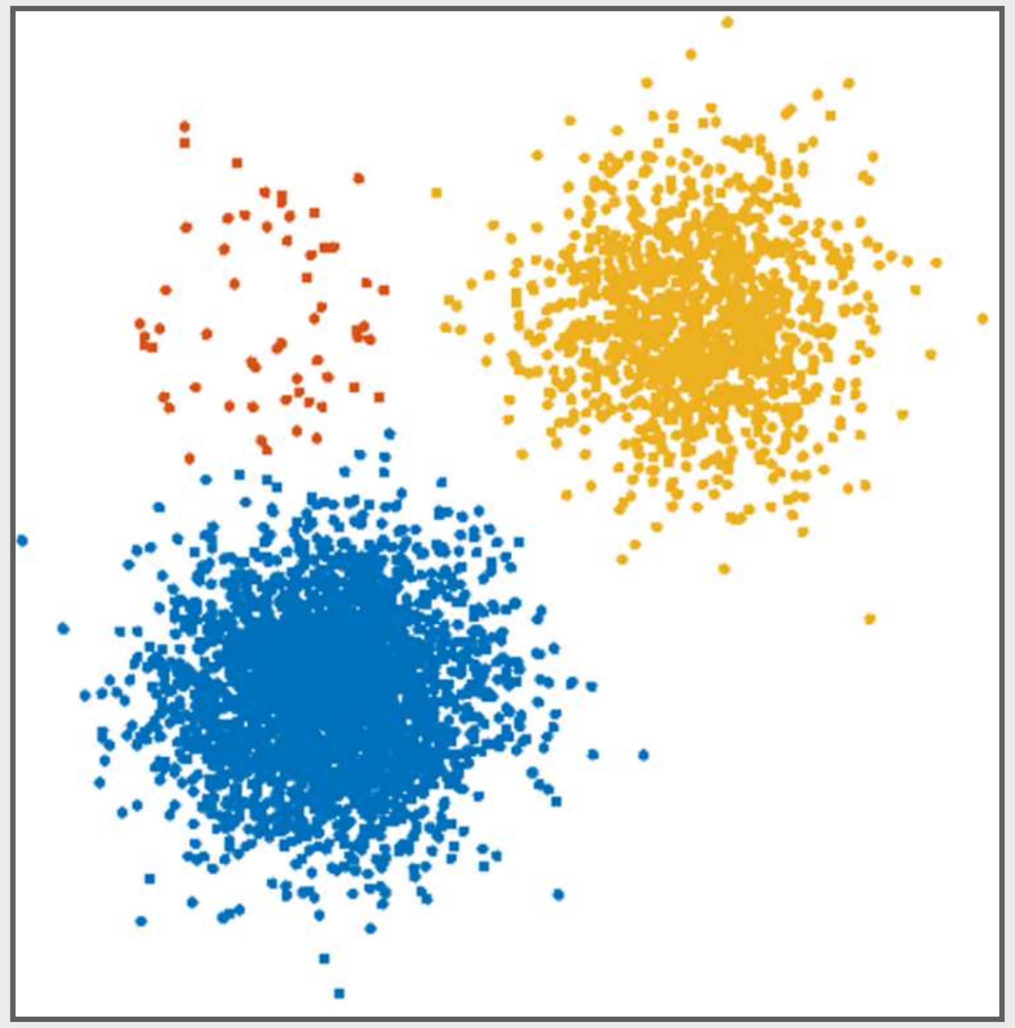
Gaussian mixture model clustering

find a mixture of K gaussian distributions that fit the data



(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*

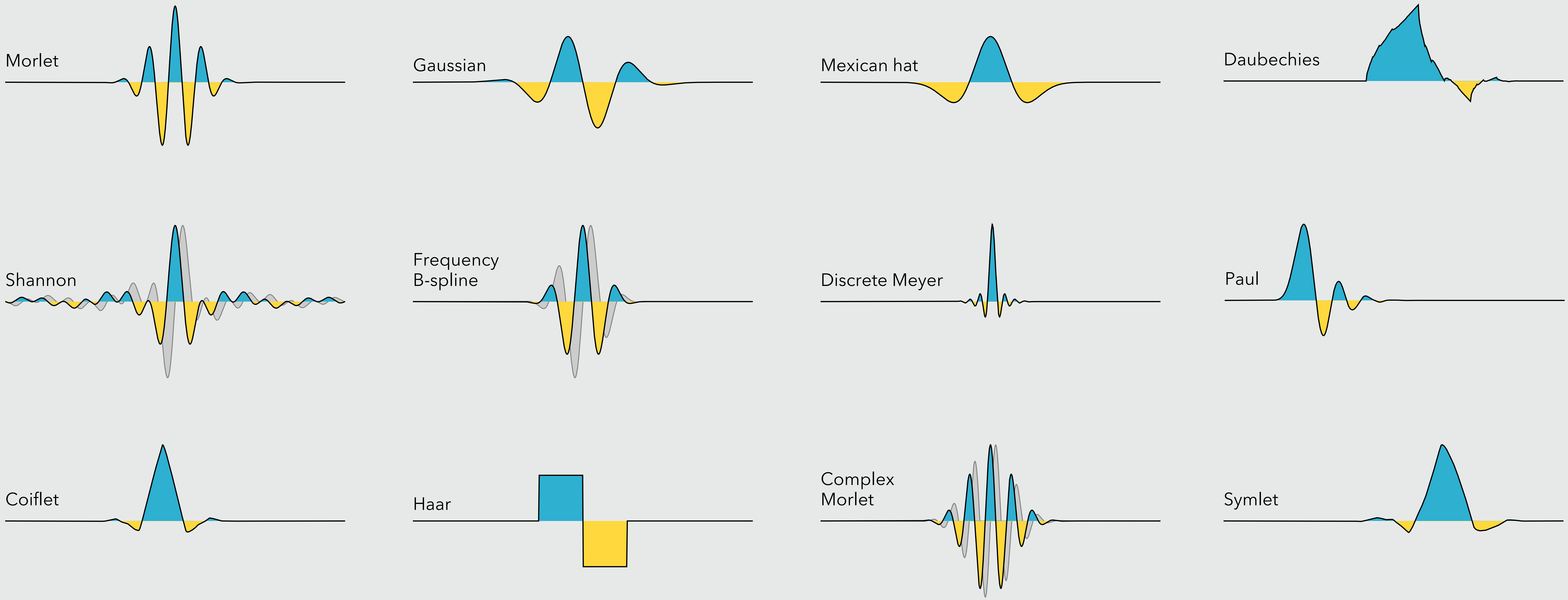
modified from Raykov et al. PONE (2016)

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

GMM can resolve clusters with **unbalanced populations and covariances**

Wavelet shape

all are wavelets with different properties

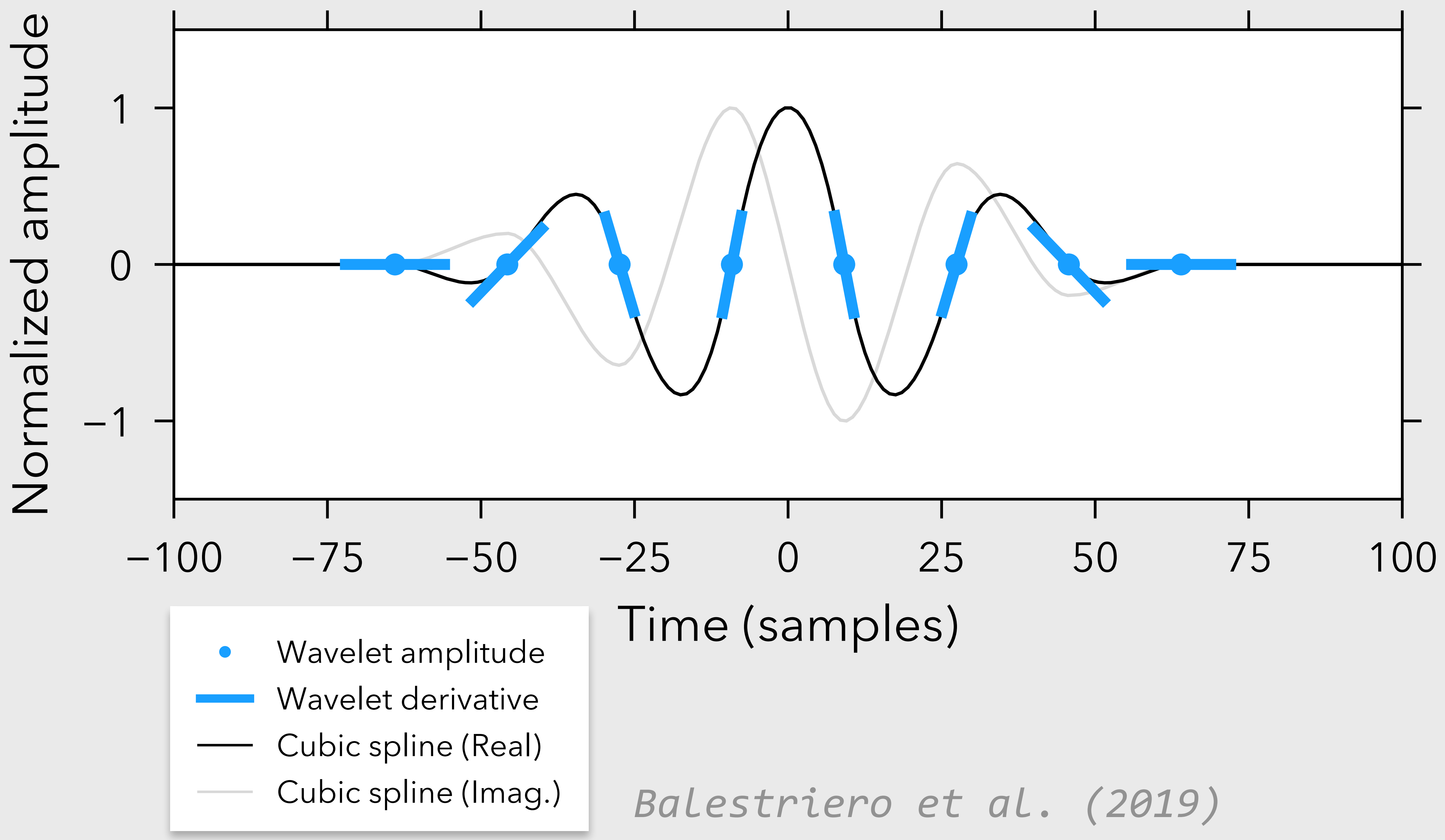


We can test different wavelet shapes and see which one does the best

We can also learn the wavelet according to a given task

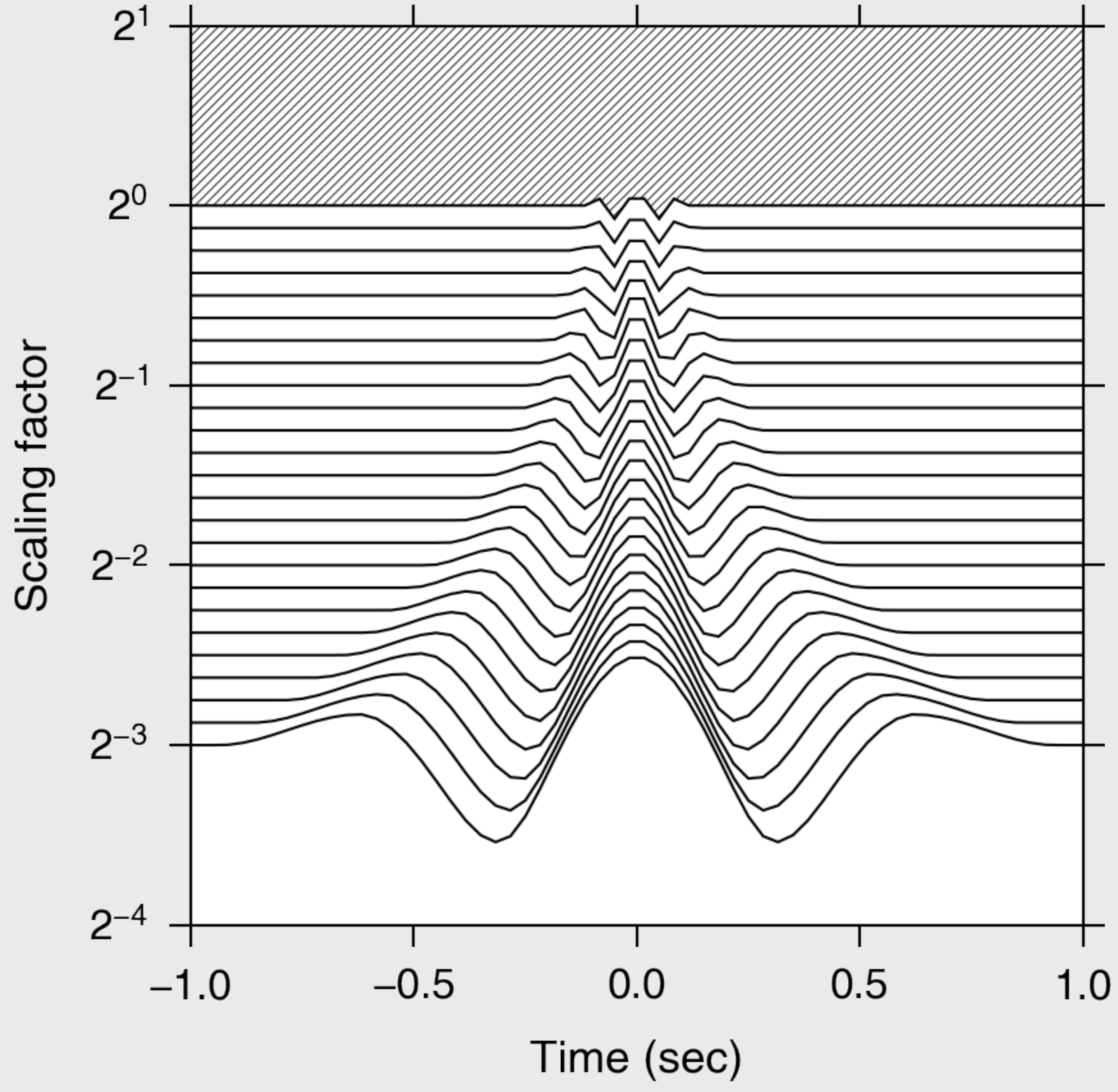
Learnable wavelets from Hermite cubic spline interpolation

