Prévision non parametrique : quelques exemples

A. Philippe

Laboratoire de mathématiques Jean Leray
Université de Nantes
Anne.Philippe@univ-nantes.fr

Hiver 2008

Présentation

Le prédicteur à noyau non paramétrique dépend de

- la taille du bloc r
- la vitesse de fermeture de la fenêtre $h_n = \frac{C_h}{n^{1/(4+r)}}$.
- Le noyau utilisé est le noyau gaussien.

On calibre les constantes r, C_h par validation croisée. Plus précisément, ayant observé $X_1, \ldots, X_n$, et cherchant à prévoir $X_{n+1}, \ldots, X_{n+K_{prev}}$, on détermine les valeurs de r, C_h qui minimisent la quantité suivante (erreur quadratique)

$$\sum_{t=[n/2]}^{n-K_{prev}} \sum_{i=1}^{K_{prev}} (\hat{X}_{t:i} - X_{t+i})^2$$

où $\hat{X}_{t:i}$ est la prévision de X_{t+i} ayant observé $X_1, \ldots, X_t$.

Fonction R :

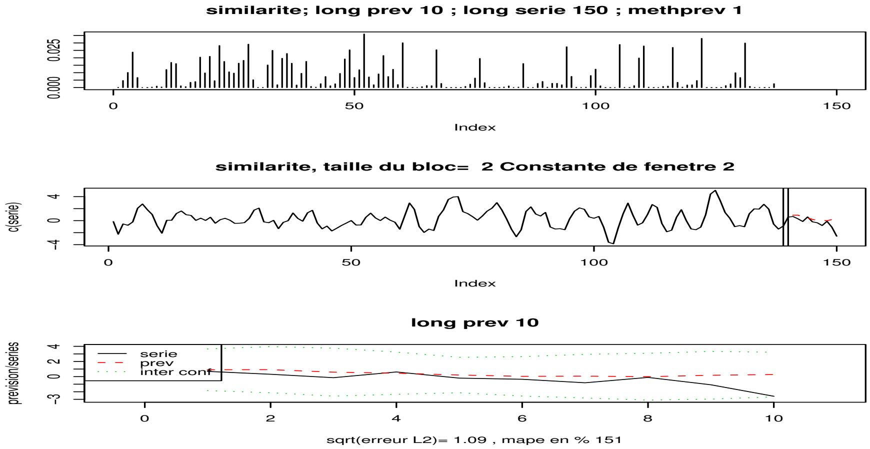
```
 function(serie, kprev = 15, r =
10,cte=1,typtub=0.1,passim=1,debbloc1=1, methprev=1, dessin
= 1)
```

Un exemple stationnaire et linéaire

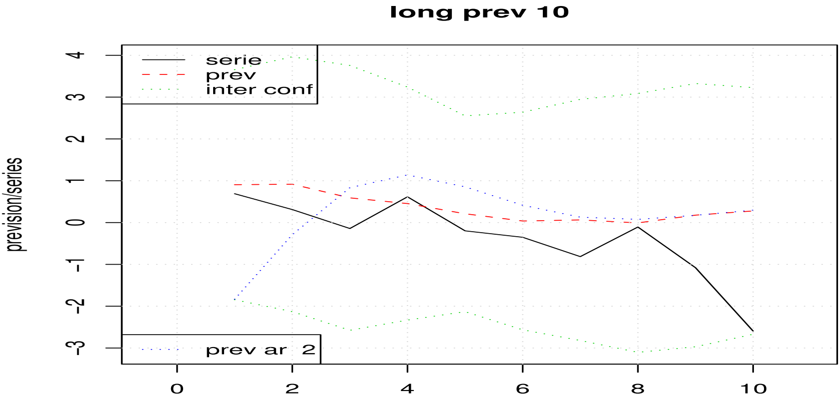
Les données sont simulées suivant un modèle AR(2) :

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \epsilon_t$$

avec $a_1 = 0.9$ et $a_2 = -0.45$.
On dispose de 140 observations et on prévoit à l'horizon 10. Les paramètres obtenus par validation croisée sont $r = 2$ et $C_h = 2$. On retrouve bien l'ordre de l'AR comme choix optimal de r.

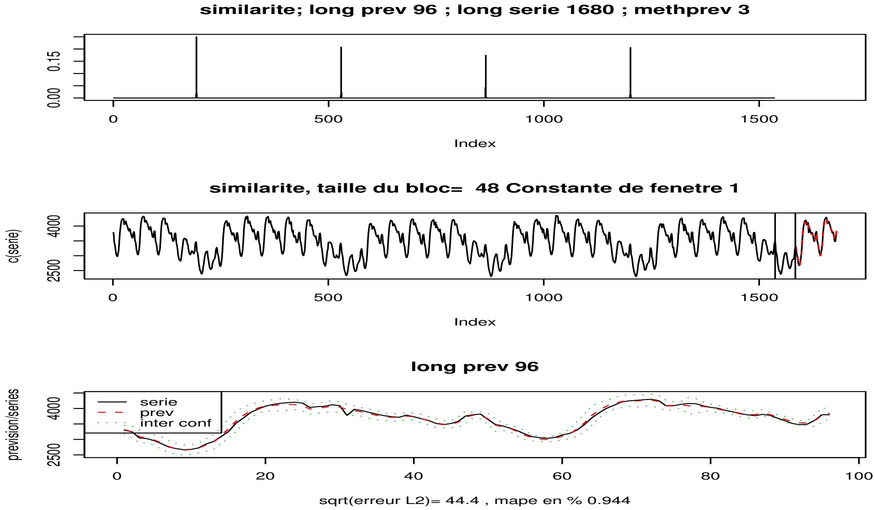


Les graphiques représentent les similarités [*haut*], la série et la prévision [*milieu*], la prévision et la région de confiance [*bas*]



