Conditional Autoregressif Hilbertian process

Application to the electricity demand

Jairo Cugliari

Inria
SELECT Research Team

Sustain workshop, Bristol
11 September, 2012

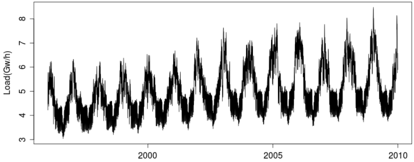
Joint work with
Anestis ANTONIADIS (Univ. Grenoble)
Jean-Michel POGGI (Univ. Paris Descartes)
Xavier BROSSAT (EDF R&D)

Outline

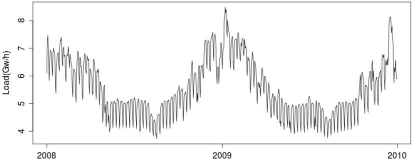
Outline

Electricity demand data

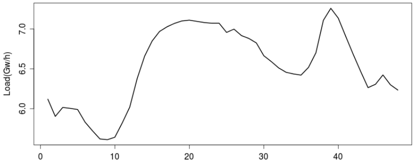
Some salient features



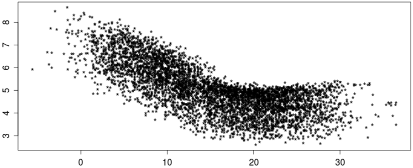
(a) Long term trend.



(b) Annual and week cycles.



(c) Daily pattern.



(d) Demand (in Gw/h) as a function of temperature (in °C)

Electricity demand forecast

Short-term electricity demand forecast in literature

- Time series analysis: SARIMA(X), Kalman filter [Dordonnat *et al.* (2009)]
- Machine learning. [Devaine *et al.* (2010)]
- Similarity search based methods. [Poggi (1994), Antoniadis *et al.* (2006)]
- Regression: EDF modelisation scheme [Bruhns *et al.* (2005)] , GAM [Pierrot and Goude (2011)], Bayesian approach [Launay , Philippe and Lamarche (2012)]

New challenges

- Market liberalization: may produce variations on clients' perimeter that risk to induce nonstationarities on the signal.
- Development of smart grids and smart meters.

But, almost all the models rely on a monoscale representation of the data

Electricity demand forecast

Short-term electricity demand forecast in literature

- Time series analysis: SARIMA(X), Kalman filter [Dordonnat *et al.* (2009)]
- Machine learning. [Devaine *et al.* (2010)]
- Similarity search based methods. [Poggi (1994), Antoniadis *et al.* (2006)]
- Regression: EDF modelisation scheme [Bruhns *et al.* (2005)] , GAM [Pierrot and Goude (2011)], Bayesian approach [Launay , Philippe and Lamarche (2012)]

New challenges

- Market liberalization: may produce variations on clients' perimeter that risk to induce nonstationarities on the signal.
- Development of smart grids and smart meters.

But, almost all the models rely on a monoscale representation of the data

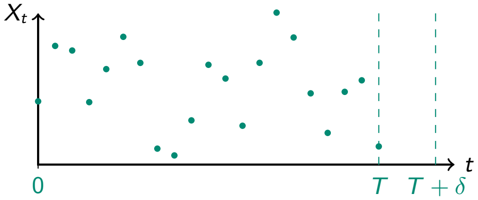
Outline

FD as slices of a continuous process

[Bosq, (1990)]

The prediction problem

- Suppose one observes a square integrable continuous-time stochastic process $X = (X(t), t \in \mathbb{R})$ over the interval $[0, T]$, $T > 0$;
- We want to predict X all over the segment $[T, T+\delta], \delta > 0$
- Divide the interval into n subintervals of equal size δ.
- Consider the functional-valued discrete time stochastic process $Z = (Z_k, k \in \mathbb{N})$, where $\mathbb{N} = \{1, 2, \dots\}$, defined by

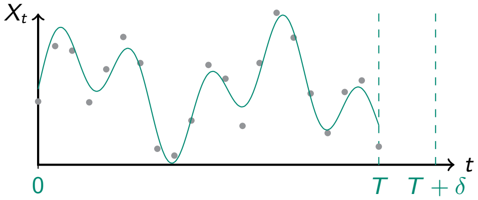


FD as slices of a continuous process

[Bosq, (1990)]