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

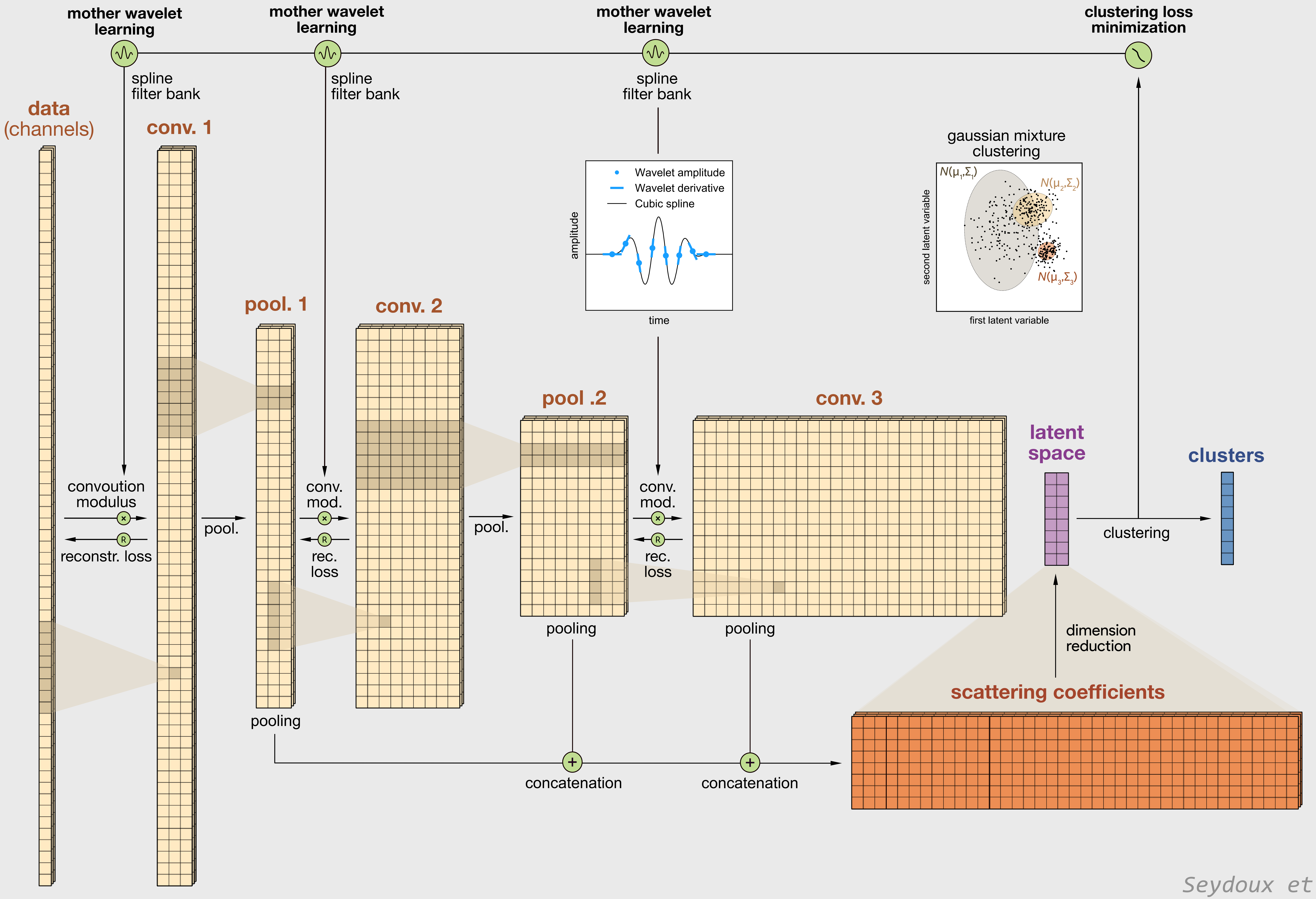


Balestriero et al. (2019)

3. Filter bank obtained from dilation of the mother wavelet

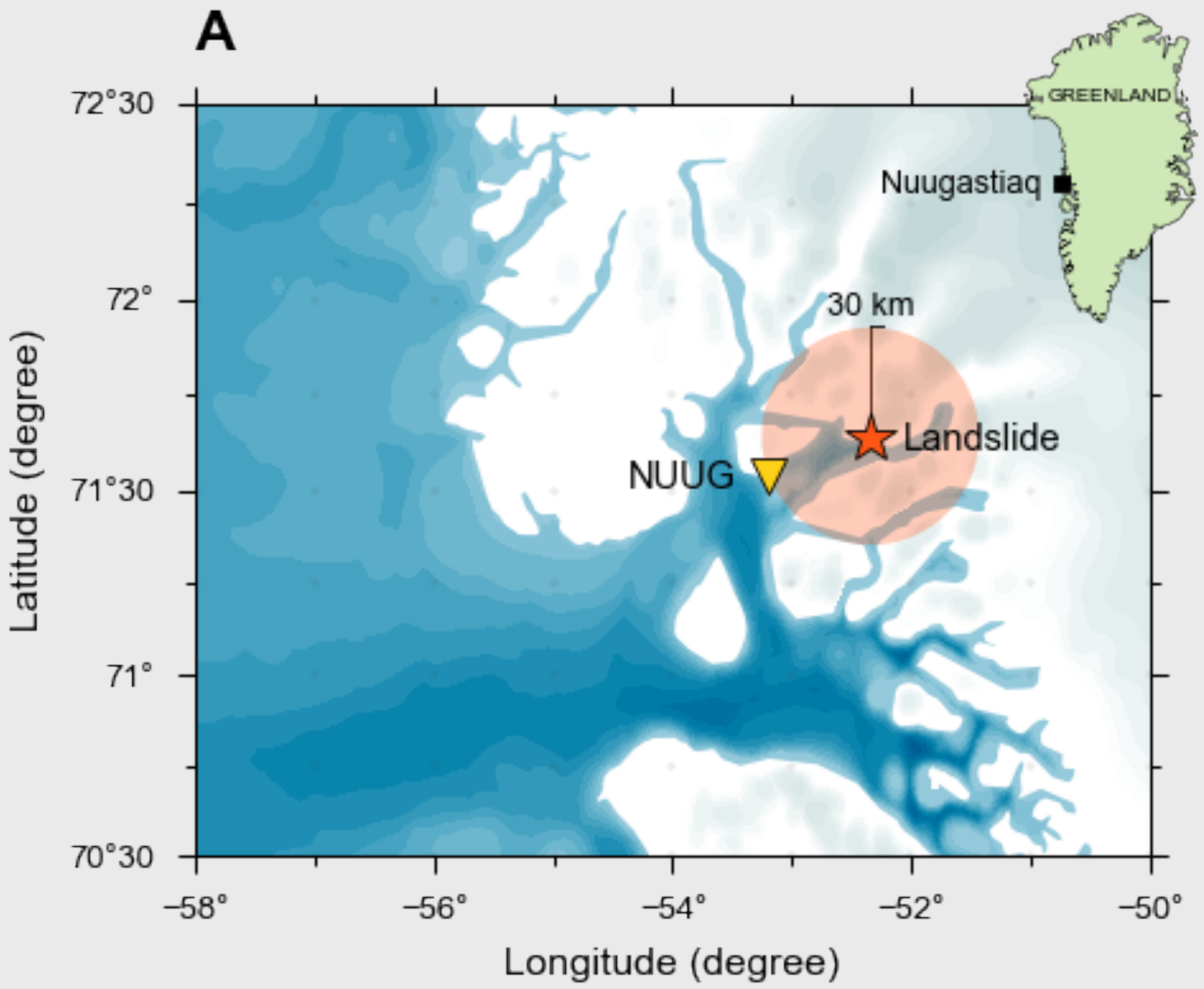


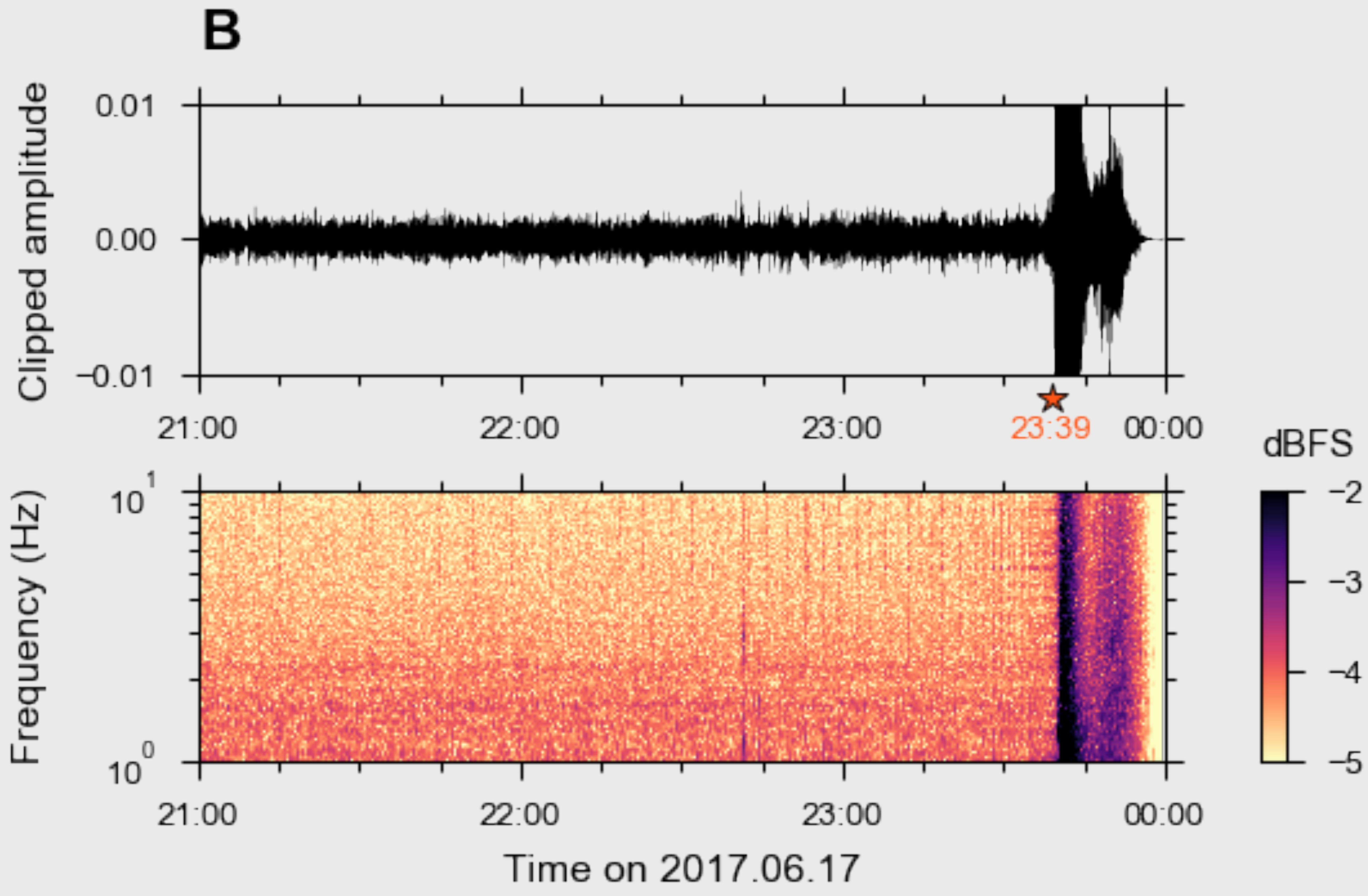
We can learn the wavelets given any task (e.g. clustering, classification, …).
Only a few coefficients are learned compared with classical convolutional nets



Seydoux et al. (rev.)

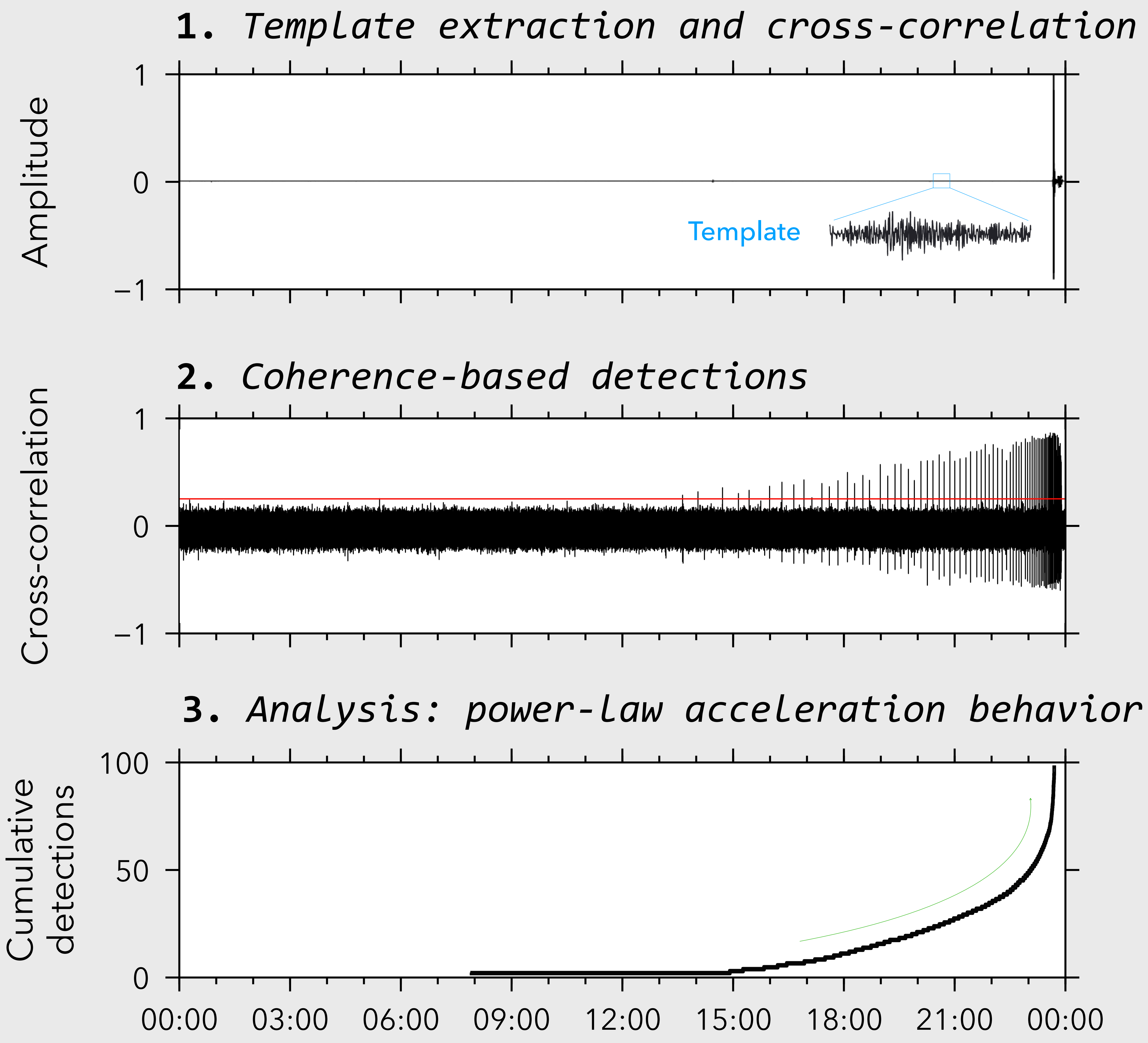
2017 Nuugaastiaq landslide **– weak high-frequency precursors?**





Seydoux et al. (rev.)

