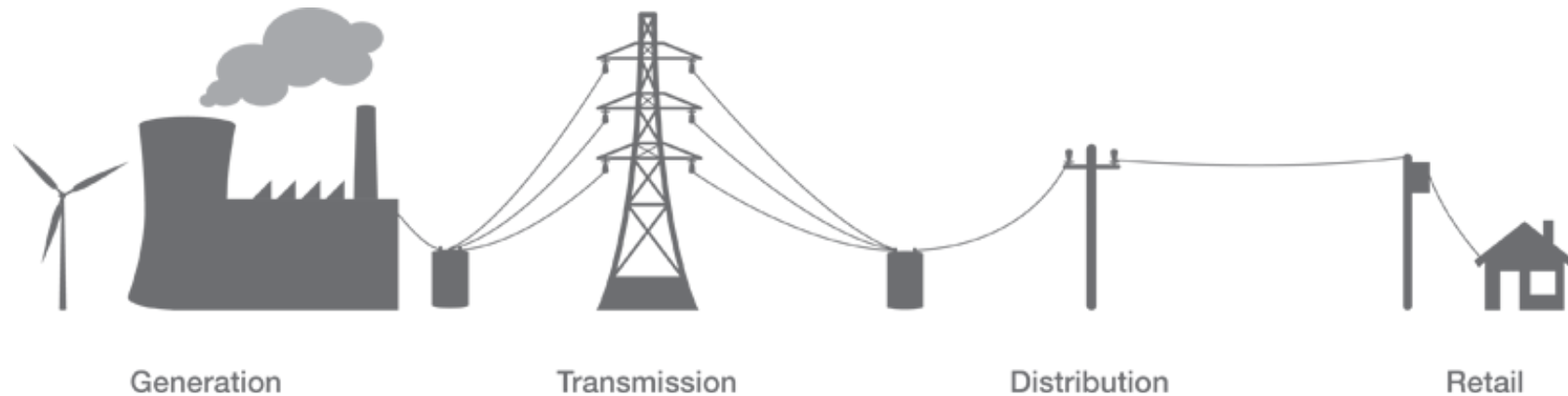


# Méthodes de Prévision - Production & Consommation Électrique

Arthur Charpentier, 2017

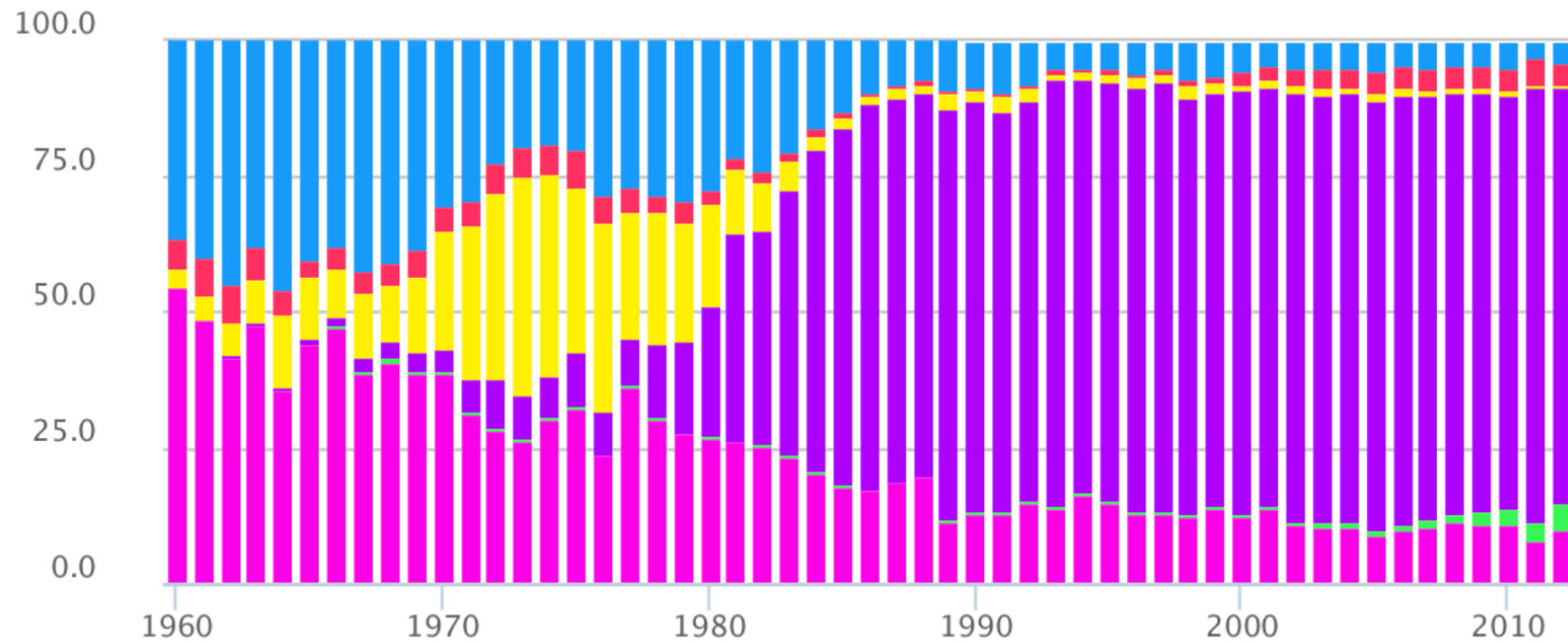
## Production d'électricité (en France)



- Production (**concurrence**) : EDF (**Électricité de France**), Electrabel-Énergie du Rhône, éolien, etc

En 2015, production française = 545 TWh (nucléaire 77.0%, hydrolique 10.7%, thermique (gaz, charbon, fioul) 7.3%, éolien 3.7%, photovoltaïque 1.3%)

## Production d'électricité (en France)



Coal (4.08%) Oil (0.58%) Non-hydro renew. (4.47%) Natural gas (3.69%) Nuclear (76.6%) Hydro (10.2%)

Source World Bank

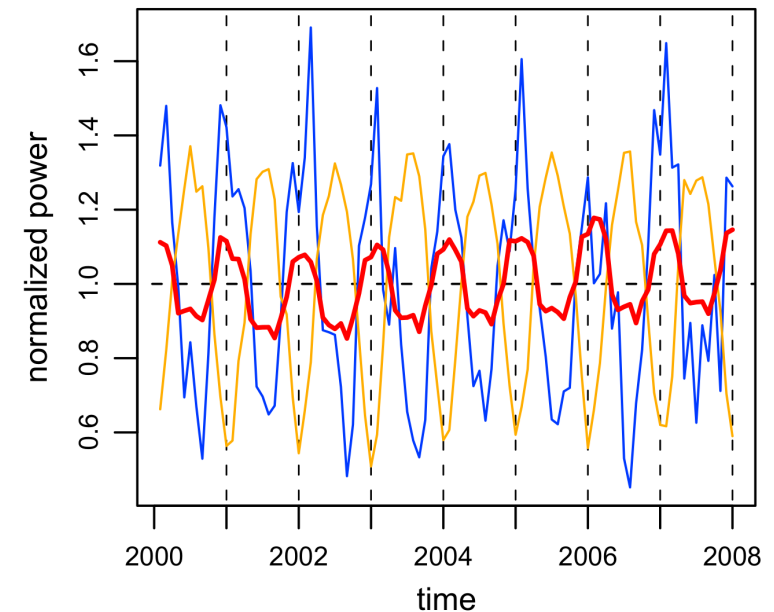
## Production d'électricité renouvelable

Un parc d'éolennes (entre 3 et 100) peut être très long à construire.

- $\sim 10$  ans de prospection
- $\sim 1$  an de mesure *in situ* pour estimer le potentiel
- construction du parc
- production d'électricité pendant  $\sim 25$  ans
- démantèlement

La production d'énergie renouvelable (éolien et solaire) est très sensible aux conditions météorologiques.

La courbe agrégée (et normalisée) est en rouge, via Tankov (2016) et Heide *et al.* (2010)

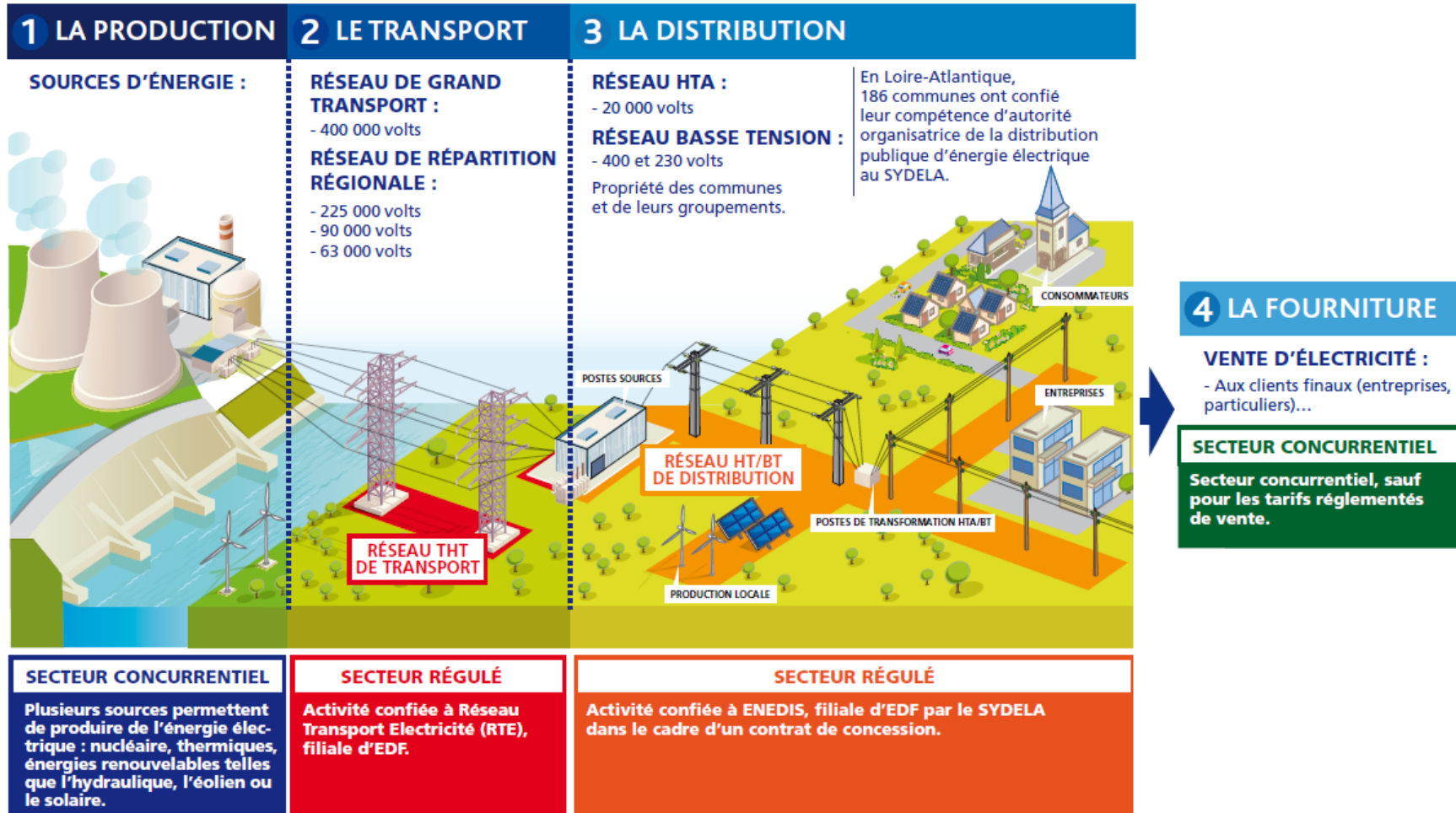


## Le réseau de transport d'électricité

- Transport ( $\sim$ monopole) RTE, Réseau de transport d'électricité, au delà de 50kV.
- Distribution (monopole) Enedis, anciennement ERDF, Électricité Réseau Distribution France
- Fourniture (concurrence) : Entreprises intégrées (producteurs, EDF) ou négociants ( Direct Énergie-Poweo, Enercoop, Planète Oui, etc).

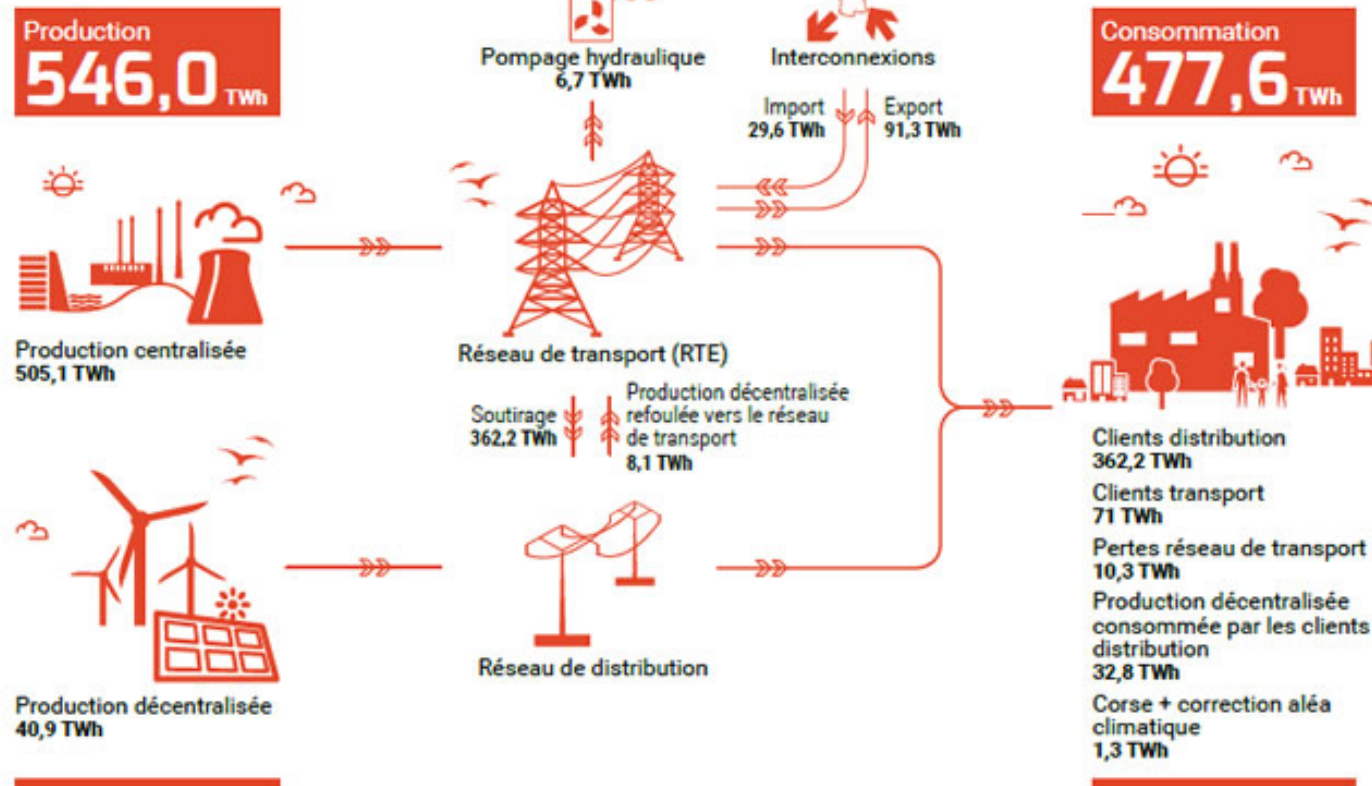
En 2015, consommation française = 475 TWh

# Le réseau de transport d'électricité



## Production et consommation d'électricité

\_ ÉLECTRICITÉ : DE LA PRODUCTION À LA CONSOMMATION



Source : CRE Rapport d'activité 2015  
15 ans de régulation

Source Commission de Régulation de l'Énergie (2015)

## Consommation d'électricité (en France)

Structure de la consommation en France :

- $1/3$  = quelques centaines de clients industriels
- $1/3$  = quelques dizaines de milliers de clients professionnels,
- $1/3$  = 26 millions de clients particuliers.

Informations à prendre en compte pour la prévision :

- variables calendaires et activité économique
- température, instantanée et passée,
- consommations récentes (prévision CT).

Prévision précise à quelques pourcents près, 1000MW en France.



## Réseau, production, consommation

L'électricité peut difficilement être stockée  
(la somme des flux entrants égale la somme  
des flux sortants, à chaque noeud)

Source carte de France: [data.gouv.fr](https://data.gouv.fr)



## Risque et Blackout

Dimanche 28 septembre 2003, Black Out en Italie (nuit Blanche à Rome)

défaillance d'une ligne à haute tension qui aurait lâché vers 3h00 du matin à la frontière italo-suisse.

Très loin de la saturation du réseau (45% de la demande de pointe), mais les capacités d'échange avec la Suisse l'étaient.

*"à 3h20, en pleine nuit, les autoroutes électriques traversant les Alpes auraient subitement cessé de fonctionner pour des raisons indéterminées"*.

le réseau italien a perdu les 6 000 mégawatts d'électricité importé de l'étranger. Des Alpes à la Sicile, tout l'Italie a été plongée dans le noir, à l'exception de la Sardaigne.

## Marché de l'électricité

Le prix d'un contrat à terme, d'horizon  $T$ , pour un sous-jacent valant  $S_t$  (aujourd'hui) est  $F_t(T) = e^{r(T-t)} S_t$ , par absence d'opportunité d'arbitrage.

Avec un coût (fixe par unité de temps) de stockage  $c$ , il devient  $F_t(T) = e^{(r+c)(T-t)} S_t$ .