The prediction problem

- Suppose one observes a square integrable continuous-time stochastic process $X = (X(t), t \in \mathbb{R})$ over the interval $[0, T]$, $T > 0$;
- We want to predict X all over the segment $[T, T + \delta], \delta > 0$
- Divide the interval into n subintervals of equal size δ.
- Consider the functional-valued discrete time stochastic process $Z = (Z_k, k \in \mathbb{N})$, where $\mathbb{N} = \{1, 2, \ldots\}$, defined by

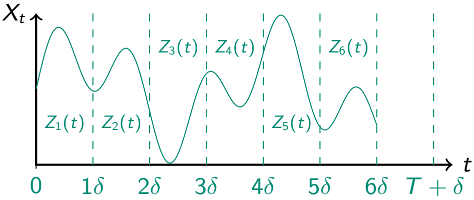


FD as slices of a continuous process

[Bosq, (1990)]

The prediction problem

- Suppose one observes a square integrable continuous-time stochastic process $X = (X(t), t \in \mathbb{R})$ over the interval $[0, T]$, $T > 0$;
- We want to predict X all over the segment $[T, T+\delta], \delta > 0$
- Divide the interval into n subintervals of equal size δ.
- Consider the functional-valued discrete time stochastic process $Z = (Z_k, k \in \mathbb{N})$, where $\mathbb{N} = \{1, 2, \ldots\}$, defined by



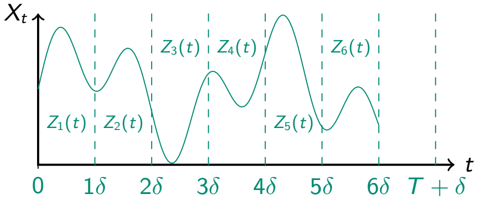
$$Z_k(t) = X(t + (k-1)\delta)$$
$$k \in \mathbb{N} \quad \forall t \in [0, \delta)$$

FD as slices of a continuous process

[Bosq, (1990)]

The prediction problem

- Suppose one observes a square integrable continuous-time stochastic process $X = (X(t), t \in \mathbb{R})$ over the interval $[0, T]$, $T > 0$;
- We want to predict X all over the segment $[T, T + \delta], \delta > 0$
- Divide the interval into n subintervals of equal size δ.
- Consider the functional-valued discrete time stochastic process $Z = (Z_k, k \in \mathbb{N})$, where $\mathbb{N} = \{1, 2, \ldots\}$, defined by



$$Z_k(t) = X(t + (k-1)\delta)$$
$$k \in \mathbb{N} \quad \forall t \in [0, \delta)$$

If X contents a $\delta-$seasonal component, Z is particularly fruitful.

Prediction of functional time series

Let $(Z_k, k \in \mathbb{Z})$ be a stationary sequence of H-valued r.v. Given $Z_1, \ldots, Z_n$ we want to predict the future value of Z_{n+1}.

- A predictor of Z_{n+1} using $Z_1, Z_2, \ldots, Z_n$ is

$$\widetilde{Z_{n+1}} = \mathbb{E}[Z_{n+1} | Z_n, Z_{n-1}, \ldots, Z_1].$$

Autoregressive Hilbertian process of order 1

The $\mathrm{ARH}(1)$ centred process states that at each k,

$$Z_k = \rho(Z_{k-1}) + \epsilon_k \tag{1}$$

where ρ is a compact linear operator and $\{\epsilon_k\}_{k \in \mathbb{Z}}$ is an $H-$valued strong white noise.
Under mild conditions, equation (1) has a unique solution which is a strictly stationary process with innovation $\{\epsilon_k\}_{k \in \mathbb{Z}}$. [Bosq, (1991)]
When Z is a zero-mean $\mathrm{ARH}(1)$ process, the best predictor of Z_{n+1} given $\{Z_1, \ldots, Z_{n-1}\}$ is:

$$\widetilde{Z_{n+1}} = \rho(Z_n).$$

Prediction of functional time series

Let $(Z_k, k \in \mathbb{Z})$ be a stationary sequence of H-valued r.v. Given $Z_1, \ldots, Z_n$ we want to predict the future value of Z_{n+1}.