2017 Nuugaastiaq landslide – weak high-frequency precursors **revealed by template matching**



modified from Poli (2017)

Advantages of template matching

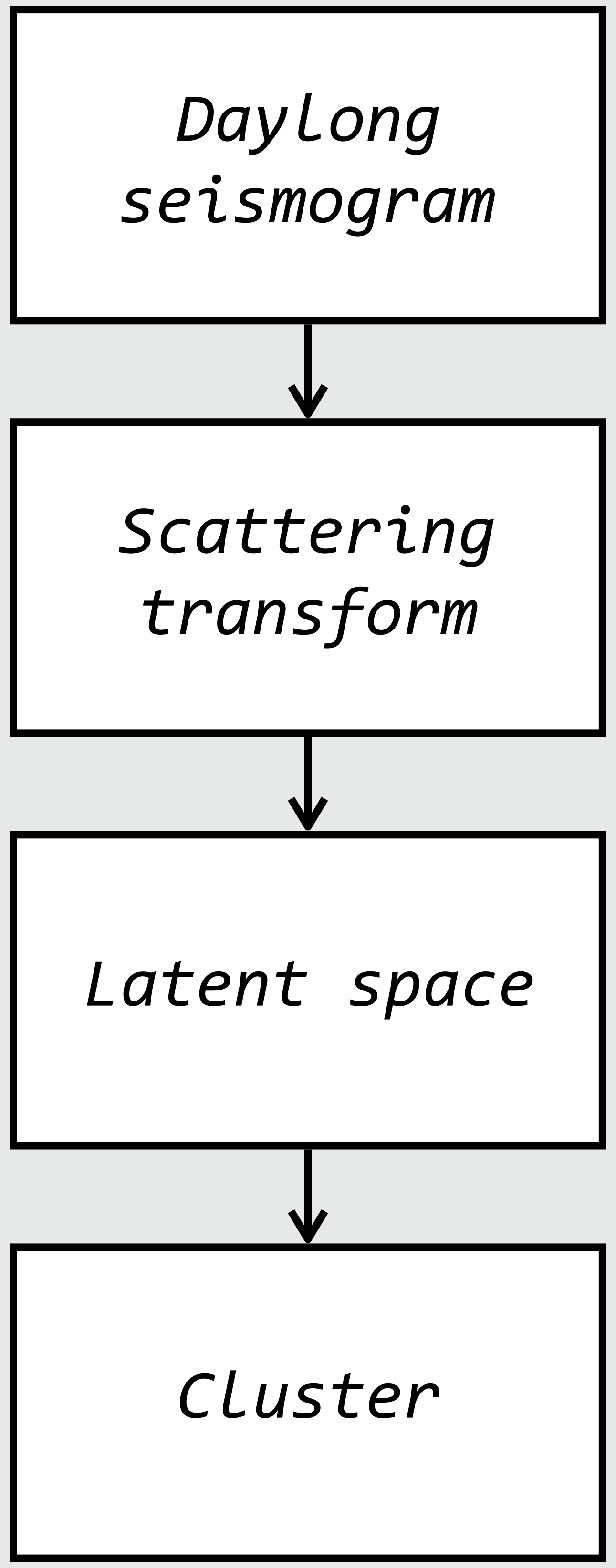
- Good robustness to noise
- Computationally fast

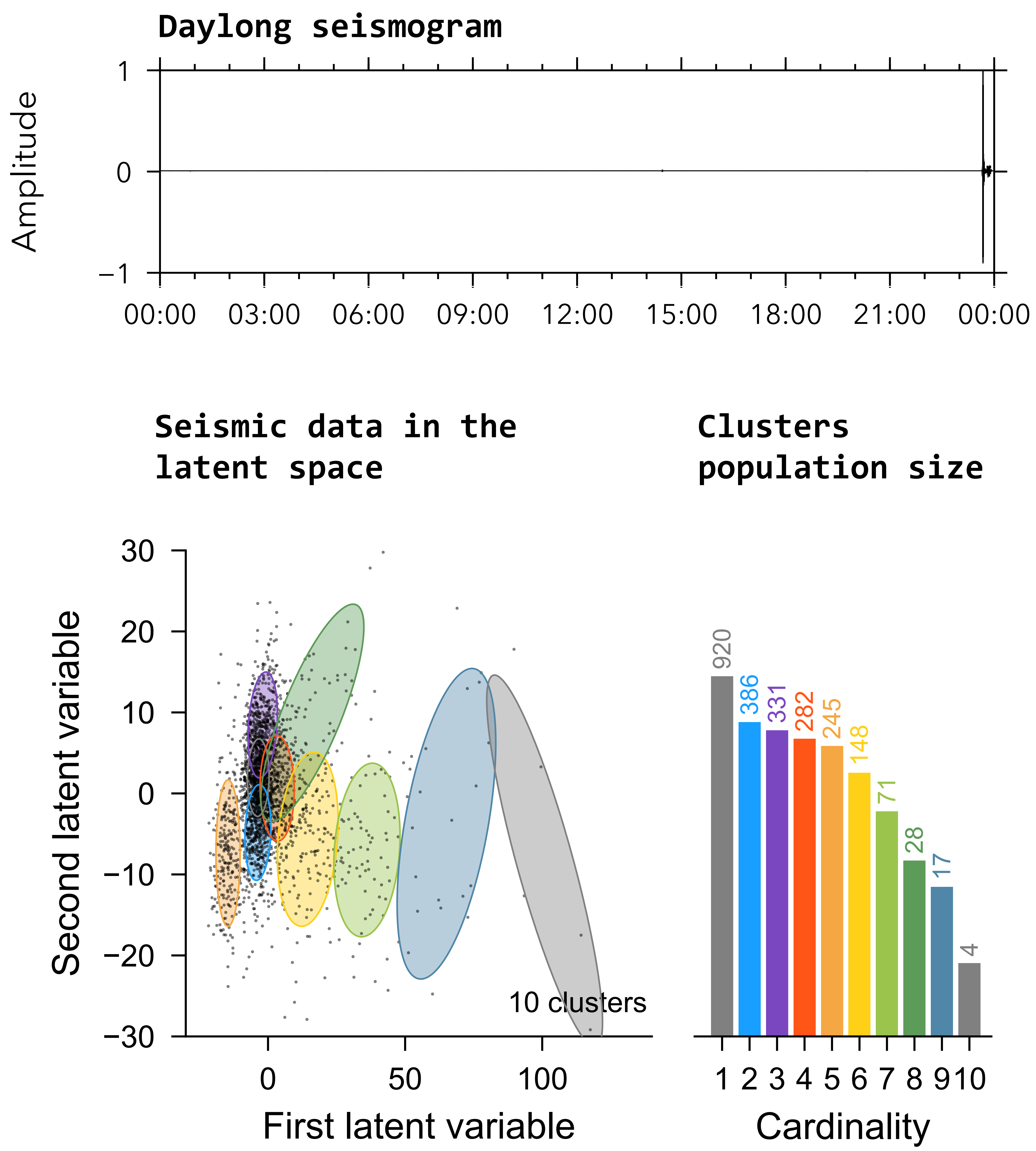
But, …

- It is sensitive to the selected template
- It is sensitive to several parameters (duration, frequency)
- It is limited to known signal (classification with two-classes)

Could this result be retrieved with an unsupervised strategy?

Unsupervised clustering of a daylong seismic record

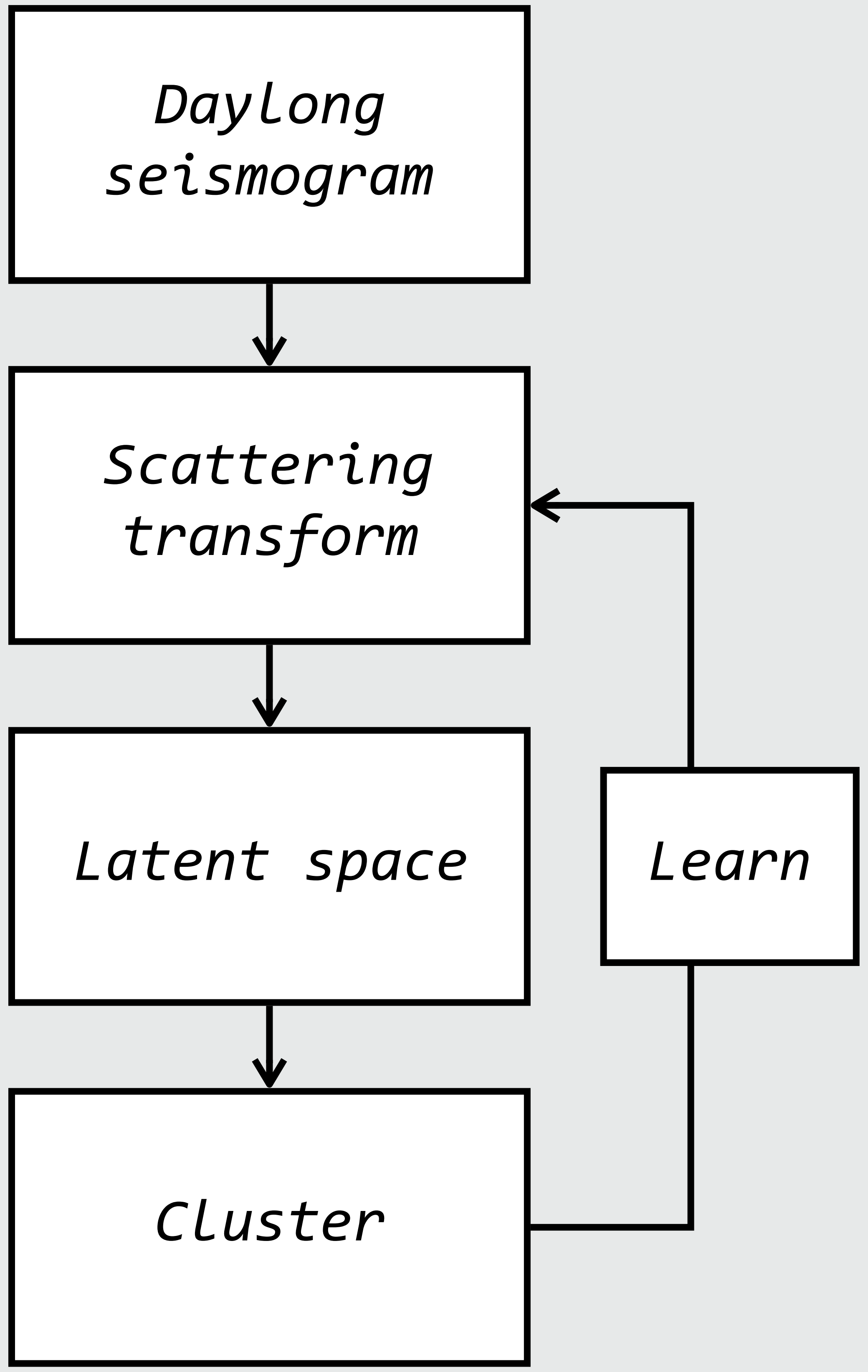


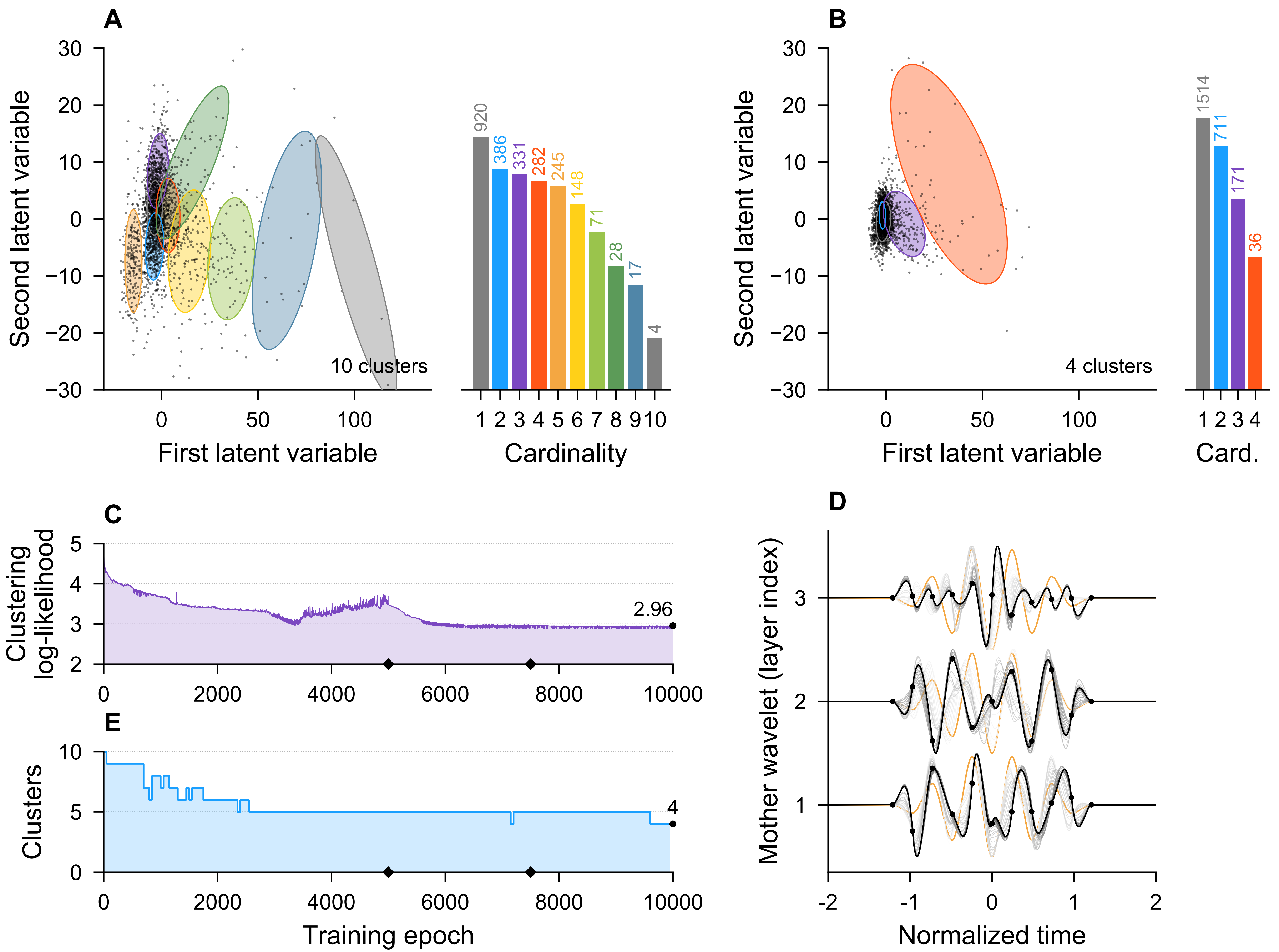


At first, a lot of clusters are found, the data is scattered

Seydoux et al. (rev.)

