Advanced data, signal and image processing tools for biomedical and medical applications

Ali Mohammad-Djafari

Groupe Problèmes Inverses
Laboratoire des Signaux et Systèmes
UMR 8506 CNRS - SUPELEC - Univ Paris Sud 11
Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette, FRANCE.

djafari@lss.supelec.fr
http://djafari.free.fr
http://www.lss.supelec.fr

SPIS2015, December 16-17, 2015, Amirkabir University of Technology (AUT)

Summary 1

- Data, signals, images in Biological and medical applications
  - Individual cells, Population of cells, Small animals, Human
  - In vitro and In Vivo
- A great number of data, variables, time series, signals, images, ...
  - Genes expression, Hormones, temperature, ECG, EMG, ...
  - Tomographic images (X rays, PET, SPECT, IRM),
  - 3D body volume, fMRI, Holographic, multi- and Hyper-spectral images, ...
- Need for Visualization tools
  - multicomponent, multivariate and multidimensional
  - Time domain
  - Transformed domain: Fourier, Wavelets, Time-Frequency...
  - Scatter plots, histograms, statistics, ...

Summary 2

- Modeling time series
  - Parametric:
    Superposition of sinusoids, Gaussians shapes, ...
  - Non Parametric:
    Fourier, Wavelets, Time-frequency, time-scale,...
  - Probabilistic:
    Moving Average (MA), Autoregressive (AR), ARMA, Markovian models, ...
- Modeling images
  - Simple Markovian models (intensity, texture,...)
  - Hierarchical Markovian models with hidden variables of contours and regions
- Modeling the relation between observed data and unknowns
  - Linear / Non linear
  - Training and test data

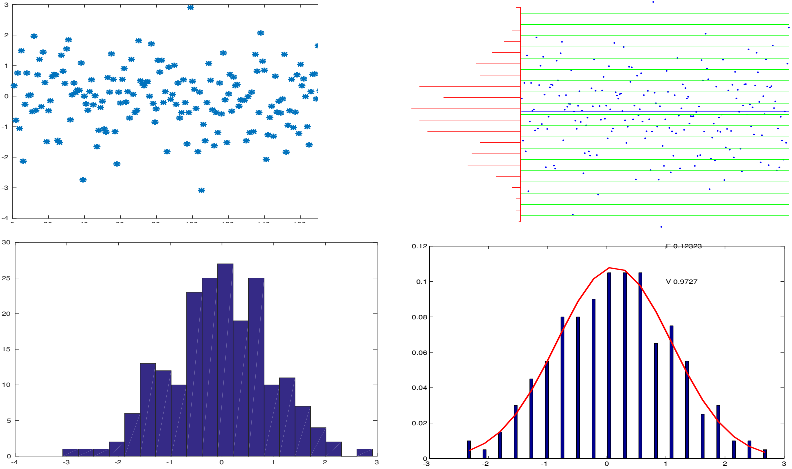
Summary 3

- Simple Analysis:
  Estimating periods, Computing harmonic components, spectra, , ...
- Multicomponent/Multivariate data analysis:
  Dimensional Reduction
  PCA, FA, ICA, Sparse PCA for dimensional reduction and main factors extraction
- Multicomponent/Multivariate Discriminant Analysis with classification:
  LDA, EDA, RDA, Sparse LDA for finding the most discriminant factors
- Blind sources separation
- Correlation (Pearson or Spearman) computation and dependency graph visualization
- Modelling input-output relations

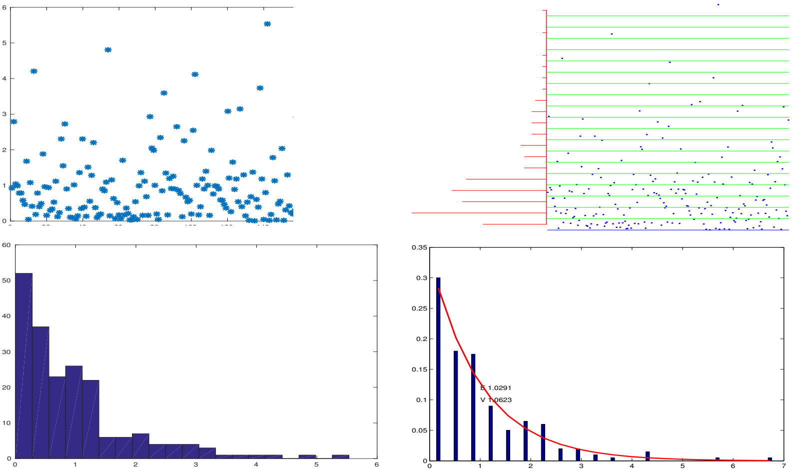
1D data: one variable

- Data: $x_i, i = 1, \cdots, M$
- 1D plot, mean, median, variance
- No order: exchangeable
- histogram, probability distribution,
- Statistical modelling: expected value, variance, mode, median, Higher order moments, entropy
- Parametric, semi-parametric and Non Parametric modelling
- Parameter estimation: MM, ML, Bayesian
- Model selection: AIC, BIC, ...