- A predictor of Z_{n+1} using $Z_1, Z_2, \ldots, Z_n$ is

$$\widetilde{Z_{n+1}} = \mathbb{E}[Z_{n+1} | Z_n, Z_{n-1}, \ldots, Z_1].$$

Autoregressive Hilbertian process of order 1

The ARH(1) centred process states that at each k,

$$Z_k = \rho(Z_{k-1}) + \epsilon_k \tag{1}$$

where ρ is a compact linear operator and $\{\epsilon_k\}_{k\in\mathbb{Z}}$ is an $H-$valued strong white noise.

Under mild conditions, equation (1) has a unique solution which is a strictly stationary process with innovation $\{\epsilon_k\}_{k\in\mathbb{Z}}$. [Bosq, (1991)]

When Z is a zero-mean ARH(1) process, the best predictor of Z_{n+1} given $\{Z_1, \ldots, Z_{n-1}\}$ is:

$$\widetilde{Z_{n+1}} = \rho(Z_n).$$

Outline

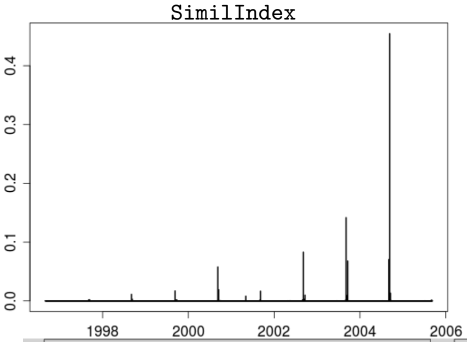
Let us predict Saturday 10 September 2005

We use Antoniadis *et al.*, (2006) prediction method with corrections to cope with non stationarity.

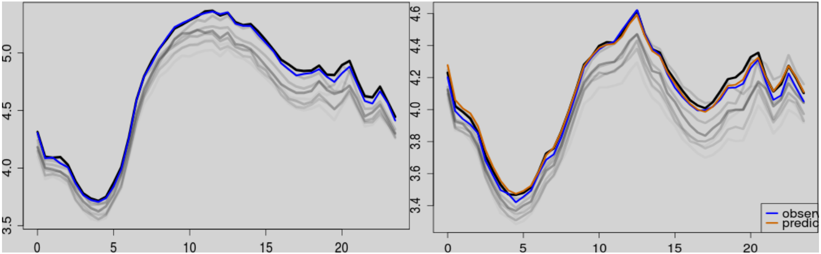
- Use the last observed segment ($n = 9/\text{Sept}/2005$) to look for similar segments in past.
- Construct a similarity index `SimilIndex` (using a kernel).
- Prediction can be written as

$$\widehat{\text{Load}}_{n+1}(t) = \sum_{m=1}^{n-1} \text{SimilIndex}_{m,n} \times \text{Load}_{m+1}(t)$$

- First difference correction of the approximation part.
- Use of groups to anticipate calendar transitions.



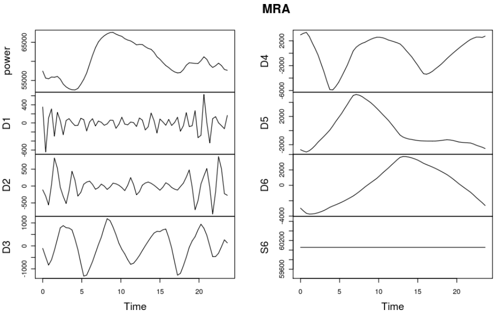
date	SimilIndex
2004-09-10	0.455
2003-09-05	0.141
2002-09-06	0.083
2004-09-03	0.070
2003-09-19	0.068
2000-09-08	0.058
2000-09-15	0.019
1999-09-10	0.017



similar past

similar future

Wavelets to cope with FD



- domain-transform technique for hierarchical decomposing finite energy signals
- description in terms of a broad trend (approximation part), plus a set of localized changes kept in the details parts.

Discrete Wavelet Transform

If $z \in L_2([0,1])$ we can write it as

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

where $c_{j,k} =< g, \phi_{j,k} >$, $d_{j,k} =< g, \varphi_{j,k} >$ are the scale coefficients and wavelet coefficients respectively, and the functions ϕ et φ are associated to a orthogonal MRA of $L_2([0,1])$.

