Nonparametric forecasting of the French load curve

An overview

Vincent Lefieux

RTE & UPMC-ISUP

ISNPS 2014, Cadix

June 15, 2014

Table of contents

Electrical context

- Generation must be strictly equal to consumption at all times
- Electricity cannot be stored on an industrial scale, and demand fluctuates heavily.
- The French Electricity Transmission System Operators (RTE) shall ensure the balance of electricity flows on the network at all times, and an optimum management of electricity flows across the network.
- There are lot of forecasting needs: consumption, electricity losses, market prices, wind and solar generation, from short (up to 7 days) to long (few years) term, at different scales.
- The daily coordination is facilitated by having short term demand-supply balance predictions on a day ahead horizon.

Short term load curve forecasting

- The French global electricity consumption (“load”) increases in winter due to the electrical heating, and also in summer due to the air-conditioning.
- RTE drafts its load curve forecasting taking into account historical consumption patterns, weather forecast and daily pricing information.
- The dispatchers make the final load balancing decisions, taking into account the most recent information, including unexpected modifications of consumption patterns (e.g. strikes, national sporting events) and of generation (e.g. thermal generation unavailabilities, wind and solar generation variations).
- Modifying forecasting (decision-support) tools for a TSO is very sensitive and costly.

Statistics context

Actually, RTE uses a 3 steps procedure:

- Step 1 RTE “corrects” the half hourly load curve by modeling the impact of climate (temperature and cloud cover measured at 32 weather stations) and prices, in order to work on a time series that doesn't depend on exogenous variables. This step is done by using a regression model with dependent variables based on climate and tariff. We denote the corrected series by Z_t.
- Step 2 RTE uses a SARIMA model to forecast Z_t at the horizon H: $\widehat{Z}_{t+H}$.
- Step 3 RTE adds the forecasts given in Step 2 with the estimation given by the regression model using prices and forecasts for the temperature and cloud cover.

Electrical and statistics challenges

- Modeling problematics
  - Parametric paradigm: rigid and not adapted to the structure of some time series.
  - Nonparametric paradigm: curse of dimensionality, interpretability.
- Big data. . . and huge variability (smart grids context)

MAVE modeling

Xia *et al.* (2002) proposed the Minimum Average (conditional) Variance Estimation:

$$Y = g\left(B^T X\right) + \varepsilon$$

where :

- $Y \in \mathbb{R}$ is the response variable
- $X \in \mathbb{R}^p$ are p covariates
- ε is a random variable in $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$.
- $\mathbb{E}(\varepsilon / X) = 0$ almost surely
- $g : \mathbb{R}^D \to \mathbb{R}$ is the unknown link function
- $B = \left(\beta^1, \ldots, \beta^D\right)_{p \times D}$, with $D \in \{1, \ldots, p\}$, is an unknown matrix such $B^T B = I_D$

MAVE estimation

Given D, Effective Dimension Reduction (EDR) B is solution of :

$$\min_{B/B^TB=I_D} \mathbb{E}\left[\left(Y - \mathbb{E}\left(Y \middle/ B^TX\right)\right)^2\right]$$

Using the conditional variance of Y given B^TX, $\sigma_B^2\left(B^TX\right)$, minimisation problem is equivalent to :

$$\min_{B/B^TB=I_D} \mathbb{E}\left[\sigma_B^2\left(B^TX\right)\right]$$

With a local linear expansion, one can obtain finally:

$$\min_{B/B^TB=I_D,\left(a_j,b_j\right)_{j\in\{1,\dots,n\}}} \sum_{j=1}^{n}\sum_{i=1}^{n}\left\{Y_i - \left[a_j + b_j^T B^T\left(X_i - X_j\right)\right]\right\}^2 \omega_{ij}$$

where $\left(\omega_{il}\right)_{i\in\{1,\dots,n\}}$ are weights between X_i and X_l.