1D data (Gaussian)



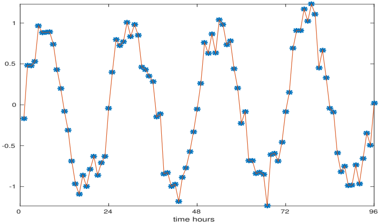
1D data (Gamma)



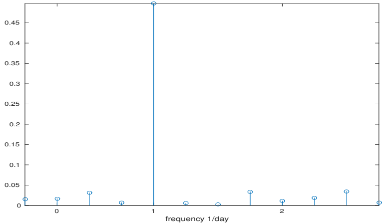
1D signals: Time series

- 1D Signal: Time series: $x_i = f(t_i)$
- In general no exchangeable.
- Time representation $f(t)$, Fourier Transform and Fourier representation $F(\nu)$, Auto Correlation function $R(\tau)$, Power Spectral Density $S(\nu)$
- Stationary and non stationary
- STFT, Time-Frequency, Time-Scale, Wavelets, ...
- Smoothing, Noise removing, Filtering
- Periodic signals, estimation of the period, Fourier series
- Modeling:
  - Sum of sinusoids model and parameter estimation
  - Moving average (MA) model
  - Autoregressive (AR) model

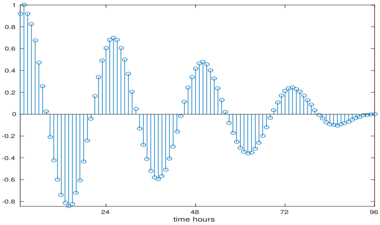
1D signals $f(t)$ and its FT $F(\nu)$



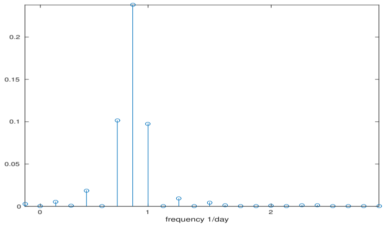
signal $f(t)$



Modulus of its FT $|F(\nu)|^2$



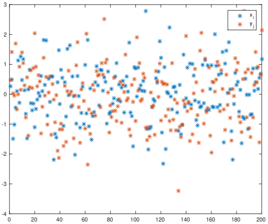
Auto-Correlation function $R(\tau)$

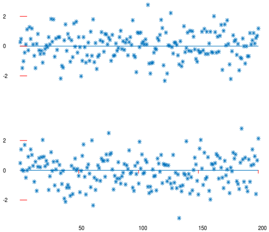


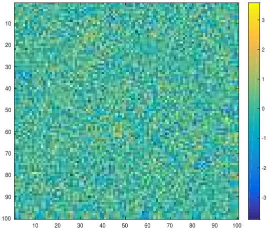
Spectral density $S(\nu)$

Multi-component, multi-variate, multi-dimensional data

- bi-component: $\{(x_i, y_i)\}, i = 1, \cdots, M\}$, $2M$ elements
- bi-variate: $\{x_i, i = 1, \cdots, M\}, \{y_j, j = 1, \cdots, N\}$, $M + N$ elements
- bi-dimensional: Images: $x_{i,j}, i = 1, \cdots, M, j = 1, \cdots, N$, $M * N$ elements

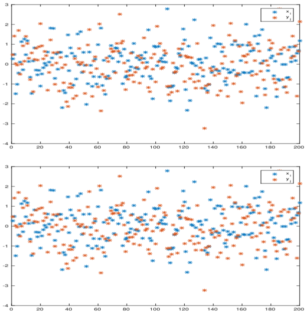


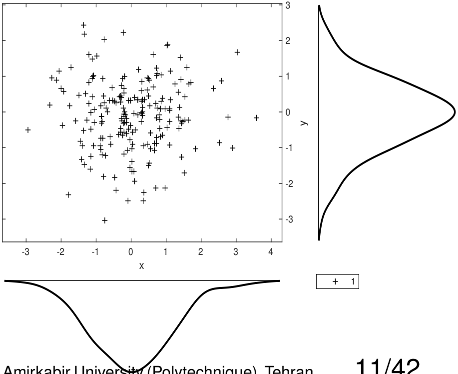




Bi-component or Bi-variate data

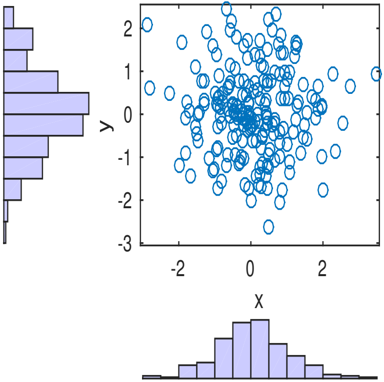
- 2D distribution: joint probability distribution $p(x, y)$
- Conditionals $p(x|y), p(y|x)$
- Marginal distributions $p(x)$, $p(y)$
- Expected values $E(X), E(Y)$, variances $V(X), V(Y)$, and Covariances, Higher order moments, entropy
- Independence tests
- Copula, ...

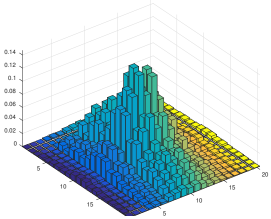


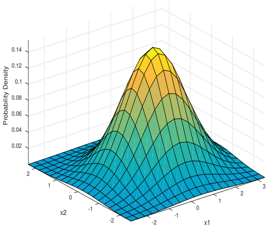


Bivariate data

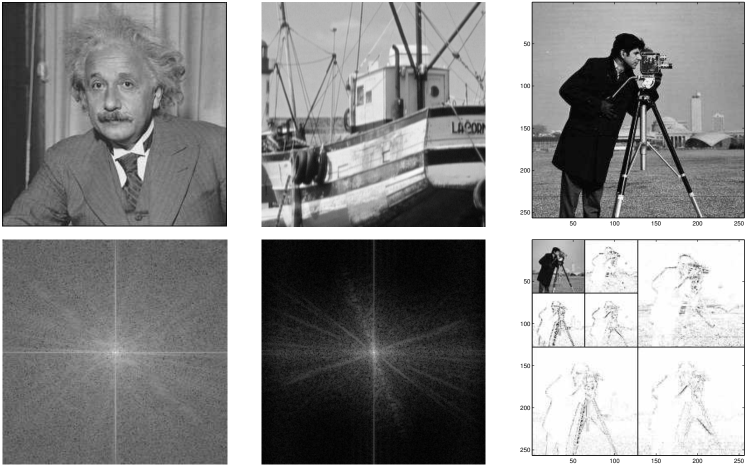
- Joint, marginals and conditional probability density functions



