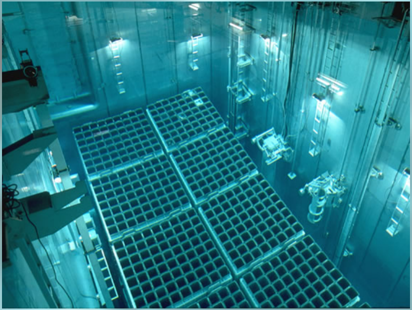


Clustering functional data with wavelets

Jairo Cugliari

R39 - OSIRIS - EDF R & D

RESP.: XAVIER BROSSAT

August 2010

ADVISORS: Anestis Antoniadis[a] and Jean-Michel Poggi[b]

[a] Joseph Fourier University, Grenoble
[b] Paris-Sud University

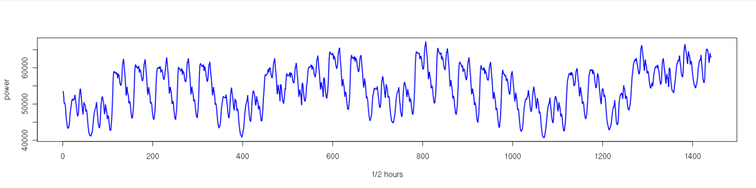
Plan

EDF data

Functional data from a time series

Consider a square integrable continuous time stochastic process $X = (X(t), t \in \mathbb{R})$ observed over the interval $[0, T]$, $T > 0$ at a relatively high sampling frequency. A commonly used approach is to divide the interval $[0, T]$ into subintervals $[l\delta, (l+1)\delta]$, $l = 1, \ldots, n$ with $\delta = T/n$, and to consider the functional-valued discrete time stochastic process $Z = (Z_i, i \in \mathbb{N})$, associated to X by

$$Z_i(t) = X(i\delta + t) \qquad t \in [0, \delta) \tag{1}$$

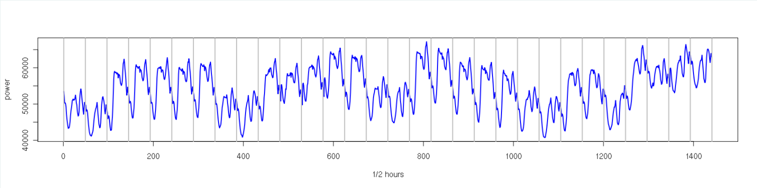


EDF data

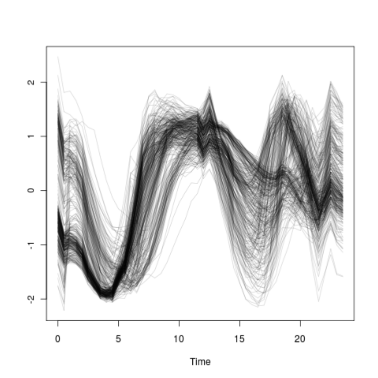
Functional data from a time series

Consider a square integrable continuous time stochastic process $X = (X(t), t \in \mathbb{R})$ observed over the interval $[0, T]$, $T > 0$ at a relatively high sampling frequency. A commonly used approach is to divide the interval $[0, T]$ into subintervals $[l\delta, (l+1)\delta]$, $l = 1, \ldots, n$ with $\delta = T/n$, and to consider the functional-valued discrete time stochastic process $Z = (Z_i, i \in \mathbb{N})$, associated to X by

$$Z_i(t) = X(i\delta + t) \qquad t \in [0, \delta) \tag{1}$$



Clustering and FD



- Given a sample of curves, we search for homogeneous subgroups of individuals.
- Clustering is a process for partitioning a dataset into sub-groups
- The instances within a group are similar to each other and are very dissimilar to the instances of other groups.
- In a functional context clustering helps to identify representative curve patterns and individuals who are very likely involved in the same or similar processes.

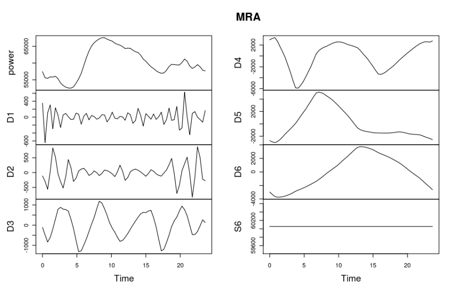
Plan

Wavelets

Wavelet transform

- domain-transform technique for hierarchical decomposing finite energy signals
- description in terms of an approximation plus a set of details
- the broad trend is preserved in the approximation part, while the localized changes are kept in the detail parts.

