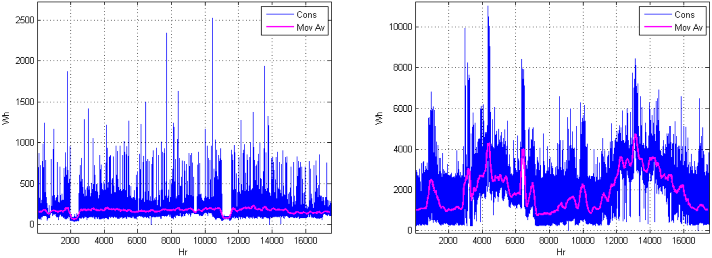
A SARIMAX coupled modelling applied to individual load curves intraday forecasting

Frédéric Proïa

Workshop EDF
Institut Henri Poincaré - Paris
05 avril 2012

INRIA Bordeaux Sud-Ouest
Institut de Mathématiques de Bordeaux

Plan



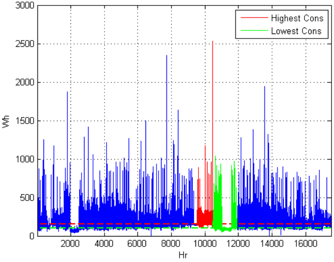
- **Motivations.**
  - New energy meters to gather individual consumption with high frequencie.
  - Economic issue for EDF : anticipate to optimize.

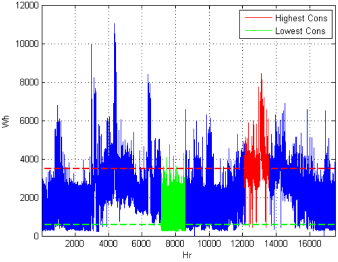
- **Objectives.**
  - Distinguish nonthermosensitive from thermosensitive customers.
  - Intraday daily forecasting.
  - Introduce temperature as an exogenous contribution.

- **Detecting thermosensitivity.**
  - Deterministic criterion.
  - Thermosensitivity if

$$\max_{\Delta \le t \le N} M_t - \min_{\Delta \le t \le N} M_t > \delta$$

    where M_t is the empirical median of $\left(C_t \quad \dots \quad C_{t-\Delta+1}\right)$.
  - $\Delta = 1440$ et $\delta = 1000$.





- **More pertinent than SR/DR.**
  - 70% of nonthermosensitive customers, only including 75% of SR.

Plan

- **Stationary ARMA processes.**

Definition (Stationarity)

A time series (Y_t) is said to be weakly stationary if, for all $t \in \mathbb{Z}$, $\mathbb{E}[Y_t^2] < \infty$, $\mathbb{E}[Y_t] = m$ and, for all $s, t \in \mathbb{Z}$, $\mathrm{Cov}(Y_t, Y_s) = \mathrm{Cov}(Y_{t-s}, Y_0)$.

Definition (ARMA)

Let (Y_t) be a stationary time series with zero mean. It is said to be an ARMA(p, q) process if, for every $t \in \mathbb{Z}$,

$$Y_t - \sum_{k=1}^{p} a_k Y_{t-k} = \varepsilon_t + \sum_{k=1}^{q} b_k \varepsilon_{t-k}$$

where (ε_t) is a white noise of variance $\sigma^2 > 0$, $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$.

- **Causality of ARMA processes.**
  - Compact expression, for all $1 \leq t \leq T$,

$$\mathcal{A}(B)Y_t = \mathcal{B}(B)\varepsilon_t$$

where the polynomials

$$\mathcal{A}(z) = 1 - a_1 z - \ldots - a_p z^p \qquad \text{and} \qquad \mathcal{B}(z) = 1 + b_1 z + \ldots + b_q z^q.$$

Definition (Causality)

Let (Y_t) be an ARMA(p, q) process for which the polynomials $\mathcal{A}$ and $\mathcal{B}$ have no common zeroes. Then, (Y_t) is causal if and only if $\mathcal{A}(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.

- **Implications.**
  - Causality implies the existence of a MA(∞) structure for (Y_t).
  - Causality implies stationarity of the process.
  - On $\mathbb{N}^*$, causality often coincides with stationarity.