MAVE conclusions

- Some computational (iterative algorithm for weights) and specification difficulties (correlated data).
- Tests of a partially linear MAVE modeling.
- No good enough results to continue the project.
- 2 alternatives: IBR & sparse regression.

IBR modeling

Joint work with P-A. Cornillon (Rennes 2), N. Hengartner (LANL), Eric Matzner-Løber (Rennes 2)

Suppose the data $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$ are related via the following regression model:

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n$$

where the errors are independent of all the covariates.

It can be rewritten in vector form:

$$Y = m + \varepsilon$$

where $Y = (Y_1, \ldots, Y_n)^t$, $m = (m(X_1), \ldots, m(X_n))^t$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^t$

Linear smoothers

Linear smoothers, which depend on a tuning parameter λ (trade-off between the smoothness of the estimate and the goodness-of-fit), can be written as:

$$\widehat{m} = S_\lambda(X)Y,$$

where $S_\lambda(X)$ is an $n \times n$ smoothing matrix and $\widehat{m} = \widehat{Y}$ denotes the fitted values.

The IBR procedure

Step 1 The pilot smoother is applied to the data (oversmoothing the data):

$$\widehat{m}_1 = S_\lambda(X)Y.$$

Step 2 The bias $\mathbb{E}(\widehat{m}_1|X) - m = (S_\lambda(X) - I)m$ is estimated by replacing m by a smooth linear estimate (possibly using the same pilot smoother). The initial estimator is then corrected by removing the estimated bias:

$$\widehat{m}_2 = [S_\lambda(X) + S_\lambda(X)(I - S_\lambda(X))]Y.$$

Step k The bias estimation and bias correction steps can be iterated to generate a sequence of bias corrected smoothers, which gives at step k

$$\widehat{m}_k = [I - (I - S_\lambda(X))^k]Y.$$

The number of iterations' choice

The number of iterations is obtained by a Generalized Cross Validation criterium:

$$GCV(k) = \log \widehat{\sigma_k}^2 - 2 \log \left(1 - \frac{\mathrm{tr}([I - (I - S_\lambda(X))^k])}{n}\right),$$

where $\widehat{\sigma_k}^2$ corresponds to the estimated variance of the current residuals at step k.

Practical Implementation

Based on half-hourly data $(Z_1, \cdots, Z_T)$ (T corresponds to 12am), we use the following model to predict $Z_{T+25}, \cdots, Z_{T+72}$:

$$Z_{(T-48i)+25} = f(Z_{T-48i}, \ldots, Z_{T-48i-p}) + \varepsilon \quad i > 0$$

where p is the memory (48 or 96).
We might consider also different models for any $H \in [25, \cdots, 72]$:

$$Z_{(T-48i)+H} = f_H(Z_{T-48i}, \ldots, Z_{T-48i-p}) + \varepsilon_H \quad i > 0. \quad (1)$$

The IBR procedure was applied using the R package **ibr** (available on CRAN).

IBR conclusions

- Some good empirical results (e.g. in 2010, the IBR MAPE is 0.98% against 1.12 % for SARIMA).
- But for the moment no exogenous variable included.

Electrical Consumption Time series

Joint work with M. Mougeot (Paris 7), D. Picard (Paris 7), K. Tribouley (Paris 7), L. Maillard-Teyssier (RTE)

