Intelligent Signal Processing

Non-Linear PCA
and
ICA

Angelo Ciaramella

Principal Component Analysis

- Principal Component Analysis (PCA) is a statistical technique
  - Dimensionality reduction
  - Lossy data compression
  - Feature extraction
  - Data visualization
- It is also known as the *Karhunen-Loeve* transform
- PCA can be defined as the principal subspace such that the variance of the projected data is maximized

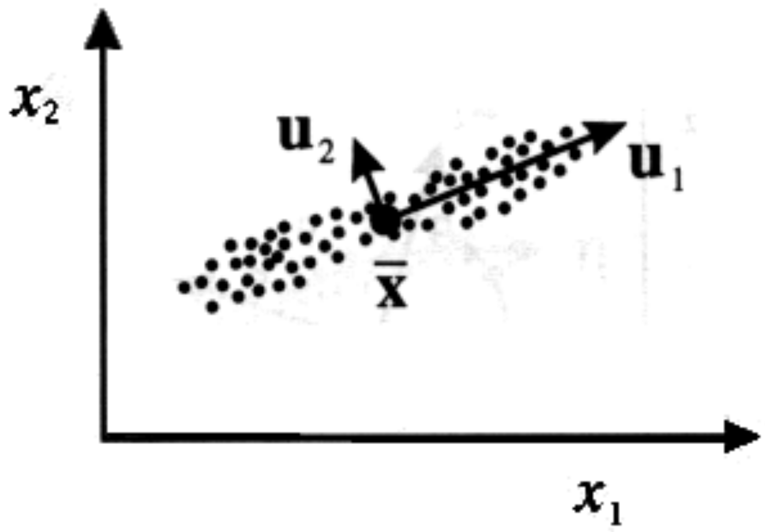
Second-Order methods

- The second-order methods are the most popular methods to find a linear transformation

- This methods find the representation using only the information contained in the covariance matrix of the data vector **x**

- PCA is widely used in signal processing, statistics, and neural computing

Principal Components



In a linear projection down to one dimension, the optimum choice of projection, in the sense of minimizing the sum-of-squares error, is obtained first subtracting off the mean of the data set, and then projecting onto the first eigenvector $\mathbf{u}_1$ of the covariance matrix.

Projection error minimization

- We introduce a complete orthonormal set of *D*-dimensional basis vectors (i=1,…,*D*)

$$\boxed{\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}}$$

- Because this basis is complete, each data point can be represented by a linear combination of the basis vectors

$$\boxed{\mathbf{x}_n = \sum_{i=1}^{D} \alpha_{ni} \mathbf{u}_i}$$

Projection error minimization

- We can write also that

$$\boxed{\mathbf{x}_n = \sum_{i=1}^{D} \left(\mathbf{x}_n^T \mathbf{u}_i\right)\mathbf{u}_i} \longleftarrow \boxed{\alpha_{nj} = \mathbf{x}_n^T \mathbf{u}_j}$$

- Our goal is to approximate this data point using a representation involving a restricted number *M* < *D* of variables corresponding to a projection onto a lower-dimensional subspace

$$\boxed{\tilde{\mathbf{x}}_n = \sum_{i=1}^{M} z_{ni}\mathbf{u}_i + \sum_{i=M+1}^{D} b_i\mathbf{u}_i}$$

Projection error minimization

- As our distortion measure we shall use the squared distance between the original point and its approximation averaged over the data set so that our goal is to minimize

$$\boxed{J = \frac{1}{N}\sum_{n=1}^{N}\left\|\mathbf{x}_n - \tilde{\mathbf{x}}_n\right\|^2}$$

- The general solution is obtained by choosing the basis to be eigenvectors of the covariance matrix given by

$$\boxed{\mathbf{S}\mathbf{u}_i = \lambda_i \mathbf{u}_i}$$

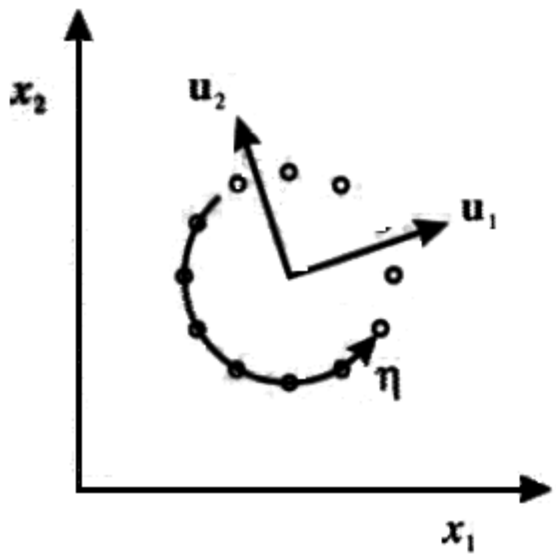
Projection error minimization

- The corresponding value of the distortion measure is then given by

$$\boxed{J = \sum_{i=M+1}^{D} \lambda_i}$$

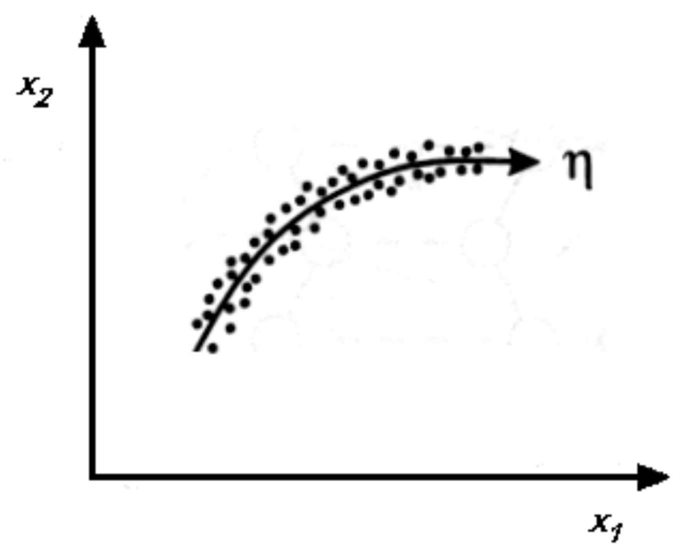
- We minimize this error selecting the eigenvectors defining the principal subspace are those corresponding to the *M* largest eigenvalues

Complex distributions



A linear dimensionality reduce technique, such as PCA, is unable to detect the lower dimensionality. In this case PCA gives two eigenvectors with equal eigenvalues. The data can described by a single eigenvalue

Addition of a small level of noise to data having an intrinsic. Dimensionality to 1 can increase its intrinsic dimensionality to 2. The data can be represented to a good approximation by a single variable η and can be regarded as having an intrinsic dimensionality of 1.



Unsupervised Neural Networks

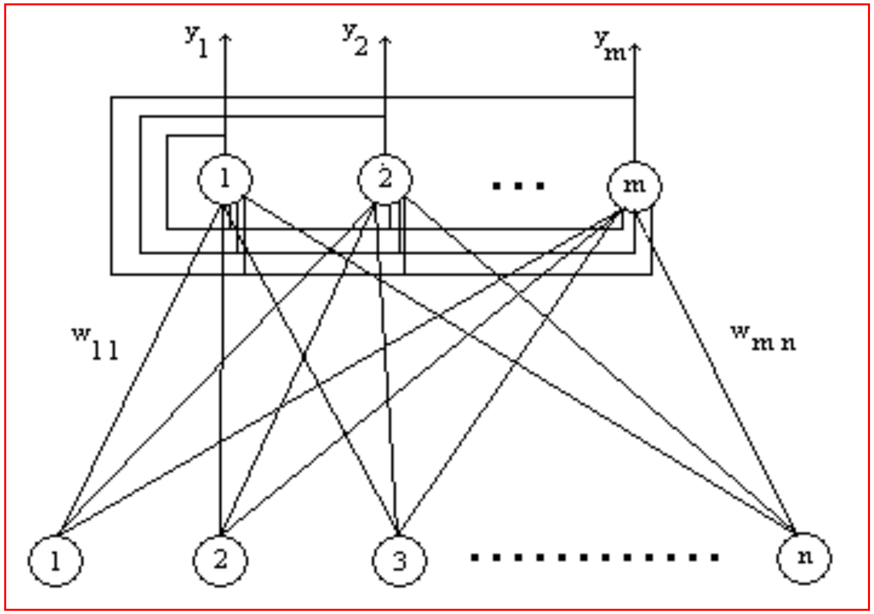
- Typically Hebbian type learning rules are used

- There are two type of NN able to extract the Principal Components:

  - Symmetric (Oja, 1989)

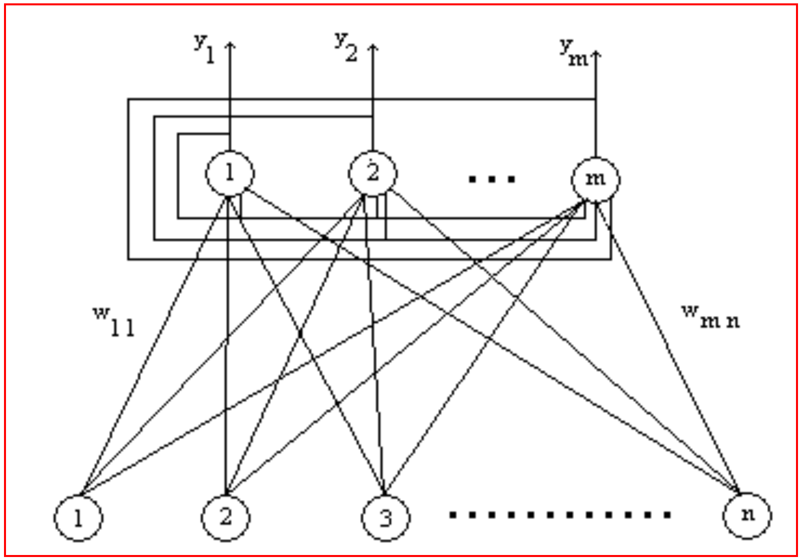
  - Hierarchical (Sanger, 1989)

PCA and Unsupervised Neural Network



Symmetric PCA NN

Single layer Neural Network

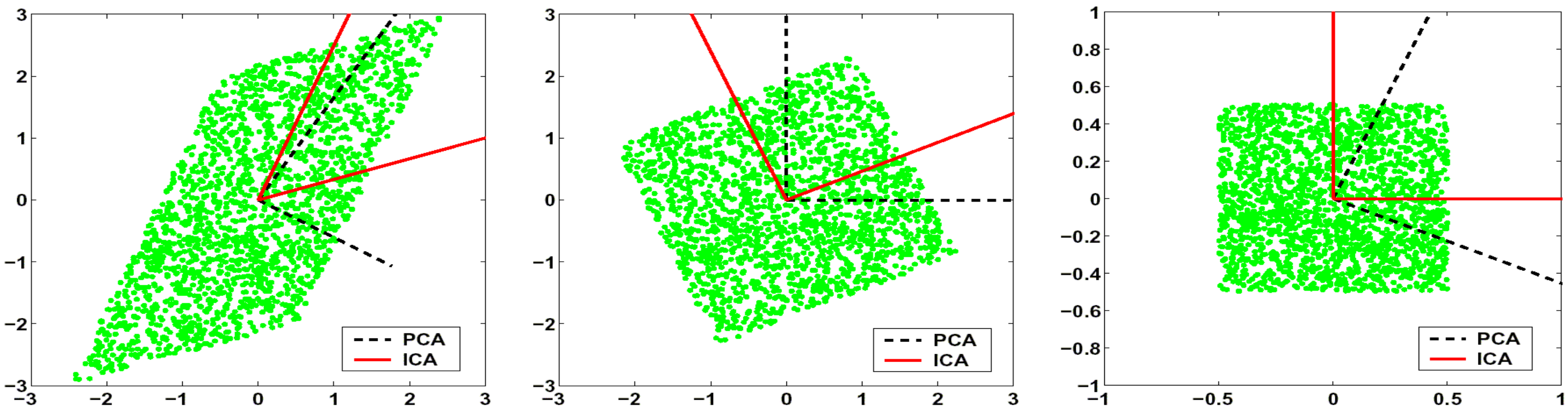


Hierarchical PCA NN

$$\boxed{E\left[\mathbf{y}^2\right] = E\left[\left(\mathbf{w}^T\mathbf{x}\right)^2\right]}$$

Objective function

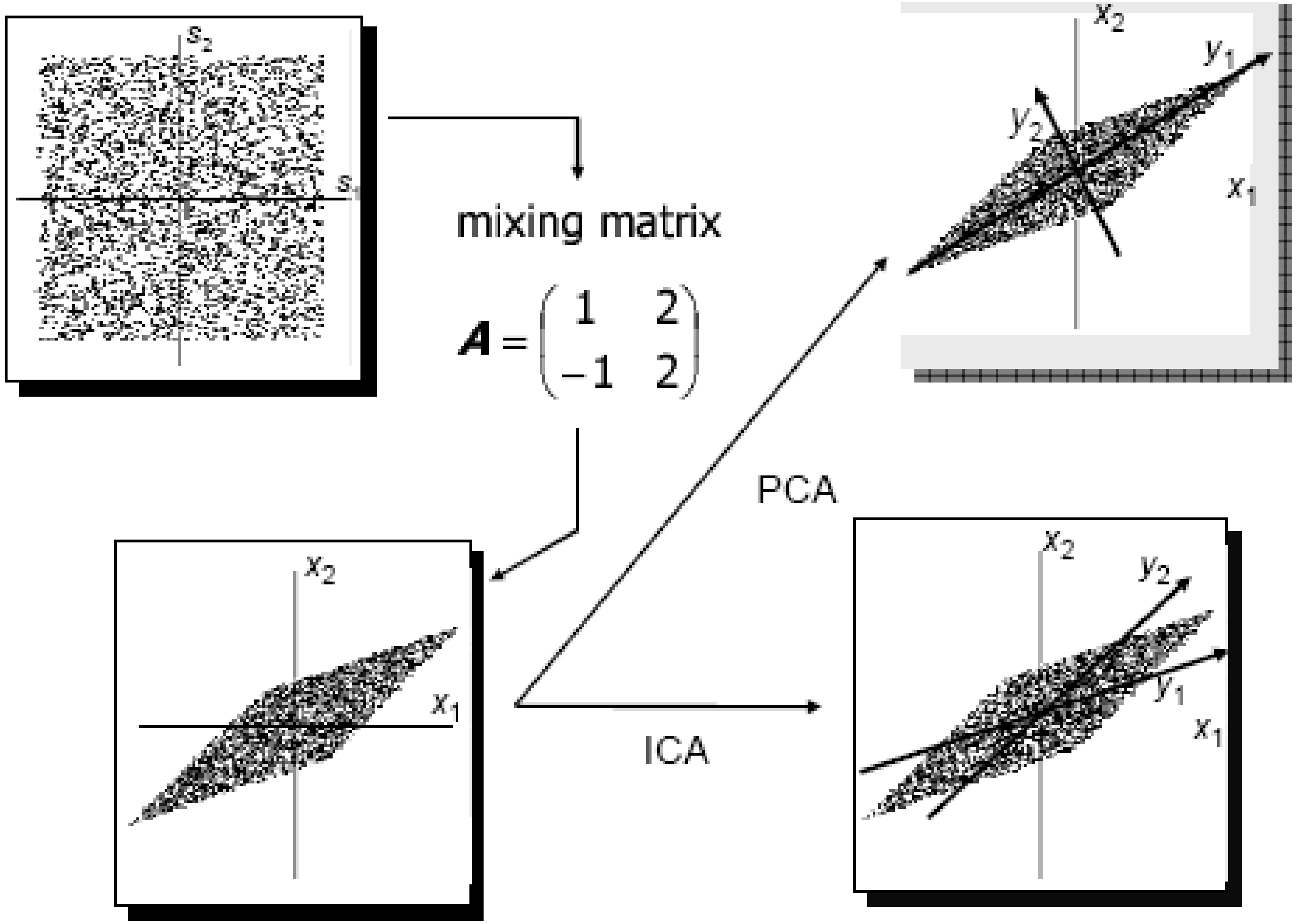
ICA versus PCA



(a) Original (b) After PCA pre-whitening (c) After ICA projection

PCA maximises the variance and projections onto the basis vectors are mixtures. ICA correctly finds the two vectors onto which the *projections are independent*.

Mixing matrix



Non-linear objective function

Maximization

$$\mathbf{x} \xrightarrow{\mathbf{w}\ \text{(weights)}} \boxed{E\left\{f\left(\mathbf{w}^T\mathbf{x}\right)\right\}}$$

L-dimensional vector

where E is the expectation with respect to the (unknown) density of **x** and f(.) is a continue function (e.g. ln cosh(.))

Taylor series

$$\ln\cosh(y) = \frac{1}{2}y^2 - \frac{1}{12}y^4 + \frac{1}{45}y^6 + O(y^8)$$

$$E\{\ln\cosh(y)\} = \tfrac{1}{2}E\left\{(w^Tx)^2\right\} - \tfrac{1}{12}E\left\{(w^Tx)^4\right\} + \tfrac{1}{45}E\left\{(w^Tx)^6\right\} + E\left\{O((w^Tx)^2)\right\}$$

$$C = I \quad \text{and} \quad \tfrac{1}{2}E\left\{(w^Tx)^2\right\} = \tfrac{1}{2}$$

$$-\frac{1}{12}E\left\{(w^Tx)^4\right\}$$

That is dominating, and the kurtosis is optimized

Unsupervised Neural Network

Standard PCA

$$E\left[\mathbf{y}^2\right] = E\left[\left(\mathbf{w}^T\mathbf{x}\right)^2\right]$$

Robust PCA

Nonlinear PCA

$$J_1(\mathbf{w}) = E\left[f\left(\mathbf{x}^T\mathbf{w}\right)\right] + \sum_{j=1}^{I(i)} \lambda_{ij}\left[\mathbf{w}_i^T\mathbf{w}_j - \delta_{ij}\right]$$

$$J_2(e_i) = 1^T E\left[f(\mathbf{x} - \hat{\mathbf{x}}_i)\right]$$

$$E\left[\left|b_i(k)\right|^2\right]$$

$$b_i(k) = \mathbf{x}_k - \sum_{j=1}^{I(i)} g\left(\mathbf{y}_j(k)\right)\mathbf{w}_j(k)$$

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu_k g(\mathbf{y}_i(k))e_i(k)$$

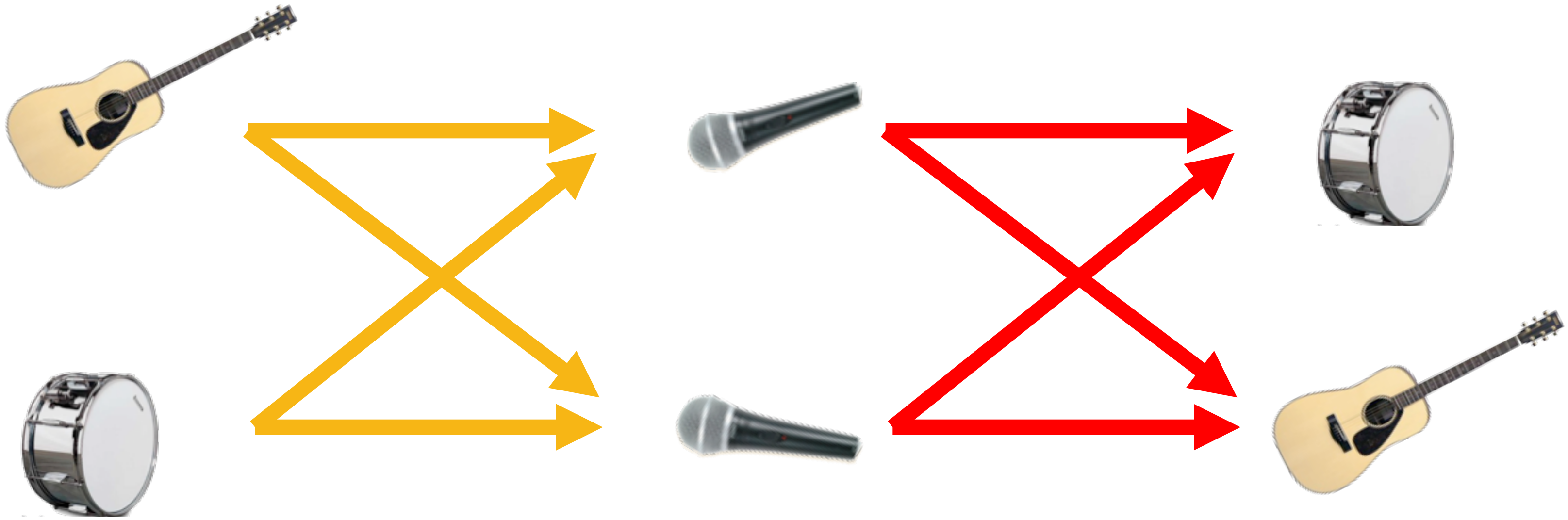
$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu_k\begin{pmatrix}\mathbf{w}_i(k)^T g(e_i(k))\mathbf{x}_k + \\ + \mathbf{x}_k^T\mathbf{w}_i(k)g(e_k(i))\end{pmatrix}$$

$$\mathbf{w}_i(k+1) = \mathbf{w}_i(k) + \mu g\left(\mathbf{y}_i(k)\right)b_i(k)$$

Descendent gradient algorithm

$$e(k) = \mathbf{x}_k - \sum_{j=1}^{I(i)} \mathbf{y}_j(k)\mathbf{w}_j(k)$$

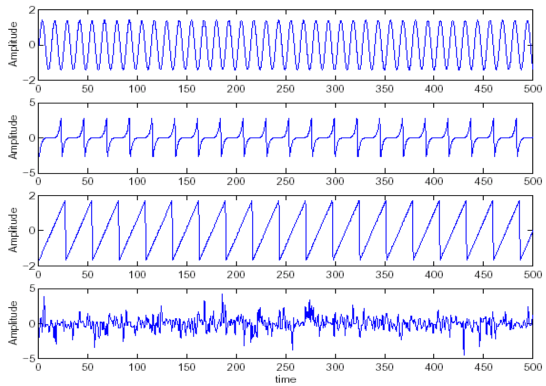
Cocktail party



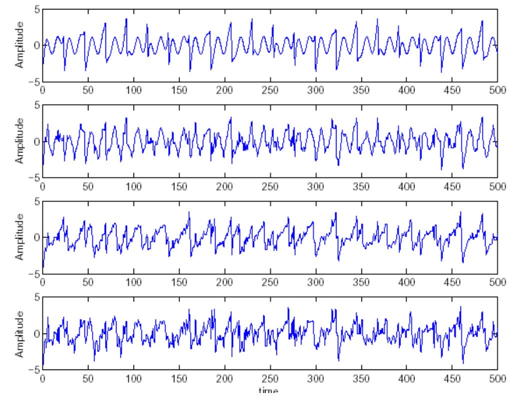
Sources Mixtures Estimated-Sources

$\mathbf{s}$ $\mathbf{A}$ $\mathbf{x}$ $\mathbf{W}$ $\mathbf{y}$

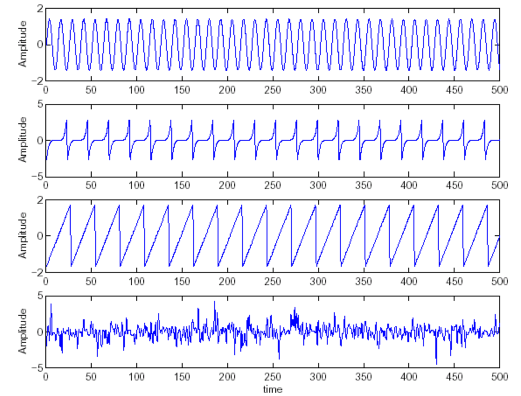
Source estimation



Source signals



Mixed signals



Estimated signals

$$\boxed{\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \\ x_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) \end{aligned}}$$

$$\boxed{\begin{aligned} y_1(t) &= w_{11}x_1(t) + w_{12}x_2(t) + w_{13}x_3(t) \\ y_2(t) &= w_{21}x_1(t) + w_{22}x_2(t) + w_{23}x_3(t) \\ y_3(t) &= w_{31}x_1(t) + w_{32}x_2(t) + w_{33}x_3(t) \end{aligned}}$$

x_1(t), $x_2(t)$,$x_3(t)$ are the observed signals,
$s_1(t)$, $s_2(t)$, $s_3(t)$ the source signals

y_1(t), $y_2(t)$,$y_3(t)$ are the separated signals

Independent Component Analysis

- Independent Component Analysis (ICA)
  - statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals
- ICA can be seen an extension of Principal Component Analysis (PCA) and Factor Analysis (FA)
- The technique of ICA was firstly introduced in early 1980s in the context of the Neural Networks (NNs) modeling
- ICA is becoming one of the exciting new topics, both in the field of NNs, mainly unsupervised learning, and in advanced statistics and signal processing

Probability distributions and densities

- random variable (rv) or stochastic variable is a variable whose value results from a measurement on some type of random process

- The cumulative distribution function (cdf) F_x of a random variable x at point $x = x_0$ is defined as the probability

$$\boxed{F_x(x_0) = P(x \le x_0)}$$

- For continuous rv the cdf is a nonnegative, nondecreasing continuous function

$$\boxed{0 \le F_x(x_0) \le 1}$$

Probability distributions and densities

- The probability density function (pdf) $p_x(x)$ is obtained as the derivative of its cumulative distribution function

$$\boxed{p_x(x_0) = \left.\frac{dF_x(x)}{dx}\right|_{x=x_0}}$$

- The cdf is computed by using

$$\boxed{F_x(x_0) = \int_{-\infty}^{x_0} p_x(\xi)\, d\xi}$$

Distribution of a random vector

- Assume now that **x** is a n-dimensional random vector of continuous random variables

$$\boxed{\mathbf{x} = \left(x_1, x_2, \ldots, x_n\right)^T}$$

- The cdf is computed by using

$$\boxed{F_{\mathbf{x}}(\mathbf{x_0}) = P(\mathbf{x} \le \mathbf{x_0})}$$

$$\boxed{p_{\mathbf{x}}(\mathbf{x_0}) = \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}\cdots\frac{\partial}{\partial x_n}F_{\mathbf{x}}(\mathbf{x})\bigg|_{\mathbf{x}=\mathbf{x}_0}}$$

Joint and marginal distributions

- The cdf called the joint distribution function of $\mathbf{x}$ and $\mathbf{y}$ is

$$\boxed{F_{\mathbf{x},\mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0) = P(\mathbf{x} \le \mathbf{x}_0, \mathbf{y} \le \mathbf{y}_0)}$$

- The joint density function $p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})$ is defined by differentiating the joint distribution function

- The marginal densities are (e.g. on $\mathbf{x}$)

$$\boxed{p_{\mathbf{x}}(\mathbf{x}) = \int_{-\infty}^{\infty} p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \eta)\, d\eta}$$

Expectation and moments

- Let **g**(**x**) denote any quantity derived from the random vector **x** the expectation of **g**(**x**) is

$$E\{\mathbf{g}(\mathbf{x})\} = \int_{-\infty}^{\infty} \mathbf{g}(\mathbf{x})\, p_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x}$$

- Moments are expectations used to characterize a random vector. The mean vector is

$$\mathbf{m}_{\mathbf{x}} = E\{\mathbf{x}\} = \int_{-\infty}^{\infty} \mathbf{x}\, p_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x}$$

- The n x n correlation matrix is

$$\mathbf{R}_{\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{C}_{\mathbf{x}} + \mathbf{m}_{\mathbf{x}}\mathbf{m}_{\mathbf{x}}^T$$

$$\mathbf{C}_{\mathbf{x}} = E\{(\mathbf{x} - \mathbf{m}_{\mathbf{x}})(\mathbf{x} - \mathbf{m}_{\mathbf{x}})^T\}$$

Covariance matrix

Uncorrelatedness and independence

- Two random vectors **x** and **y** are uncorrelated if their cross-covariance matrix is a zero matrix

$$\boxed{\mathbf{C}_{\mathbf{xy}} = E\left\{(\mathbf{x}-\mathbf{m}_{\mathbf{x}})(\mathbf{y}-\mathbf{m}_{\mathbf{y}})^T\right\} = 0}$$

- The rvs x and y are said independent if and only if

$$\boxed{p_{x,y}(x,y) = p_x(x)\,p_y(y)}$$

- For random vectors is

$$\boxed{p_{\mathbf{x},\mathbf{y},\mathbf{z},\ldots}(\mathbf{x},\mathbf{y},\mathbf{z},\ldots) = p_{\mathbf{x}}(\mathbf{x})\,p_{\mathbf{y}}(\mathbf{y})\,p_{\mathbf{z}}(\mathbf{z})\ldots}$$

- Uncorrelated Gaussian rvs are also independent. This property is not shared by other distributions in general

Higher-order statistics

- Consider a scalar rv x, the *j*-th moment is defined as (j=1,2,…)

$$\boxed{\alpha_j = E\{x^j\} = \int_{-\infty}^{\infty} \xi^j p_x(\xi) d\xi}$$

- The j-th central moment

$$\boxed{\mu_j = E\{(x-\alpha_1)^j\} = \int_{-\infty}^{\infty} (\xi - m_x)^j p_x(\xi) d\xi}$$

Skewness

- The third central moment is called the skewness (asymmetricity of the pdf)

$$\boxed{\mu_3 = E\left\{\left(x - m_x\right)^3\right\}}$$

- The 4-th moment and central moment are applied in ICA

Kurtosis

- Usually the 4-order statistic (i.e. cumulants) is employed and it is called Kurtosis

$$\boxed{\mathbf{kurt}(x) = E\left\{x^4\right\} - 3\left[E\left\{x^2\right\}\right]^2}$$

- A distribution having kurtosis
  - Zero is called mesocurtic
  - Negative platykurtic (subgaussian)
  - Positive leptokurtic (supergaussian)

Differential entropy

- The differential entropy of a rv is defined as

$$H(x) = -\int p_x(\xi) \log p_x(\xi) d\xi = -E\{\log p_x(x)\}$$

- Can be interpreted as a measure of randomness. If the rv is concentrated on certain small intervals, its differential entropy is small

Mutual Information

- Mutual information is a measure of the information that members of a set of random variables have on other random variables in the set

$$\boxed{I(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} H(x_i) - H(\mathbf{x})}$$

where **x** is the vector containing all the x_i

- If x_i are independent they give no information on each other

Kullback-Leibler divergence

- Mutual information can be considered a distance using the Kullback-Leibler divergence

$$\delta(p^1, p^2) = \int p^1(\xi) \log \frac{p^1(\xi)}{p^2(\xi)} d\xi$$

- Can be considered as a distance between pdfs
  - Is always nonnegative
  - Is zero if and only if the two distributions are equal
  - Can be symmetrized

Negentropy

- The Negentropy is a measure that is zero for a Gaussian variable and always nonnegative

$$\boxed{J(\mathbf{x}) = H(\mathbf{x}_{Gauss}) - H(\mathbf{x})}$$

- A simple approximation is (standardized rv)

$$\boxed{J(x) \approx \frac{1}{12} E\left\{x^3\right\}^2 + \frac{1}{48} \mathbf{kurt}(x)^2}$$

Negentropy

- A more robust approximation is

$$J(x) \approx k_1\left(E\left\{G^1(x)\right\}\right)^2 + k_2\left(E\left\{G^2(x)\right\} - E\left\{G^2(\nu)\right\}\right)^2$$

where k_1 and k_2 are positive constants, G^1 and G^2 are odd and even function, respectively (e.g. $G^1(x) = x^3$ and $G^2(x) = x^4$)

Newton's method

- Newton's method is one of the most efficient ways for function minimization $F(\mathbf{w})$

- The updating rule is (by using the gradient and the *Hessian*)

$$\boxed{\Delta \mathbf{w} = -\left[\frac{\partial^2 F(\mathbf{w})}{\partial \mathbf{w}^2}\right]^{-1} \frac{\partial F(\mathbf{w})}{\partial \mathbf{w}}}$$

- The convergence of the Newton's method is quadratic

The Lagrange method

- In many cases we have constrained optimizations

$$\boxed{\begin{array}{l}\min F(\mathbf{w}) \\ \text{subject to } H_i(\mathbf{w}) = 0, \quad i = 1,\dots,k\end{array}}$$

