

Introduction aux séries chronologiques et à la prévision

1. Définition et domaines d'application

2. Lissage

3. Processus stochastiques

1. Définition et domaines d'application

1. Définition et domaines d'application

SERIE CHRONOLOGIQUE

=

Suite d'observations indiquées par le temps

$$\mathbf{y}_t, \quad t = 1, \dots, T$$

En général, dates d'observation régulières :

t = heure ...

$jour$...

$mois$...

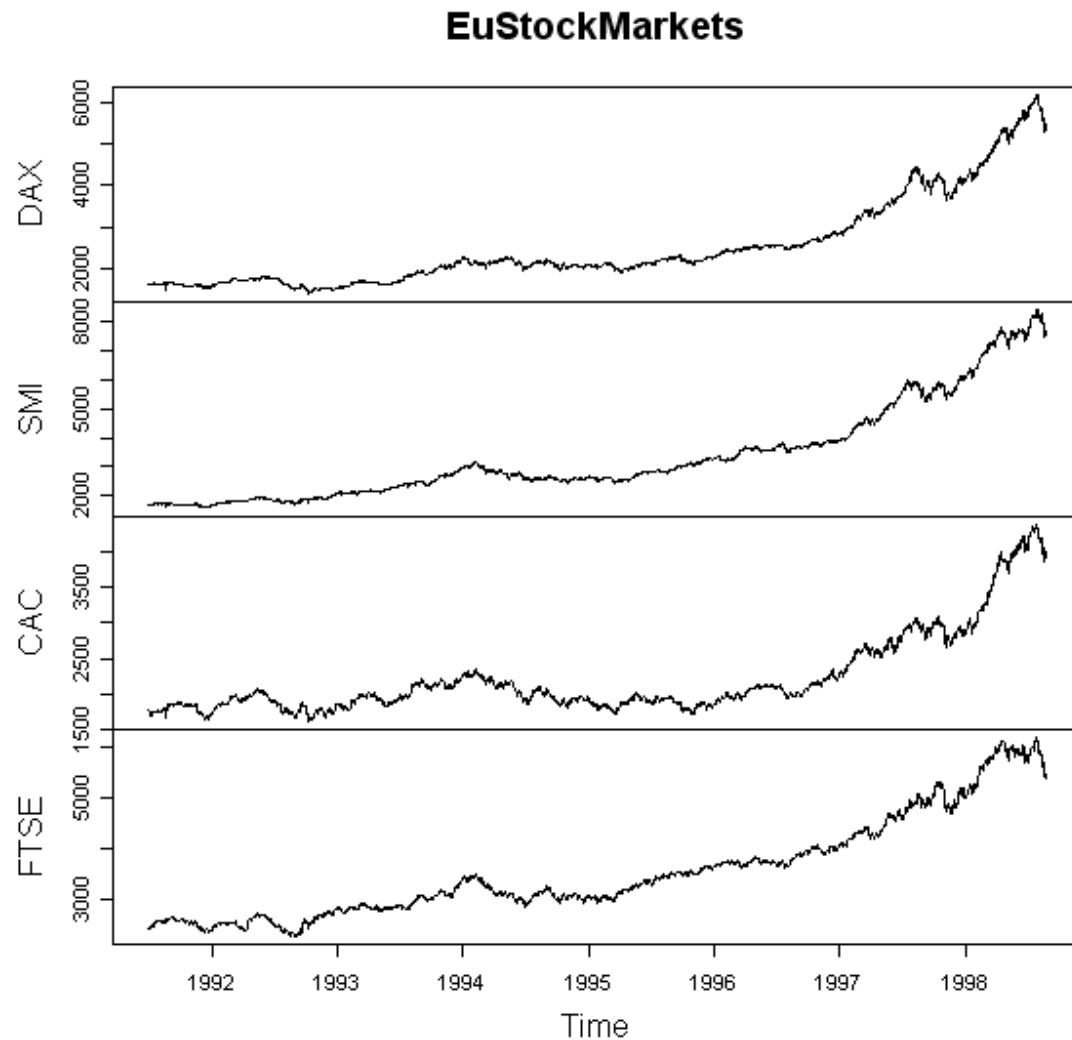
$année$...

1. Définition et domaines d'application

Domaines d'application

Finance

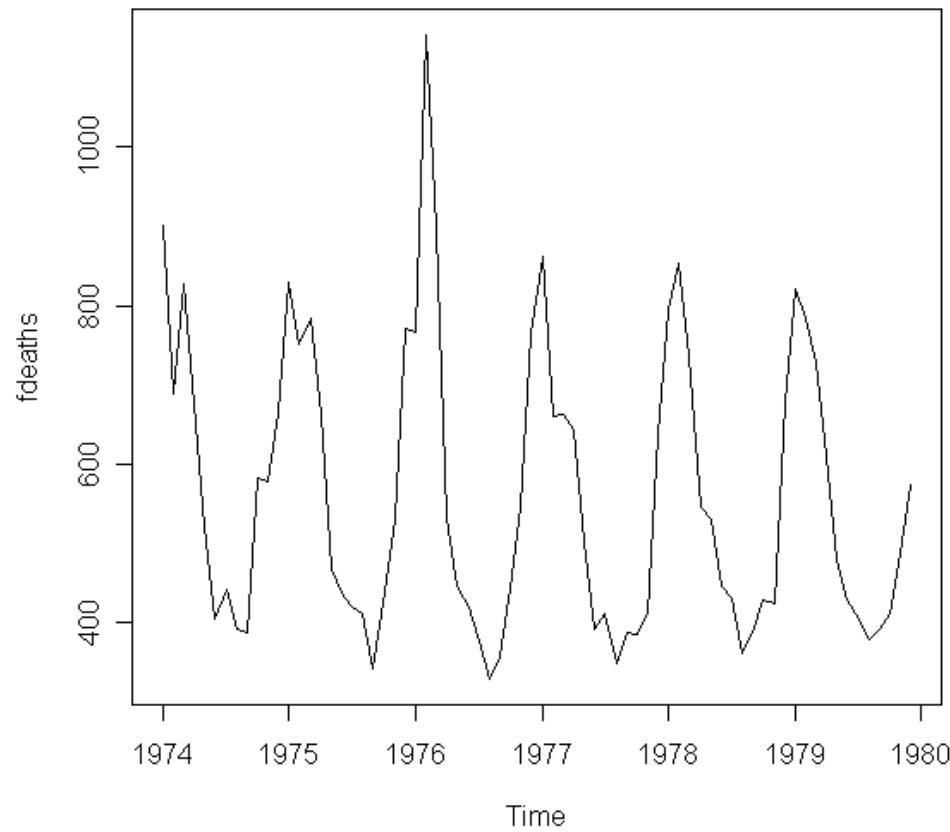
Cours de bourse
quotidiens



1. Définition et domaines d'application

Epidémiologie

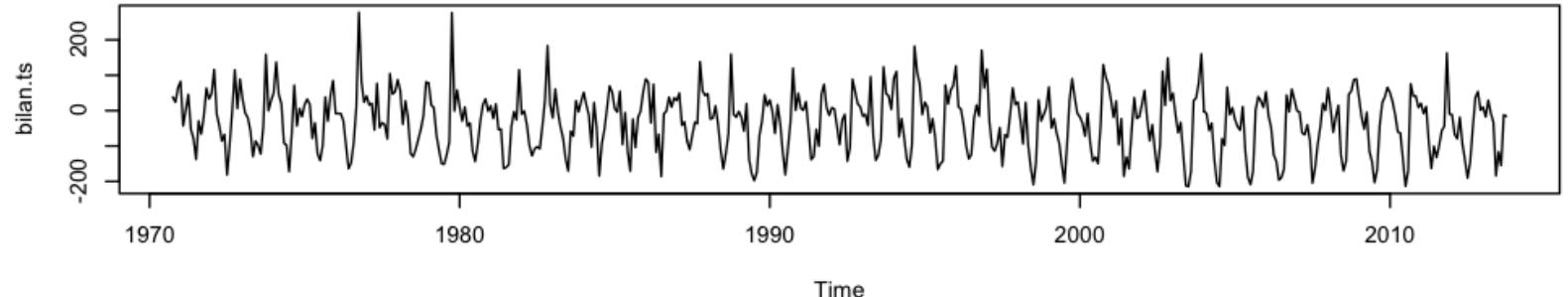
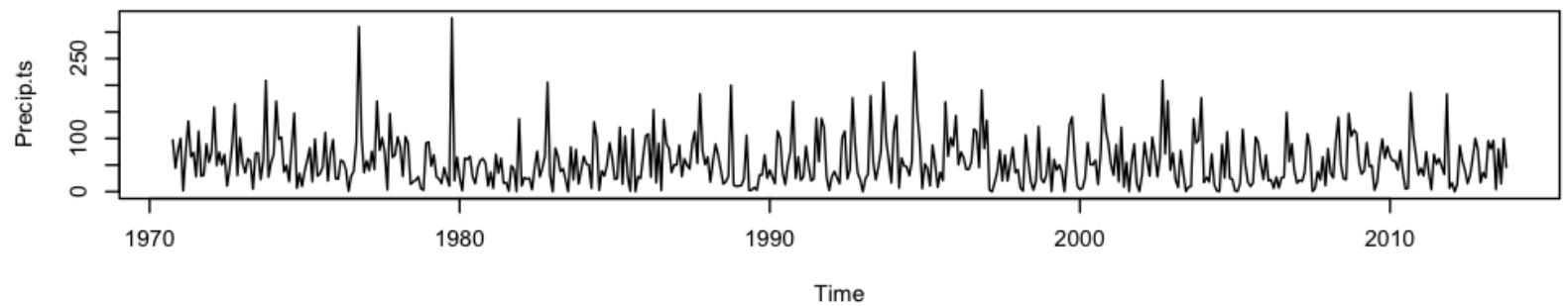
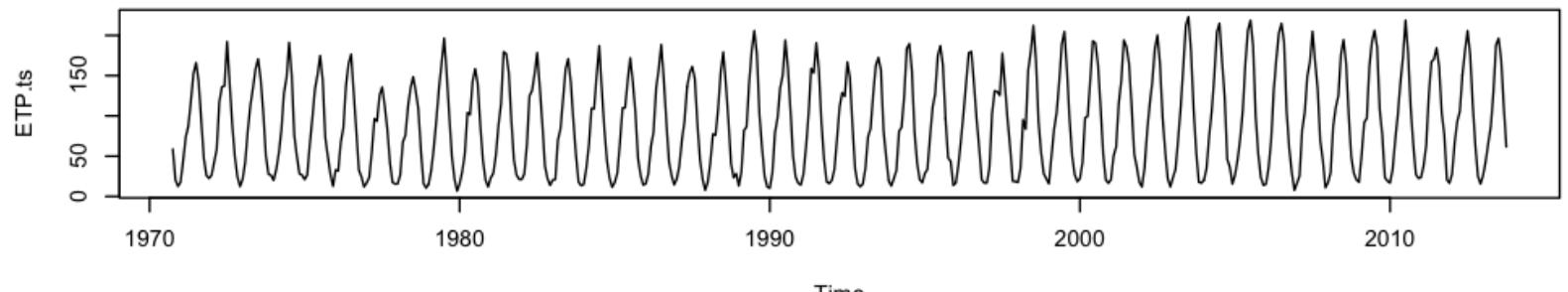
Décès mensuels
dus à des
maladies pulmonaires
(UK)



1. Définition et domaines d'application

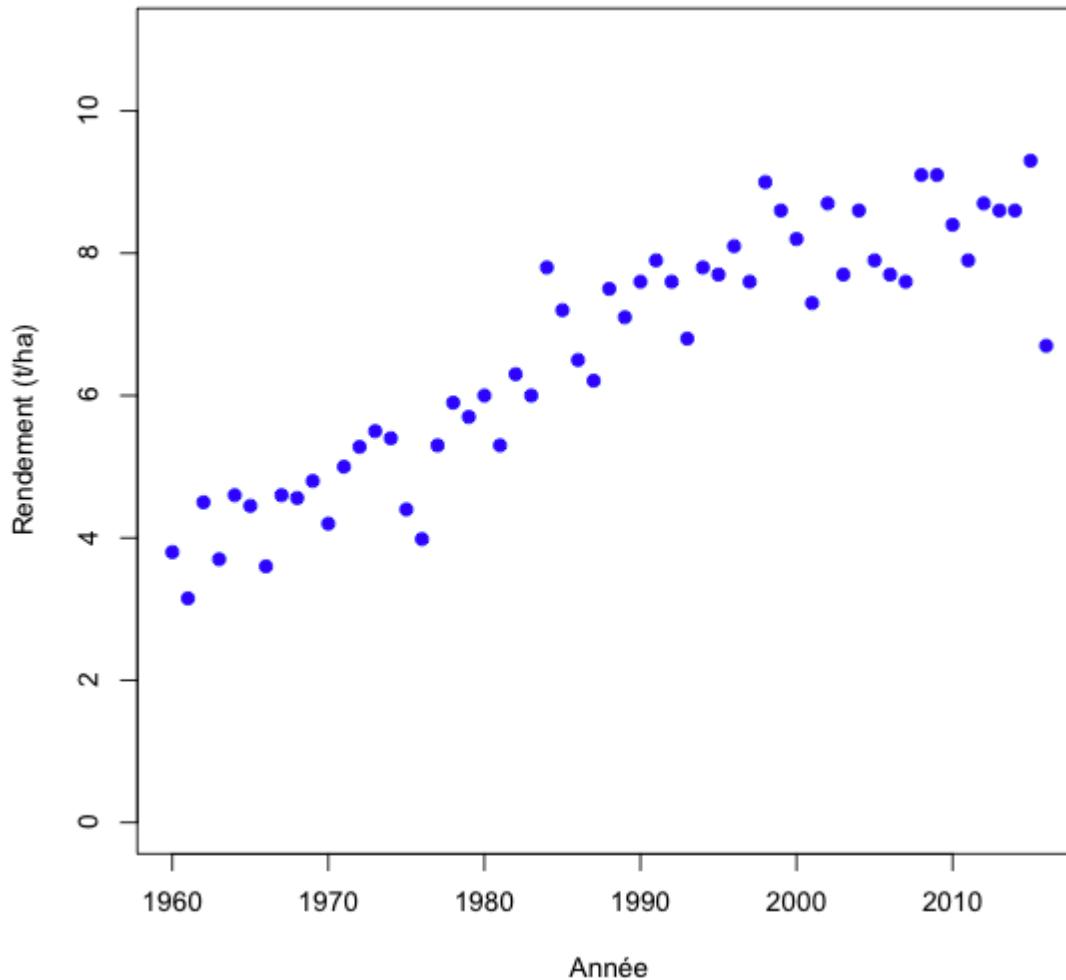
Météorologie/climatologie :

ETP (mm), Précipitation (mm), et bilan (P-ETP) dans le Vaucluse



1. Définition et domaines d'application

Écologie
Agronomie
Environnement
Hydrologie
etc.



1. Définition et domaines d'application

Objectifs de l'étude d'une série chronologique

- **Analyser un phénomène**
« Décrire/comprendre/juger l'évolution de la série »
- **Prévision**
« Prévoir les valeurs futures de Y_t »

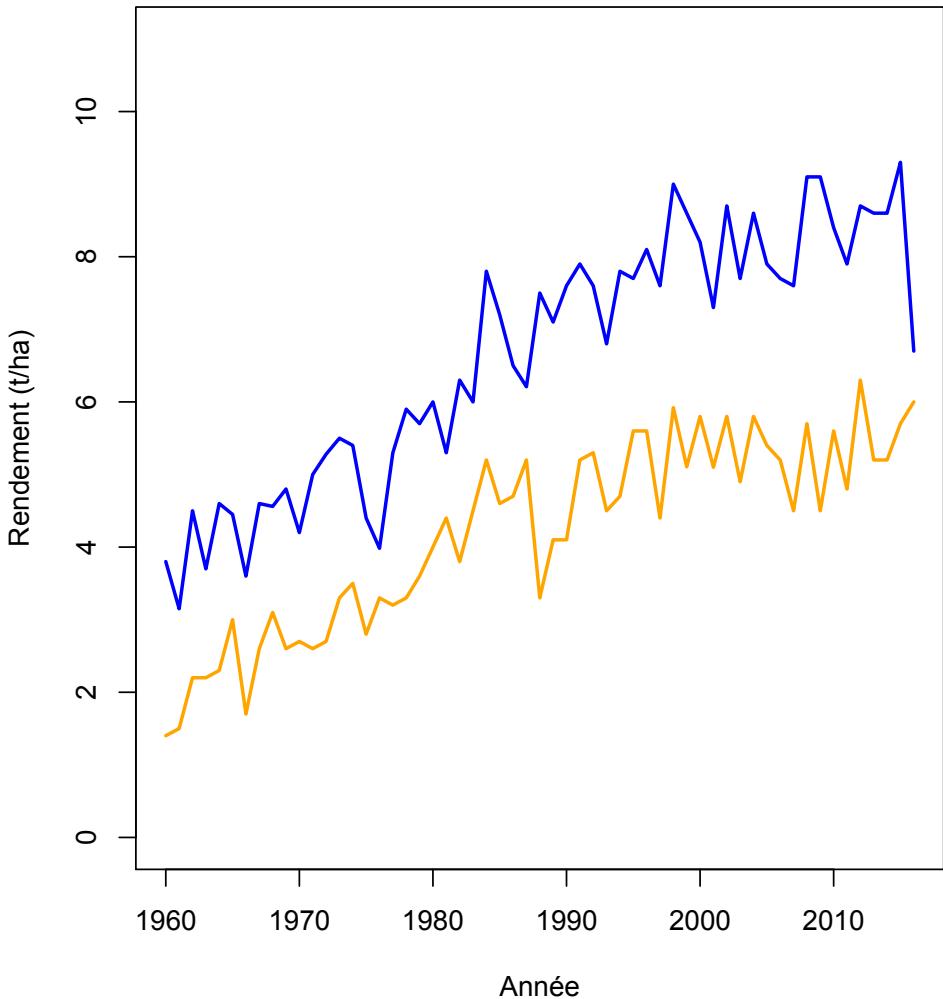
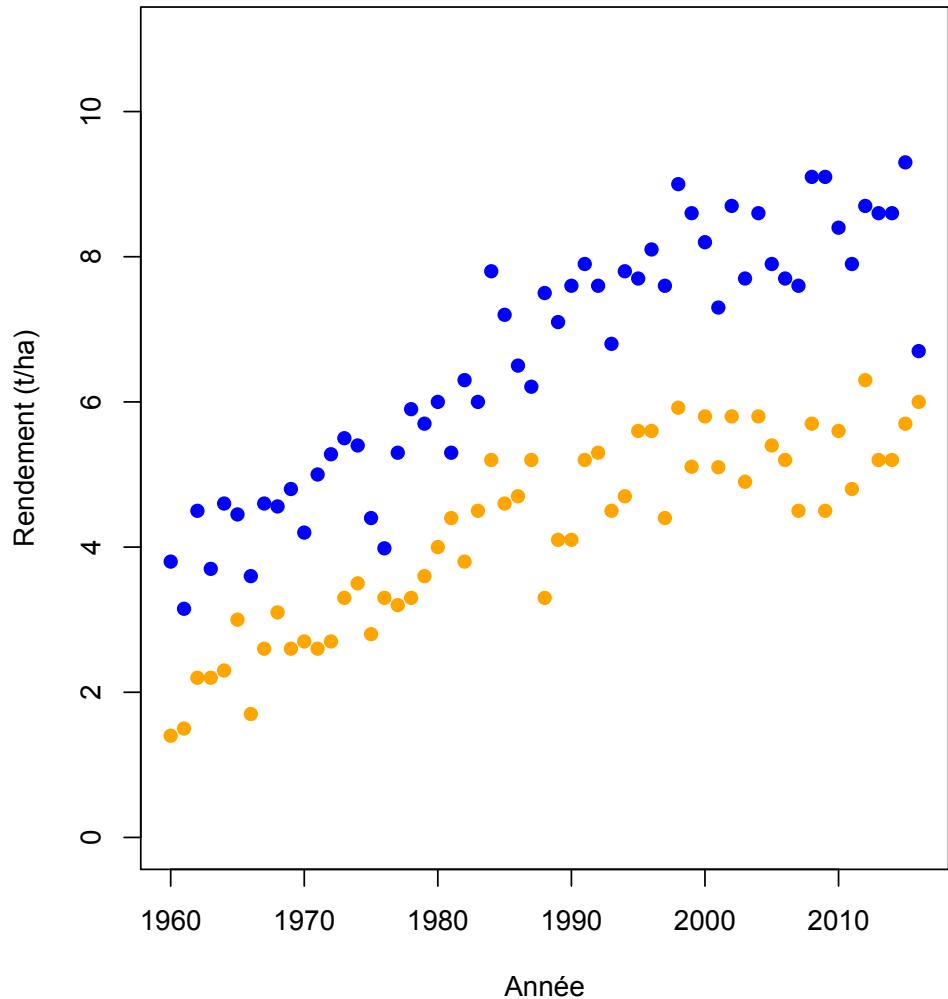
1. Définition et domaines d'application

Exemples de questions

- Quelle tendance ?
- Quelle saisonnalité ?
- Quelle valeur future ?

