Forecasting the french electricity consumption

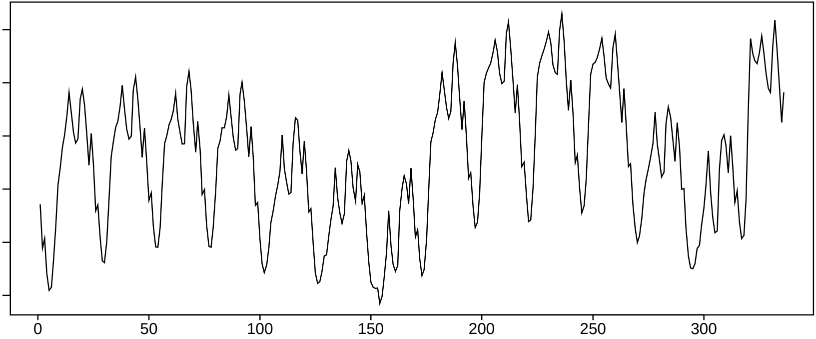
E. Matzner-Løber (eml@uhb.fr)

in collaboration with P-A. Cornillon, N. Hengartner and V. Lefieux

CNLS 2010

Goal

Every day at noon, the FTS operator has to forecast the next hours of the french electricity consumption for the electricity balance.



At any time the production should equal the consumption.

Data

Data available :

- consumption every half hour
- temperature every half hour
- type of the day : week day, week end
- holliday...
- and in a recent futur price ?

Actually

Different methods

- Parametric such as ARIMA-SARIMA models
  - 20 years old
  - more and more complicated
  - become a huge black box

New methodology

- to compare with existing one
- to simplify if possible
- to build local forecasting models

Régression

Transform the time series problem in a regression problem

$$C_t = m(C_{t-1}, \dots, C_{t-??}, ??, ??) + \varepsilon$$

Questions :
how to estimate which variables to use ?
how to estimate the unknown function mapping $\mathbb{R}^?$ to $\mathbb{R}$.

Non parametric regression model

$(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}$ n pairs of observations

$$Y_i = m(X_i) + \varepsilon_i$$

$$Y = m + \varepsilon.$$

Main goal : estimate m unknown function.

Smoothing

We estimate m non-parametrically (smoothing) :

$$\hat{m} \quad = \quad S_\lambda Y$$

where S_λ is the smoothing matrix and λ is the smoothing parameter (size of the bin, bandwidth, penalty...).

Classical smoother

- moving average $S_{ij} = 1/\text{nbr de } X$
- Bin $S_{ij} = 1/\text{nbr de } X$ in the bin
- Kernel $S_{ij} = K_h(X_i - X_j)/\sum_l K_h(X_i - X_l)$
- $k$nn $S_{ij} = 1/K$ si $X_j \in kpp(X_i)$
- Regression spline $S = B(B'B)^{-1}B'$
- Smoothing spline $S = N(N'N + \lambda\Omega_N)^{-1}N'$

Main idea

Assume λ big, so that the smoother is very smooth ; then

- estimate the bias
- correct the previous smoother

and iterate.

Bias of a linear smoother

We choose a smooth pilot S_1. The estimation is

$$\hat{m}_1 = S_1 Y.$$

The bias is

$$B(\hat{m}_1) = \mathbb{E}[\hat{m}_1|X] - m = (S_1 - I)m$$
$$= -\mathbb{E}[(I - S_1)Y] = -\mathbb{E}R_1.$$

Estimating the bias

- plug-in

$$\hat{b}_1 \quad = \quad (S_1 - I)S_2 Y$$

- residuals

$$\tilde{b}_1 \quad = \quad -S_2 R = S_2(I - S_1)Y.$$

Bias correction

The corrected estimators are

- plug-in

$$\begin{aligned}\hat{m}_k &= S_1 Y + (I - S_1) S_2 Y + \cdot + (I - S_1)(I - S_2) \cdots S_k Y \\ &= [I - (I - S_1)(I - S_2) \ldots (I - S_k)] Y.\end{aligned}$$

- residuals

$$\begin{aligned}\widehat{m}_k &= S_1 Y + S_2 (I - S_1) Y + \cdots + S_k (I - S_{k-1}) \cdots (I - S_1) \\ &= [I - (I - S_k)(I - S_{k-1}) \cdots (I - S_1)] Y.\end{aligned}$$

Bias correction

If the same S is used at each iteration

$$\begin{aligned} \widehat{m}_k &= S[I + (I - S) + (I - S)^2 + \cdots + (I - S)^{k-1}]Y \\ &= [I - (I - S)^k]Y. \end{aligned}$$