- The most used way to take the constraints into account is the method of Lagrange multipliers $(\lambda_1, \dots, \lambda_k)$

- We form the Lagrange function

$$\boxed{L(\mathbf{w}, \lambda_1, \dots \lambda_k) = F(\mathbf{w}) + \sum_{i=1}^{k} \lambda_i H_i(\mathbf{w})} \Rightarrow \boxed{\frac{\partial L(\mathbf{w}, \lambda_1, \dots \lambda_k)}{\partial \mathbf{w}} = 0}$$

ICA

- We want to maximize the negentropy using this approximation

$$\boxed{J_G(\mathbf{w}) = \left[E\{G(\mathbf{w}^T\mathbf{x})\} - E\{G(\nu)\}\right]^2}$$

- The multi-unit problem is

$$\boxed{\begin{array}{l} \max \sum_{i=1}^{n} J_G(\mathbf{w}_i) \quad \mathbf{w}_i, i = 1,..,n \\ \text{such that } E\{(\mathbf{w}_k^T\mathbf{x})(\mathbf{w}_j^T\mathbf{x})\} = \delta_{jk} \end{array}}$$

- A fixed point algorithm is obtained by applying the Newton's method to the Lagrangian of this optimization problem (FastICA)

ICA

- In many cases we have constrained optimizations

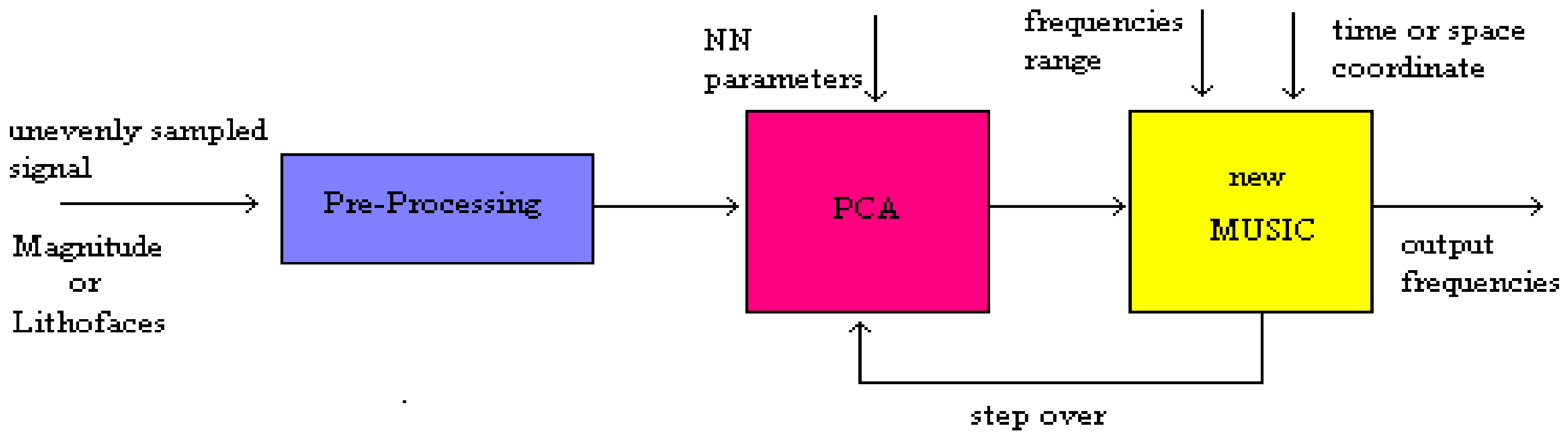
$$\boxed{\begin{aligned} G(y) &= \frac{1}{a_1}\log\cosh a_1 y \\ G(y) &= -\exp(-y^2/2) \\ G(y) &= y^4 \end{aligned}}$$

$$\boxed{\begin{aligned} g(y) &= \tanh(a_1 y) \\ g(y) &= y\exp(-y^2/2) \\ g(y) &= y^3 \end{aligned}}$$

Frequency estimator

- Approach
  - Reformulation of the MUSIC (MUlti SIgnal Classificator) frequency estimator for unevenly sampled data
  - Rosbust PCA Neural Network to extract signal information
- Periodicities estimation
  - Light curves
    - W UMa-system (e.g. U Pegasi), Cepheid (e.g. Su Cygni), blazars, ecc.
  - Stratigraphic sequences
    - Tobenna, Raggeto, San Lorenzello montains

STIMA approach

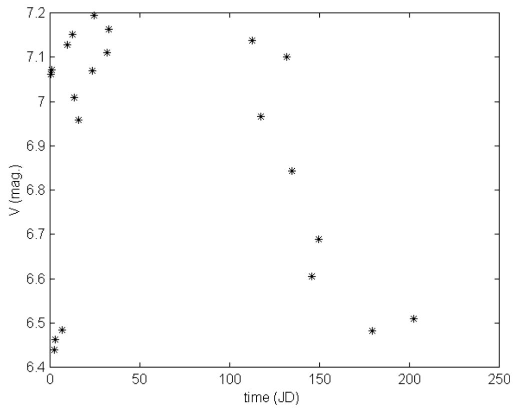


$$P_M = \frac{1}{M - \sum_{i=1}^{M} |\mathbf{e}_f^H \mathbf{w}(i)|^2}$$

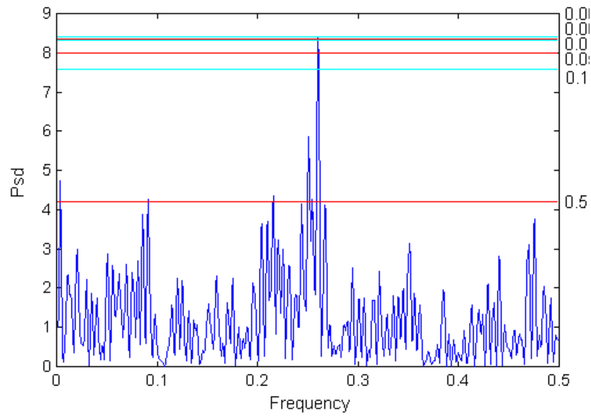
MUSIC estimator for unevenly sampled data

$$\mathbf{e}_f^H = \left[e_f^{t_0}, e_f^{t_1}, \ldots, e_f^{t_{L-1}} \right]^H$$

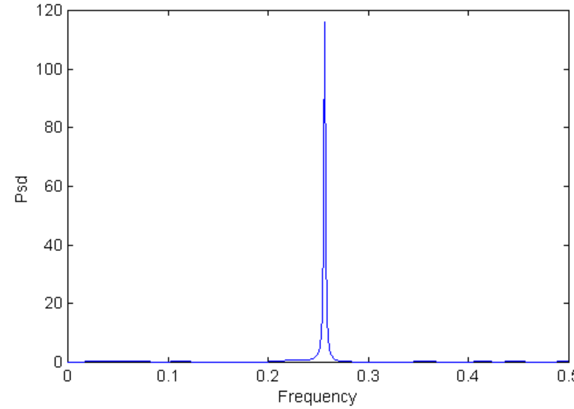
Light curve SU Cygni



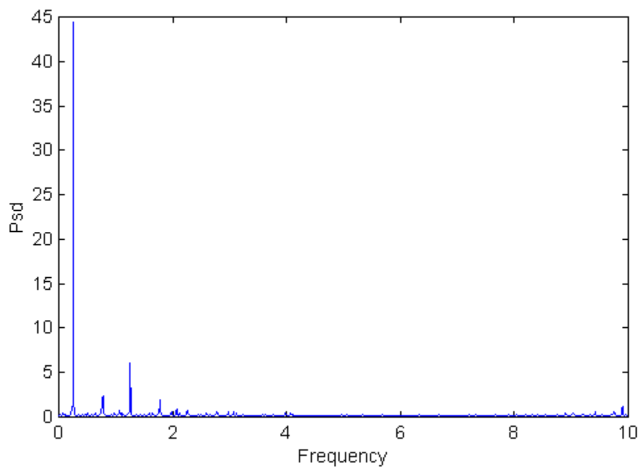
Light curve



Lomb's Periodogram

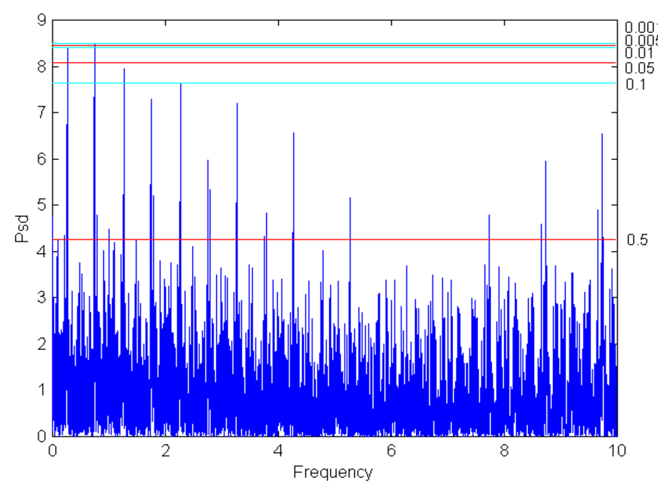


STIMA approach



STIMA approach

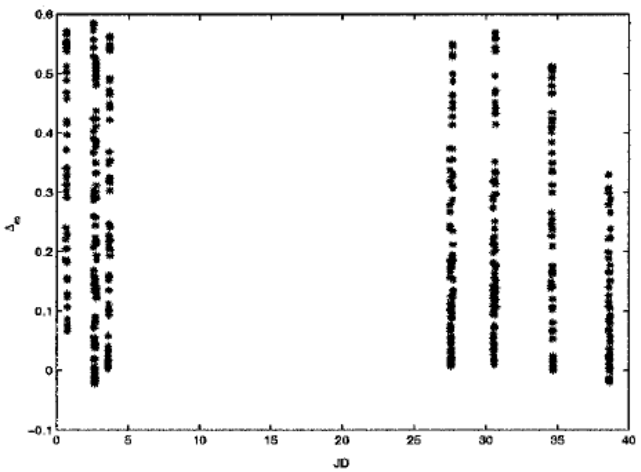
Frequency variation



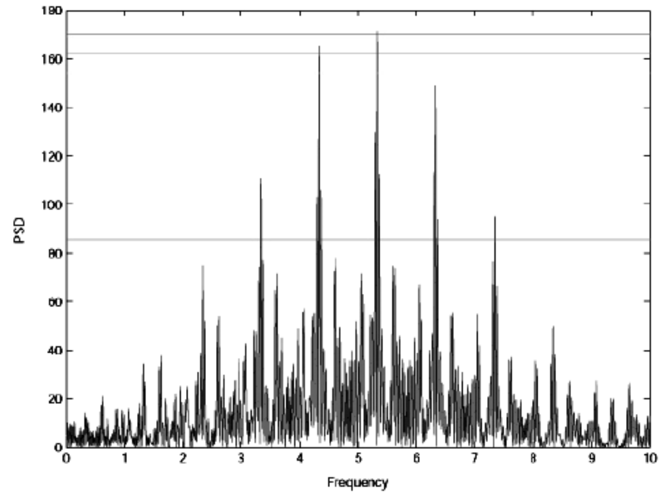
Scargle's Periodogram

Collaboration with the Department of Physics (Unversity of Naples "Federico II") and Astronomic Observatory of Capodimonte

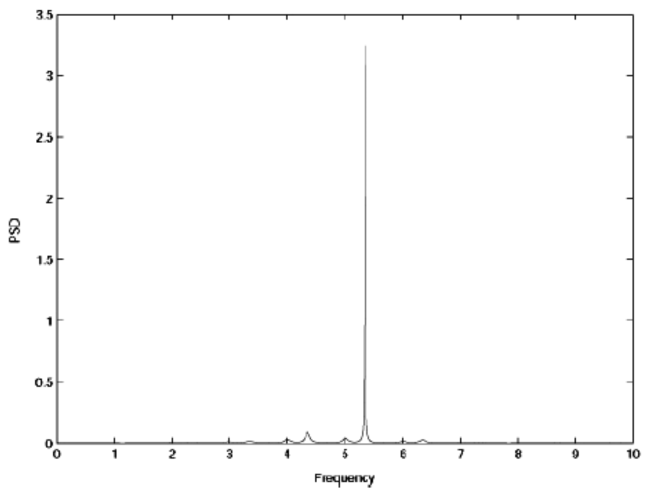
Light curve U Pegasi (I data set)



Light curve

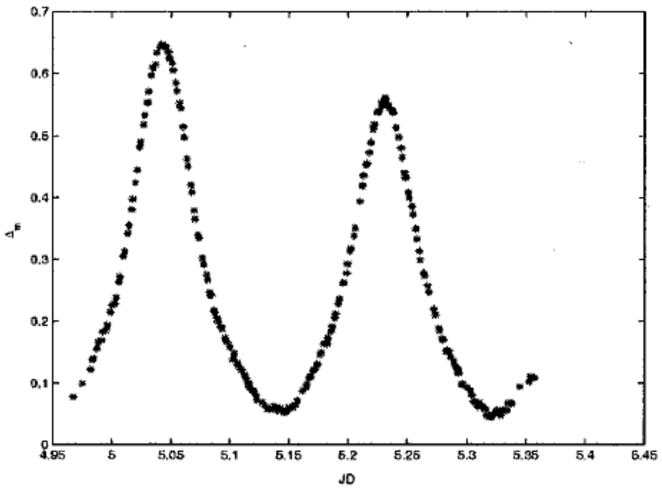


Lomb's Periodogram

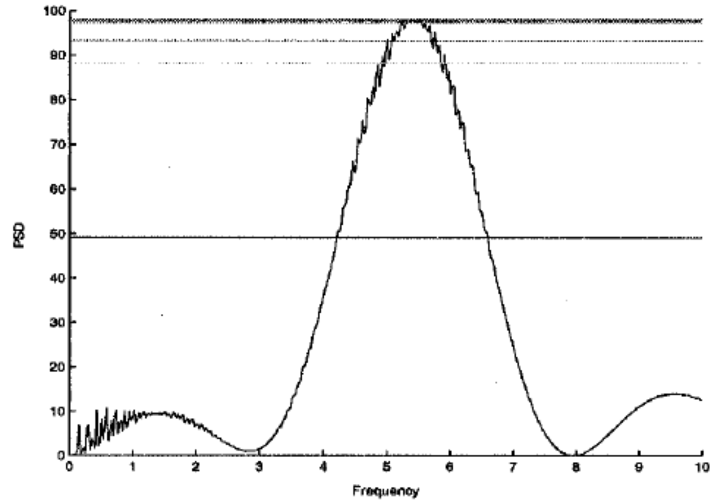


STIMA approach

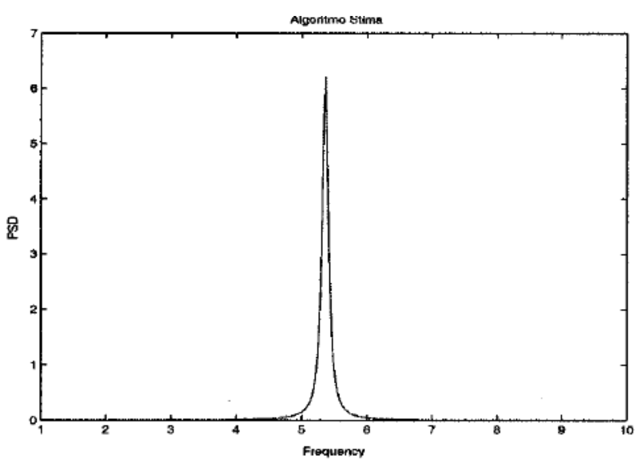
Light curve U Pegasi (II data set)



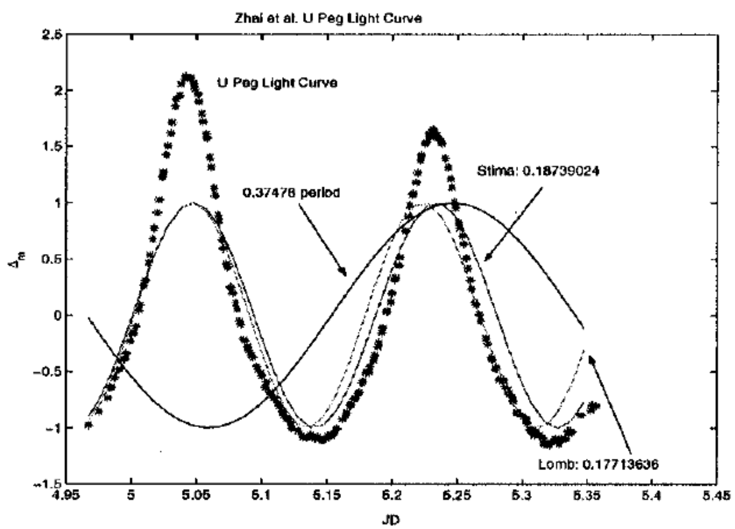
Light curve



Lomb's Periodogram

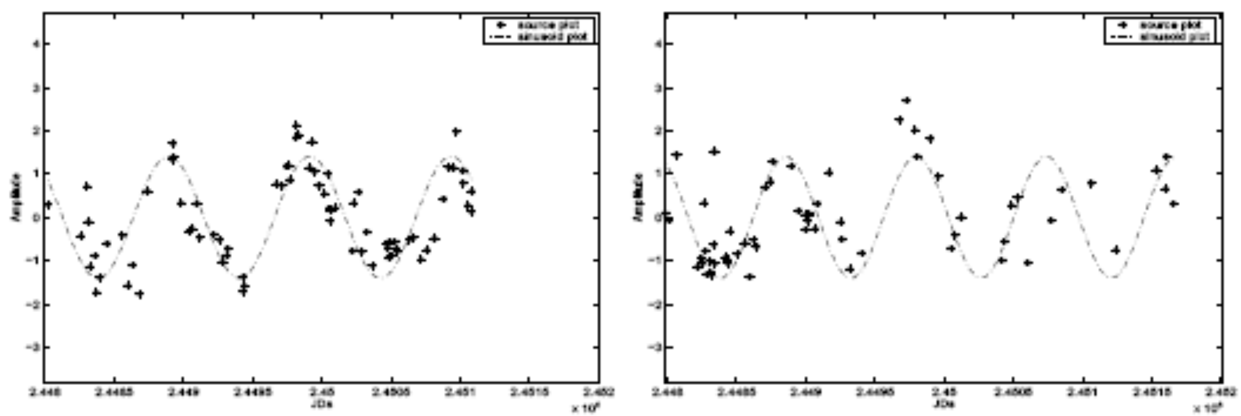


STIMA approach

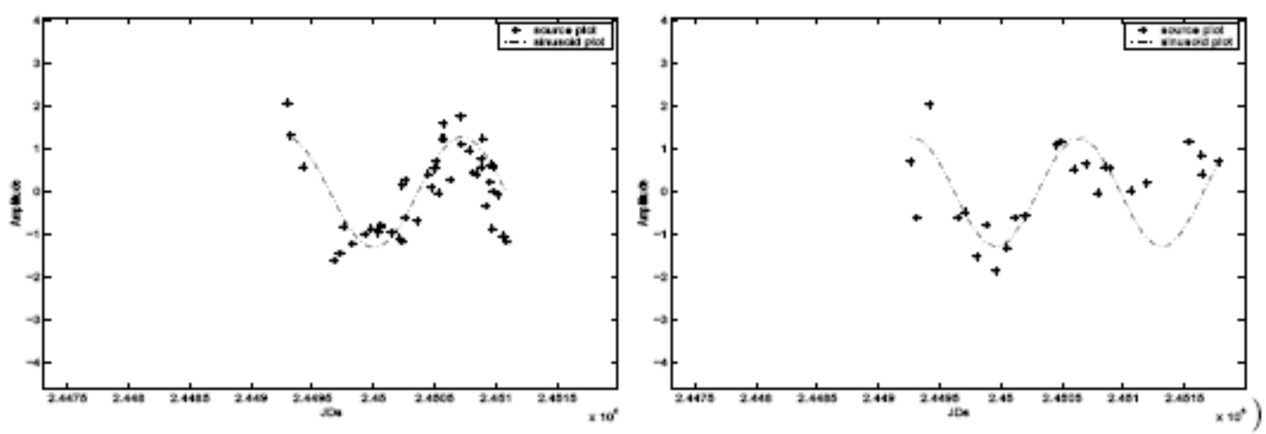


Comparison

Radio variability of blazars



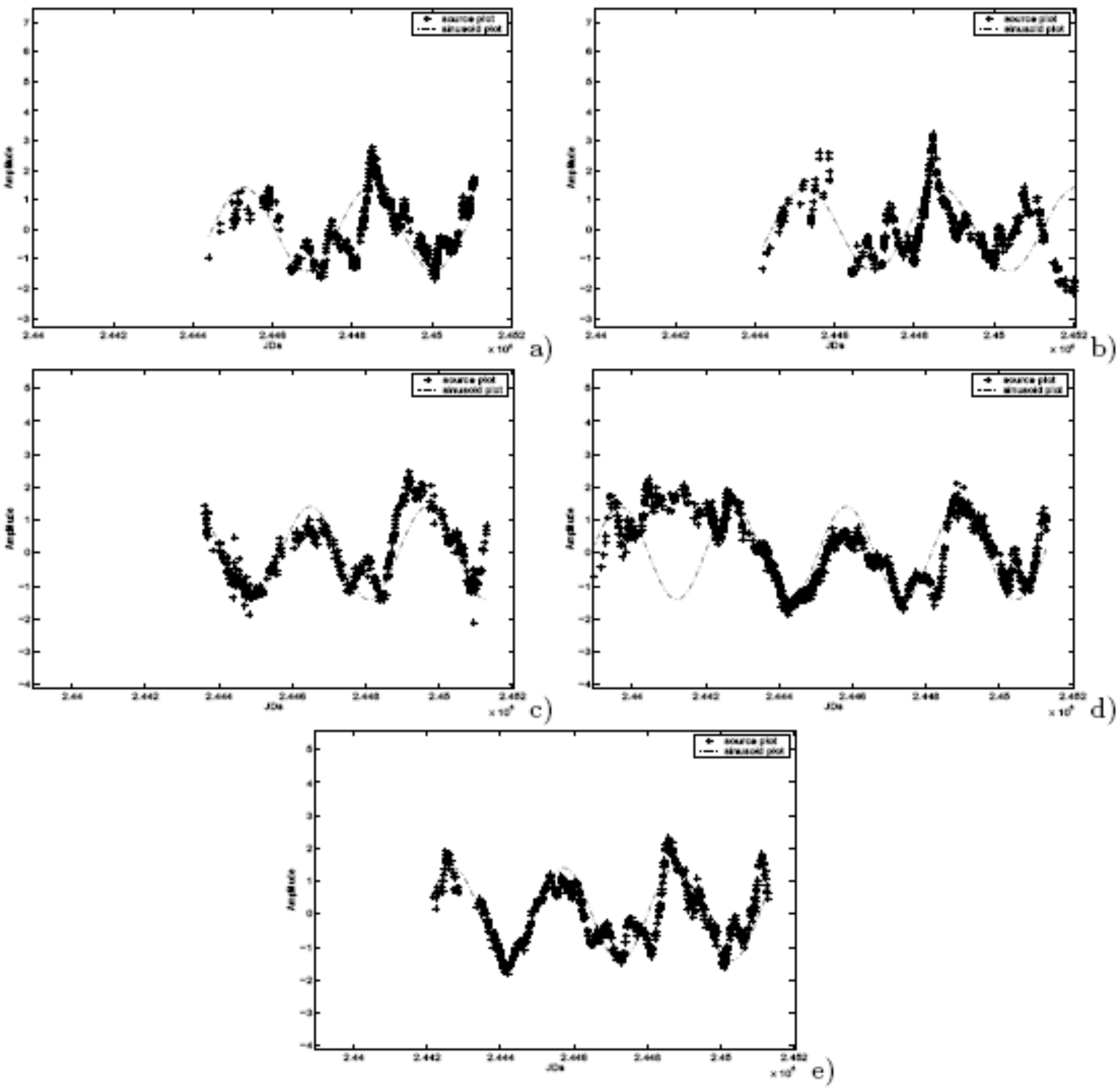
A0224+671. Results for the 22 GHz (left panel) and 37 (right panel), daily averaged datasets. Sinusoids with a period equal to those provided by STIMA



0945 + 408. Results for the 22 (left) and 37 GHz (right)

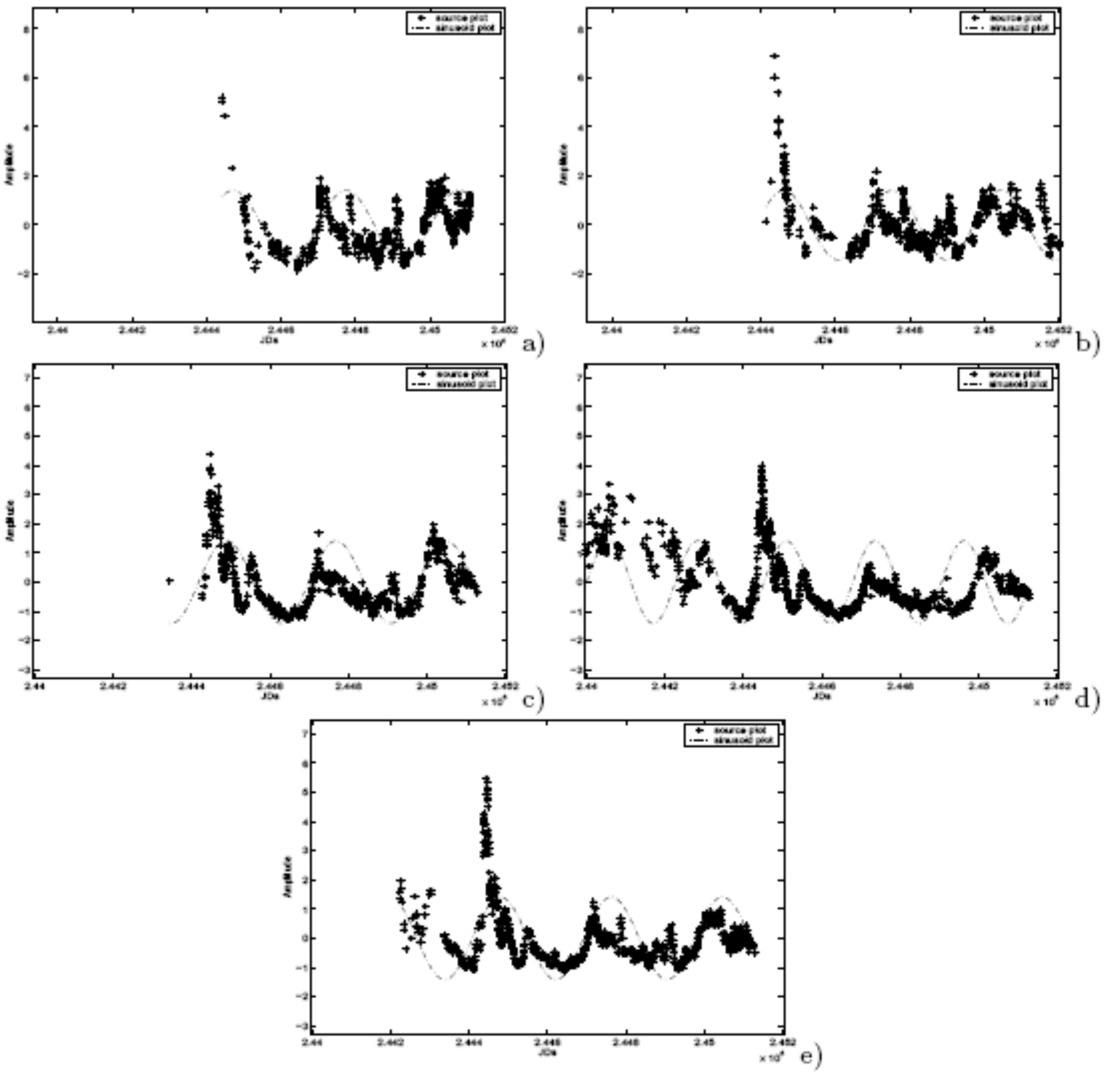
Collaboration with the Department of Physics of Unversity of Naples “Federico II”, Astronomic Observatory of Padova, Dept. Of Astronomy of the University of Michigan, Metsahovi Radio Observatory, Kylmala, Finland

Radio variability of blazars



1226+023. Panels a and b: results for the 22 GHz and the 37 GHz daily averaged radio curves, respectively. Panels c, d and e: the same for the 4.8, 8 and 14.5 GHz data, respectively

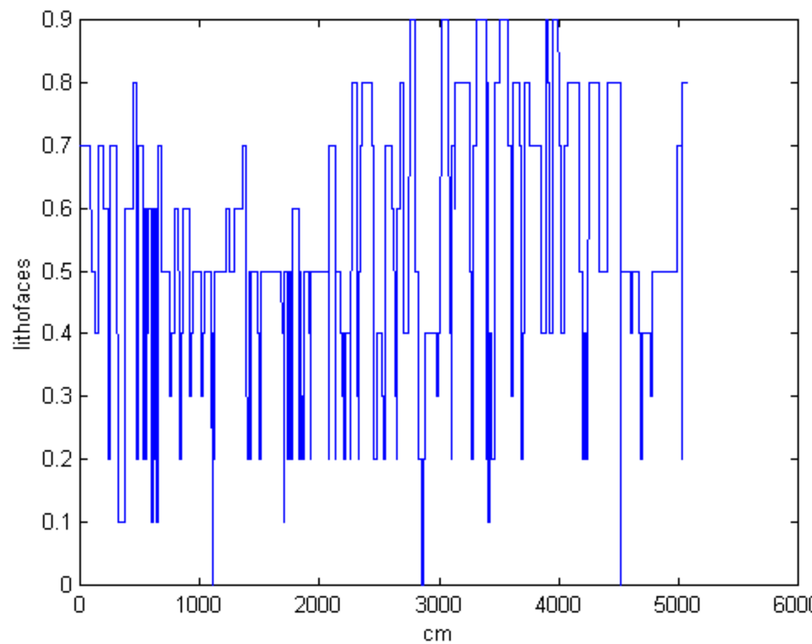
Radio variability of blazars



2200 + 420. Panels have the same meaning as in the previous figure

Stratigraphic sequences

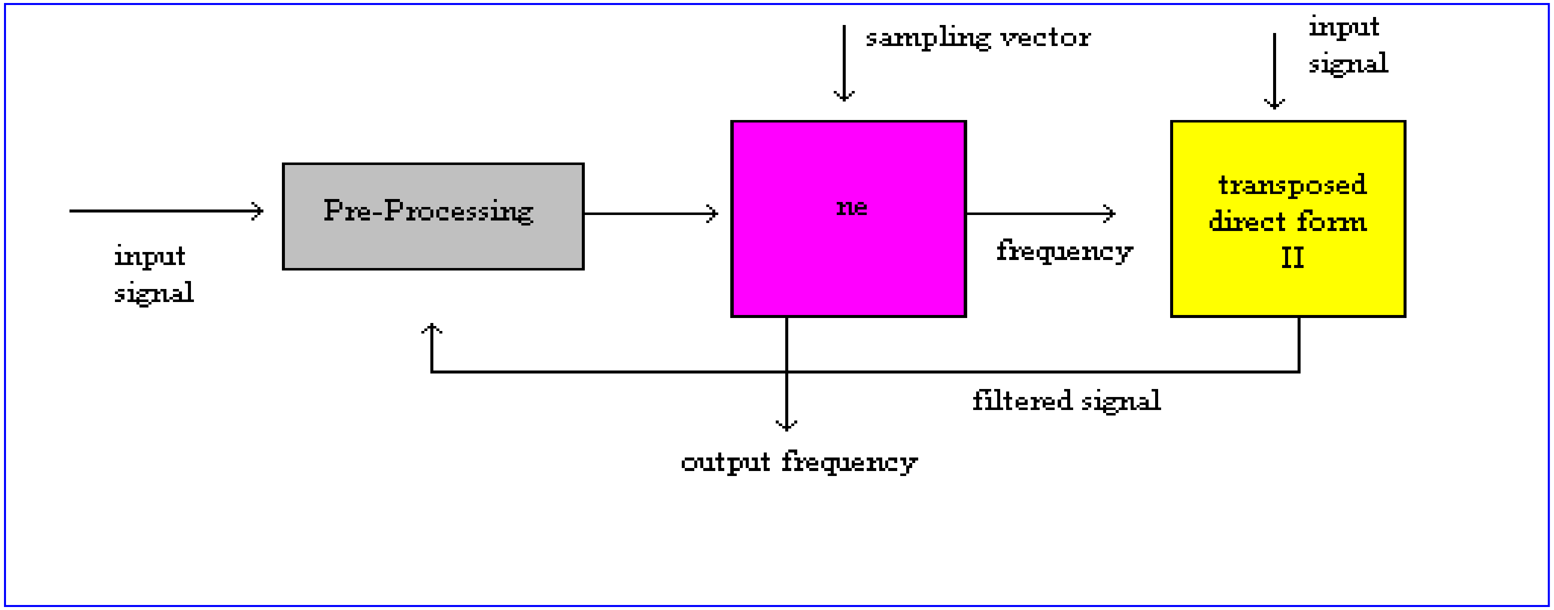
An example of
stratigraphy in a mountain



Mount San Lorenzello stratigraphic sequence

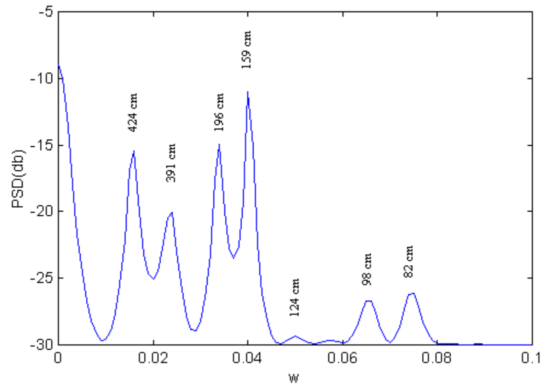
Collaboration with the Istituto Geomare (NA)

Estimation process

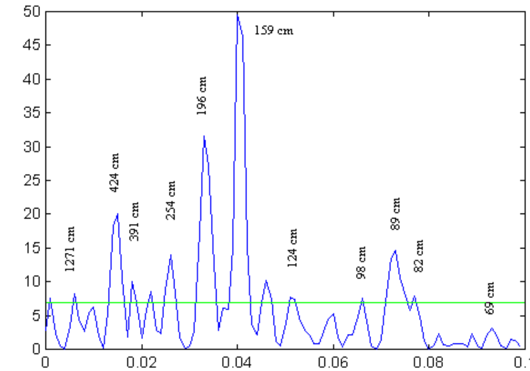


The method is the most powerful tool because it accurately finds the real spectral features, comprehensive of frequencies which the others methods do not find or confuse with the noise. They also require low computing time.