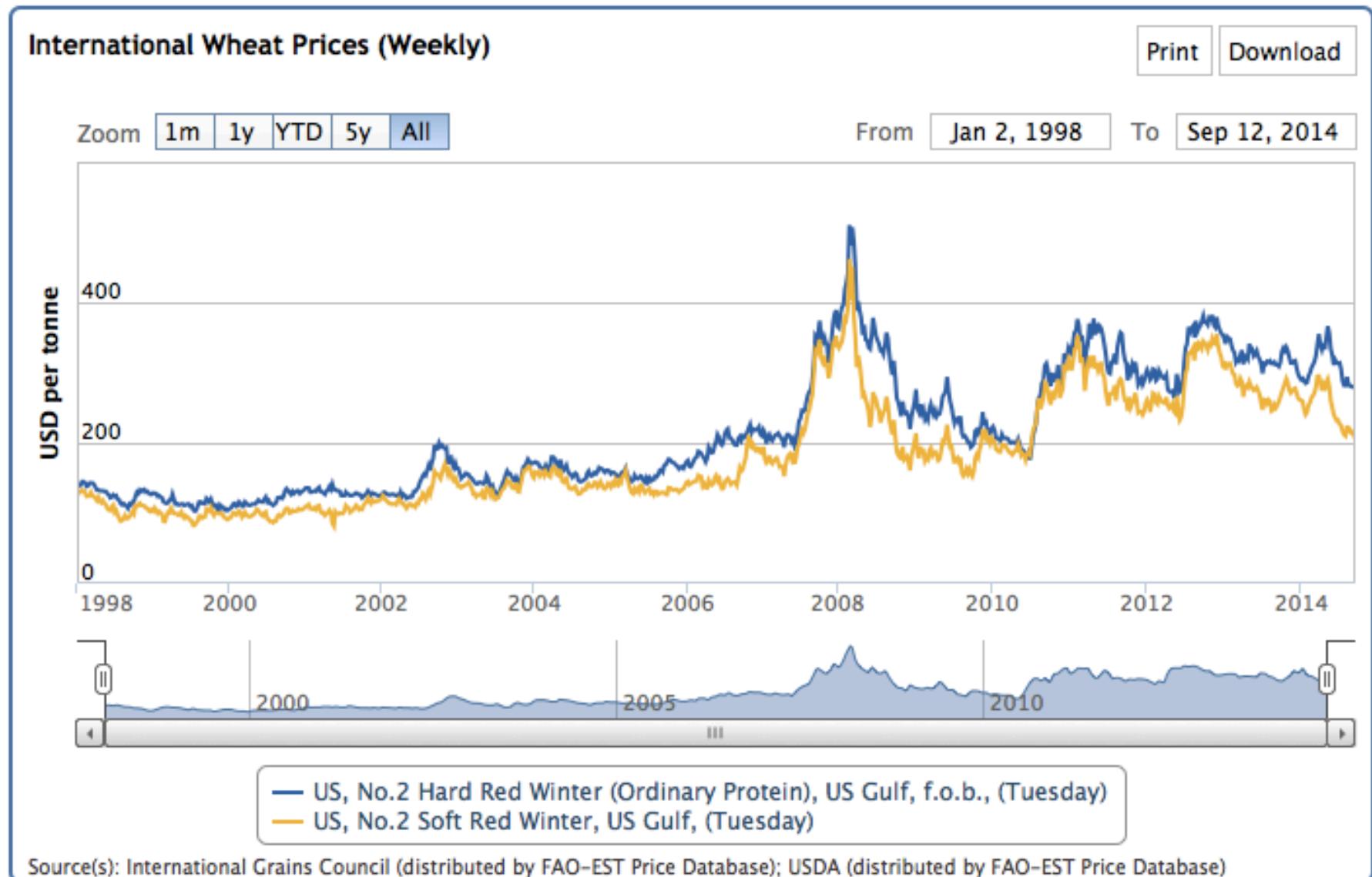
1. Définition et domaines d'application

Quelles tendances ?



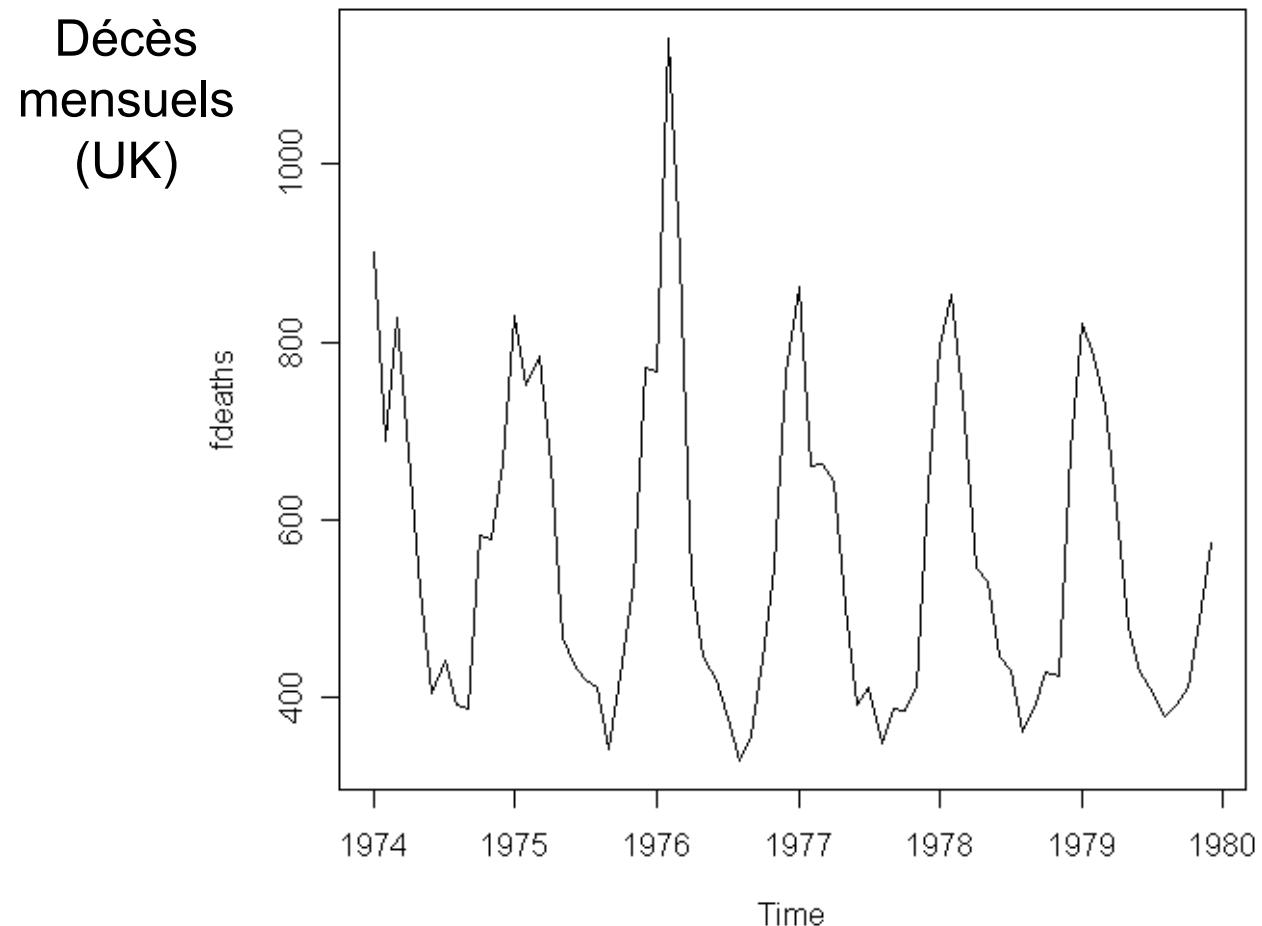
1. Définition et domaines d'application

**Comment les prix du blé ont-ils évolués ?
Quel est leur « volatilité » ?**



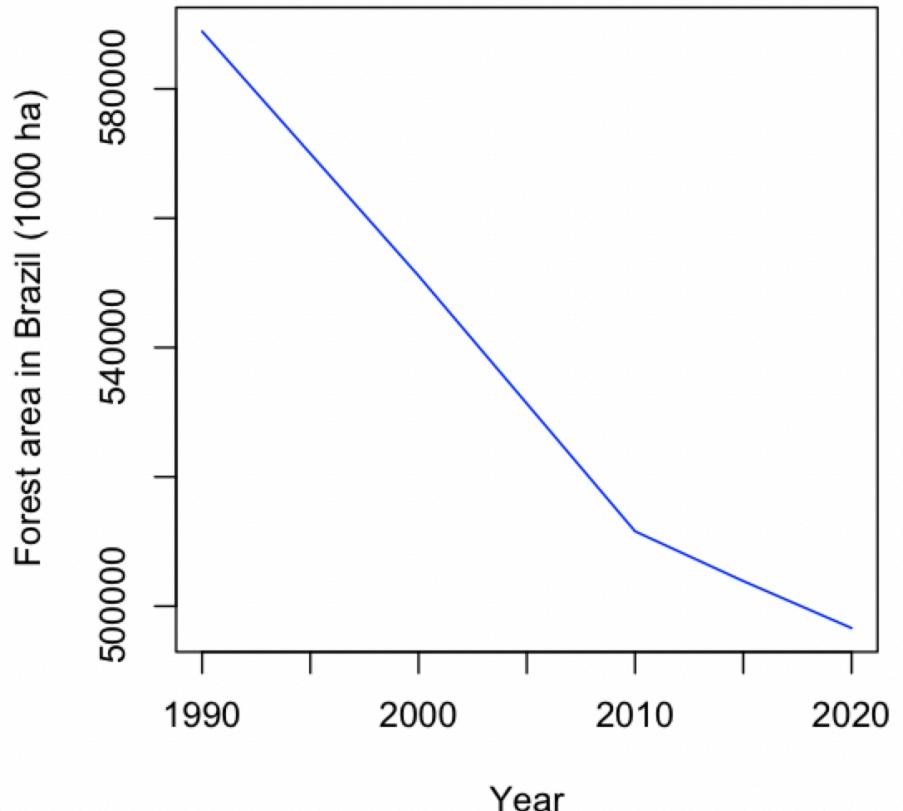
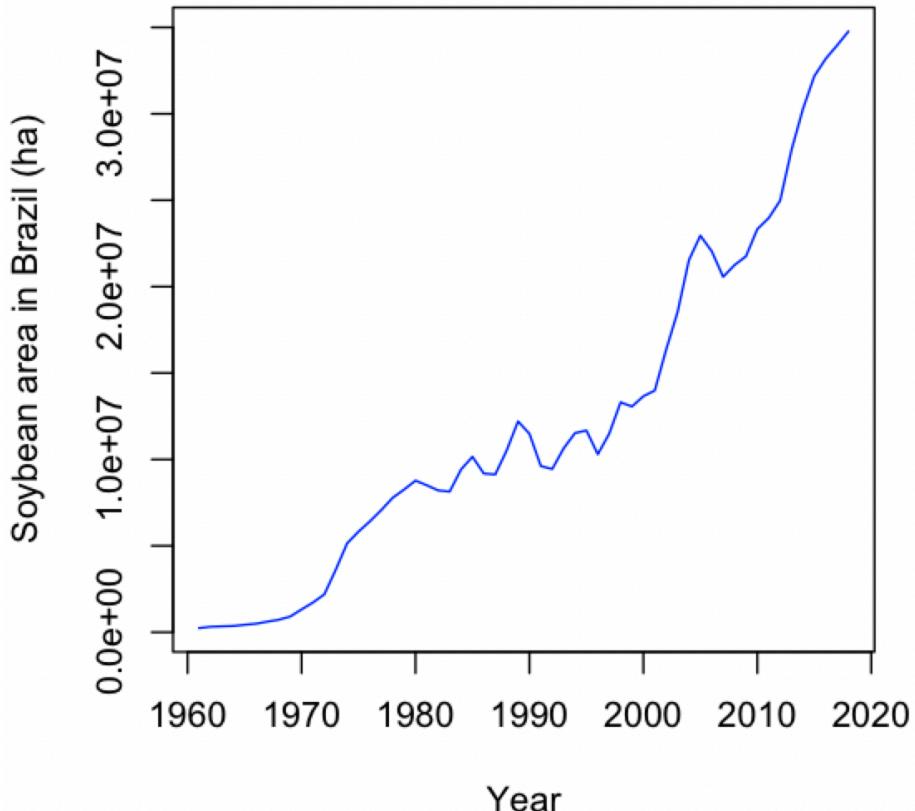
1. Définition et domaines d'application

Quels sont les effets des saisons sur le nombre de décès ?



1. Définition et domaines d'application

Quelles valeurs futures ?



Surfaces de soja et de forêt au Brésil en 2025 ?

1. Définition et domaines d'application

Les étapes de l'analyse

- i. Modélisation (définition des équations du modèle)**
- ii. Estimation des paramètres**
- iii. Prévision avec le modèle et évaluation des erreurs**

1. Définition et domaines d'application

Des méthodes très diversifiées

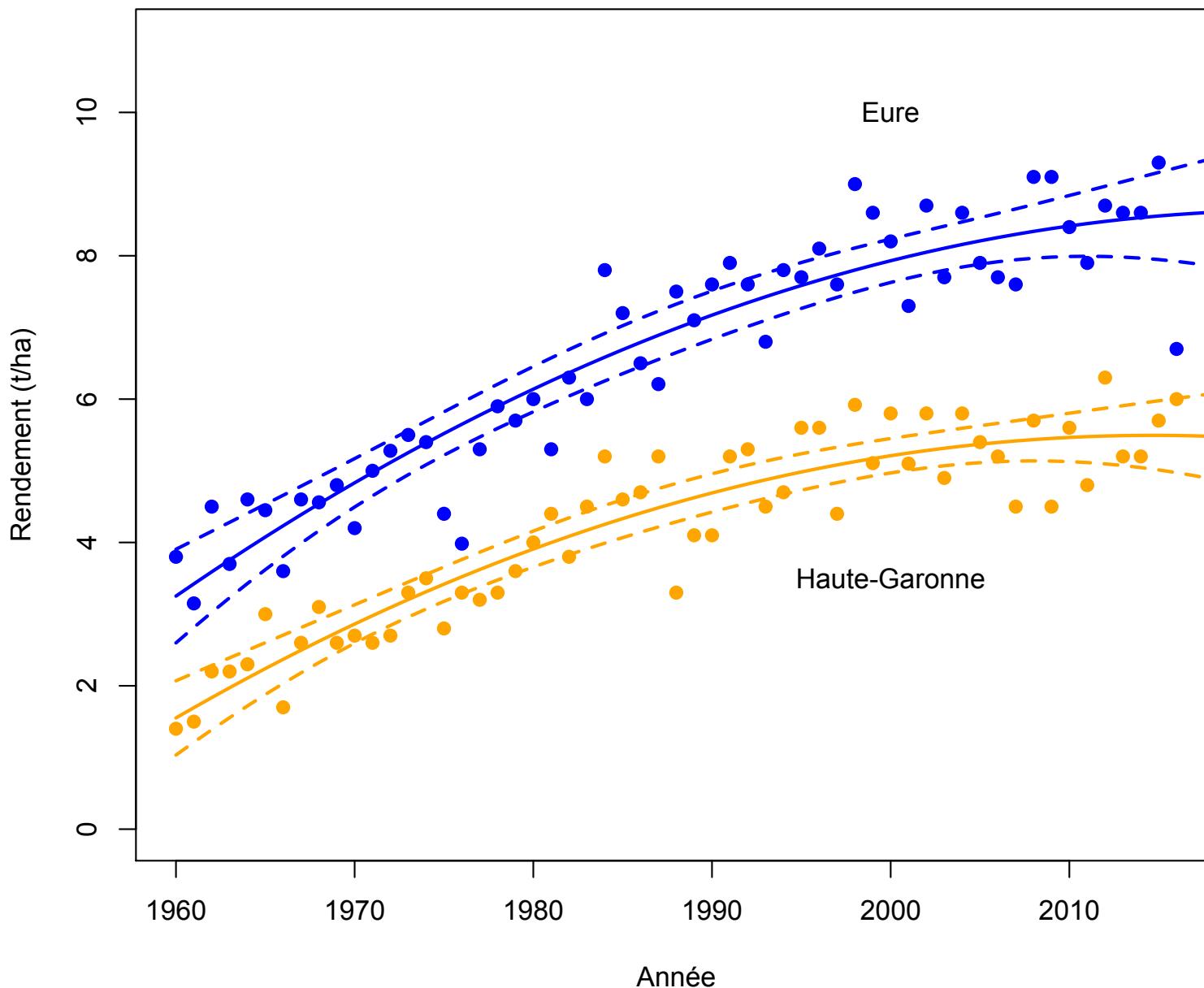
des méthodes très classiques...

Régression polynomiale

$$Y_i = \alpha + \beta X_i + \gamma X_i^2 + \varepsilon$$

Estimation avec la fonction lm() ou glm()

Régression quadratique



```
Mod_1<-lm(Rendement_eure~Annee+I(Annee^2), data=DATA)
summary(Mod_1)
```

```
Mod_2<-lm(Rendement_hautegaronne~Annee+I(Annee^2), data=DATA)
summary(Mod_2)
```

```
plot(DATA$Année,DATA$Rendement_eure,xlab="Année",ylab="Rendement (t/ha)",  
pch=19,col="blue",ylim=c(0,11))  
X<-1960:2020  
pred<-coef(Mod_1)[1]+coef(Mod_1)[2]*X  
predIC<-predict(Mod_1,newdata=data.frame(Année=X),interval="confidence",level=0.99)  
predIC  
lines(X,predIC[,1],lty=1,lwd=2, col="blue")  
lines(X,predIC[,2],lty=2,lwd=2, col="blue")  
lines(X,predIC[,3],lty=2,lwd=2, col="blue")
```

1. Définition et domaines d'application

Des méthodes très diversifiées

des méthodes très classiques...