Predictive smoother

In order to evaluate the iterated smoother at any point, write

$$\begin{aligned}\widehat{m}_k &= S[I + (I - S) + (I - S)^2 + \cdots + (I - S)^{k-1}]Y \\ &= S\widehat{\beta}_k\end{aligned}$$

and

$$\widehat{m}_k(x) = S(x)^t \widehat{\beta}_k.$$

Theoretical properties

The squared bias and variance of the k^{th} iterated bias corrected smoother $\widehat{m}_k$ are

$$B^2(\hat{m}_k) = m^t \left((I - S)^k\right)^t (I - S)^k m$$

and

$$\mathrm{Var}(\hat{m}_k) = \sigma^2 (I - (I - S)^k) \left((I - (I - S)^k)\right)^t .$$

The behaviour of $\widehat{m}_k$ can be related to the spectrum of $I - S$.

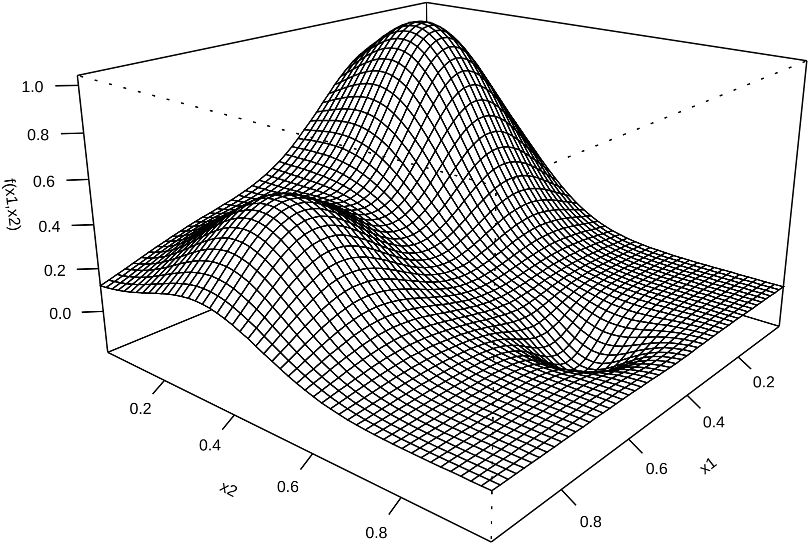
Thin plate splines

Given a smoothing parameter λ, the thin-plate smoother of degree ν_0 minimises

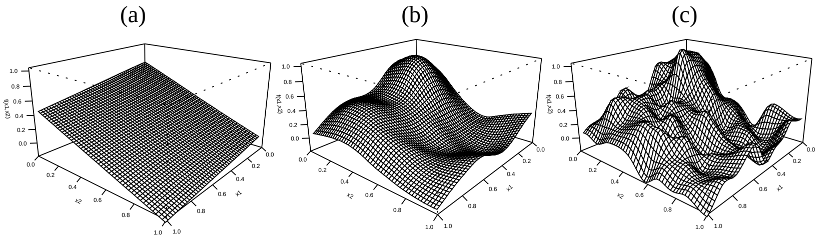
$$\sum_{i=1}^{n}(Y_i - f(X_i))^2 + \lambda \left[\sum_{\substack{i_1,\dots,i_d = 0 \\ i_1+\dots+i_d \leq \nu_0}} \int_{\mathbb{R}^d} \left| \frac{\partial^{i_1+\dots+i_d}}{\partial x_{i_1} \dots \partial x_{i_{\nu_0}}} f(x) \right|^2 dx \right].$$

TPS are attractive class of multivariate smoothers because the solution of is numerically tractable and the eigenvalues of the smoothing matrix are approximatively known.

Example



Example



Smoother with 0 (respectively 500 and 50 000) iterations.
All the smoothers are evaluated on a regular grid $[0,1] \times [0,1]$.
The sample size is $n = 100$.

Theoretical properties

Suppose $m \in \mathcal{H}^{(\nu)}$, Sobolev space. If the pilot smoother $\hat{m}_1 = SY$ is obtained with a TPS of degree ν_0, with $\lceil d/2 \rceil \leq \nu_0 < \nu$ and with $\lambda_0 > 0$ independent with n,
then it exists an optimal number of iterations $k(n)$ such that

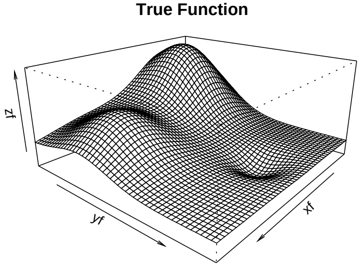
$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n}(\hat{m}_k(X_j) - m(X_j))\right)^2\right] = O\left(n^{-\frac{2\nu}{2\nu+d}}\right),$$