Si les ventes à découvert sont impossibles, il existe un rendement de commodité  $y \geq 0$  et  $F_t(T) = e^{(r+c-y)(T-t)} S_t$

Comme on ne peut pas stocker l'électricité à grande échelle, l'horizon  $T$  est en réalité un intervalle de temps  $[T_1, T_2]$ , et on pose

$$F_t(T_1, T_2) = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} F_t(T) dT$$

## Marché de l'électricité

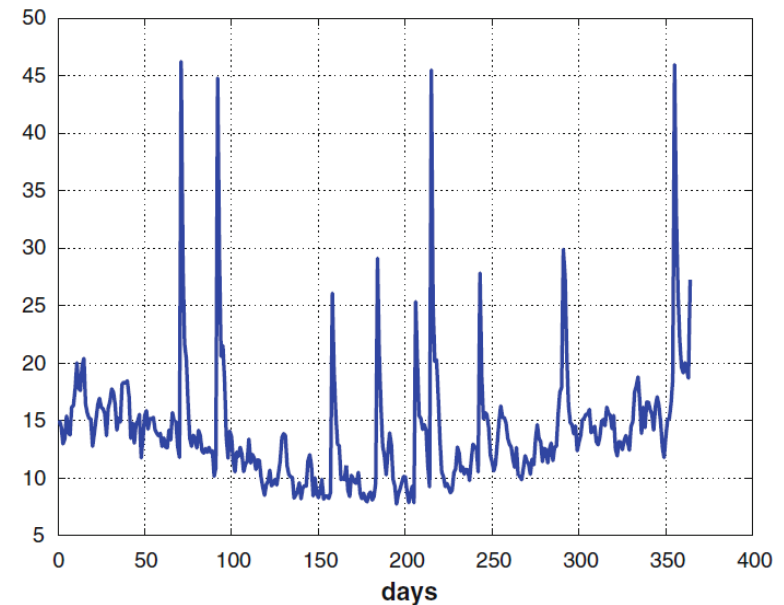
On cherche alors un modèle pour  $(S_t)$ ,  
cf [Cartea & Figueroa \(2005\)](#) et [Tankov \(2016\)](#)

$$\log S_t = g(t) + Y_t$$

où  $g(t)$  est la composante saisonnière et

$$dY_t = -\alpha Y_t dt + \sigma(t) dW_t + J dN_t$$

où  $(N_t)$  est un processus de Poisson, et  $J$  l'amplitude de sauts, supposés lognormaux.



On peut aussi utiliser des modèles structuraux pour expliquer le prix de fabrication d'énergie, cf [Posner \(2016\)](#)

En s'inspirant de [Barlow \(2002\)](#) et [Kanamura & Ohashi \(2004\)](#), on peut aussi modéliser la demande. On suppose que  $D_t = g(t) + X_t$  où  $g(t)$  est une composante saisonnière, et  $X_t$  la composante stochastique, vérifiant  $dX_t = (\mu - \lambda X_t)dt + \sigma dW_t$ . Le prix  $P_t$  associé à une demande  $D_t$  soit être non-linéaire.

## Marché de l'électricité

Barlow (2002) a proposé

$$P_t = (a_0 - b_0 D_t)^{1/\alpha}$$

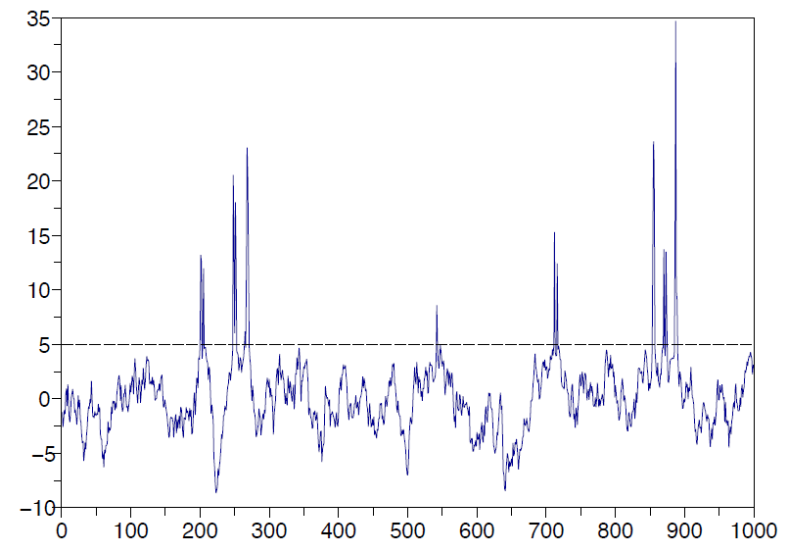
avec  $\alpha > 0$ .

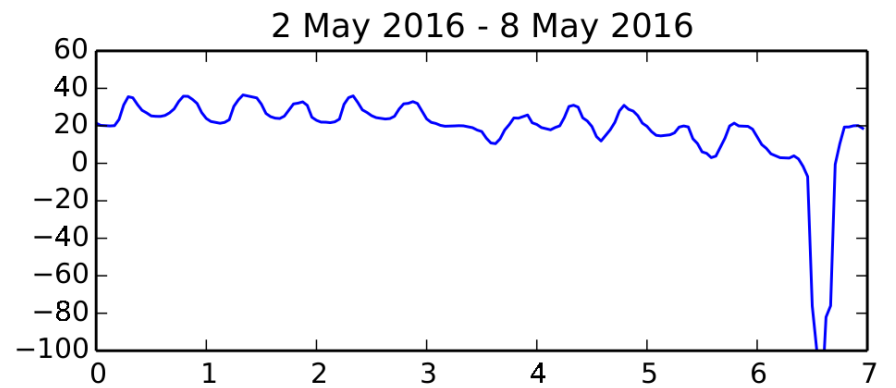
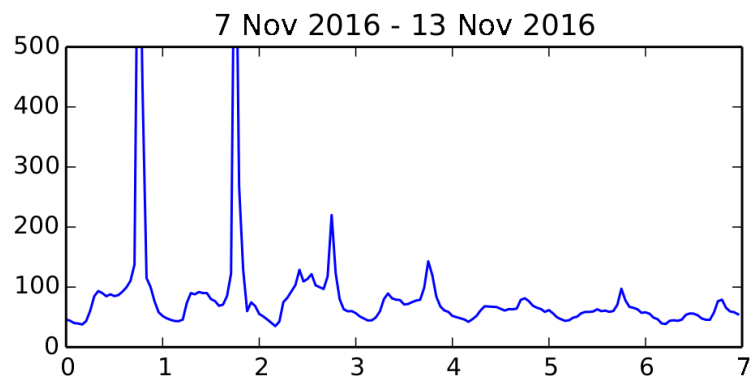
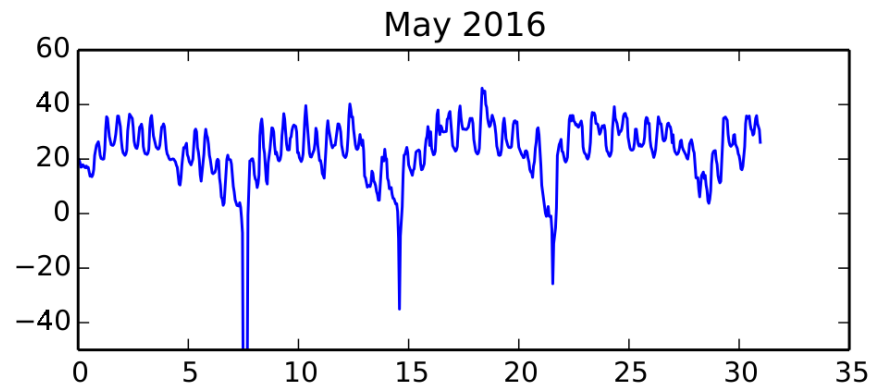
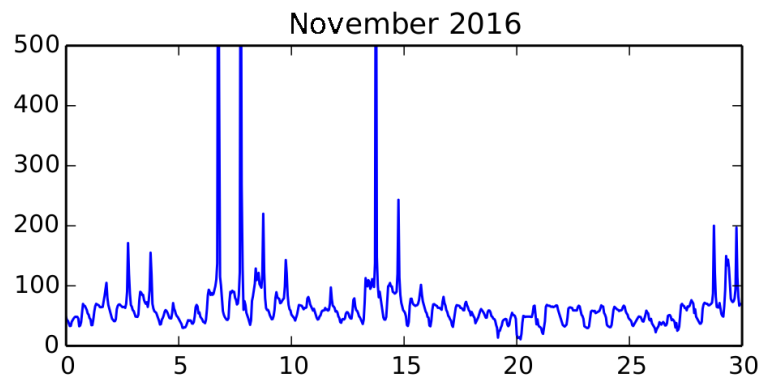
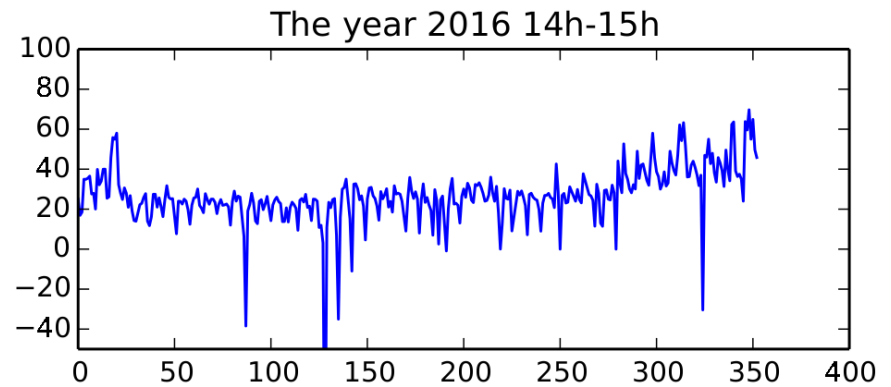
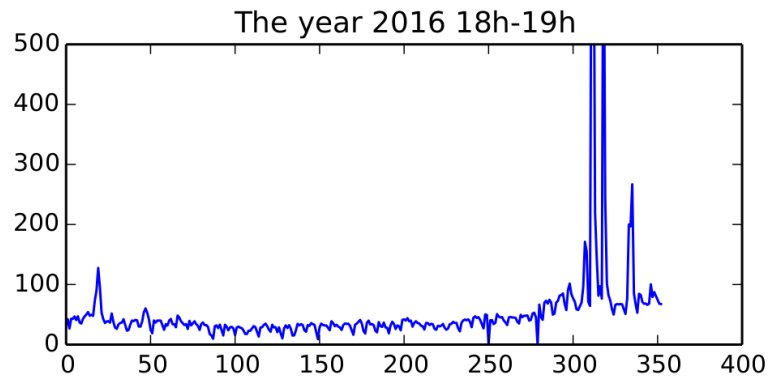
Kanamura & Ohashi (2004) a proposé

$$P_t = (a_1 + b_1 D_t) \mathbf{1}_{D_t \leq D_0} + (a_2 + b_2 D_t) \mathbf{1}_{D_t > D_0}$$

(*hockey stick*)

Pour les prix du marché spot de l'électricité, [transparency.entsoe.eu](https://transparency.entsoe.eu)





## Énergie(s) Renouvelable(s)

Pour estimer la production d'un parc d'éoliennes, il convient de modéliser correctement la vitesse du vent (ou mieux sa puissance).

Analyse spectrale de la vitesse du vent,

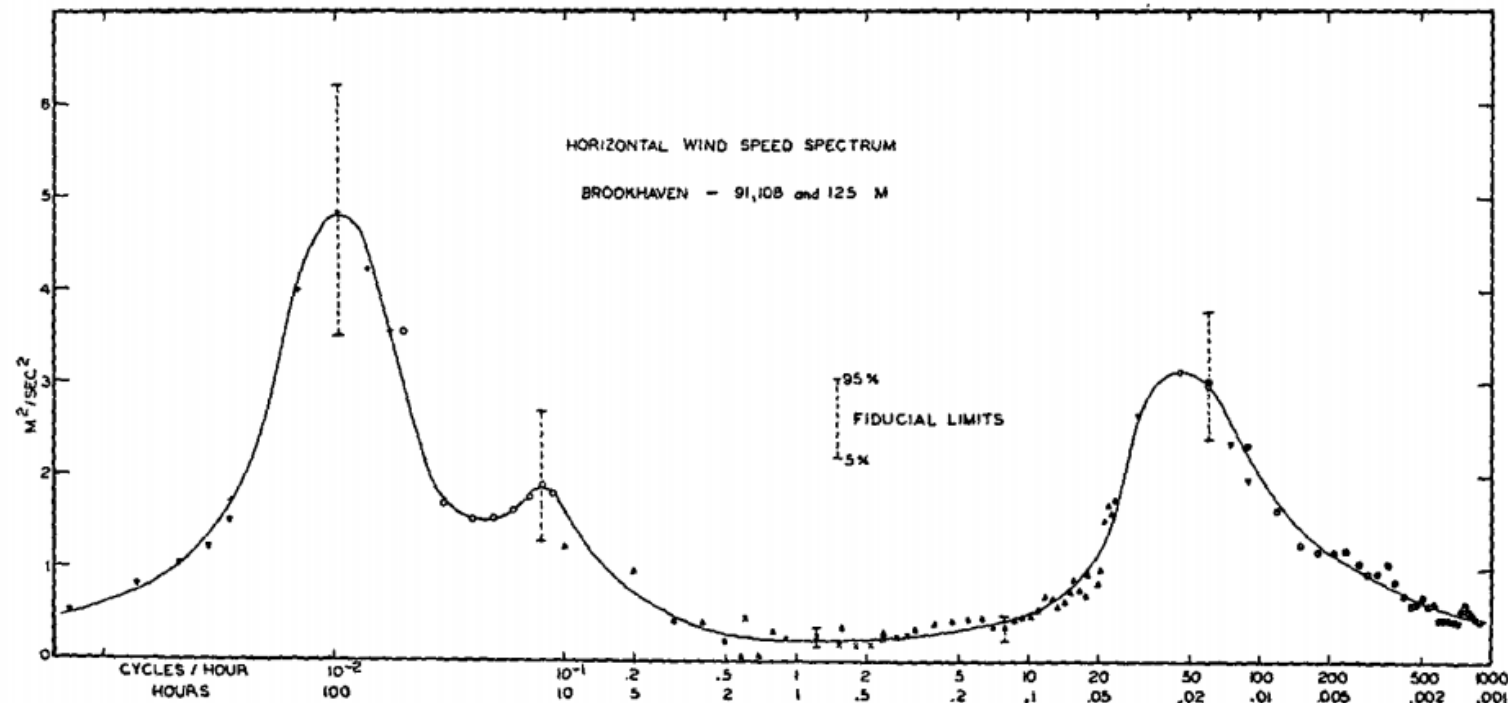


FIG. 1. Horizontal wind-speed spectrum at Brookhaven National Laboratory at about 100-m height. (See table 1 for date and time.)

## Énergie(s) Renouvelable(s)

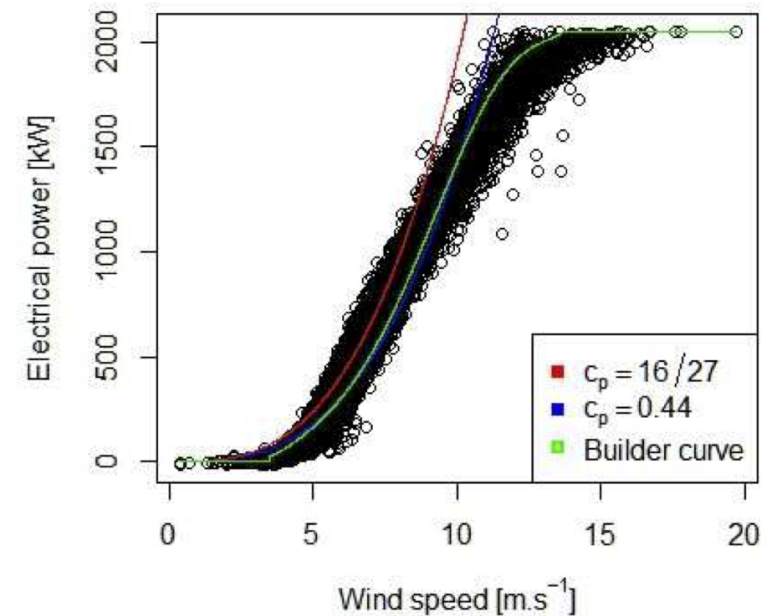
Il y a aussi de (très) forte corrélations spatiales), [Haslett & Raftery \(1989\)](#)

There is a theoretical formula to predict the output power of a turbine as a function of wind speed,

$$p(V) = \frac{1}{2} \rho A c_p V^3$$

où  $\rho$  est la densité de l'air (en  $kg/m^3$ ,  $\sim 1.225$  au niveau de la mer),  $A$  est la surface balayé par le rotor (en  $m^2$ ), et  $c_p$  un coefficient de puissance (inférieur à  $16/27$  - limite de Betz - et souvent de l'ordre de 0.45).

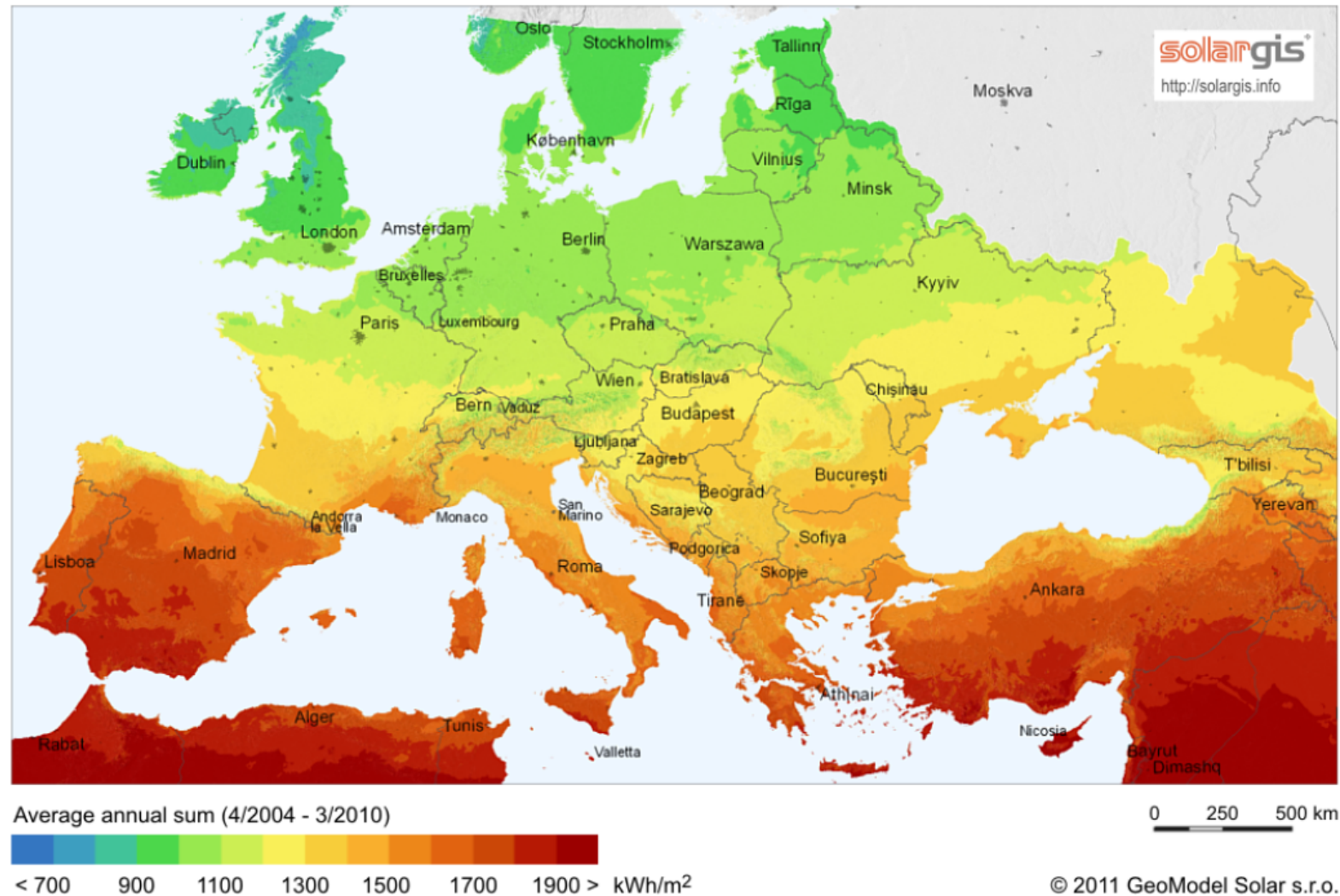
Il existe souvent en pratique un *cut-out* au delà duquel la turbine est éteinte (par sécurité). En pratique, c'est plus compliqué, cf [Fischer et al. \(2016\)](#)