**... mais aussi des méthodes spécifiques aux
séries chronologiques !**

1. Définition et domaines d'application

2. Lissage

3. Processus stochastiques

2. Le lissage

2. Lissage

Principes

- « Lisser » les aléas afin d'observer la tendance
- Ajuster une fonction du temps à des données et extrapoler
- Donner plus de poids aux observations les plus récentes

De nombreuses méthodes

- Moyenne mobile
- Lissage exponentielle
- Méthode de Holt-Winters

...

2. Lissage – moyenne mobile

Moyenne mobile

Permet de **décomposer** la série initiale en fonction d'une tendance et d'une composante saisonnière

Décomposition additive

$$Y_t = \text{Tendance}(t) + \text{Saison}(t) + \text{Résidu}(t)$$

Décomposition multiplicative

$$Y_t = \text{Tendance}(t) * \text{Saison}(t) * \text{Résidu}(t)$$

2. Lissage – moyenne mobile

Moyenne mobile k (k impair)

$$Z_t = \frac{y_{t-m} + \dots + y_t + \dots + y_{t+m}}{k}$$

avec $m = (k - 1)/2$

Pour une décomposition additive :

- La moyenne des résidus $y_t - Z_t$ obtenus pour une saison permet d'estimer l'effet de cette saison estim(s_t)
- $y_t - \text{estim}(s_t) = \text{série corrigée des variations saisonnières}$

2. Lissage – moyenne mobile

Moyenne mobile k (k pair)

$$Z_t^a = \frac{y_{t-m} + \dots + y_t + \dots + y_{t+m-1}}{k}$$

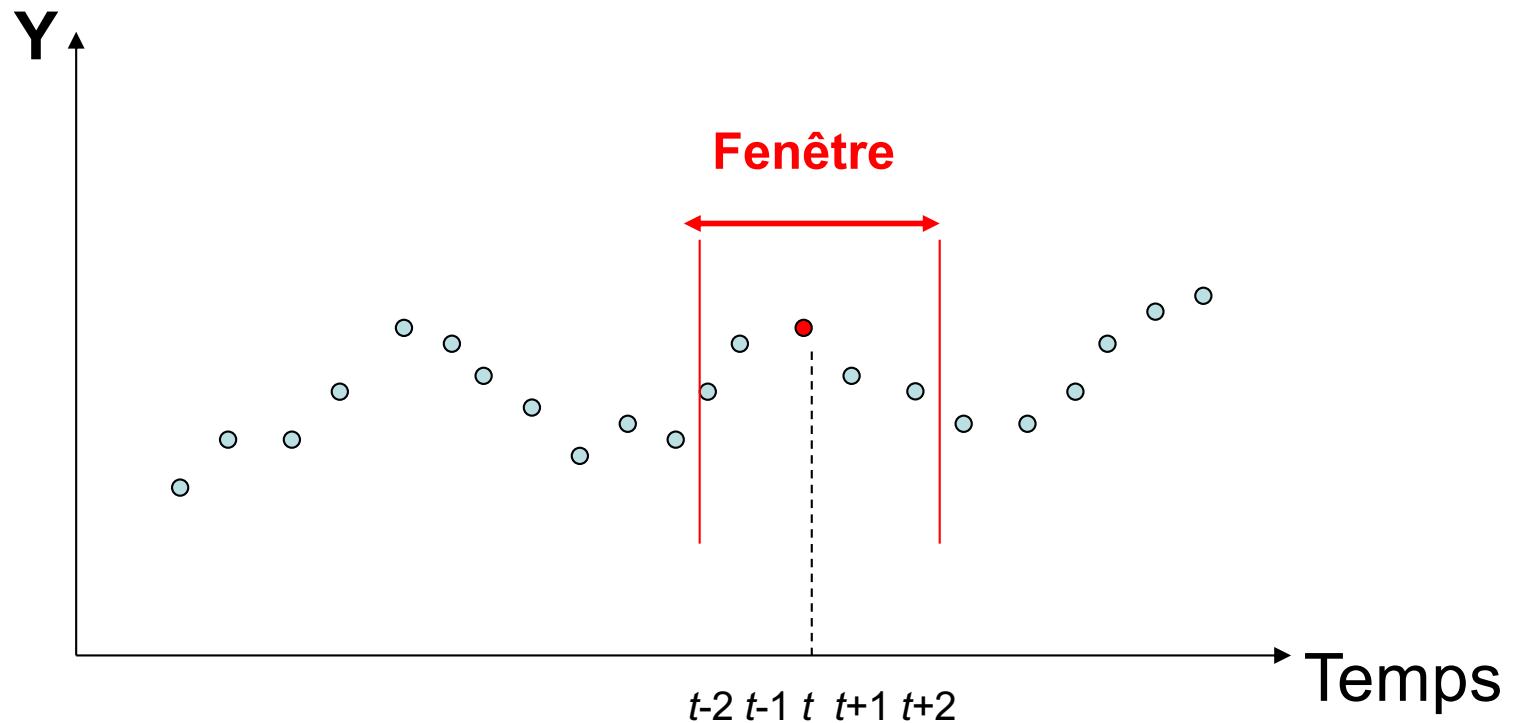
$$Z_t^b = \frac{y_{t-m+1} + \dots + y_t + \dots + y_{t+m}}{k}$$

$$Z_t = \frac{Z_t^a + Z_t^b}{2}$$

avec $m = k / 2$

2. Lissage – moyenne mobile

Moyenne mobile 5



Quel est l'effet du coefficient de lissage de la moyenne mobile ?

2. Lissage – moyenne mobile

Expression utile pour la prévision

$$Z_{T+1} = \frac{\sum_{j=0}^k y_{T-j}}{k + 1}$$

2. Lissage – moyenne mobile

Moyenne mobile - estimation des paramètres

Un paramètre à estimer: k

- k faible: lissage faible.
- k fort: lissage fort.
- Estimation en fonction de la saisonnalité ou en minimisant les erreurs de prédiction.

2. Lissage – moyenne mobile

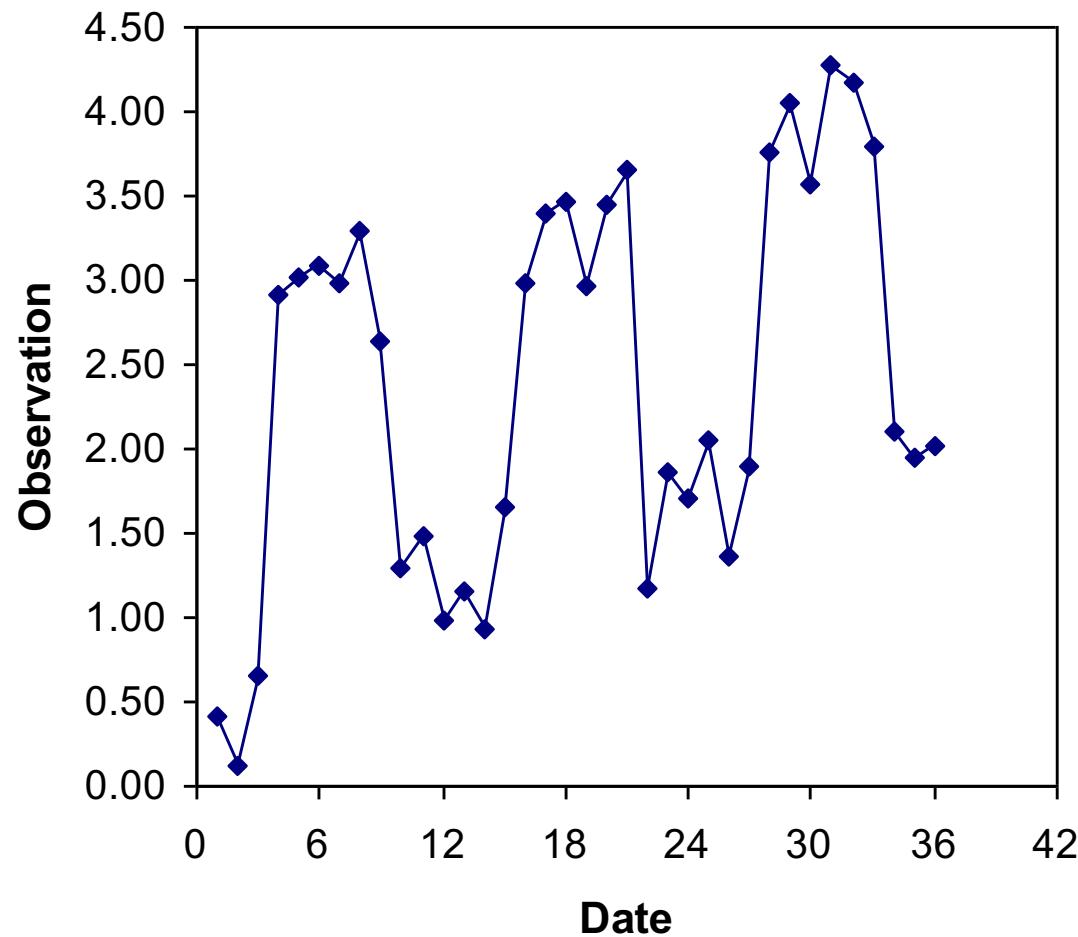
Un exemple artificiel

Nombre de glaces consommées par personne chaque mois.

Trois années de données.

	Mois	Tendance	Saisonnalité	Aléa	Série
1	1	1.50	-1.00	-0.08	0.42
2	2	1.55	-1.00	-0.44	0.11
3	3	1.60	-1.00	0.05	0.65
4	4	1.65	1.00	0.27	2.92
5	5	1.70	1.00	0.31	3.01
6	6	1.75	1.00	0.33	3.08
7	7	1.80	1.00	0.18	2.98
8	8	1.85	1.00	0.44	3.29
9	9	1.90	1.00	-0.27	2.63
10	10	1.95	-1.00	0.34	1.29
11	11	2.00	-1.00	0.49	1.49
12	12	2.05	-1.00	-0.07	0.98
13	1	2.10	-1.00	0.06	1.16
14	2	2.15	-1.00	-0.22	0.93
15	3	2.20	-1.00	0.45	1.65
16	4	2.25	1.00	-0.27	2.98
17	5	2.30	1.00	0.10	3.40
18	6	2.35	1.00	0.11	3.46
19	7	2.40	1.00	-0.44	2.96
20	8	2.45	1.00	0.00	3.45
21	9	2.50	1.00	0.16	3.66

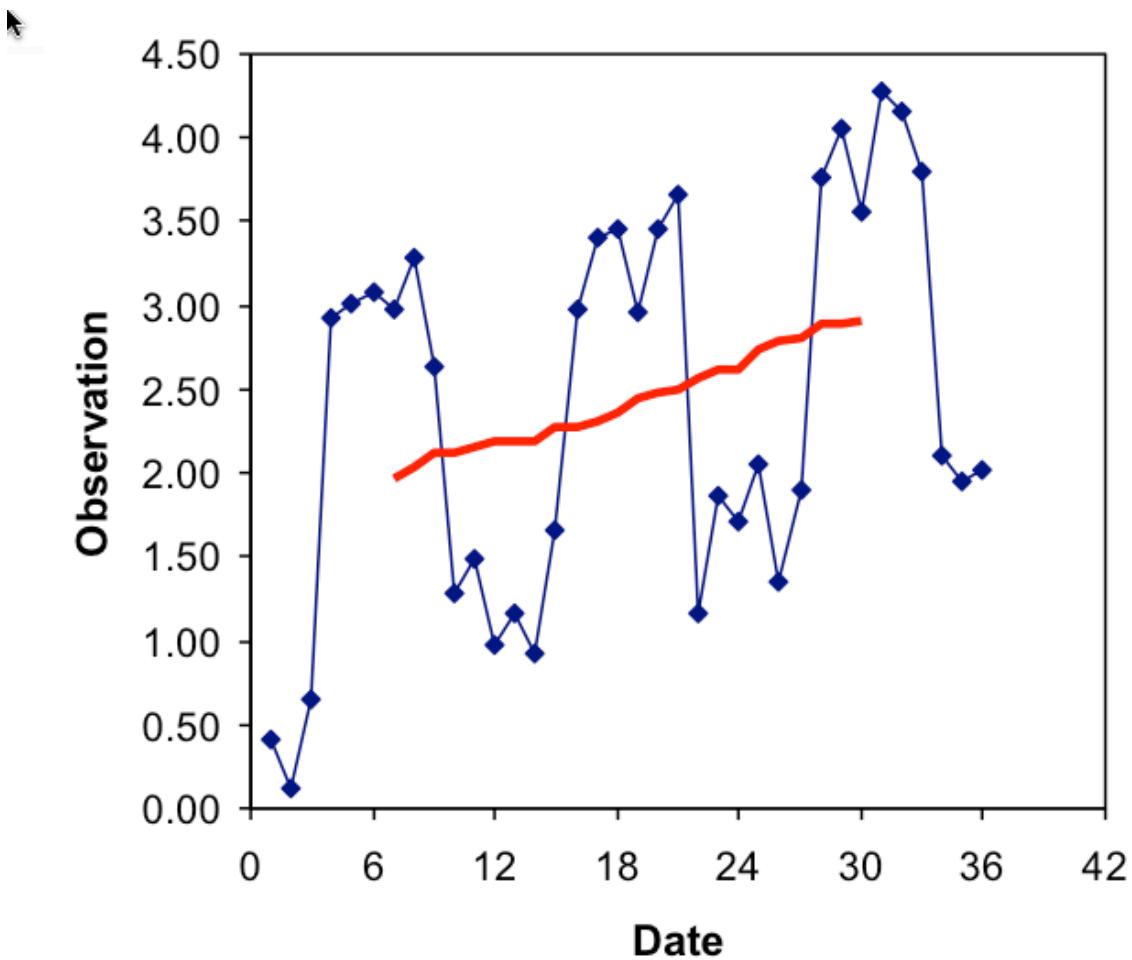
2. Lissage – moyenne mobile



2. Lissage – moyenne mobile

Série	MM12a	MM12b	MMf	Delta	Série-MMf
0.42	-			-	-
0.11	-			-	-
0.65	-			-	-
2.92	-			-	-
3.01	-			-	-
3.08	1.90			-	-
2.98	1.97	1.90	1.93	-	1.05
3.29	2.03	1.97	2.00	0.06	1.29
2.63	2.12	2.03	2.08	0.08	0.56
1.29	2.12	2.12	2.12	0.04	-0.83
1.49	2.15	2.12	2.14	0.02	-0.65
0.98	2.19	2.15	2.17	0.03	-1.19
1.16	2.18	2.19	2.18	0.01	-1.03

2. Lissage – moyenne mobile



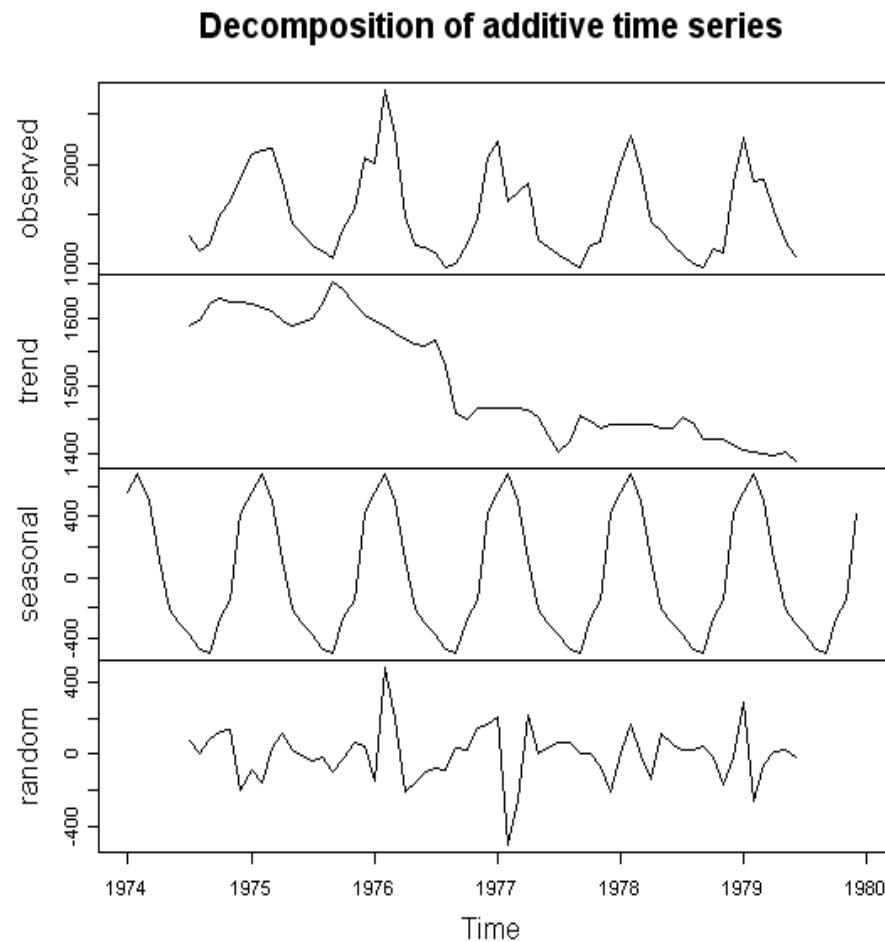
Moyenne des Delta = 0.042

Moyenne résidus hiver = -0.96

Moyenne résidus été = 0.94

2. Lissage – moyenne mobile

Données de décès UK

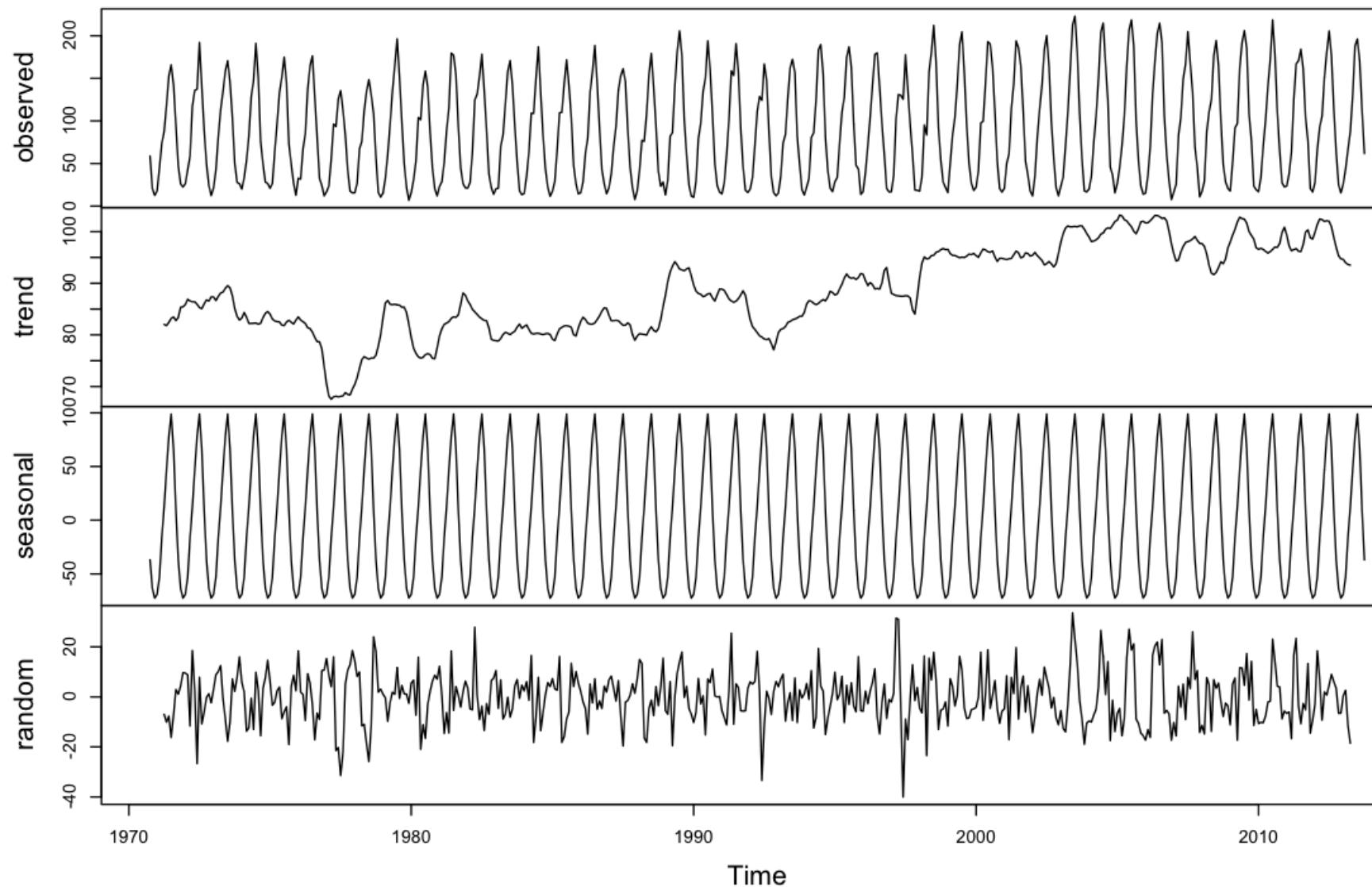


Approche moyennes mobiles
(fonction decompose de R)