Unsupervised clustering of a daylong seismic record

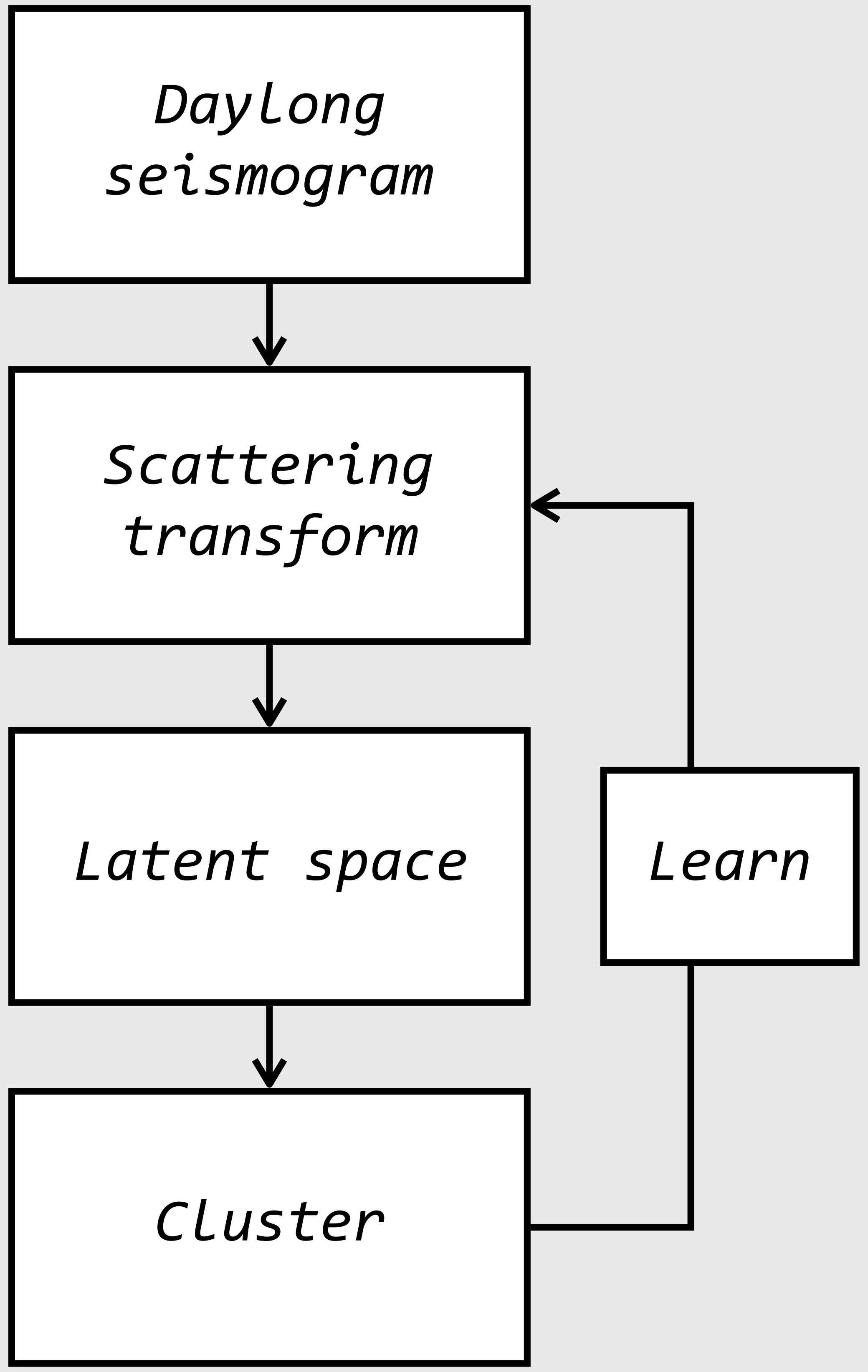


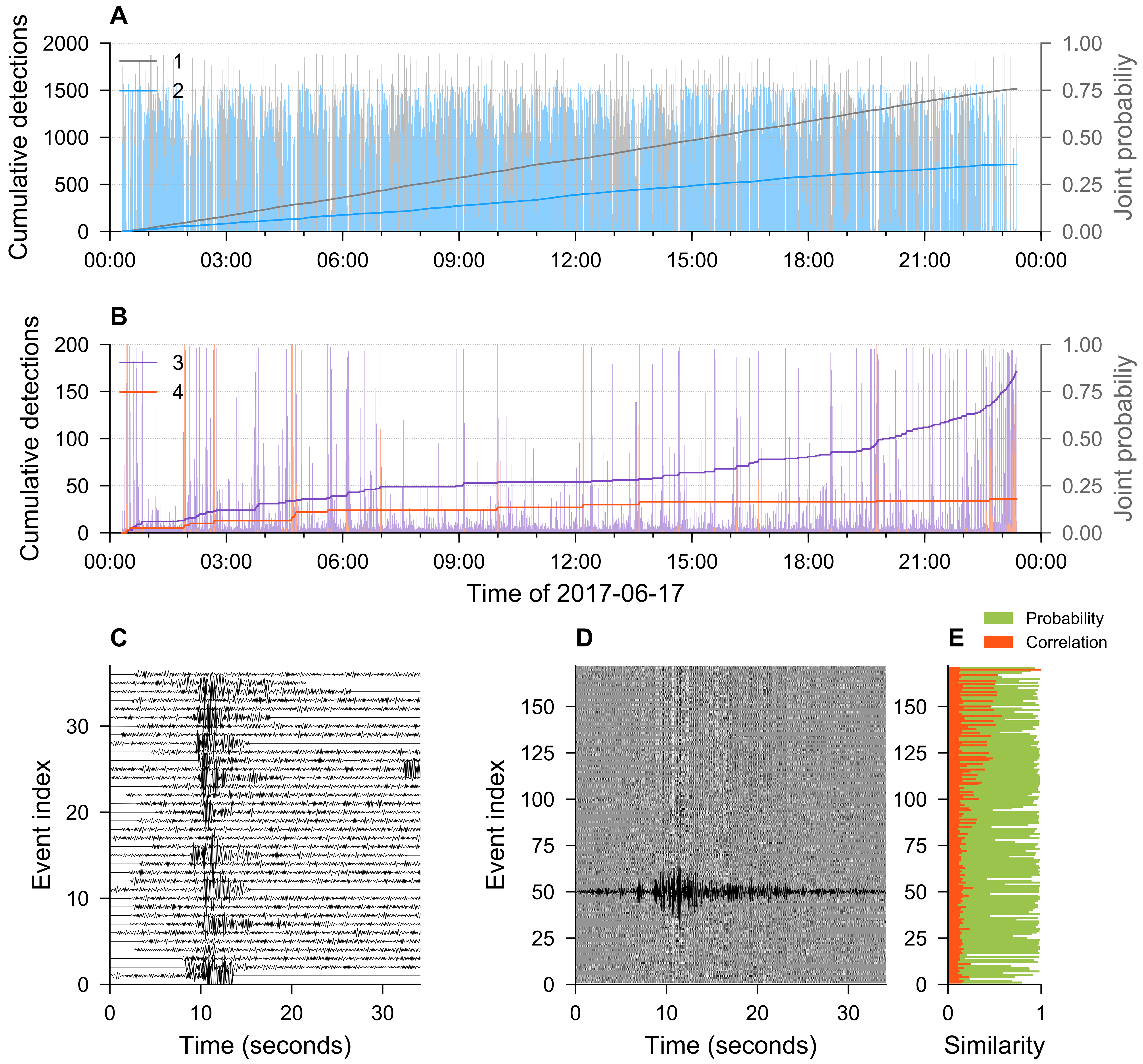


The training concentrate the data around poles improving clusters

Seydoux et al. (rev.)