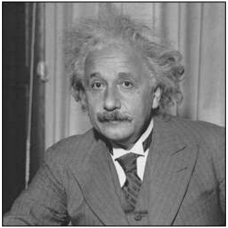




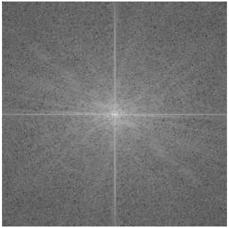
Images: Space, Fourier and Wavelets representations



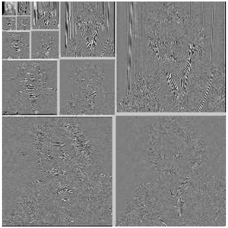
Sparse images (Fourier and Wavelets domain)



Image



Fourier



Wavelets

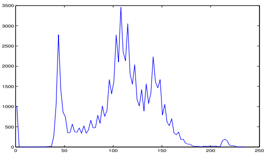
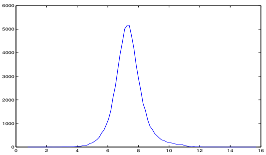
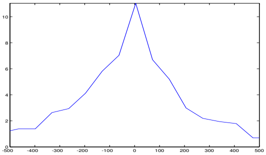


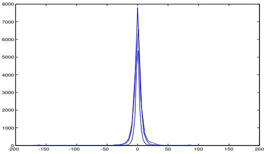
Image hist.



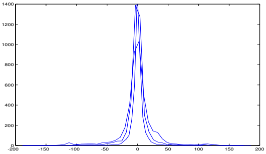
Fourier coeff. hist.



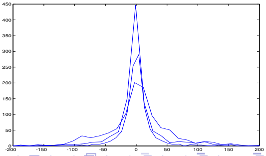
Wavelet coeff. hist.



bands 1-3



bands 4-6

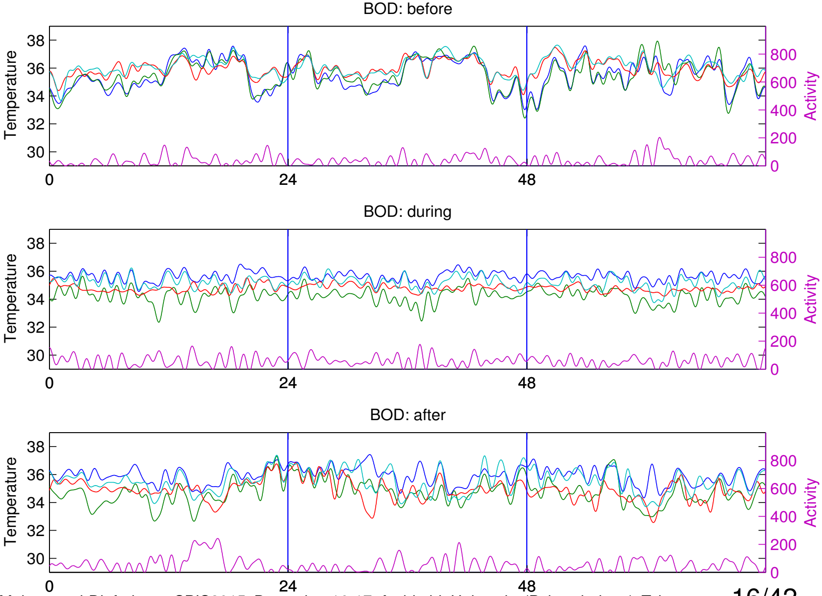


bands 7-9

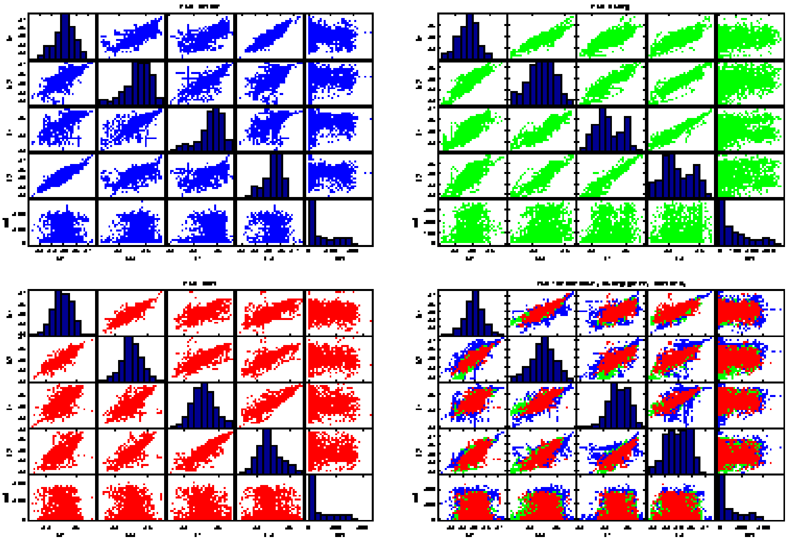
Multi-variate data and signals

- Multi-variate data:
  - Scatterplots,
  - Correlation coefficients (Pearson and Spearman)
  - Multivariate probability distribution,
  - means, variances, covariances,
  - Joint, marginals and conditional probability density functions
- Multi-variate signals:
  - Multi-variate time series,
  - Auto and Inter Correlation functions,
  - Auto and Inter Power Spectral Densities
- Needs for advances visualization, Dimensionality Reduction, Factorial analysis, modelling, parameter estimation, classification, ... and Knowledge extraction

Temperature and activity Time series before, during and some treatment



Temperatures, before, during and after changes



Simple questions for 1D time series

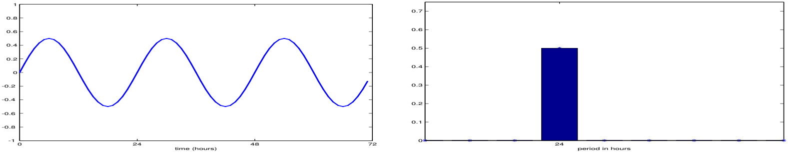
- Main question: Has something changed during and some medical action?
- In this study: effects of circadian cycle on cancer cells.