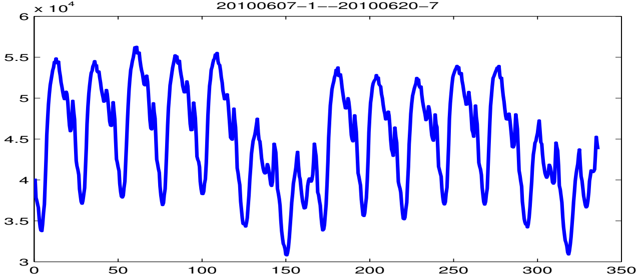


Figure : Two weeks of electrical consumption

Intraday load curves

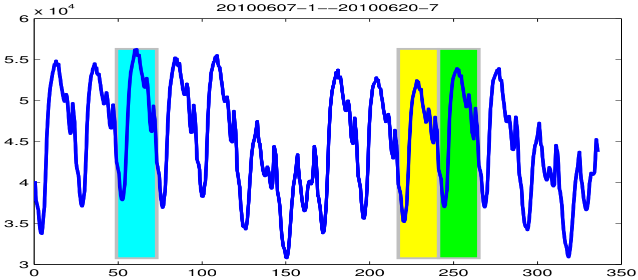
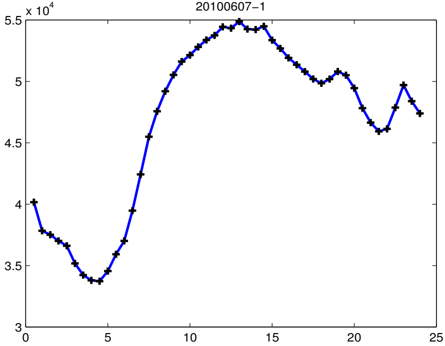


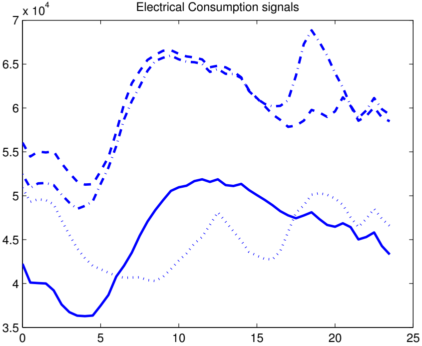
Figure : Functional data, Intra day load curves

Intraday load curve



Intra day load curve, 30' sampling (48 pts),
$Y \in \mathbb{R}^{n=48}$ (Y_t, $1 \leq t \leq 2800$)

Intra-day load curves



Intraday load curves for some days.
2003-10-27: dashed dot line, 2003-08-28: solid line, 2003-01-01: dot line, 2003-04-10: dashed line.

Spot of Temperatures, Cloud Cover and Wind information

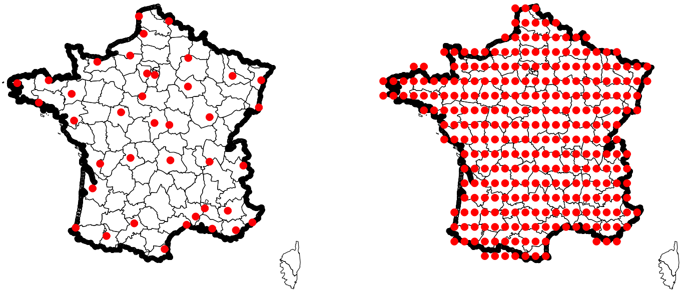


Figure : Temp., Cloud Cover spots (♯39) and wind data (♯293)

Modeling each signal as a function

We investigate the problem in a supervised learning setting.

- We consider each time unit signal

$$(U_i, Y_i),\ i = 1, \dots, n$$

- The generic consumption signal observed on the time unit:

$$Y_i,\ i = 1, \dots, n$$

- The design (here equi-distributed):

$$U_i = \frac{i}{n}$$

- We want to identify f (for each signal) in such a way that the model

$$Y_i = f(U_i) + \epsilon_i.$$

makes sense (has "small" errors ϵ_i's).

Using a dictionary

Consider a dictionary $\mathcal{D}$ of functions $\mathcal{D} = \{g_1, \ldots g_p\}$ and Assume that f can be well fitted by this dictionary

$$f = \sum_{\ell=1}^{p} \beta_\ell \, g_\ell + h$$

where h is a "small" function (in absolute value).
The model is

$$Y_i = \sum_{\ell=1}^{p} \beta_\ell g_\ell(U_i) + h(U_i) + \epsilon_i^{'}, \; i = 1, \ldots, n$$

which coincides with the linear model :

$$Y = X\beta + \epsilon \qquad \textit{with} \;\; X(n \times p)$$

putting $\epsilon_i = h(U_i) + \epsilon_i^{'}$ and $G_{i\ell} = g_\ell(U_i)$.