Unsupervised clustering of a daylong seismic record

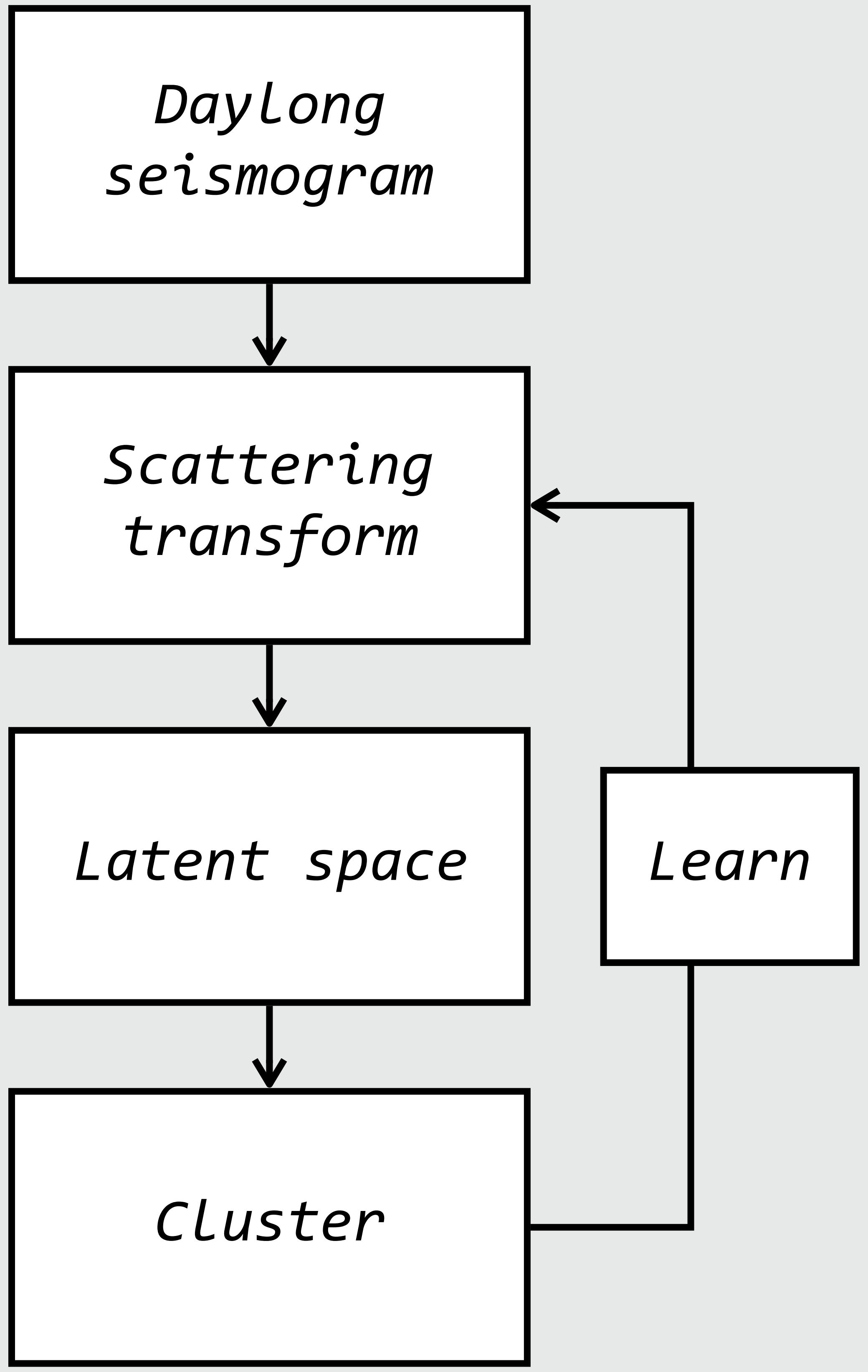


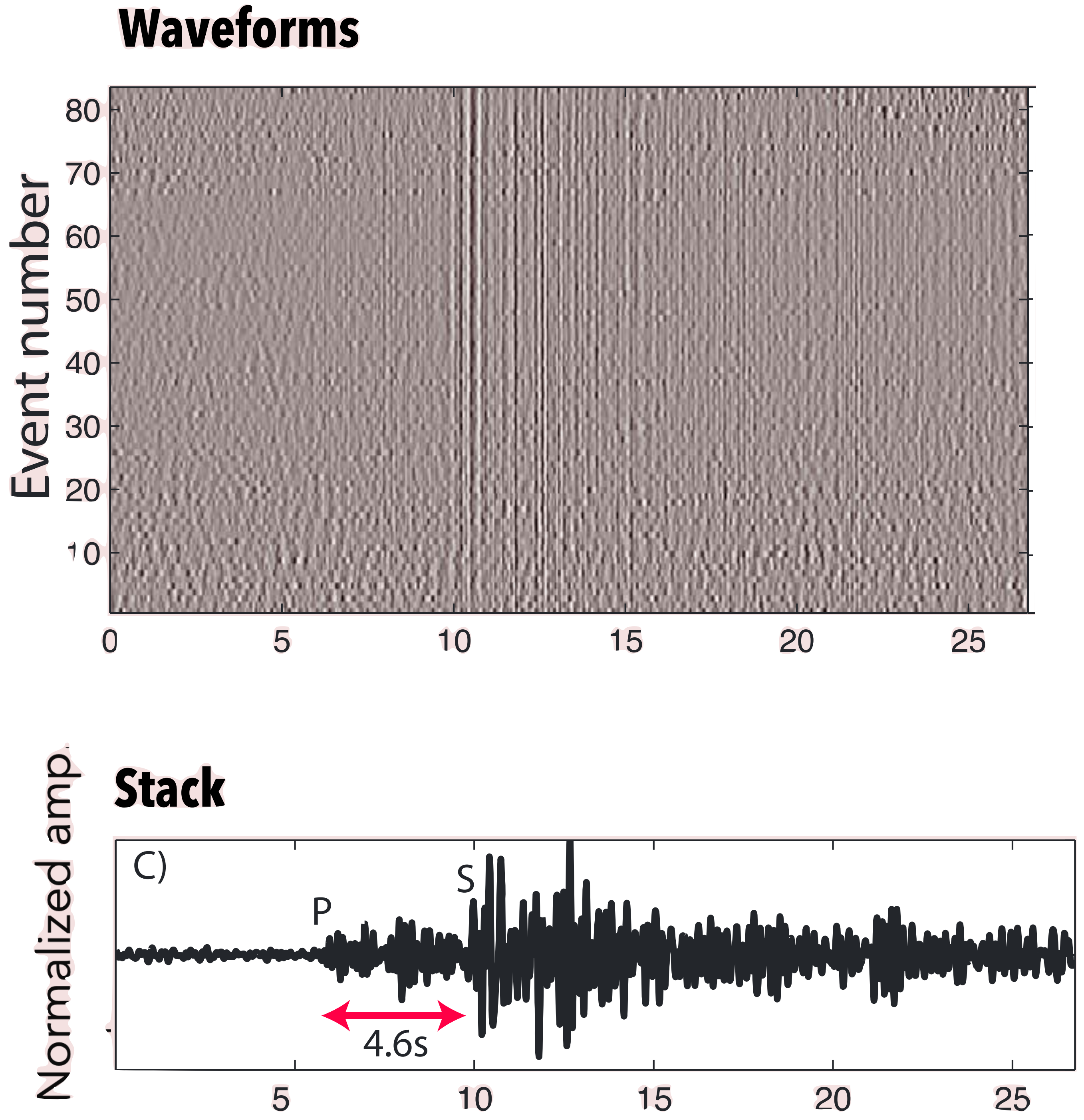


Cumulative detection curves and waveforms analysis

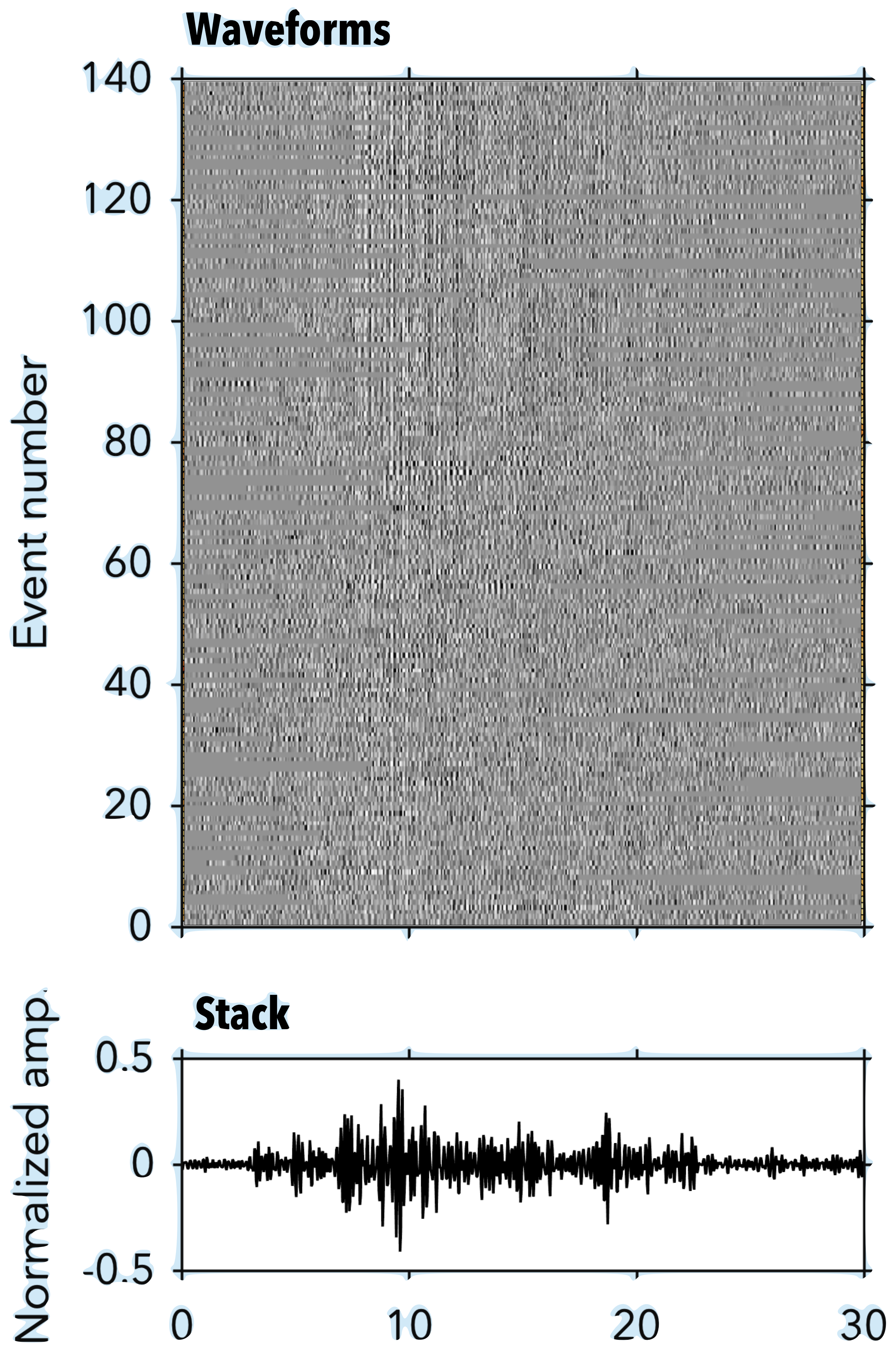
Seydoux et al. (rev.)

Vs. template matching





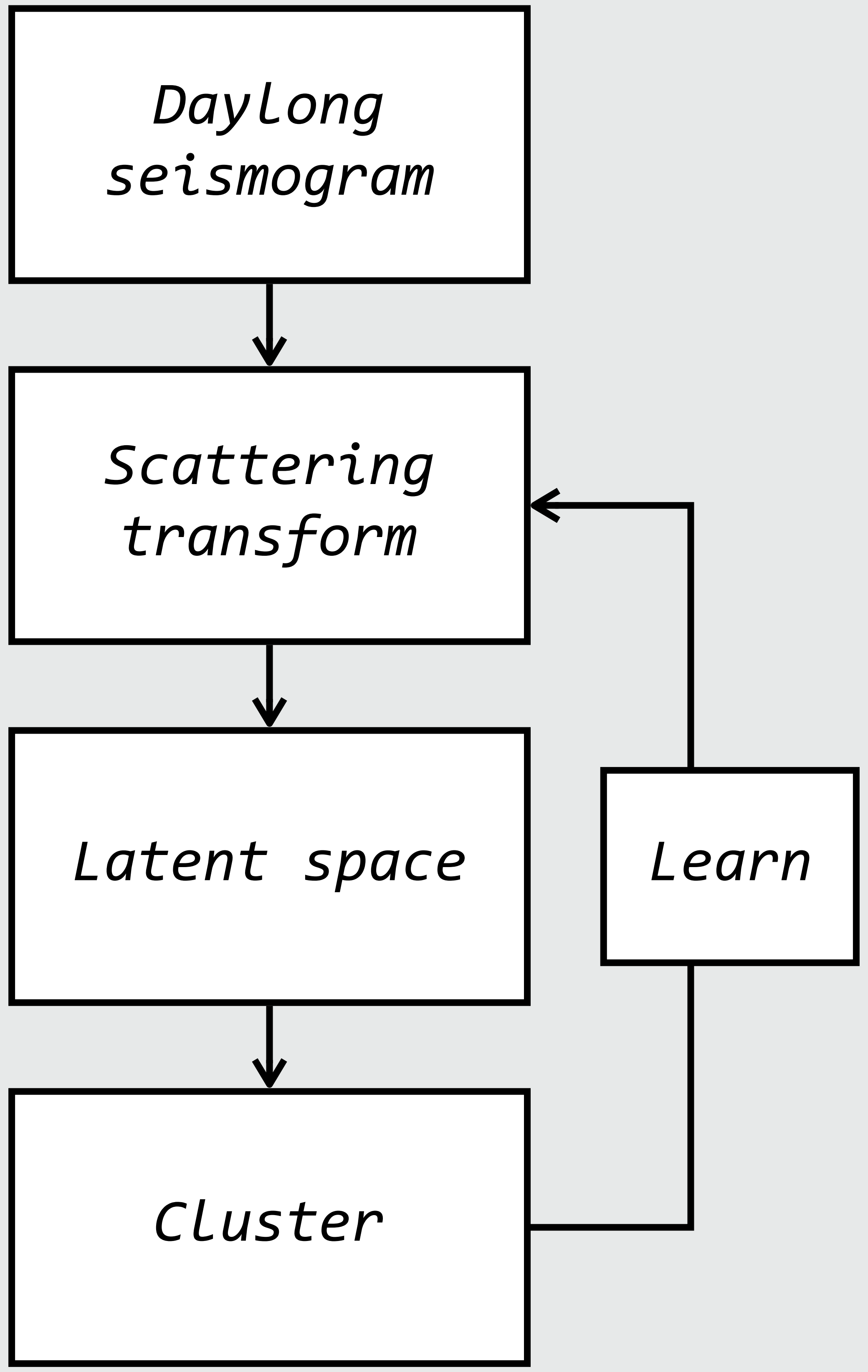
Poli (2017)



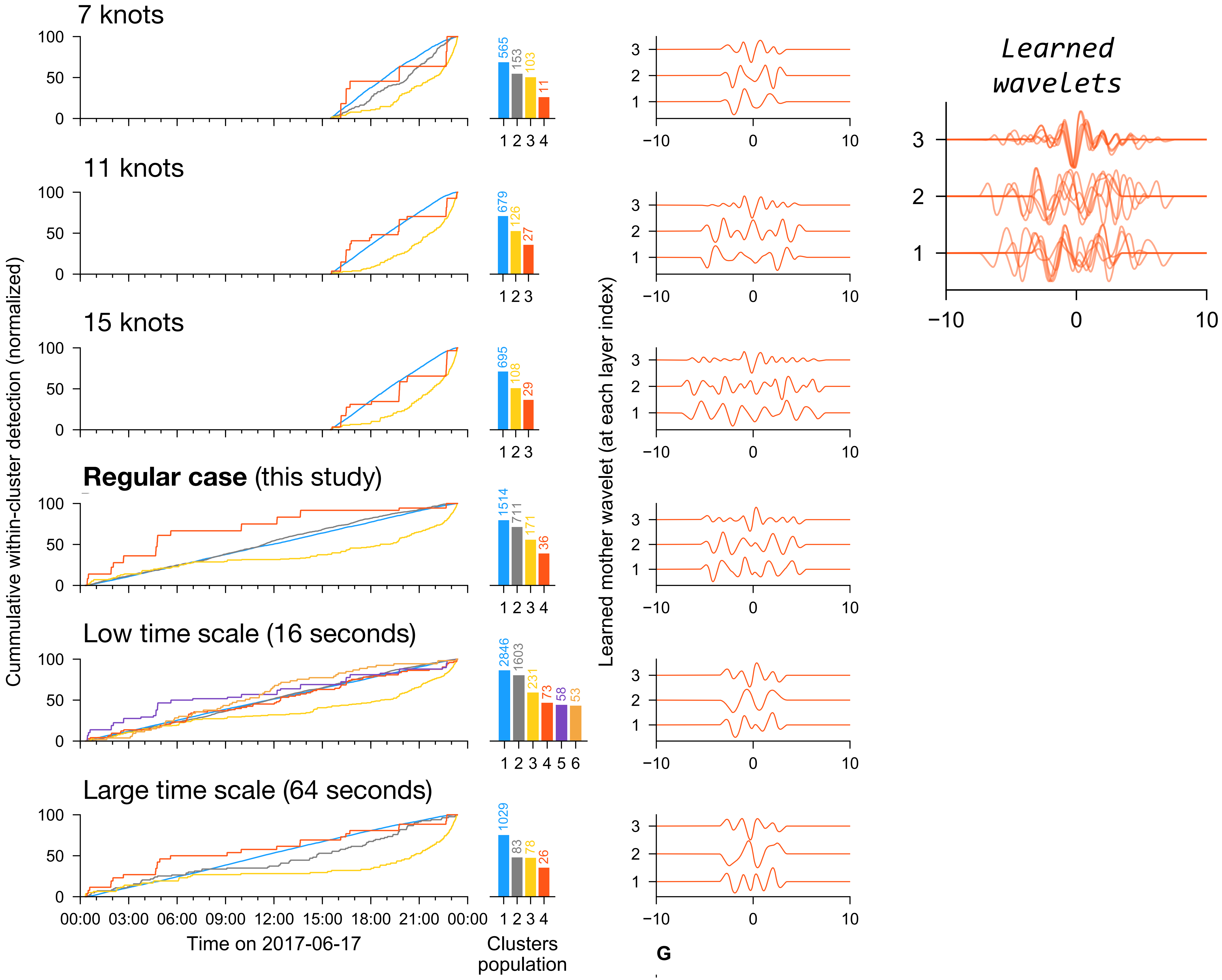
Seydoux et al. (soon)

We recover the precursory template in an unsupervised way

Robustness to scattering network design

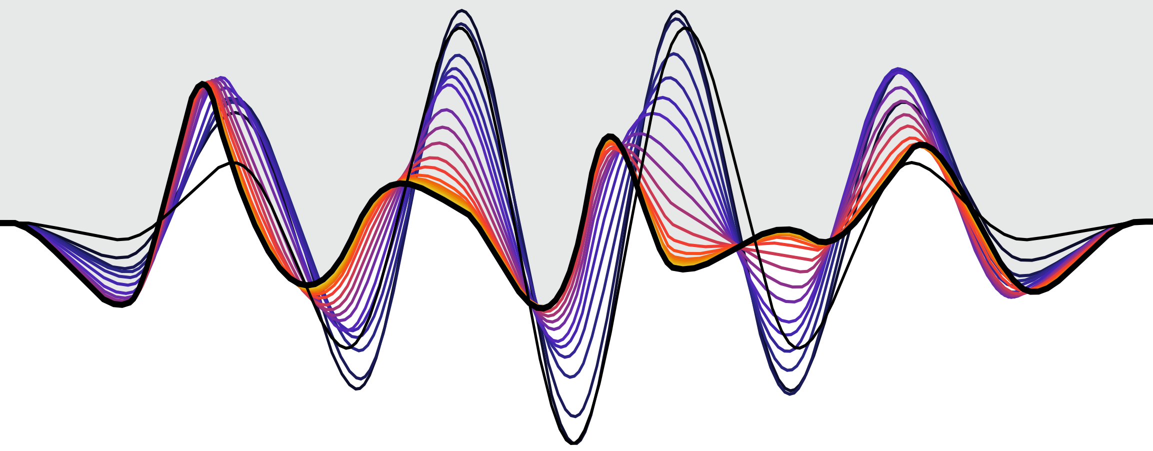


Seydoux et al. (rev.)



Different parameters always recover the precursory pattern

Conclusions



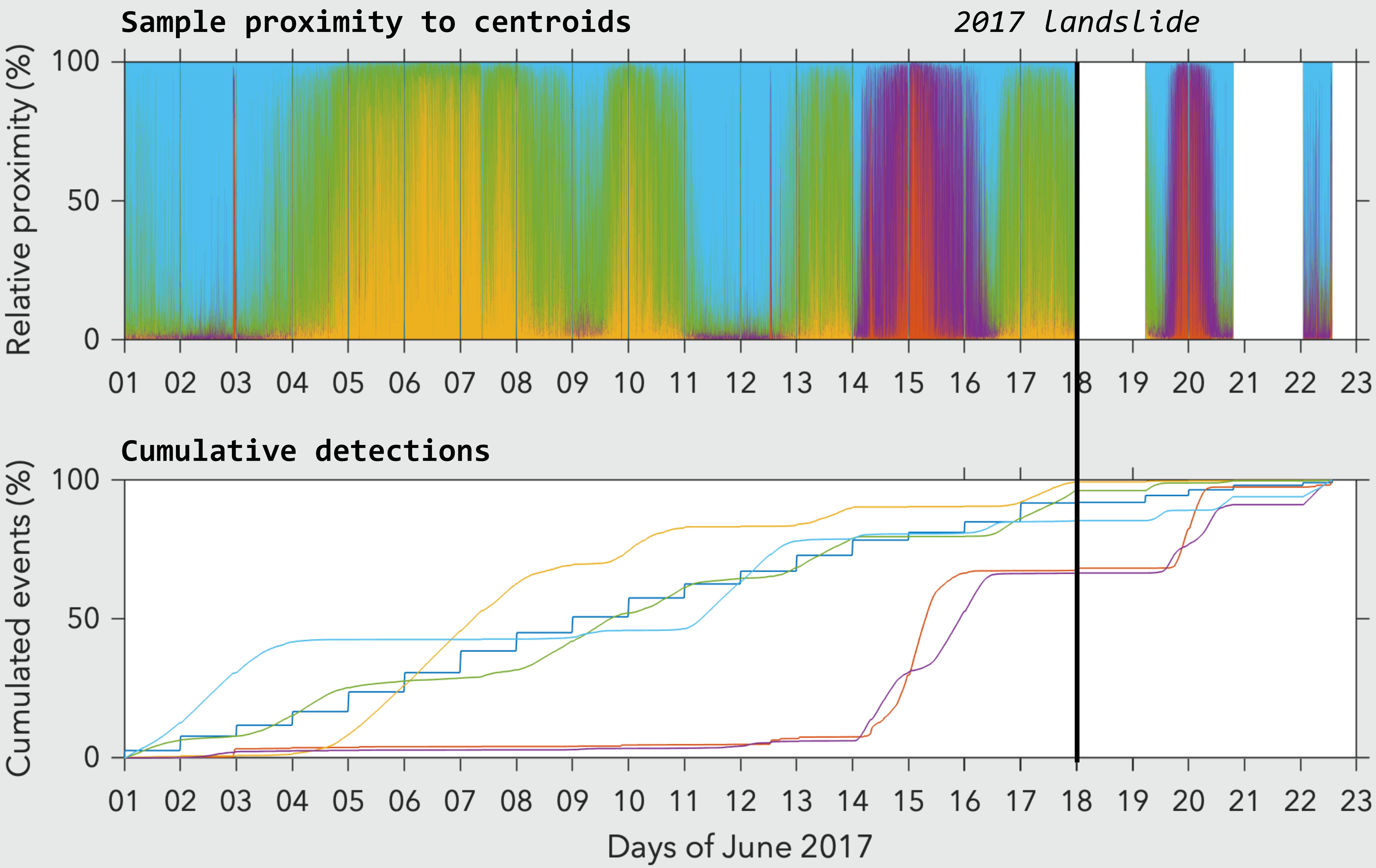
Scattering network is used as a **stable multiple time-scale** representation of the seismic data