1. Is there any periodic components in these signals? Yes/No (Detection)? Confidence ?
2. If Yes, How many?
3. What are those components (Periods p_i or Frequencies ν_i , Amplitudes a_i)?

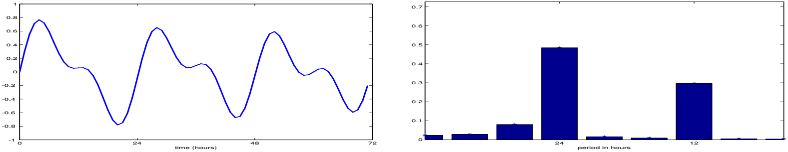
- When questions 1 and 2 are answered, the problem becomes easier: Parameter estimation
- Trying to answer all the three questions at the same time: semi- or Non-Parametric modelling
- Biologists always need uncertainties $\longrightarrow$ Bayesian inference

Simple Analysis tools may not be successful even in very simple cases

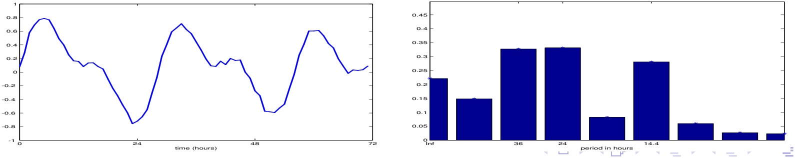
Case of 1 sinusoid



Case of 2 sinusoids+noise



Case of 3 sinusoids+noise



Classical methods: Spectral estimation $S(\omega)$?

- Fast Fourier Transform (FFT):
  $g(t) \longrightarrow FFT \longrightarrow f(\omega) \longrightarrow S(\omega) = |f(\omega)|^2$
  - Advantages: Well-known and understood, fast
  - Drawbacks: linear in frequencies ν,
    but not equidistance in periods
    $\nu = [0, \cdots, N-1] \longrightarrow p = [\infty, 1, \cdots, 1/(N-1)]$
- Autocorrelation function: $\gamma(\tau)$
  - If $g(t)$ is periodic, then $\gamma(\tau)$ is also periodic,
    but much smoother
  - $\gamma(0) = 1 \; \gamma(\tau) \leq \gamma(0), \forall \tau$
- Power spectral density: $\gamma(\tau) \longrightarrow \text{FFT} \longrightarrow S(\omega)$
- Autoregressive (AR), Moving Average (MA) and ARMA models
- Non-stationary GARCH models
- Sum of sinusoidal components

Parametric, Semi- and Non-Parametric models

- Parametric:

$$g(t) = \sum_{k=1}^{K} a_k \sin(2\pi\nu_k t + \phi_k) + \epsilon(t), \quad \boldsymbol{\theta} = \{a_k, \phi_k, \nu_k\}$$

$$g(t) = \sum_{k=1}^{K} a_k \cos(2\pi\nu_k t) + b_k \sin(2\pi\nu_k t) + \epsilon(t), \; \boldsymbol{\theta} = \{a_k, b_k, \nu_k\}$$

$$g(t) = \sum_{k=1}^{K} c_k \exp\left[j2\pi\nu_k t\right] + \epsilon(t), \quad \boldsymbol{\theta} = \{c_k, \nu_k\}, t = 0, \cdots, T$$

- Semi-Parametric: $\nu_k = k\nu_0, \nu_0 = 1/T, K = T \longrightarrow$ DFT
- Non-Parametric: ν_k fixed in a given interval with given precision, so K is fixed but can be as large as necessary.

$$g(t) = \sum_{k=1}^{K} c_k \exp\left[j2\pi\nu_k t\right] + \epsilon(t), \; \boldsymbol{\theta} = \{c_k\} \text{ Linear model}$$

Can we propose a unifying approach for all these problems?

My answer is Yes:

- Identify what your are looking for. (red color $\boldsymbol{f}$)
- Identify what are the data : (blue color $\boldsymbol{g}$)
- Consider the errors (modeling and measurement $\boldsymbol{\epsilon}$)
- Write the Forward model relating them: $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$
- Write the expression of the likelihood $p(\boldsymbol{g}|\boldsymbol{f})$
- Translate your prior knowledge on the unknowns in $p(\boldsymbol{f})$
- Use the Bayes rule:

$$p(\boldsymbol{f}|\boldsymbol{g}) = \frac{p(\boldsymbol{g}|\boldsymbol{f})\, p(\boldsymbol{f})}{p(\boldsymbol{g})} \propto p(\boldsymbol{g}|\boldsymbol{f})\, p(\boldsymbol{f})$$

- Infer on $\boldsymbol{f}$ using the posterior $p(\boldsymbol{f}|\boldsymbol{g})$:
  - Maximum A Posteriori (MAP): $\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\}$
  - Posterior Mean (PM): $\widehat{\boldsymbol{f}} = \int \boldsymbol{f}\, p(\boldsymbol{f}|\boldsymbol{g})\, \mathrm{d}\boldsymbol{f}$