High dimensional framework

Solution: $\hat{\beta} = Argmin||Y - X\beta||^2$

- More variables than observations $n << p$

$$\begin{bmatrix} y_1 \\ y_2 \\ \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & \dots & & x_{1p} \\ & & & & \\ x_{n1} & & & \dots & x_{np} \end{bmatrix} * \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \dots \\ \\ \beta_p \end{bmatrix} + \epsilon$$

"Fat matrix"

$\rightarrow$ Infinity of $\hat{\beta}$ solutions.

$\rightarrow$ Need more assumptions on β to solve the problem

$\rightarrow$ e.g. Lasso, Ridge

Alternative procedure

- Learning Out of Leaders*: based on 2 Thresholding steps,
- weak complexity, sparse solution,
- Algorithm in 3 steps (normalized X column):

step		compute	size
1. SELECTION (threshold)	Find b Leaders $b < n << p$	X_b	(n, b)
2. REGRESSION	on Leaders	$\tilde{\beta} = (X_b^T X_b)^{-1} X_b^T Y$	$(1, b)$
3. THRESHOLD	the coefficients	$\hat{\beta}$	$(1, \hat{S})$

Generic Dictionary

For daily load curves, a good choice happened finally to be a mixture of the Fourier basis and the Haar basis ($p = 62$):

1. (1:1) constant function (1)
2. (2:24) cosine functions (with increasing frequencies) (23)
3. (25:47) sine functions (with increasing frequencies)(23)
4. (48:62) Haar functions (with increasing frequencies)(15)

November 18^{th} 2007

$S = 12$, $MAPE = 0.0057 = 0,57\%$.
$MAPE = \frac{1}{n}\sum_i^{n=48} |Y_i - \hat{Y}_i|/Y_i$

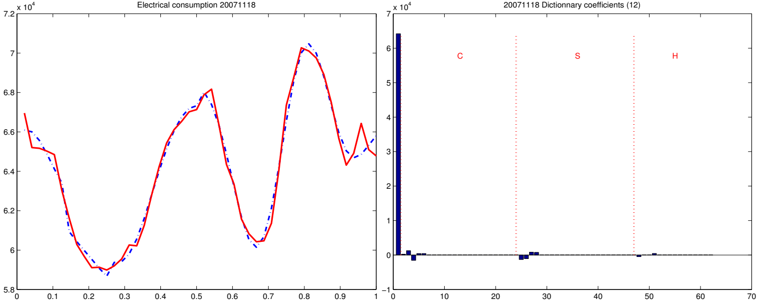


Figure : 2007 11 18

left: observed signal - red line, approximated signal -blue line
right: S coefficients on the dictionary

June 17^{th}, 2009

$S = 5$, $MAPE = 0.0147$

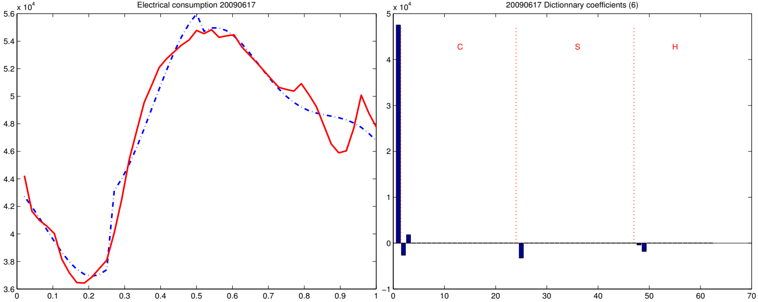


Figure : 2003 04 30

left: observed signal - red line, approximated signal -blue line
right: S coefficients on the dictionary