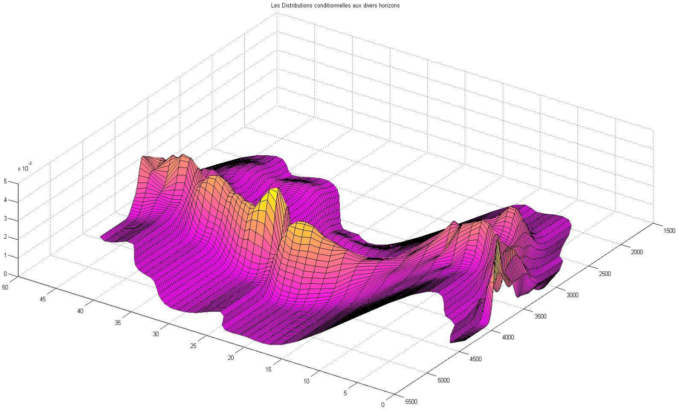
Sur les données précédentes. On compare la prévision non paramétrique avec le prédicteur linéaire construit à partir d'un modèle AR(2).

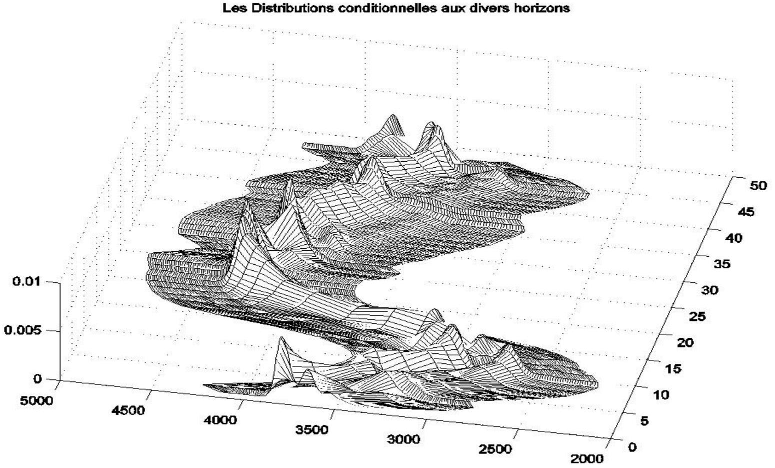
Situation non stationnaire : consommation d'électricité



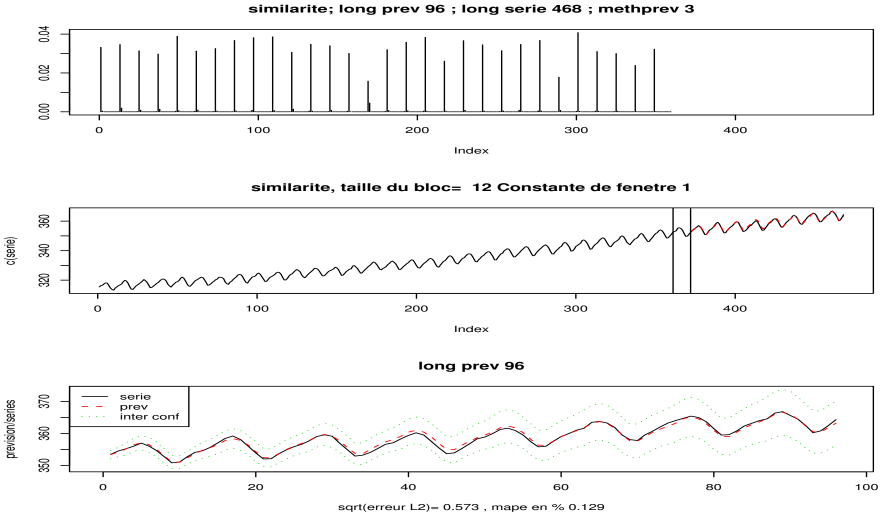
$r = 48$ (une journée) et $C_h = 2$.

Estimation des distributions conditionnelles aux différents horizons.