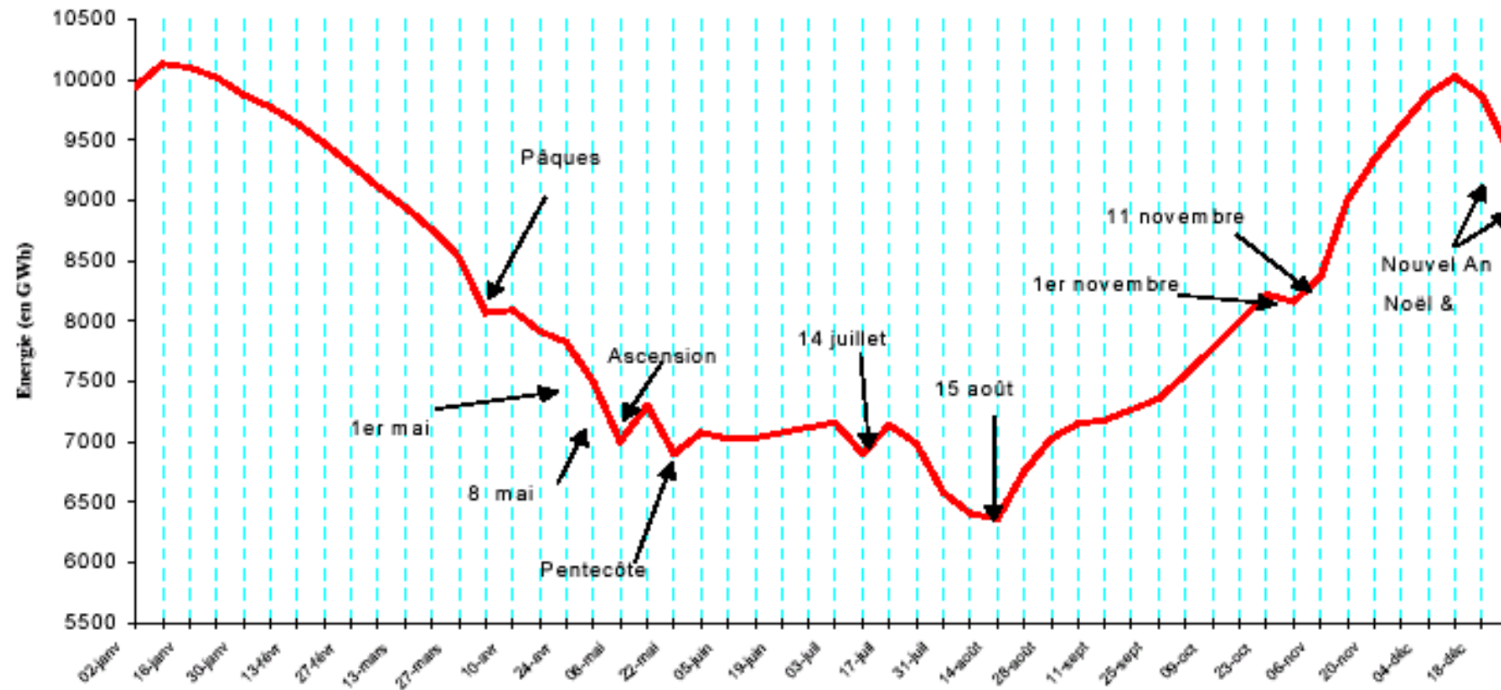


## Énergie(s) Renouvelable(s)

Dans le cas de l'énergie solaire,

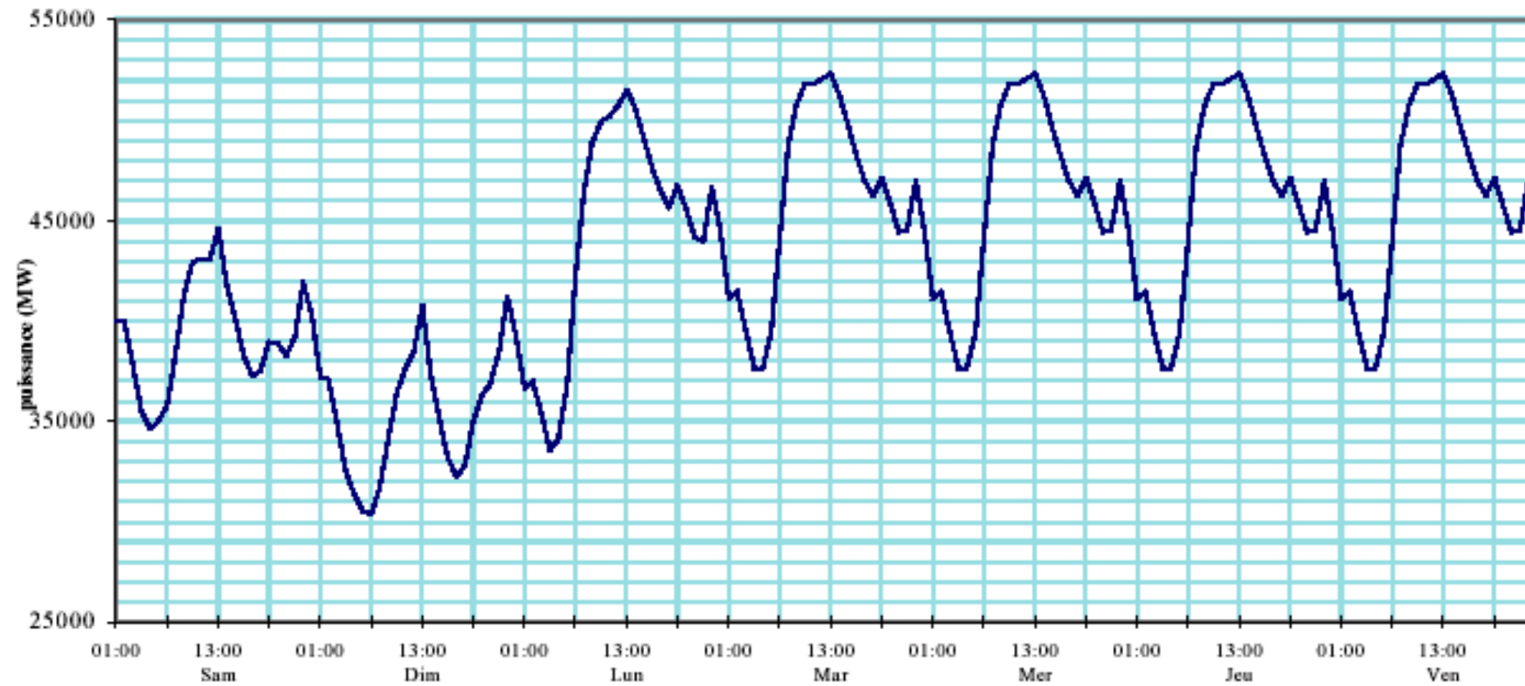


## Consommation électrique : année



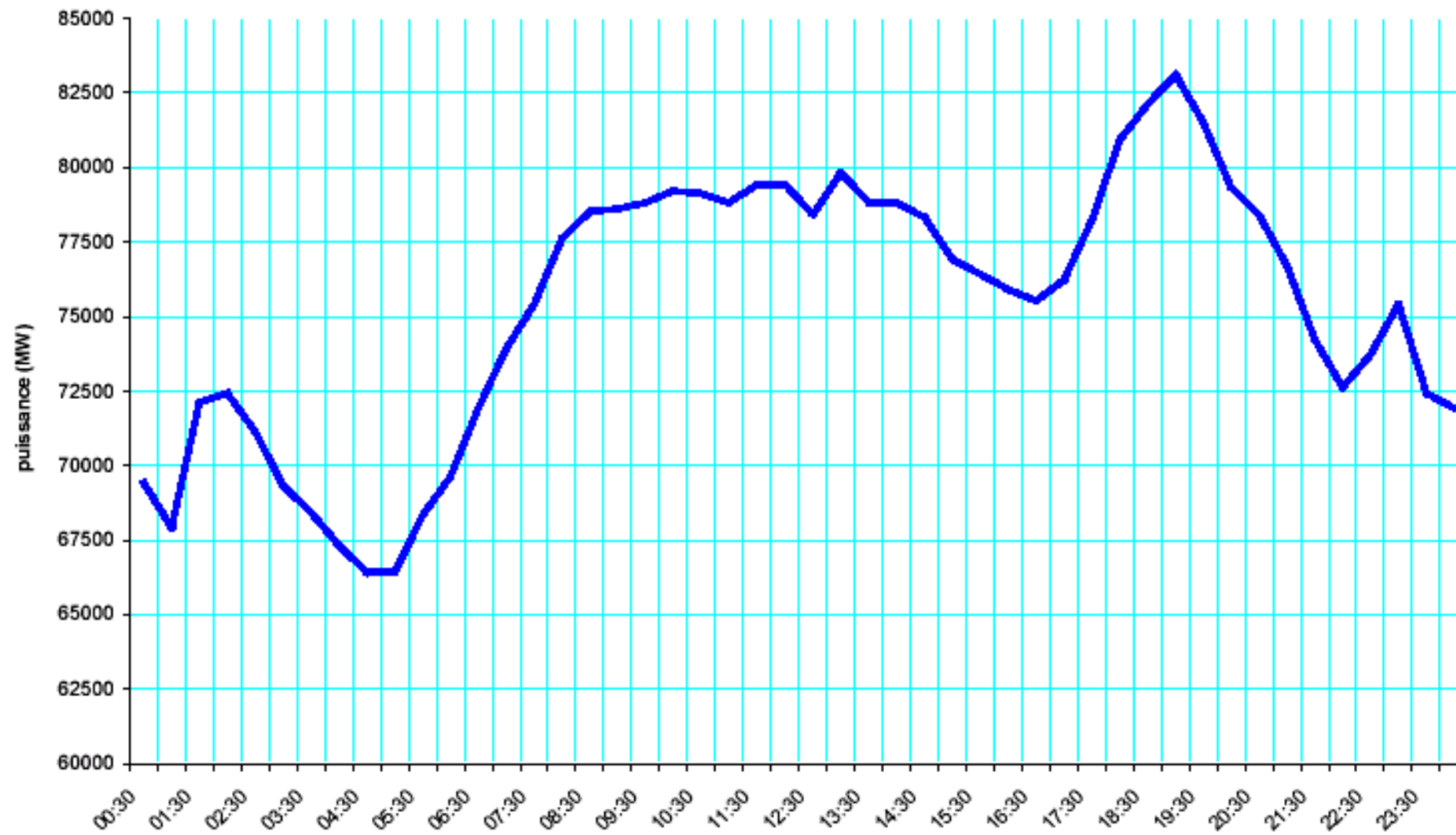
## Consommation électrique : semaine

Semaine pendant l'été 2003



## Consommation électrique : journée

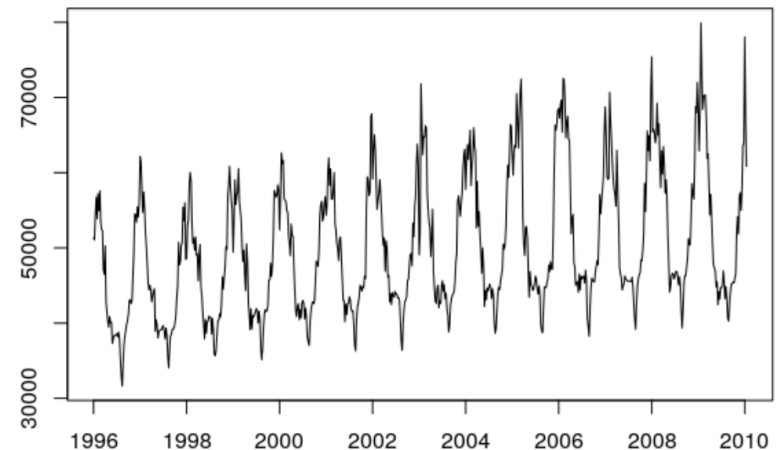
Mercredi, Janvier 2003



## Données de Consommation Électrique

Données hebdomadaires, entre 1996 et 2010, série  $(Y_t)$ ,  $\mathbf{y} = \{y_1, \dots, y_n\}$

```
6 url = "http://freakonometrics.free.fr/  
  elec.RData"  
7 download.file(url, "elec.RData")  
8 load("elec.RData")  
9 plot(ts(data = electricite$Load, start  
  = 1996, frequency = 52))
```



## Consommation d'électricité (en France)

Les variables, et la modélisation de leur impact :

- jour de la semaine (discret),
- heure (discret ou non-paramétrique),
- position dans l'année (continu).

Interactions :

- entre jour et heure,
- entre position dans l'année et heure (éclairage).

Faible activité

- jours fériés (11), veilles, lendemains, ponts : types de jours particuliers ;
- périodes de Noël et première quinzaine d'août : variable.

## Impact de la température

Caractéristiques à modéliser :

effet des températures chaudes, froides et très froides,

retard dans l'effet de la température sur la consommation (inertie des murs),

interaction avec l'heure.

Modélisations :

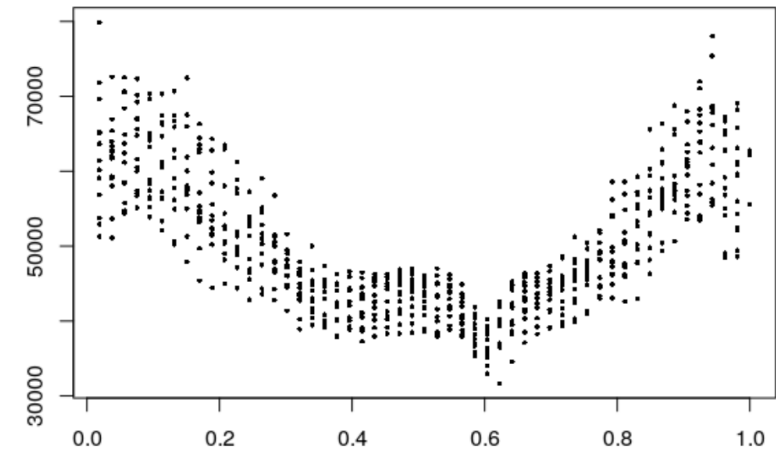
fonctions linéaires par morceaux,

splines dans un modèle GAM.

## Données

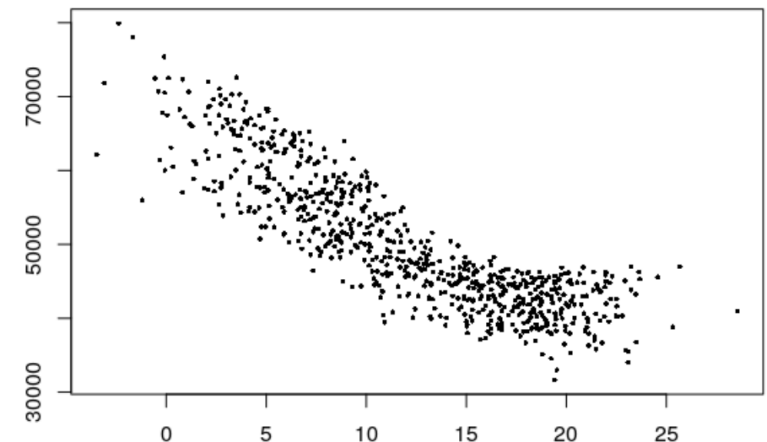
Impact de la **position dans l'année** sur la production électrique

```
6 > plot(electricite$NumWeek, electricite  
$Load, pch = 16, cex = 0.5)
```



Impact de la **température** sur la production électrique,  $(T_t, Y_t)$

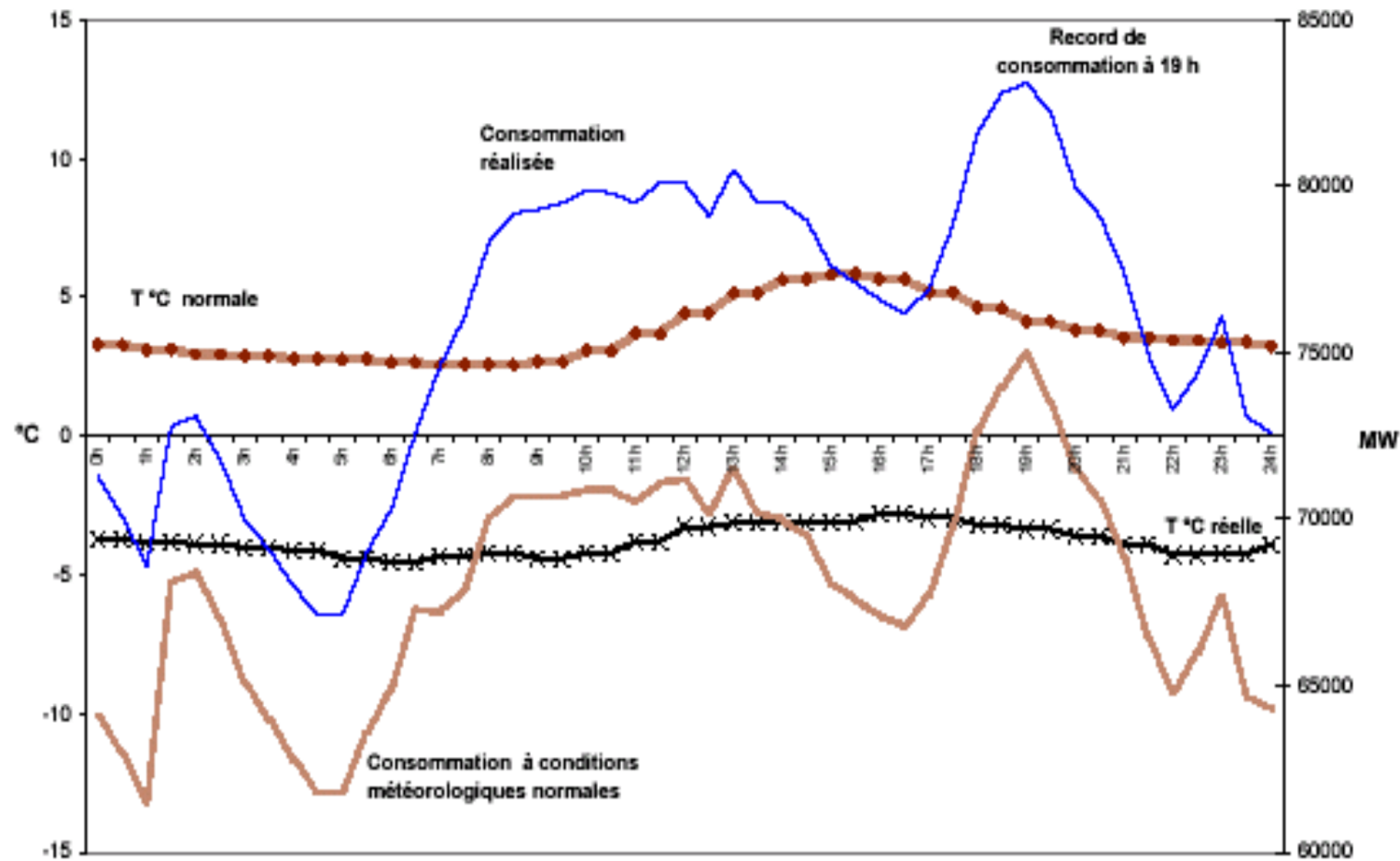
```
6 > plot(electricite$Temp, electricite$  
Load, pch = 16, cex = 0.5)
```





## Impact de la température

Vague de froid, hiver 2013, En France, une variation de 1° C en hiver correspond à une variation de 2500MW.