2. Lissage – moyenne mobile

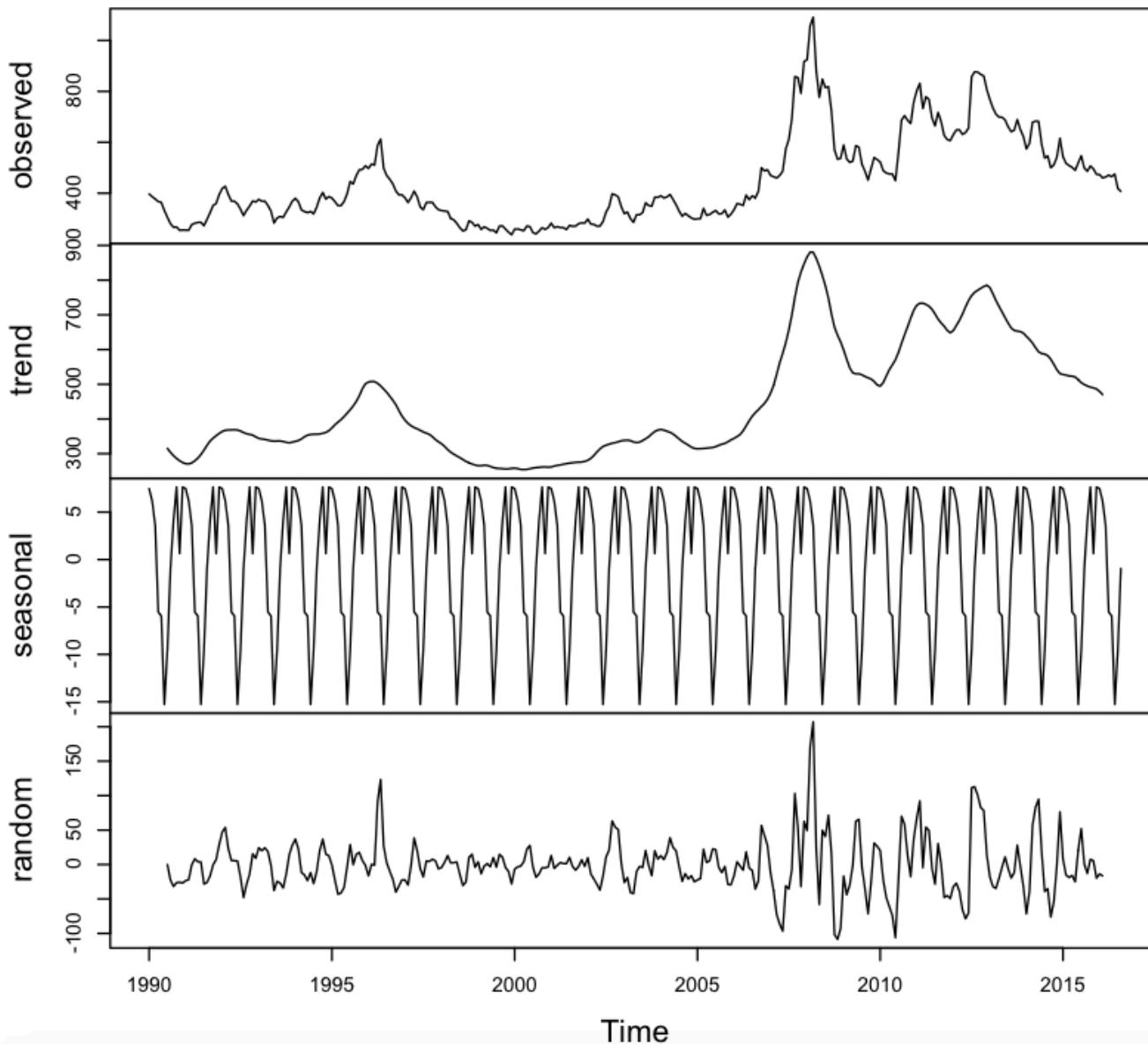
Données d'évapotranspiration du Vaucluse

Decomposition of additive time series



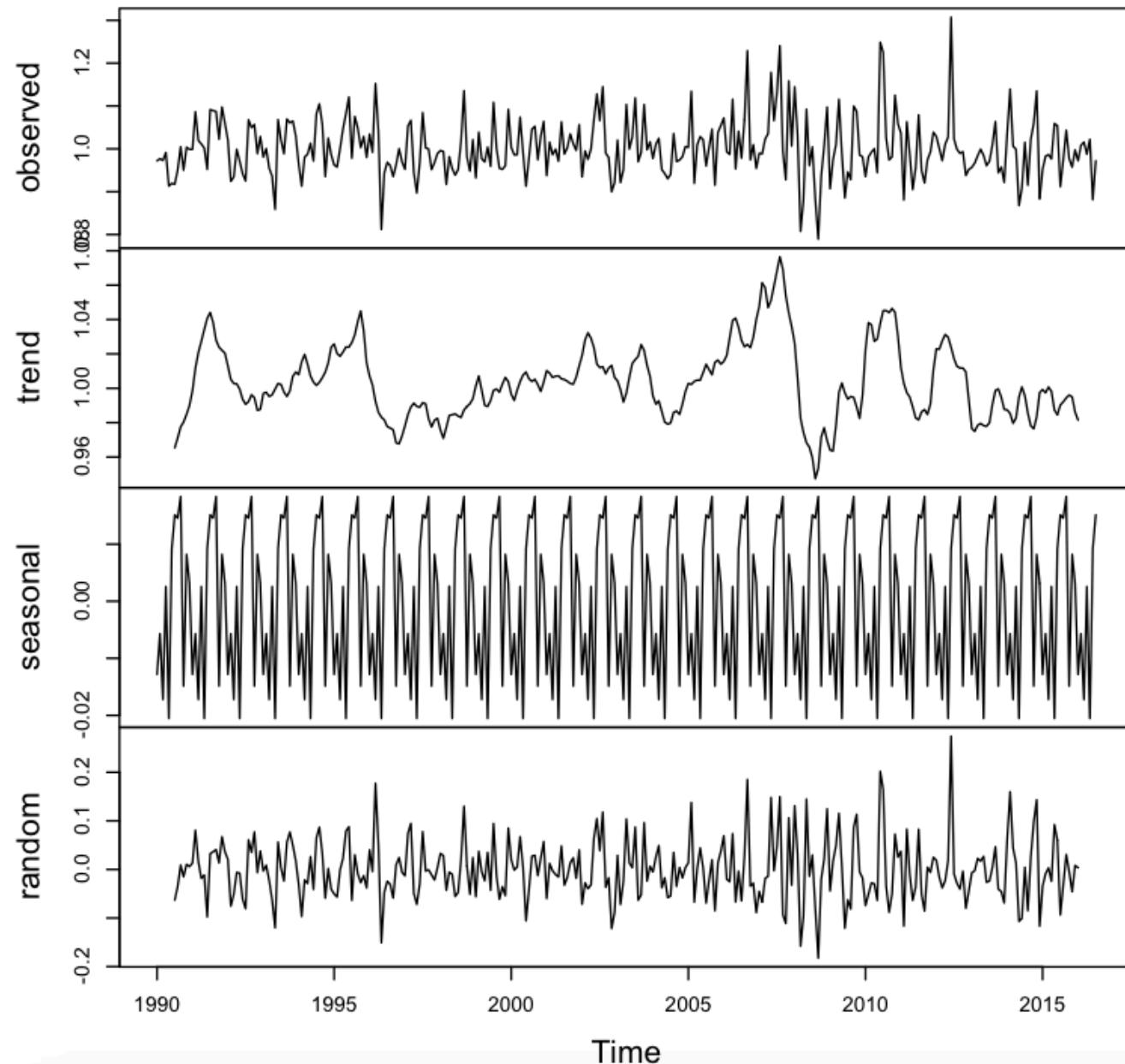
2. Lissage – moyenne mobile

Prix du blé (Chicago, US cent per gallon)



2. Lissage – moyenne mobile

Ratio du prix du blé (Chicago)



Un outil simple mais puissant: la fonction decompose() de R

```
TAB<-read.table("PrixBle.txt", sep="\t", header=T)
Prix<-TAB$Prix
Prix.ts=ts(Prix,start=c(1990,1),frequency = 12)
plot(decompose(Prix.ts))

l_Prix<-log(Prix)

# exp( ln(x(t)) – ln(x(t-1)) )= exp( ln(x(t)/x(t-1)) ) = x(t)/x(t-1)

dl_Prix<-exp(diff(l_Prix))

dl_Prix.ts = ts(dl_Prix,start=c(1990,1),frequency = 12)

plot(decompose(dl_Prix.ts))
```

Exercice

Décomposer la série chronologique
mensuelle du bilan climatique du Vaucluse

```
##Lecture du fichier  
EVT=read.csv("ETPpluvioVC.csv",sep=";",skip=9)
```

```
#Evapotranspiration potentielle
```

```
ETP = EVT[,5]
```

```
#Precipitation
```

```
Precip = EVT[,4]
```

```
##Création d'objets "time series"
```

```
ETP.ts = ts(ETP,start=c(1970,10),frequency = 12)
```

```
Precip.ts = ts(Precip,start=c(1970,10),frequency = 12)
```

```
bilan = Precip - ETP
```

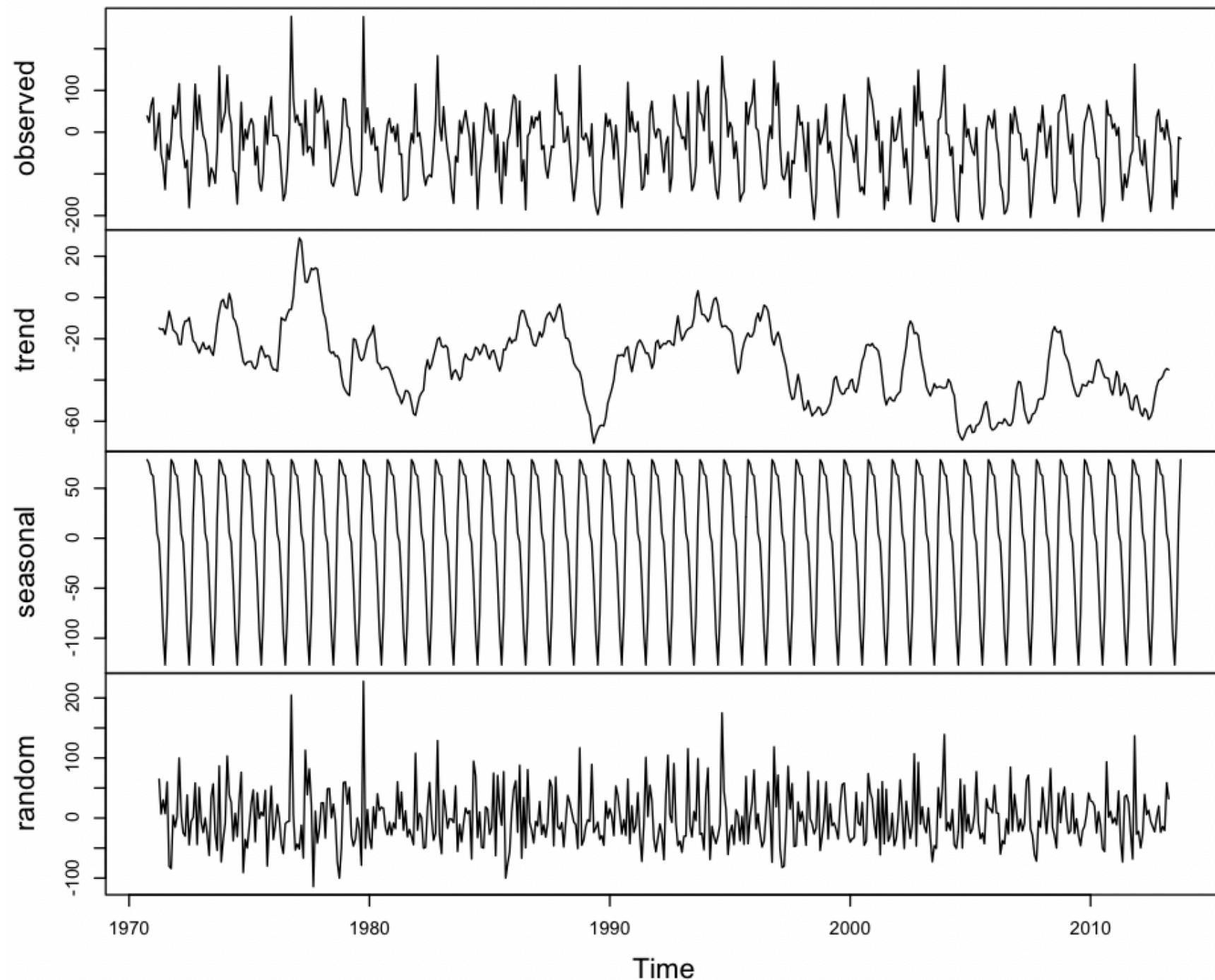
```
bilan.ts = ts(bilan,start=c(1970,10),frequency = 12)
```

```
##Décomposition des series (tendance, saison, résidus)
```

```
dev.new()
```

```
plot(decompose(bilan.ts))
```

Decomposition of additive time series



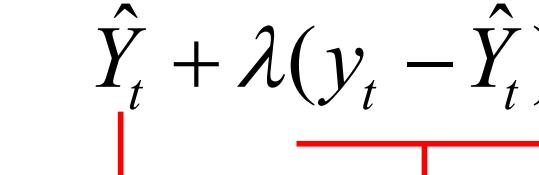
2. Lissage – Lissage exponentiel simple

Lissage exponentiel simple

La prédiction au temps $t+1$ est une somme pondérée de la prédiction au temps t et de la mesure obtenue à cette date.

$$\hat{Y}_{t+1} = \lambda y_t + (1 - \lambda) \hat{Y}_t =$$

$$\hat{Y}_t + \lambda(y_t - \hat{Y}_t)$$



$$0 < \lambda \leq 1$$

$$\hat{Y}_1 = y_1$$

Comment évolue la prévision après la dernière date de mesure ?

2. Lissage – Lissage exponentiel simple

Lissage exponentiel simple

- Le résultat dépend de la valeur choisie pour la constante de lissage λ
- La valeur peut être déterminée en cherchant à minimiser les erreurs de prédiction
- La prédiction est constante après la dernière observation

2. Lissage – Lissage exponentiel simple

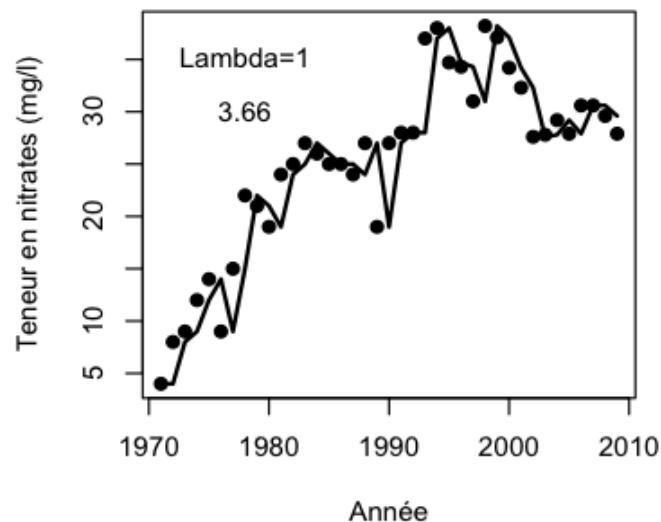
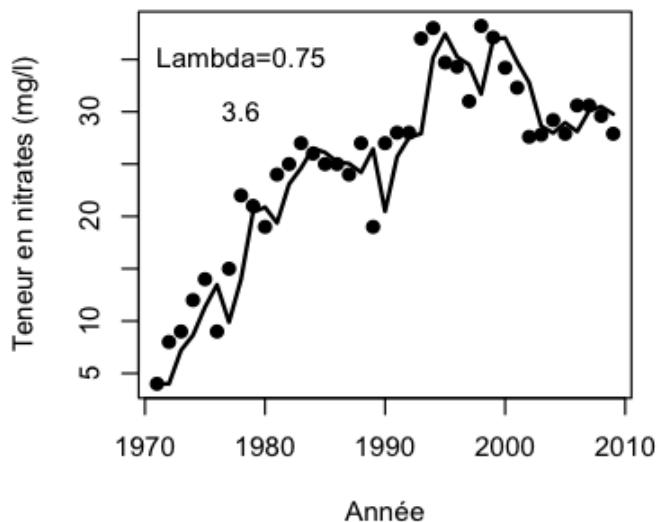
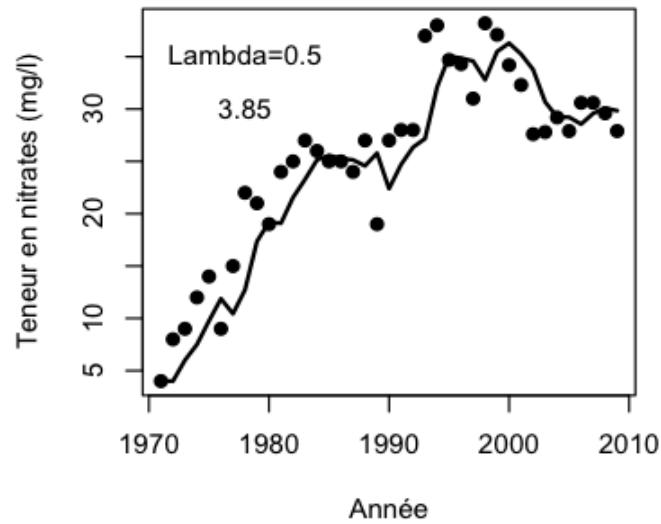
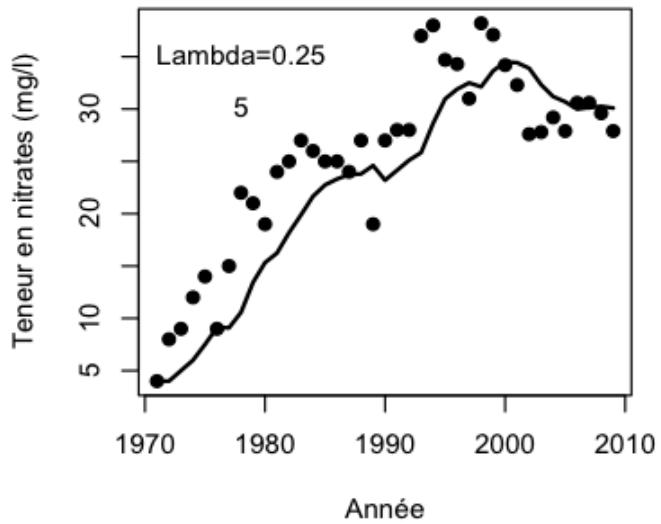
Mean Square Error

$$MSE = \frac{1}{T} \sum_{i=1}^T (y_t - \hat{Y}_t)^2$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (y_t - \hat{Y}_t)^2}$$