- **Existence and unicity of a stationary solution.**

Proposition

If $\mathcal{A}(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$, then the ARMA equation $\mathcal{A}(B)Y_t = \mathcal{B}(B)\varepsilon_t$ have the unique stationary solution

$$Y_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k},$$

and the coefficients $(\psi_k)_{k\in\mathbb{N}}$ are determined by the relation

$$\mathcal{A}^{-1}(z)\mathcal{B}(z) = \sum_{k=0}^{\infty} \psi_k z^k \qquad \textit{with} \qquad \sum_{k=0}^{\infty} \psi_k^2 < \infty.$$

- **Explosive cases.**
  - On $\mathbb{Z}$, no zeroes on the unit circle is a sufficient condition.
  - Irrelevant for practical purposes.

- **Autocorrelation function.**
  - To identify q.

Definition (ACF)

Let (Y_t) be a stationary time series. The autocorrelation function ρ associated with (Y_t) is defined, for all $t \in \mathbb{Z}$, as

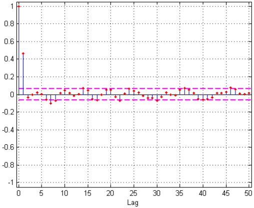
$$\rho(t) = \frac{\gamma(t)}{\gamma(0)}$$

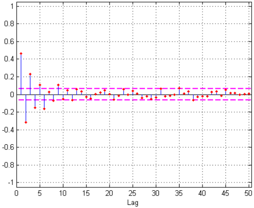
where the autocovariance function $\gamma(t) = \mathrm{Cov}(Y_t, Y_0)$.

Proposition

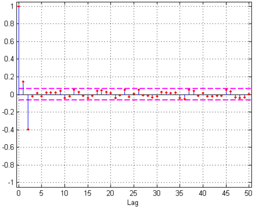
The stationary time series (Y_t) with zero mean is a MA(q) process such that $b_q \neq 0$ if and only if $\rho(q) \neq 0$ and $\rho(t) = 0$ for all $|t| > q$.

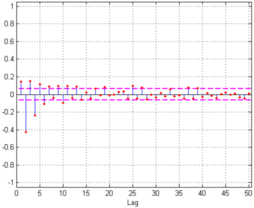
- **Examples.**
  - MA(1) : only $\rho(1)$ nonzero, exponential decay of $\alpha(t)$.





  - MA(2) : only $\rho(1)$ and $\rho(2)$ nonzero, damped exponential and sine wave for $\alpha(t)$.





- **Partial autocorrelation function.**
  - To identify p.

Definition (PACF)

Let (Y_t) be a stationary time series with zero mean. The partial autocorrelation function α is defined as $\alpha(0) = 1$, and, for all $t \in \mathbb{N}^*$, as

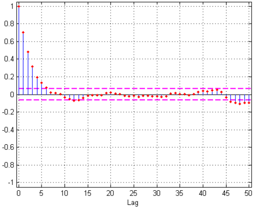
$$\alpha(t) = \phi_{t,\,t}$$

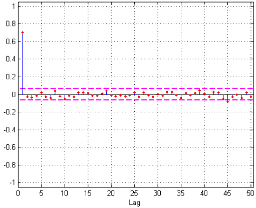
where $(\phi_{t,\,t})_{t \in \mathbb{N}^*}$ are computed via the Durbin-Levinson recursion.

Proposition

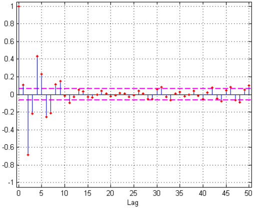
If there exists a square-integrable sequence $(\psi_k)_{k \in \mathbb{N}}$ such that (Y_t) has a MA(∞) expression with $\psi_0 = 1$, then the stationary time series (Y_t) with zero mean is an AR(p) process such that $a_p \neq 0$ if and only if $\alpha(p) \neq 0$ and $\alpha(t) = 0$ for all $t > p$.

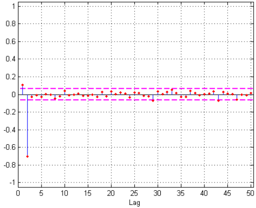
- **Examples.**
  - AR(1) : only $\alpha(1)$ nonzero, exponential decay of $\rho(t)$.



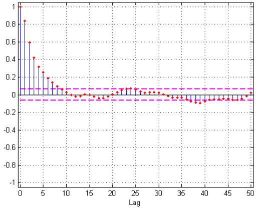


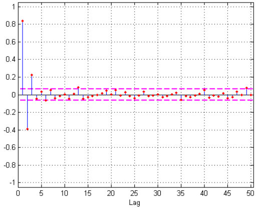
  - AR(2) : only $\alpha(1)$ and $\alpha(2)$ nonzero, damped exponential and sine wave for $\rho(t)$.