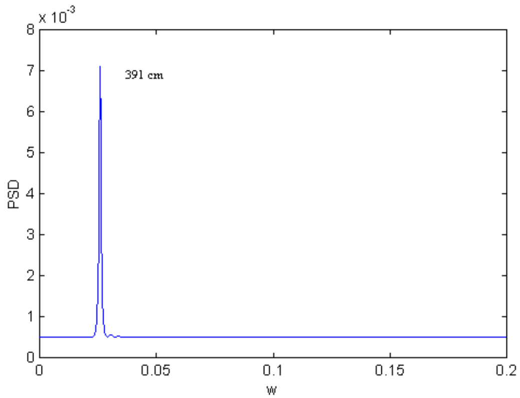
Periodicities estimation

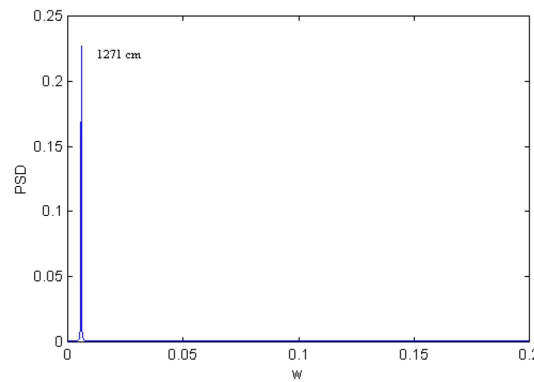


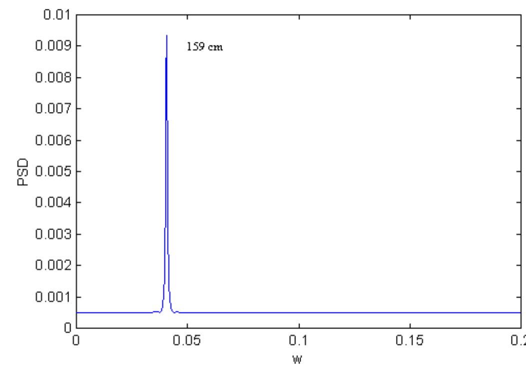
STIMA approach

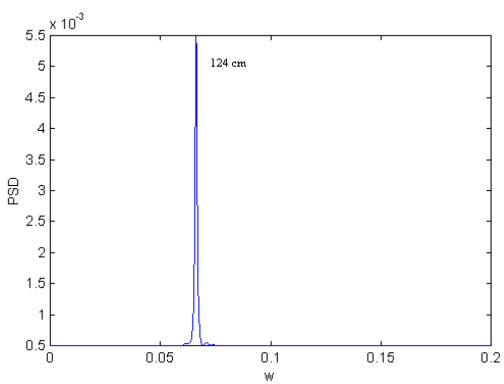


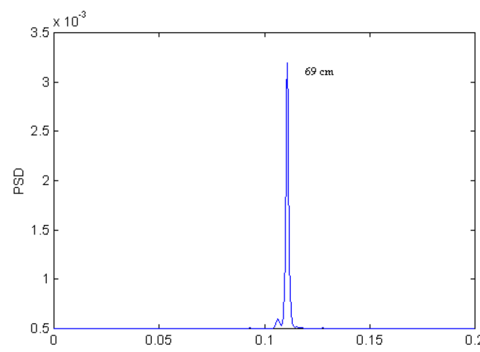
Scargle's Periodogram estimation





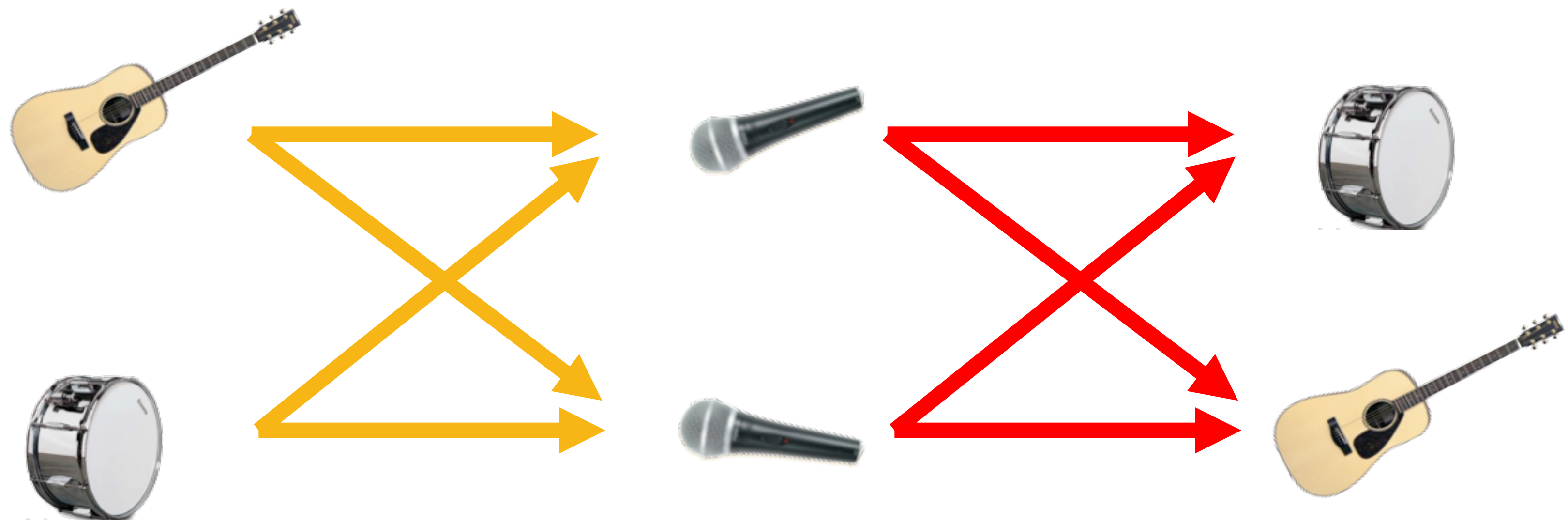






STIMA and filtering

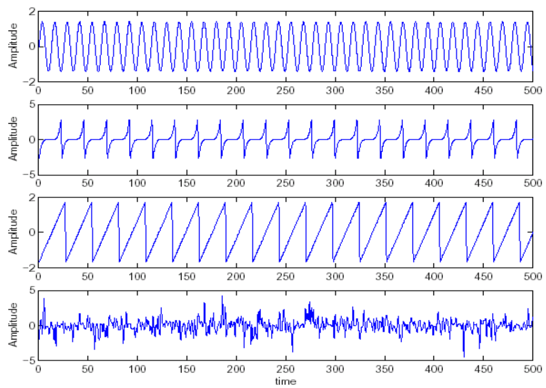
Cocktail party



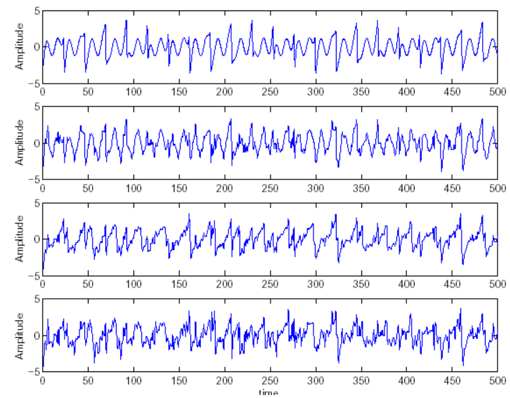
Sources | Mixtures | Estimated-Sources

$\mathbf{s}$ $\mathbf{A}$ $\mathbf{x}$ $\mathbf{W}$ $\mathbf{y}$

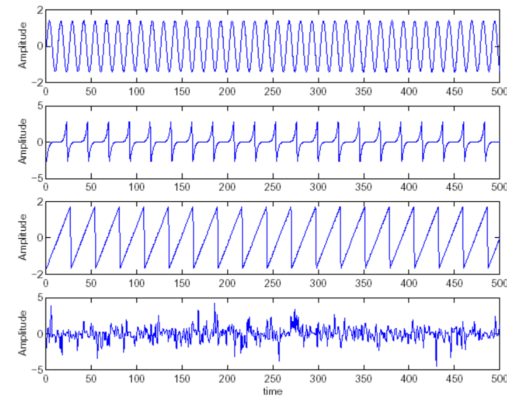
Source estimation



Source signals



Mixed signals



Estimated signals

$$\begin{aligned}
x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t)\\
x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t)\\
x_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t)
\end{aligned}$$

$$\begin{aligned}
y_1(t) &= w_{11}x_1(t) + w_{12}x_2(t) + w_{13}x_3(t)\\
y_2(t) &= w_{21}x_1(t) + w_{22}x_2(t) + w_{23}x_3(t)\\
y_3(t) &= w_{31}x_1(t) + w_{32}x_2(t) + w_{33}x_3(t)
\end{aligned}$$

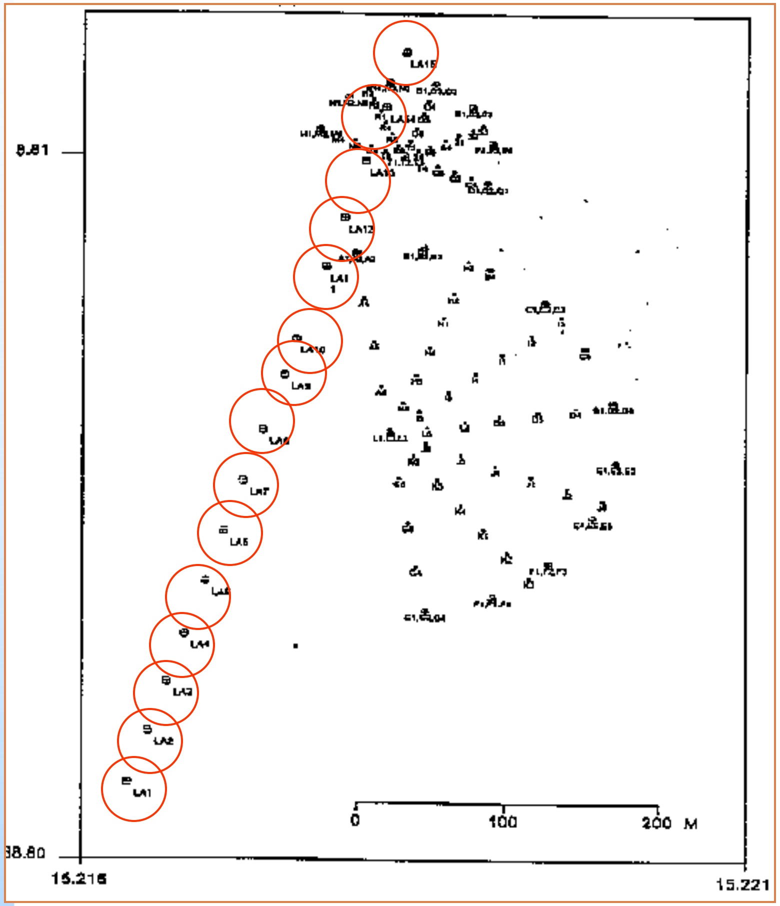
x_1(t), $x_2(t)$,$x_3(t)$ are the observed signals,
$s_1(t)$, $s_2(t)$, $s_3(t)$ the source signals

y_1(t), $y_2(t)$,$y_3(t)$ are the separated signals

ICA and seismic signals

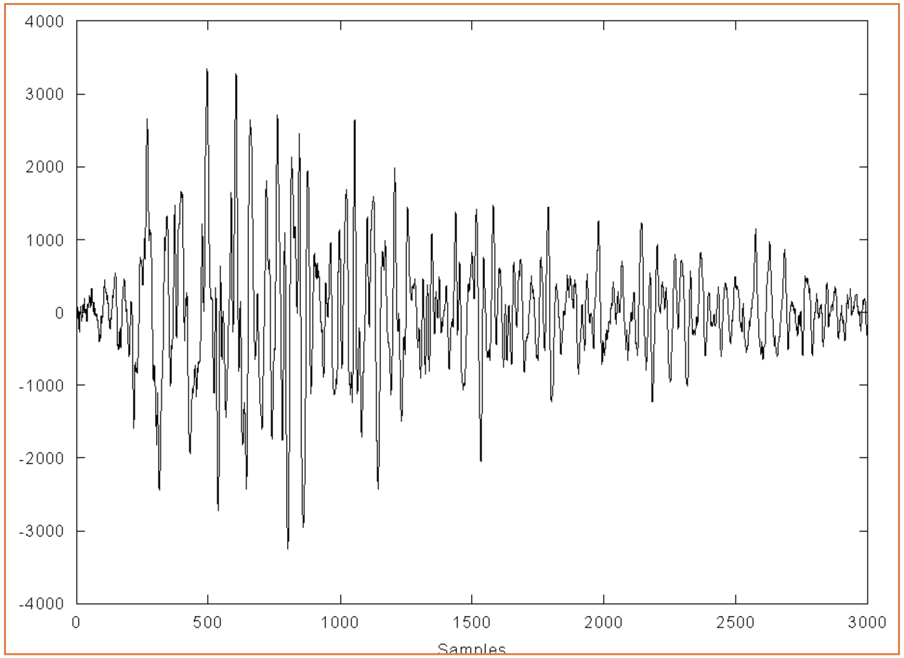
- ICA is used to analyze the seismic signals produced by explosions of the Stromboli volcano
- It has been experimentally proved that it is possible to extract the *most significant components* from seismometer recorders
- In particular, the signal, eventually thought as generated by the source, is corresponding to the higher power spectrum, isolated by our analysis
- Furthermore, the *amplitude of the source signals has been found* by using a simple trick and so overcoming, for this specific case, the classical problem of ICA regarding the amplitude loss of the separated signals

Seismic signals



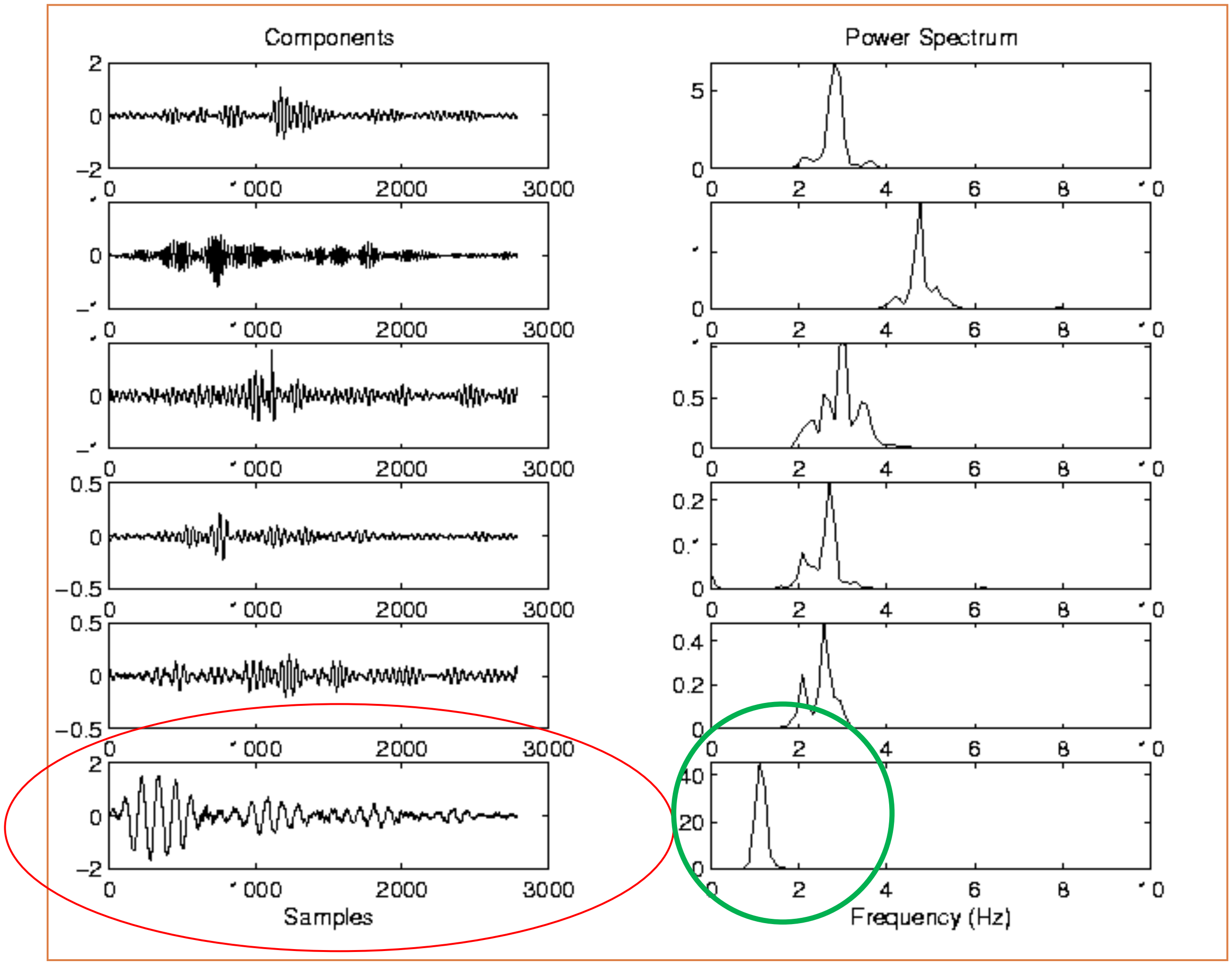
Array position

An example of explosion-quake



Collaboration with the Department of Physics of the Unversity Salerno and Gran Sasso Observatory

Source estimation



Radial separated signals

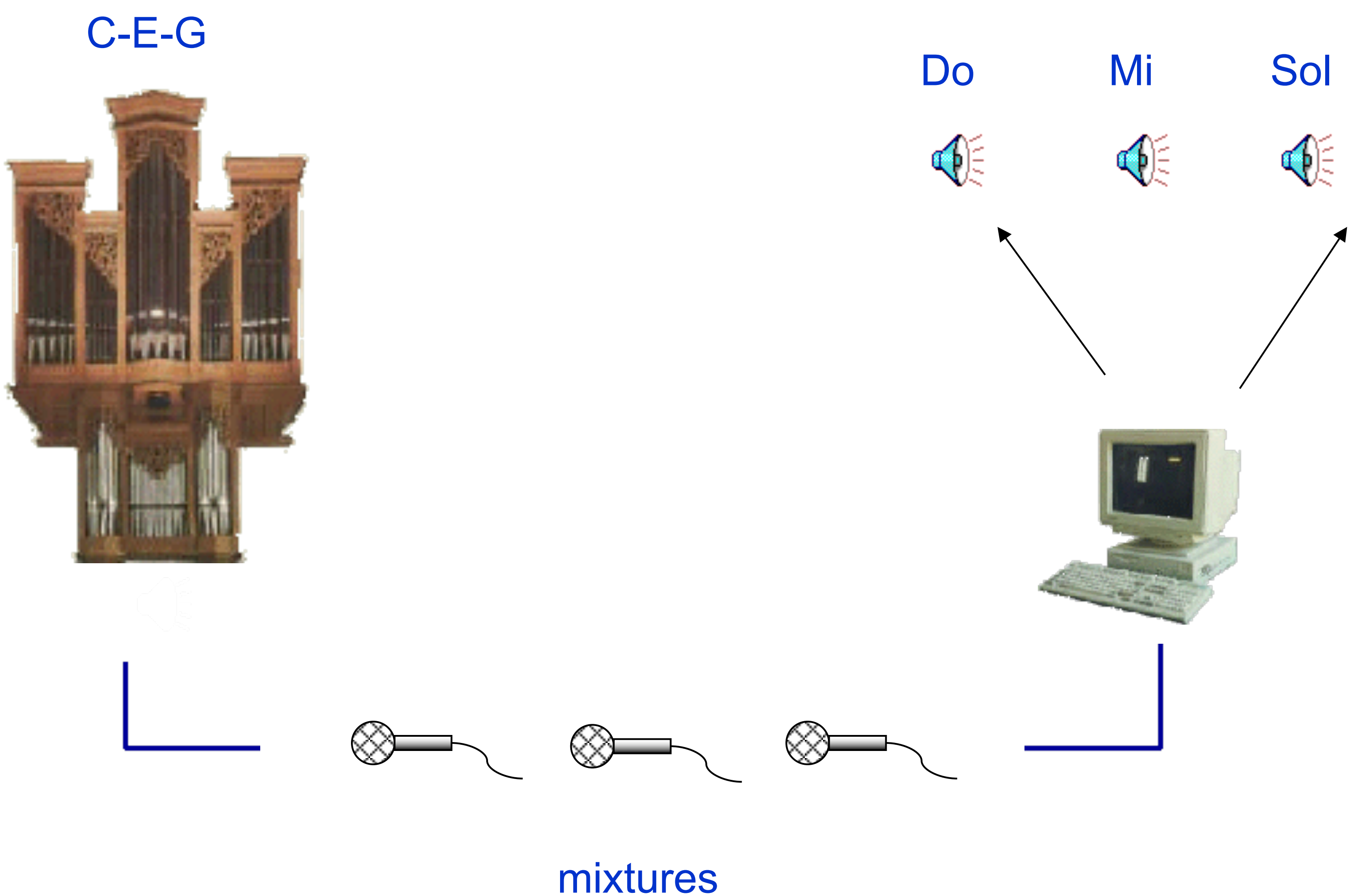
ICA and pipe organ sound

- The aim of this work is to analyze and to resynthesize acoustic signals emitted by organ pipes recorded in real environments

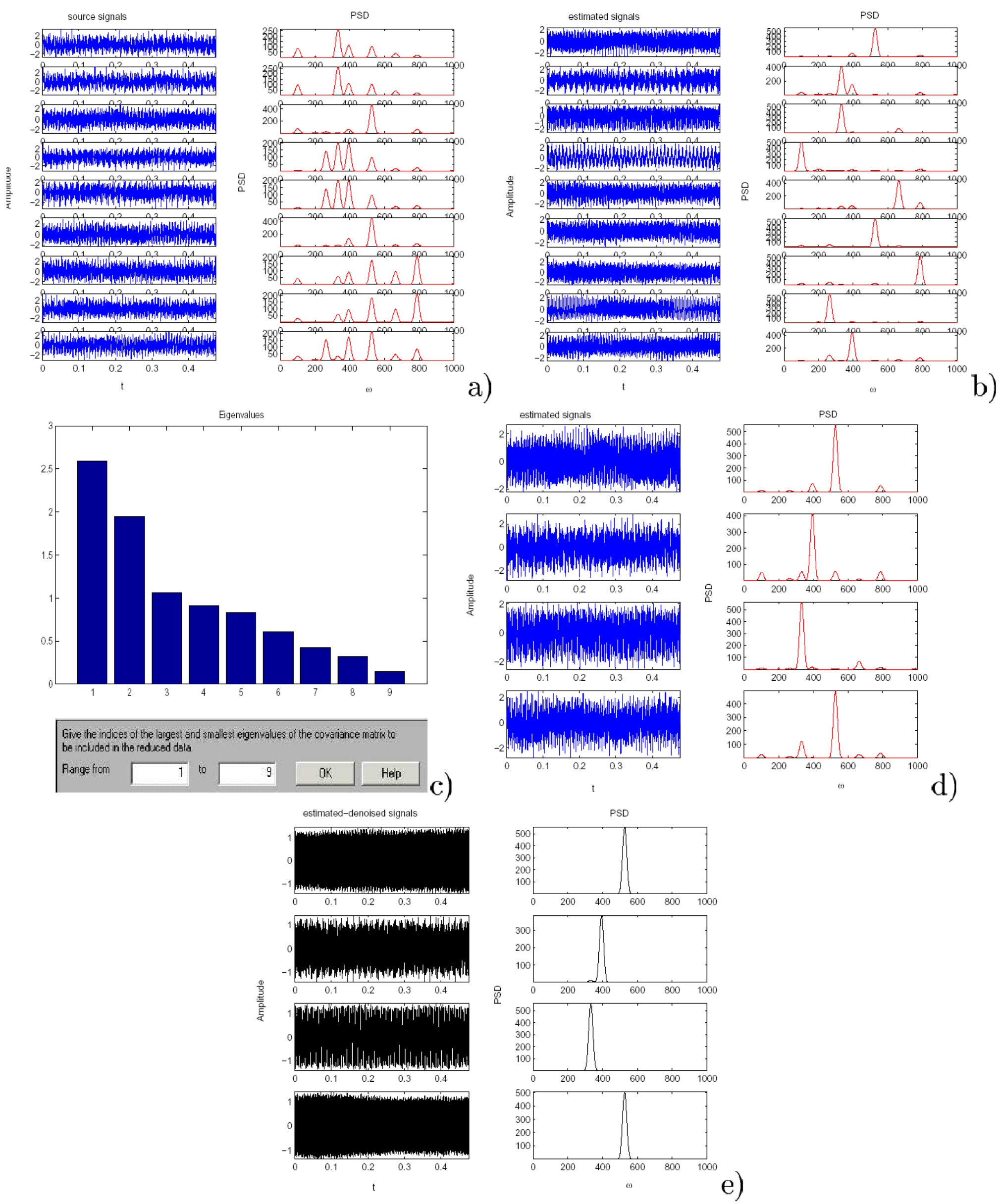
- By applying ICA to the recordings, we have established that a single note is itself composed of three self-oscillating signals (Andronov oscillator) with well defined frequencies
  - the pipe acoustic signals can be described by a mixture of non-linear oscillations obtained by a self-sustained feedback system
  - Considering this non-linear system, an additive synthesis model is proposed

- Suitable analogical and integrate circuit models, able to reproduce the registered waveform and sound in listening, have been designed

Pipe organ

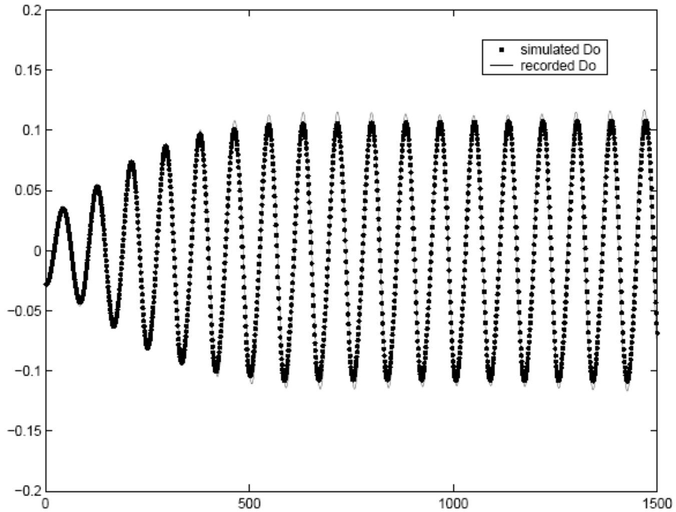


Source estimation

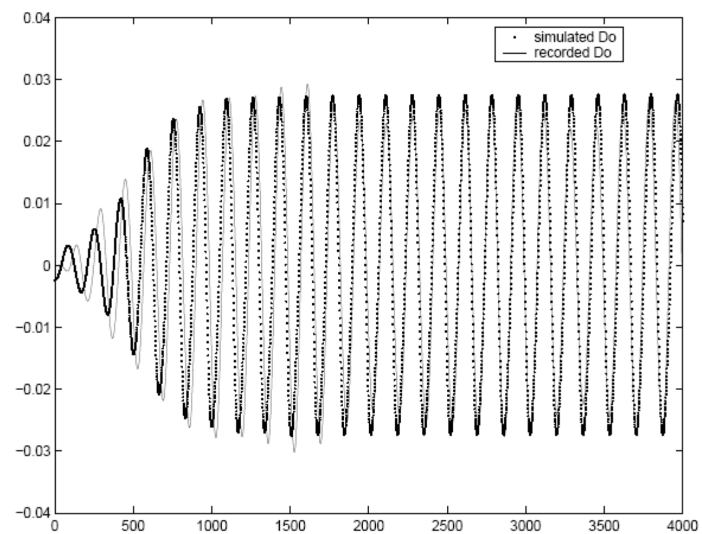


Do note: a) C-E-G chord signals and their PSD;
b) FastICA separation;
c) Principal components of the covariance matrix;
d) FastICA separation with four components
e) Estimated signals after denoising

Simulated notes

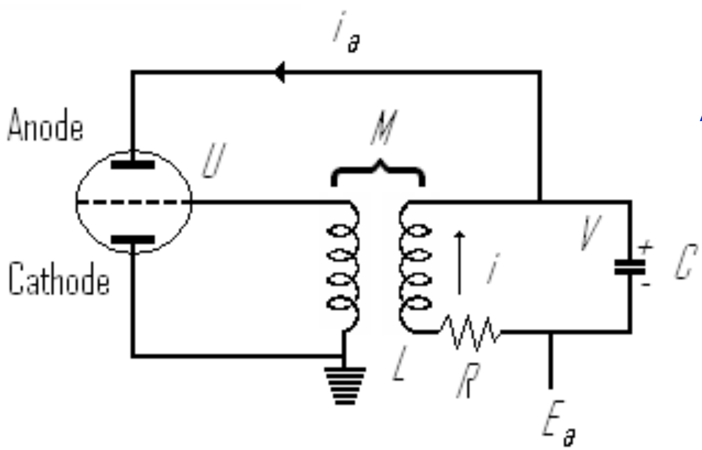


Comparison between simulated and real C note (523 Hz)