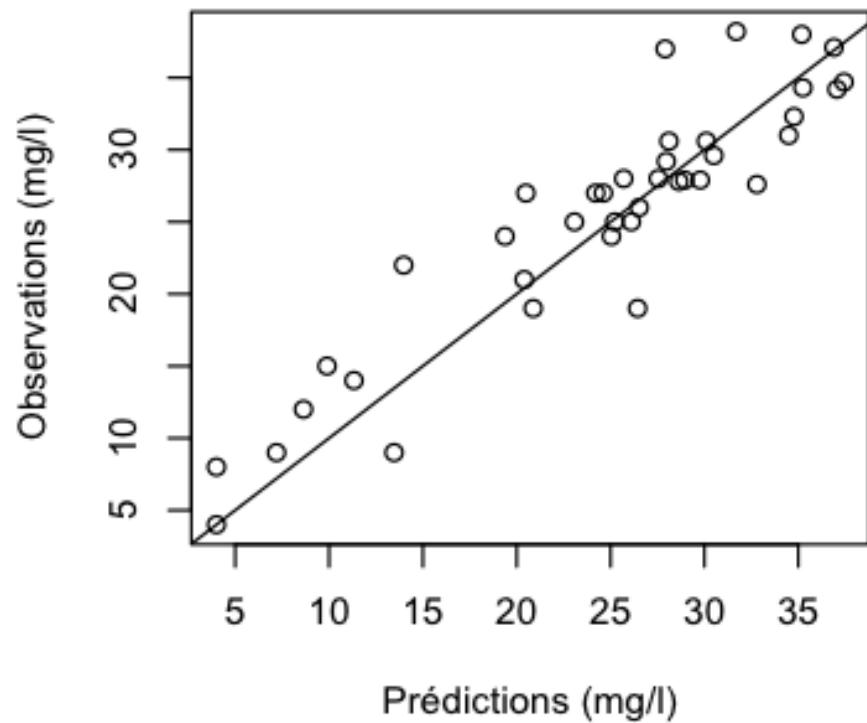
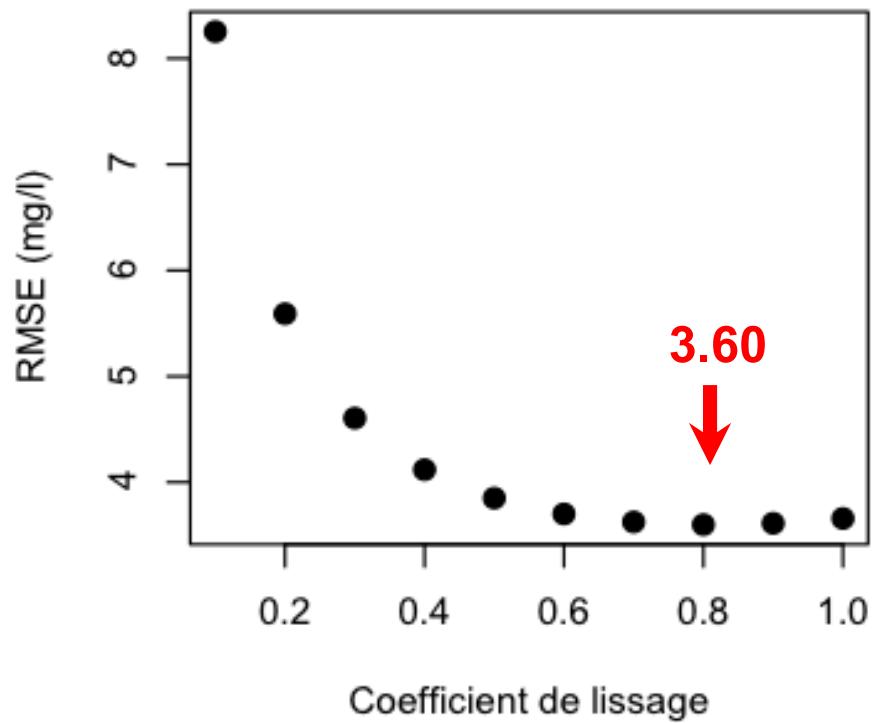
2. Lissage – Lissage exponentiel simple

Prédictions des moyennes annuelles de concentration en nitrates et RMSE



2. Lissage – Lissage exponentiel simple

RMSE en fonction du coefficient de lissage



2. Lissage – Lissage exponentiel simple

Prédiction post 2009 avec $\lambda = 0.8$: 28.3 mg/l

Très proche de la mesure 2009 (27.9 mg/l)

2. Lissage –Méthode de Holt-Winters

Méthode de Holt-Winters

- Permet de prendre en compte une tendance et/ou une saisonnalité
- Basée sur deux ou trois équations récursives
- Deux ou trois constantes de lissage à estimer

2. Lissage –Méthode de Holt-Winters

Prise en compte d'une tendance linéaire

Niveau

$$m_t = \lambda_0 y_t + (1 - \lambda_0) \hat{Y}_t$$

Pente

$$b_t = \lambda_1 (m_t - m_{t-1}) + (1 - \lambda_1) b_{t-1}$$

Initialisation

$$m_2 = y_2 \quad b_2 = y_2 - y_1$$

Prévision à la date suivante

$$\hat{Y}_{t+1} = m_t + b_t$$

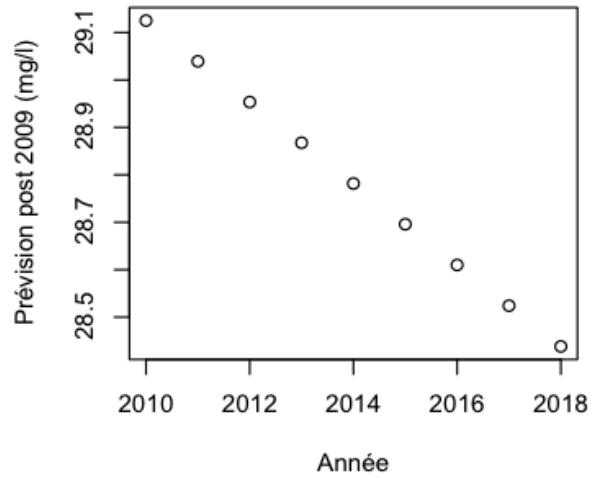
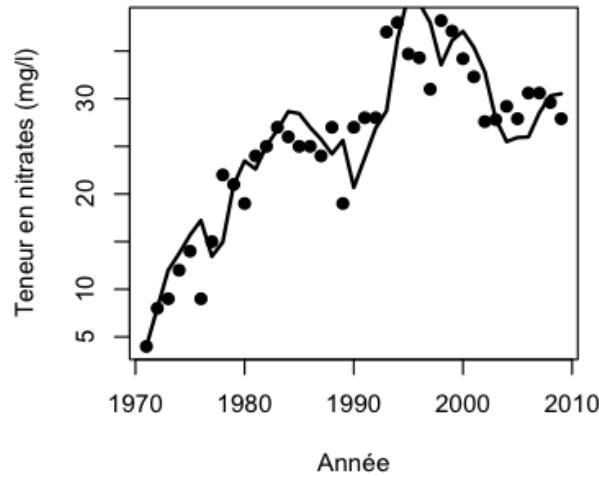
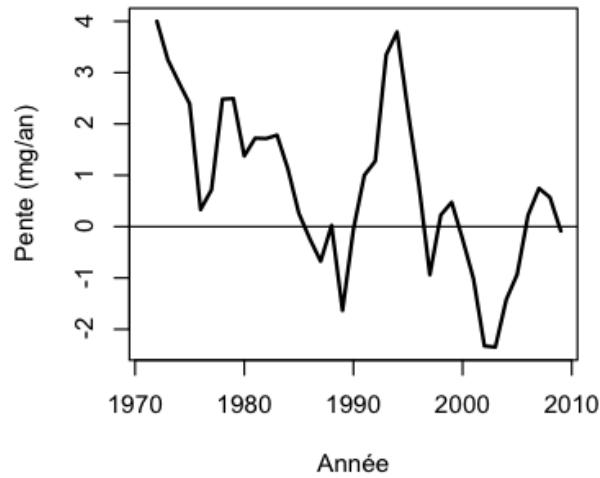
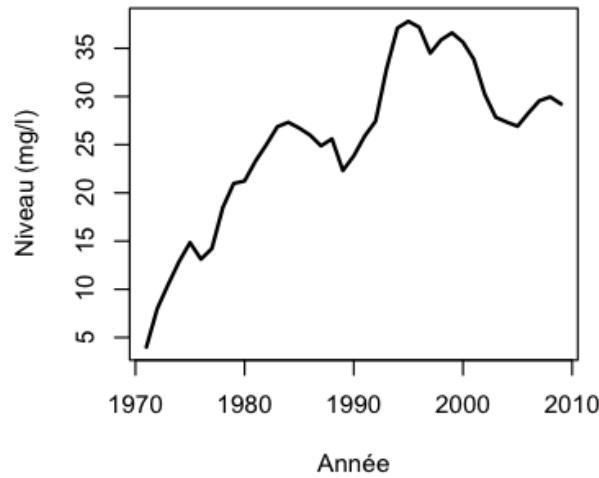
Prévision à un temps k après la dernière mesure

$$\hat{Y}_{T+k} = m_T + b_T \times k$$

2. Lissage –Méthode de Holt-Winters

Niveau, pente, prévisions des moyennes annuelles de concentration en nitrates.

$$\lambda_0 = \lambda_1 = 0.5. \text{ RMSE} = 3.93 \text{ mg/l}$$

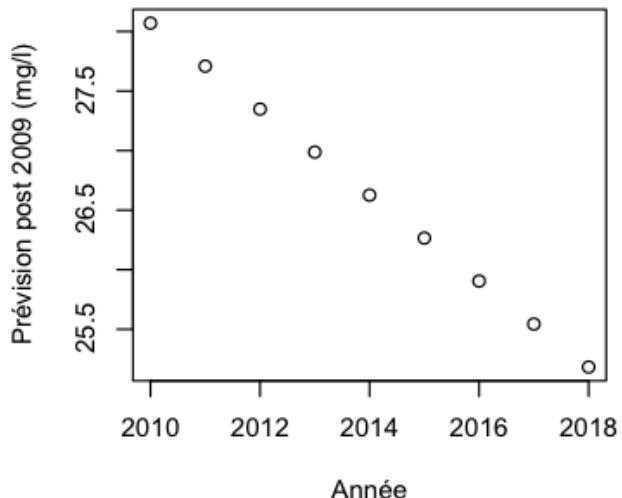
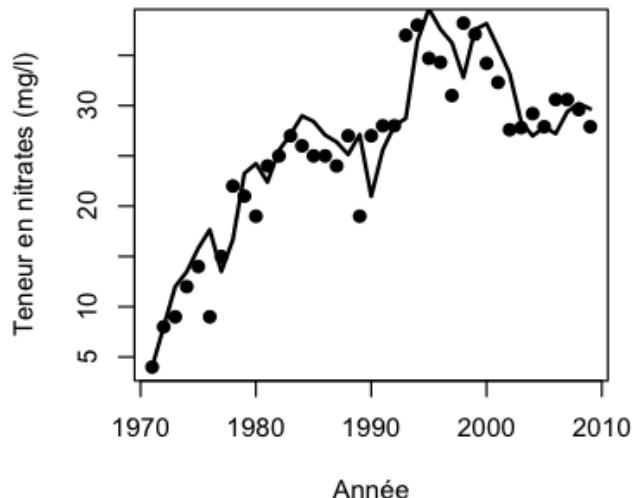
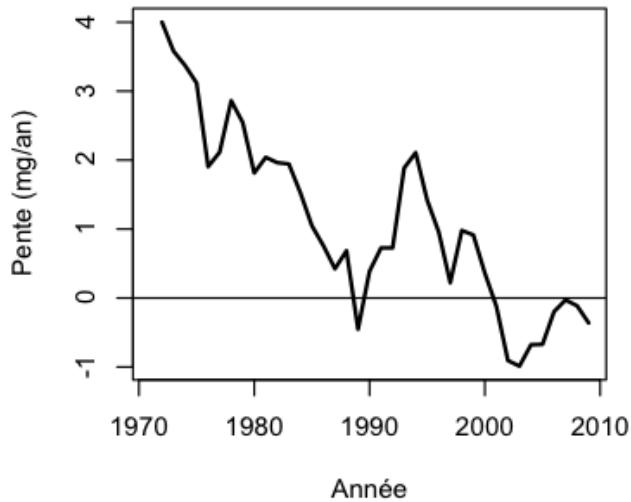
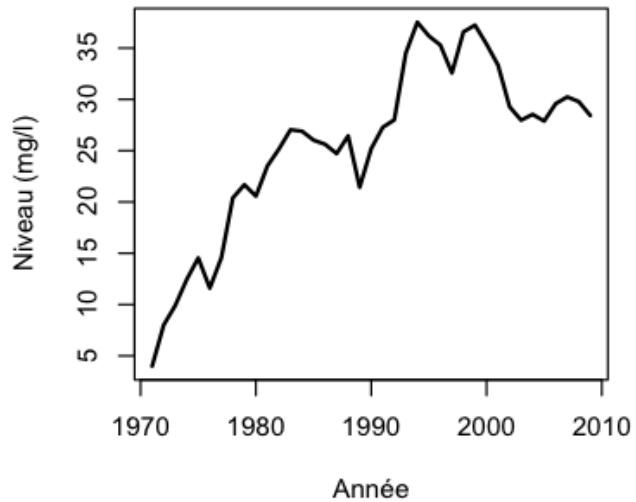


2. Lissage –Méthode de Holt-Winters

Niveau, pente, prévisions des moyennes annuelles de concentration en nitrates.

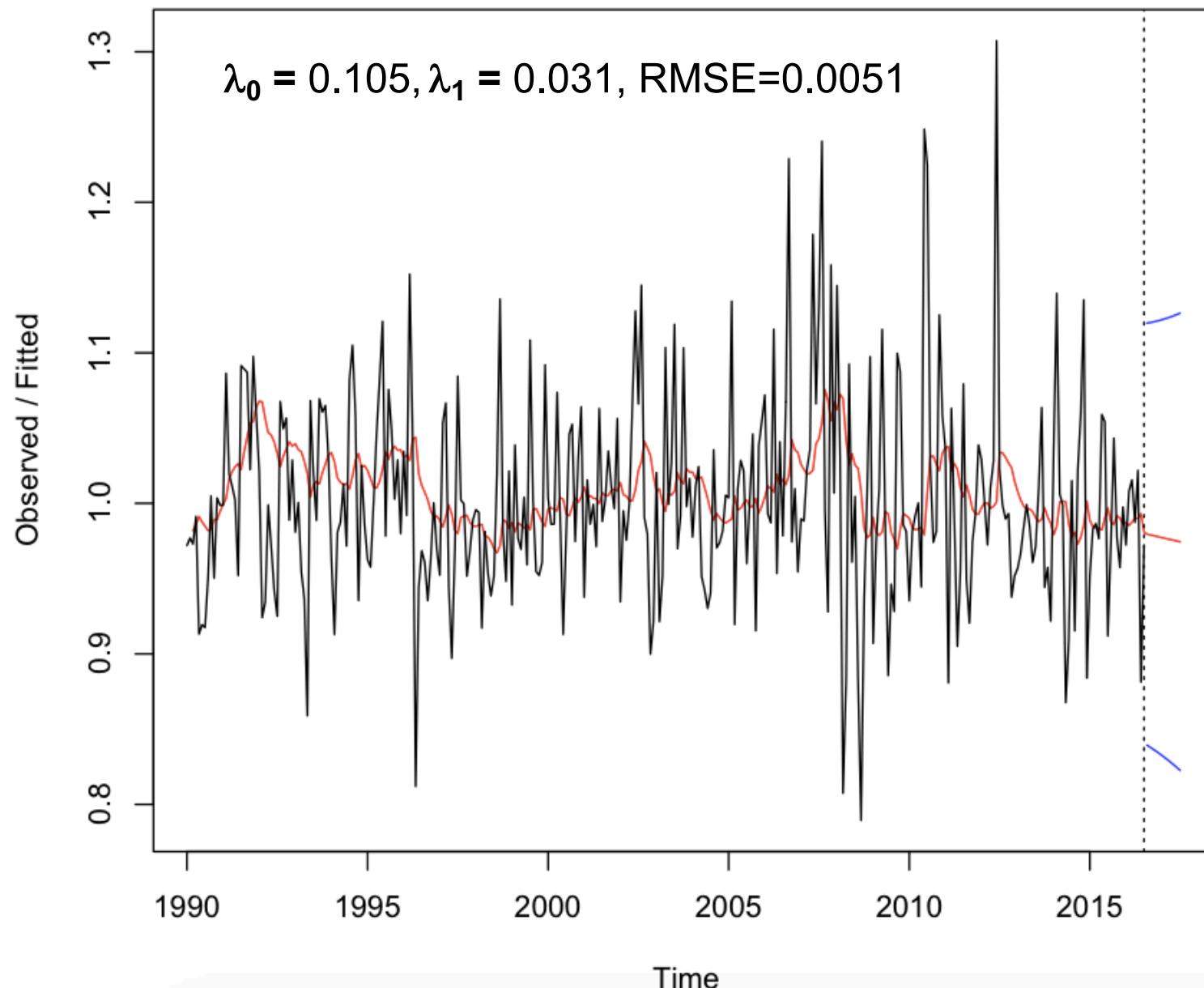
$$\lambda_0 = 0.7 \text{ et } \lambda_1 = 0.2.$$

$$\text{RMSE} = 3.73 \text{ mg/l}$$



2. Lissage –Méthode de Holt-Winters

Ratio du prix du blé (Chicago)



Exercice

Utiliser le lissage exponentiel et la méthode de Holt-Winters pour prédire le rendement du blé dans l'Eure et la Haute-Garonne jusqu'en 2025

Package R « forecast »

- Ajuste de nombreux modèles utiles pour prédire
- Sélection automatique de modèles
- Réalise des prédictions et calcule des intervalles de confiance
- Graphiques

Exponential smoothing state space model

Description

Returns ets model applied to y.