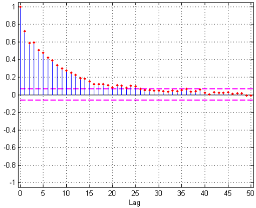


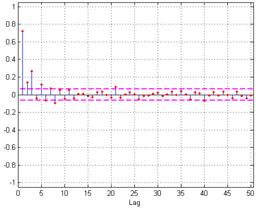
- **Examples.**
  - ARMA(1,1) : exponential decay of $\rho(t)$ and $\alpha(t)$ from first lag.





  - ARMA(2,2) : exponential decay of $\rho(t)$ and $\alpha(t)$ from second lag.

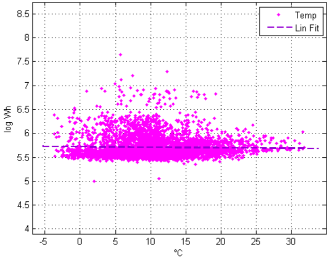


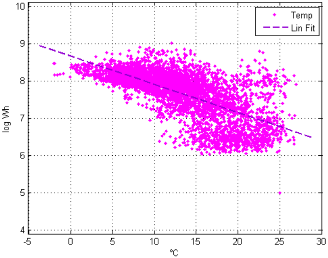


- **Variance-stabilizing Box-Cox transformation.**
  - Logarithmic transform given, for all $1 \leq t \leq T$, by

$$Y_t = \log\left(C_t + e^m\right)$$

where m ensures that $Y_t = m$ when $C_t = 0$.





- **Linear relationship between consumption and temperature.**
  - On a thermosensitive load curve, for all $1 \leq t \leq T$,

$$Y_t = c_0 + \mathcal{C}(B)U_t + \varepsilon_t.$$

  - For all $z \in \mathbb{C}$,

$$\mathcal{C}(z) = \sum_{k=0}^{r-1} c_{k+1} z^k.$$

  - Unknown vector $c \in \mathbb{R}^{r+1}$ estimated by OLS.

- **Seasonal residuals.**
  - Residuals (ε_t) regarded as a seasonal time series.

- **Residuals (ε_t) as a seasonal time series.**
  - SARIMA$(p, d, q) \times (P, D, Q)_s$ modelling, for all $1 \leq t \leq T$,

$$(1-B)^d (1-B^s)^D \mathcal{A}(B)\mathcal{A}_s(B)\varepsilon_t = \mathcal{B}(B)\mathcal{B}_s(B)V_t,$$

    where (V_t) is a white noise of variance $\sigma^2 > 0$.
  - Polynomials defined, for all $z \in \mathbb{C}$, as

$$\mathcal{A}(z) = 1 - \sum_{k=1}^{p} a_k z^k, \qquad \mathcal{A}_s(z) = 1 - \sum_{k=1}^{P} \alpha_k z^{sk},$$

$$\mathcal{B}(z) = 1 - \sum_{k=1}^{q} b_k z^k, \qquad \mathcal{B}_s(z) = 1 - \sum_{k=1}^{Q} \beta_k z^{sk},$$

  - Parameters $a \in \mathbb{R}^p$, $b \in \mathbb{R}^q$, $\alpha \in \mathbb{R}^P$ and $\beta \in \mathbb{R}^Q$ estimated by GLS.
  - $\mathcal{A}$ and $\mathcal{A}_s$ are causal.

- **The dynamic coupled modelling.**

Definition (SARIMAX)

In the particular framework of the study, a random process (Y_t) will be said to follow a SARIMAX$(p, d, q, r) \times (P, D, Q)_s$ coupled modelling if, for all $1 \leq t \leq T$, it satisfies

$$\begin{cases} Y_t = c_0 + \mathcal{C}(B)U_t + \varepsilon_t, \\ (1-B)^d(1-B^s)^D \mathcal{A}(B)\mathcal{A}_s(B)\varepsilon_t = \mathcal{B}(B)\mathcal{B}_s(B)V_t. \end{cases}$$

- As soon as $d + D > 0$,

$$(1-B)^d(1-B^s)^D \mathcal{A}(B)\mathcal{A}_s(B)\left(Y_t - \mathcal{C}(B)U_t\right) = \mathcal{B}(B)\mathcal{B}_s(B)V_t.$$

- **Existence of a stationary solution.**
  - Let I be the identity matrix of order T and

$$U = \begin{pmatrix} 1 & U_T & U_{T-1} & \dots & U_{T-r+1} \\ 1 & U_{T-1} & U_{T-2} & \dots & U_{T-r} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & U_1 & U_0 & \dots & U_{-r+2} \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}.$$

Theorem

Assume that $U'U$ is invertible. Then, the differenced process $(\nabla^d \nabla_s^D \varepsilon_t)$ where ε_t is given, for all $1 \leq t \leq T$, by the vector form

$$\varepsilon = \left(I - U(U'U)^{-1}U'\right) Y$$

is a stationary solution of the coupled model suitably specified.