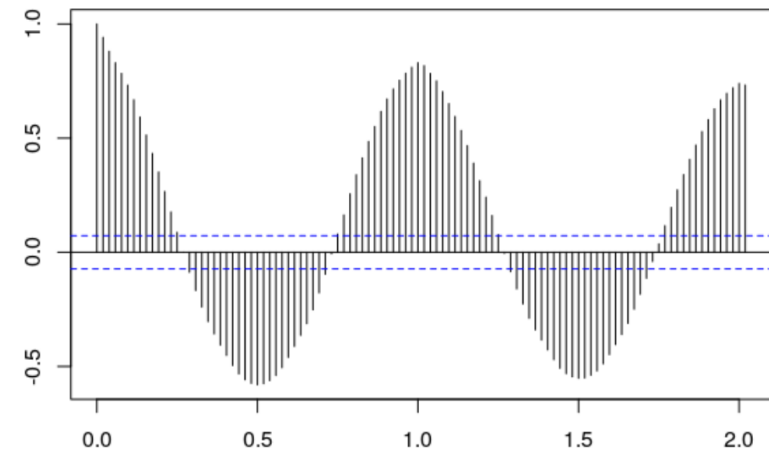


## Impact de la commutation récente

Autocorrélation de la série de production ( $Y_t$ ),

$$\rho(h) = \text{cor}(Y_t, Y_{t-h})$$

```
6 > acf(electricite$Load)
```



## Prévision ?

Notion de **risque 1%** : contrainte posée par la CRE **Commission de Régulation de l'Énergie** : marge nécessaire pour que la probabilité de coupures soit inférieure à 1%,

obligation réglementaire pour RTE, **Réseau de transport d'électricité**

conséquence : achats pratiquement à tout prix pour conserver la marge.

Interprétation : dépasser la prévision de quantile 1% dans 1% des cas,

“**Probabilistic forecasting** aims to maximize the sharpness of the predictive distributions, subject to calibration, on the basis of the available information set”,  
**Gneiting & Katzfuss (2014)**

## Régression Quantile

Pour  $\mathbf{y} \in \mathbb{R}^d$ ,  $\bar{y} = \operatorname{argmin}_{m \in \mathbb{R}} \left\{ \sum_{i=1}^n \frac{1}{n} \underbrace{[y_i - m]^2}_{\varepsilon_i} \right\}$ . Il s'agit de la version empirique de

$$\mathbb{E}[Y] = \operatorname{argmin}_{m \in \mathbb{R}} \left\{ \int \underbrace{[y - m]^2}_{\varepsilon} dF(y) \right\} = \operatorname{argmin}_{m \in \mathbb{R}} \left\{ \mathbb{E} \left[ \underbrace{\|Y - m\|_{\ell_2}}_{\varepsilon} \right] \right\}$$

où  $Y$  est une variable aléatoire.

Aussi,  $\operatorname{argmin}_{m(\cdot): \mathbb{R}^k \rightarrow \mathbb{R}} \left\{ \sum_{i=1}^n \frac{1}{n} \underbrace{[y_i - m(\mathbf{x}_i)]^2}_{\varepsilon_i} \right\}$  est la version empirique  $\mathbb{E}[Y | \mathbf{X} = \mathbf{x}]$ .

On parle d'estimation par **moindres carrés**, Legendre (1805) et Gauß (1809).

## Régression Quantile

Sketch of proof: (1) Let  $h(x) = \sum_{i=1}^d (x - y_i)^2$ , then

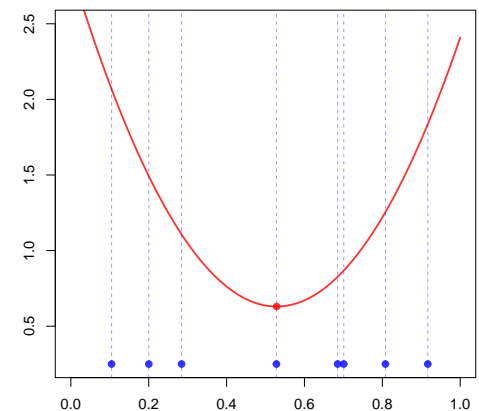
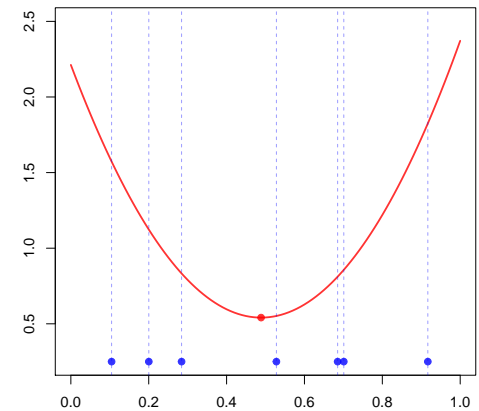
$$h'(x) = \sum_{i=1}^d 2(x - y_i)$$

and the FOC yields  $x = \frac{1}{n} \sum_{i=1}^d y_i = \bar{y}$ .

(2) If  $Y$  is continuous, let  $h(x) = \int_{\mathbb{R}} (x - y)^2 f(y) dy$  and

$$h'(x) = \frac{\partial}{\partial x} \int_{\mathbb{R}} (x - y)^2 f(y) dy = \int_{\mathbb{R}} \frac{\partial}{\partial x} (x - y)^2 f(y) dy$$

$$\text{i.e. } x = \int_{\mathbb{R}} x f(y) dy = \int_{\mathbb{R}} y f(y) dy = \mathbb{E}[Y]$$



## Régression Quantile

Pour  $\mathbf{y} \in \mathbb{R}^d$ ,  $\text{median}[\mathbf{y}] \in \underset{m \in \mathbb{R}}{\text{argmin}} \left\{ \sum_{i=1}^n \frac{1}{n} \underbrace{|y_i - m|}_{\varepsilon_i} \right\}$ . Il s'agit de la version empirique de

$$\text{median}[Y] \in \underset{m \in \mathbb{R}}{\text{argmin}} \left\{ \int \underbrace{|y - m|}_{\varepsilon} dF(y) \right\} = \underset{m \in \mathbb{R}}{\text{argmin}} \left\{ \mathbb{E}[\underbrace{\|Y - m\|_{\ell_1}}_{\varepsilon}] \right\}$$

où  $Y$  est une variable aléatoire,  $\mathbb{P}[Y \leq \text{median}[Y]] \geq \frac{1}{2}$  and  $\mathbb{P}[Y \geq \text{median}[Y]] \geq \frac{1}{2}$ .

$\underset{m(\cdot): \mathbb{R}^k \rightarrow \mathbb{R}}{\text{argmin}} \left\{ \sum_{i=1}^n \frac{1}{n} \underbrace{|y_i - m(\mathbf{x}_i)|}_{\varepsilon_i} \right\}$  est la version empirique  $\text{median}[Y|\mathbf{X} = \mathbf{x}]$ .

See [Boscovich \(1757\)](#) et [Laplace \(1793\)](#).

## Régression Quantile

### Median Regression, $\ell_1$ norm and Median

Sketch of proof: (1) Let  $h(x) = \sum_{i=1}^d |x - y_i|$

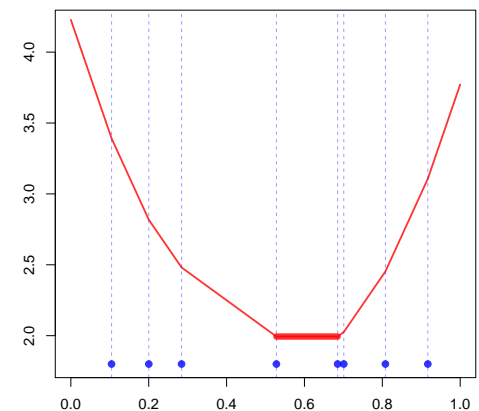
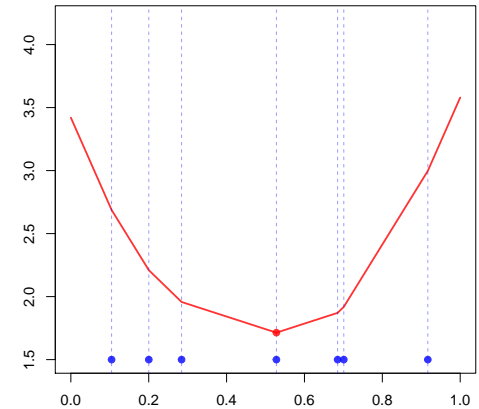
(2) If  $F$  is absolutely continuous,  $dF(x) = f(x)dx$ , and the median  $m$  is solution of  $\int_{-\infty}^m f(x)dx = \frac{1}{2}$ .

Set  $h(y) = \int_{-\infty}^{+\infty} |x - y| f(x) dx$

$$= \int_{-\infty}^y (-x + y) f(x) dx + \int_y^{+\infty} (x - y) f(x) dx$$

Then  $h'(y) = \int_{-\infty}^y f(x) dx - \int_y^{+\infty} f(x) dx$ , and FOC yields

$$\int_{-\infty}^y f(x) dx = \int_y^{+\infty} f(x) dx = 1 - \int_{-\infty}^y f(x) dx = \frac{1}{2}$$



## OLS vs. Median Regression (Least Absolute Deviation)

Consider some linear model,  $y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$ , and define

$$(\hat{\beta}_0^{\text{ols}}, \hat{\boldsymbol{\beta}}^{\text{ols}}) = \operatorname{argmin} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \mathbf{x}_i^\top \boldsymbol{\beta})^2 \right\}$$

$$(\hat{\beta}_0^{\text{lad}}, \hat{\boldsymbol{\beta}}^{\text{lad}}) = \operatorname{argmin} \left\{ \sum_{i=1}^n |y_i - \beta_0 - \mathbf{x}_i^\top \boldsymbol{\beta}| \right\}$$

Assume that  $\varepsilon|\mathbf{X}$  has a symmetric distribution,  $\mathbb{E}[\varepsilon|\mathbf{X}] = \operatorname{median}[\varepsilon|\mathbf{X}] = 0$ , then  $(\hat{\beta}_0^{\text{ols}}, \hat{\boldsymbol{\beta}}^{\text{ols}})$  and  $(\hat{\beta}_0^{\text{lad}}, \hat{\boldsymbol{\beta}}^{\text{lad}})$  are consistent estimators of  $(\beta_0, \boldsymbol{\beta})$ .

Assume that  $\varepsilon|\mathbf{X}$  does not have a symmetric distribution, but  $\mathbb{E}[\varepsilon|\mathbf{X}] = 0$ , then  $\hat{\boldsymbol{\beta}}^{\text{ols}}$  and  $\hat{\boldsymbol{\beta}}^{\text{lad}}$  are consistent estimators of the slopes  $\boldsymbol{\beta}$ .

If  $\operatorname{median}[\varepsilon|\mathbf{X}] = \gamma$ , then  $\hat{\beta}_0^{\text{lad}}$  converges to  $\beta_0 + \gamma$ .



## OLS vs. Median Regression

Median regression is stable by monotonic transformation. If

$$\log[y_i] = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i \text{ with } \text{median}[\varepsilon | \mathbf{X}] = 0,$$

then

$$\text{median}[Y | \mathbf{X} = \mathbf{x}] = \exp(\text{median}[\log(Y) | \mathbf{X} = \mathbf{x}]) = \exp(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})$$

while

$$\mathbb{E}[Y | \mathbf{X} = \mathbf{x}] \neq \exp(\mathbb{E}[\log(Y) | \mathbf{X} = \mathbf{x}]) \quad (= \exp(\mathbb{E}[\log(Y) | \mathbf{X} = \mathbf{x}]) \cdot [\exp(\varepsilon) | \mathbf{X} = \mathbf{x}])$$

```

1 > ols <- lm(y~x, data=df)
2 > library(quantreg)
3 > lad <- lm(y~x, data=df, tau=.5)

```

## Notations

Cumulative distribution function  $F_Y(y) = \mathbb{P}[Y \leq y]$ .

Quantile function  $Q_X(u) = \inf \{y \in \mathbb{R} : F_Y(y) \geq u\}$ ,

also noted  $Q_X(u) = F_X^{-1}u$ .

One can consider  $Q_X(u) = \sup \{y \in \mathbb{R} : F_Y(y) < u\}$

For any increasing transformation  $t$ ,  $Q_{t(Y)}(\tau) = t(Q_Y(\tau))$

$F(y|\mathbf{x}) = \mathbb{P}[Y \leq y | \mathbf{X} = \mathbf{x}]$

$Q_{Y|\mathbf{x}}(u) = F^{-1}(u|\mathbf{x})$

## Quantile regression ?

In OLS regression, we try to evaluate  $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}] = \int_{\mathbb{R}} y dF_{Y|\mathbf{X}=\mathbf{x}}(y)$

In quantile regression, we try to evaluate

$$Q_u(Y|\mathbf{X} = \mathbf{x}) = \inf \{y : F_{Y|\mathbf{X}=\mathbf{x}}(y) \geq u\}$$

as introduced in [Newey & Powell \(1987\)](#).

[Li & Racine \(2007\)](#) suggested

$$\hat{Q}_u(Y|\mathbf{X} = \mathbf{x}) = \inf \{y : \hat{F}_{Y|\mathbf{X}=\mathbf{x}}(y) \geq u\}$$

where  $\hat{F}_{Y|\mathbf{X}=\mathbf{x}}(y)$  can be some kernel-based estimator.

## Quantiles and Expectiles

Consider the following **risk functions**

$$\mathcal{R}_\tau^q(u) = u \cdot (\tau - \mathbf{1}(u < 0)), \quad \tau \in [0, 1]$$

with  $\mathcal{R}_{1/2}^q(u) \propto |u| = \|u\|_{\ell_1}$ , and

$$\mathcal{R}_\tau^e(u) = u^2 \cdot (\tau - \mathbf{1}(u < 0)), \quad \tau \in [0, 1]$$

with  $\mathcal{R}_{1/2}^e(u) \propto u^2 = \|u\|_{\ell_2}^2$ .

$$Q_Y(\tau) = \operatorname{argmin}_m \{ \mathbb{E}(\mathcal{R}_\tau^q(Y - m)) \}$$

which is the median when  $\tau = 1/2$ ,

$$E_Y(\tau) = \operatorname{argmin}_m \{ \mathbb{E}(\mathcal{R}_\tau^e(X - m)) \}$$

which is the expected value when  $\tau = 1/2$ .

## Quantiles and Expectiles

One can also write

**quantile:**  $\operatorname{argmin} \left\{ \sum_{i=1}^n \omega_{\tau}^q(\varepsilon_i) \underbrace{|y_i - q_i|}_{\varepsilon_i} \right\}$  where  $\omega_{\tau}^q(\epsilon) = \begin{cases} 1 - \tau & \text{if } \epsilon \leq 0 \\ \tau & \text{if } \epsilon > 0 \end{cases}$

**expectile:**  $\operatorname{argmin} \left\{ \sum_{i=1}^n \omega_{\tau}^e(\varepsilon_i) \underbrace{(y_i - q_i)^2}_{\varepsilon_i} \right\}$  where  $\omega_{\tau}^e(\epsilon) = \begin{cases} 1 - \tau & \text{if } \epsilon \leq 0 \\ \tau & \text{if } \epsilon > 0 \end{cases}$

Expectiles are unique, not quantiles...

Quantiles satisfy  $\mathbb{E}[\operatorname{sign}(Y - Q_Y(\tau))] = 0$

Expectiles satisfy  $\tau \mathbb{E}[(Y - e_Y(\tau))_+] = (1 - \tau) \mathbb{E}[(Y - e_Y(\tau))_-]$

## Empirical Quantiles

Consider some i.i.d. sample  $\{y_1, \dots, y_n\}$  with distribution  $F$ . Set

$$Q_\tau = \operatorname{argmin} \left\{ \mathbb{E} \left[ \mathcal{R}_\tau^q(Y - q) \right] \right\} \text{ where } Y \sim F \text{ and } \widehat{Q}_\tau \in \operatorname{argmin} \left\{ \sum_{i=1}^n \mathcal{R}_\tau^q(y_i - q) \right\}$$

Then as  $n \rightarrow \infty$

$$\sqrt{n}(\widehat{Q}_\tau - Q_\tau) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \frac{\tau(1-\tau)}{f^2(Q_\tau)} \right)$$

Sketch of the proof:  $y_i = Q_\tau + \varepsilon_i$ , set  $h_n(q) = \frac{1}{n} \sum_{i=1}^n (\mathbf{1}(y_i < q) - \tau)$ , which is a non-decreasing function, with

$$\mathbb{E} \left[ Q_\tau + \frac{u}{\sqrt{n}} \right] = F_Y \left( Q_\tau + \frac{u}{\sqrt{n}} \right) \sim f_Y(Q_\tau) \frac{u}{\sqrt{n}}$$

$$\operatorname{Var} \left[ Q_\tau + \frac{u}{\sqrt{n}} \right] \sim \frac{F_Y(Q_\tau)[1 - F_Y(Q_\tau)]}{n} = \frac{\tau(1-\tau)}{n}.$$

## Quantile Regression

We want to solve, here,  $\min \left\{ \sum_{i=1}^n \mathcal{R}_{\tau}^q(y_i - \mathbf{x}_i^{\top} \boldsymbol{\beta}) \right\}$

$$y_i = \mathbf{x}_i^{\top} \boldsymbol{\beta} + \varepsilon_i \text{ so that } \widehat{Q}_{y|\mathbf{x}}(\tau) = \mathbf{x}^{\top} \widehat{\boldsymbol{\beta}} + F_{\varepsilon}^{-1}(\tau)$$

## Geometric Properties of the Quantile Regression

Observe that the median regression will always have two supporting observations.