Estimating Periodic Components: Inverse Problems Approach

$$g(t) = \sum_{k=1}^{K} c_k \exp\left[j2\pi\nu_k t\right] + \epsilon(t), \; \boldsymbol{\theta} = \{c_k\} \text{ Linear model}$$

Slight changes of notations: use of periods p_n in place of frequencies ν_k and f_n in place of c_k:

$$g(t) = \sum_{n=1}^{N} f_n \exp\left[j2\pi/p_m t\right] + \epsilon(t), \; t = m\Delta t, m = 1, \cdots, M$$

Defining the vectors: $\boldsymbol{g} = [g_1, \cdots, g_M]'$, $\boldsymbol{\epsilon} = [\epsilon_1, \cdots, \epsilon_M]'$, $\boldsymbol{f} = [f_1, \cdots, f_N]'$ and the matrix $\boldsymbol{H}$: $H_{m,n} = \exp\left[j2\pi/p_m m\Delta t\right]$, we obtain:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$

The objective is to infer on $\boldsymbol{f}$.

Inverse Problems Approach

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$$

Bayesian approach:

- Assign the Likelihood : $p(\boldsymbol{g}|\boldsymbol{f})$
- Assign the prior law: $p(\boldsymbol{f})$
- Use the Bayes rule : $p(\boldsymbol{f}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f})\, p(\boldsymbol{f})$
- MAP:

$$\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\}$$

- Assuming Gaussian noise and Gaussian prior

$$p(\boldsymbol{f}|\boldsymbol{g}) = \mathcal{N}(\boldsymbol{f}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\Sigma}}) \text{ with } \widehat{\boldsymbol{\Sigma}} = (\boldsymbol{H}'\boldsymbol{H} + \lambda\boldsymbol{I})^{-1} \text{ and } \widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{J(\boldsymbol{f})\}$$

$$J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{f}\|^2$$

- Other priors (Generalized Gaussian, Student-t or Cauchy)
  $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\Omega(\boldsymbol{f})$

Bayesian estimation with priors enforcing sparsity

- Sparsity: For any periodic signal, the spectrum is a set of Diracs
- Biological signals related to clock genes: a few independent oscillators
- Spectrum has a few non zero elements in any given interval

$$g(m\Delta t) = \sum_{n=1}^{N} f_n \exp\left[-j2\pi/p_m\, m\Delta t\right] + \epsilon(t),\; m = 1, \cdots, M$$

$\boldsymbol{g} = \boldsymbol{H} \boldsymbol{f} + \boldsymbol{\epsilon}$ with $\boldsymbol{f}$ sparse

- The question is now: How to translate sparsity?
- Two solutions: L1 regularization and Bayesian sparsity enforcing priors.
- Three main options in Bayesian: Genealized Gaussian, Student-t, mixtures models

Bayesian estimation with priors enforcing sparsity

- $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$ with $\boldsymbol{f}$ sparse
- To translate this information use the heavy tailed prior law Student-t with its hierarchical structure and hidden variables

$$\mathcal{S}t(f_j|\nu) \propto \exp\left[-\frac{\nu+1}{2}\log\left(1+f_j^2/\nu\right)\right]$$

- Infinite Gaussian Scaled Mixture (IGSM) property:

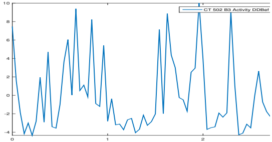
$$\mathcal{S}t(f_j|\nu) \propto= \int_0^\infty \mathcal{N}(f_j|,0,1/z_j)\,\mathcal{G}(z_j|\alpha,\beta)\,\mathrm{d}z_j, \quad \text{with } \alpha=\beta=\nu/2$$

- Hiearchical prior model:

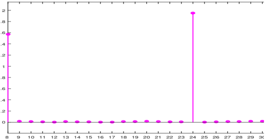
$$p(f_j|z_j) = \mathcal{N}(f_j|0,1/z_j), \quad p(z_j) = \mathcal{G}(z_j|\alpha,\beta)$$

$$\left\{\begin{array}{ll} p(\boldsymbol{f}|\boldsymbol{z}) & = \prod_j p(f_j|z_j) \\ p(\boldsymbol{z}) & = \prod_j p(z_j) \longrightarrow \\ p(\boldsymbol{g}|\boldsymbol{f}) & = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f},\sigma_\epsilon^2\boldsymbol{I}) \end{array}\right. \qquad p(\boldsymbol{f},\boldsymbol{z}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f})p(\boldsymbol{f}|\boldsymbol{z})p(\boldsymbol{z})$$

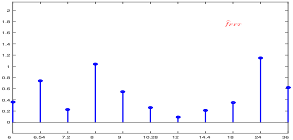
Results on simulated and real activity data



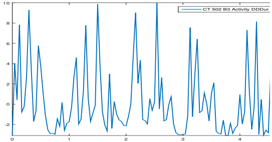
Data BEFORE



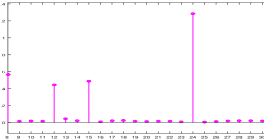
Proposed method



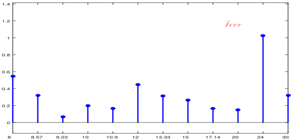
FFT



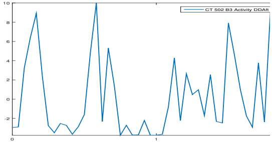
Data DURING



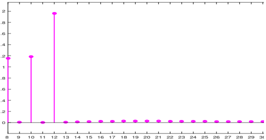
Proposed method



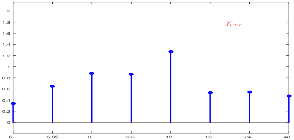
FFT



Data AFTER



Proposed method



FFT

Dimension reduction, PCA, Factor Analysis, ICA

- M variables $\boldsymbol{g}(t)$ are observed. They are redundant.
  Can we express them with $N \leq M$ factors $\boldsymbol{f}$?
  How many factors (Principal Components, Independent Components) can describe the observed data?