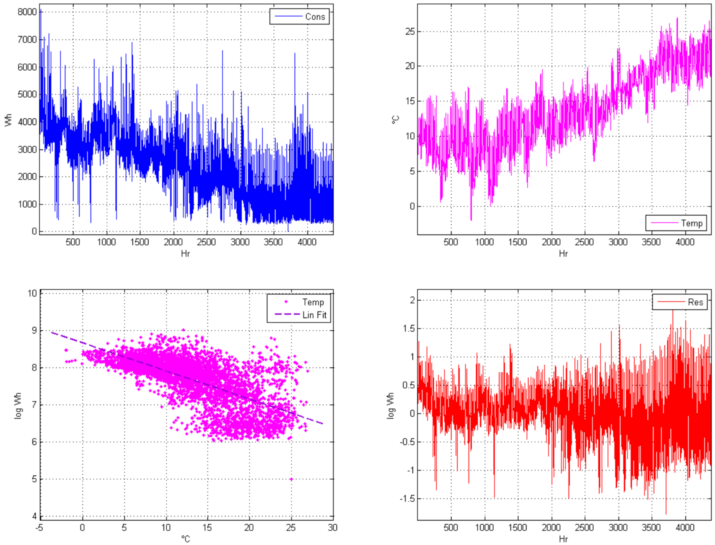
- ADF and KPSS tests : d and D.
- Box and Jenkins methodology : p, q, r, P, Q, s.

- **Forecasting using time series analysis.**
  - Let $\widetilde{\varepsilon}_{T+1}$ be the predictor of (ε_t) at stage $T+1$.
  - Let $\widehat{c}_T$ be the OLS estimate of c.
  - Assume that $\widehat{r}$ has been evaluated.
  - The predictor at horizon 1 is given by

$$\widetilde{Y}_{T+1} = \widehat{c}_{0,T} + \sum_{k=1}^{\widehat{r}} \widehat{c}_{k,T} U_{T-k+2} + \widetilde{\varepsilon}_{T+1}.$$

  - The predictor at horizon H is given by

$$\widetilde{Y}_{T+H} = \widehat{c}_{0,T} + \sum_{k=1}^{\widehat{r}} \widehat{c}_{k,T} U_{T-k+H+1} + \widetilde{\varepsilon}_{T+H}.$$

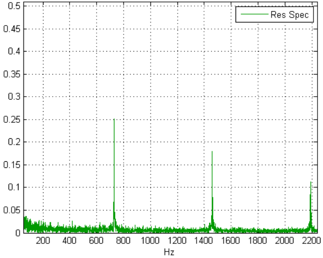

Cons
Wh
Hr
Temp
°C
Hr
Temp
Lin Fit
log Wh
°C
Res
log Wh
Hr

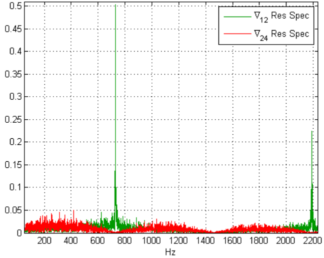
- **Box and Jenkins methodology.**
  - For a given $\widehat{r}$, estimation of the residual set $(\widehat{\varepsilon}_t)$, for all $1 \leq t \leq T$,

$$\widehat{\varepsilon}_t = Y_t - \widehat{c}_{0,T} - \sum_{k=1}^{\widehat{r}} \widehat{c}_{k,T} U_{t-k+1}.$$

  - Select $\widehat{s}$ by investigating the seasonality of $(\widehat{\varepsilon}_t)$.
  - Select $\widehat{d}$ and $\widehat{D}$ by investigating the stationarity of $(\nabla^d \nabla_{\widehat{s}}^D \widehat{\varepsilon}_t)$.
  - Select $\widehat{p}$, $\widehat{q}$, $\widehat{P}$ and $\widehat{Q}$ by looking at ACF and PACF on $(\nabla^{\widehat{d}} \nabla_{\widehat{s}}^{\widehat{D}} \widehat{\varepsilon}_t)$.
  - Adjust $\widehat{p}$, $\widehat{q}$, $\widehat{r}$, $\widehat{P}$ and $\widehat{Q}$ by minimizing bayesian or prediction criteria.
  - Test of white noise on the fitted innovations.