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)

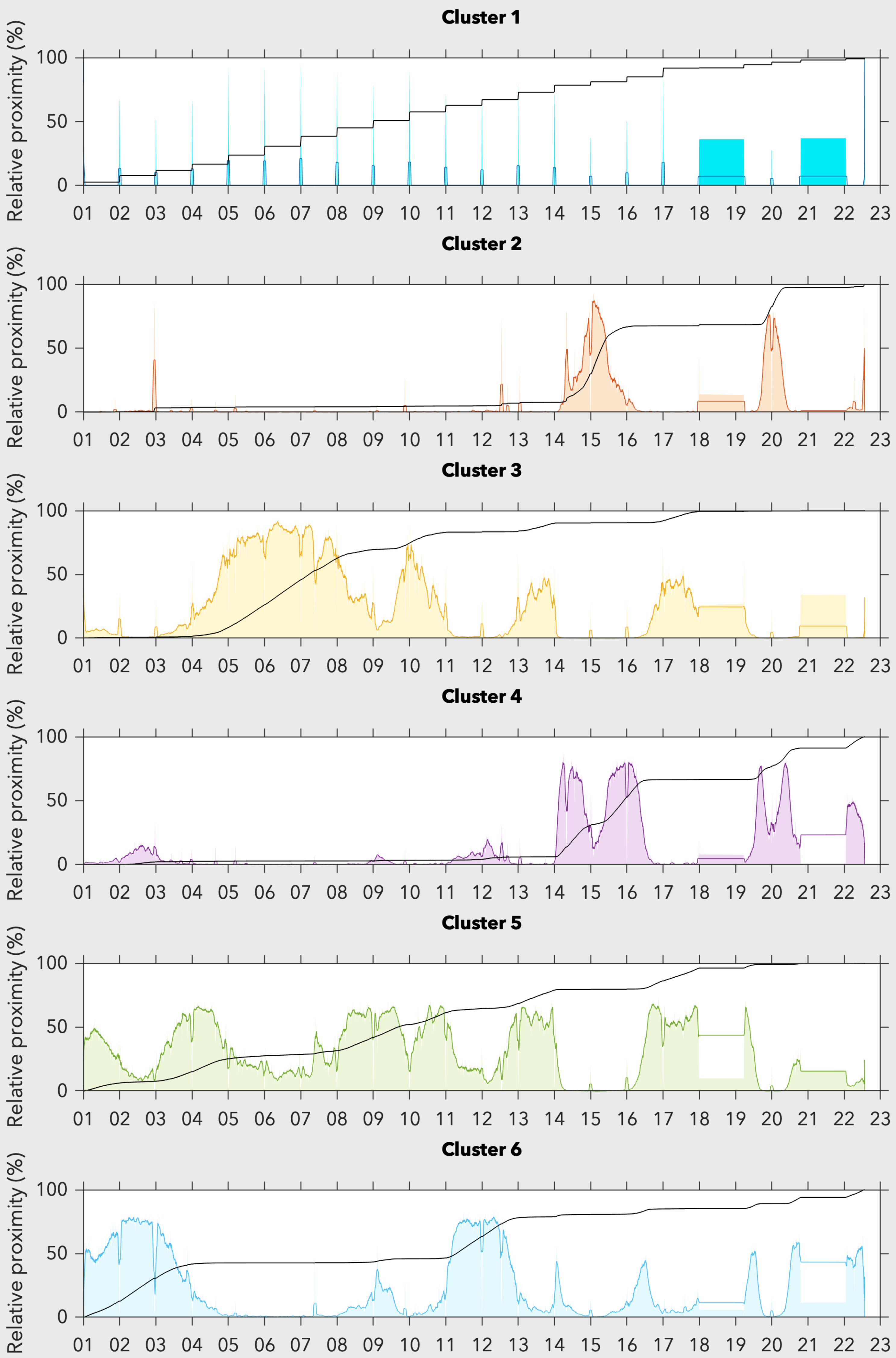
We were able to **blindly recover the precursory repeater** preceding the main landslide rupture

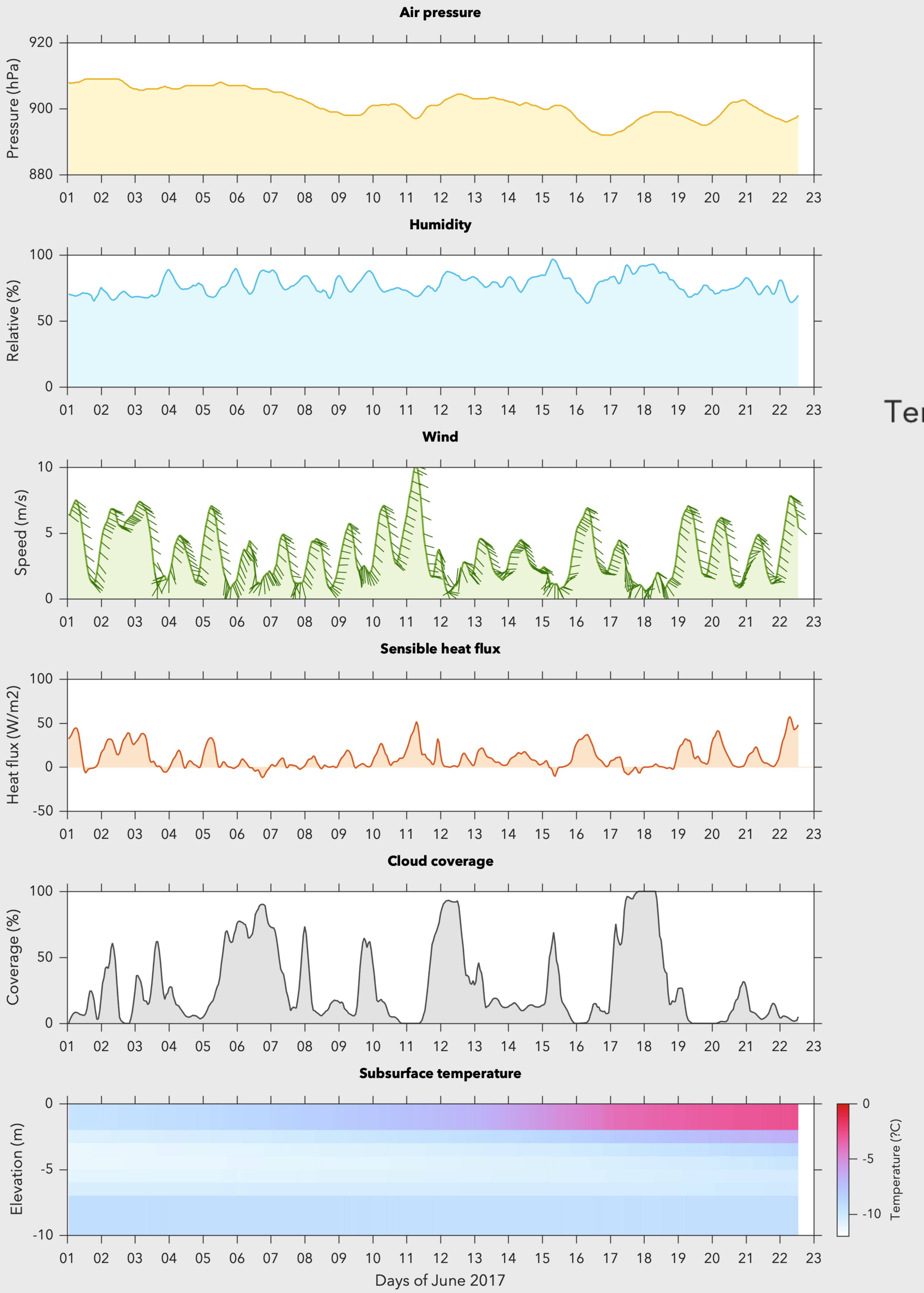
Discussion – at larger time scales

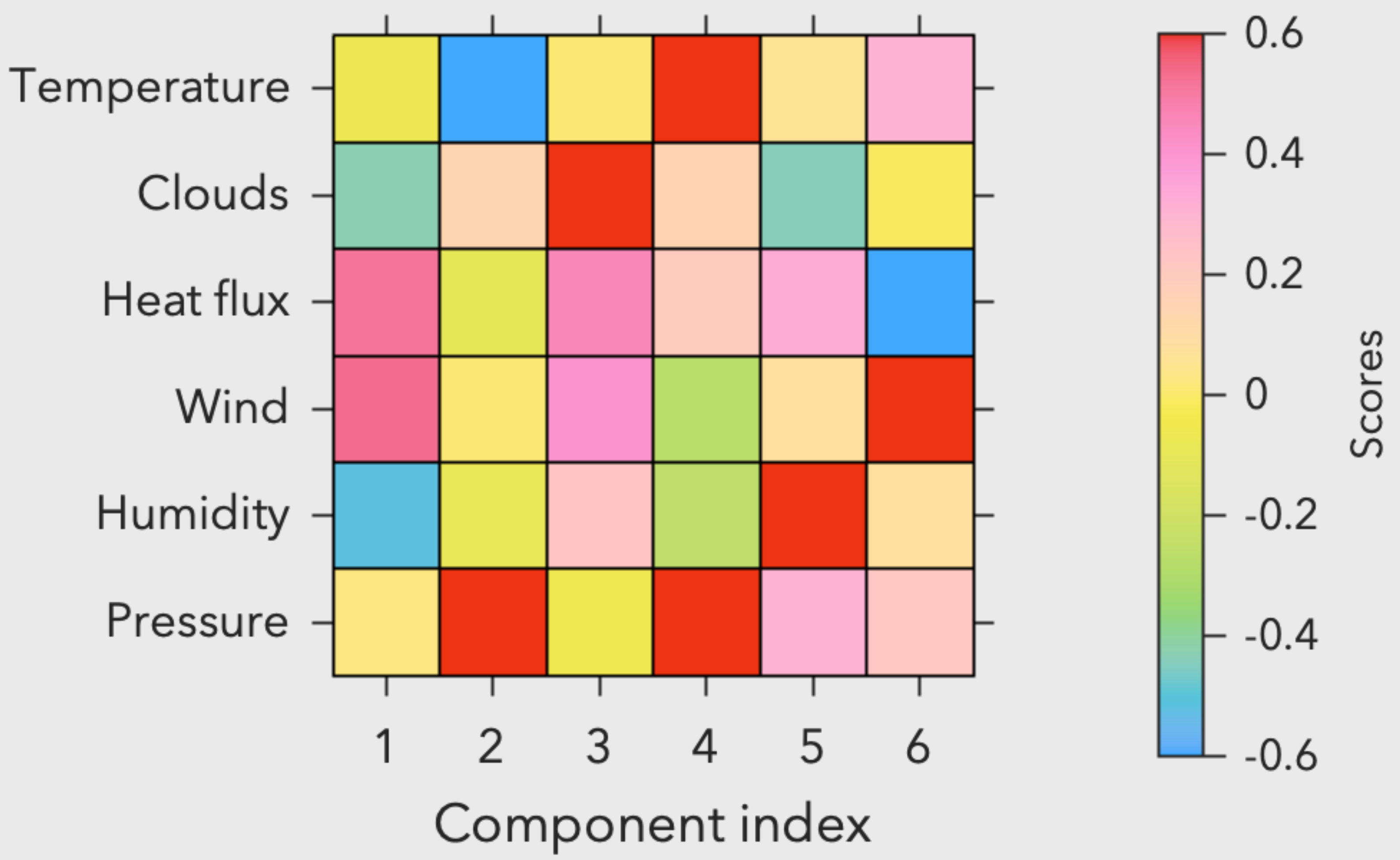


At very large time scales with do not see the precursory signal anymore (unbalanced) and observe different clusters. What are they related to?