For short, a wavelet is a smooth and quickly vanishing oscillating function with good localisation properties in both frequency and time.
Specially interesting for approximating time series curves that contain localized structures !!!



Discret Wavelet Transform

We consider an orthonormal basis of waveforms derived from scaling and translations of a compactly supported scaling function ϕ and a compactly supported mother wavelet ψ. We let

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k), \quad \psi_{j,k}(t) = 2^{j/2}\phi(2^j t - k).$$

For any $j_0 \geq 0$, the collection

$$\{\phi_{j_0,k}, k = 0, 1, \ldots, 2^{j_0} - 1; \psi_{j,k}, j \geq j_0, k = 0, 1, \ldots, 2^j - 1\}, \tag{2}$$

is an orthonormal basis of $\mathcal{H}$ a real separable Hilbert space.
Any $z \in \mathcal{H}$ can be written as

$$z(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t), \tag{3}$$

where $c_{j,k}$ and $d_{j,k}$ are the scale and the wavelet coefficients (resp.) of z at the position k of the scale j defined as

$$c_{j,k} =< z, \phi_{j,k} >_{\mathcal{H}} \qquad d_{j,k} =< z, \psi_{j,k} >_{\mathcal{H}} .$$

Discret Wavelet Transform

We consider an orthonormal basis of waveforms derived from scaling and translations of a compactly supported scaling function ϕ and a compactly supported mother wavelet ψ. We let

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k), \quad \psi_{j,k}(t) = 2^{j/2}\phi(2^j t - k).$$

For any $j_0 \geq 0$, the collection

$$\{\phi_{j_0,k}, k = 0, 1, \ldots, 2^{j_0} - 1; \psi_{j,k}, j \geq j_0, k = 0, 1, \ldots, 2^j - 1\}, \tag{2}$$

is an orthonormal basis of $\mathcal{H}$ a real separable Hilbert space.
Any $z \in \mathcal{H}$ can be written as

$$\widetilde{z}_J(t) = c_0\phi_{0,0}(t) + \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t). \tag{3}$$

where $c_{j,k}$ and $d_{j,k}$ are the scale and the wavelet coefficients (resp.) of z at the position k of the scale j defined as

$$c_{j,k} =< z, \phi_{j,k} >_{\mathcal{H}} \qquad d_{j,k} =< z, \psi_{j,k} >_{\mathcal{H}} .$$

Energy decomposition of the DWT

Since DWT is based on an L_2-orthonormal basis decomposition we have conservation of the signal's energy.
We can then write for a discretized function $\widetilde{\mathbf{z}}$ a characterization by the set of channel variances estimated at the output of the corresponding filter bank:

$$\mathcal{E}_z \approx \|\mathbf{z}\|_2^2 = c_0^2 + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k}^2 = c_0^2 + \sum_{j=0}^{J-1} \|\mathbf{d}_j\|_2^2. \tag{4}$$

where $\mathcal{E}_z = \|z\|_{\mathcal{H}}^2$.

Scale specific AC and RC Contributions

We will use $j_0 = 0$ and we will concentrate on the wavelet coefficients $d_{j,k}$.
We have conservation of the energy

$$||z(t)||^2 = ||c_{0,0}||^2 + \sum_j ||\mathbf{d_j}||^2$$

.
For each $j = 1, \ldots, J$, we compute the absolute and relative contribution representations (ACR and RCR rp.) by

$$\underbrace{\mathrm{cont}_j = ||\mathbf{d_j}||^2}_{\boxed{\text{ACR}}} \qquad \underbrace{\mathrm{rel}_j = \frac{||\mathbf{d_j}||^2}{\sum_j ||\mathbf{d_j}||^2}}_{\boxed{\text{RCR}}}$$

These coefficients resume the relative importance of each scale to the global dynamic of a trajectory.

Plan

Simulated data

We simulate $K = 3$ clusters of 25 observations sampled by 1024 points each.

a 2-sinus model

b FAR with diagonal covariance operator

c FAR with non diagonal covariance operator

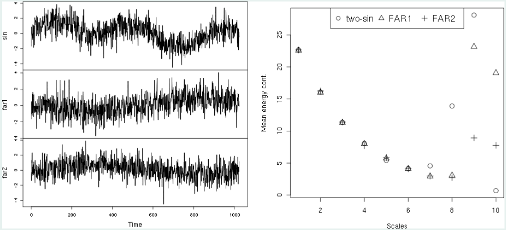
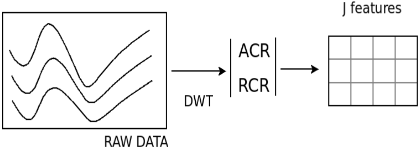
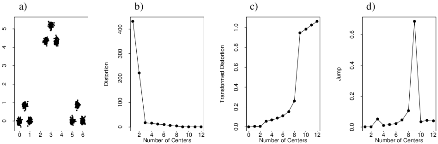


Figure: Mean energy scale's contribution by model.