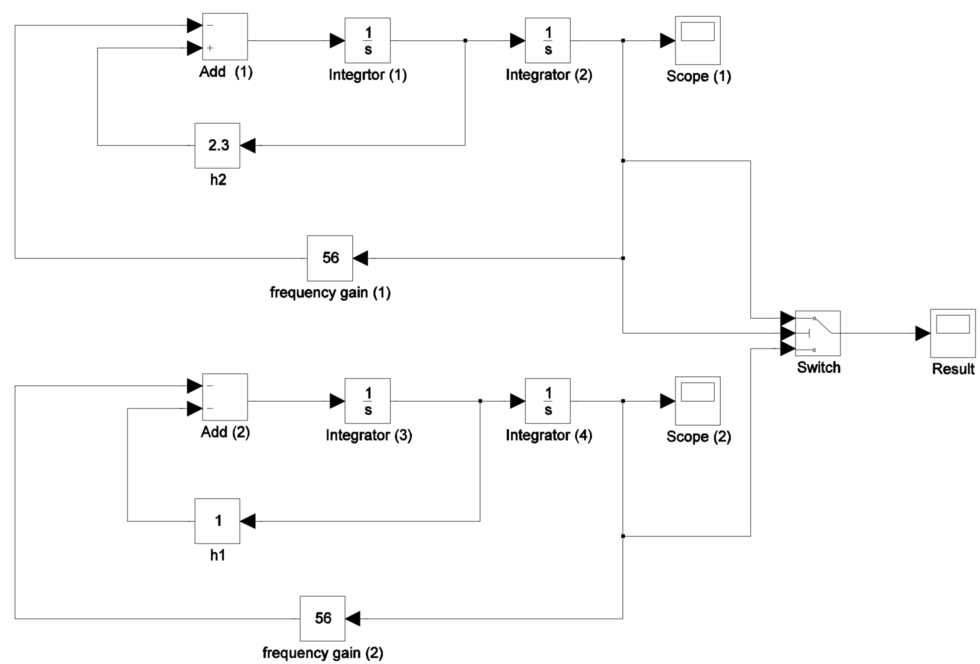


Comparison between simulated and real C note (263 Hz)

Analogi circuit



Andronov analogical circuit



Andronov system by using Matlab Simulink

ICA and real environment

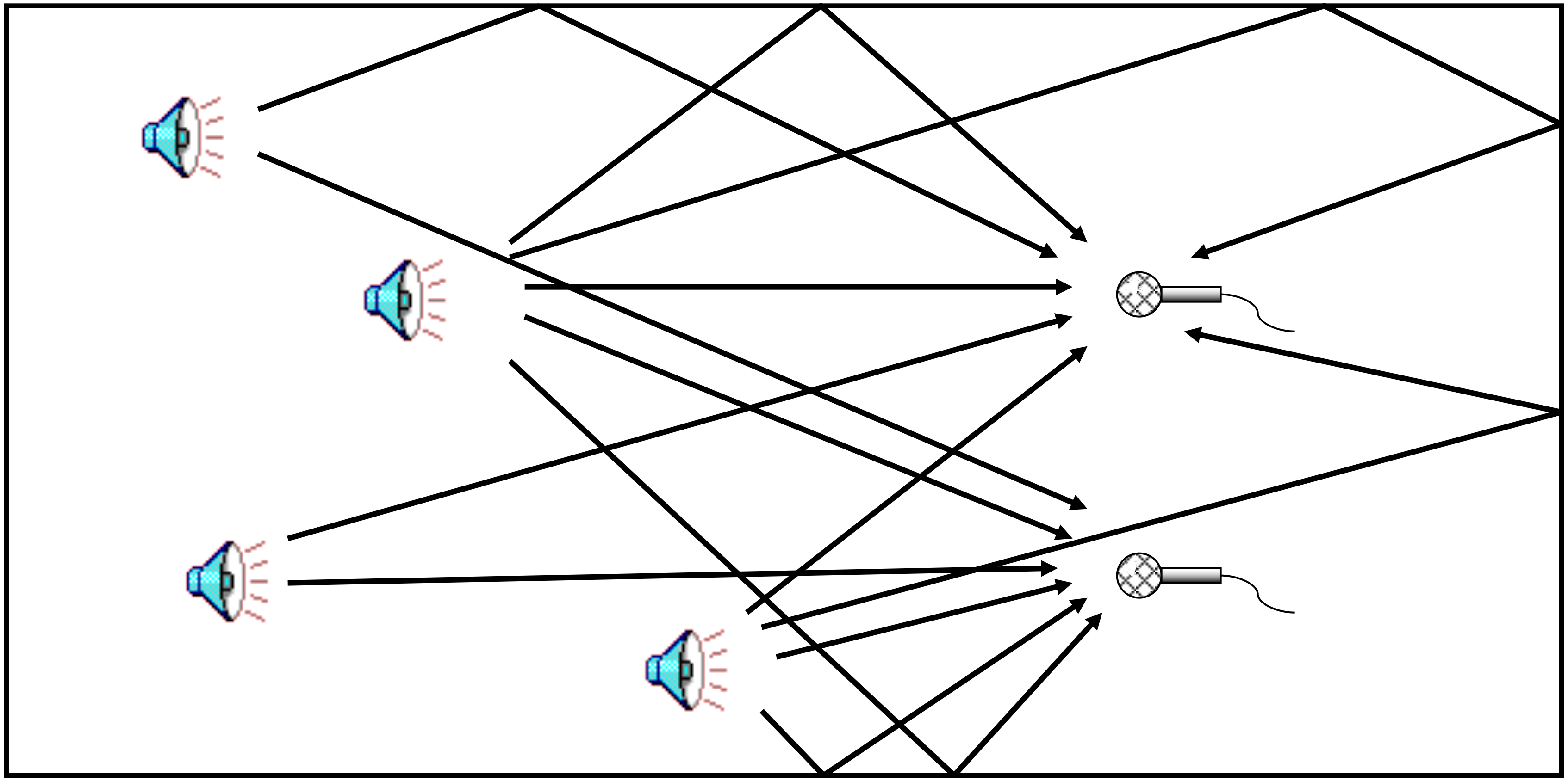
- Now we consider a novel approach to solve the permutation indeterminacy in the separation of convolved mixtures in frequency domain

- These are obtained applying a Short Time Fourier Transform on a set of fixed frames

- To solve the ambiguity of the amplitude dilation, a simple disassemble method is proposed

- The permutation indeterminacy is solved using an approach based on the *Hungarian algorithm* that solves an Assignment Problem and an algorithm of Dynamic Programming
  - To obtain the distances in the Assignment Problem, a Kullback-Leibler divergence is adopted

Real environment



In real cases we have to consider the reverberation and delay

Convolved mixtures

$$x_i(t) = \sum_{j=1}^{n} \sum_{k} a_{ikj} s_j(t-k) \qquad \text{for } i = 1, \ldots, n$$

Convolved mixtures
(FIR filter)

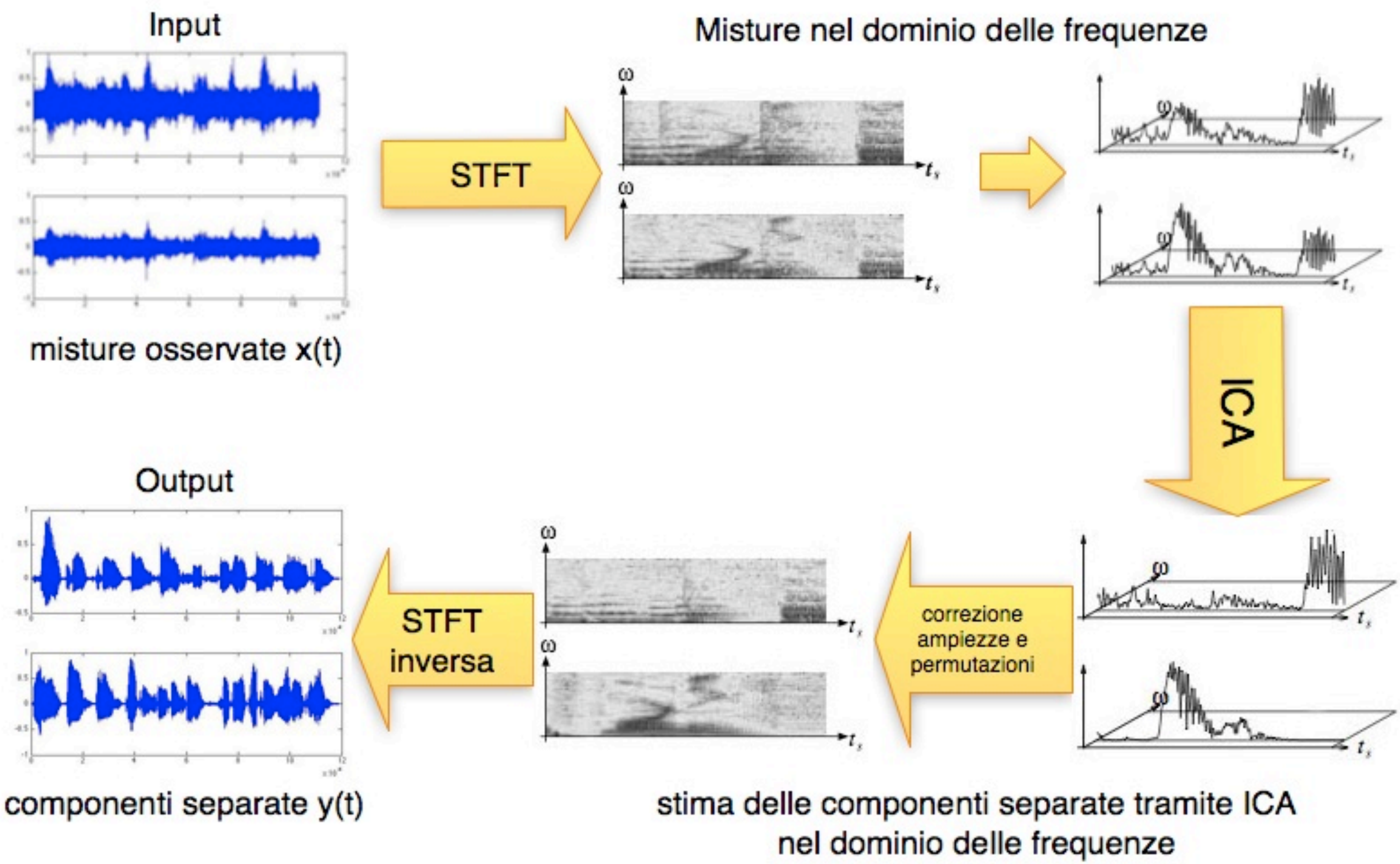
$$y_i(t) = \sum_{j=1}^{n} \sum_{k} w_{ikj} x_j(t-k) \qquad \text{for } i = 1, \ldots, n$$

Estimate sources
(FIR filter)

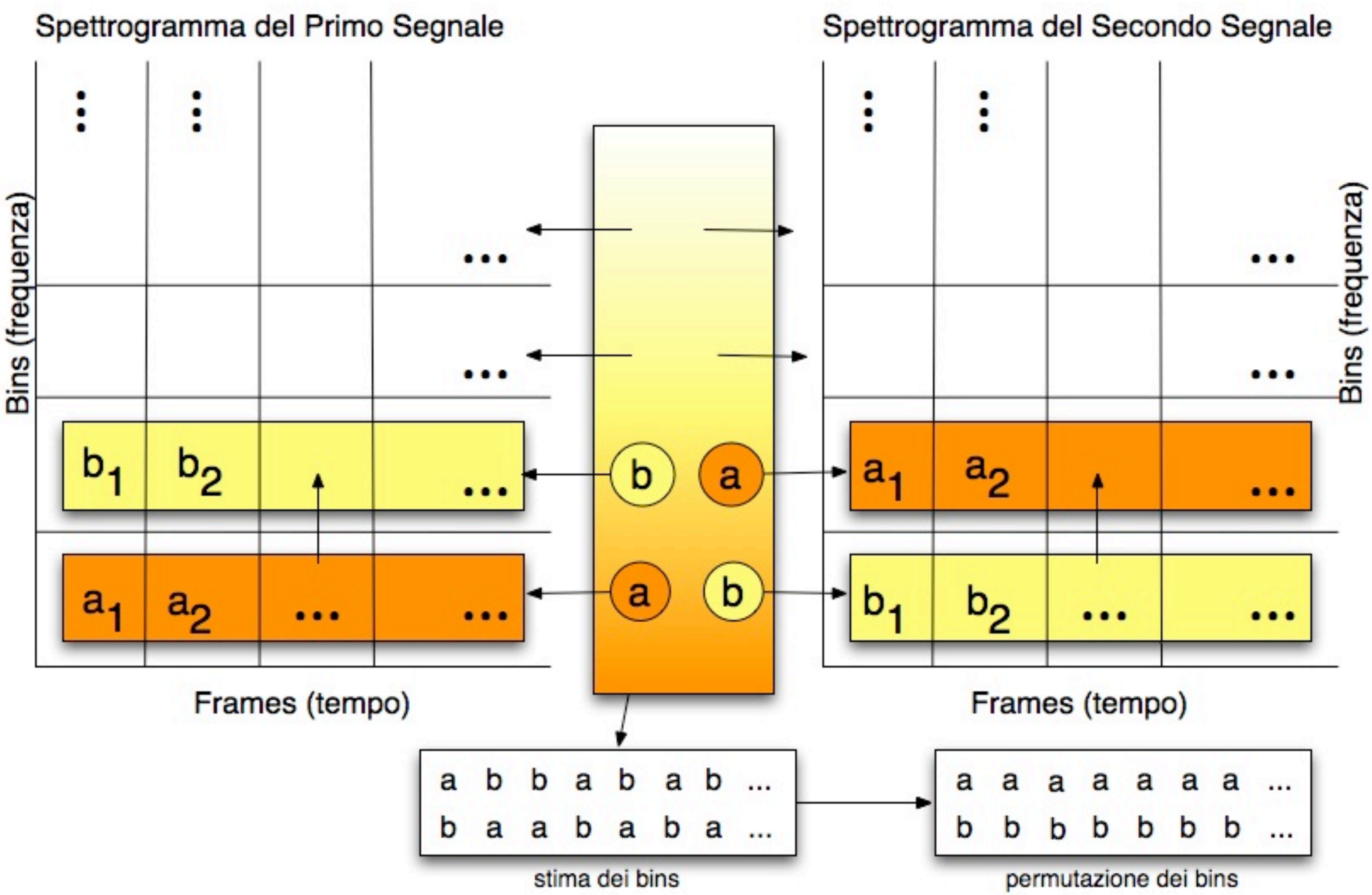
$$X_i(\omega) = \sum_{j=1}^{m} A_{ij}(\omega) S_j(\omega), \quad \text{for } i = 1, ..., n$$

Product in frequency domain (STFT)

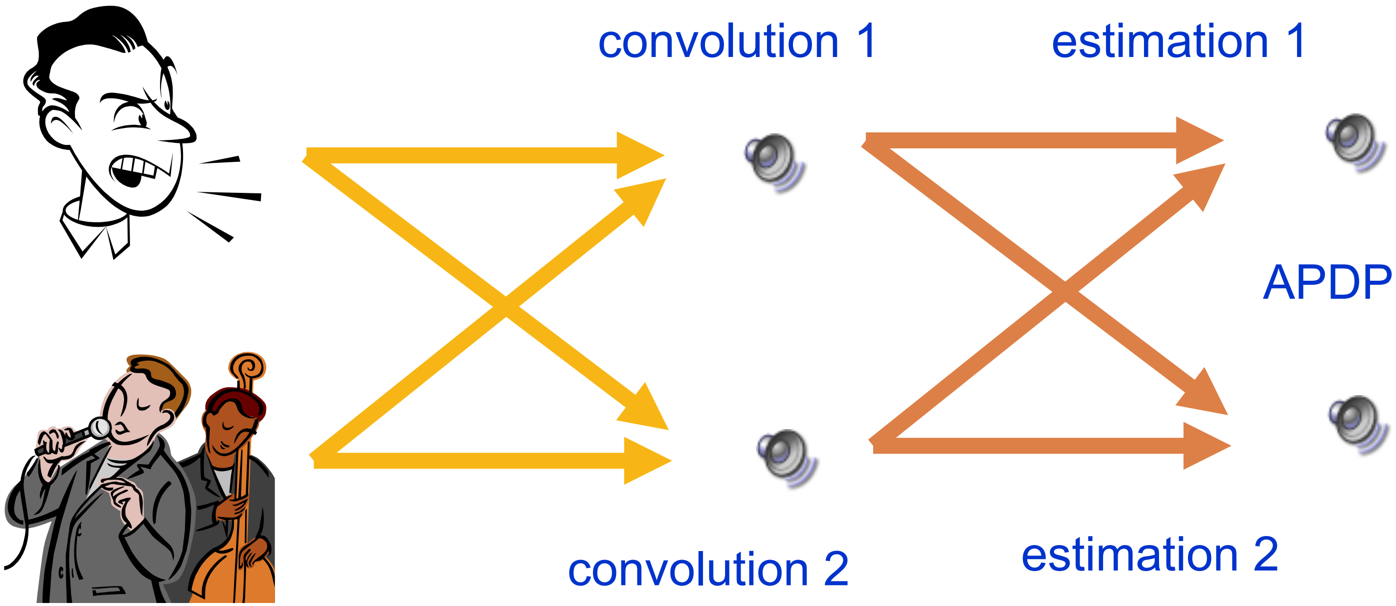
Convolved mixtures

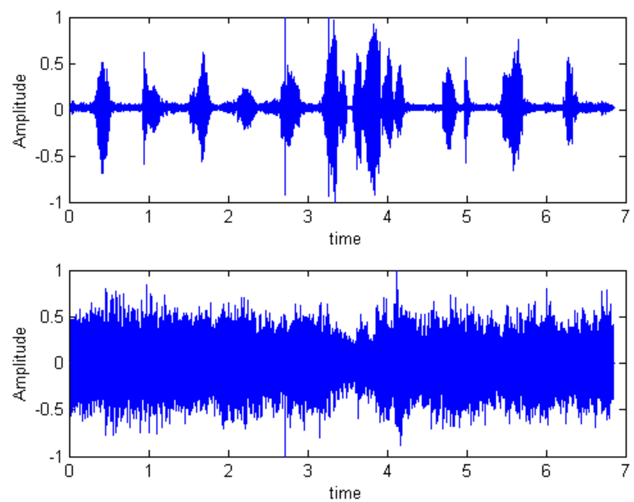


Permutation



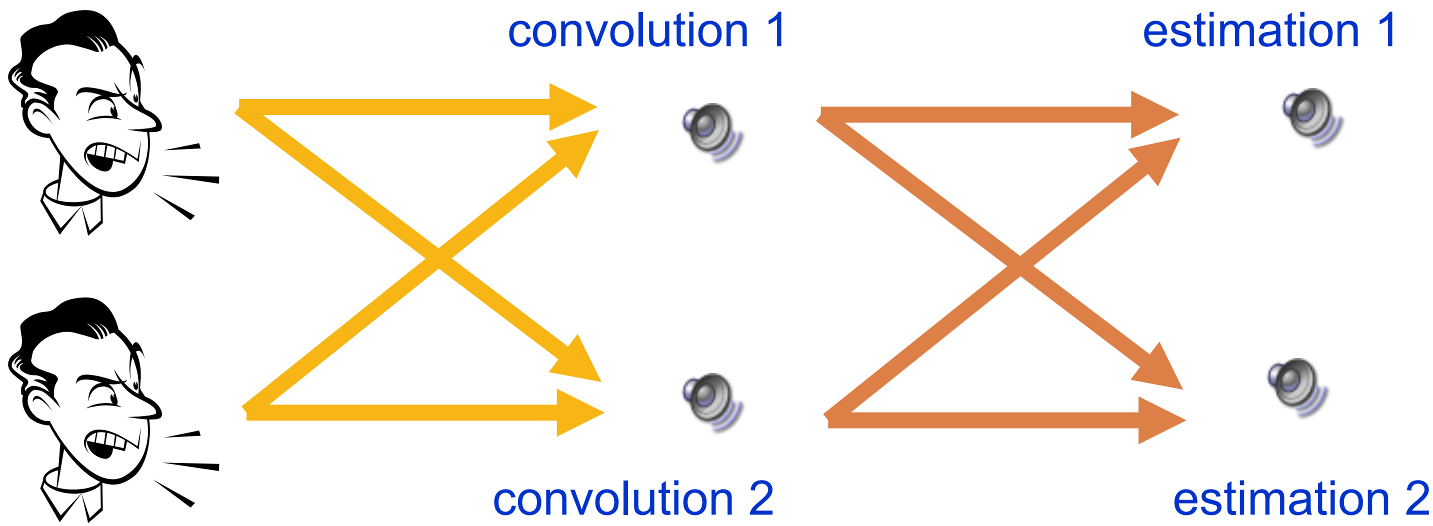
Speech-Music

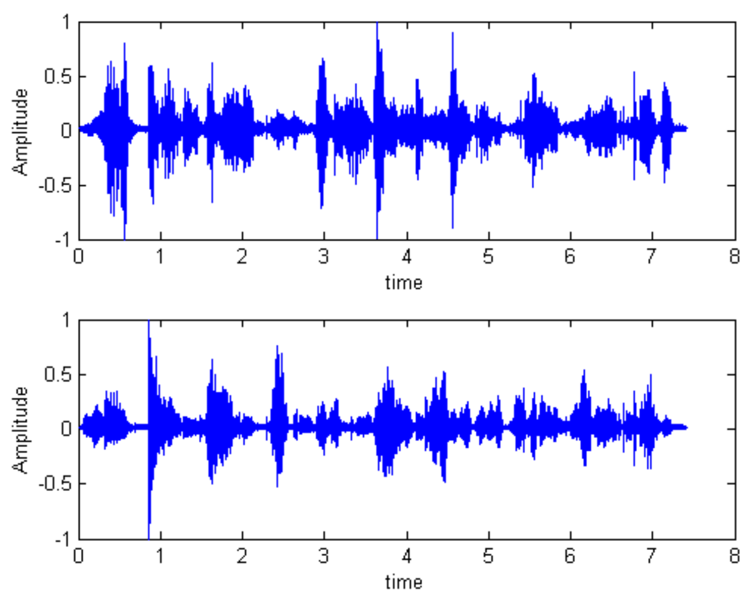




Separated signals

Speech-Speech





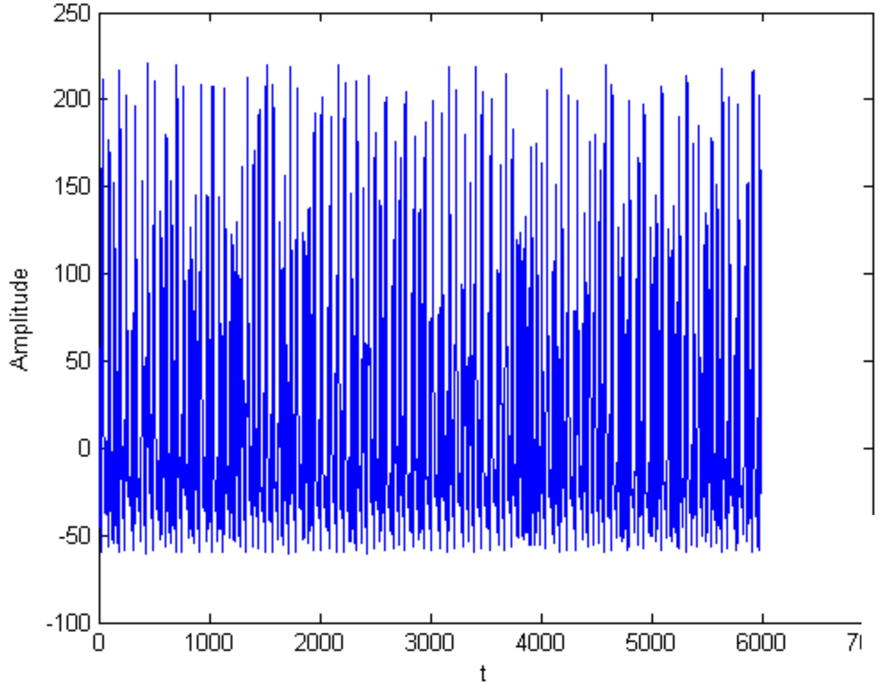
Separated signals

Single channel music transcription

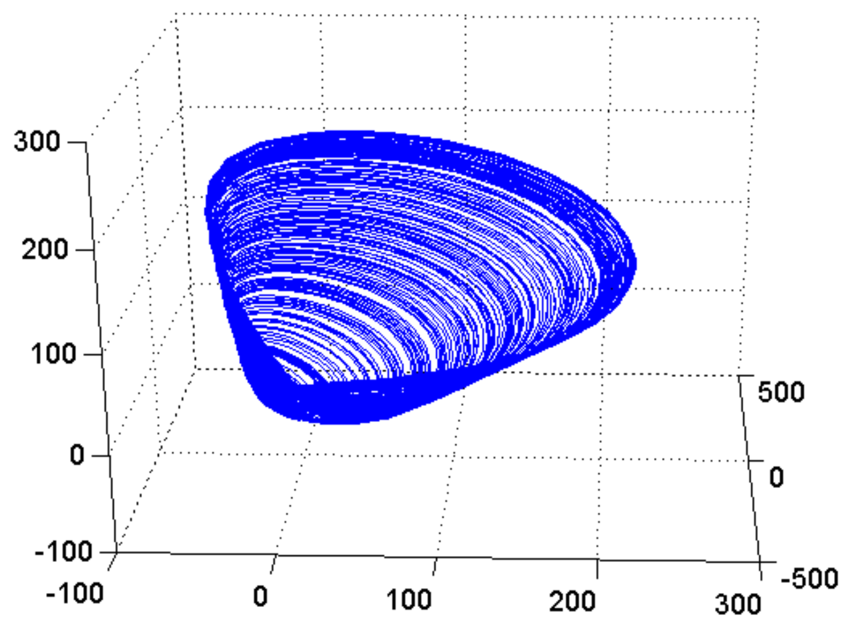
- Extracting multiple source signals from a single channel mixture is an open problem and a challenging research field with several applications

- We present a method based on the phase space reconstruction of the mixture
  - we estimate the embedding dimension and the time lag of a mixture
  - we use the estimated parameters to obtain the architecture (input and output neurons) of a Robust PCA (RPCA) Neural Network

Chaotic features

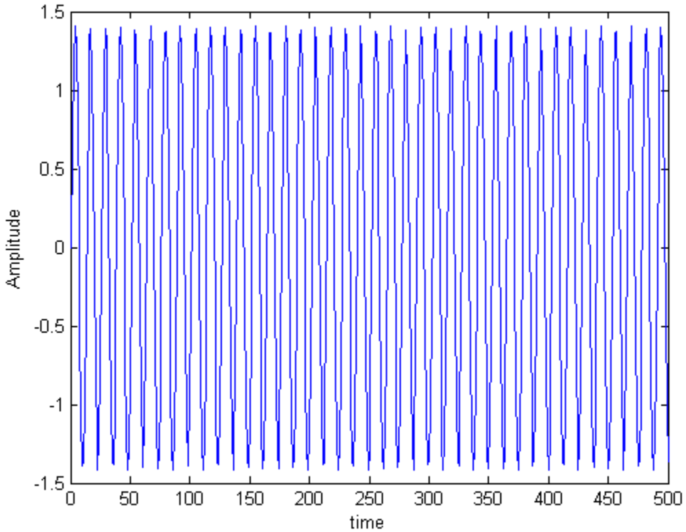


Time-series generated by an electronical Colpitts circuit that oscillates chaotically.