Start with some regression line,  $y_i = \beta_0 + \beta_1 x_i$

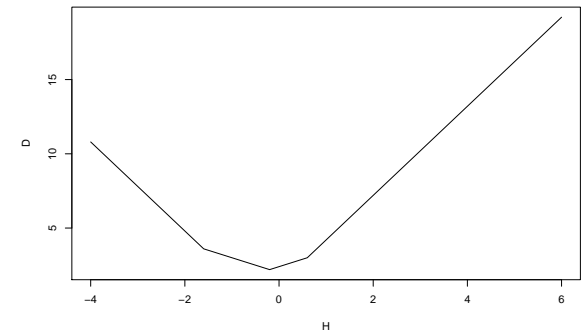
Consider small translations  $y_i = (\beta_0 \pm \epsilon) + \beta_1 x_i$

We minimize  $\sum_{i=1}^n |y_i - (\beta_0 + \beta_1 x_i)|$

From line blue, a shift up decrease the sum by  $\epsilon$  until we meet point on the left

an additional shift up will increase the sum

We will necessarily pass through one point (observe that the sum is piecewise linear in  $\epsilon$ )





## Geometric Properties of the Quantile Regression

Consider now rotations of the line around the support point

If we rotate up, we increase the sum of absolute difference (large impact on the point on the right)

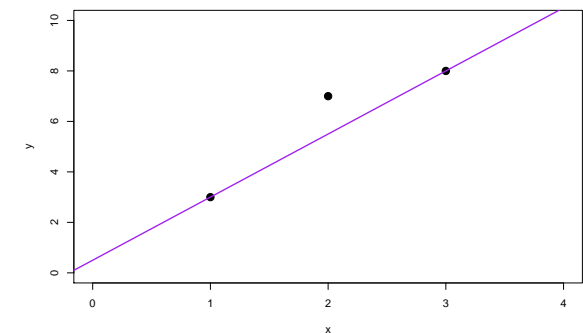
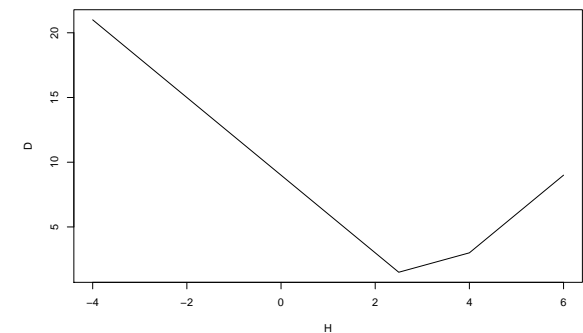
If we rotate down, we decrease the sum, until we reach the point on the right

Thus, the median regression will always have two supporting observations.

```

1 > library(quantreg)
2 > fit <- rq(dist~speed, data=cars, tau=.5)
3 > which(predict(fit)==cars$dist)
4 1 21 46
5 1 21 46

```



## Quantile Regression and Iterative Least Squares

start with some  $\beta^{(0)}$  e.g.  $\beta^{\text{ols}}$

at stage  $k$  :

$$\text{let } \varepsilon_i^{(k)} = y_i - \mathbf{x}_i^\top \beta^{(k-1)}$$

$$\text{define weights } \omega_i^{(k)} = \mathcal{R}'_{\tau}(\varepsilon_i^{(k)})$$

compute weighted least square to estimate  $\beta^{(k)}$

One can also consider a **smooth approximation** of  $\mathcal{R}_{\tau}^q(\cdot)$ , and then use Newton-Raphson.

## Optimization Algorithm

Primal problem is

$$\min_{\beta, \mathbf{u}, \mathbf{v}} \{ \tau \mathbf{1}^\top \mathbf{u} + (1 - \tau) \mathbf{1}^\top \mathbf{v} \} \text{ s.t. } \mathbf{y} = \mathbf{X}\beta + \mathbf{u} - \mathbf{v}, \text{ with } \mathbf{u}, \mathbf{v} \in \mathbb{R}_+^n$$

and the dual version is

$$\max_{\mathbf{d}} \{ \mathbf{y}^\top \mathbf{d} \} \text{ s.t. } \mathbf{X}^\top \mathbf{d} = (1 - \tau) \mathbf{X}^\top \mathbf{1} \text{ with } \mathbf{d} \in [0, 1]^n$$

Koenker & D'Orey (1994) suggest to use the **simplex method** (default in R)

Portnoy & Koenker (1997) suggest to use the **interior point method**

## Assumptions to get Consistency of Quantile Regression Estimators

As always, we need some assumptions to have consistency of estimators.

- observations  $(Y_i, \mathbf{X}_i)$  must (conditionally) i.i.d.
- regressors must have a bounded second moment,  $\mathbb{E}[\|\mathbf{X}_i\|^2] < \infty$
- error terms  $\varepsilon$  are continuously distributed given  $\mathbf{X}_i$ , centered in the sense that their median should be 0,

$$\int_{-\infty}^0 f_{\varepsilon}(\epsilon) d\epsilon = \frac{1}{2}.$$

- “local identification” property :  $[f_{\varepsilon}(0)\mathbf{X}\mathbf{X}^{\top}]$  is positive definite

## Quantile Regression Estimators

Under those weak conditions,  $\widehat{\beta}_\tau$  is asymptotically normal:

$$\sqrt{n}(\widehat{\beta}_\tau - \beta_\tau) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau(1 - \tau)D_\tau^{-1}\Omega_x D_\tau^{-1}),$$

where

$$D_\tau = \mathbb{E}[f_\varepsilon(0)\mathbf{X}\mathbf{X}^\top] \text{ and } \Omega_x = \mathbb{E}[\mathbf{X}^\top\mathbf{X}].$$

hence, the asymptotic variance of  $\widehat{\beta}$  is

$$\widehat{\text{Var}}[\widehat{\beta}_\tau] = \frac{\tau(1 - \tau)}{[\widehat{f}_\varepsilon(0)]^2} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^\top \mathbf{x}_i \right)^{-1}$$

where  $\widehat{f}_\varepsilon(0)$  is estimated using (e.g.) an histogram, as suggested in [Powell \(1991\)](#), since

$$D_\tau = \lim_{h \downarrow 0} \mathbb{E} \left( \frac{\mathbf{1}(|\varepsilon| \leq h)}{2h} \mathbf{X}\mathbf{X}^\top \right) \sim \frac{1}{2nh} \sum_{i=1}^n \mathbf{1}(|\varepsilon_i| \leq h) \mathbf{x}_i \mathbf{x}_i^\top = \widehat{D}_\tau.$$

## Quantile Regression Estimators

There is **no first order condition**, in the sense  $\partial V_n(\boldsymbol{\beta}, \tau) / \partial \boldsymbol{\beta} = \mathbf{0}$  where

$$V_n(\boldsymbol{\beta}, \tau) = \sum_{i=1}^n \mathcal{R}_\tau^q(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})$$

There is an **asymptotic first order condition**,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \psi_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) = \mathcal{O}(1), \text{ as } n \rightarrow \infty,$$

where  $\psi_\tau(\cdot) = \mathbf{1}(\cdot < 0) - \tau$ , see [Huber \(1967\)](#).

One can also define a Wald test, a Likelihood Ratio test, etc.

## Quantile Regression Estimators

Then the confidence interval of level  $1 - \alpha$  is then

$$\left[ \hat{\beta}_\tau \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}[\hat{\beta}_\tau]} \right]$$

An alternative is to use a bootstrap strategy (see #2)

- generate a sample  $(y_i^{(b)}, \mathbf{x}_i^{(b)})$  from  $(y_i, \mathbf{x}_i)$
- estimate  $\beta_\tau^{(b)}$  by

$$\hat{\beta}_\tau^{(b)} = \operatorname{argmin} \left\{ \mathcal{R}_\tau^q(y_i^{(b)} - \mathbf{x}_i^{(b)\top} \beta) \right\}$$

- set  $\widehat{\text{Var}}^*[\hat{\beta}_\tau] = \frac{1}{B} \sum_{b=1}^B (\hat{\beta}_\tau^{(b)} - \hat{\beta}_\tau)^2$

For confidence intervals, we can either use Gaussian-type confidence intervals, or empirical quantiles from bootstrap estimates.

## Quantile Regression Estimators

If  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)$ , one can prove that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} - \boldsymbol{\beta}_{\boldsymbol{\tau}}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\tau}}),$$

where  $\boldsymbol{\Sigma}_{\boldsymbol{\tau}}$  is a block matrix, with

$$\Sigma_{\tau_i, \tau_j} = (\min\{\tau_i, \tau_j\} - \tau_i \tau_j) D_{\tau_i}^{-1} \Omega_x D_{\tau_j}^{-1}$$

see [Kocherginsky et al. \(2005\)](#) for more details.



## Quantile Regression: Transformations

### Scale equivariance

For any  $a > 0$  and  $\tau \in [0, 1]$

$$\hat{\beta}_{\tau}(aY, \mathbf{X}) = a\hat{\beta}_{\tau}(Y, \mathbf{X}) \text{ and } \hat{\beta}_{\tau}(-aY, \mathbf{X}) = -a\hat{\beta}_{1-\tau}(Y, \mathbf{X})$$

### Equivariance to reparameterization of design

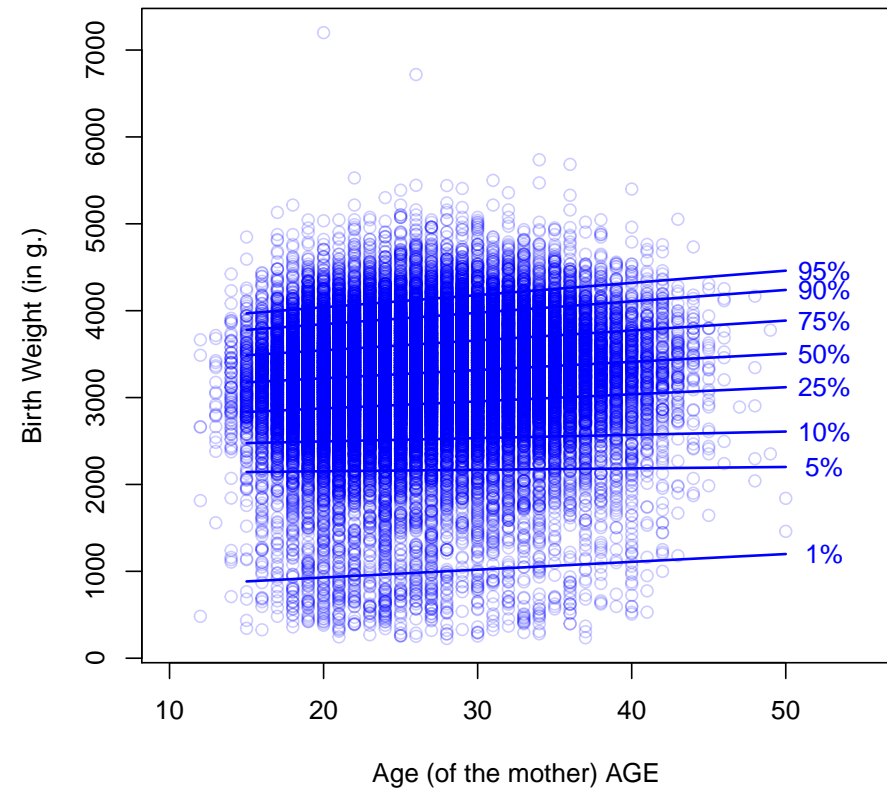
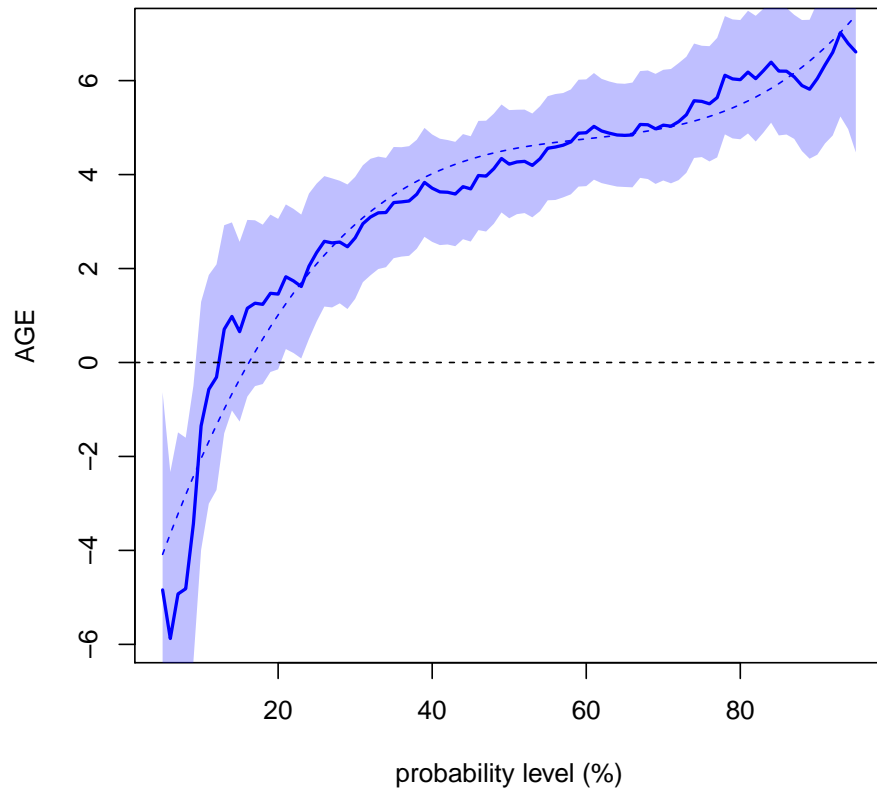
Let  $\mathbf{A}$  be any  $p \times p$  nonsingular matrix and  $\tau \in [0, 1]$

$$\hat{\beta}_{\tau}(Y, \mathbf{X}\mathbf{A}) = \mathbf{A}^{-1}\hat{\beta}_{\tau}(Y, \mathbf{X})$$

## Visualization, $\tau \mapsto \hat{\beta}_\tau$

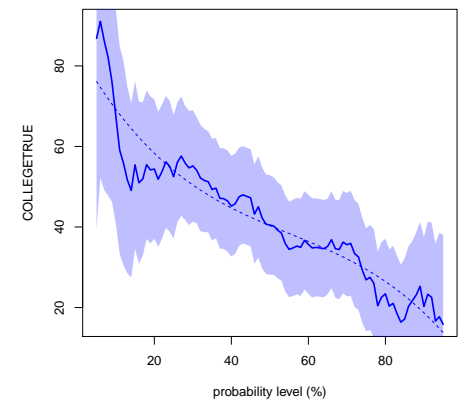
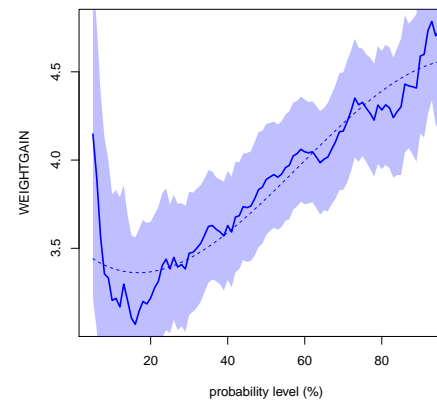
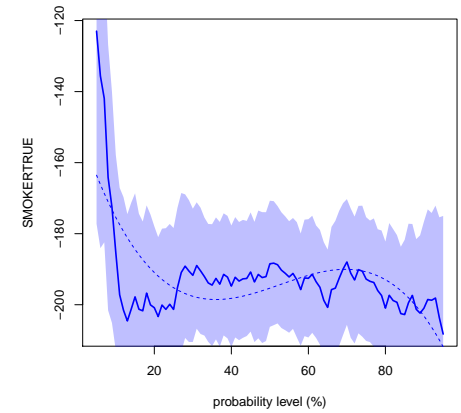
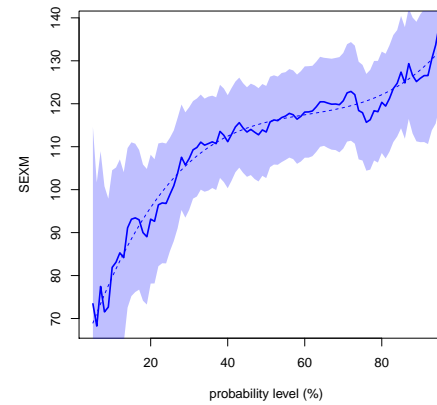
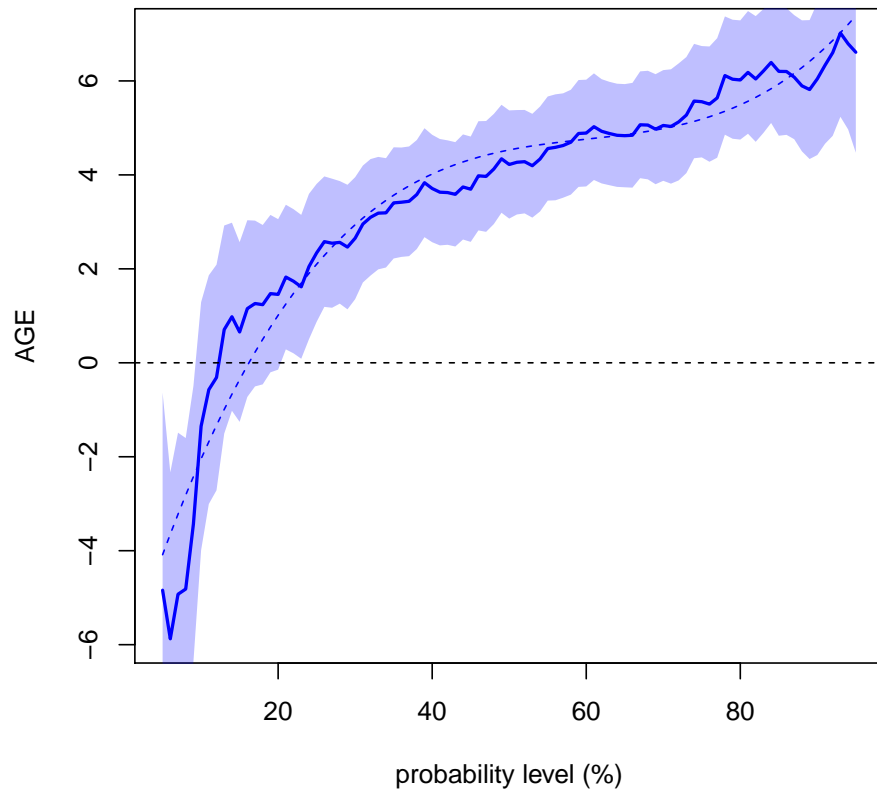
See [Abreveya \(2001\)](#)

```
1 > base=read.table("http://freakonometrics.free.fr/natality2005.txt")
```