$$\begin{cases} g_i(t) = \sum_{j=1}^{N} a_{ij} f_j(t) + \epsilon_i(t) \\ \boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t) + \boldsymbol{\epsilon}(t) \end{cases} \quad \begin{cases} \boldsymbol{A} : (M \times N) \text{ Loading matrix }, N \leq M \\ \boldsymbol{f}\,(\mathrm{t}) \;:\; \textit{factors, sources} \end{cases}$$

- How to find both $\boldsymbol{A}$ and factors $\boldsymbol{f}(t)$?
- Bayesian methods:

$$(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{f}}) = \arg\max_{(\boldsymbol{A},\boldsymbol{f})} \{p(\boldsymbol{A}, \boldsymbol{f}|\boldsymbol{g})\} = \arg\min_{(\boldsymbol{A},\boldsymbol{f})} \{\ln p(\boldsymbol{g}|\boldsymbol{A}, \boldsymbol{f}) - \ln p(\boldsymbol{A}) - \ln p(\boldsymbol{f}$$

How to determine the number of factors

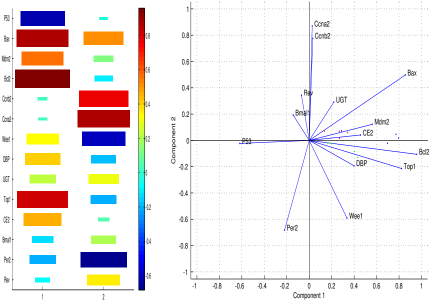
- When N is given:

$$p(\boldsymbol{A}, \boldsymbol{f}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{A}, \boldsymbol{f})\, p(\boldsymbol{A})\, p(\boldsymbol{f})$$

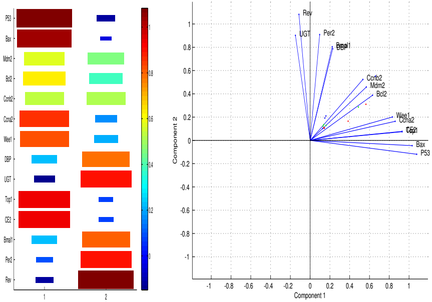
  Different choices for $p(\boldsymbol{A})$ and $p(\boldsymbol{f})$ and
  Different methods to estimate both $\boldsymbol{A}$ and $\boldsymbol{f}$:
  JMAP, EM, Variational Bayesian Approximation
- When N is not known:
  - Model selection
  - Bayesian or Maximum likelihood methods
  - To determine the number of factors we do the analyze with different N factors and use two criteria:
  - -log likelihood $-\ln p(\boldsymbol{g}|\boldsymbol{A}, N)$ of the observations and
  - DFE: Degrees of freedom error $(N-M)^2-(N+M))/2$ related to AIC or BIC model selection criteria.

Factor Analysis: 13 variables 2 factors

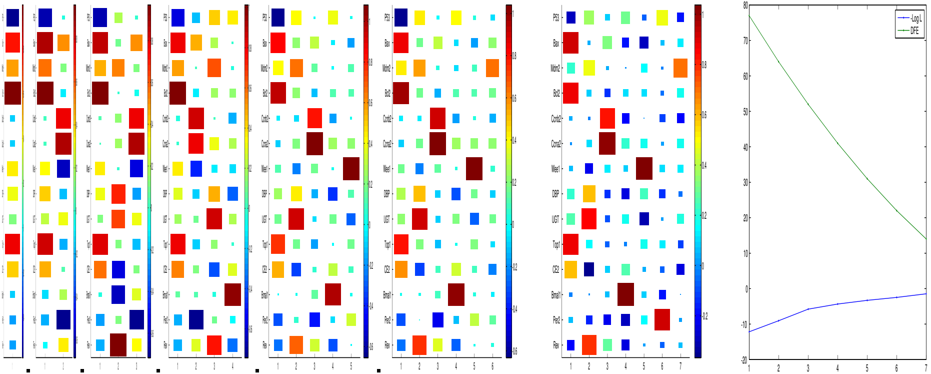
Time series



FT Amplitudes

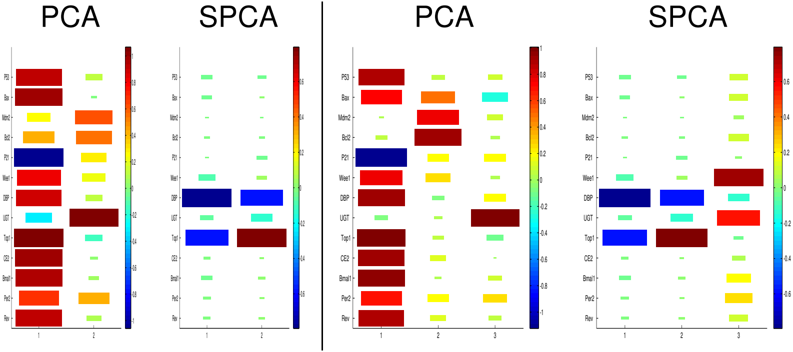


Factor Analysis: Time series, Number of factors



Sparse PCA

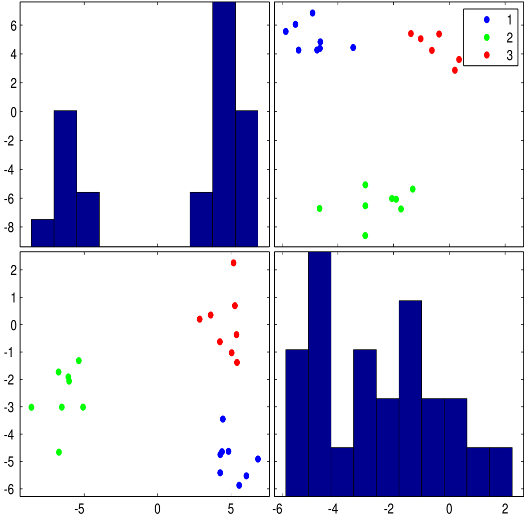
- In classical PCA, FA and ICA, one looks to obtain principal (uncorrelated or independent) components.
- In Sparse PCA or FA, one looks for the loading matrix $\boldsymbol{A}$ with sparsest components.
- This can be imposed via the prior $p(\boldsymbol{A})$. This leads to least variables selections.