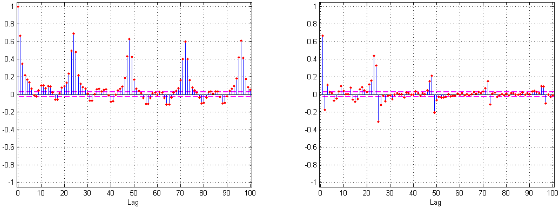
- **Seasonality.**
  - Fourier spectrogram on $(\widehat{\varepsilon}_t)$, $(\nabla_{12}\widehat{\varepsilon}_t)$ and $(\nabla_{24}\widehat{\varepsilon}_t)$, for $T = 730 \times 24$ and $\widehat{r} = 1$.



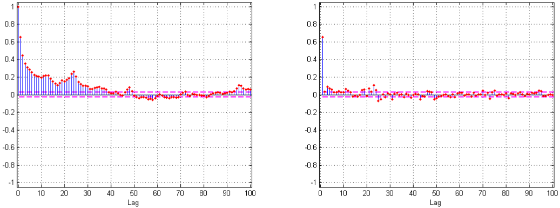


- **Stationarity.**
  - $(\widehat{\varepsilon}_t)$ is not stationary.
  - $(\nabla\widehat{\varepsilon}_t)$, $(\nabla_{24}\widehat{\varepsilon}_t)$ and $(\nabla\nabla_{24}\widehat{\varepsilon}_t)$ are stationary around a deterministic trend.

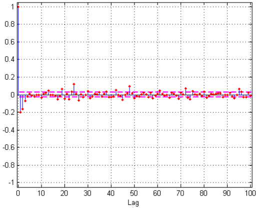
- **Sample autocorrelations.**
  - On the estimated residuals $(\widehat{\varepsilon}_t)$.

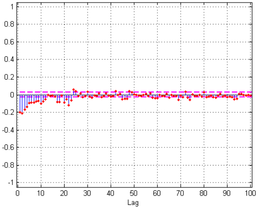


  - On the seasonally differenced residuals $(\nabla_{24}\widehat{\varepsilon}_t)$.



- **Sample autocorrelations.**
  - On the doubly differenced residuals ($\nabla \nabla_{24} \widehat{\varepsilon}_t$).





- **Identified models.**
  - SARIMAX$(p, 0, 0, r) \times (P, 1, Q)_{24}$ with $p \leq 5$, $r \leq 2$, $P \leq 1$ and $Q = 1$.
  - SARIMAX$(p, 1, q, r) \times (P, 1, Q)_{24}$ with $p \leq 1$, $q = 2$, $r \leq 2$, $P \leq 1$ and $Q = 1$.

- **Bayesian criteria.**
  - Akaike information criterion and Schwarz bayesian criterion,

$$\mathrm{AIC} = -2\log\mathcal{L} + 2k \quad \text{and} \quad \mathrm{SBC} = -2\log\mathcal{L} + k\log T$$

    where $\mathcal{L}$ is the model likelihood and k the number of parameters.
  - Log-likelihood.
  - Overall randomness of successive innovations.

- **Prediction criteria.**
  - We define C_A and C_R as follows,

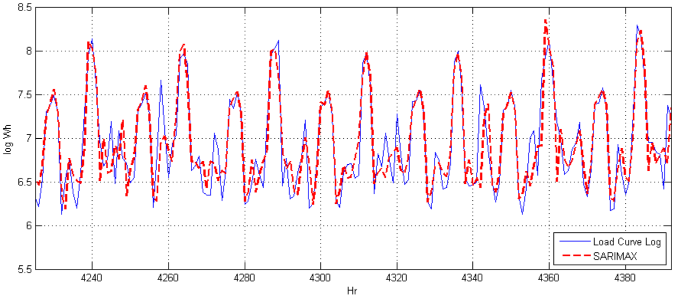
$$C_A = \frac{1}{NH}\sum_{k=1}^{NH}\left|\widetilde{C}_{T+k} - C_{T+k}\right| \quad \text{and} \quad C_R = \left(\sum_{k=1}^{NH} C_{T+k}\right)^{-1}\left(\sum_{k=1}^{NH}\left|\widetilde{C}_{T+k} - C_{T+k}\right|\right)$$

    where $(\widetilde{C}_{T+1}, \ldots, \widetilde{C}_{T+NH})$ are N consecutive predictions at horizon H from time T.