Schema of procedure





- After approximating functions by discretized data, we obtain J handy features.
- We use Steinley & Brusco's feature selection algorithm
- In order to use $k-$means we estimate the number of clusters K by detecting jumps in the distortion energy curve d_K(Sugar & James, 2003):

Simulated data

Confusion matrix

Model	K_1	K_2	K_3
2-sinus	25	–	–
FAR1	–	20	5
FAR2	–	13	12

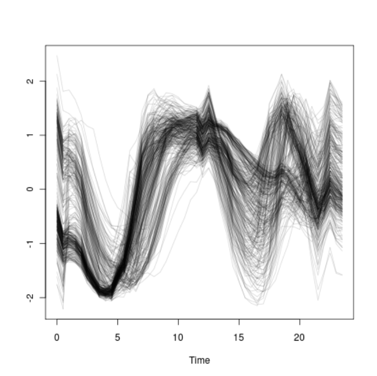
- Good overall missclafication rate (18/75)
- Perfect distinction of 2-sinus model
- Relatively good performance on the FAR models

EDF application

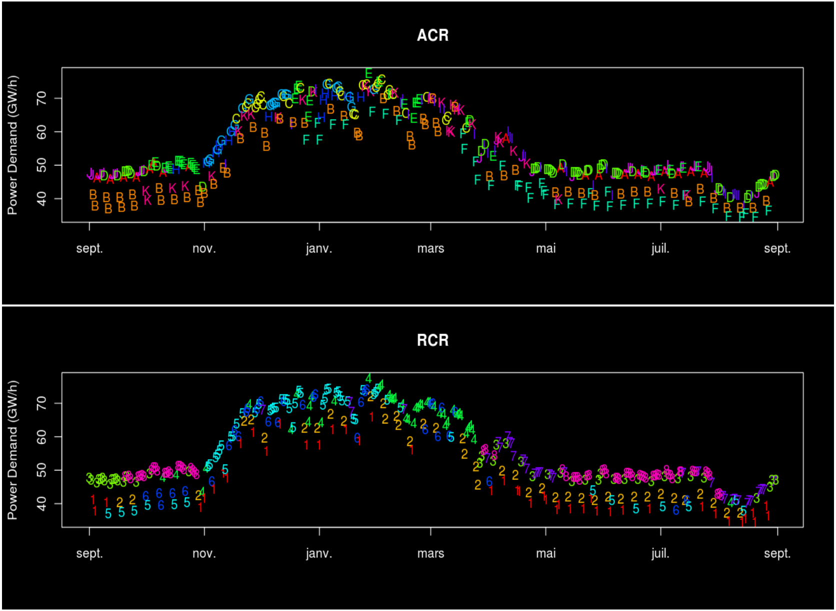
Data: 365 daily power demand profiles of french national consumption (48 points per day)

Some well known facts of electricity demand:

- 2 well defined seasons with transitions
- Weekly cycle due to calendar (WE vs working days)
- Daily cycle: day vs night
- Other features that affect electricity consumtion: bank holidays, special priced days, strikes, financial crisis, storms



Aim: Detect daily profiles of french national electricity load demand.



ACR
Power Demand (GW/h)
40
50
60
70
sept.
nov.
janv.
mars
mai
juil.
sept.
RCR
Power Demand (GW/h)
40
50
60
70
sept.
nov.
janv.
mars
mai
juil.
sept.

Plan

Conclusion

- We have presented a way of efficiently clustering functions using wavelet-based dissimilarities.
- Wavelets give a well suited plateform because of their capacity on detecting highly localized events.
- Feature extraction and feature selection give additional explanaitory capacity to unsupervised learning.

Antoniadis, A. - Fan, J.
Regularization of wavelet approximations.
Journal of the American Statistical Association, 96:455, 2001.

Antoniadis, A. - Paparoditis, E. - Sapatinas, T.
A functional wavelet-kernel approach for time series prediction.
Royal Statistical Society, 68:837–857, 2008.

Antoniadis, A. - Paparoditis, E. - Theofanis, S.
Bandwidth selection for functional time series prediction.

Poggi, J.M.
Prévision non paramétrique de la consommation électrique.
Rev. Statistiqué Appliquée, XLII(4):93–98, 1994.