Usage

```
ets(  
  y,  
  model = "ZZZ",  
  damped = NULL,  
  alpha = NULL,  
  beta = NULL,  
  gamma = NULL,  
  phi = NULL,  
  additive.only = FALSE,  
  lambda = NULL,  
  biasadj = FALSE,  
  lower = c(rep(1e-04, 3), 0.8),  
  upper = c(rep(0.9999, 3), 0.98),  
  opt.crit = c("lik", "amse", "mse", "sigma", "mae"),  
  nmse = 3,  
  bounds = c("both", "usual", "admissible"),  
  ic = c("aicc", "aic", "bic"),  
  restrict = TRUE,  
  allow.multiplicative.trend = FALSE,  
  use.initial.values = FALSE,  
  na.action = c("na.contiguous", "na.interp", "na.fail"),  
  ...  
)
```

Arguments

y

a numeric vector or time series of class ts

model

Usually a three-character string identifying method using the framework terminology of Hyndman et al. (2002) and Hyndman et al. (2008). The first letter denotes the error type ("A", "M" or "Z"); the second letter denotes the trend type ("N", "A", "M" or "Z"); and the third letter denotes the season type ("N", "A", "M" or "Z"). In all cases, "N"=none, "A"=additive, "M"=multiplicative and "Z"=automatically selected. So, for example, "ANN" is simple exponential smoothing with additive errors, "MAM" is multiplicative Holt-Winters' method with multiplicative errors, and so on.

```
library(forecast)

#####Donnees
DATA<-read.table("Donnees_Wheat.txt", header=T, sep=" ")
DATA
summary(DATA)

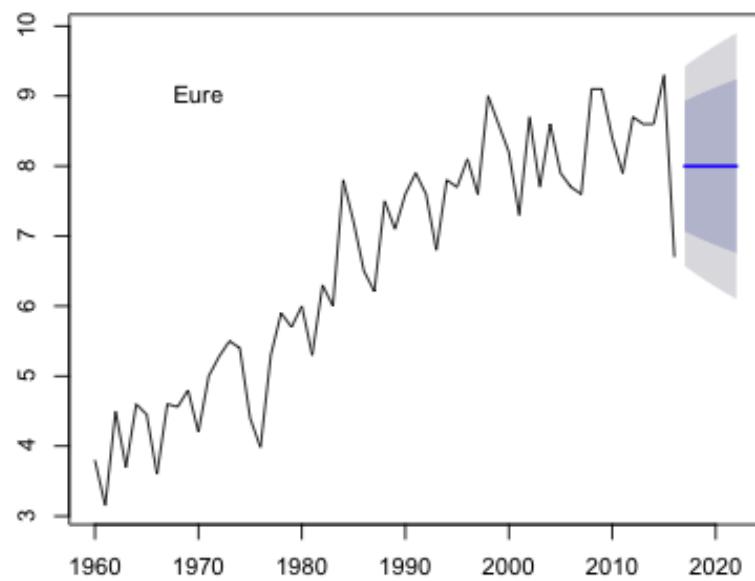
Rdt_eure.ts<-ts(DATA$Rendement_eure, start=c(1960))
plot(Rdt_eure.ts)

Rdt_hg.ts<-ts(DATA$Rendement_hautegaronne, start=c(1960))
plot(Rdt_hg.ts)

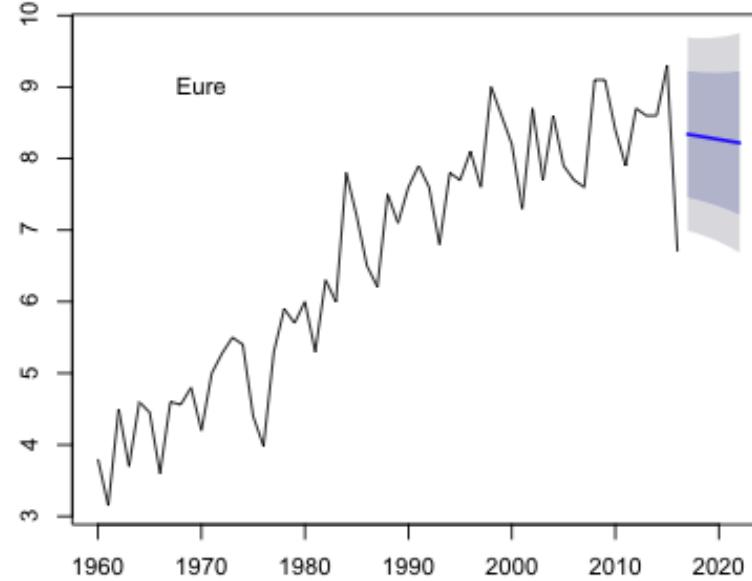
par(mfrow=c(2,2))

#####Lissage exponentiel simple
Fit_1<-ets(Rdt_eure.ts, model="ANN")
#plot(Fit_1)
summary(Fit_1)
plot(forecast(Fit_1, h=6))
text(1970, 9, "Eure")
```

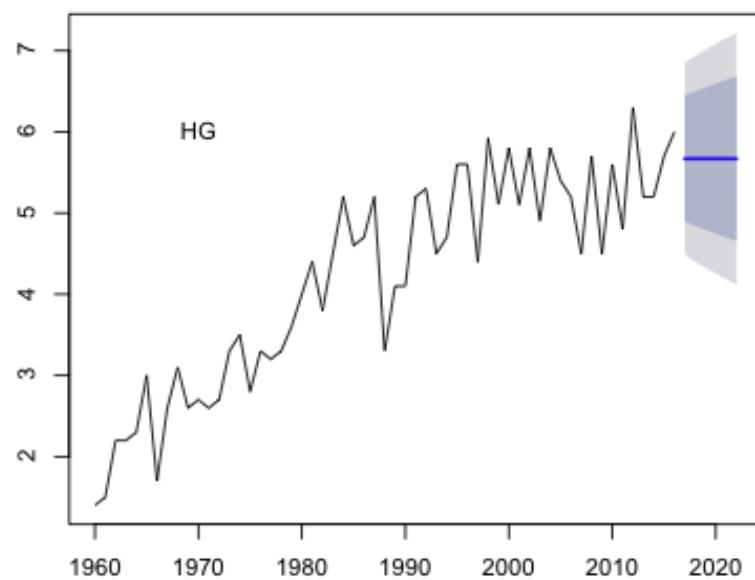
Forecasts from ETS(A,N,N)



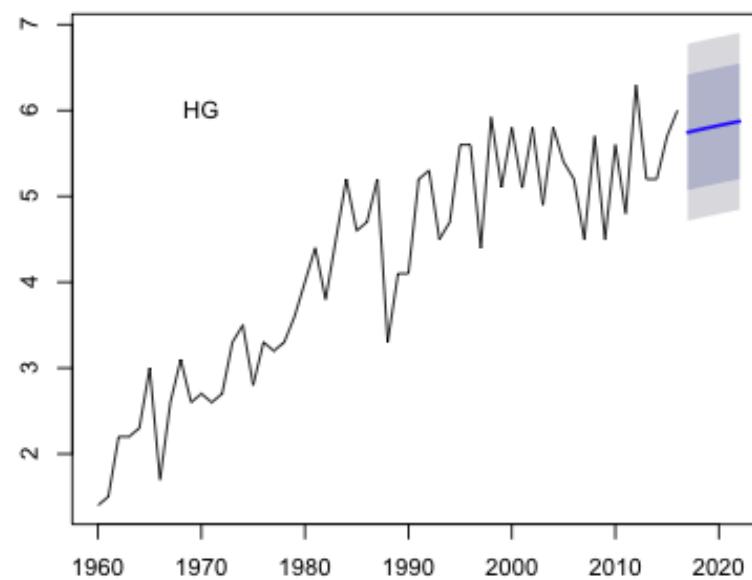
Forecasts from ETS(A,A,N)



Forecasts from ETS(A,N,N)



Forecasts from ETS(A,Ad,N)



Selection automatique du « meilleur » modèle

```
Fit_2<-ets(Rdt_hg.ts)  
summary(Fit_2)
```

1. Définition et domaines d'application

2. Lissage

3. Processus stochastiques

3. Processus stochastiques

3. Processus stochastiques

Définitions

- Décrit l'évolution dans le temps d'une variable aléatoire Y_t
- Un processus stochastique est un mécanisme qui permet de générer des observations y_t sur une période $t = 1, \dots, T$
- Peut être utilisé pour générer une infinité de jeux de données. Chacun correspond à une réalisation du processus
- Les processus stationnaires constituent un cas particulier important

3. Processus stochastiques

Processus stationnaire

Stationnarité faible

Certaines propriétés du processus ne changent pas dans le temps

- Moyenne constante (pas de tendance)
- Variance constante (pas d'hétéroscedasticité)
- Corrélations fonction de l'écart de temps

3. Processus stochastiques

Processus stationnaire

Stationnarité faible

$$E(Y_t) = \mu$$

$$\text{var}(Y_t) = E[(Y_t - \mu)^2] = \gamma(0)$$

$$E[(Y_t - \mu)(Y_{t-\tau} - \mu)] = \gamma(\tau)$$

Les propriétés du processus ne changent pas dans le temps.

3. Processus stochastiques

Intérêt de la stationnarité

On peut estimer l'espérance, la variance et les autocovariances à partir d'une seule réalisation.

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t$$

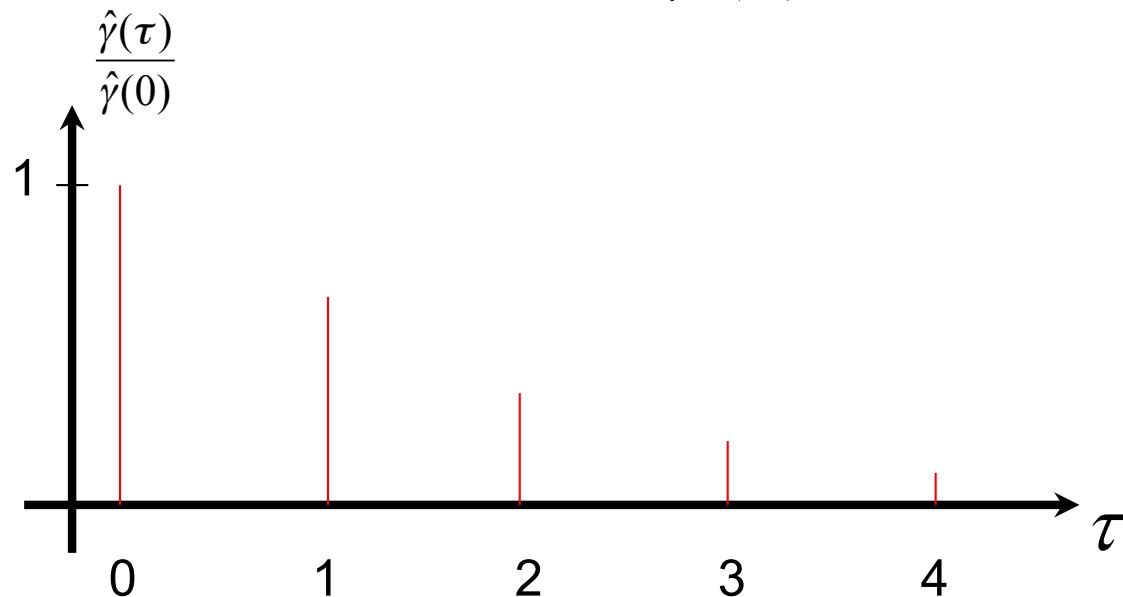
$$\hat{\gamma}(0) = \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2$$

$$\hat{\gamma}(\tau) = \frac{1}{T} \sum_{t=\tau+1}^T (y_t - \bar{y})(y_{t-\tau} - \bar{y})$$

3. Processus stochastiques

Corrélogramme

Graphique des autocorrélations $\frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}$ en fonction de τ



- Ce graphique est très utile pour identifier la nature du processus.
- Les **autocorrélations partielles** sont également utiles.

3. Processus stochastiques

Corrélations partielles

$$E\left[(Y_t - Y_t^{(t-1)})(Y_{t-\tau} - Y_{t-\tau}^{(t-1)})\right]$$

$$Y_t^{(t-1)} = \beta_1 Y_1 + \dots + \beta_{t-1} Y_{t-1}$$

3. Processus stochastiques

Les différents types de processus stationnaires

- Les modèles « autorégressif » (AR)
- Les modèles « moyenne mobile » (MA)
- Les modèles « mixtes » (ARMA)

3. Processus stochastiques

Les processus non stationnaires

- L'espérance et/ou la variance évoluent dans le temps
- On va chercher à éliminer la tendance et à stabiliser la variance

3. Processus stochastiques

Eliminer la tendance

- Différenciation → Modèle ARIMA
- Ajustement d'une tendance déterministe → Régression

3. Processus stochastiques

Stabiliser la variance

Transformation des données, par exemple en prenant le logarithme des mesures ou une puissance (transformation de Box-Cox).

$$B(x, \lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \log(x) & \text{si } \lambda = 0 \end{cases}$$

3. Processus stochastiques

Démarche de Box-Jenkins (1979)

- Identifier un modèle
- Estimer les paramètres
- Evaluer
- Prévoir



3. Processus stochastiques

Georges Box (né en 1919 en Angleterre)



George Edward Pelham Box (18 October 1919 –) is one of the most influential statisticians of the 20th century and a pioneer in the areas of quality control, time series analysis, design of experiments and Bayesian inference.

He was born in Gravesend, Kent, England and originally trained as a chemist. During World War II, he worked on biochemical experiments on the effect of poison gases on small animals for the British Army. He needed statistical advice to analyze the results of his experiments but could not find a statistician who could give him guidance, so he taught himself statistics from available texts. After the war, he enrolled at University College London and obtained a bachelor's degree in mathematics and statistics. He received a Ph.D. from the University of London in 1953.

From 1948 to 1956, Box worked as a statistician for Imperial Chemical Industries (ICI). While at ICI, he took a leave of absence for a year and served as a visiting professor at the University of North Carolina at Chapel Hill. He later went to Princeton University where he served as Director of the Statistical Research Group.

In 1960, Box moved to the University of Wisconsin-Madison to create the Department of Statistics. He was appointed Vilas Research Professor of Statistics (the highest honor accorded to any faculty member at the University of Wisconsin-Madison) in 1980. George Box and Bill Hunter co-founded the Center for Quality and Productivity Improvement at the University of Wisconsin-Madison in 1984. Box officially retired in 1992, becoming an Emeritus Professor.

Throughout his career, George Box has written numerous research papers and published many books. His most important books include *Statistics for Experimenters* (1978), *Time Series Analysis: Forecasting and Control* (1979, with Gwilym Jenkins) and *Bayesian Inference in Statistical Analysis*. (1973, with George C. Tiao). Today, his name is associated with important results in statistics such as Box-Jenkins models, Box-Cox transformations, Box-Behnken designs and numerous others.

He served as President of the American Statistical Association in 1978 and of the Institute of Mathematical Statistics in 1979. He received the Shewhart Medal from the American Society for Quality Control in 1968, the Wilks Memorial Award from the American Statistical Association in 1972, the R. A. Fisher Lectureship in 1974, and the Guy Medal in Gold from the Royal Statistical Society in 1993. He was elected a member of the American Academy of Arts and Sciences in 1974 and a Fellow of the Royal Society in 1979.

The often quoted phrase, "Essentially, all models are wrong, but some are useful", is attributed to George Box.

Box married Joan Fisher, second of Ronald Fisher's five daughters. In 1978, Joan Fisher Box published a very well-received biography of her father. He supervised Lars Pallesen, who was appointed rector (president) of the Technical University of Denmark in 2007.

3. Processus stochastiques

Démarche de Box-Jenkins (1979)

- i. Identifier un modèle
- ii. Estimer les paramètres
- iii. Evaluer
- iv. Prévoir



3. Processus stochastiques

i. Identifier un modèle

- Déterminer si la stationnarité est plausible (analyse graphique, test)
- Si ce n'est pas le cas, différencier les données ou ajouter une tendance.
- Déterminer la nature du processus (AR, MA, ARMA) (analyse des corrélogrammes)
- Etape délicate, plus ou moins subjective

3. Processus stochastiques

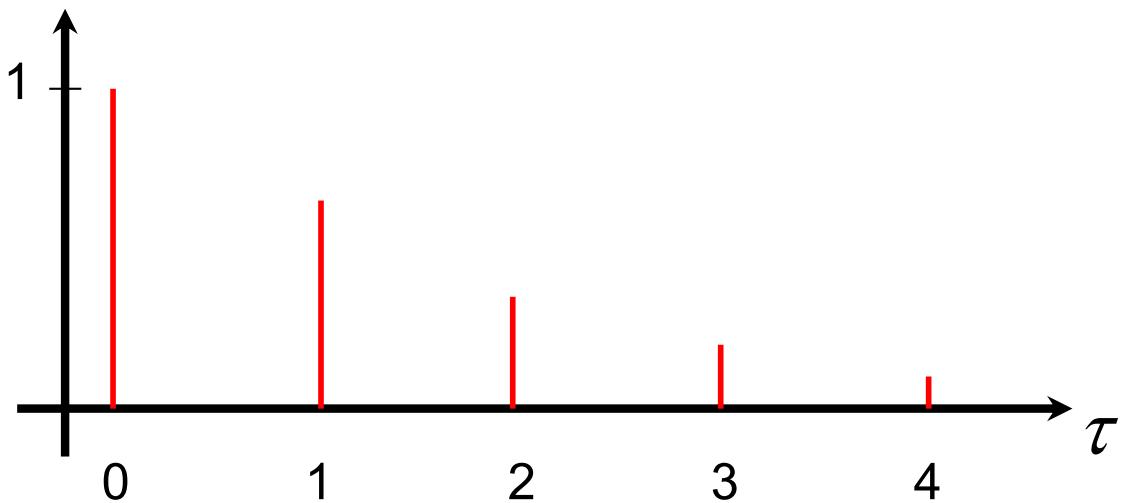
Modèle AR(1) (autorégressif d'ordre 1)

$$Y_t = \varphi Y_{t-1} + \xi_t$$

$$E(\xi_t) = 0$$

$$\text{var}(\xi_t) = \sigma^2$$

$$\varphi > 0$$



3. Processus stochastiques

Modèle MA(1) (moyenne mobile d'ordre 1)

$$Y_t = \xi_t + \theta \xi_{t-1}$$

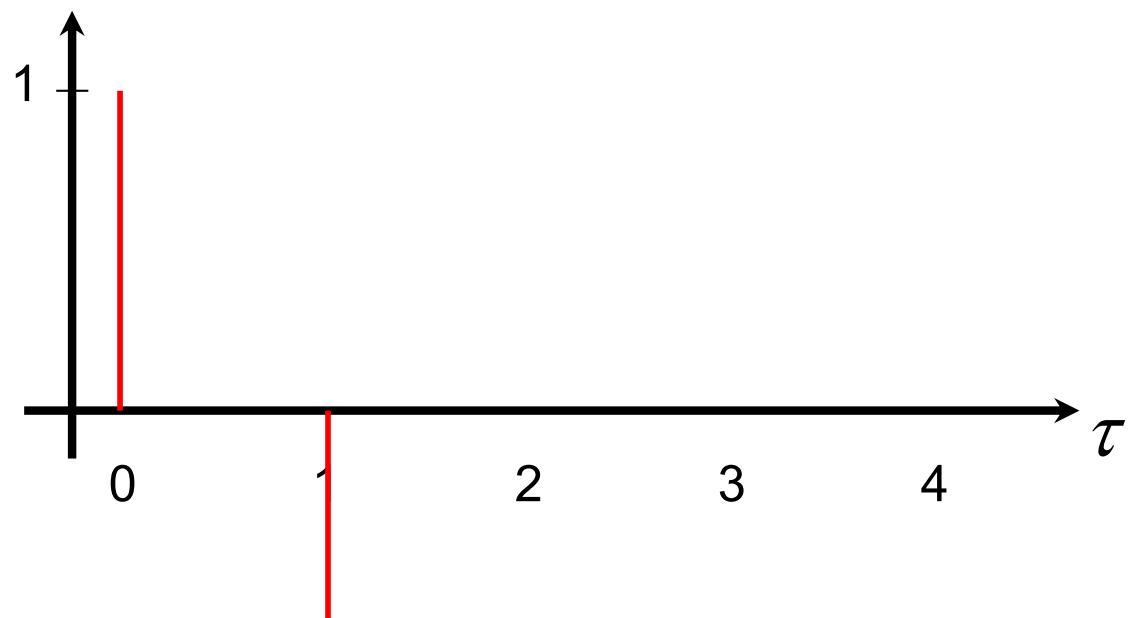
$$E(\xi_t) = 0$$

$$\text{var}(\xi_t) = \sigma^2$$

$$\theta > 0$$



$$\theta < 0$$



3. Processus stochastiques

Modèle ARMA(p, q)

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \xi_t + \theta_1 \xi_{t-1} + \dots + \theta_q \xi_{t-q}$$

Des conditions sur $\varphi_1, \dots, \varphi_p$ sont requises pour assurer la stationnarité.