Discussion – clusters versus meteorological data

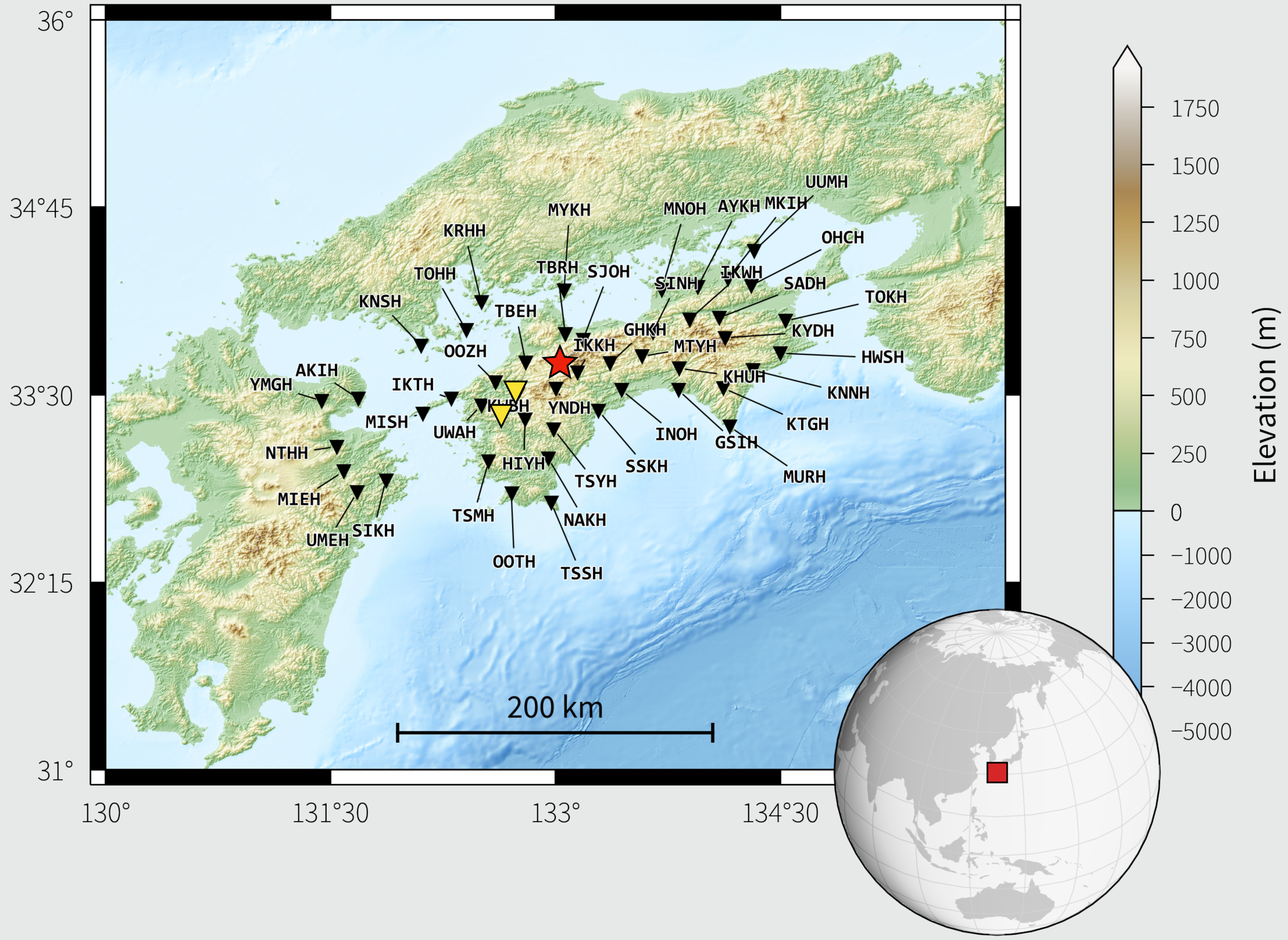




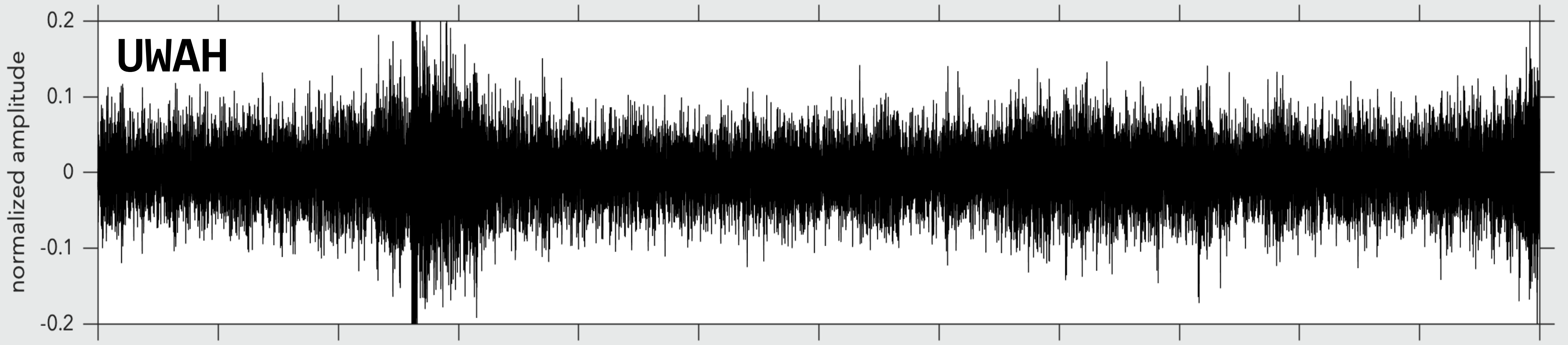


Discussion – towards single-station detection of non-volcanic tremors

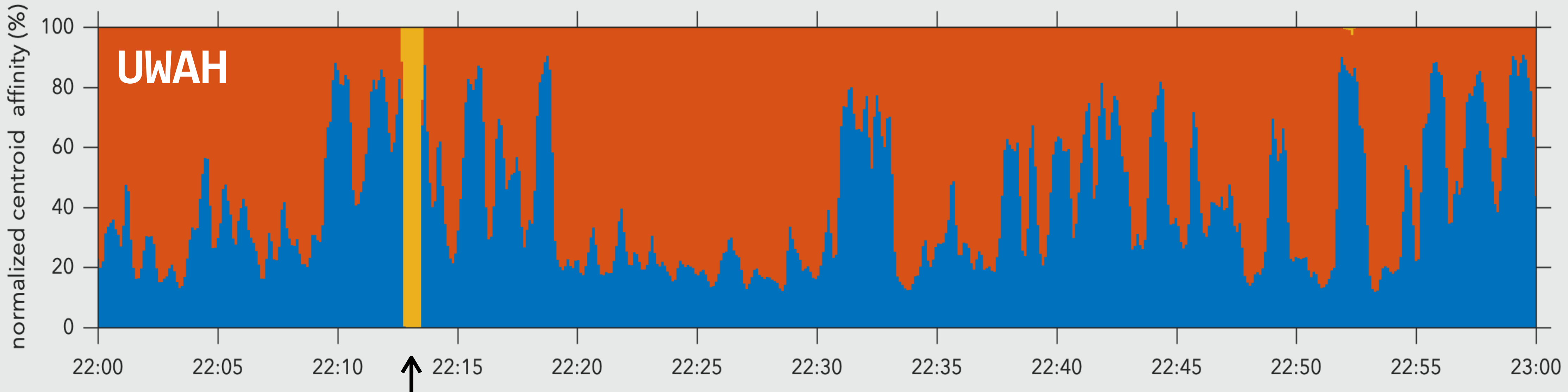
Japan hi-net



Broadband records at two stations located 50 km apart

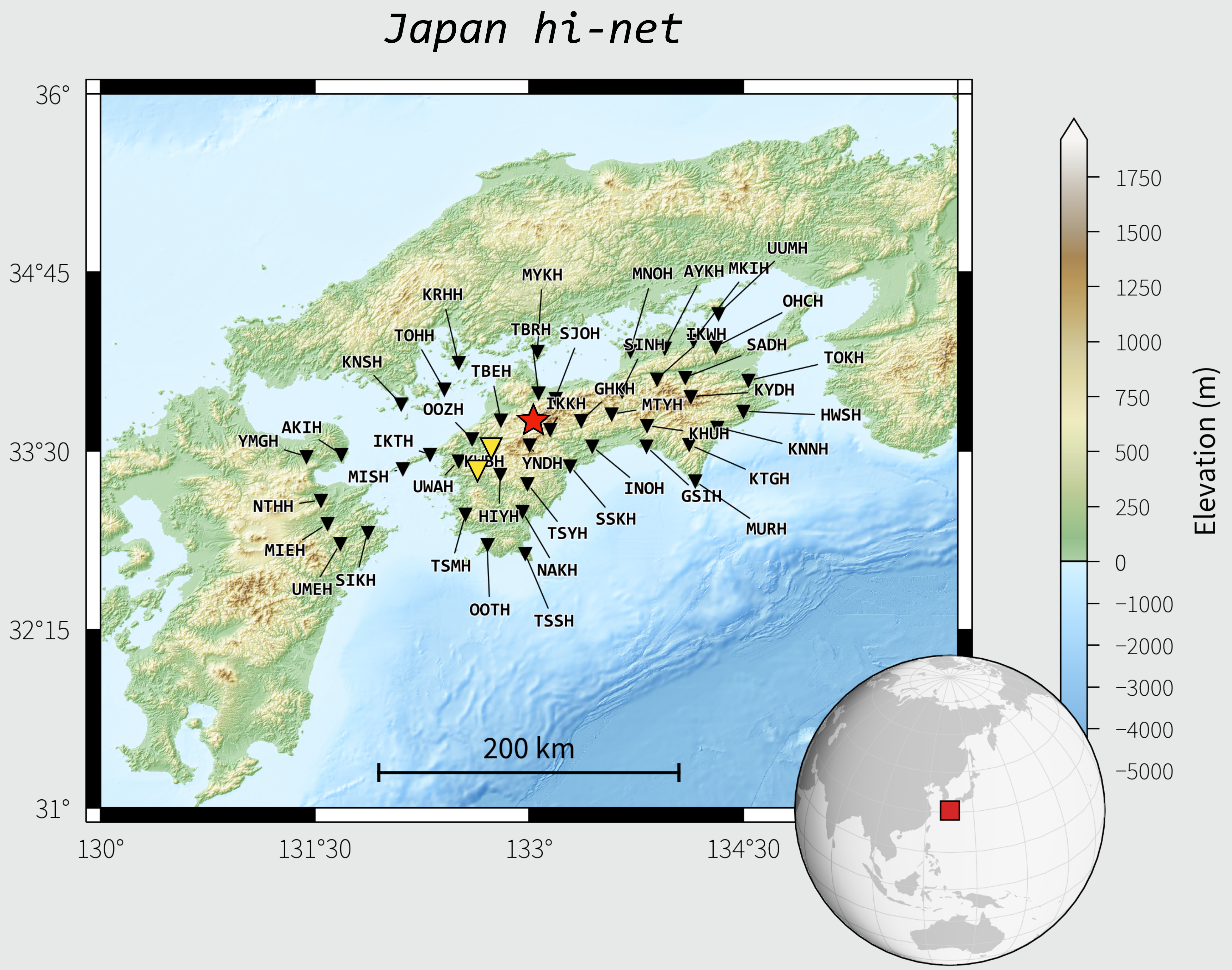


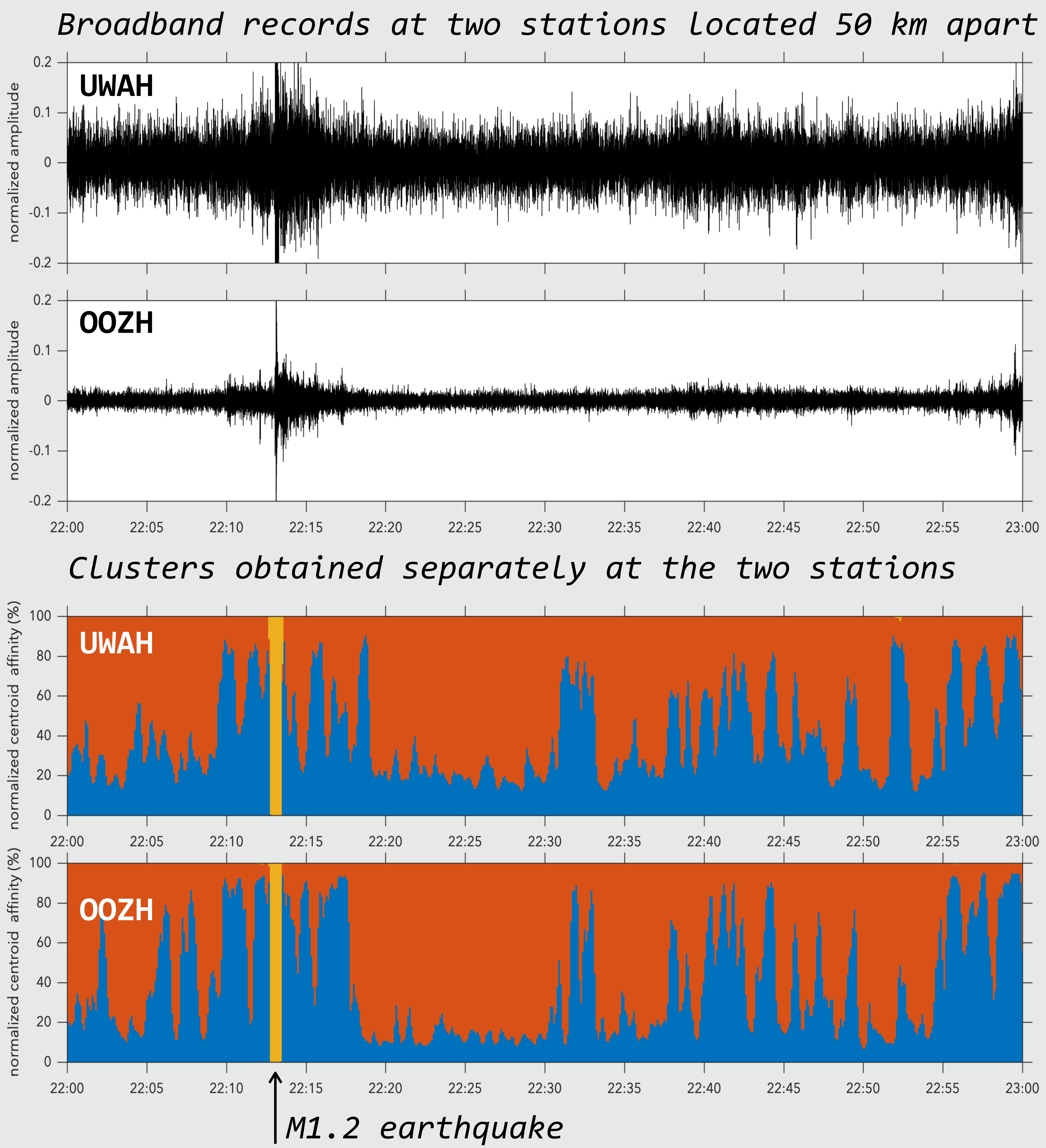
Clusters obtained separately at the two stations



M1.2 earthquake

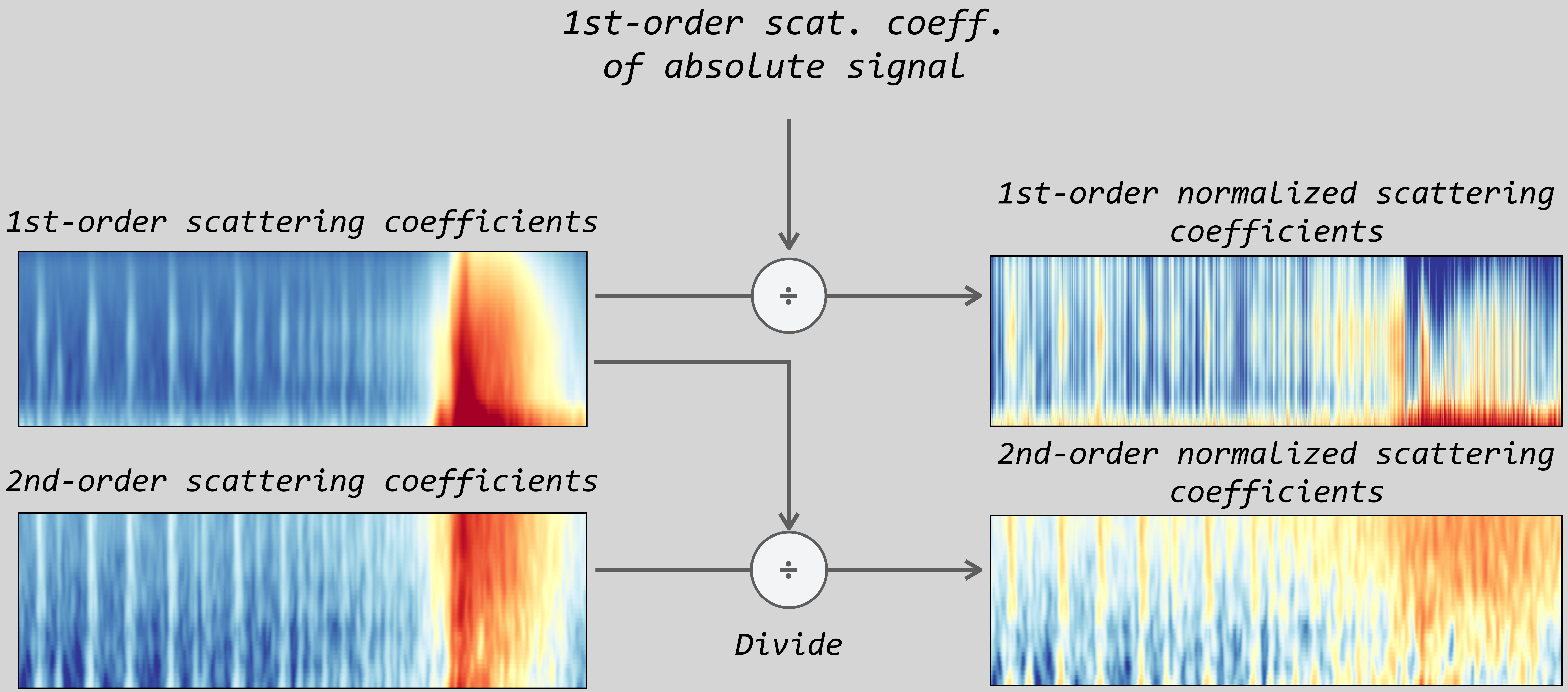
Discussion – towards single-station detection of non-volcanic tremors