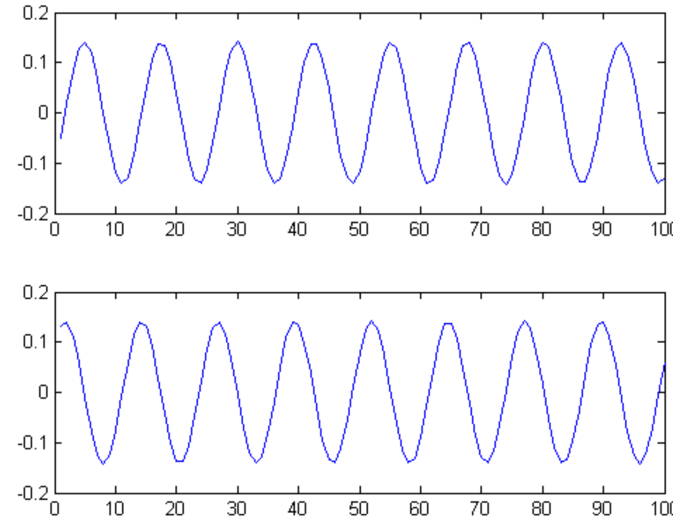


Phase space representation with
$\tau = 4$ and embedding dimension = 3

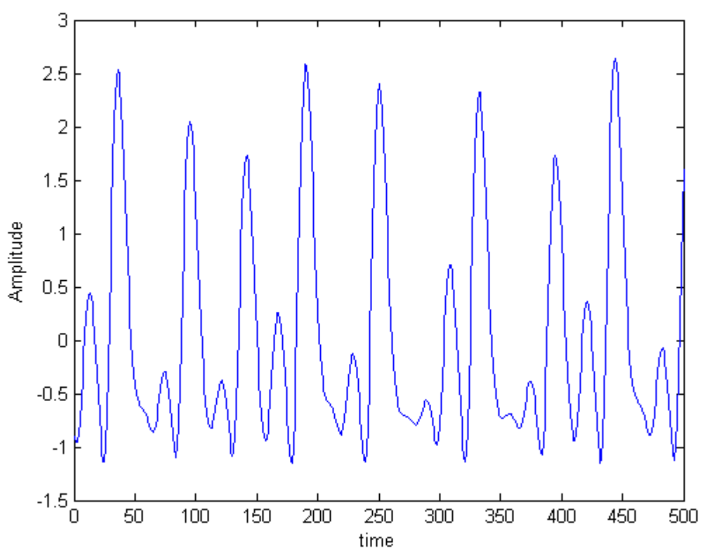
Oscillator and collpits



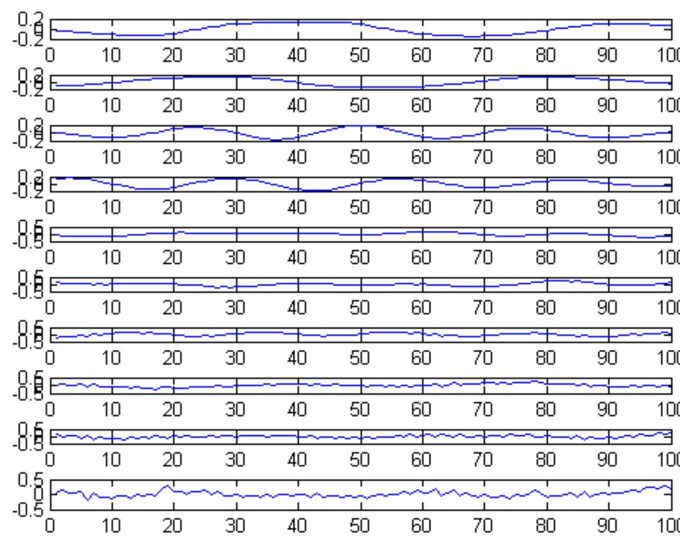
Harmonic oscillator



Indepedent basis obtained by using RPCA

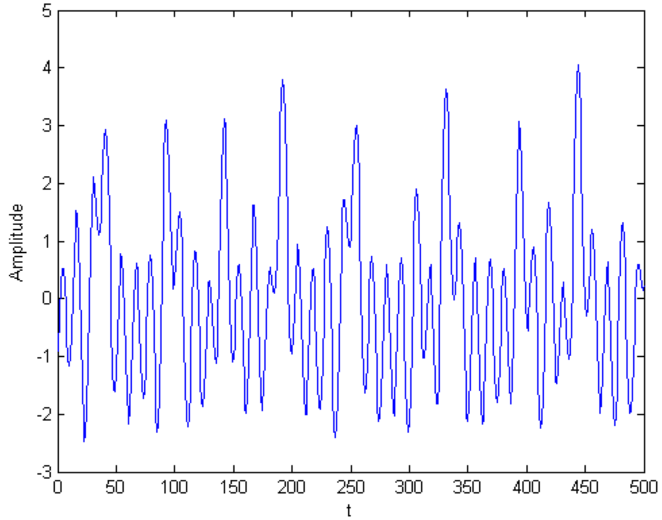


Colpitts time series

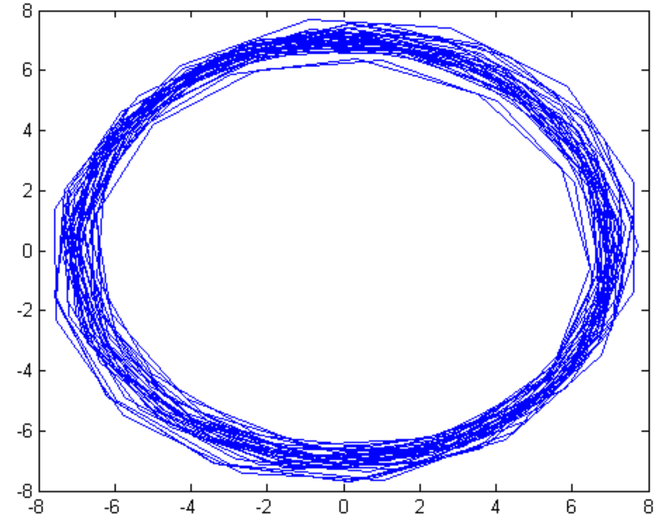


Independent basis obtained by using RPCA

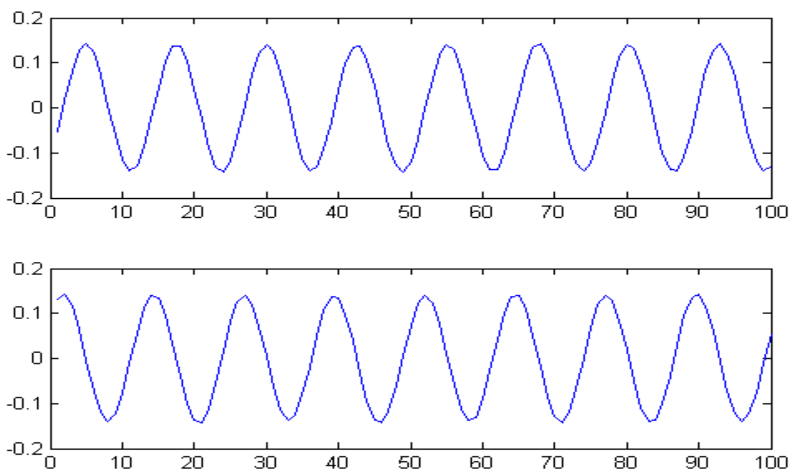
Oscillator



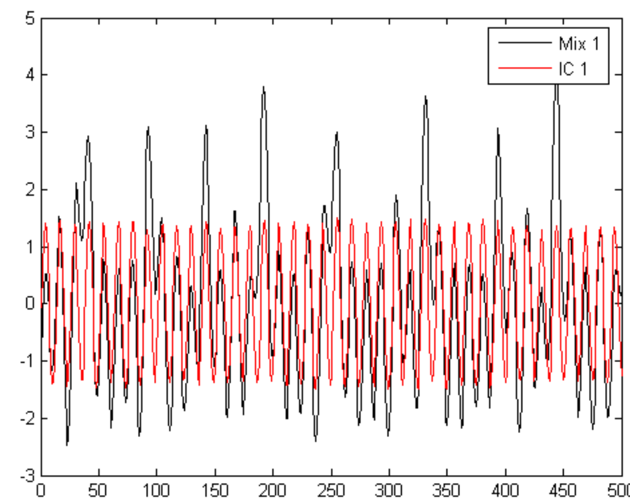
Mixture



Phase space on the first two components

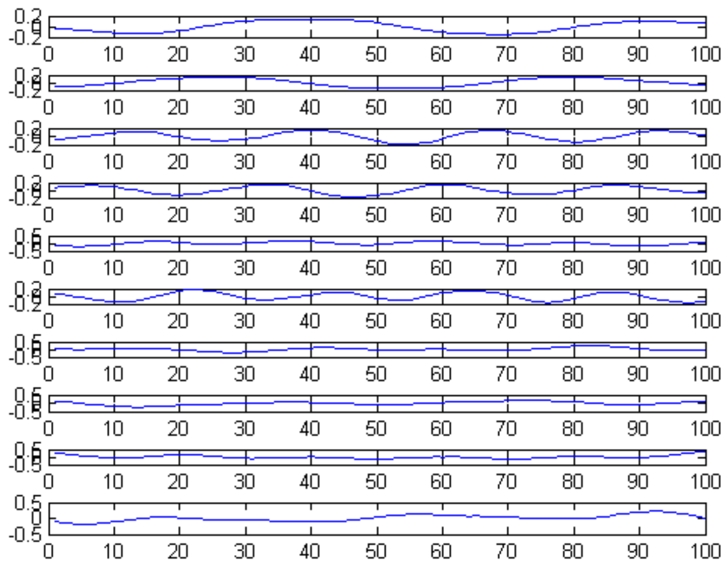


First two independent basis

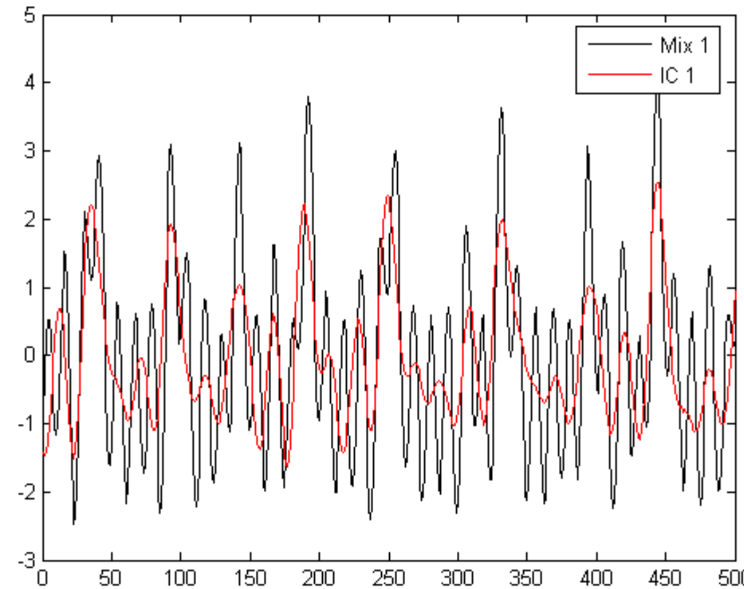


Comparison between the mixture and the reconstructed signal

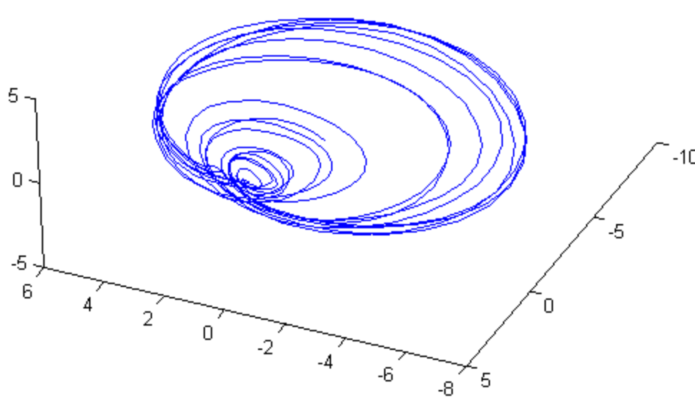
Colpitts



10 independent basis components



Comparison between the mixture and reconstructed signal



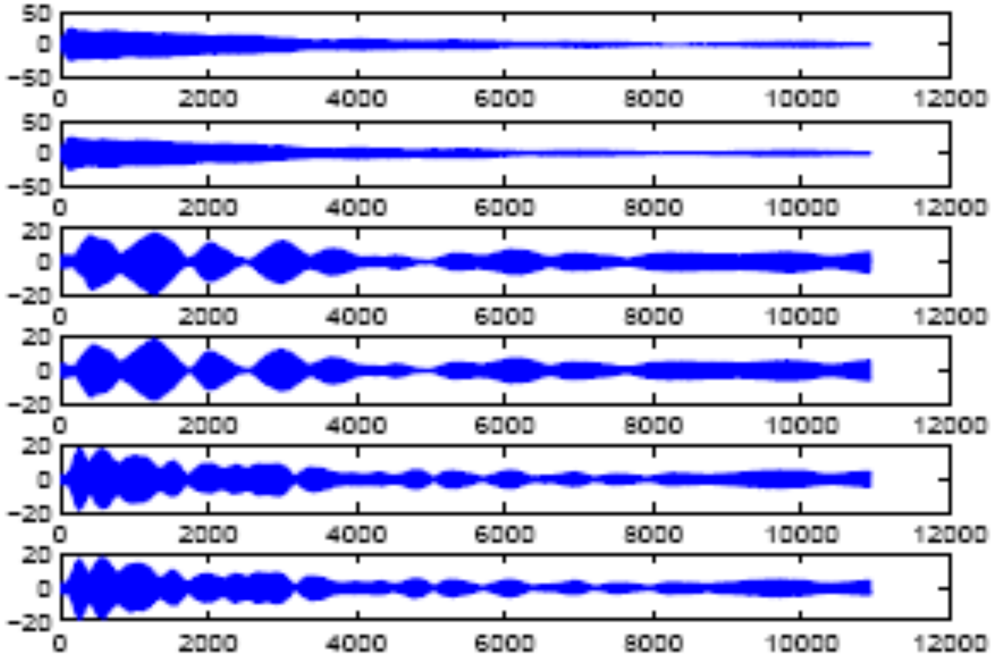
Phase space reconstruction on 3 of the 10 estimated components

Transcription

- It is based on
  - Robust PCA
  - data base of features
  - Kullback-Leibler
  - MUSIC frequency estimator

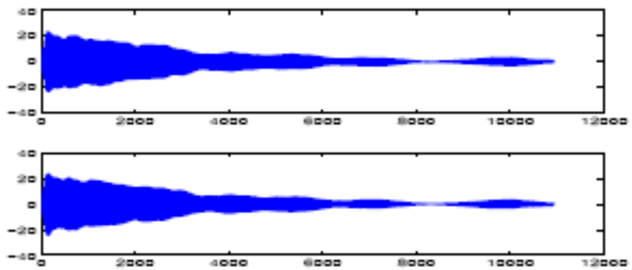
- Several results are proposed
  - Synthetic
    - Piano and trumpet
  - Real recording
    - Blue room – Chet Baker
  - Real recording in a real environment
    - Shadows – Kismet

Experimental results



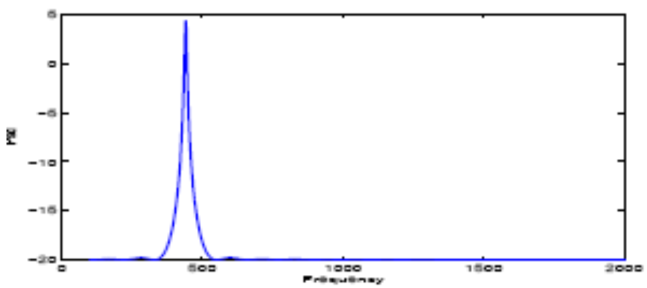
Non-linear PCA separated sources

Experimental results



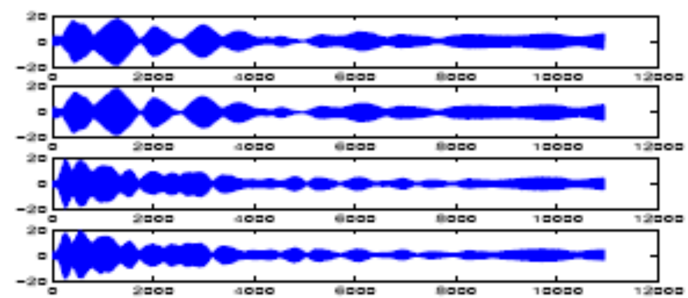
Piano classified components

Time domain



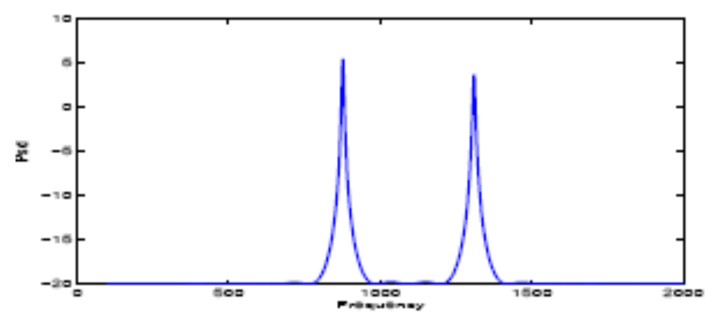
MUSIC frequency estimation

trumpet classified components



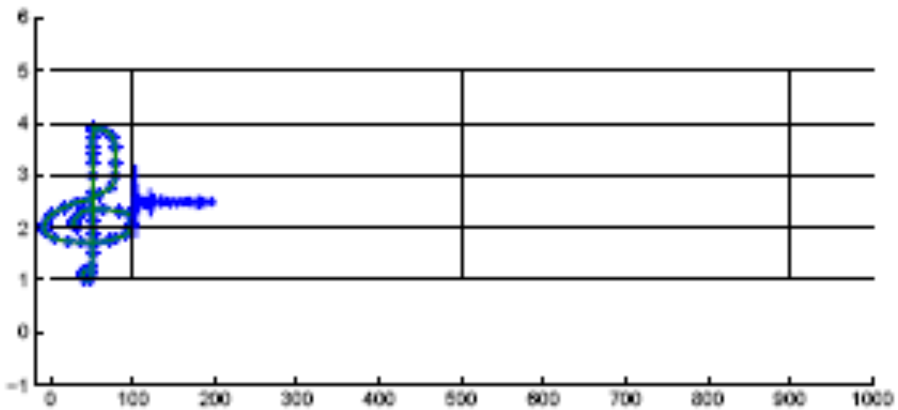
Time domain

MUSIC frequency estimation

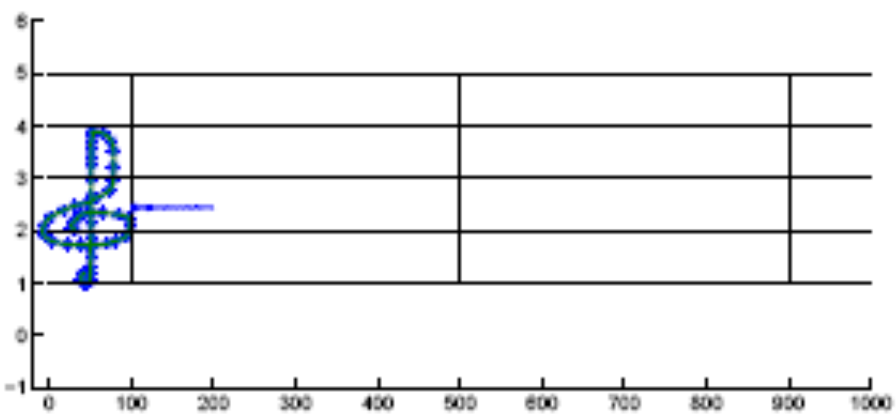


Experimental results

Transcription



piano



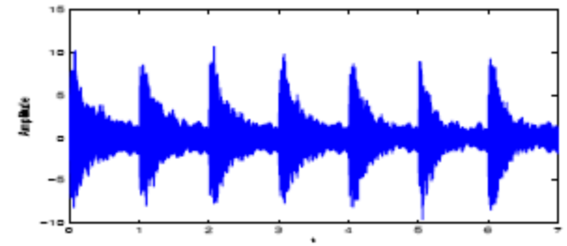
Trumpet

Experimental results

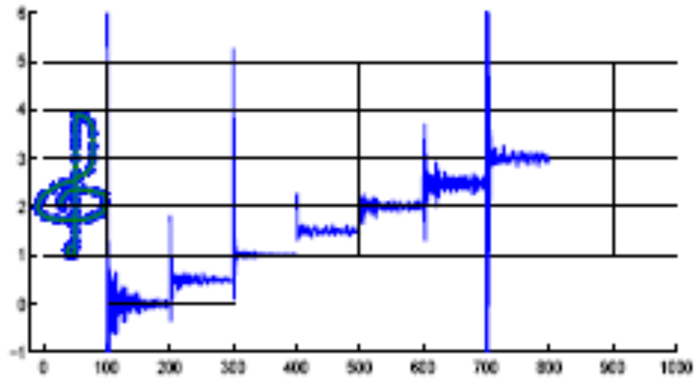
- We consider different notes for the piano (A_4, G_4, C_4, D_4, E_4, F_4, B_4) and the single note A_4 for the trumpet

- Aim of this experiment is to show that we can obtain a perfect separation also varying the fundamental frequency of the waveform

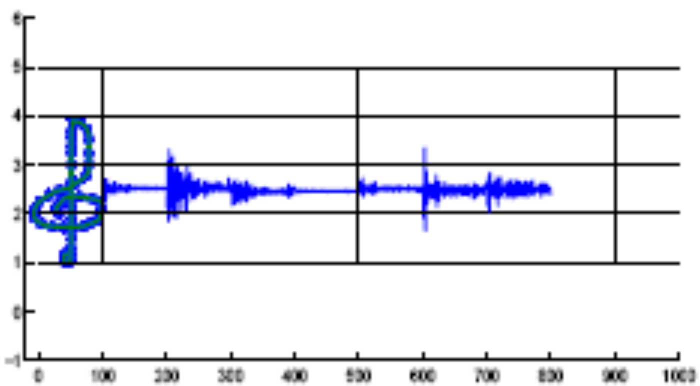
Experimental results



Transcription



piano

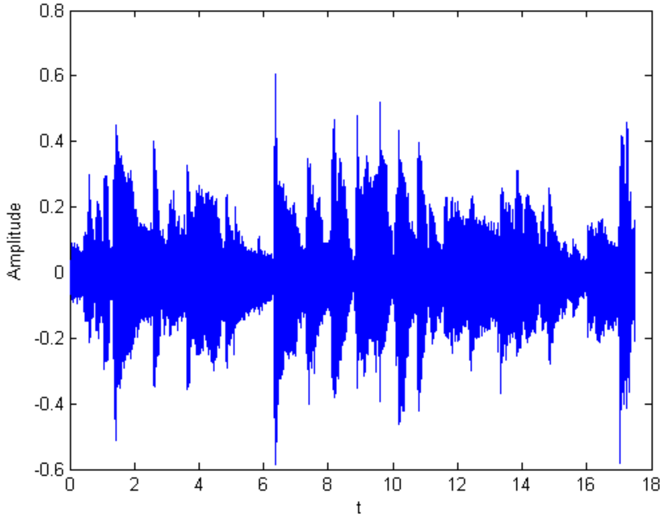


trumpet

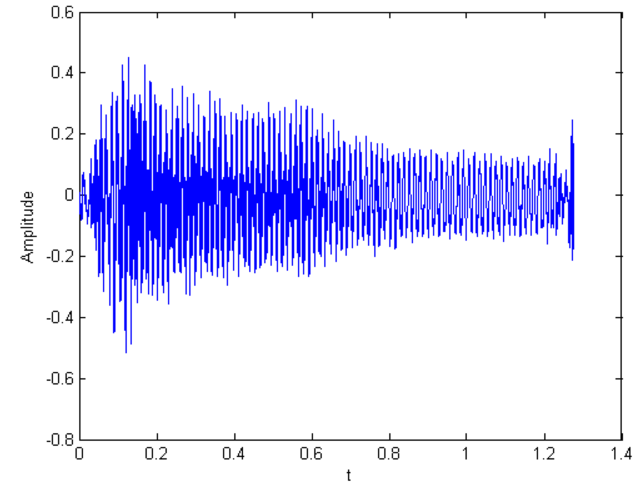
Experimental results

- We consider a real recorded song
- Blue Room song played by Chet Baker
  - Drum (shuffle)
  - Piano
  - Bass
  - Trumpet
- We present the extraction of the trumpet score

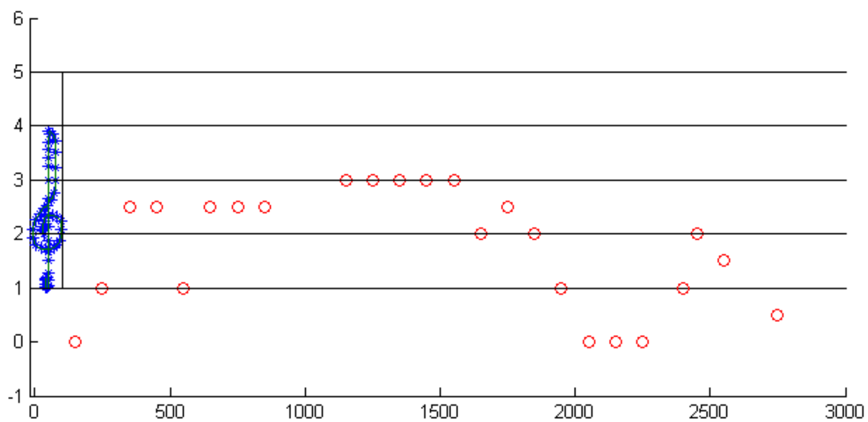
Experimental results



Solo sequence



Wafeform in the database



score

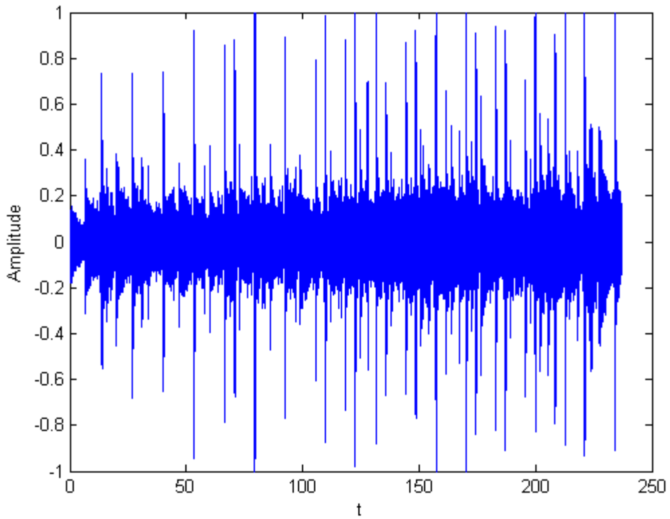
Experimental results

- We consider a song recorded in a real environment

- Shadows song played by Kismet
  - One drum
  - Bass
  - 2 Guitars (accompaniment and solo)
  - Voice

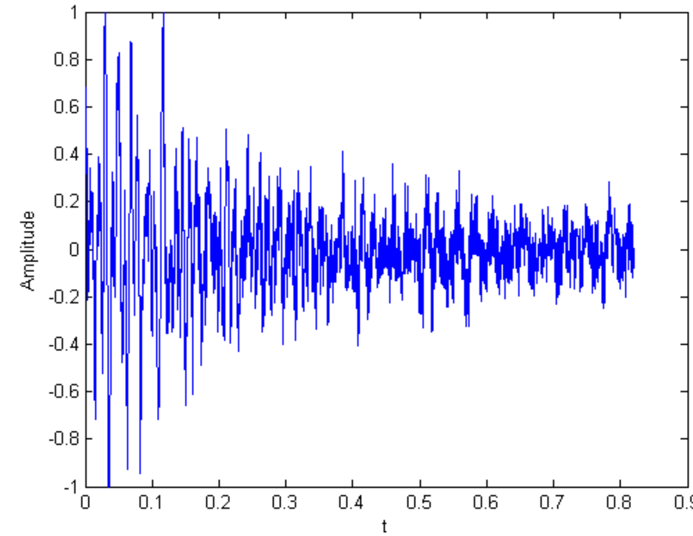
- We present the extraction of the snare and the kick bass drum scores

Experimental results

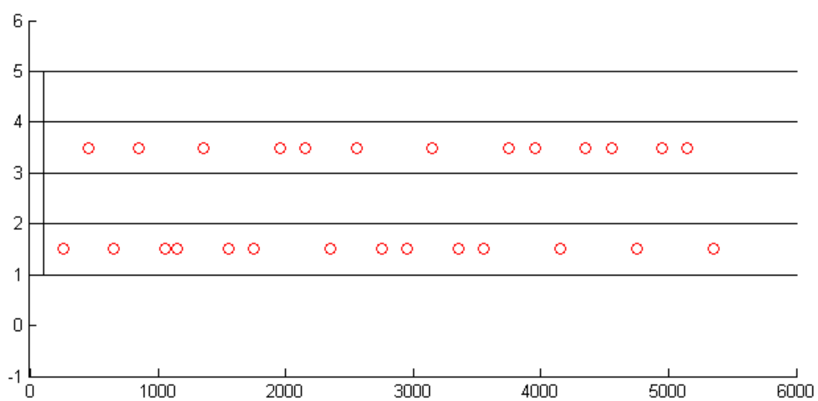


Sequence

snare

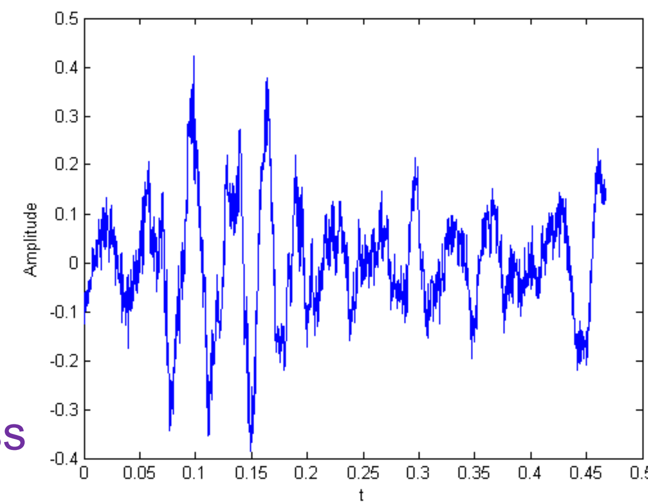


Wafeforms in the database



score

kick bass



NEgentropy based Clustering (NEC)

- The proposed approach divides the clustering phase in several steps
  - Pre-processing
    - Robust PCA for unevenly sampled data
  - Pre-clustering
    - Competitive learning (Winner Take All, Self Organizing Maps, Probabilistic Principal Surfaces, ecc.)
  - Agglomerative clustering
    - Hirarchical agglomerative clustering based on Fisher and Negentropy information
  - Visual and interactive data exploration
    - Multi-Dimensional Scaling (MDS)

Fisher's Linear Discriminant

- Fisher's linear discriminant is a classification method that project high-dimensional data onto a line

- The projection maximizes the distance between the means of the two classes while minimizing the variance within each class

NEC approach

- We agglomerate the regions using the following objective function

$$\boxed{J_{NEC}(\mathbf{x}) \propto \alpha_F J_F(\mathbf{x}) + \alpha_N J_N(\mathbf{x})}$$

Fisher

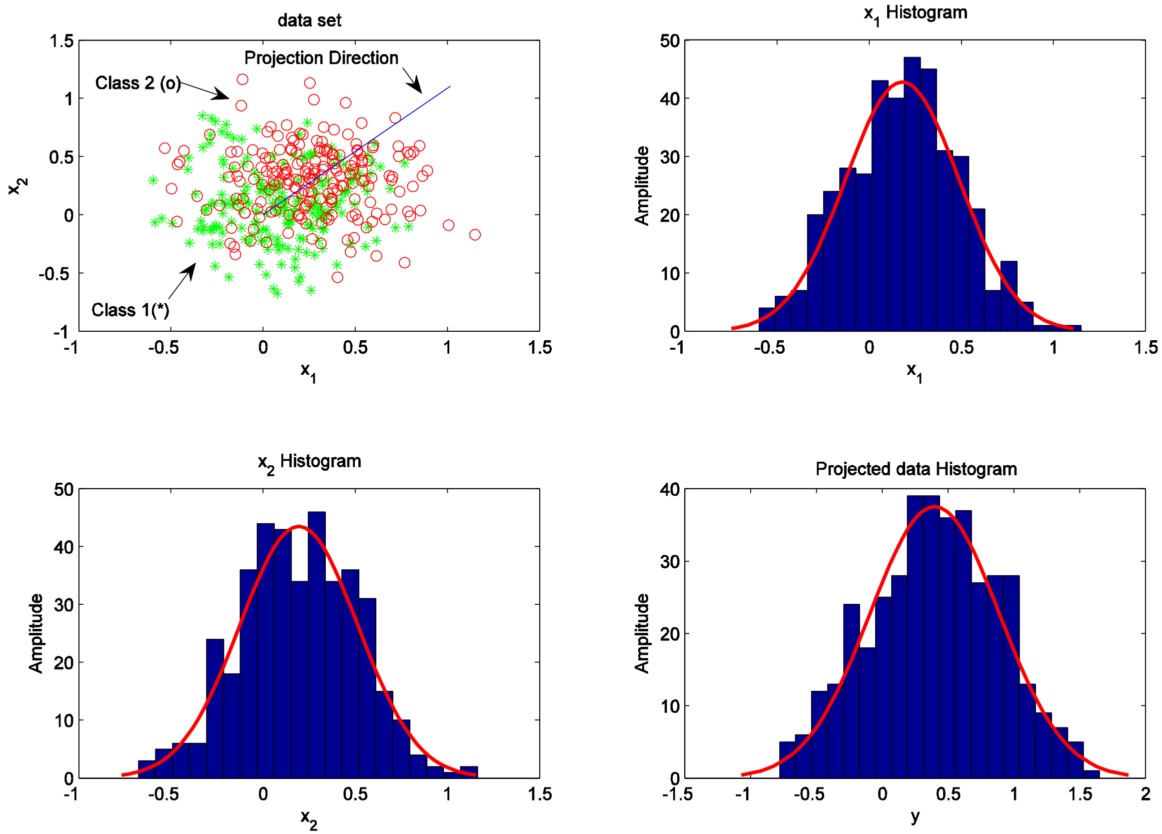
$$\boxed{J_F(\mathbf{x}) = \frac{\mathbf{w}^T \mathbf{S}_{\mathrm{B}} \mathbf{w}}{\mathbf{w}^T \mathbf{S}_{\mathrm{W}} \mathbf{w}}}$$

Negentropy

$$J_N(\mathbf{x}) \propto k_1 E\left\{\mathbf{x}\exp\left(-\mathbf{x}^2/2\right)\right\}^2 + k_2^a\left(E\{|\mathbf{x}|\} - \sqrt{2/\pi}\right)^2$$

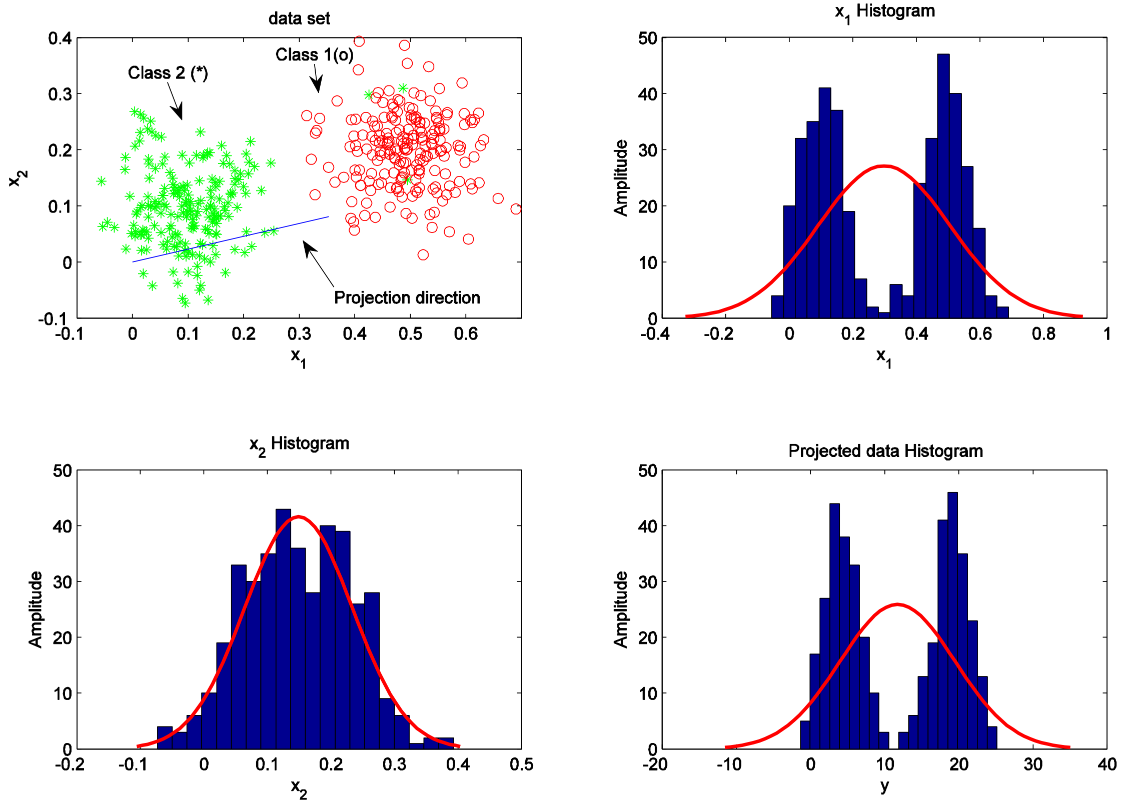
$$J_N(\mathbf{x}) \propto k_1 E\left\{\mathbf{x}\exp\left(-\mathbf{x}^2/2\right)\right\}^2 + k_2^b\left(E\left\{\exp\left(-\mathbf{x}^2/2\right)\right\} - \sqrt{1/2}\right)^2$$