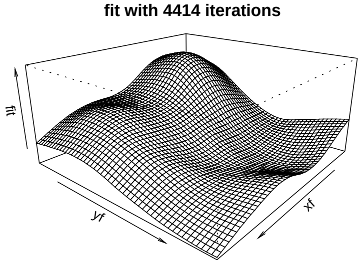
Automatic choice of k

Let $\hat{k}_{GCV} \in \mathcal{K}_n = \{1, \ldots, n^{1+\gamma}\}$, $\gamma \geq 0$, suppose that ε is Gaussian, then

$$\frac{\|\hat{m}_{\hat{k}_{GCV}} - m\|^2}{\inf_{k \in \mathcal{K}_n} \|\hat{m}_k - m\|^2} \longrightarrow 1, \quad \text{in probability.}$$

TP smoother






```
Residuals:
      Min        1Q    Median        3Q       Max
-0.234967 -0.068325 -0.007392  0.068939  0.301346
Residual standard error: 0.1197 on 73.5 degrees of freedom

Initial df: 3.03 ; Final df: 26.52

Number of iterations: 4414 chosen by gcv
Base smoother: Thin plate spline of order 2 (with 3.03 df)
```

Practical consideration

The number of eigenvalues of $(S_\lambda) = 1$ is

$$\binom{\nu_0 + d - 1}{\nu_0 - 1}$$

where $\nu_0 = \lfloor d/2 \rfloor + 1$.
If $d = 5$, the number of eigen values equal to one is 21.
If $d = 8$, the number of eigen values equal to one is 495.

Kernel smoothers

The general form of a kernel smoother is

$$S_{ij}(X,h) = K(d_h(X_i,X_j))/\sum_k K(d_h(X_i,X_j))$$

and with a Gaussian kernel

$$S_{ij}(X,h) = \frac{\exp -.5\left[(X_{i1}-X_{j1})^2/h_1^2+\ldots+(X_{id}-X_{jd})^2/h_d^2\right]}{\sum_k K(d_h(X_i,X_j))}.$$

Write

$$S = DS_S$$

where S_S is symetric.

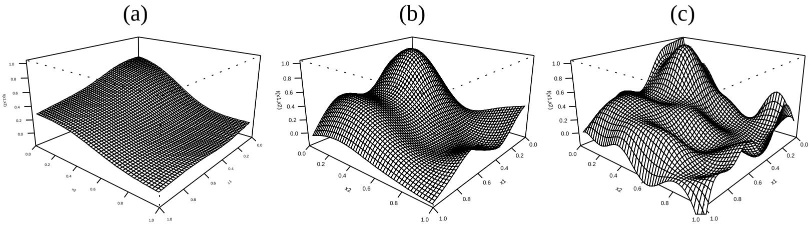
Transformation

$$\begin{aligned}\hat{m}_k &= [I-(I-S)^k]Y \\ &= [I-(D^{1/2}D^{-1/2}-D^{1/2}D^{1/2}S_sD^{1/2}D^{-1/2})^k]Y \\ &= D^{1/2}[I-(I-A)^k]D^{-1/2}Y.\end{aligned}$$

Theorem

If the Fourier transform of K is a finite positive measure then the spectrum of A is $[0,1]$.

Example Gaussian kernel



Smoother with 0 (respectively 50 and 1 000) iterations.
All the smoothers are evaluated on a regular grid $[0, 1] \times [0, 1]$.
The sample size is $n = 100$.

Simulation study

We will compare function **ibr** (R package **ibr**) with :

- Additive models (Hastie & Tibshirani, 1995), R package **mgcv**
- MARS (Friedman, 1991) R package **mda**
- Projection pursuit (Breiman & Friedman, 1985) function **ppr**
- L2-Boosting (Bühlmann & Yu, 2003) R package **mboost**.

Simulations

	gam	gamb	mars	N1.1	Tps
$\sin(5\pi x)$	1.13	1.29	66.42	1.00	1.21
	1.32	1.39	74.31	1.00	1.32
	1.24	1.49	91.03	1.00	1.05
$2(x_1-0.5)^2+\exp(-\frac{(x_2-0.3)^2}{0.09})$	1.00	1.06	2.39	1.68	2.03
	1.00	1.31	2.10	3.27	2.07
	1.01	1.00	1.95	4.10	2.12
$10x_1^2+\exp(2x_2)\{x_1<0.5\}+\exp(2x_2)$	2.98	3.14	3.41	1.55	1.00
	3.72	3.68	3.82	1.76	1.00
	3.38	3.36	3.66	1.71	1.00
$x_1x_2^2$	21.33	19.48	21.82	1.00	1.49
	33.51	34.00	35.12	1.00	1.38
	80.93	81.04	81.42	1.00	2.01
$10\sin(\pi x_1x_2)+20(x_3-0.5)^2+10x_4+5x_5$	1.00	1.61	1.28	1.16	1.35
	1.43	1.64	1.64	1.00	1.21
	3.22	3.47	3.70	1.00	1.75
$\exp(-\frac{(-2(x_1.5)+2(x_2-.5)-2(x_3-.5)-x_4-x_5+1)^2}{15})$	25.08	22.68	26.83	1.42	1.00
	47.61	48.56	52.07	1.16	1.00
	73.34	69.78	76.08	1.21	1.00