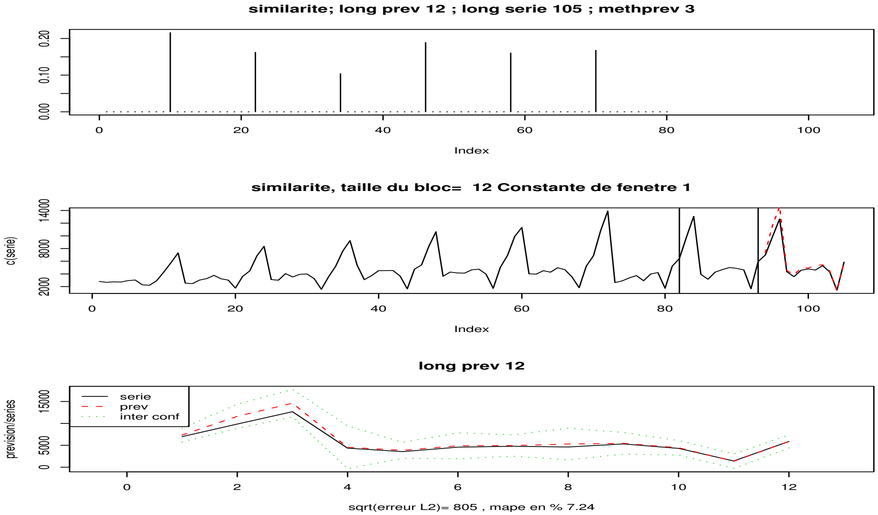

Les Distributions conditionnelles aux divers horizons
0.01
0.005
0
5000
4500
4000
3500
3000
2500
2000
0
5
10
15
20
25
30
35
40
45
50

Concentrations du CO2 dans l'atmosphère



$r = 12$ et $C_h = 2$.

Ventes de champagne



$r = 12$ et $C_h = 1$.

Modèle Markovien

On suppose que le processus (X_n) s'écrit sous la forme

$$X_n = g(X_{n-1}, \dots, X_{n-r}) + \varepsilon_n$$

où

- $(\varepsilon_n)_{n\in\mathbb{Z}}$ est un bruit blanc centré,
- On suppose que pour tout n, ε_n est indépendant de $(X_{n-1}, \dots, X_{n-r})$.
- la fonction g est inconnue.

Remarque

Dans cette approche, on ne fait pas d'hypothèse paramétrique sur g.

Construction du prédicteur

On suppose que $\mathbb{E}|X_j| < \infty$ pour tout j.
On cherche à prévoir la valeur de X_{n+h}, en se basant sur l'observation des valeurs de $X_n, \ldots, X_1$.

$$\mathbb{E}(X_{n+h}|X_n, \ldots) = \mathbb{E}(X_{n+h}|X_n, \ldots, X_{n-r+1}) = \phi(X_{n-r+1}, \ldots, X_n),$$

est (au sens de L^2) le meilleur prédicteur de X_{n+h} conditionnellement à $X_n, \ldots, X_1$.

On est donc ramené à un problème d'estimation non paramètrique d'une fonction.

Estimation de ϕ par un estimateur à noyau

L'estimateur "classique" $\hat{\phi}_n(x)$ s'écrit comme une combinaison linéaire des X_j

$$\hat{\phi}_n(x_1, \ldots, x_r) \quad = \quad \sum_{j=r}^{n-1} X_{j+1} W_{n,j}$$

où le poids $W_{n,j}(x)$ mesure la proximité/similarité entre

$$x_1^r = x_1, \ldots, x_r \quad \text{et} \quad X_{j+1-r}^j = (X_{j+1-r}, \ldots, X_j).$$

Le prédicteur est $\hat{\phi}_n(X_{n-r+1}, \ldots, X_n)$

.

Mesure de la similarité

$$\tilde{W}_{n,j} = K\left(\frac{x_1^r - X_{j+1-r}^j}{h_n}\right) \qquad W_{n,j} = \tilde{W}_{n,j} / \sum \tilde{W}_{n,j}$$

où

- $h_n \to 0$ quand $n \to \infty$
- Un exemple de noyau, le noyau gaussien :

$$K(y_1, \ldots, y_r) = (2\pi)^{-r/2} \exp\left(-\frac{1}{2}\sum_{j=1}^{r} y_j^2\right)$$

Situation non-stationnaire

Mesure des similitudes

- X et Y désignent deux vecteurs de R^r,
- m_X et m_Y les moyennes respectives
- σ_X et σ_Y les écart-types respectifs

$$K^{(1)} = K\left(\frac{(X - m_X) - (Y - m_Y)}{h}\right)$$

$$K^{(2)} = K\left(\frac{\left(\frac{X-m_X}{\sigma_X}\right) - \left(\frac{Y-m_Y}{\sigma_Y}\right)}{h}\right)$$

On se ramène alors à une même échelle en tendance et en variabilité.

Expression du prédicteur

On note

- Y le bloc témoin
- $X = X(t)$ le bloc du passé $(X_{t-r+1}, \dots X_t)$

Alors les prédicteurs s'écrivent sous la forme

$$\hat{X}(n:h) = \sum_{j=r}^{n-h} W_{n,j}^{(1)} \left(X_{j+h} + m_Y - m_{X(j)}\right)$$

$$\hat{X}(n:h) = \sum_{j=r}^{n-h} W_{n,j}^{(2)} \left(\frac{X_{j+h} - m_{X(t)}}{\sigma_{X(t)}} \cdot \sigma_Y + m_Y\right)$$