# Visualization, $\tau \mapsto \hat{\beta}_\tau$

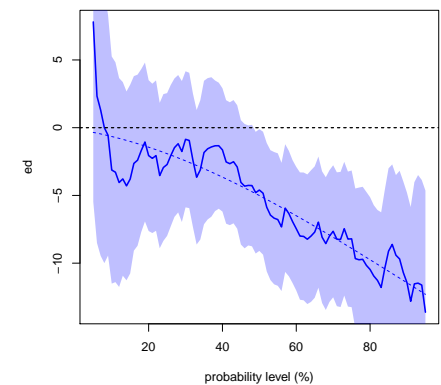
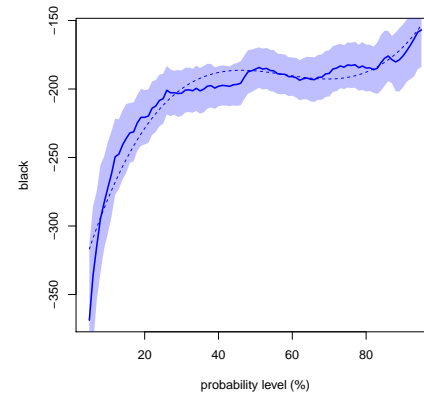
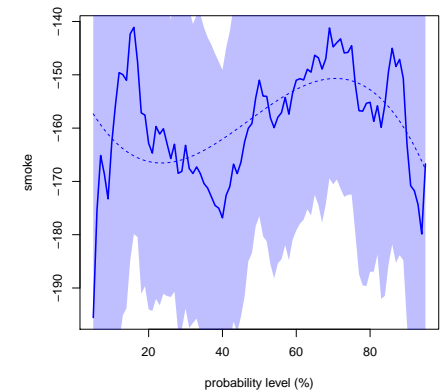
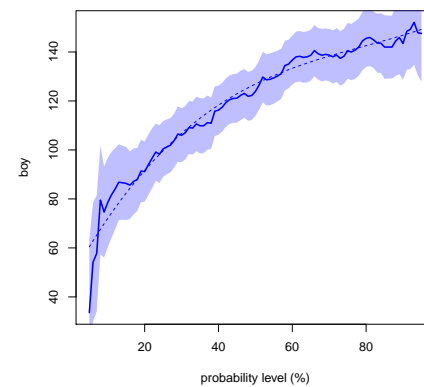
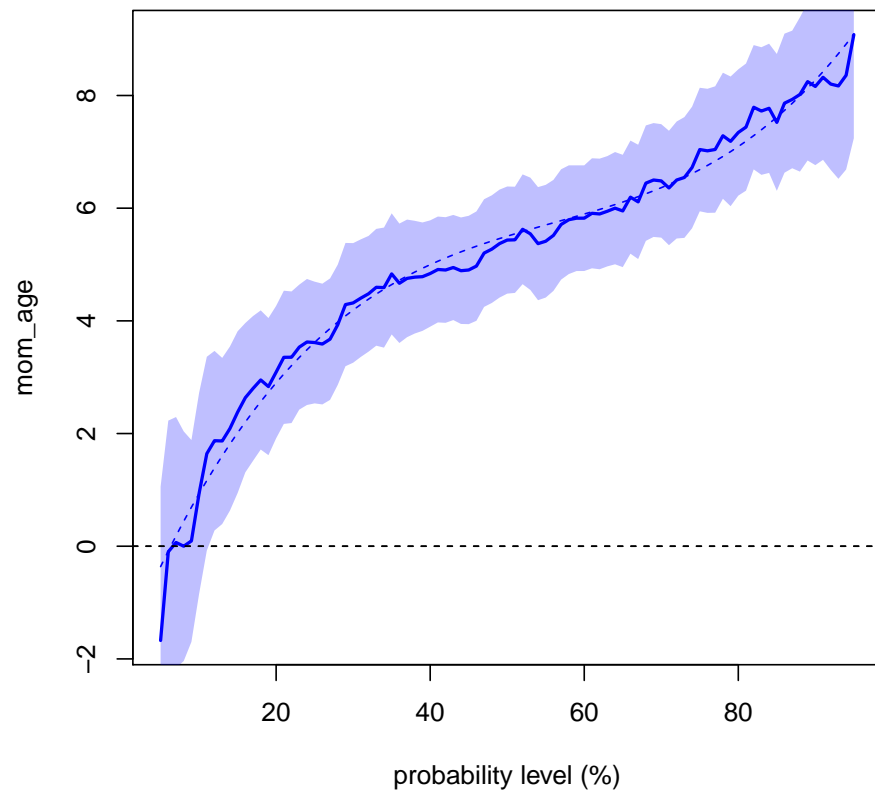
See Abreveya (2001)



## Visualization, $\tau \mapsto \hat{\beta}_\tau$

See [Abreveya \(2001\)](#)

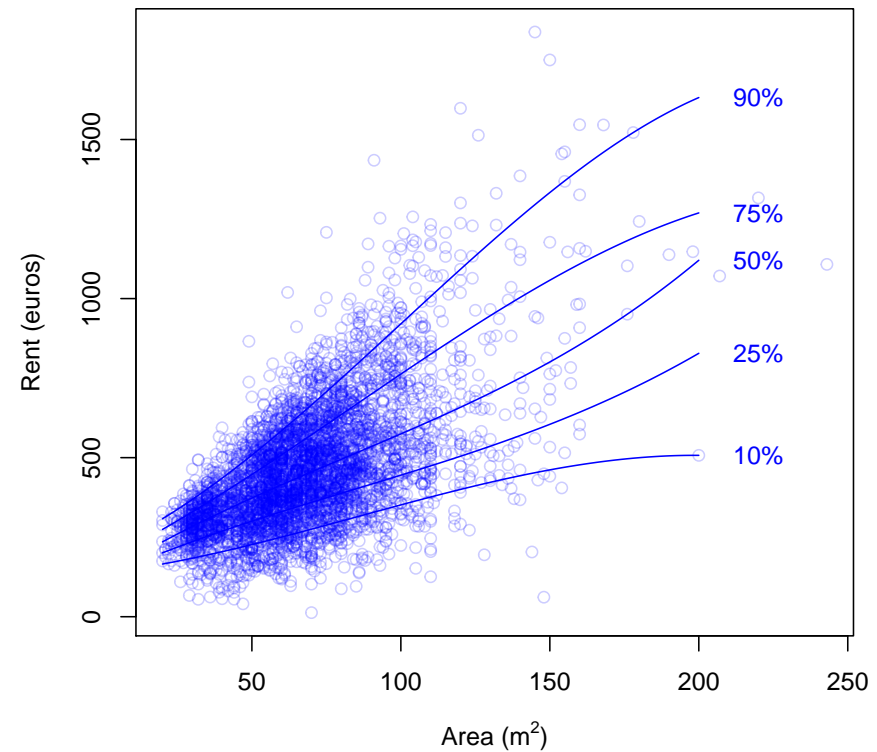
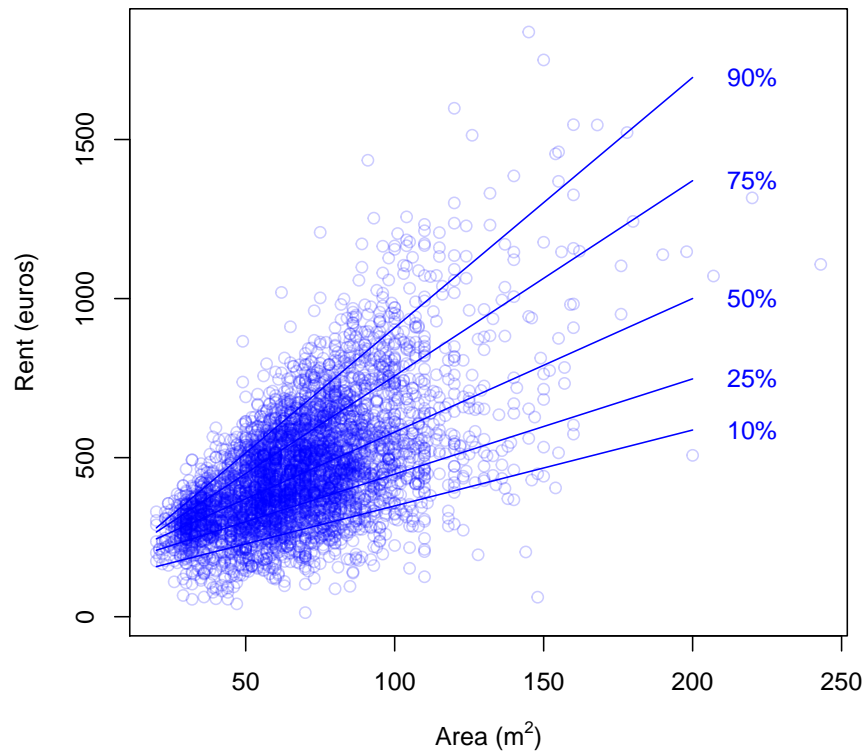
```
1 > base=read.table("http://freakonometrics.free.fr/BWeight.csv")
```



## Quantile Regression, with Non-Linear Effects

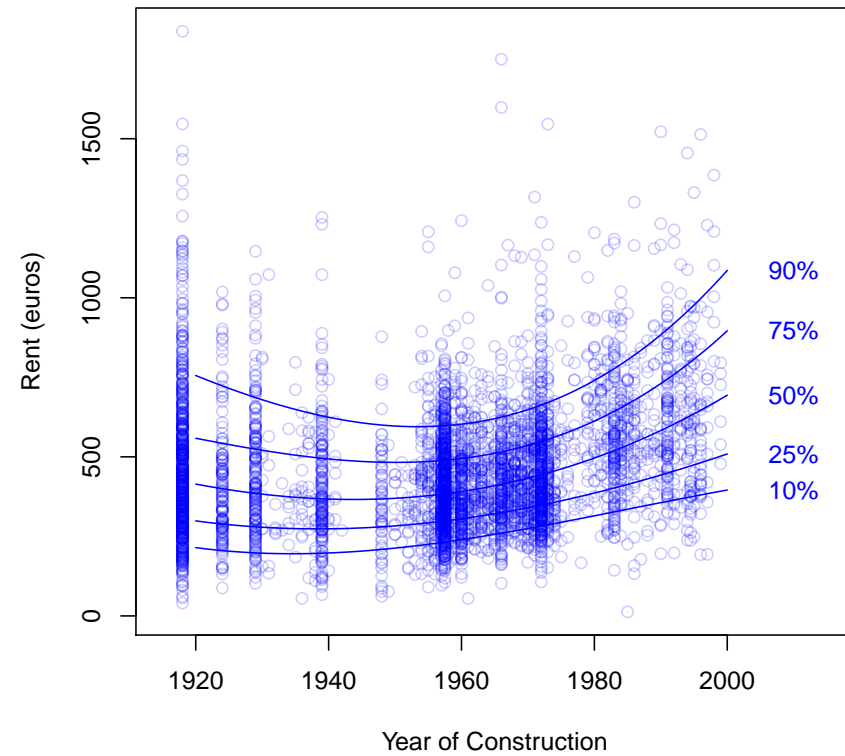
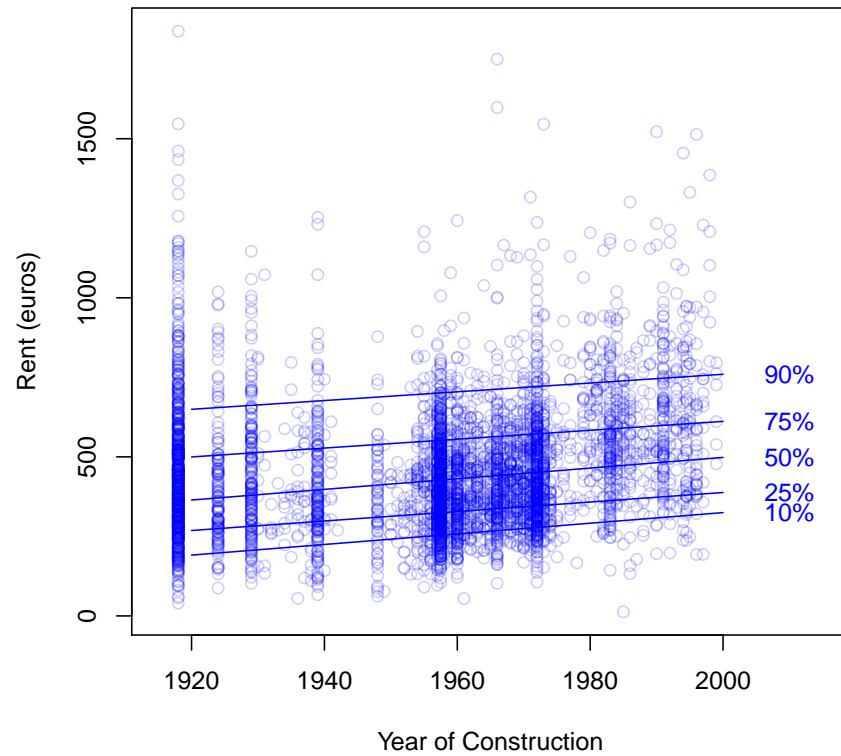
Rents in Munich, as a function of the area, from Fahrmeir *et al.* (2013)

```
1 > base=read.table("http://freakonometrics.free.fr/rent98_00.txt")
```



## Quantile Regression, with Non-Linear Effects

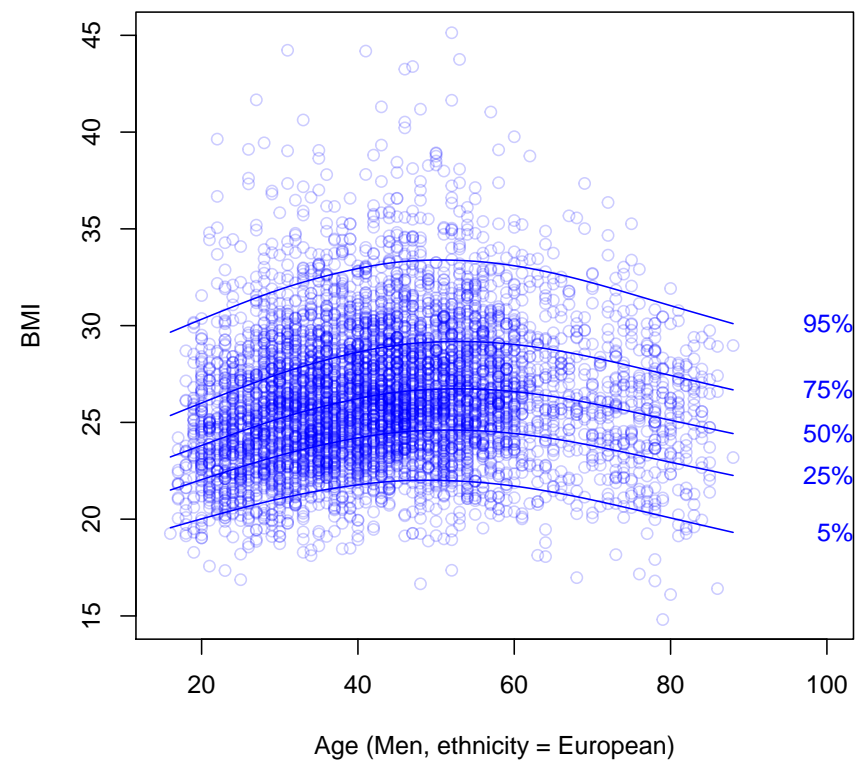
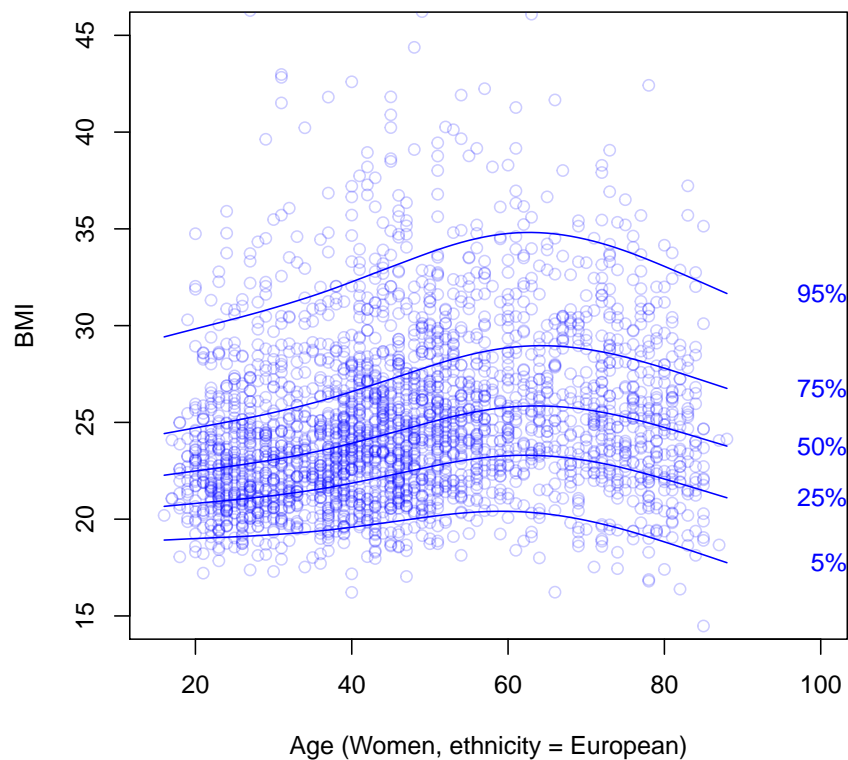
Rents in Munich, as a function of the year of construction, from Fahrmeir *et al.* (2013)