A real example : Los Angeles Ozone Data

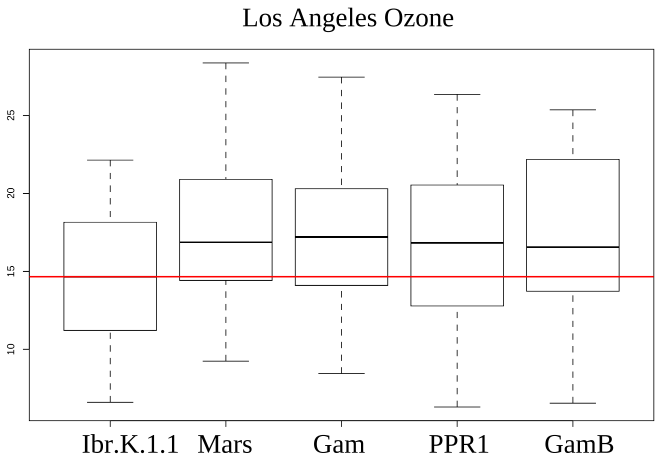
The sample size is $n = 330$ and $d = 8$ explanatory variables :

```
"Pressure.Vand"  "Wind"  "Humidity"
"Temp.Sand"  "Inv.Base.height" "Pressure.Grad"
"Inv.Base.Temp" "Visilibity"
```

Simulations

Method	MPSE
Multivariate regression	19.53
L_2Boost with component-wise spline	17.23
additive model (backfitted with R)	17.44 (67)
Projection pursuit (with R)	16.97 (1)
MARS (with R)	17.49
IBR with GCV stopping rule	
and multivariate Gaussian kernel with	
1.05 initial DDL per variable and **257** iterations	14.85
1.1 initial DDL per variable and **57** iterations	**14.83**
1.2 initial DDL per variable and **13** iterations	14.89

Comparing relative prediction mean square errors



Going back to the problem

We wish to predict a time series $\{C_t\}_{t=1}^{n}$:

$$C_H = m(C_{H-1}, C_{H-2}, \ldots, C_{H-24}) + \varepsilon$$

$$C_H = m_H(C_{H-1}, C_{H-2}, \ldots, C_{H-24}) + \varepsilon_H$$

where m or $m_H : \mathbb{R}^{24} \mapsto \mathbb{R}$ are unknown.

Simplified model

Forecast at step 1 :

$$\hat{C}_{12+1,j} = \hat{m}(C_{12,j}, C_{11,j}, \ldots, C_{12,j-1}).$$

and replace

$$\hat{C}_{14,j} = \hat{m}(\hat{C}_{13,j}, Z_{12,j}, \ldots, Z_{13,j-1})$$

and so on for the next 36 hours.

Different models

We estimate different functions for the different forecasts

$$\hat{C}_{12+1,j} = \hat{m}_1(C_{12,j}, C_{11,j}, \ldots, C_{12,j-1}).$$

and

$$\hat{C}_{12+2,j} = \hat{m}_2(C_{12,j}, C_{11,j}, \ldots, C_{12,j-1}).$$

and the temperature

$$C_H = m(C_{H-1}, C_{H-2}, \ldots, C_{H-24}, \hat{T}_H, \ldots, T_{H-24}) + \varepsilon$$

$$C_H = m_H(C_{12}, C_{11}, \ldots, C_{-12}, \hat{T}_H, \ldots, T_{H-24}) + \varepsilon_H$$

where m or $m_H : \mathbb{R}^{48} \mapsto \mathbb{R}$ are unknown.

Results for one year

We evaluate the MAPE

$$\mathrm{MAPE}(C_t) = \frac{1}{T}\sum_{t=1}^{T} |C_t - \hat{C}_t|/C_t$$

and get for one year
1.5 % for IBR
1.7 % for the existing SARIMA model.

work in progress

Selected the lagged "right" variables in consumption and in temperature to get the best models. About IBR

- Simple idea
- Good theoretical properties
- Nice practical results
- Easy to use : R package **ibr** on CRAN