NEC approach



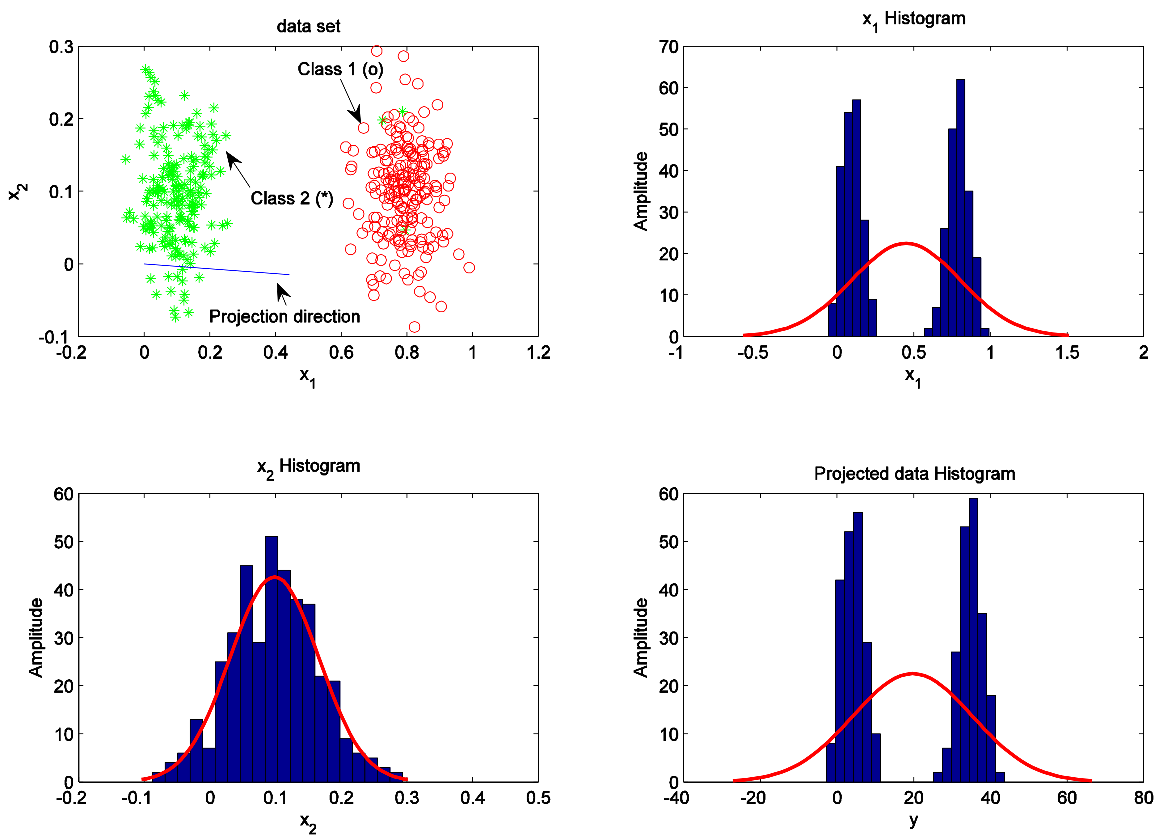
In this case we obtain $J_{NEC} = 0.0079$ in which $\alpha_F J_F = 0.0040$ and $\alpha_N J_N = 9.73\text{x}10^{-4} + 0.0029$

NEC approach



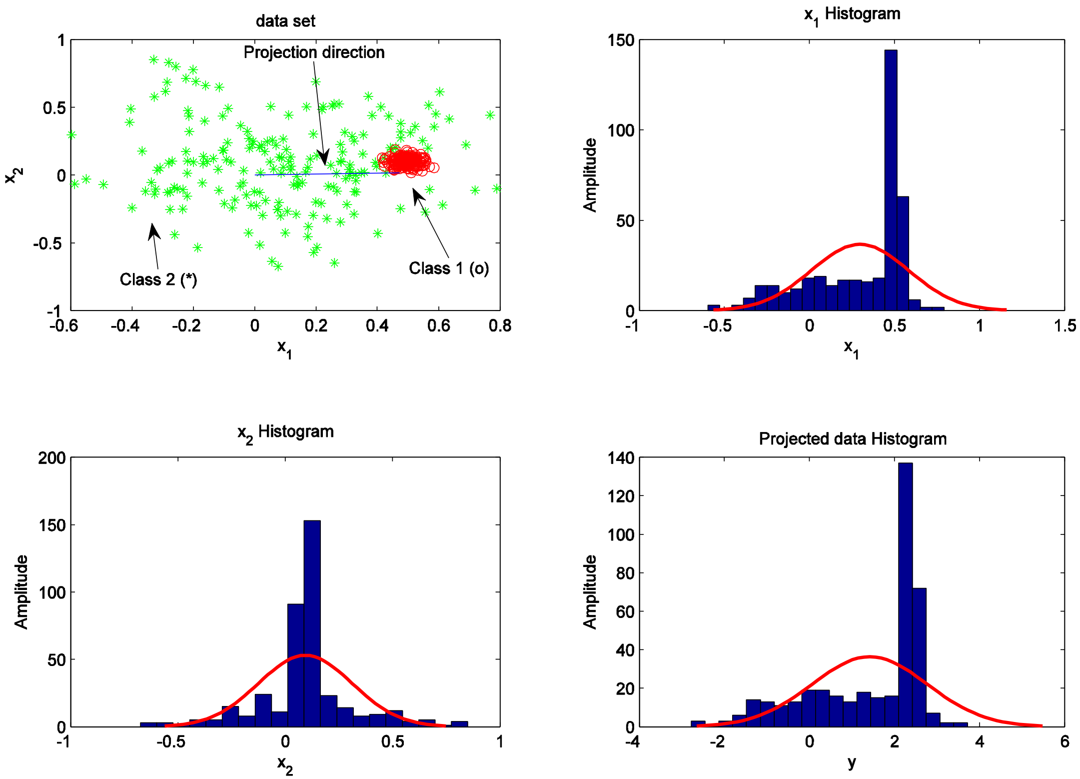
In this case we obtain J_{NEC} =0.4064 in which $\alpha_F\, J_F$ = 0.1455 and $\alpha_N\, J_N = 1.39\text{x}10^{-4} + 0.2669$

NEC approach



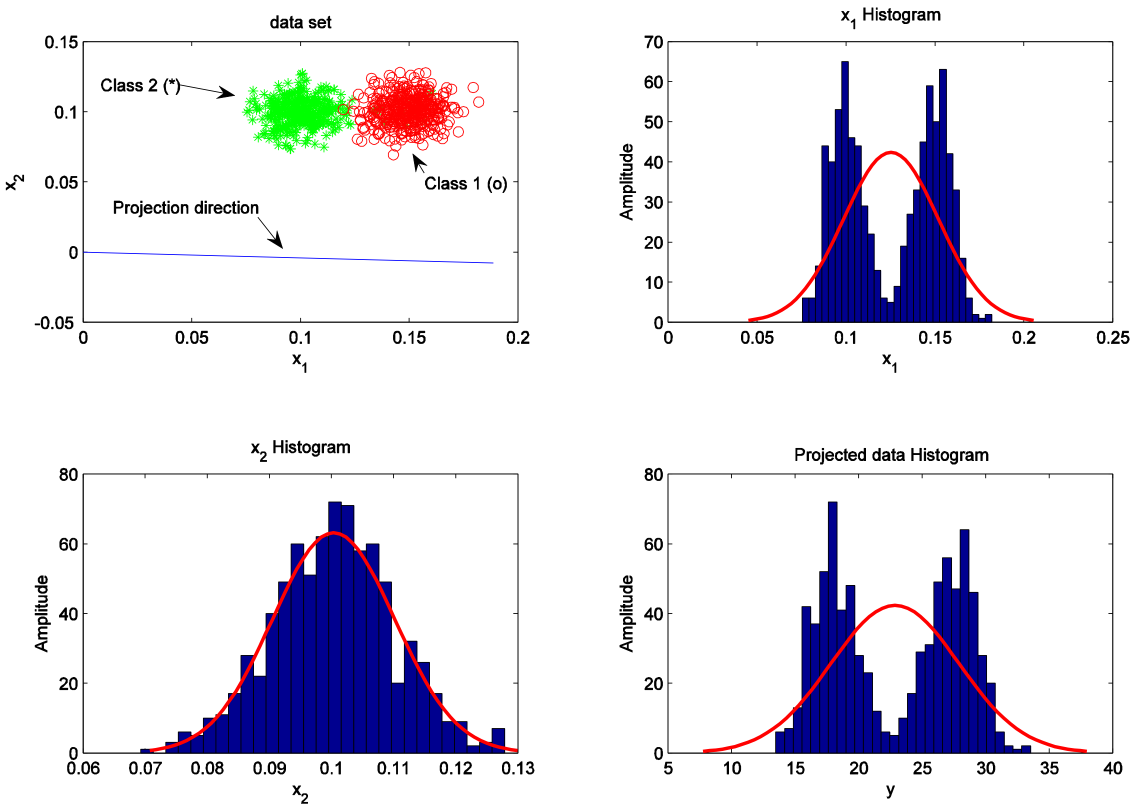
In this case we have two classes with the same direction on the *y* axes.
We obtain $J_{NEC} = 0.6794$ in which $\alpha_F\, J_F = 0.3027$ and $\alpha_N\, J_N = 3.49\text{x}10^{-4} + 0.3763$

NEC approach



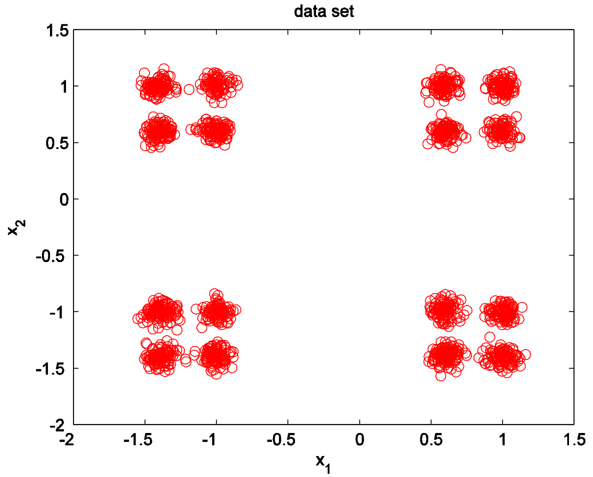
In this case we have two complete overlapped classes. We obtain J_{NEC} = 0.26603 in which $\alpha_F J_F$ = 0.0188 and $\alpha_N J_N$ = 0.2196 + 0.0275. We note that the information is high since the asymmetry information is high.

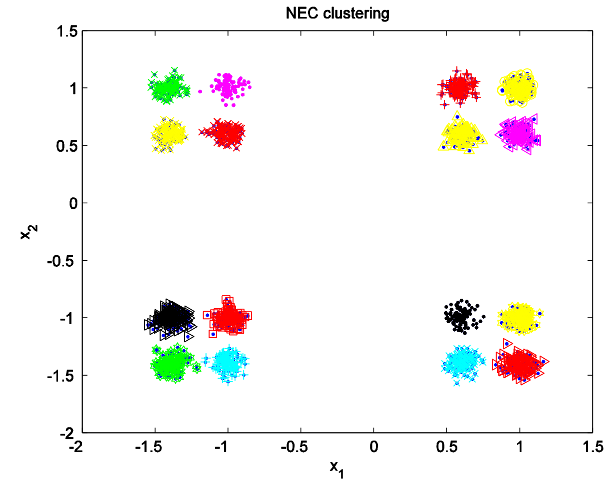
NEC approach

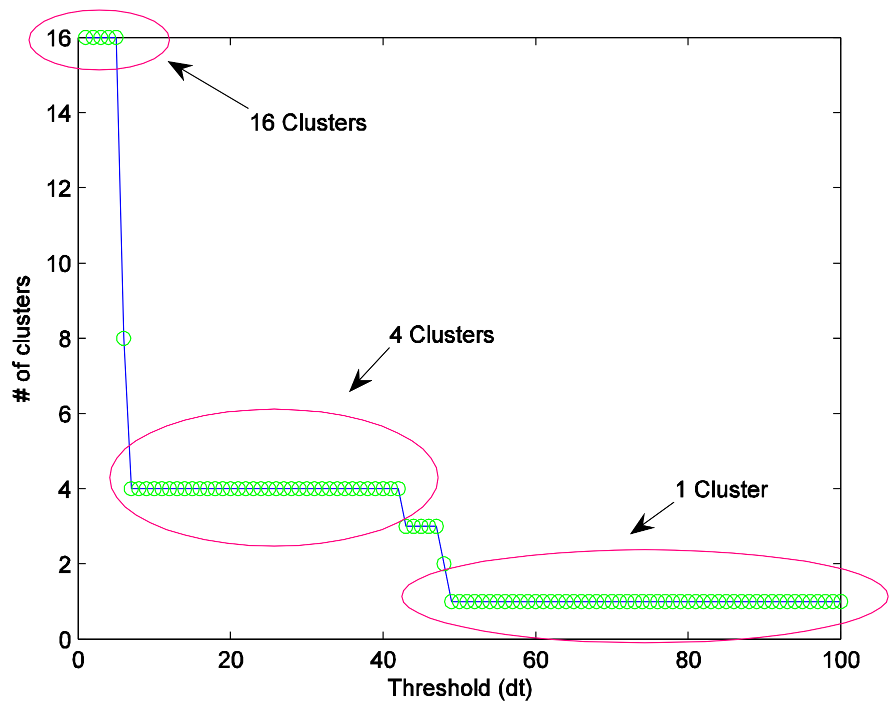


In this case we have two overlapped classes. We obtain $J_{NEC} = 0.2861$ in which $\alpha_F J_F = 0.0910$ and $\alpha_N J_N = 3.64\text{x}10^{-4} + 0.194876$. We note that the information is high since the bimodality/sparsity information is higher than the other information.

NEC approach

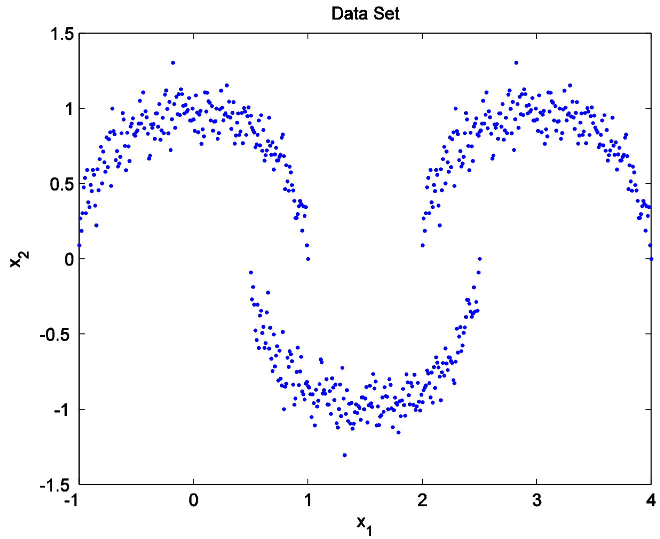


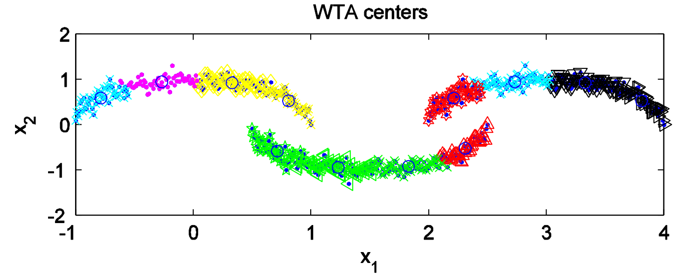


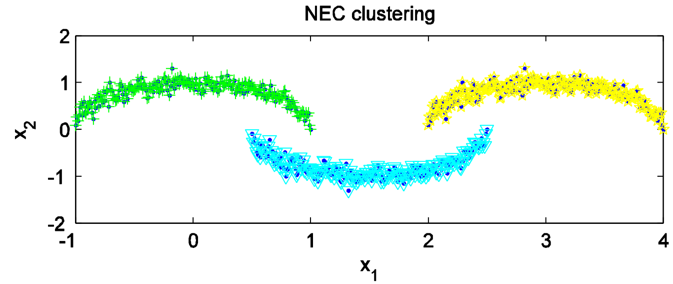


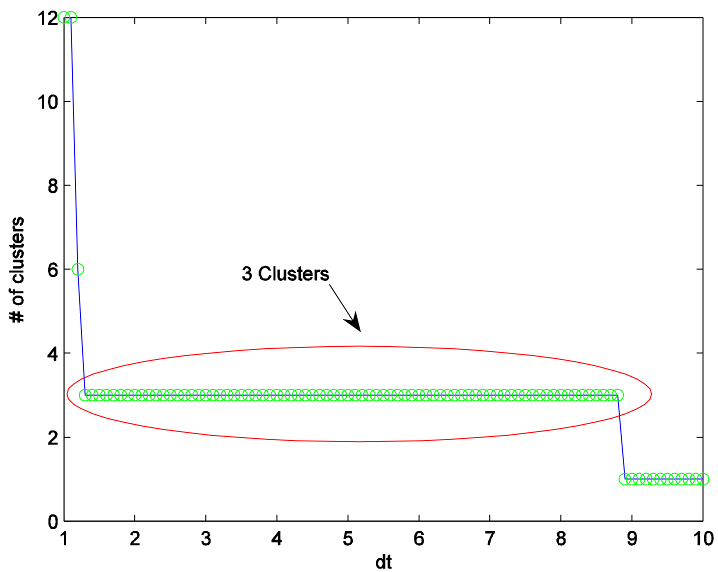
The clusters obtained varying the threshold (dt)

NEC approach



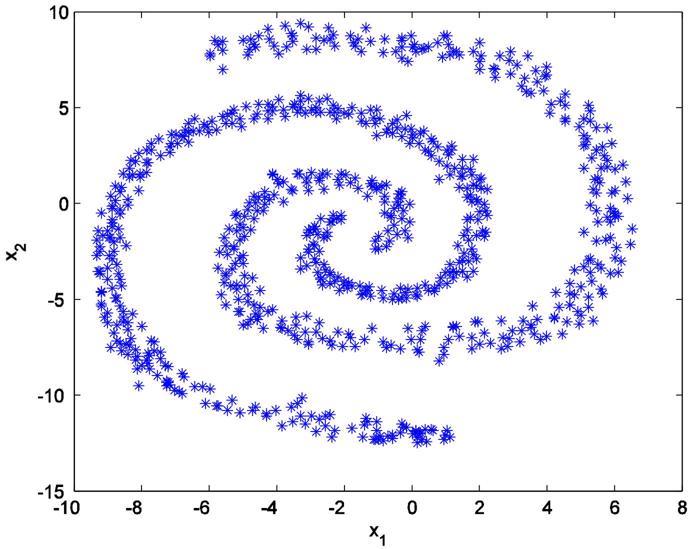


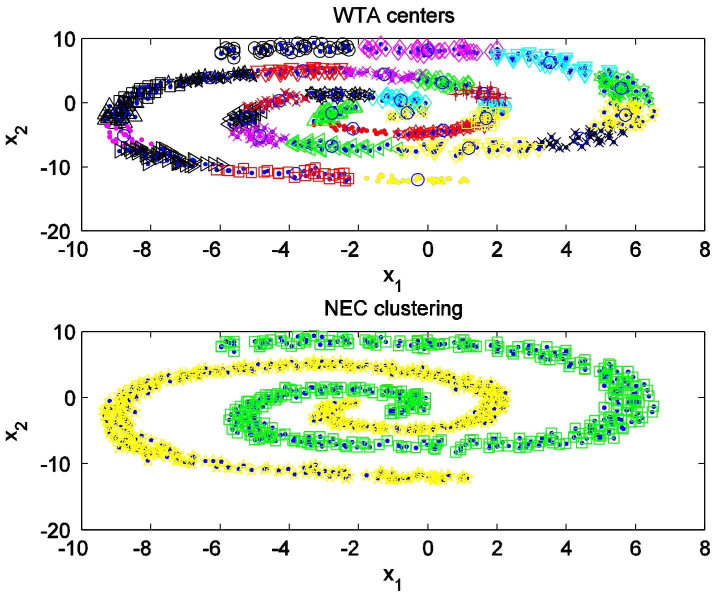


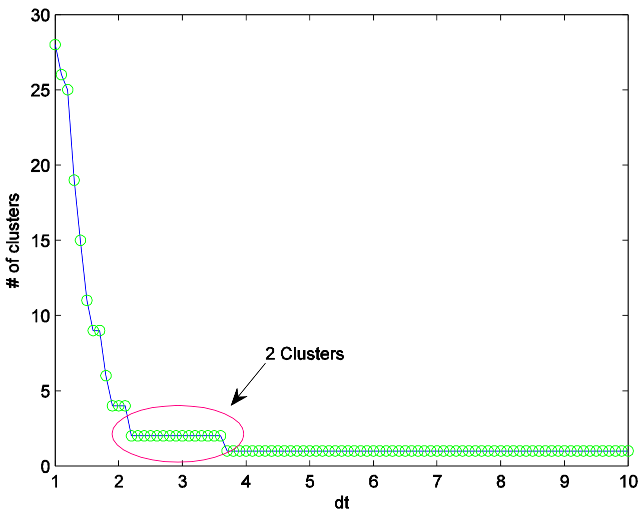


The clusters obtained varying the threshold (dt)

NEC approach

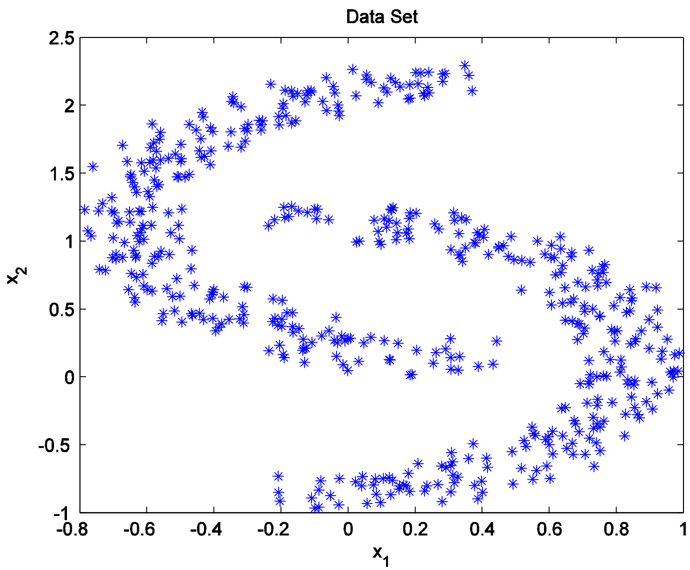






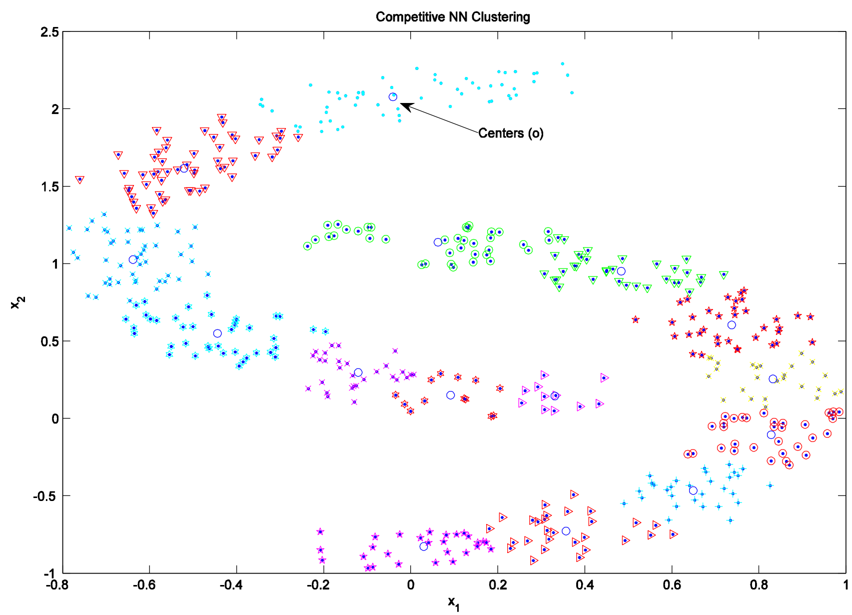
The clusters obtained varying the threshold (dt)

Clustering and Negentropy

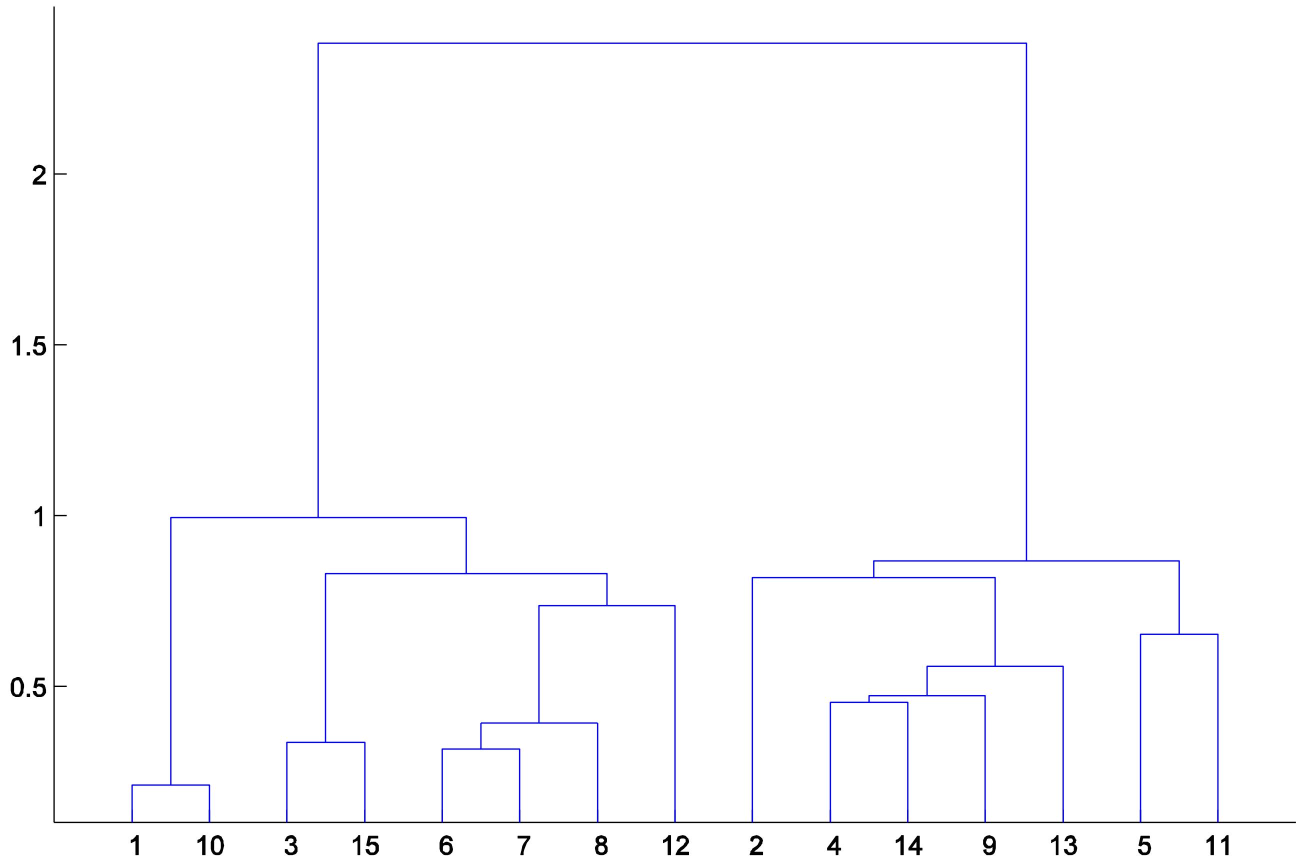


In this case we consider a 2-dimensional data set composed by two separated classes