## Quantile Regression, with Non-Linear Effects

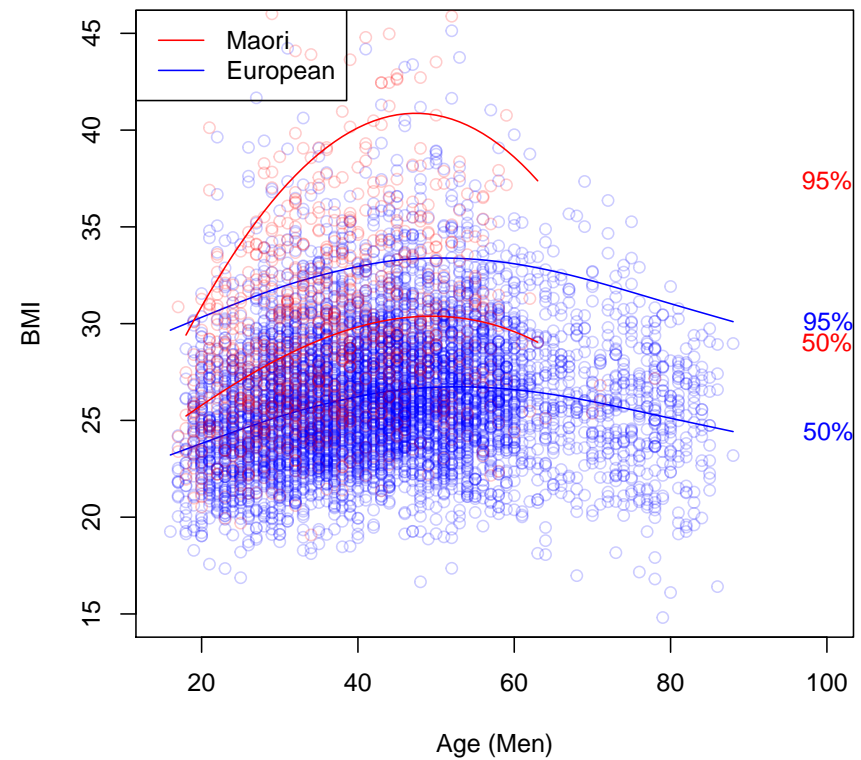
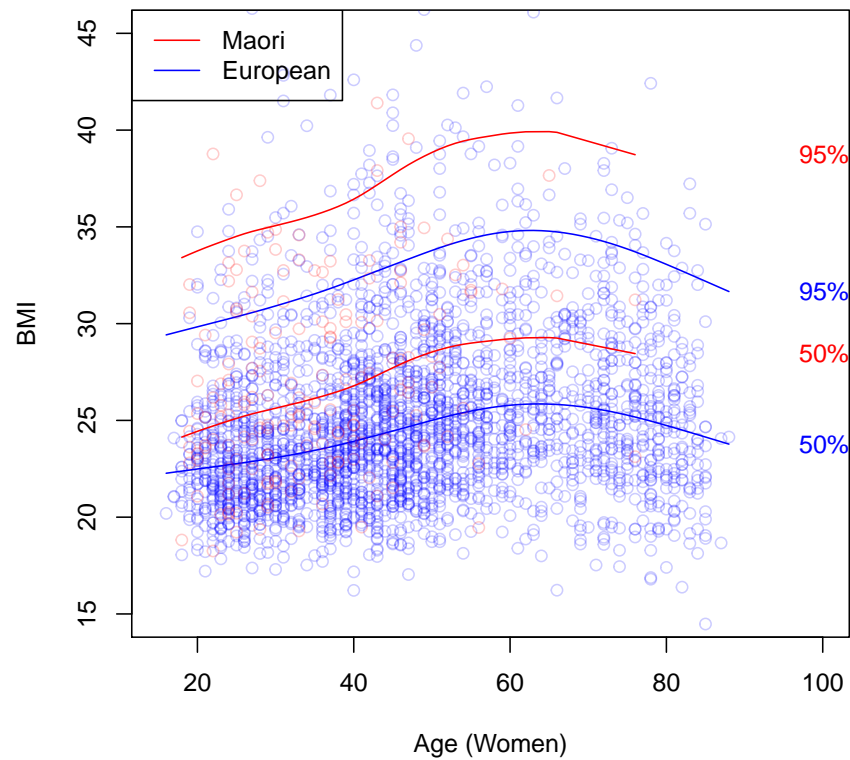
BMI as a function of the age, in New-Zealand, from Yee (2015), for Women and Men

```
1 > library(VGAMdata); data(xs.nz)
```



## Quantile Regression, with Non-Linear Effects

BMI as a function of the age, in New-Zealand, from Yee (2015), for Women and Men





## Quantile Regression, with Non-Linear Effects

One can consider some **local polynomial quantile regression**, e.g.

$$\min \left\{ \sum_{i=1}^n \omega_i(\mathbf{x}) \mathcal{R}_\tau^q(y_i - \beta_0 - (\mathbf{x}_i - \mathbf{x})^\top \beta_1) \right\}$$

for some weights  $\omega_i(\mathbf{x}) = H^{-1} K(H^{-1}(\mathbf{x}_i - \mathbf{x}))$ , see **Fan, Hu & Truong (1994)**.

## Régression Quantile et Lissage Exponentiel

Le **lissage exponentiel simple** : on approche  $y_{t+1}$  par  ${}_t\tilde{y}_{t+1}$  en posant

$${}_t\tilde{y}_{t+h} = L_t \text{ avec } L_t = \alpha Y_t + (1 - \alpha)L_{t-1}$$

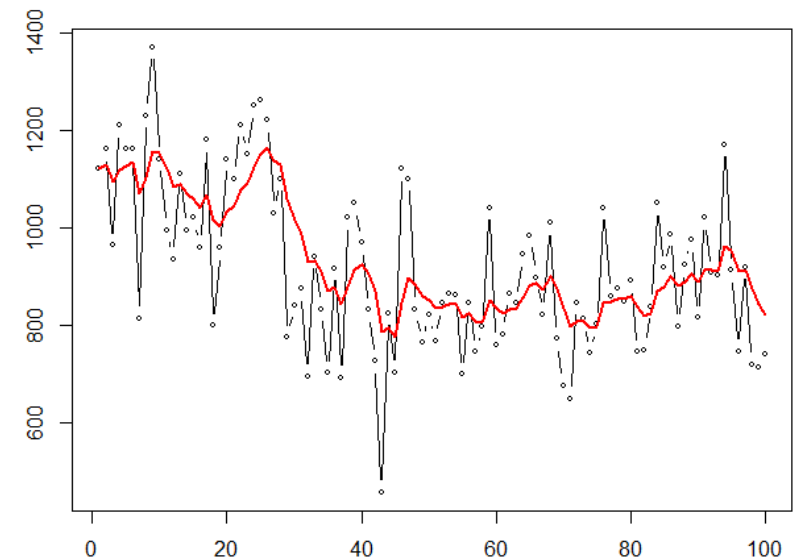
avec  $\alpha \in [0, 1]$

Classiquement, on cherche  $\alpha$  qui minimise l'erreur de prévision à horizon 1, i.e.

$$\alpha^* = \operatorname{argmin} \left\{ \sum_{t=2}^n (y_t - {}_{t-1}\tilde{y}_t)^2 \right\}$$

## Régression Quantile et Lissage Exponentiel

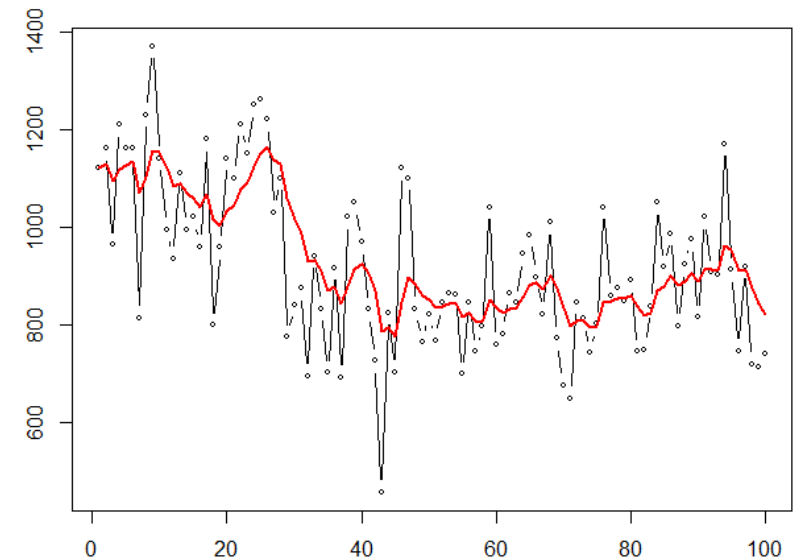
```
6 library(datasets)
7 X=as.numeric(Nile)
8 SimpleSmooth = function(a){
9   T=length(X)
10  L=rep(NA,T)
11  L[1]=X[1]
12  for(t in 2:T){L[t]=a*X[t]+(1-a)*L[t-1]}
13  return(L)
14 }
15 plot(X,type="b",cex=.6)
16 lines(SimpleSmooth(.2),col="red")
```



```

6 V=function(a){
7   T=length(X)
8   L=erreur=rep(NA,T)
9   erreur[1]=0
10  L[1]=X[1]
11  for(t in 2:T){
12    L[t]=a*X[t]+(1-a)*L[t-1]
13    erreur[t]=X[t]-L[t-1] }
14  return(sum(erreur^2))
15 }
16 optim(.5,V)$par
17 [1] 0.2464844
18 hw=HoltWinters(X,beta=FALSE,gamma=FALSE,
19   1.start=X[1])
19 hw$alpha
20 [1] 0.2465579

```



## Régression Quantile et Lissage Exponentiel

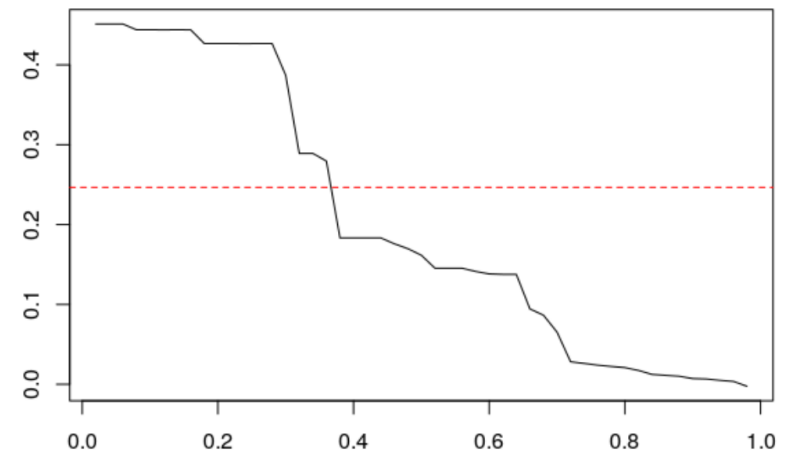
On peut considérer une erreur quantile à horizon 1,

$$\alpha_\tau^* = \operatorname{argmin} \left\{ \sum_{t=2}^n \mathcal{R}_\tau(y_t - {}_{t-1}\tilde{y}_t) \right\} \text{ where } \mathcal{R}_\tau(\varepsilon) = \varepsilon(\tau - \mathbf{1}_{\varepsilon \leq 0})$$

```

6 HWtau=function(tau){
7 loss=function(e) e*(tau-(e<=0)*1)
8 V=function(a){
9 T=length(X)
10 L=erreur=rep(NA,T)
11 erreur[1]=0
12 L[1]=X[1]
13 for(t in 2:T){
14 L[t]=a*X[t]+(1-a)*L[t-1]
15 erreur[t]=X[t]-L[t-1] }
16 return(sum(loss(erreur)))
17 }
18 optim(.5,V)$par
19 }

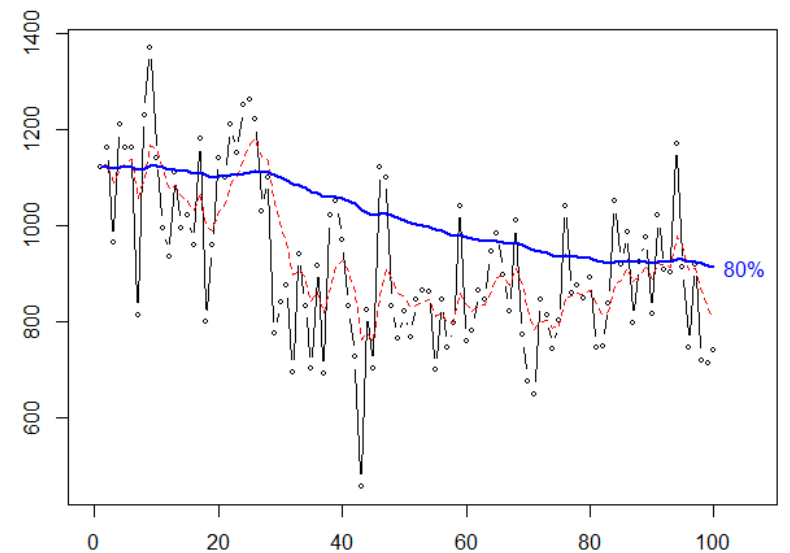
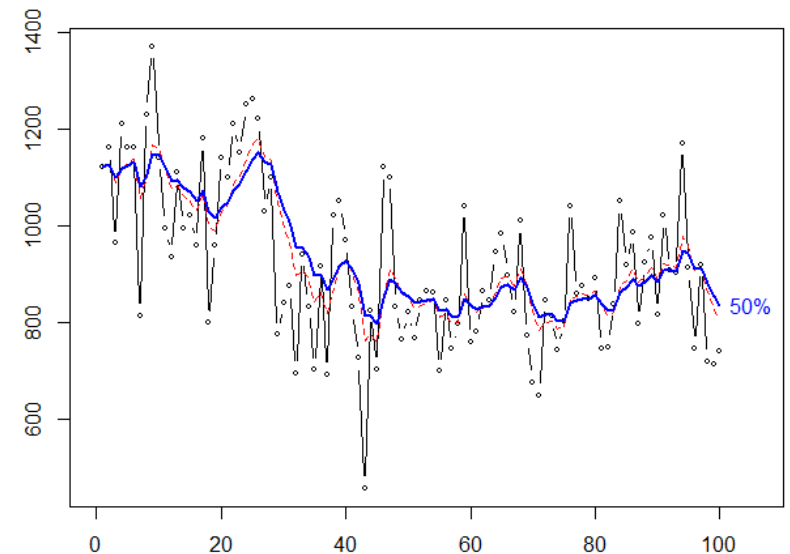
```



```

6 plot(X,type="b",cex=.6,xlim=c(0,106))
7 lines(SimpleSmooth(.2465),col="red",lty
      =2)
8 lines(SimpleSmooth(HWtau(.5)),col="blue",
      lwd=2)
9
10 plot(X,type="b",cex=.6,xlim=c(0,106))
11 lines(SimpleSmooth(.2465),col="red",lty
      =2)
12 lines(SimpleSmooth(HWtau(.8)),col="blue",
      lwd=2)

```



## Régression Quantile et Lissage Exponentiel

Le lissage exponentiel double : on approche  $y_{t+1}$  par  ${}_t\tilde{y}_{t+1}$  en posant

$${}_t\tilde{y}_{t+h} = L_t + hB_t \text{ avec } \begin{cases} L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + B_{t-1}) \\ B_t = \beta(L_t - L_{t-1}) + (1 - \beta)B_{t-1} \end{cases}$$

Là encore, on considère une erreur quantile à horizon 1,

$$(\alpha^*, \beta^*) = \operatorname{argmin} \left\{ \sum_{t=2}^n \ell(y_t - {}_{t-1}\tilde{y}_t) \right\}$$

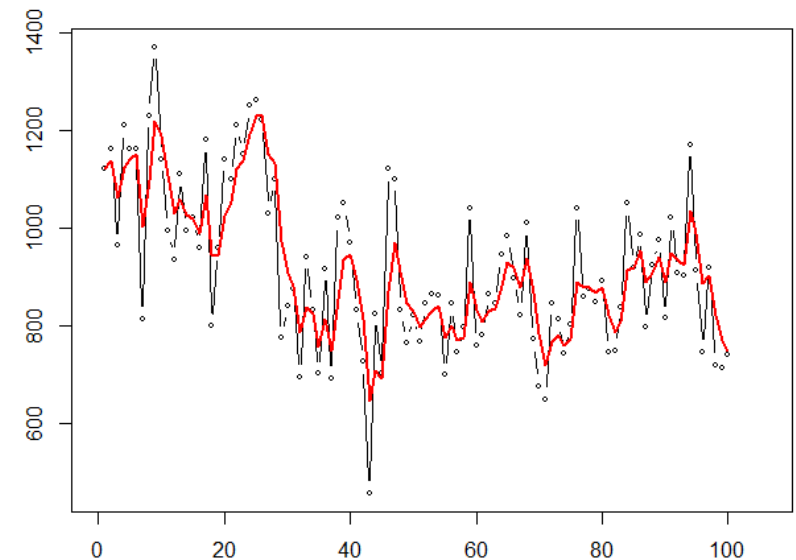
avec  ${}_{t-1}\tilde{y}_t = L_{t-1} + B_{t-1}$ .



```

6 hw=HoltWinters(X, gamma=FALSE, l.start=X
      [1])
7 hw$alpha
8     alpha
9 0.4200241
10 hw$beta
11     beta
12 0.05973389
13 DoubleSmooth = function(a,b){
14     T=length(X)
15     L=B=rep(NA, T)
16     L[1]=X[1]; B[1]=0
17     for(t in 2:T){
18         L[t]=a*X[t]+(1-a)*(L[t-1]+B[t-1])
19         B[t]=b*(L[t]-L[t-1])+(1-b)*B[t-1] }
20     return(L+B)
21 }

```



## Prévision

Statistique mathématique construite autour de l'**estimation** et des **tests d'hypothèses** sur les paramètres de modèles probabilistes.

Prévision en statistique : lorsque le modèle permet de reproduire les observations, il a toutes les chances de fournir de bonnes prévisions.

On va ici plutôt tenter d'utiliser un ensemble de scénarios susceptibles de se réaliser sur un horizon donné (par exemple en tirant suivant une distribution de probabilité, dite prédictive)

Classiquement, plus la prévision  $\hat{y}_t$  est proche de l'observation  $y_t$ , meilleur sera le modèle, soit pour la norme  $\ell_1$  avec  $|\hat{y}_t - y_t|$ , soit pour la norme  $\ell_2$  avec  $(\hat{y}_t - y_t)^2$

Il permet d'évaluer la tendance centrale, pas l'habilité à anticiper les événements extrêmes.

## Prévision

Le défi réel est de comparer deux objets de nature différentes : une fonction (la distribution de probabilité prédictive) et un nombre réel (l'observation)

Approche naturelle : introduire un score, cf [Good \(1952\)](#) ou [Winkler \(1969\)](#), appliqué en météorologie par [Murphy & Winkler \(1987\)](#).

Formellement la loi prédictive sera notée  $F$ . Elle exprime le degré d'incertitude qu'on accorde aux valeurs à venir compte tenu de l'information disponible.

## Prévision Probabiliste

Notion de [probabilistic forecasts](#), [Gneiting, T. and Raftery, A.E. \(2007\)](#)

Dans un contexte général, on souhaite prévoir la valeur prise par une variable aléatoire (réelle)  $Y$ .

Soit  $F$  une fonction de répartition.