Choix de p et q par analyse des corrélogrammes

ou...

Choix automatique par optimisation

3. Processus stochastiques

ii. Estimer les paramètres

Méthode du maximum de vraisemblance

$$L = \text{prob}\left(y_1, \dots, y_T \mid \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q\right)$$

3. Processus stochastiques

iii. Evaluer le modèle

- **Analyse des résidus**
- **Critères de comparaison de modèles**

3. Processus stochastiques

iii. Evaluer le modèle

- **Analyse des résidus**
 - **Auto-corrélogramme des résidus**
 - **Test d'indépendance des résidus (ex: Ljung-Box)**

3. Processus stochastiques

iii. Evaluer le modèle

- **Critères de comparaison de modèles**
 - **Critère d'Akaïke**
 - **BIC**

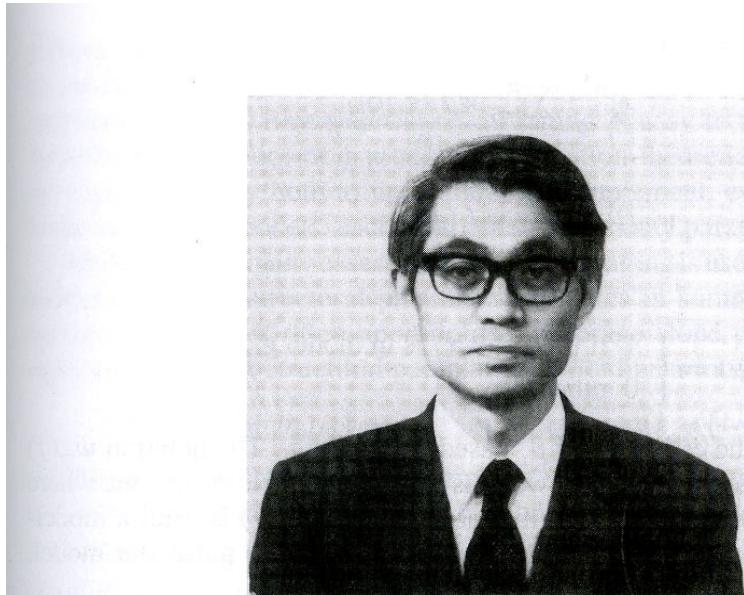
3. Processus stochastiques

Le critère d'Akaïke

$$AIC = -2 \times \log(L) + 2 \times (p + q)$$

**Estimation de la distance entre le vrai processus ayant
généré les données et le modèle.**

Hirotugu Akaike (né en 1927 au Japon)



Hirotugu Akaike was born in 1927 in Fujinomiya-shi, Shizuoka-jen, in Japan. He received B.S. and D.S. degrees in mathematics from the University of Tokyo in 1952 and 1961, respectively. He worked at the Institute of Statistical Mathematics for over 30 years, becoming its Director General in 1982. He has received many awards, prizes, and honors for his work in theoretical and applied statistics (deLeeuw 1992, Parzen 1994). The three-volume set, *"Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach"* (Bozdogan 1994) commemorated Professor Hirotugu Akaike's 65th birthday. Bozdogan (1994) records that the idea of a connection between the Kullback–Leibler discrepancy and the empirical log-likelihood function occurred to Akaike on the morning of March 16, 1971, as he was taking a seat on a commuter train.

Burnham and Anderson, 2002

3. Processus stochastiques

iv. Prévoir

Espérance de Y conditionnelle aux dernières observations

$$\hat{Y}_{T+k} = \hat{E}\left(Y_{T+k} \mid y_T, y_{T-1}, \dots, y_{T-p}\right)$$

Exemple avec un AR(1)

$$\hat{Y}_{T+1} = \hat{\varphi} y_T$$

$$\hat{Y}_{T+k} = \hat{\varphi}^k y_T$$

3. Processus stochastiques

Des variantes utiles

3. Processus stochastiques

Une autre façon d'éliminer une tendance: la différenciation et le modèle ARIMA

Travailler sur des différences plutôt que sur les valeurs brutes

$$\nabla Y_t = Y_t - Y_{t-1} \longrightarrow \text{ARIMA}(p, 1, q)$$

Cette démarche se généralise.

$$\text{Différenciation d'ordre } d \longrightarrow \text{ARIMA}(p, d, q)$$

3. Processus stochastiques

Prise en compte d'un effet saisonnier

- Ajout d'une composante saisonnière
- Différenciation

3. Processus stochastiques

Ajout d'une composante saisonnière

$$Y_t = a + b.t + c.t^2 \quad (\text{tendance})$$
$$+ d_{\text{mois}(t)} \quad (\text{composante saisonnière})$$
$$+ e_t \quad (\text{aléa})$$

Variante possible

Exemple: composante saisonnière sinusoïdale

$$d_{\text{mois}(t)} = d \cdot \sin(t/12) + e \cdot \cos(t/12)$$

3. Processus stochastiques

Elimination des effets saisonniers par différenciation : le modèle SARIMA

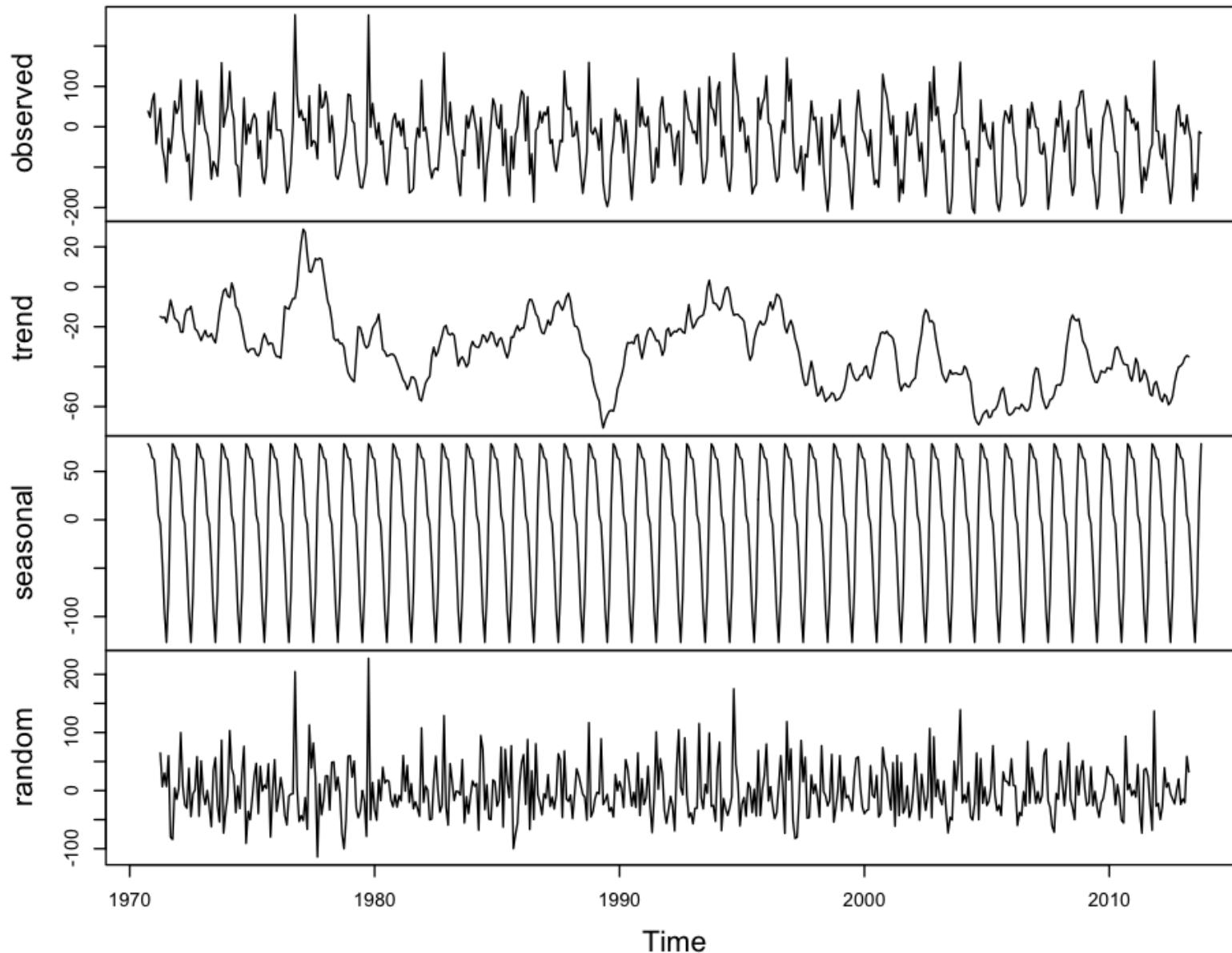
Travailler sur des différences d'ordre s plutôt que sur les valeurs brutes

$$\nabla_s Y_t = Y_t - Y_{t-s}$$

$$\nabla_{12} Y_t = Y_t - Y_{t-12}$$

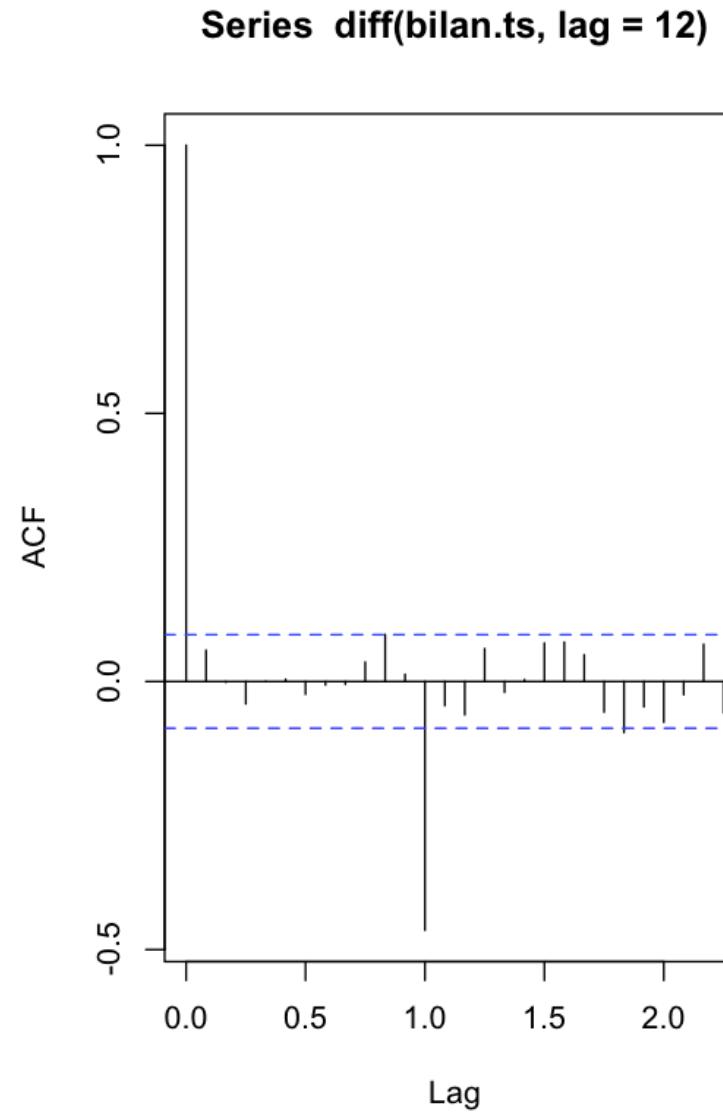
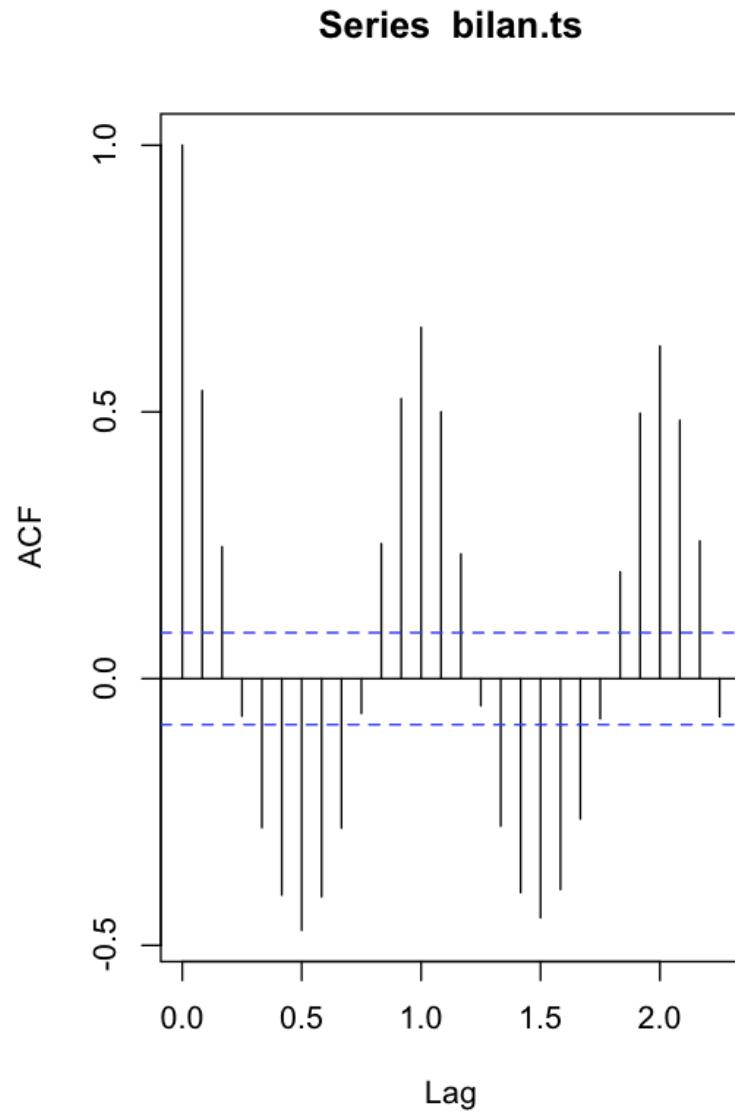
3. Processus stochastiques

Exemple : données de bilan climatique du Vaucluse



3. Processus stochastiques

Exemple : données de bilan climatique du Vaucluse



3. Processus stochastiques

Modèle MA(1) (moyenne mobile d'ordre 1)

$$\nabla_{12} Y_t = Y_t - Y_{t-12} = \xi_t + \theta \xi_{t-12}$$

$$E(\xi_t) = 0$$

$$\text{var}(\xi_t) = \sigma^2$$

3. Processus stochastiques

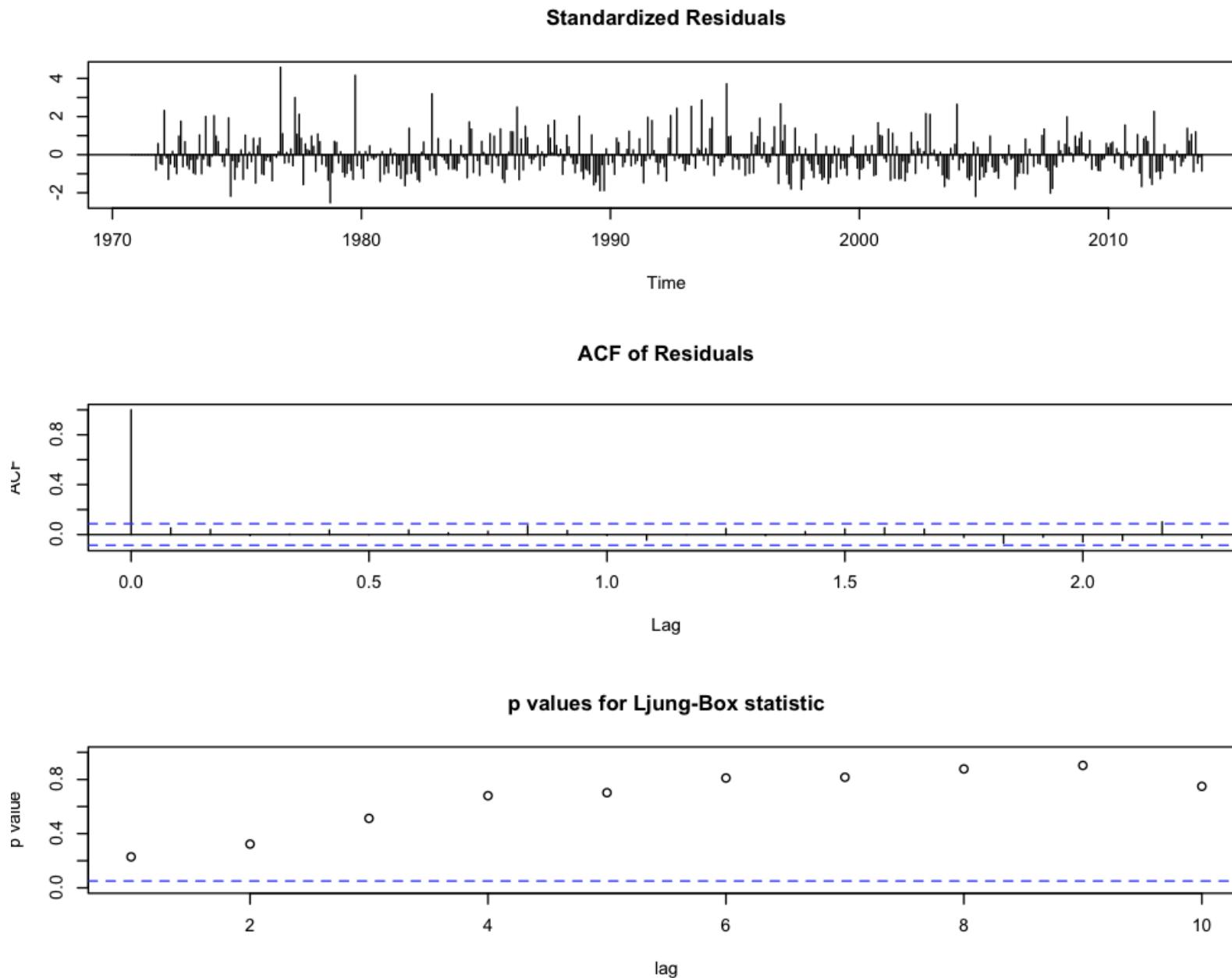
```
> Mod.1<-arima(bilan.ts, order=c(0,0,0), seasonal=list(order = c(0,1,1), period = 12))
> Mod.1

Call:
arima(x = bilan.ts, order = c(0, 0, 0), seasonal = list(order = c(0, 1, 1),
    period = 12))

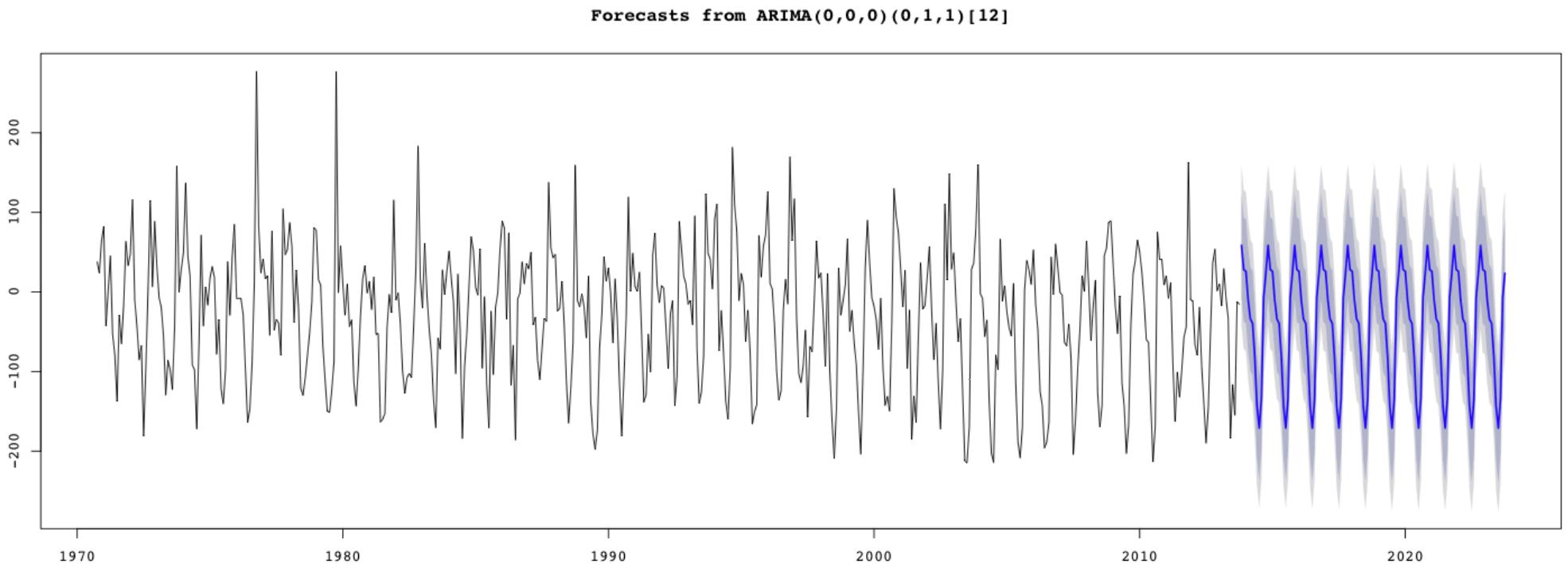
Coefficients:
      smal
      -0.8984
s.e.  0.0232

sigma^2 estimated as 2593:  log likelihood = -2711.23,  aic = 5426.47
.
```