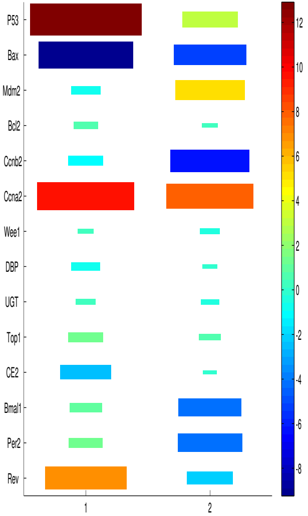


Discriminant Analysis

- When we have data and classes, the question to answer is:
  What are the most discriminant factors?
- There are many variants:
  - Linear Discriminant Analysis (LDA),
  - Quadratic Discriminant Analysis (QDA),
  - Exponential Discriminant Analysis (EDA),
  - Regularized LDA (RLDA), ...
- One can also ask for Sparsest Linear Discriminant factors (SLDA)
- Deterministic point of view (Geometrical distances)
- Probabilistic point of view (Mixture densities)
- Mixture of Gaussians models:
  Each classe is modelled by a Gaussian pdf

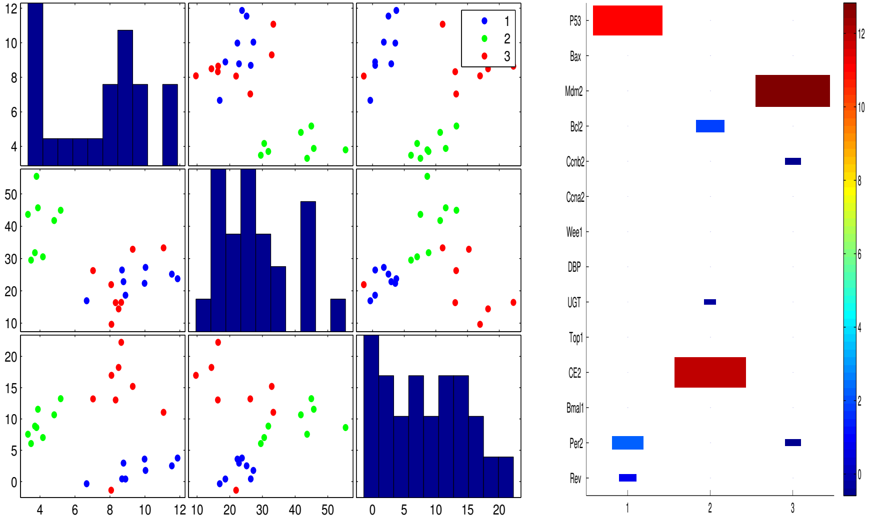
Discriminant Analysis: Time series, Colon



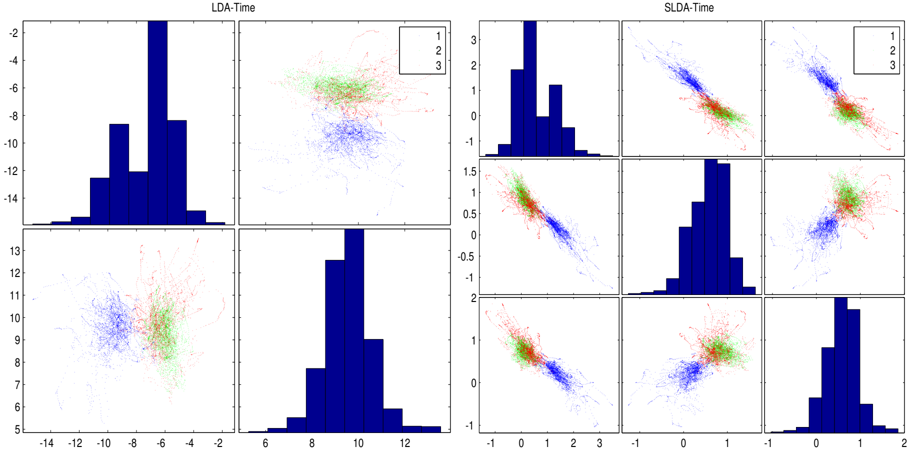


Sparse Discriminant Analysis: Time series, colon

What are the sparsest discriminant factors?