Segmentation of the intra-day load curves using Sparse Approximation on a Generic Dictionary

8 years of data: $T = 2800$ intra day load curves ($n = 48$)

- Using the sparse approximation (same support, $S = 8$)
- Using a clustering algorithm in 2 steps (k-means algorithm)
- Segmentation of the daily signals in clusters
- ...
- From Cluster to groups using calendar interpretation

Two step Clustering results

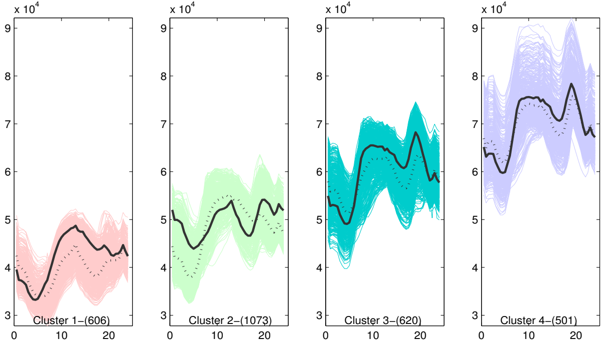
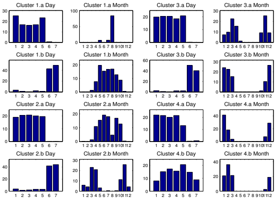


Figure : $T = 2800$ intra day load curves of size $n = 48$ (clustering using $S = 7$ approximated coefficients)

From clusters to groups

From clusters:



To groups: calendar interpretation of the clusters

	Months											
Days	1	2	3	4	5	6	7	8	9	10	11	12
1	7	8	5	3	3	3	3	1	3	3	5	7
2	7	8	5	3	3	3	3	1	3	3	5	7
3	7	8	5	3	3	3	3	1	3	3	5	7
4	7	8	5	3	3	3	3	1	3	3	5	7
5	7	8	5	3	3	3	3	1	3	3	5	7
6	6	8	4	4	2	2	2	2	2	2	4	6
7	6	6	4	4	2	2	2	2	2	2	4	6

Intraday Specific Dictionary

Each day t,

- $\boxed{Y_t = X_t \beta_t + \epsilon_t}$
- With Dictionary of p functions $\mathcal{D}_t = \{g_1^t, \ldots g_p^t\}$
  Final model, $p = 10$ ($p = 14$)
  1. **2**, Shape fonctions (group centroid, previous week day Y_{t-7})
  2. **8**, Climate fonctions (Temperature and Cloud Cover Indicators computed over the 39 meteorological spots. (and Wind...(+4))

Forecast: Plug in estimated coefficients $\mathcal{M}(t)$, with $\mathcal{M}$ expert

Conclusion

- Universal approach for functional data (with "intra day" pattern)
- Sparse approximation using
  - A Generic dictionary for compression and pattern extraction
  - Intra day specific dictionaries for approximation and prediction
- Forecasting
  - Various experts for prediction
  - Agregation using exponential weights
- Competitive approach compared to usual time serie analysis with much less parameters.

Specialized Experts focus on

1. (2) Time depending (t-1, t-7)
2. (2) climatic configuration of the day (Temperature)
3. (2) constrained climatic configuration of the day (Temperature/Cloud Covering)
4. group constraint climatic configuration of the day (Temperature/group)
5. climatic configuration of the day constrained by the type of the day (Temperature/day)
6. climatic configuration of the day constrained by a calendar group (Temperature/calendar)
7. climatic configuration of the day (Nebulosity)
8. group constraint climatic configuration of the day (Cloud Covering/group)
9. climatic configuration of the day constrained by the type of the day (Cloud Covering/day)
10. climatic configuration of the day constrained by a calendar group (Cloud Covering/calendar)