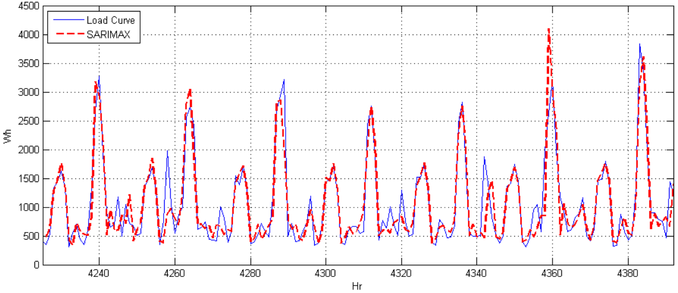
- **Selection on bayesian criteria.**

	p	*d*	*q*	*r*	*P*	*D*	*Q*	*s*	AIC	SBC	LL	VAR	WN
SARIMAX	1	0	0	1	0	1	1	24	-362.4	-330.4	186.2	0.053	
SARIMAX	3	0	0	1	0	1	1	24	-393.6	-348.8	203.8	0.053	✓
SARIMAX	5	0	0	1	0	1	1	24	-424.7	-367.2	221.4	0.053	✓
SARIMAX	3	0	2	1	0	1	1	24	-446.9	-389.4	232.5	0.052	✓
SARIMAX	3	0	2	2	0	1	1	24	-457.2	-393.3	238.6	0.052	✓
SARIMAX	0	1	1	1	0	1	1	24	100.4	132.3	-45.2	0.059	
SARIMAX	0	1	2	1	0	1	1	24	-238.7	-200.4	125.4	0.055	
SARIMAX	1	1	1	1	0	1	1	24	-389.5	-351.2	200.8	0.053	
SARIMAX	2	1	2	1	0	1	1	24	-435.1	-384.0	225.6	0.052	✓

- Modelling with SARIMAX$(3,0,2,2) \times (0,1,1)_{24}$ with $T = 730 \times 24$.



- **Selection on bayesian criteria.**
  - Modelling with SARIMAX$(3,0,2,2) \times (0,1,1)_{24}$ with $T = 730 \times 24$.



  - Least squares estimation, for all $28 \le t \le T$,

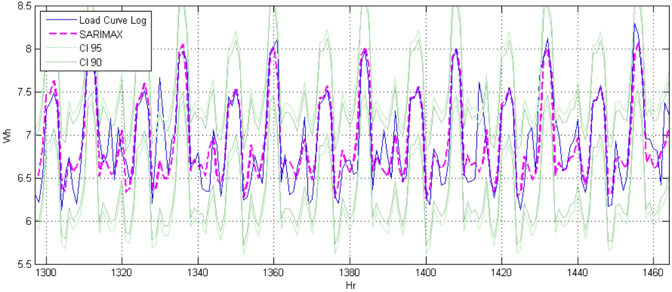
$$\begin{cases} C_t = \exp\left(\widehat{c}_0 + \widehat{c}_1 U_t + \widehat{c}_2 U_{t-1} + \varepsilon_t\right) - \exp(5), \\ \varepsilon_t = \varepsilon_{t-24} + \widehat{a}_1(\varepsilon_{t-1} - \varepsilon_{t-25}) + \widehat{a}_2(\varepsilon_{t-2} - \varepsilon_{t-26}) + \widehat{a}_3(\varepsilon_{t-3} - \varepsilon_{t-27}) \\ \qquad + (V_t - \widehat{b}_1 V_{t-1} - \widehat{b}_2 V_{t-2}) - \widehat{\beta}_1(V_{t-24} - \widehat{b}_1 V_{t-25} - \widehat{b}_2 V_{t-26}), \end{cases}$$

in which $\widehat{c}_0 = 7.9871$, $\widehat{c}_1 = 0.0166$, $\widehat{c}_2 = -0.0420$, $\widehat{a}_1 = 0.4776$, $\widehat{a}_2 = 0.9030$, $\widehat{a}_3 = -0.4305$, $\widehat{b}_1 = 0.0801$, $\widehat{b}_2 = -0.8524$, $\widehat{\beta}_1 = -0.8125$.