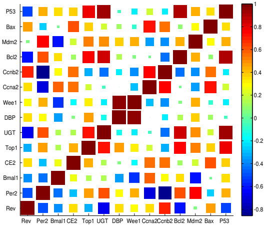
LDA and SLDA study on time serie: 1:before, 2:during, 3:



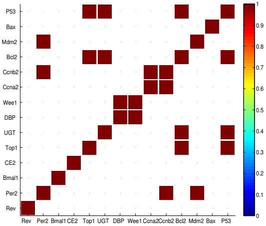
Dependancy graphs

- The main objective here is to show the dependencies between variables
- Three different measures can be used: Pearson ρ, Spearman ρ_s and Kendall τ
- In this study we used ρ_s
- A table of 2 by 2 mutual ρ_s are computed and used in different forms: Hinton, Adjacency table and Graphical network representation

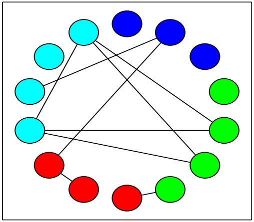
Hinton



Adjacency

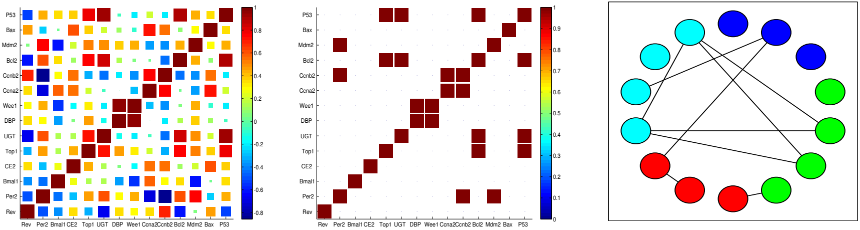


Network

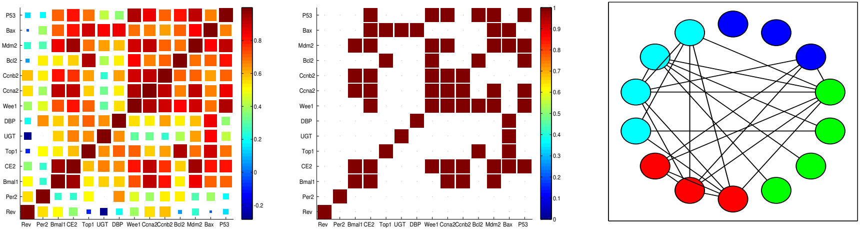


Graph of Dependancies: Colon, Class 1

Time series



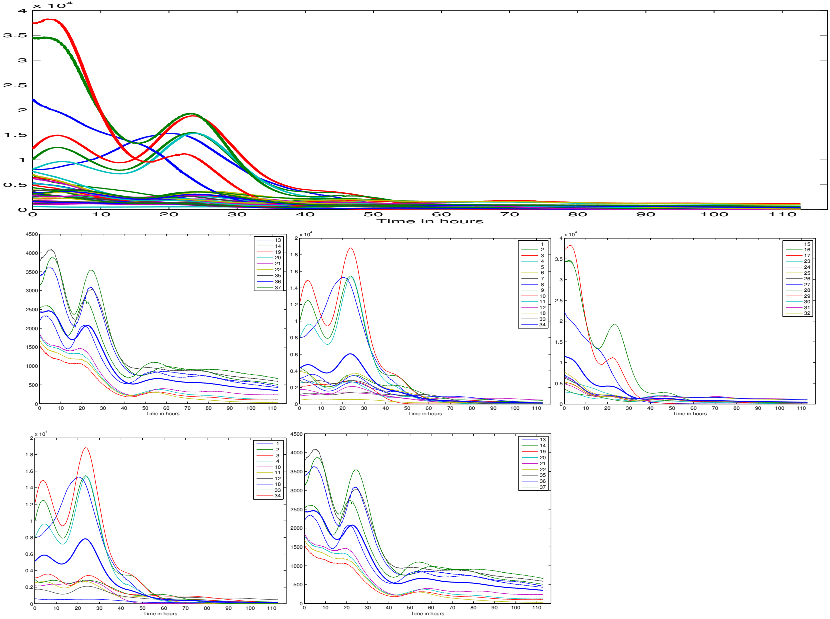
FT amplitudes



Classification tools

- Supervised classification
  - K nearest neighbors methods
  - Needs Training sets data
  - Must be careful to measure the performances of the classification on a different set of data (Test set)
- Unsupervised classification
  - Mixture models
  - Expectation-Maximization methods
  - Bayesian versions of EM
  - Bayesian Variational Approximation (VBA)

Classification tools



Input-Output modeling using training data and test data

- Linear models: $\boldsymbol{g}_k = \boldsymbol{A}\boldsymbol{f}_k + \epsilon_k, \quad k = 1, \cdot, K$
- Bayesian framework, MAP estimation with hyperparameter estimation

$$p(\boldsymbol{A}|\{\boldsymbol{g}_k, \boldsymbol{f}_k\}) \propto \prod_k p(\boldsymbol{g}_k|\boldsymbol{A}, \boldsymbol{f}_k) p(\boldsymbol{A})$$

- Gaussian priors for ϵ_k and for $\boldsymbol{A}$ and MAP solution: $\widehat{\boldsymbol{A}} = \arg\max_{\boldsymbol{A}} \{p(\boldsymbol{A}|\{\boldsymbol{g}_k, \boldsymbol{f}_k\})\}$

$$\widehat{\boldsymbol{A}} = \left(\sum_k \boldsymbol{g}_k \boldsymbol{f}_k'\right) \left(\sum_k \boldsymbol{f}_k \boldsymbol{f}_k' + \lambda \boldsymbol{I}\right)^{-1}$$

- Other priors to enforce sparsity or bloc-sparsity of the prediction matrix $\boldsymbol{A}$
- See the poster of Mircea Dumitru et al for weight loss prediction from the two genes expressions of Bmal1 and Rev-erb-alpha

Conclusions

- A lot to do to answer the questions of biologists
- Forward modeling and Bayesian inference are natural tools to answer these questions
- Very often the questions are ill-posed inverse problems which need prior knowledge
- Appropriate translation of prior knowledge to prior laws is very important
- Carefull computational algorithms have to be developped
- Carfeful presentation and interpretation of the inference results are very important
- Constant dialogue between ”biologists” and ”Data and Signal processors” is of great importance.