Approximation and details

- In practice, we don't dispose of the whole trajectory but only with a (possibly noisy) sampling at 2^J points, for some integer J.
- Each approximated segment $Z_{i,J}(t)$ is broken up into two terms:
  - a smooth approximation $\mathcal{S}_i(t)$ (lower freqs)
  - a set of details $\mathcal{D}_i(t)$ (higher freqs)

$$Z_{i,J}(t) = \underbrace{\sum_{k=0}^{2^{j_0}-1} c_{j_0,k}^{(i)} \phi_{j_0,k}(t)}_{\mathcal{S}_i(t)} + \underbrace{\sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k}^{(i)} \psi_{j,k}(t)}_{\mathcal{D}_i(t)}$$

- The parameter j_0 controls the separation. We set $j_0 = 0$.

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

A two step prediction algorithm

Step I: Dissimilarity between segments

Search the past for segments that are similar to the last one.
For two observed series of length 2^J say Z_m and Z_l we set for each scale $j \geq j_0$:

$$\text{dist}_j(Z_m, Z_l) = \left(\sum_{k=0}^{2^j-1} (d_{j,k}^{(m)} - d_{j,k}^{(l)})^2 \right)^{1/2}$$

Then, we aggregate over the scales taking into account the number of coefficients at each scale

$$D(Z_m, Z_l) = \sum_{j=j_0}^{J-1} 2^{-j/2} \text{dist}_j(Z_m, Z_l)$$

A two step prediction algorithm

Step 2: Kernel regression

Obtain the prediction of the scale coefficients at the finest resolution $\Xi_{n+1} = \{c_{J,k}^{(n+1)} : k = 0, 1, \ldots, 2^J - 1\}$ for Z_{n+1}

$$\widehat{\Xi_{n+1}} = \sum_{m=1}^{n-1} w_{m,n} \Xi_{m+1}$$

$$w_{m,n} = \frac{K\left(\frac{D(Z_n, Z_m)}{h_n}\right)}{\sum_{m=1}^{n-1} K\left(\frac{D(Z_n, Z_m)}{h_n}\right)}$$

Finally, the prediction of Z_{n+1} can be written

$$\widehat{Z_{n+1}(t)} = \sum_{k=0}^{2^J-1} \widehat{c_{J,k}^{(n+1)}} \phi_{J,k}(t)$$

Outline

CARH process

$(Z, V) = \{(Z_k, V_k) \in H \times \mathbb{R}^d, k \in \mathbb{Z}\}$ is a CARH(1) process if it is stationary and and such that,

$$Z_k = a + \rho_{V_k}(Z_{k-1} - a) + \epsilon_k, \qquad k \in \mathbb{Z}, \tag{2}$$

where for each $v \in \mathbb{R}^d$, $a_v = \mathbb{E}^v[Z_0|V]$, $\{\epsilon_k\}_{k\in\mathbb{Z}}$ is an $H-$white noise independent of V, and $\{\rho_{V_k}\}_{k\in\mathbb{Z}}$ is a sequence of linear compact operators.

Theorem (Existence and uniqueness)

If $\sup_n \|\rho_{V_n}\|_{\mathcal{L}} < 1$ *a.s., then* (2) *defines a* CARH *process with an unique stationary solution given by*

$$Z_k = a + \sum_{j=0}^{\infty} \left(\prod_{p=0}^{j-1} \rho_{V_{k-p}} \right) (\epsilon_{k-j}),$$

with the convention $\prod_{p=0}^{j-1} \rho_{V_{k-p}}$ *is the identity operator for* $j = 0$.

Conditional covariance operators

Conditional covariance and cross covariance operators (on V at the point $v \in \mathbb{R}^d$) are respectively defined by

$$z \in H \mapsto \Gamma_v z = \mathbb{E}^v[(Z_0 - a) \otimes (Z_0 - a)(z)|V] \qquad \text{and}$$
$$z \in H \mapsto \Delta_v z = \mathbb{E}^v[(Z_0 - a) \otimes (Z_1 - a)(z)|V],$$

where $x \in H \to (u \otimes v)(x) =< u, x > v$.

- For each $v \in \mathbb{R}^d$: these are trace-class operators, thus Hilbert-Schimdt (additionally Γ_v is positive definite and selfadjoint)
- Spectral decomposition of Γ_v, $\Gamma_v = \sum_{j \in \mathbb{N}} \lambda_{v,j}(e_{v,j} \otimes e_{v,j})$, where $\lambda_{v,1} \geq \lambda_{v,2} \geq \ldots \geq 0$ are the eigenvalues and $(e_{v,j})_{j \in \mathbb{N}}$ the associated eigenfunctions.