On note  $\mathcal{A}$  l'information disponible lorsque l'on fait la prévision.

$F$  sera la prévision idéale ([ideal forecast](#)) pour  $Y$  sachant  $\mathcal{A}$  si la loi de  $Y|\mathcal{A}$  admet pour distribution  $F$ .

Supposons  $F$  continue. Posons  $Z_F = F(Y)$ , la [probability integral transform](#) de  $F$ .

$F$  est [probabilistically calibrated](#) si  $Z_F \sim \mathcal{U}([0, 1])$

$F$  est [marginally calibrated](#) si  $\mathbb{E}[F(y)] = \mathbb{P}[Y \leq y]$  pour tout  $y \in \mathbb{R}$ .

## Prévision Probabiliste

Notons que pour une **ideal forecast**,  $F(y) = \mathbb{P}[Y \leq y | \mathcal{A}]$ , qlors

- $\mathbb{E}[F(y)] = \mathbb{E}[\mathbb{P}[Y \leq y | \mathcal{A}]] = \mathbb{P}[Y \leq y]$

Cette prévision est marginally calibrated

- $\mathbb{P}[Z_F \leq z] = \mathbb{E}[\mathbb{P}[Z_F \leq z | \mathcal{A}]] = z$

Cette prévision est probabilistically calibrated

Supposons que  $\mu \sim \mathcal{N}(0, 1)$ . Et que la prévision idéale soit  $Y | \mu \sim \mathcal{N}(\mu, 1)$ .

Par exemple  $Y_t \sim \mathcal{N}(0, 1)$  et  $Y_{t+1} = y_t + \varepsilon_t \sim \mathcal{N}(y_t, 1)$ .

On peut proposer comme prévision  $F = \mathcal{N}(0, 2)$  (dite naïve). Cette distribution est marginally calibrated, probabilistically calibrated et ideal.

On peut proposer comme prévision  $F$  un mélange  $\mathcal{N}(\mu, 2)$  et  $\mathcal{N}(\mu \pm 1, 2)$  où "±1" signifie qu'on a +1 ou -1 avec probabilité 1/2 (prévision dite hésitante). Cette distribution est probabilistically calibrated, mais pas marginally calibrated.

## Prévision Probabiliste

En effet,  $\mathbb{P}[F(Y) \leq u] = u$ ,

$$\mathbb{P}[F(Y) \leq u] = \frac{\mathbb{P}[\Phi(Y) \leq u] + \mathbb{P}[\Phi(Y + 1) \leq u]}{2} + \frac{\mathbb{P}[\Phi(Y) \leq u] + \mathbb{P}[\Phi(Y - 1) \leq u]}{2}$$

On peut proposer comme prévision  $F = \mathcal{N}(-\mu, 1)$ . Cette distribution est marginally calibrated, mais pas probabilistically calibrated.

En pratique, on dispose d'une suite  $(Y_t, F_t)$  de couples, aussi notée  $(\mathbf{Y}, \mathbf{F})$ .

L'ensemble de prévisions  $\mathbf{F}$  est dit **performant** si pour toutes les dates  $t$ , les distributions prédictives  $F_t$  sont précises (notion de **sharpness**) tout en étant bien calibrées.

La **précision** est liée à la concentration de la densité de probabilité prédictive autour d'une valeur centrale (degré d'incertitude).

La **calibration** est liée à la cohérence entre les distributions prédictives  $F_t$  et les observations  $y_t$ .

## Prévision Probabiliste

On a une mauvaise calibration si les intervalles de confiance à 80% déduits des distributions prédictives (i.e.  $[F_t^{-1}(\alpha), F_t^{-1}(1 - \alpha)]$ ) ne contiennent pas en moyenne l'observation  $y_t$  8 fois sur 10 (environ).

Pour tester la calibration marginale, comparons la fonction de répartition empirique

$$\widehat{G}(y) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{1}_{Y_t \leq y}$$

et la moyenne des fonctions de prévision,

$$\overline{F}(y) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n F_t(y)$$

Pour tester la calibration probabiliste, testons si l'échantillon  $\{F_t(Y_t)\}$  suit une loi uniforme (approche dite PIT, [Dawid \(1984\)](#) )

## Prévision Probabiliste

On peut aussi créer un score  $S(F, y)$  pour toute prévision  $F$  et toute réalisation  $y$ .

Le score est propre si

$$\forall F, G, \mathbb{E}[S(G, Y)] \leq \mathbb{E}[S(F, Y)] \text{ où } Y \sim G.$$

En pratique, on approche cette espérance par  $\frac{1}{n} \sum_{t=1}^n S(F_t, Y_t)$

Parmi les règles classiques, on peut utiliser le **score logarithmique**

$S(F, y) = -\log[F'(y)]$  si la fonction  $F$  est à densité.

Un autre score classique est le **continuous ranked probability score** (CRPS)

$$S(F, y) = \int_{-\infty}^{+\infty} (F(x) - \mathbf{1}_{x \geq y})^2 dx = \int_{-\infty}^y F(x)^2 dx + \int_y^{+\infty} (F(x) - 1)^2 dx$$



## Prévision Probabiliste

dont la version empirique est

$$\hat{S} = \frac{1}{n} \sum_{t=1}^n S(F_t, y_t) = \frac{1}{n} \sum_{t=1}^n \int_{-\infty}^{+\infty} (F_t(x) - \mathbf{1}_{x \geq y_t})^2 dx$$

(étudié par [Murphy \(1970\)](#)).

Cette règle est propre puisque

$$\begin{aligned} \mathbb{E}[S(F, Y)] &= \int_{-\infty}^{\infty} \mathbb{E}[F(x) - \mathbf{1}_{x \geq Y}]^2 dx \\ &= \int_{-\infty}^{\infty} [[F(x) - G(x)]^2 + G(x)[1 - G(x)]]^2 dx \end{aligned}$$

est minimal quand  $F = G$ .

## Prévision Probabiliste

Si  $F$  correspond à la loi  $\mathcal{N}(\mu, \sigma^2)$

$$S(F, y) = \sigma \left[ \frac{y - \mu}{\sigma} \left( 2\Phi \left( \frac{y - \mu}{\sigma} \right) - 1 \right) + 2 \frac{y - \mu}{\sigma} - \frac{1}{\sqrt{\pi}} \right]$$

Notons que

$$S(F, y) = \mathbb{E}|X - y| - \frac{1}{2} \mathbb{E}|X - X'| \text{ où } X, X' \sim F$$

(et  $X$  et  $X'$  sont des variables indépendantes), cf [Gneiting, T. and Raftery, A.E. \(2007\)](#).

Si on utilise pour  $F$  la fonction de répartition empirique  $\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{y_i \leq y}$

alors

$$S(\hat{F}_n, y) = \frac{2}{n} \sum_{i=1}^n (y_{i:n} - y) \left( \mathbf{1}_{y_{i:n} \leq y} - \frac{i - 1/2}{n} \right)$$

## Prévision Probabiliste

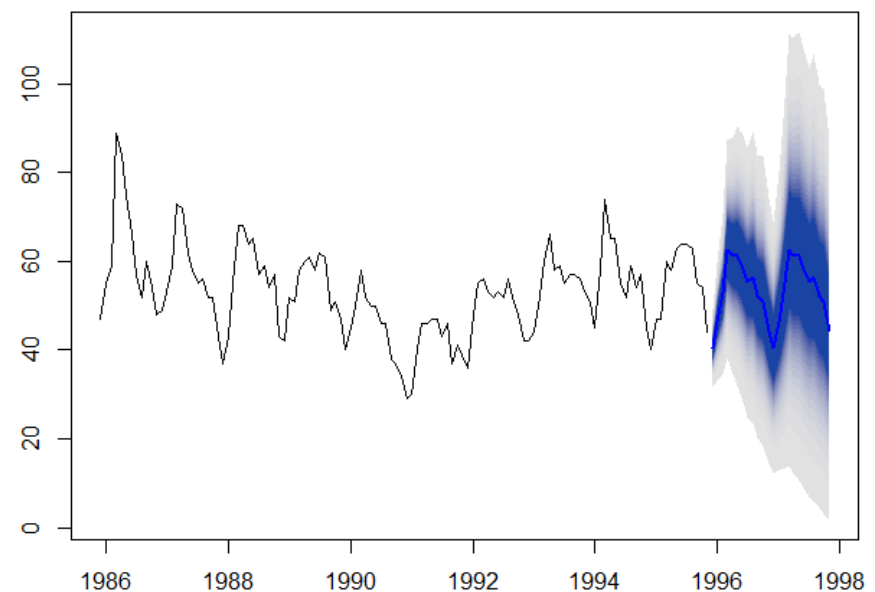
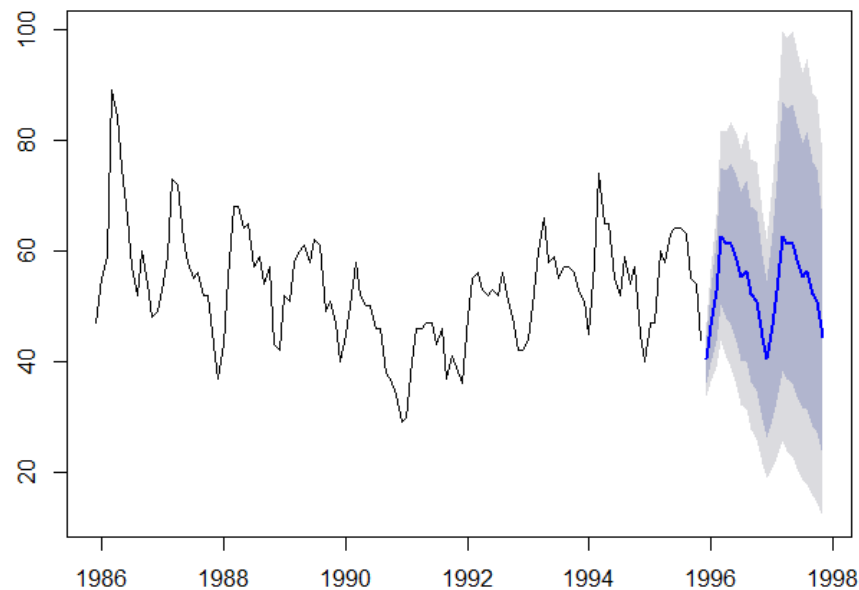
Pour un processus  $AR(p)$  Gaussien,

$$Y_t = c + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t, \text{ avec } \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

alors pour la prévision à horizon 1,

$$F_t \sim \mathcal{N}({}_{t-1}\hat{Y}_t, \sigma^2)$$

où  ${}_{t-1}\hat{Y}_t = c + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p}$ .



## Prévision Probabiliste

Supposons que  $Y$  puisse être expliqué par des variables  $\mathbf{x} = (x_1, \dots, x_m)$ .  
 Considérons une estimation à noyau de la densité conditionnelle

$$\hat{p}(y|\mathbf{x}) = \frac{\hat{p}(y, \mathbf{x})}{\hat{p}(\mathbf{x})} = \frac{\sum_{i=1}^n K_h(y - y_i) K_h(\mathbf{x} - \mathbf{x}_i)}{\sum_{i=1}^n K_h(\mathbf{x} - \mathbf{x}_i)}$$

Dans le cas d'un modèle linéaire, il existe  $\boldsymbol{\theta}$  tel que  $\hat{p}(y|\mathbf{x}) = \hat{p}(y|\boldsymbol{\theta}^\top \mathbf{x})$ , et

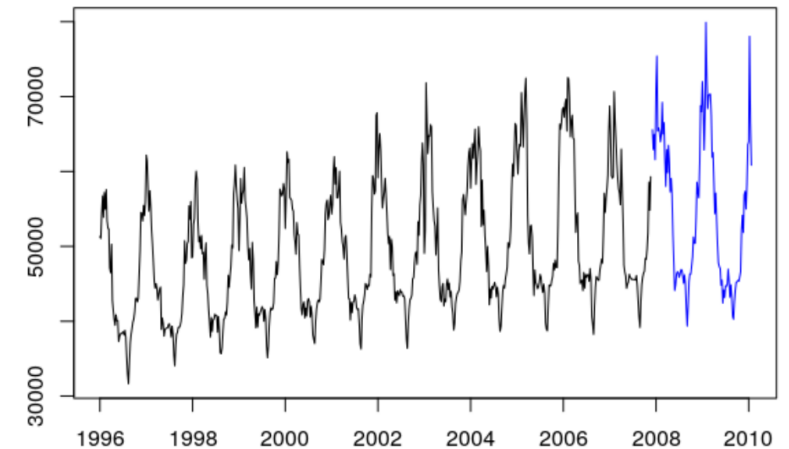
$$\hat{p}(y|\boldsymbol{\theta}^\top \mathbf{x} = s) = \frac{\sum_{i=1}^n K_h(y - y_i) K_h(s - \boldsymbol{\theta}^\top \mathbf{x}_i)}{\sum_{i=1}^n K_h(s - \boldsymbol{\theta}^\top \mathbf{x}_i)}$$

Le paramètre  $\boldsymbol{\theta}$  peut être estimé en maximisant un proxy de la log-vraisemblance

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax} \left\{ \sum_{i=1}^n \log \hat{p}(y_i | \boldsymbol{\theta}^\top \mathbf{x}_i) \right\}$$

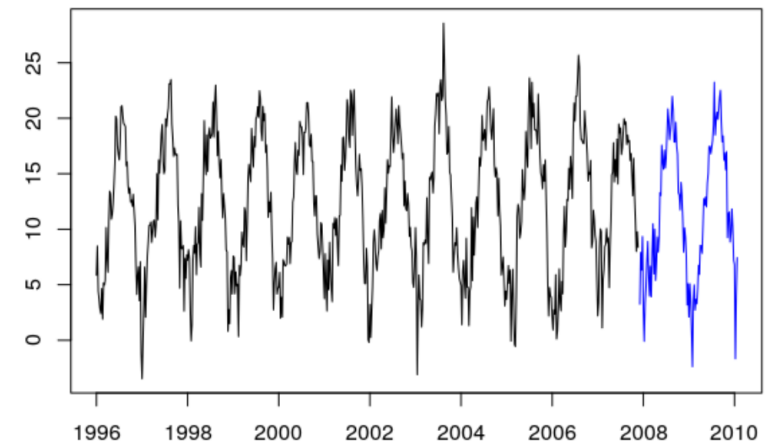
## Construction de prédicteurs

```
6 idx_data_test = 620:nrow(electricite)
7 data_train = electricite[-idx_data_test,]
8 data_test = electricite[idx_data_test,]
9 plot(ts(data = electricite$Load, start =
    1996, frequency = 52), col="white")
10 lines(ts(data = data_train$Load, start =
    1996, frequency = 52))
11 lines(ts(data = data_test$Load, start =
    1996+620/52, frequency = 52), col="
    blue")
```



## Régression

```
6 plot(ts(data = electricite$Temp, start =  
    1996, frequency = 52), col="white")  
7 lines(ts(data = data_train$Temp, start =  
    1996, frequency = 52))  
8 lines(ts(data = data_test$Temp, start =  
    1996+620/52, frequency = 52), col=""  
    blue")
```



## Régression

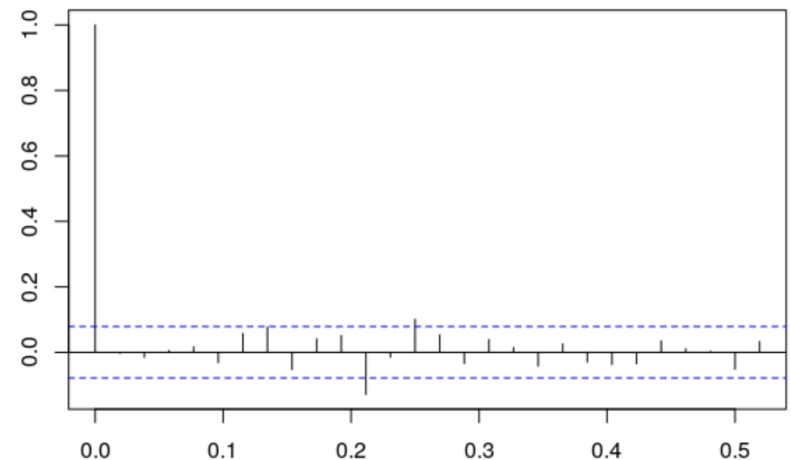
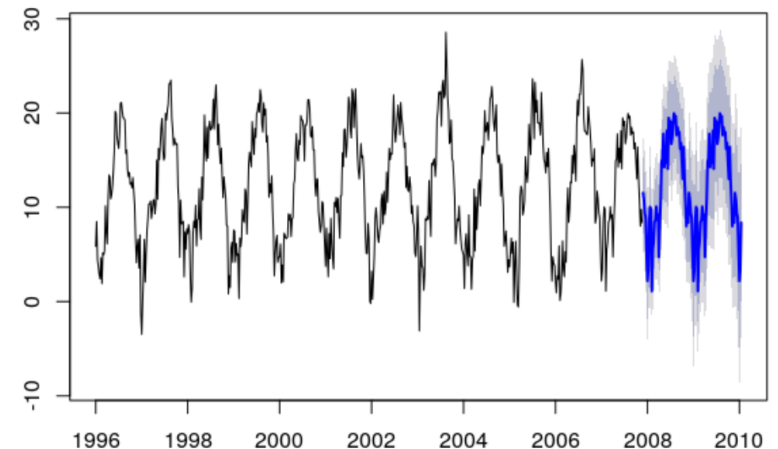
Modèle SARIMA,  $s = 52$

$$(1 - \varphi L)(1 - L^s)Y_t = \varepsilon_t.$$

```

6 z = ts(data = data_train$Temp, start =
    1996, frequency = 52)
7 ARIMA = arima(z, order=c(1,0,0), seasonal=
    list(order=c(0,1,0), period=52))
8 plot(forecast(ARIMA, h=112))
9 acf(residuals(ARIMA))

```



## Aggrégation de modèles

Discuté dans [Cesa-Bianchi and Lugisi \(2006\)](#)

On dispose d'un ensemble de  $k$  modèles,  $E = \{1, \dots, k\}$ , et on note  $f_{i,t}$  la prévision faite par l'expert  $i \in E$ . Soit  $\mathbf{p}_t = (p_{i,t}) \in \mathbb{R}^k$ , on peut alors considérer la prévision donnée par

$$\hat{y}_t = \sum_{i \in E_t} p_{i,t} f_{i,t} = \mathbf{p}_t^\top \mathbf{f}_t.$$

On parlera de poids convexes si  $\mathbf{p}_t \in \mathbb{R}_+^k$  et  $\mathbf{p}_t^\top \mathbf{1} = 1$ .

Dans un contexte générale, on se donne une fonction de perte  $\ell$  et on note la perte moyenne

$$\frac{1}{n} \sum_{t=1}^n \ell(y_t, \mathbf{p}_t^\top \mathbf{f}_t)$$