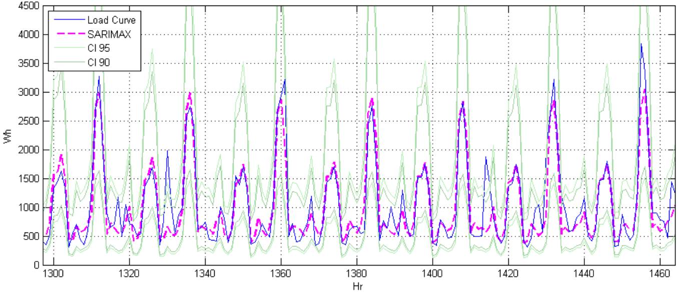
- **Selection on prediction criteria.**

	p	d	q	r	P	D	Q	s	C_A	C_R
SARIMAX	1	0	0	1	0	1	1	24	193.9	0.1814
SARIMAX	1	0	1	2	0	1	1	24	193.3	0.1808
SARIMAX	3	0	0	1	0	1	1	24	196.0	0.1833
SARIMAX	3	0	2	1	0	1	1	24	199.9	0.1870
SARIMAX	3	0	2	2	0	1	1	24	198.9	0.1861
SARIMAX	1	1	1	1	0	1	1	24	195.4	0.1828
SARIMAX	2	1	2	1	0	1	1	24	195.3	0.1828

- Forecasting with SARIMAX$(1,0,1,2)\times(0,1,1)_{24}$ with $T = 730 \times 24$.

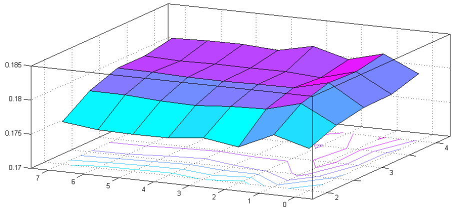


- **Selection on prediction criteria.**
  - Forecasting with SARIMAX$(1,0,1,2)\times(0,1,1)_{24}$ with $T = 730\times 24$.

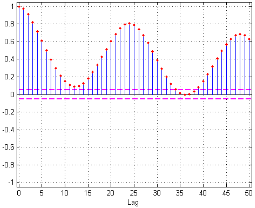


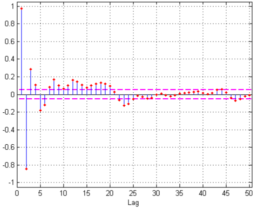
  - Parsimony is a central issue in time series analysis.
  - Results slightly improved with a sliding window of 2 months.
  - Around 2% of relative error between $(\widetilde{Y}_{T+1},\dots,Y_{T+NH})$ and $(Y_{T+1},\dots,\widetilde{Y}_{T+NH})$.

- **Refining...**
  - Influence of *r* and the size of the sliding windows *M* on C_R.



  - $M = 2$ months is the optimal sliding window.
  - No more influence of *r* as soon as $r \geq 2$.

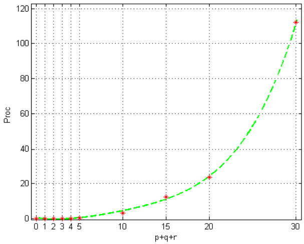


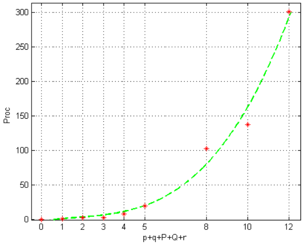


Plan

- **Huge dataset.**
  - More than 2000 load curves.
  - Around 70% nonthermosensitive.
  - High quality : more than 9 months of data per curve, very little missing values.

- **Technical problems.**
  - Exponential growth of computing time with parsimony.



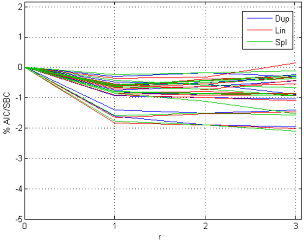


- **On a representative panel.**
  - Selection of 200 heterogeneous thermosensitive curves (size, peaks intensity, etc.)
  - Massive statistical KPSS procedures and visual first conclusions.
  - Bayesian criteria to select the best models **on average**.
  - Application to forecasting.
  - Prediction criteria to select the best models **on average**.
  - All parameters vary in their neighborhood.
  - Consider technical issues : a night of computation for some models.

- **2 more bayesian criteria.**
  - Reliability index.
  - Percentage of significance of the first exogenous coefficient.
  - Assess the relevance on large-scale.
  - Caution : main assumptions for t-test not satisfied !