Estimation of the conditional covariance operators

- Nonparametric Nadaraya-Watson like estimators.
- Context of dependent data ($\alpha-$mixing framework)

$$\hat{a}_{v,n} = \sum_{i=1}^{n} w_{n,i}(v, h_a) Z_i$$

$$\hat{\Gamma}_{v,n} = \sum_{i=1}^{n} w_{n,i}(v, h_\gamma)(Z_i - \hat{a}_n(v)) \otimes (Z_i - \hat{a}_n(v))$$

$$\hat{\Delta}_{v,n} = \sum_{i=2}^{n} w_{n,i}(v, h_\delta)(Z_{i-1} - \hat{a}_n(v)) \otimes (Z_i - \hat{a}_n(v))$$

where the weights $w_{n,i}$ are defined by

$$w_{n,i}(v, h) = \frac{K(h^{-1}(V_i - v))}{\sum_{i=1}^{n} K(h^{-1}(V_i - v))}. \tag{3}$$

Estimation of ρ_v

Two relation between the operators

$$\Delta_v = \rho_v \Gamma_v \qquad \text{and} \qquad \Delta_v^* = \Gamma_v \rho_v^*.$$

- If $\dim(H) < \infty$, the inversion of the operator Γ_v gives us a way to estimate ρ_v.
- Problem In the general case, the inverse of Γ_v is a problem: the operator is not bounded and may not be defined over the whole space H ([Mas, 2000]).

However, for a well identify ρ_v we can define a linear measurable mapping Γ_v^{-1} within a dense domain $\mathcal{D}_{\Gamma_v^{-1}} \subset H$, and using the *closed graph theorem* and the fact that the range$(\Delta_v^*) \subset \mathcal{D}_{\Gamma_v^{-1}}$, then restricted to $\mathcal{D}_{\Gamma_v^{-1}}$ we can write

$$\rho_v^* = \Gamma_v^{-1} \Delta_v^*.$$

Classical results on linear operators allow us to extend ρ_v^* by continuity to H. Then, we focus on the estimation of ρ_v^*.

Estimation of ρ_v

We extend the two class of estimators proposed by [Mas, (2000)] proposed on the ARH framework.
Let us call $P_v^{k_n}$ the projection operator from H to $H_v^{k_n}$. Then, we define the projection estimator of ρ_v^* by

$$\widehat{\rho}^*_{v,n} = (P_v^{k_n}\widehat{\Gamma}_{v,n}P_v^{k_n})^{-1}\widehat{\Delta}^*_{v,n}P_v^{k_n}. \quad (4)$$

A whole class of resolvent estimators can be obtained using the resolvent of Γ_v

$$\widehat{\rho}^*_{v,n,p} = b_{n,p}(\widehat{\Gamma}_{v,n})\widehat{\Delta}^*_{v,n}, \quad (5)$$

where we write $b_{n,p,\alpha}(\widehat{\Gamma}_{v,n}) = (\widehat{\Gamma}_{v,n} + \alpha_n I)^{-(p+1)}$ with $p \geq 0$, $\alpha_n \geq 0$, $n \geq 0$.

Almost sure convergence results are obtained for all the propose estimator. In addition, convergence on probability of both predictors $\widehat{\rho}^*_{v,n}(Z_{n+1})$ and $\widehat{\rho}^*_{v,n,p}(Z_{n+1})$.

Simulation and prediction

- We extend the simulation strategies for ARH processes [Guillas & Damon (2000)] to the simple case of an CARH process with $d = 1$ and V is a i.i.d. sequence of Beta$(\beta_1, \beta 2)$ rv.
- Numerical experience: prediction of the electricity demand using the temperature as exogenous information

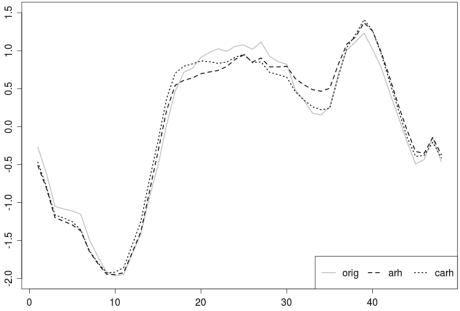


Figure: Prediction of one simulated curve of an CARH process (full line)