Two continuous records independently analyzed lead to the same clusters

Appendix – parental normalization of the scattering coefficients

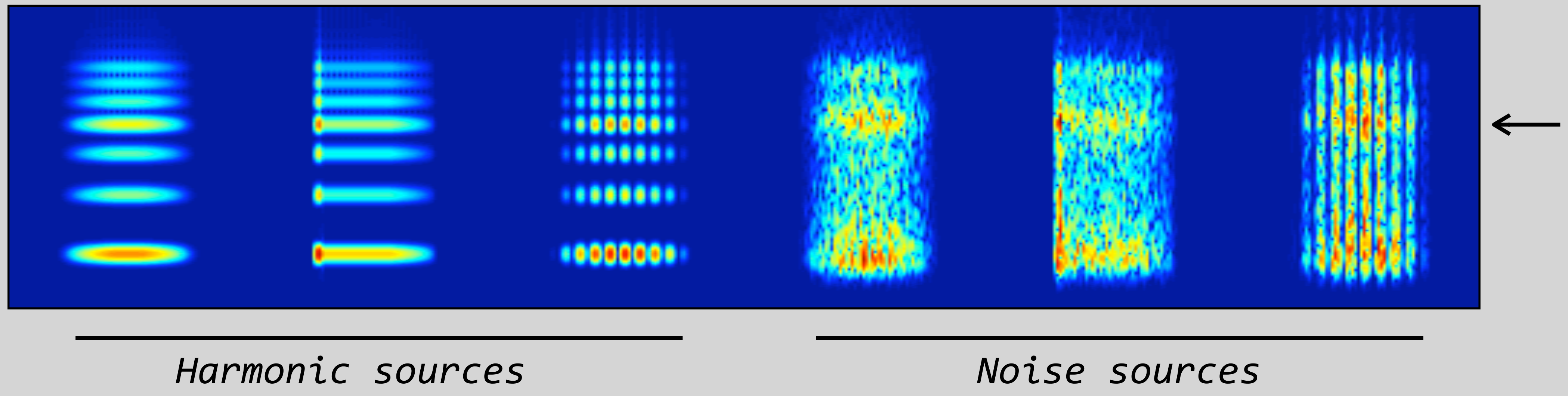


Several order of magnitude of amplitude difference between signals in the seismic data.
We normalize the amplitude w.r.t. the parent scattering coefficients.

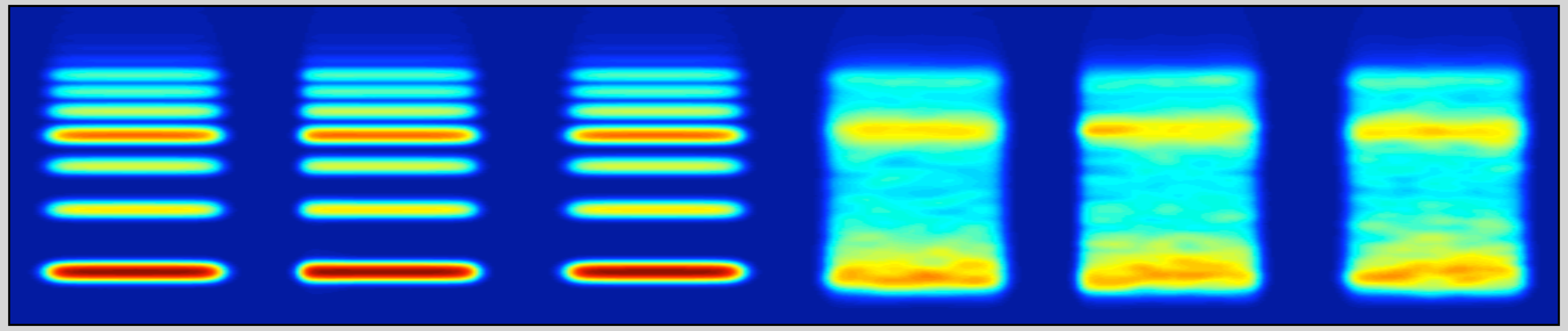
Siffre et al. (2013)

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

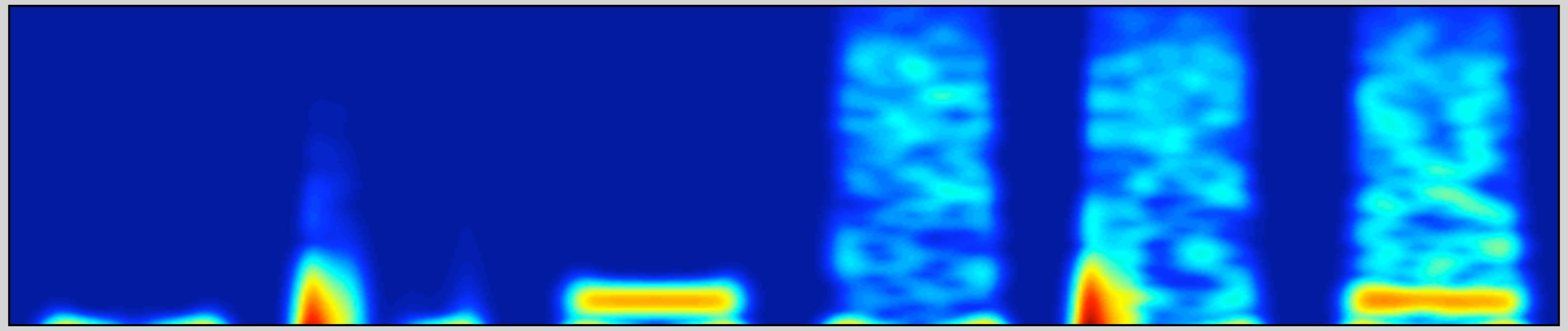
Scalogram



1st order scattering coefficients



2nd order scattering coefficients

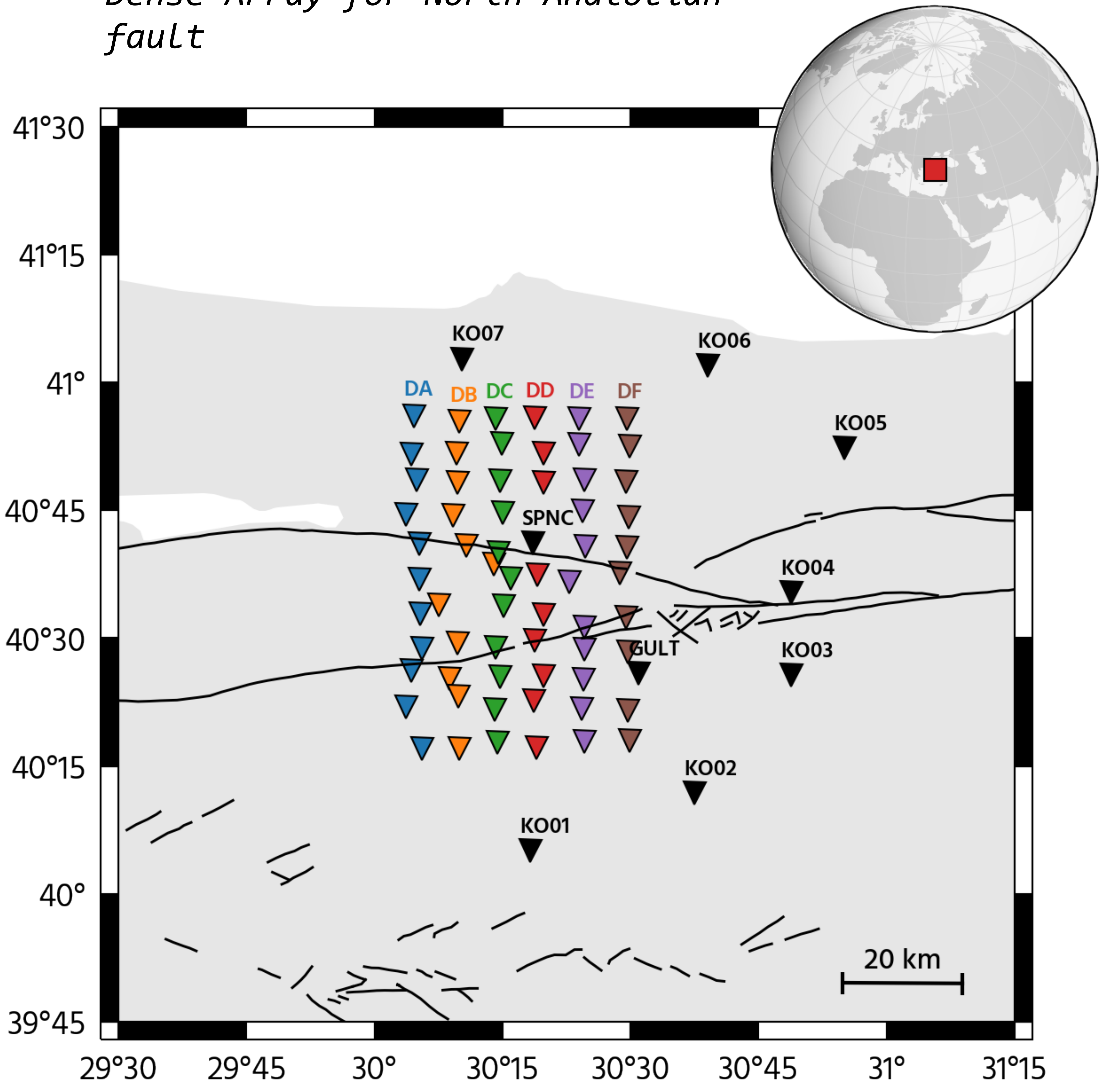


Anden & Mallat IEEE (2014)

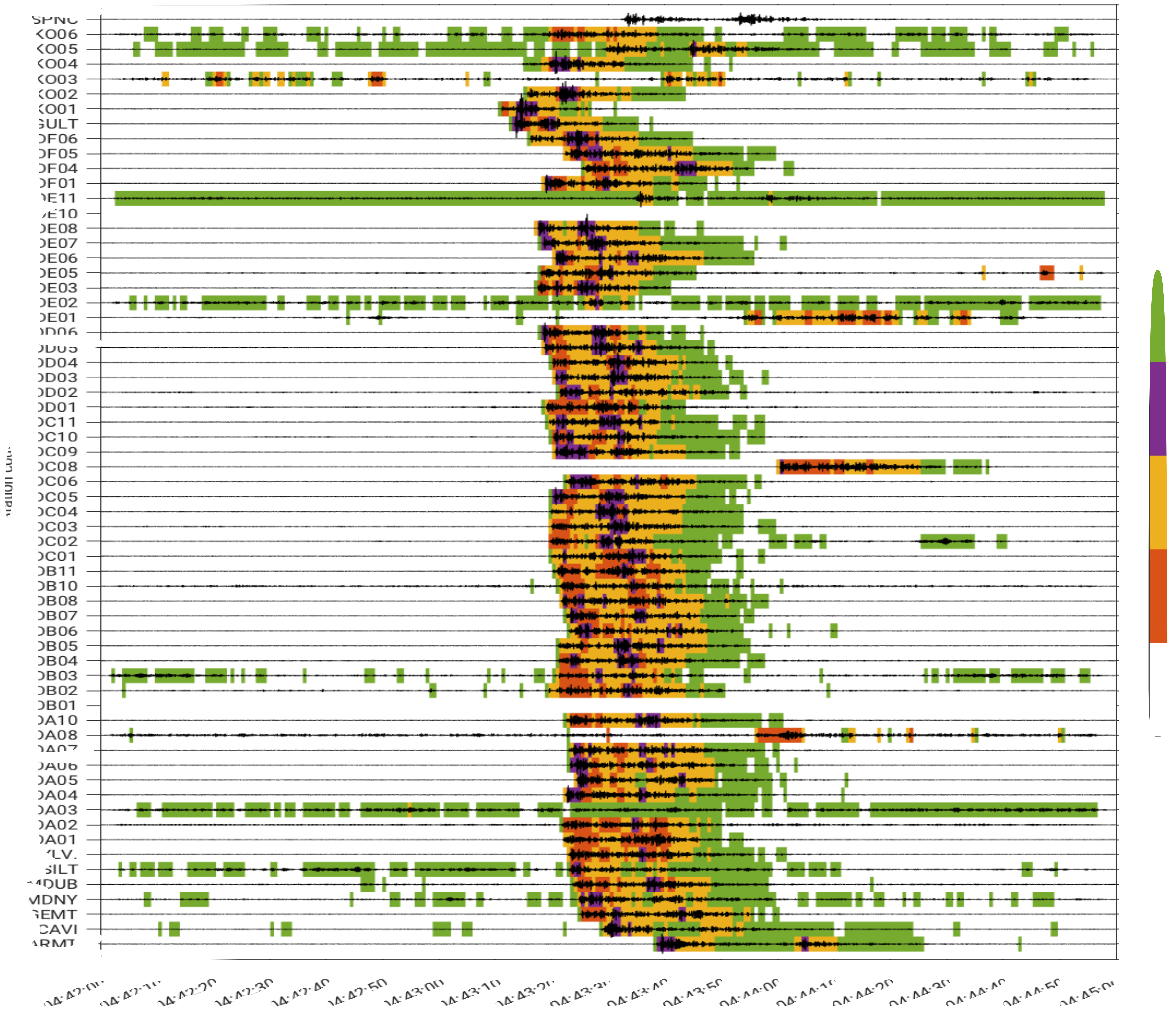
B

Ongoing work – differentiate between seismic phases

Dense Array for North Anatolian fault



Analysis of a M1.6 earthquake



C