Clusters obtained by using unsupervised clustering

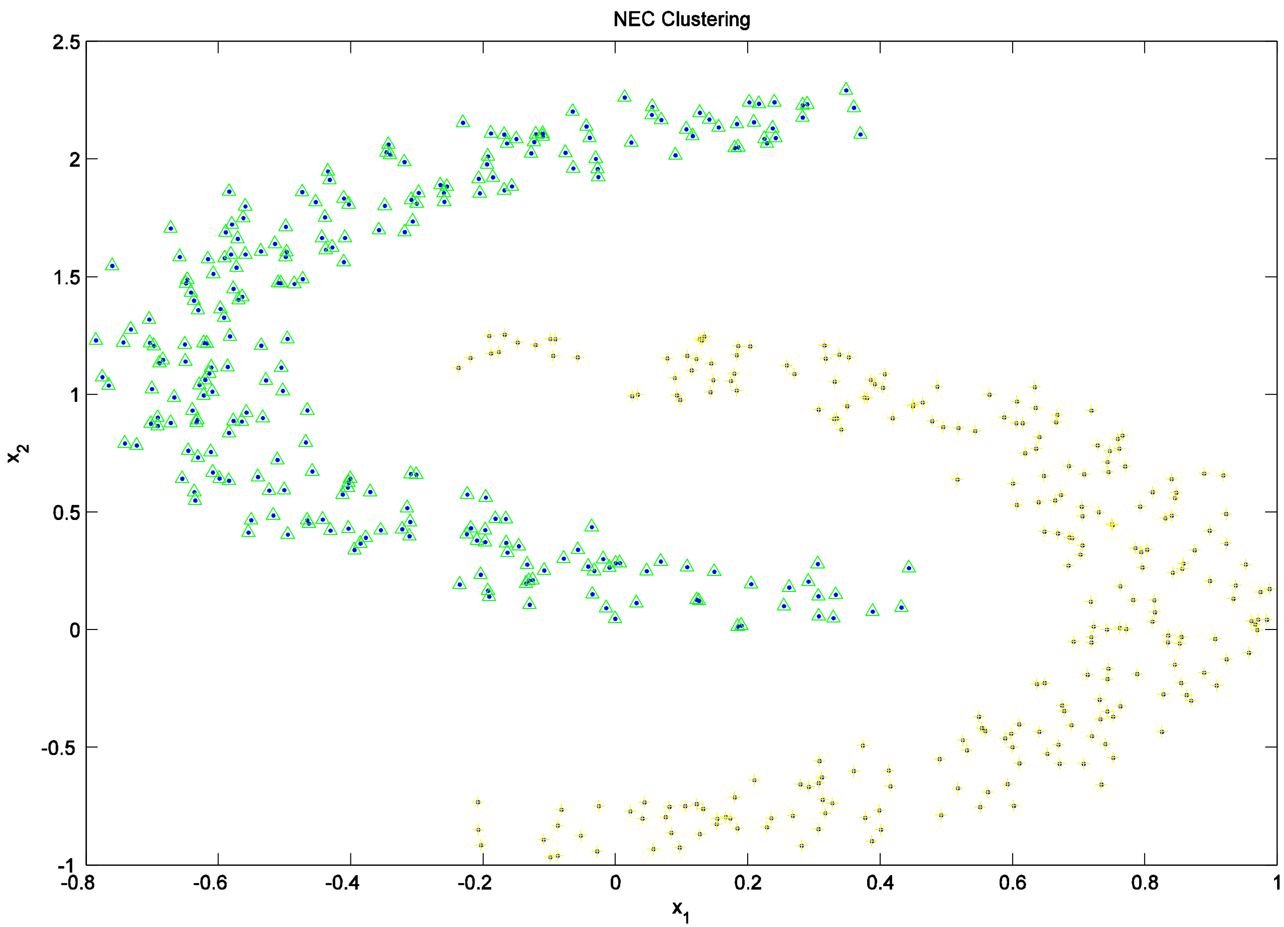


Dendogram



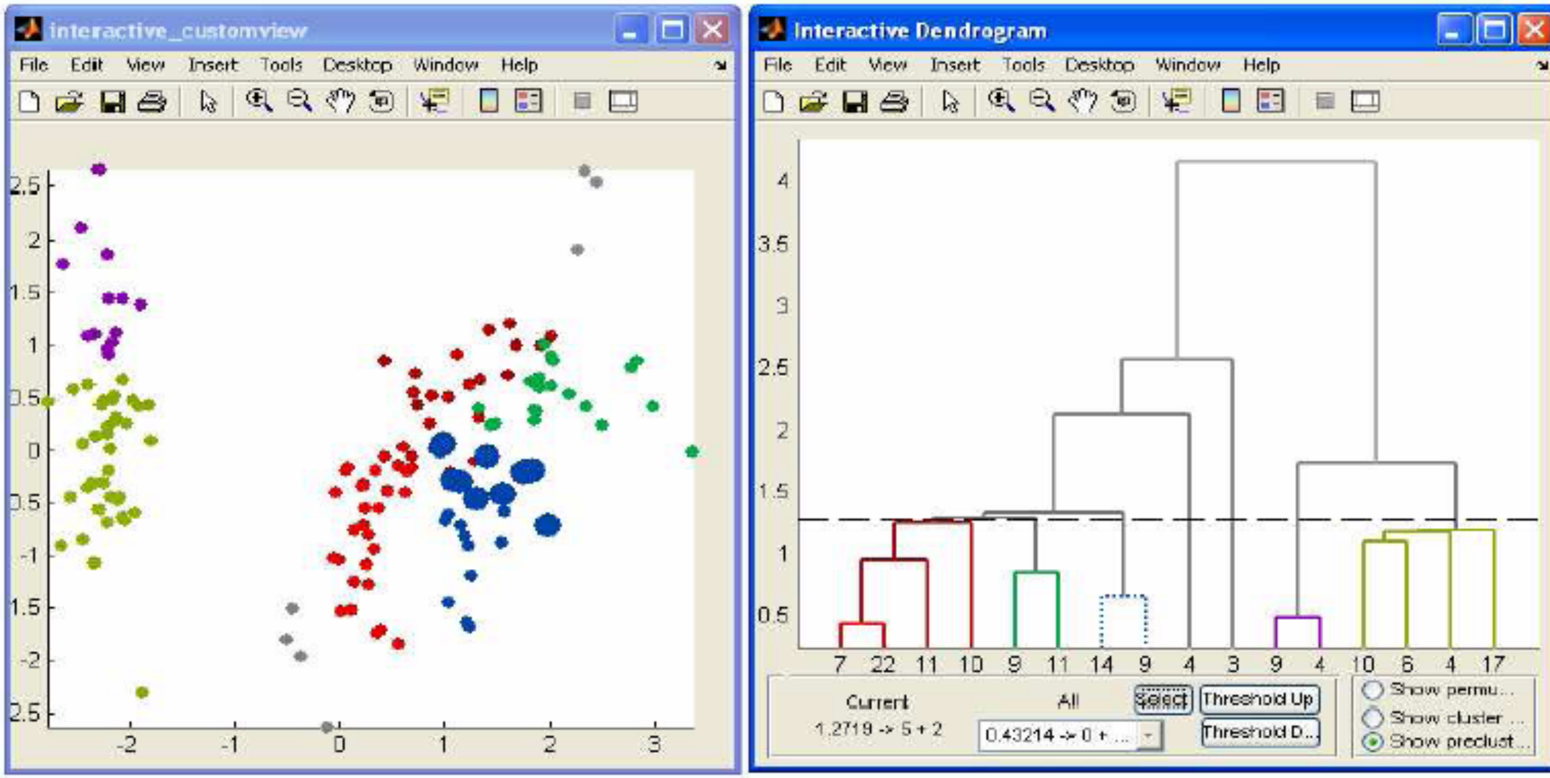
Dendogram obtained by using the NEC approach

Dendogram and clusterization



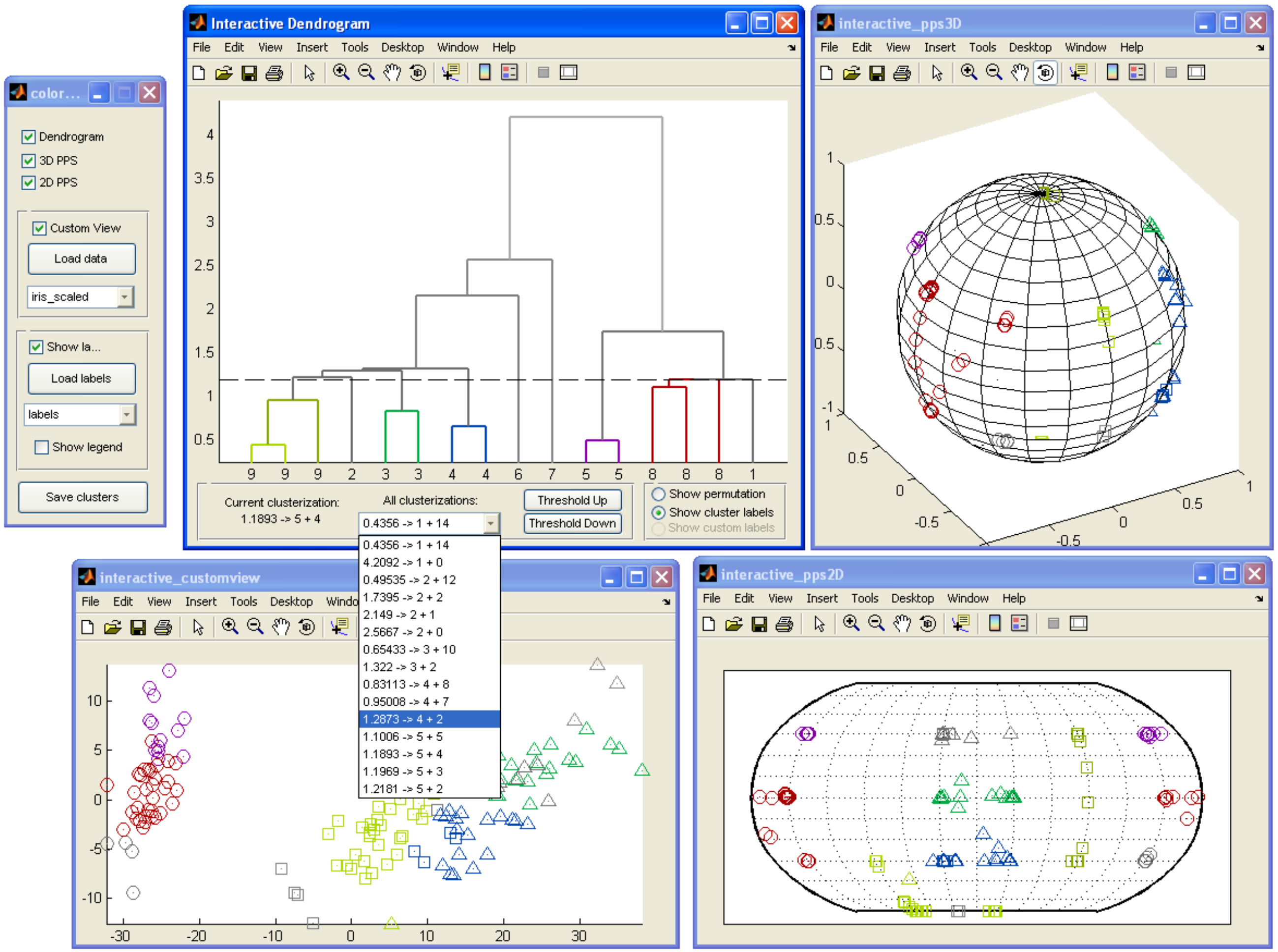
Clusters obtained by using the NEC approach

Interface



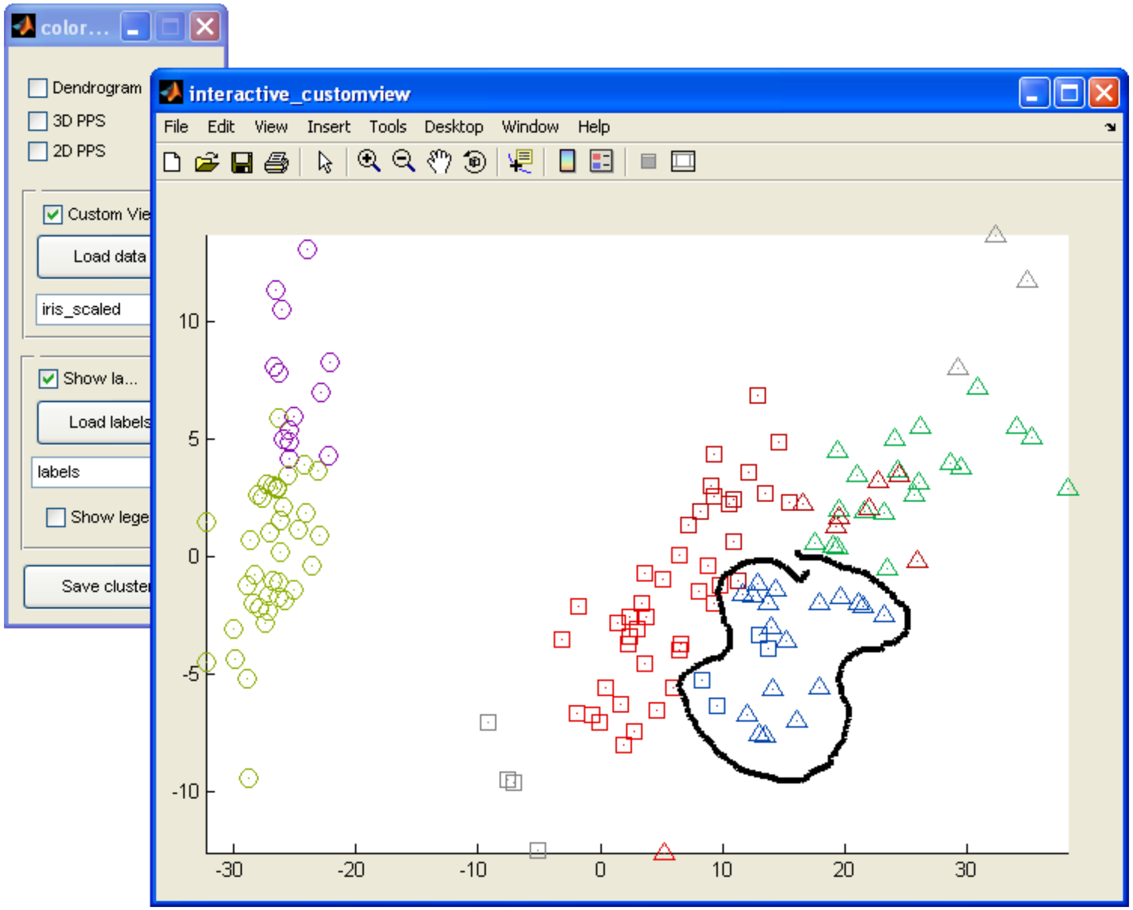
NEC toolbox

Interface



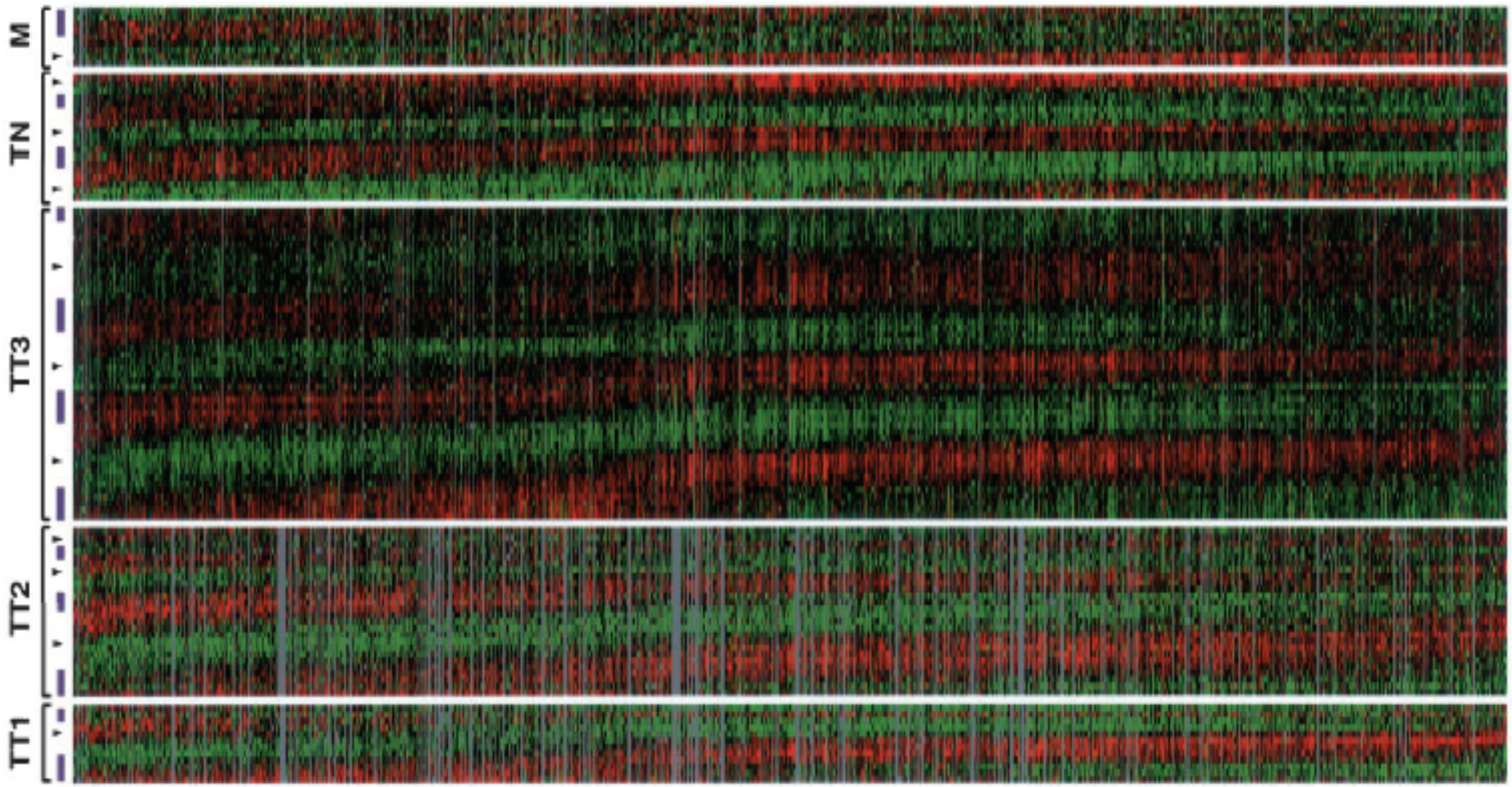
NEC toolbox

Interface



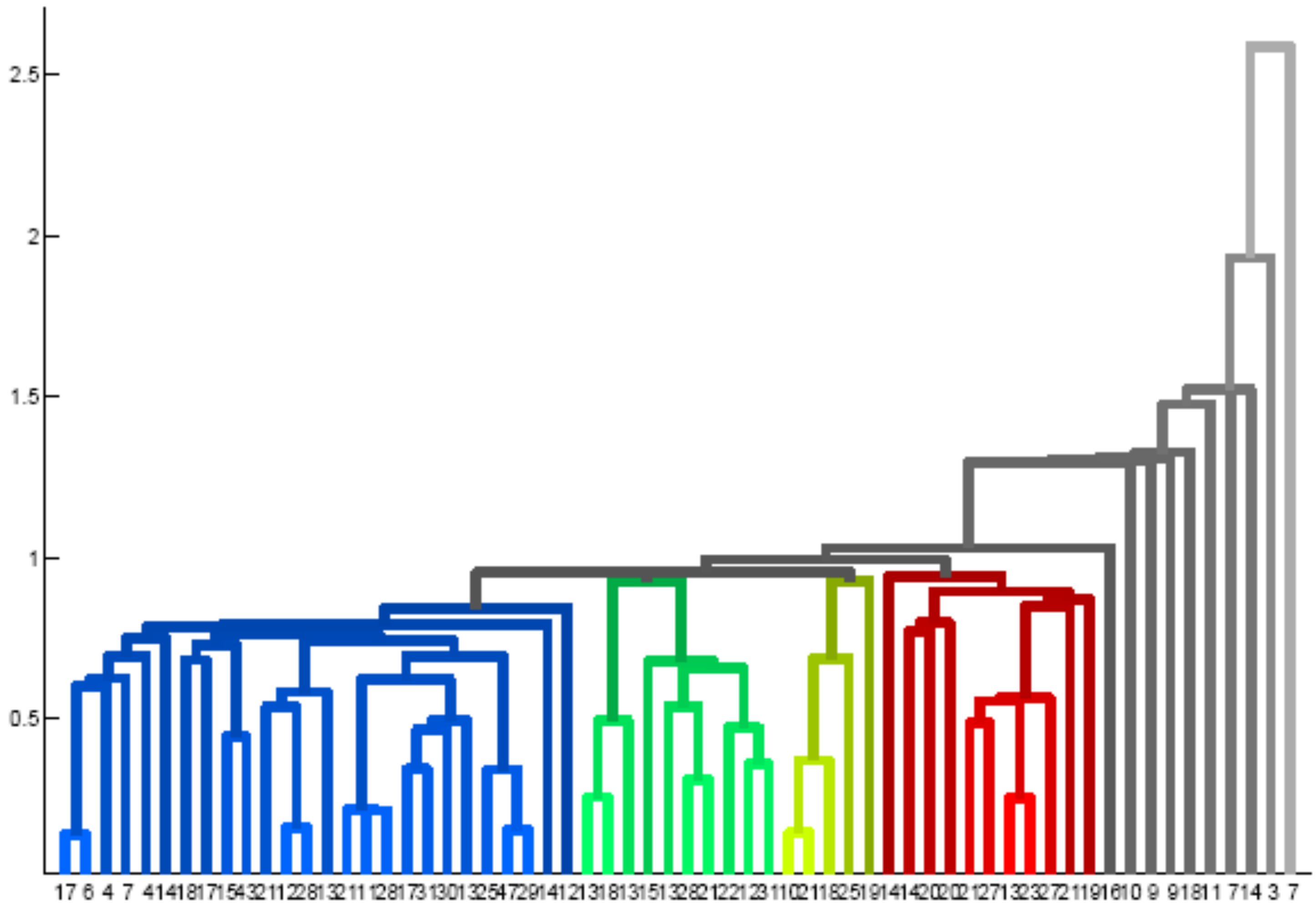
NEC toolbox

Human Cell Cycle Gene Expression Data

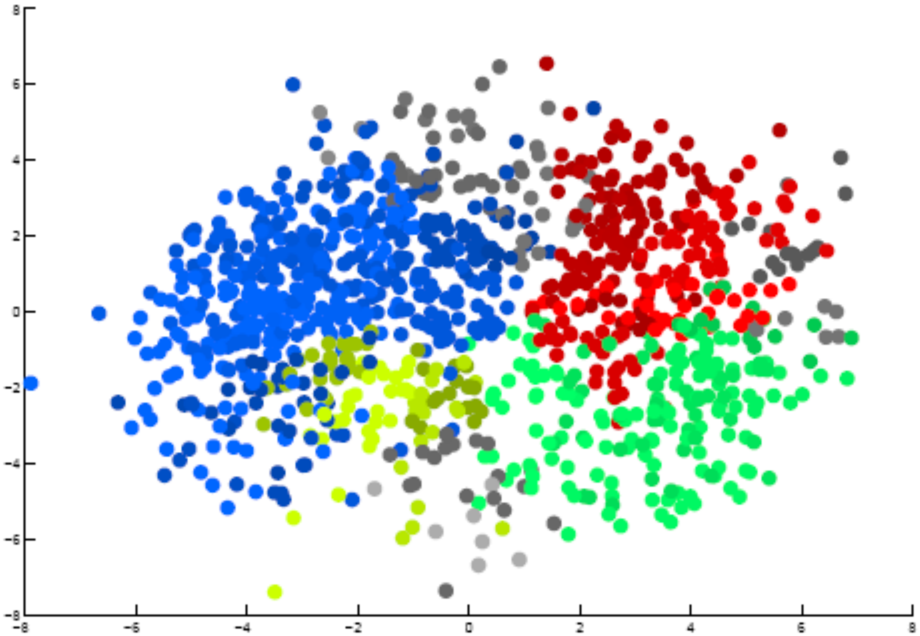


Collaboration with the Department of Physics of the Unversity of Naples "Federico II" and Telethon Institute of Genetics and Medicine (TIGEM) of Naples

NEC dendrogram

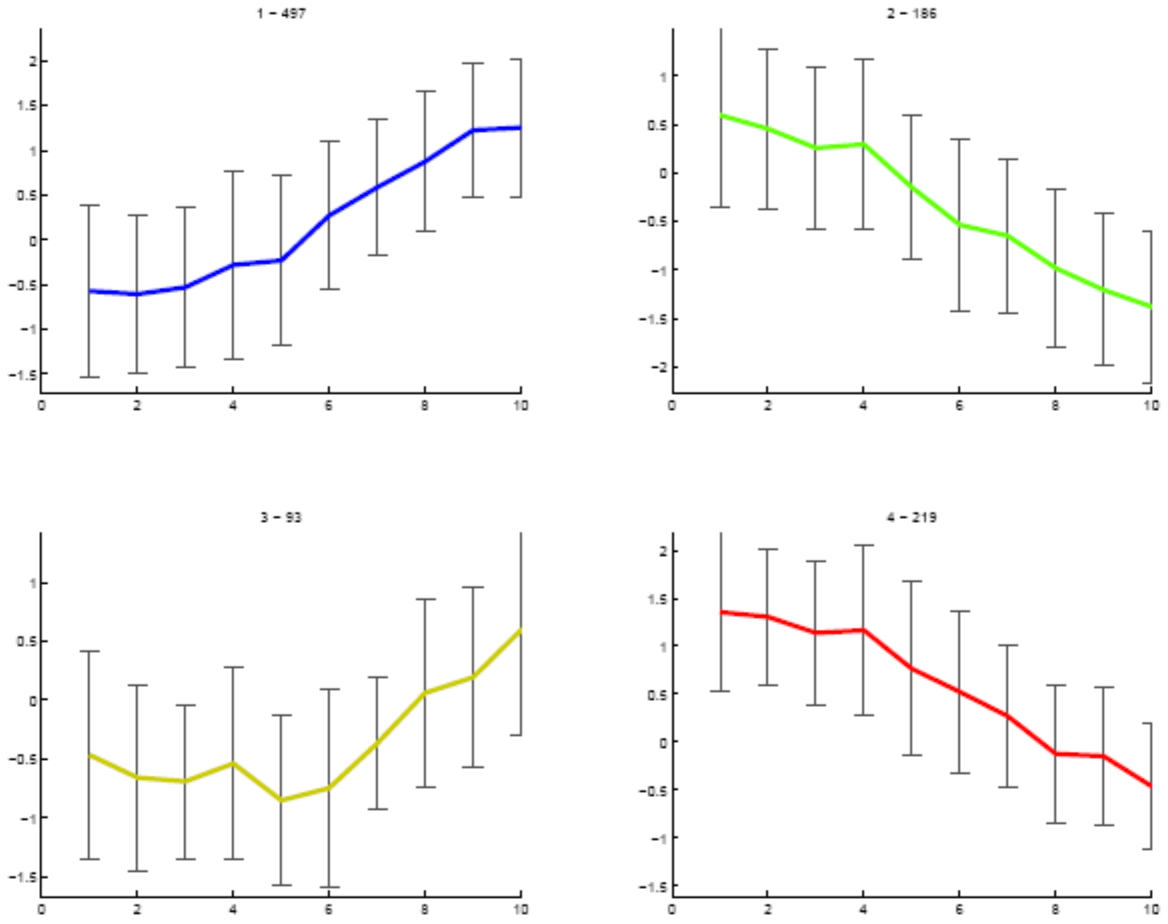


NEC dendrogram



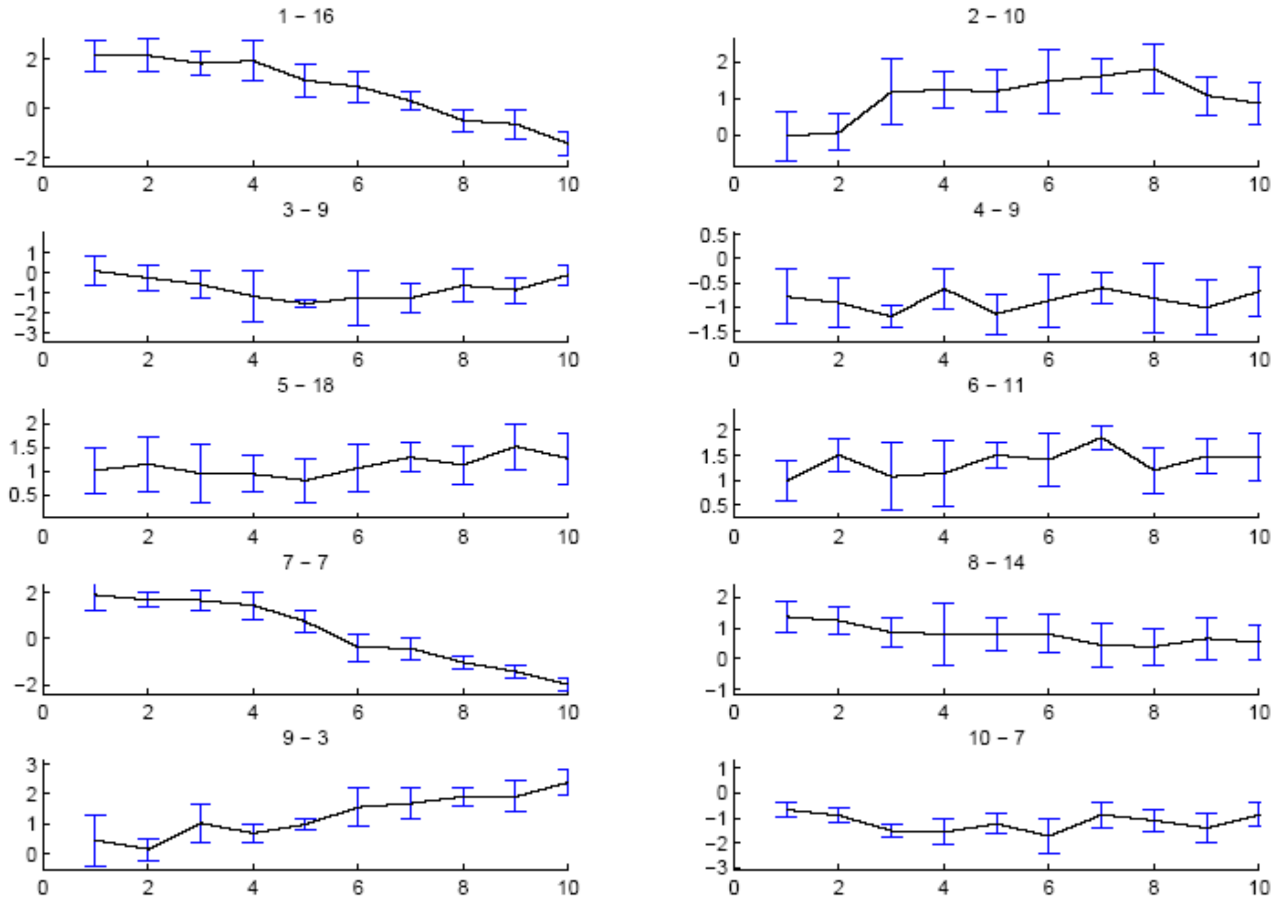
MDS Labeling

NEC dendrogram



The archetype behavior with corresponding standard deviation for the for particularly significant clusters found in our analysis

NEC dendrogram



The archetype behavior with corresponding standard deviation for the 10 outliers clusters found in our analysis.