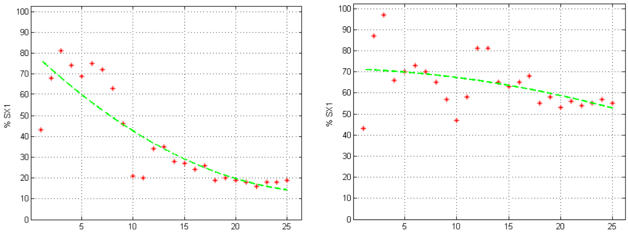
- **Stationarity, for $r = 1$ and $M = 3$ months.**
  - Less than 30% of $(\widehat{\varepsilon}_t)$ stationary.
  - 100% of $(\nabla_{24}\widehat{\varepsilon}_t)$ and $(\nabla\nabla_{24}\widehat{\varepsilon}_t)$ stationary.
  - SARIMAX$(p, 0, q, r) \times (P, 1, Q)_{24}$ and SARIMAX$(p, 1, q, r) \times (P, 1, Q)_{24}$.
  - Making r increase.

- **Relative evolution of AIC and SBC with r.**



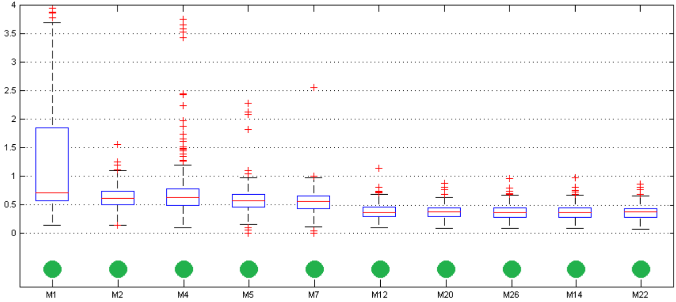
  - Clearly, $r = 1$.
  - SARIMAX$(3, 0, 2, 1) \times (0, 1, 1)_{24}$, SARIMAX$(3, 1, 2, 1) \times (0, 1, 1)_{24}$.
  - Gain over the naive model : $\approx 50\%$.
  - Gain over the SARIMA model : $\approx 2\%$.

- **Significance of c_1 for $M = 6$ months.**



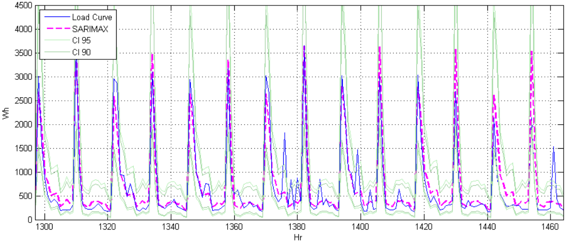
- Substantial on thermosensitive curves.

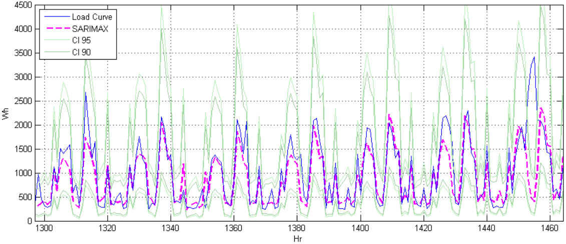
- **Evolution of C_R.**



- SARIMAX$(1,1,1,1) \times (0,1,1)_{24}$ and SARIMAX$(2,0,1,1) \times (0,1,1)_{24}$.
- $M = 2$ months.
- Gain over the naive model : $\approx$ 65%.
- Gain over the SARIMA model : $\approx$ 3%.

- **Examples.**


Load Curve
SARIMAX
CI 95
CI 90
Wh
Hr
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1300
1320
1340
1360
1380
1400
1420
1440
1460


Load Curve
SARIMAX
CI 95
CI 90
Wh
Hr
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1300
1320
1340
1360
1380
1400
1420
1440
1460

- **Nonthermosensitive curves.**
  - SARIMA$(1,0,1) \times (0,1,1)_{24}$
- **Thermosensitive curves.**
  - SARIMAX$(1,1,1,1) \times (0,1,1)_{24}$ and SARIMAX$(2,0,1,1) \times (0,1,1)_{24}$.
- **A careful study curve by curve would provide better results.**
  - Of course...
- **Caution : technical issues for huge datasets.**
  - Is a computation time $\times 1000$ for a gain of 0.1% relevant ?
  - Engineering approach, corporate vision.
- **Thank you for your attention.**
- **Comments or questions ?** ☺