3. Processus stochastiques



3. Processus stochastiques



3. Processus stochastiques

Méthode automatique pour sélectionner p, q, d
avec le critère AIC (ou d'autres critères du même type)

```
library(forecast)  
auto.arima(Y)
```

Exercice

Prédire le rendement jusqu'en 2025 avec un modèle ARIMA

```
library(forecast)

#####Donnees
DATA<-read.table("Donnees_Wheat.txt", header=T, sep=" ")
DATA
summary(DATA)

Rdt_eure.ts<-ts(DATA$Rendement_eure, start=c(1960))
plot(Rdt_eure.ts)

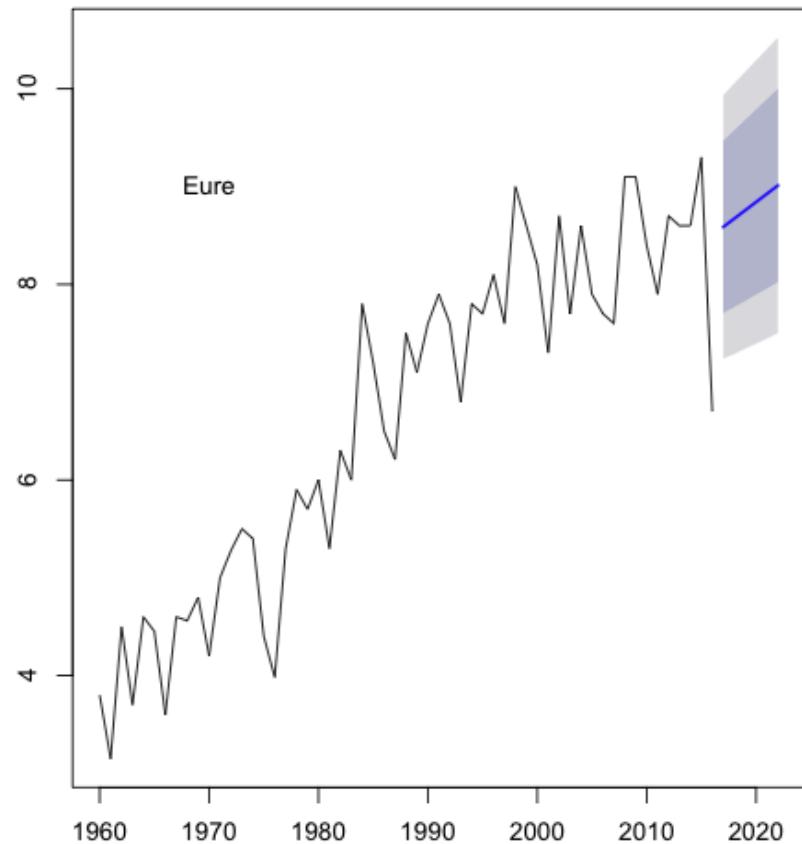
Rdt_hg.ts<-ts(DATA$Rendement_hautegaronne, start=c(1960))
plot(Rdt_hg.ts)

par(mfrow=c(1,2))

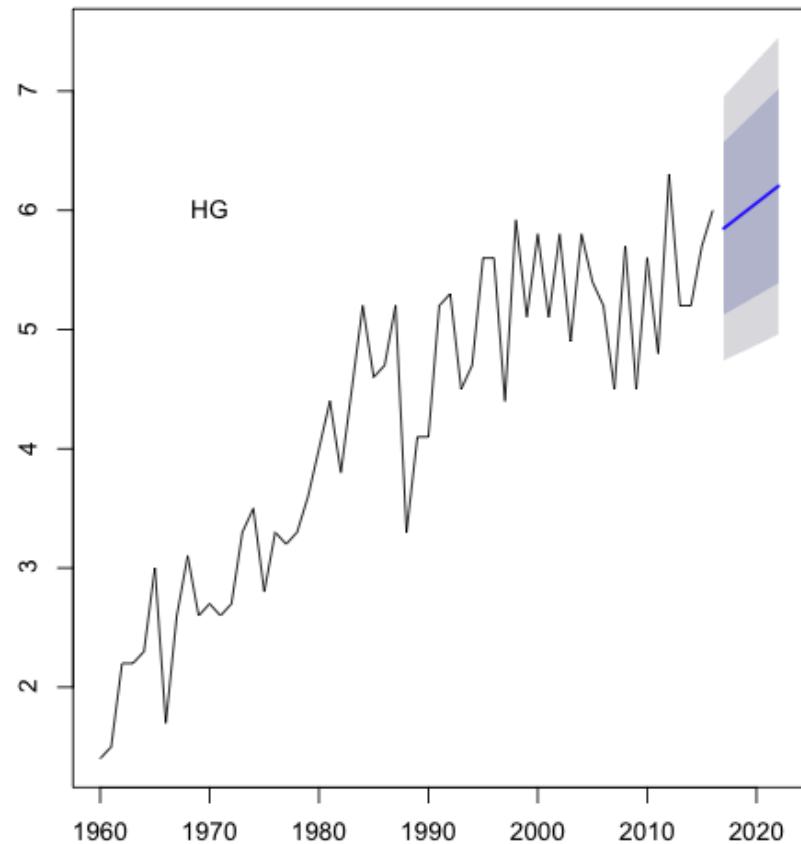
#####ARIMA
Fit_1<-auto.arima(Rdt_eure.ts,D=0)
summary(Fit_1)
plot(forecast(Fit_1,h=6))
text(1970,9,"Eure")
forecast(Fit_1,h=6)

#####ARIMA
Fit_2<-auto.arima(Rdt_hg.ts,D=0)
summary(Fit_2)
plot(forecast(Fit_2,h=6))
text(1970,6,"HG")
forecast(Fit_2,h=6)
```

Forecasts from ARIMA(0,1,1) with drift



Forecasts from ARIMA(0,1,1) with drift



3. Processus stochastiques

Combiner
tendance + lissage + saisonalité +processus stochastique

La méthode TBATS

Trigonometric seasonality, **B**ox-Cox transformation,
ARMA errors, **T**rend and **S**easonal components.

Model:

$$y_t^{(\lambda)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = \phi b_{t-1} + \beta d_t$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t$$

Where:

$y_t^{(\lambda)}$ - time series at moment t (Box-Cox transformed)

$s_t^{(i)}$ - i th seasonal component

l_t - local level

b_t - trend with damping

d_t - ARMA(p, q) process for residuals

e_t - Gaussian white noise

Model parameters:

T - Amount of seasonalities

m_i - Length of i th seasonal period

k_i - Amount of harmonics for i th seasonal period

λ - Box-Cox transformation

α, β - Smoothing

ϕ - Trend damping

φ_i, θ_i - ARMA(p, q) coefficients

$\gamma_1^{(i)}, \gamma_2^{(i)}$ - Seasonal smoothing (two for each period)

Model:

$$y_t^{(\lambda)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = \phi b_{t-1} + \beta d_t$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t$$

Where:

$y_t^{(\lambda)}$ - time series at moment t (Box-Cox transformed)

$s_t^{(i)}$ - i th seasonal component

l_t - local level

b_t - trend with damping

d_t - ARMA(p, q) process for residuals

e_t - Gaussian white noise

Seasonal part:

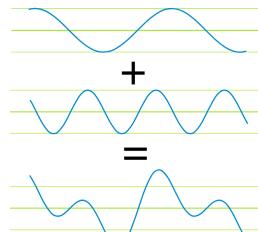
$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos(\omega_i) + s_{j,t-1}^{*(i)} \sin(\omega_i) + \gamma_1^{(i)} d_t$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \sin(\omega_i) + s_{j,t-1}^{*(i)} \cos(\omega_i) + \gamma_2^{(i)} d_t$$

$$\omega_i = 2\pi j/m_i$$

Fourier series



Model parameters:

T - Amount of seasonalities

m_i - Length of i th seasonal period

k_i - Amount of harmonics for i th seasonal period

λ - Box-Cox transformation

α, β - Smoothing

ϕ - Trend damping

φ_i, θ_i - ARMA(p, q) coefficients

$\gamma_1^{(i)}, \gamma_2^{(i)}$ - Seasonal smoothing (two for each period)

3. Processus stochastiques

Plusieurs modèles TBATS sont comparés automatiquement selon le critère AIC

- Avec et sans transformation Box Cox
- avec et sans tendance
- avec et sans amortissement de la tendance
- avec et sans ARMA(p, q) sur les résidus
- avec et sans saisonnalité
- avec différentes séries de Fourier pour la saisonnalité.

Le choix de p et q dans le processus ARMA est réalisé avec `auto.arima()`.

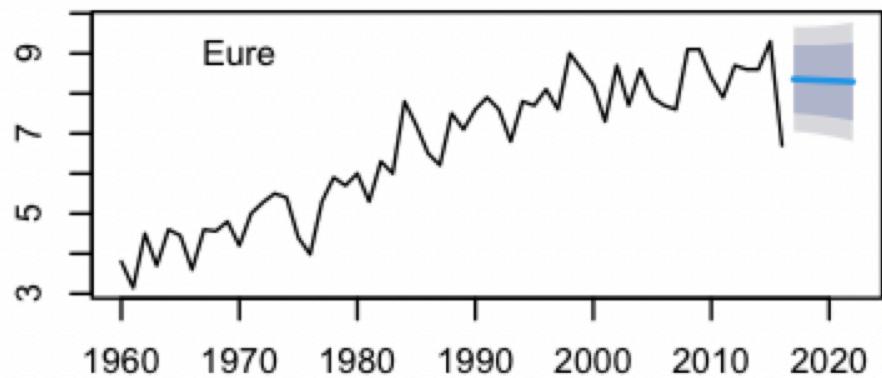
Exercice

Prédire le rendement jusqu'en 2025 avec un modèle TBATS

```
Fit_1<-tbats(Rdt_eure.ts)
summary(Fit_1)
plot(forecast(Fit_1,h=6))
text(1970,9,"Eure")
forecast(Fit_1,h=6)
```

```
Fit_2<-tbats(Rdt_hg.ts)
summary(Fit_2)
plot(forecast(Fit_2,h=6))
text(1970,6,"HG")
forecast(Fit_2,h=6)
```

Forecasts from BATS(1, {0,0}, 1, -)

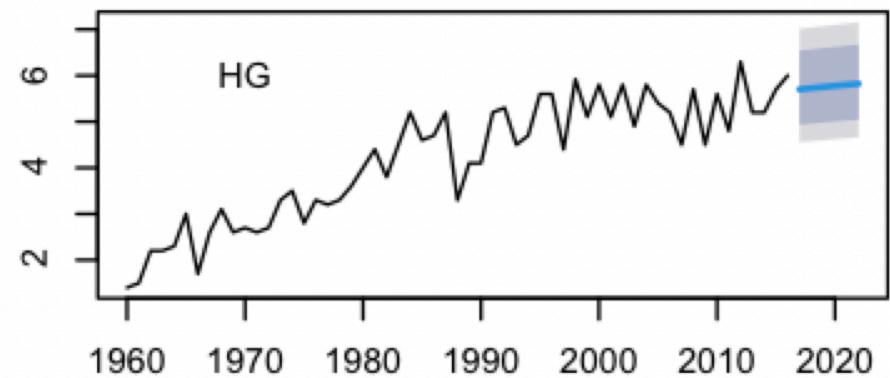


Transformation
Box Cox

p, q

damping

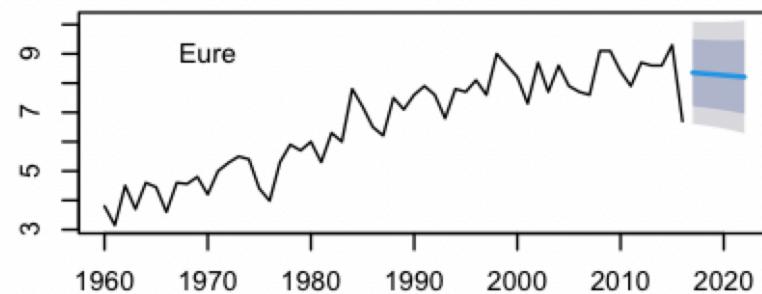
Forecasts from BATS(0.375, {0,0}, 0.964, -)



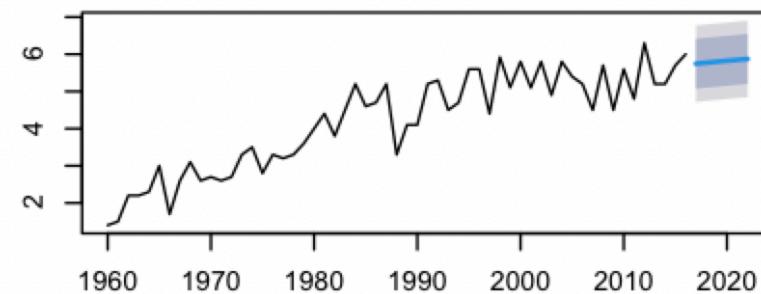
Question

Comment choisir la meilleure méthode pour prédire une quantité d'intérêt à partir d'une série chronologique ?

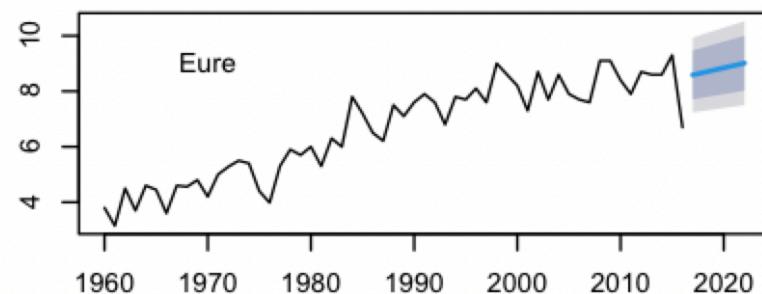
Forecasts from ETS(M,A,N)



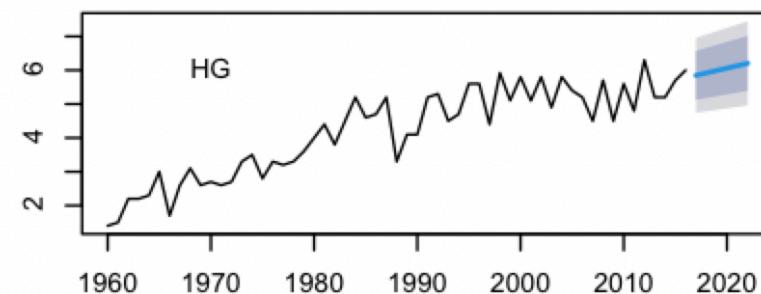
Forecasts from ETS(A,Ad,N)



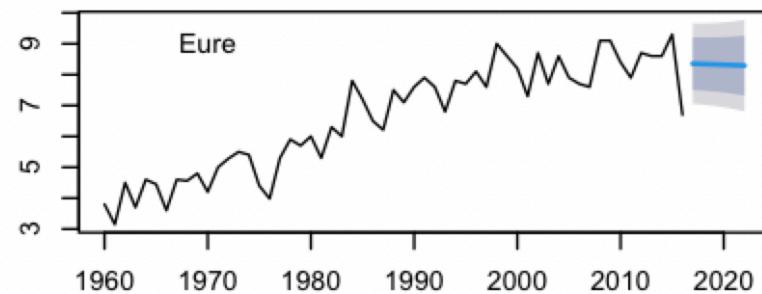
Forecasts from ARIMA(0,1,1) with drift



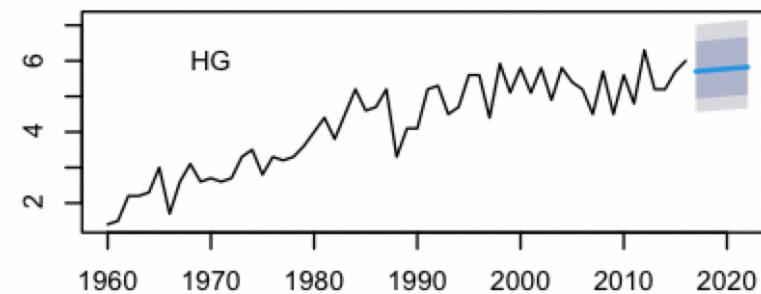
Forecasts from ARIMA(0,1,1) with drift



Forecasts from BATS(1, {0,0}, 1, -)



Forecasts from BATS(0.375, {0,0}, 0.964, -)



Quelques références

Mots clés en anglais: time series, forecasting

Brockwell P.J., David R.A. 2002. Introduction to time series and forecasting. Springer.

Chatfield C. 2004. The analysis of time series. An introduction. (6th edition). Chapman & Hall/CRC.

De Livera, A.M., Hyndman, R.J., & Snyder, R. D. (2011), Forecasting time series with complex seasonal patterns using exponential smoothing, *Journal of the American Statistical Association*, 106(496), 1513-1527.

Harvey A.C. 1991. Forecasting, structural time series models and the Kalman filter. Cambridge University Press.

Makowski D, Monod H. 2011. Analyse statistique des risques agro-environnementaux. Springer.

Michel L., Makowski D. 2013. Comparison of statistical models for analyzing wheat yield time series. *Plos one* 8(10).

Petris G. 2010. A R package for dynamic linear models. *Journal of Statistical Software* 36, 1-14.

Venables W.N., Ripley B.D. 2002. Modern applied statistics with S. Springer. Chapitre 14.