Model selection for ensemble dispersion

- This work aims at introducing an approach to analyze the independence between different model data in a multi-model ensemble context
- The models are operational long-range transport and dispersion models
  - real-time simulation of pollutant dispersion or the accidental release of radioactive nuclides in the atmosphere
- An approach based on the hierarchical agglomeration of distributions of predicted radionuclide concentrations is proposed
  - two different similarity measures: Negentropy information and Kullback-Leibler divergence
- These approaches are used to analyze the data obtained during the ETEX-1 exercise
- The approach select subsets of independent models, whose performance are comparable to those from the whole ensemble

Collaboration with the Institute for Environment and Sustainability European Commission, Joint Research Centre TP 441, 21020 Ispra - Varese (Italy)

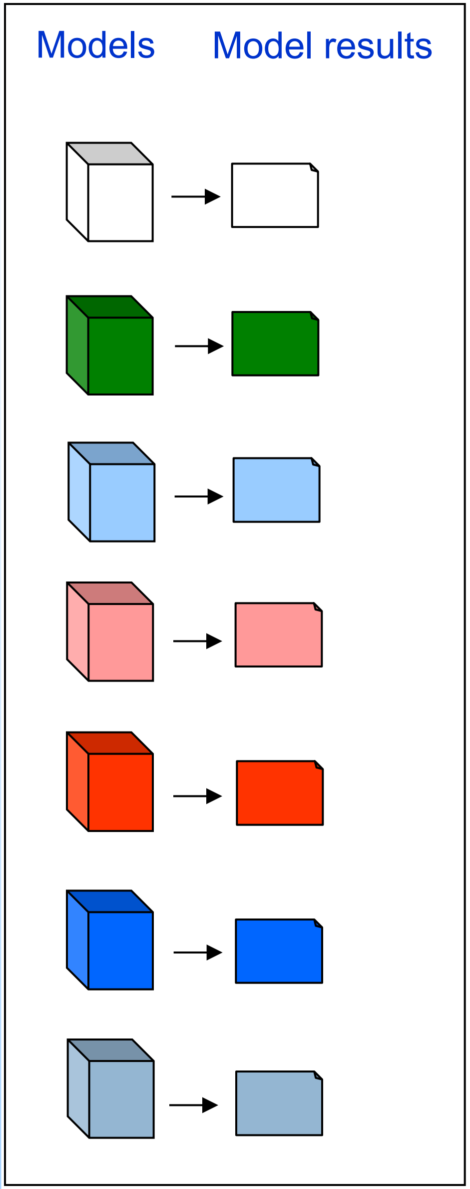
Emergency response

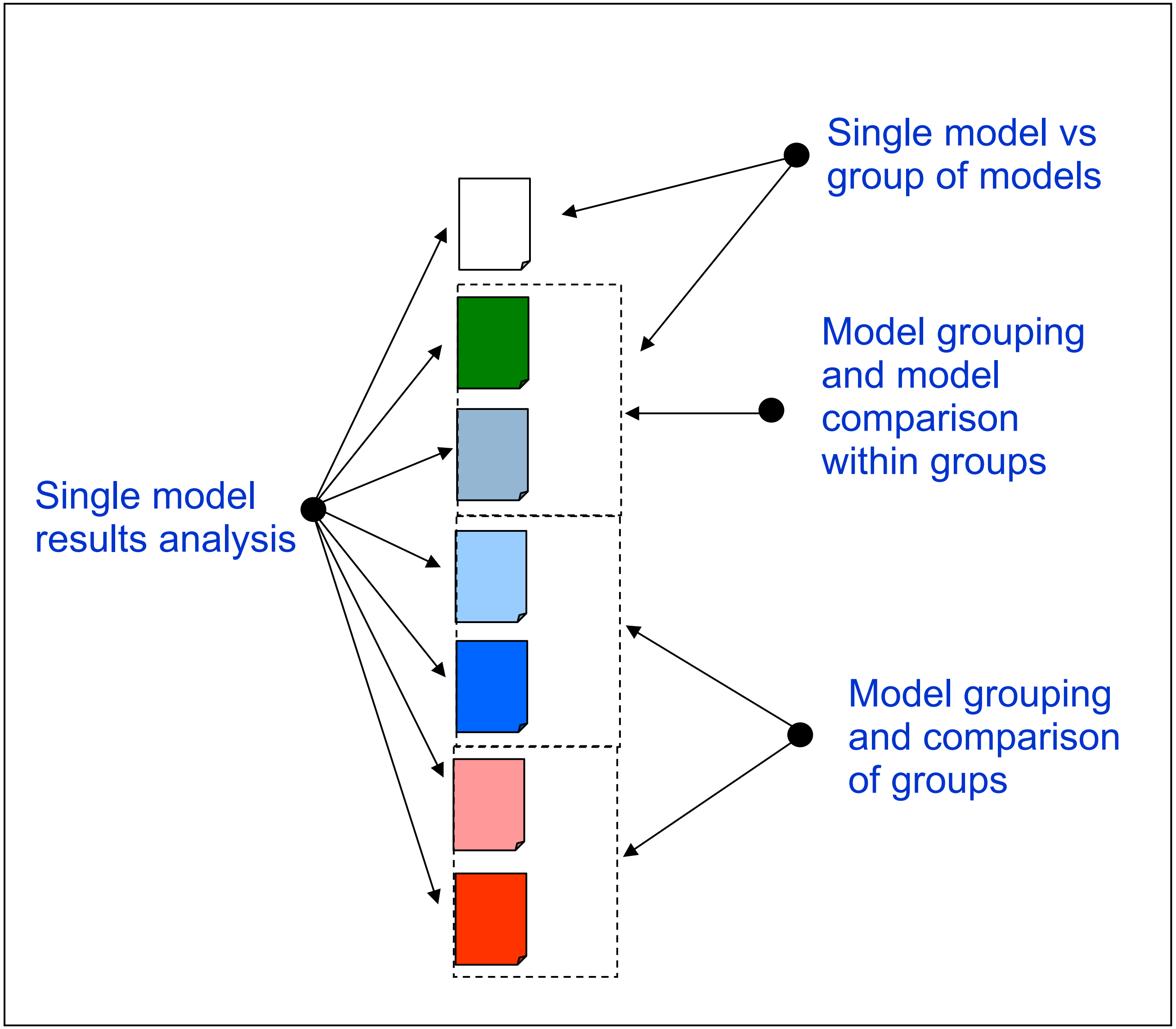
- Accidental release of radioactive material from NPP or other (i.e. Chernobyl, Algeciras, Hemel Hempstead)

- Atmospheric transport and dispersion over the continent, potential for trans-boundary character
  - Where is it?
  - When is getting there?
  - How much should I expect?

- At national level, met services and environmental protection agencies use long range transport and dispersion models to forecast concentration and deposition

Ensemble treatment - concept





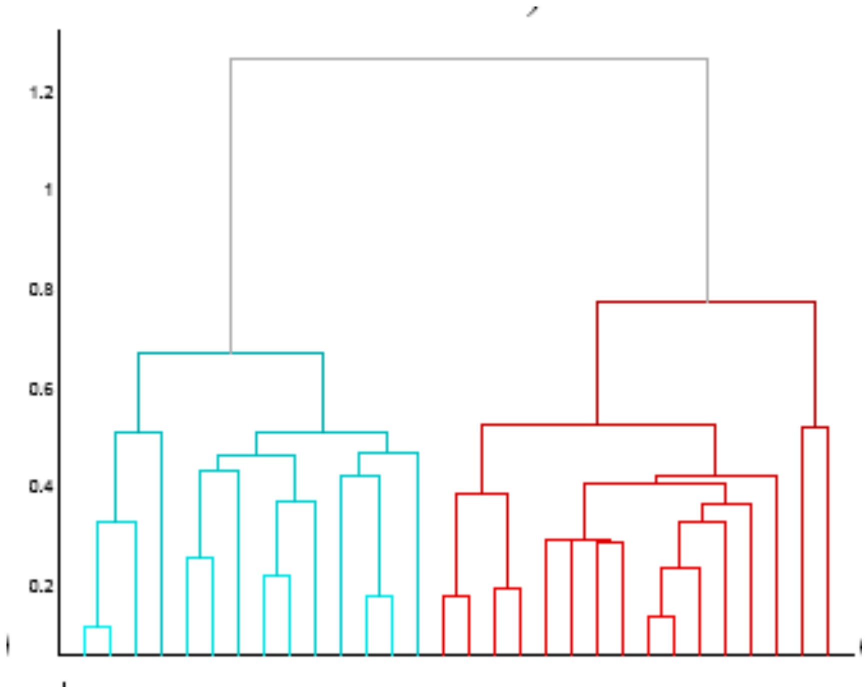
Ensemble and independence

- The Median Model was shown to outperform the results of any single deterministic model in reproducing the concentration of atmospheric pollutants
  - during the ETEX experiment

- A well-known statistical approach has been applied to multimodel data analysis
  - Bayesian Model Averaging (BMA)
  - similarities and differences between models were explored by means of correlation analysis

- Note - If different models are used to simulate the same phenomenon they probably will give similar responses
  - model ensemble results may lead to erroneous interpretations
  - this is more probable if models are strongly dependent

Model selection



Dendogram

$$KL(\mathbf{p}||\mathbf{q}) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right)$$

$$KL = \frac{KL(\mathbf{p}||\mathbf{q}) + KL(\mathbf{q}||\mathbf{p})}{2}$$

Kullback-Leibler

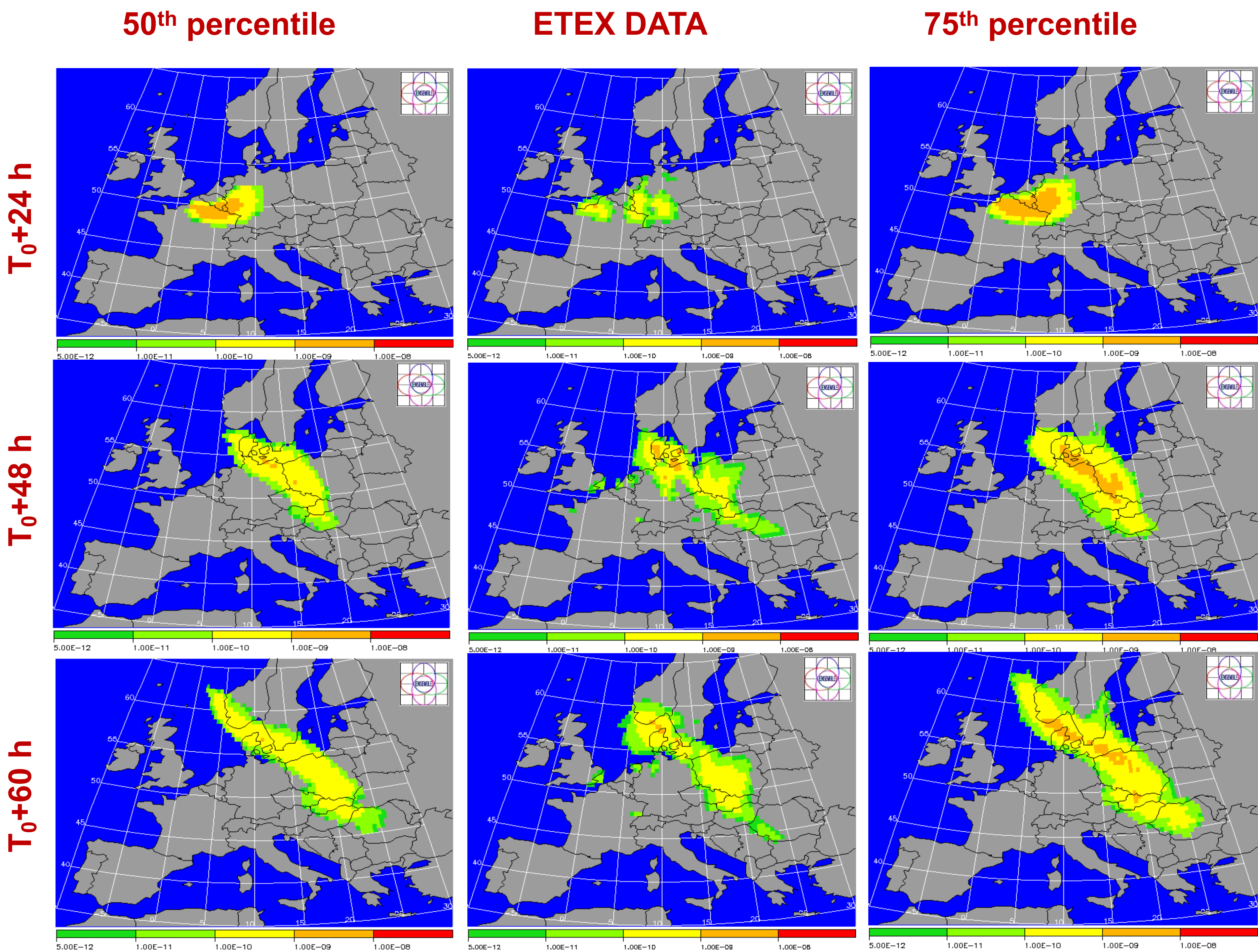
$$J_N(\mathbf{x}) = H(\mathbf{x}_{Gauss}) - H(\mathbf{x})$$

Negentropy

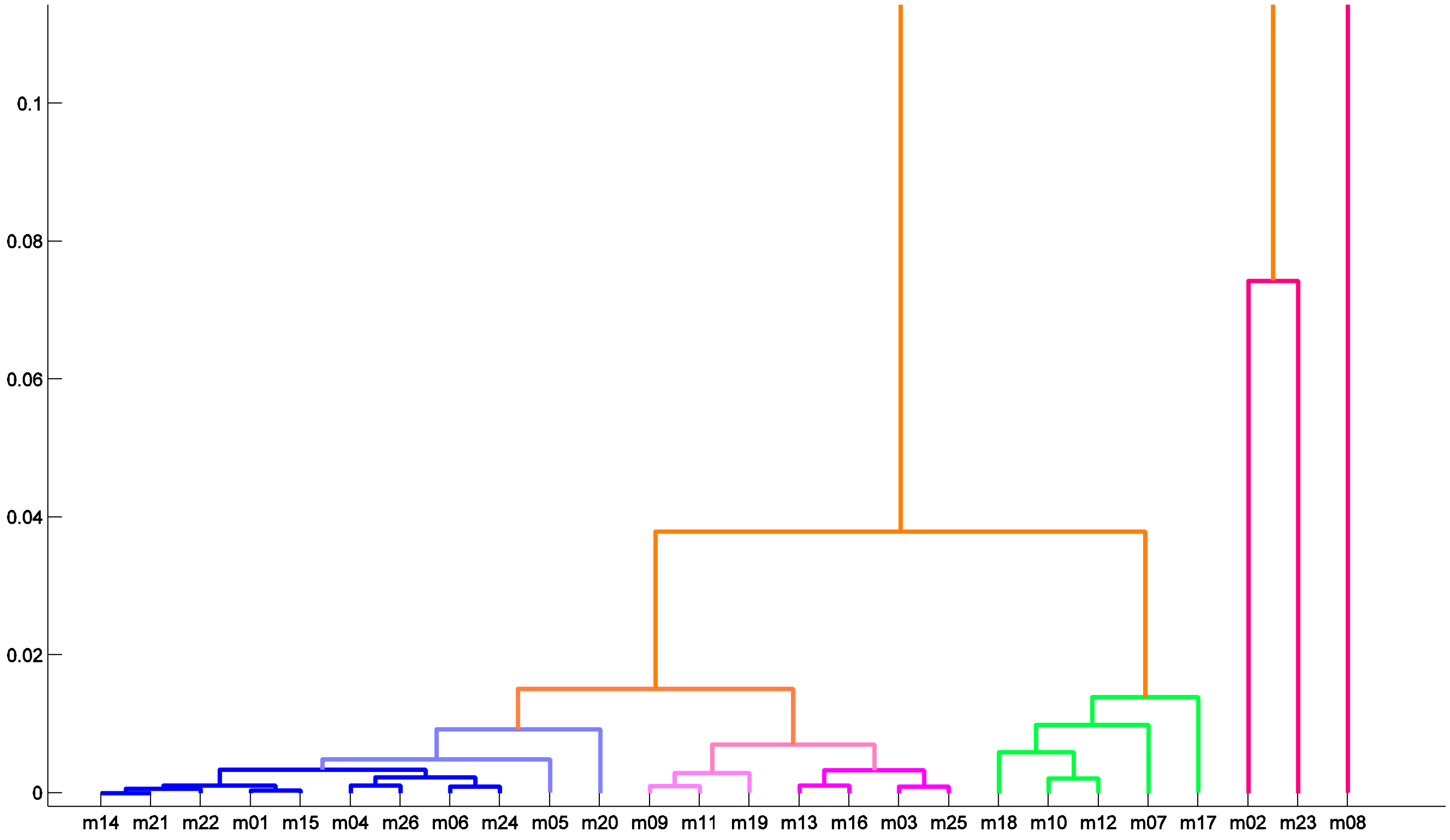
$$J_N(\mathbf{x}) \propto k_1 E\{G^1(\mathbf{x})\}^2 + k_2(E\{G^2(\mathbf{x})\} - E\{G^2(v)\})^2$$

ETEX-1

- The ETEX-1 experiment concerned the release of pseudo-radioactive material on 23 October 1994 at 16:00 UTC from Monterfil, southeast of Rennes (France)
  - A steady westerly flow of unstable air masses was present over central Europe
  - Such conditions persisted for the 90 h that followed the release with frequent precipitation events over the advection area and a slow movement toward the North Sea region
- Several independent groups worldwide tried to forecast these observations
  - Each simulation, and therefore each ensemble member, is produced with different atmospheric dispersion models and is based on weather fields generated by (most of the time) different Global Circulation Models (GCM)
  - All the simulations relate to the same release conditions

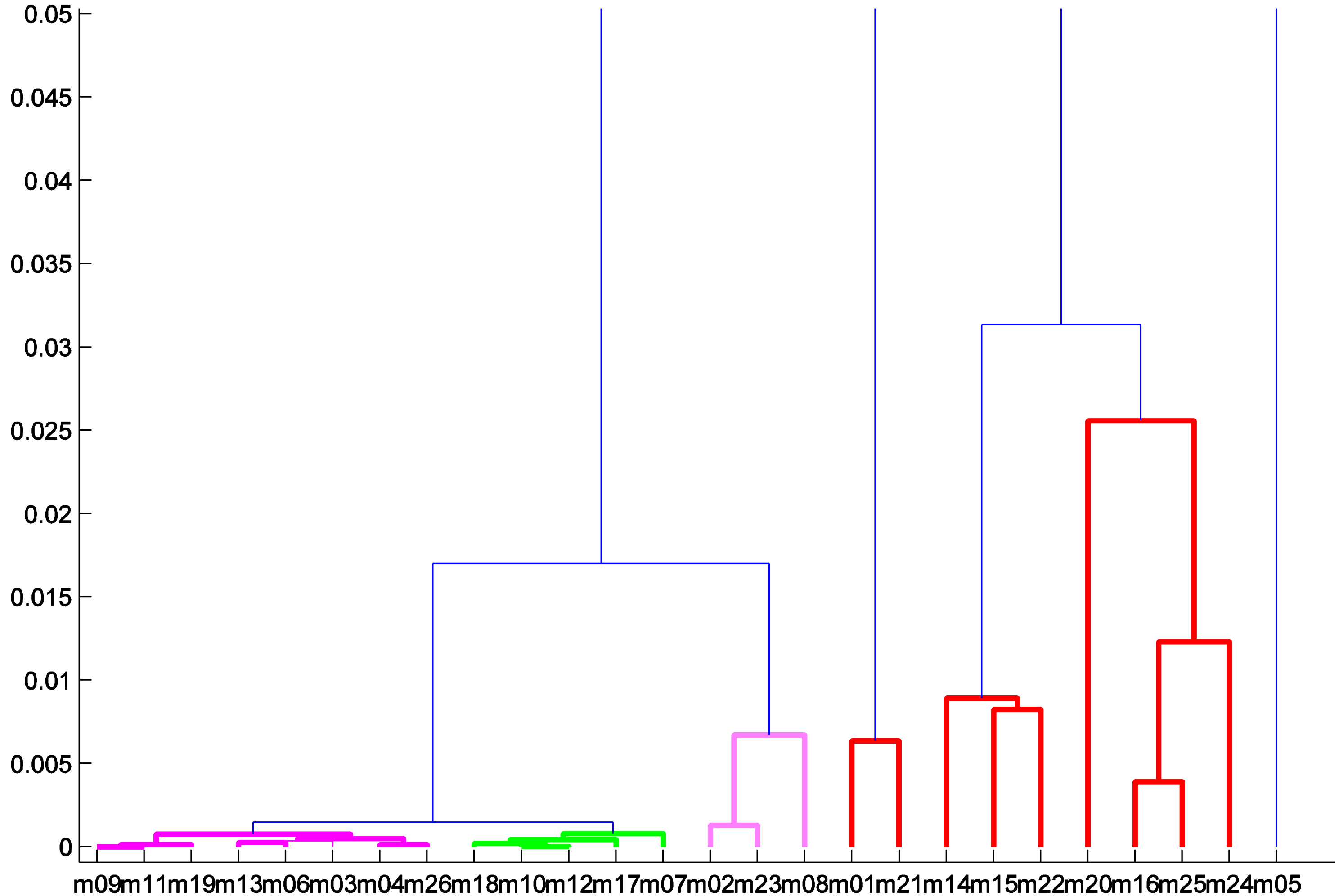

50th percentile
ETEX DATA
75th percentile
T0+24 h
T0+48 h
T0+60 h
5.00E-12
1.00E-11
1.00E-10
1.00E-09
1.00E-08

ETEX-1 results



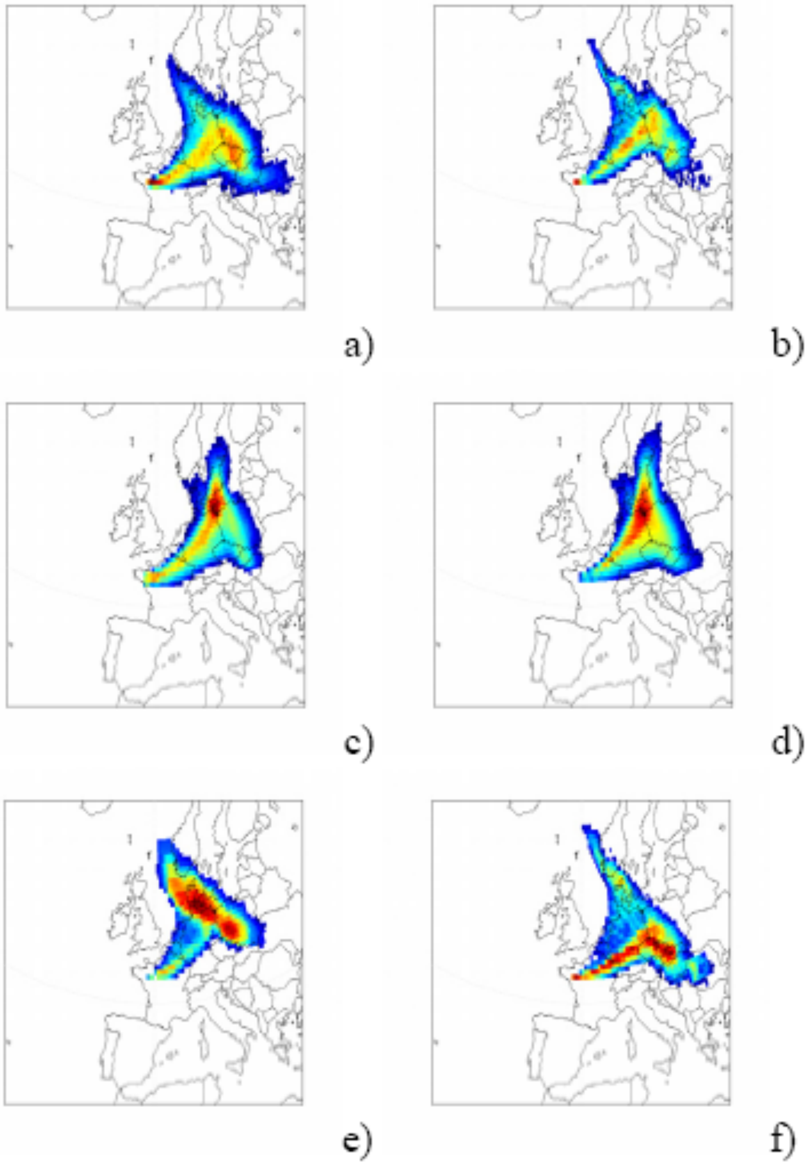
Negentropy dendrogram

ETEX-1 results



Kl divergence dendrogram

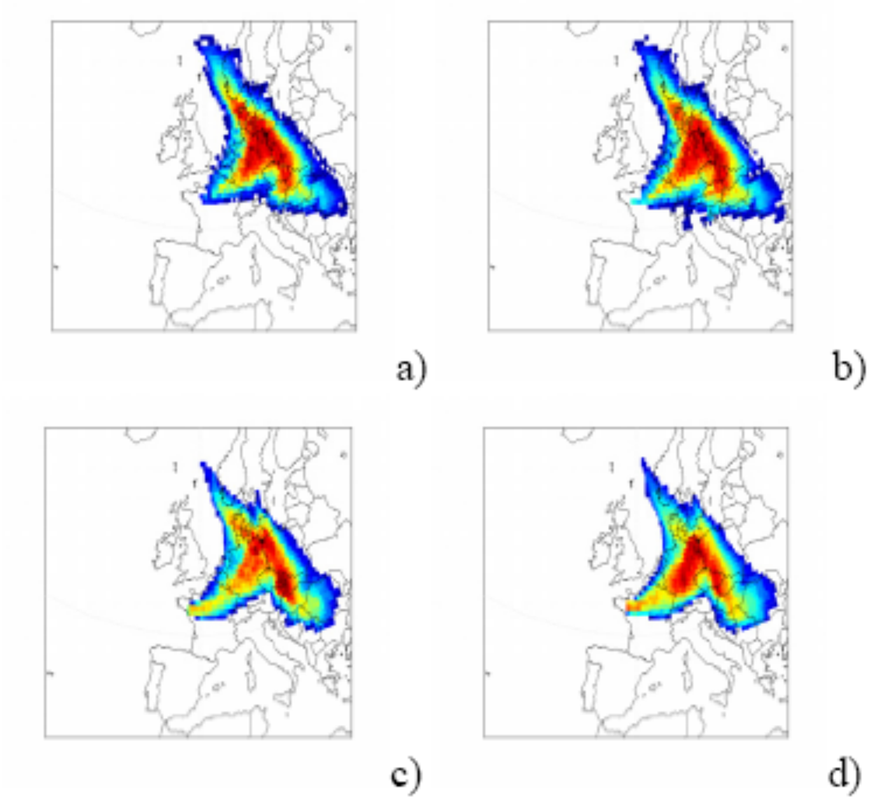
ETEX-1 results



blue cluster in the Negentropy dendrogram.
The visualized models are m_{14}, m_{21}, m_{01}, m_{15}, m_{06} and m_{24}, respectively

Distributions after 78 hours of the models m_{14} (a), m_{21} (b), m_{01} (c), m_{15} (d), m_{06} (e), m_{24} (f)

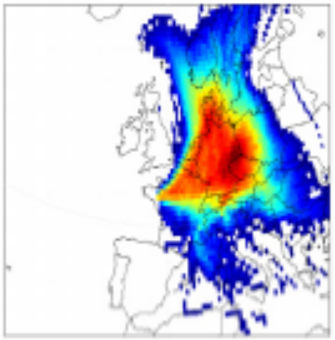
ETEX-1 results



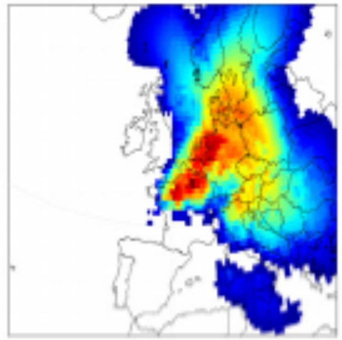
Distributions after 78 hours of the models m_{09} (a), m_{11} (b), m_{13} (c), m_{16} (d)

Distributions in the magenta cluster of the Negentropy dendrogram. In this case the models are m_{09}, m_{11}, m_{13} and m_{16}

ETEX-1 results

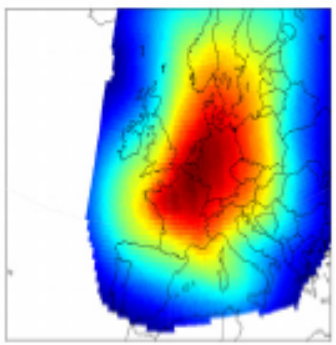


a)



b)

Distributions after 78 hours of the models m_{02} (a), m_{23} (b) and m_{08} (d)

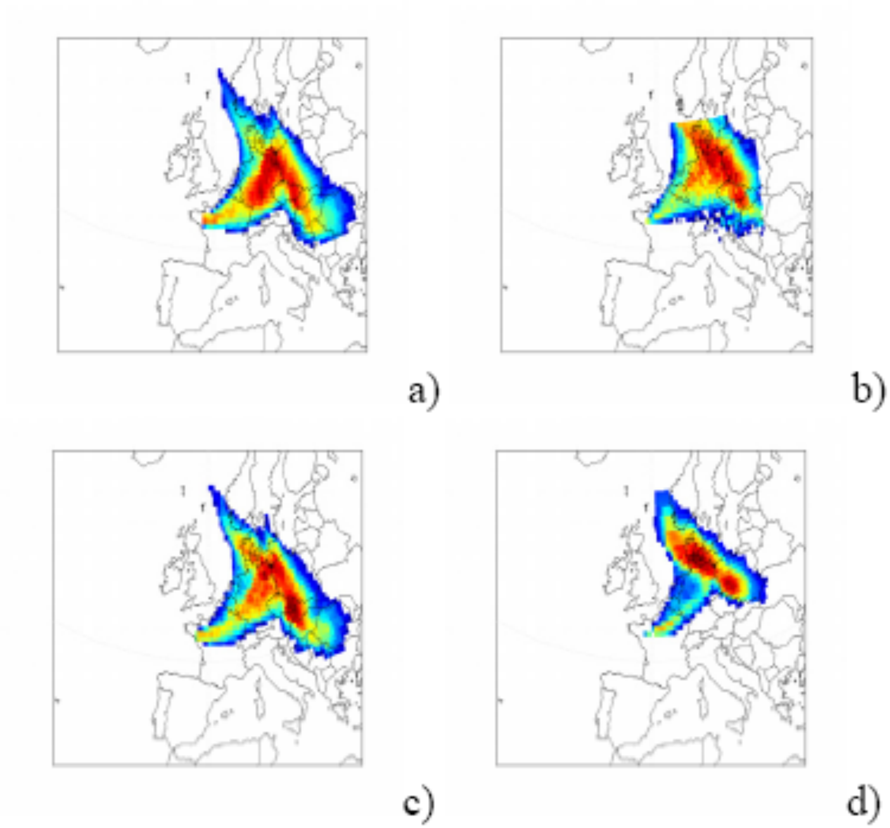


c)

Distributions in the Negentropy dendrogram. In this case the models are m_{02}, m_{23}, m_{08}

Note: the same three models are agglomerated together in the KL dendrogram but they belong to another extended cluster

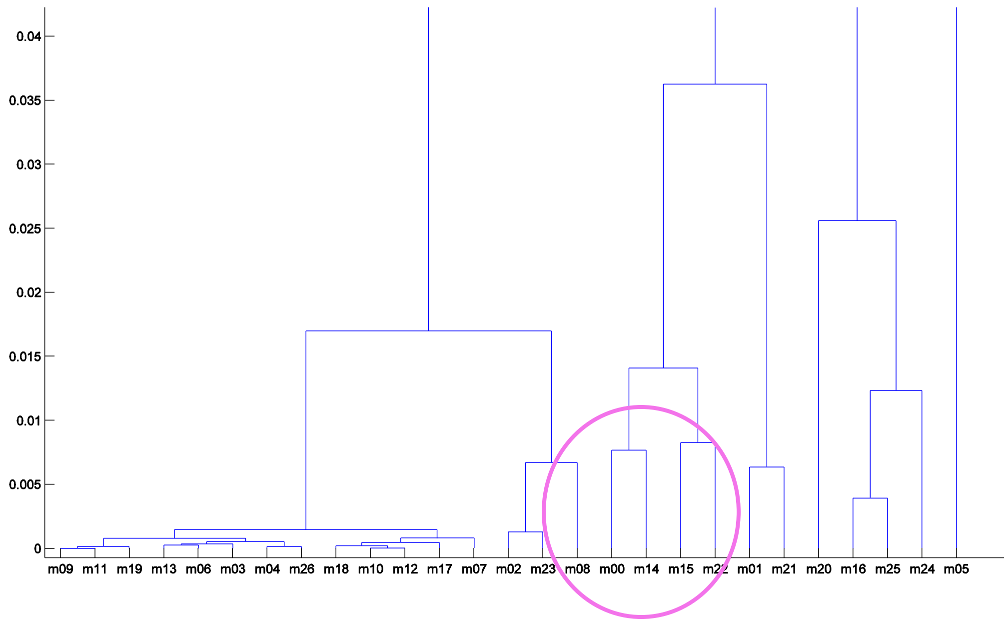
ETEX-1 results

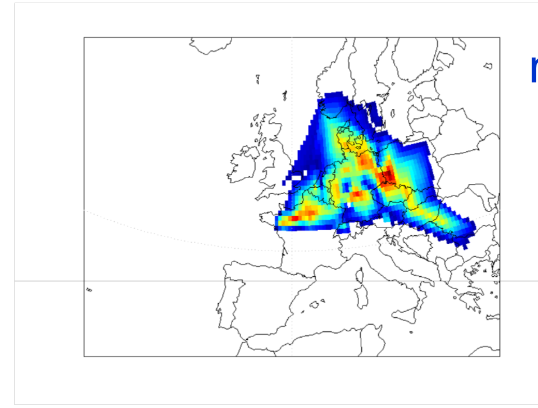


Distributions after 78 hours of the models m_{16} (a), m_{25} (b), m_{13} (c) and m_{06}

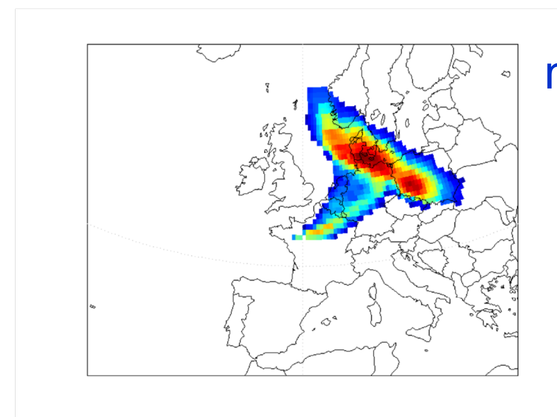
In the KL dendrogram the model m_{16} is associated together with the model m_{25} but we can note that its distribution is closer to that of model m_{13} than m_{25}.
In the Negentropy based dendrogram models m_{13} and m_{16} are agglomerated.
Moreover model m_{13} in the KL dendrogram is agglomerated with model m_{06}, but, they have a rather different distribution

ETEX-1 results





m_{00}



m_{14}

In the KL dendrogram it is associated only with model m_{14}