

Séries temporelles

2022-04-14

Contents

Introduction, découverte des données et premières idées	1
Choix de la ville	2
Première approche	2
Décomposition de la série	6
Elimination de la saisonnalité	7
Methode avec opérateur différence	7
Fonction <i>decompose</i> : moyennes mobiles	11
Test sur les résidus	12
Modélisation de la série	14
Modèle retenus	14

```
library(zoo)
library(xts)
library(ggplot2)
library(dplyr)
library(lubridate)
library(forecast)
library(gridExtra)
```

Introduction, découverte des données et premières idées

```
data <- read.csv("hungary_chickenpox.csv")
data$date <- dmy(data$date)
mean(colMeans(data[-1])) # Moyenne générale de cas de varicelle de 38.84282

## [1] 38.84282

colMeans(data[-1]) # Budapest bien supérieure à la moyenne, top 1 (logique car capitale)
```

```

##  BUDAPEST   BARANYA      BACS      BEKES      BORSOD    CSONGRAD     FEJER      GYOR
## 101.24521  34.20498  37.16667  28.91188  57.08238  31.48851  33.27203  41.43678
## HAJDU       HEVES       JASZ      KOMAROM    NOGRAD      PEST      SOMOGY    SZABOLCS
## 47.09770  29.69157  40.86973  25.64368  21.85057  86.10153  27.60920  29.85441
## TOLNA       VAS        VESZPREM    ZALA
## 20.35249  22.46743  40.63602  19.87356

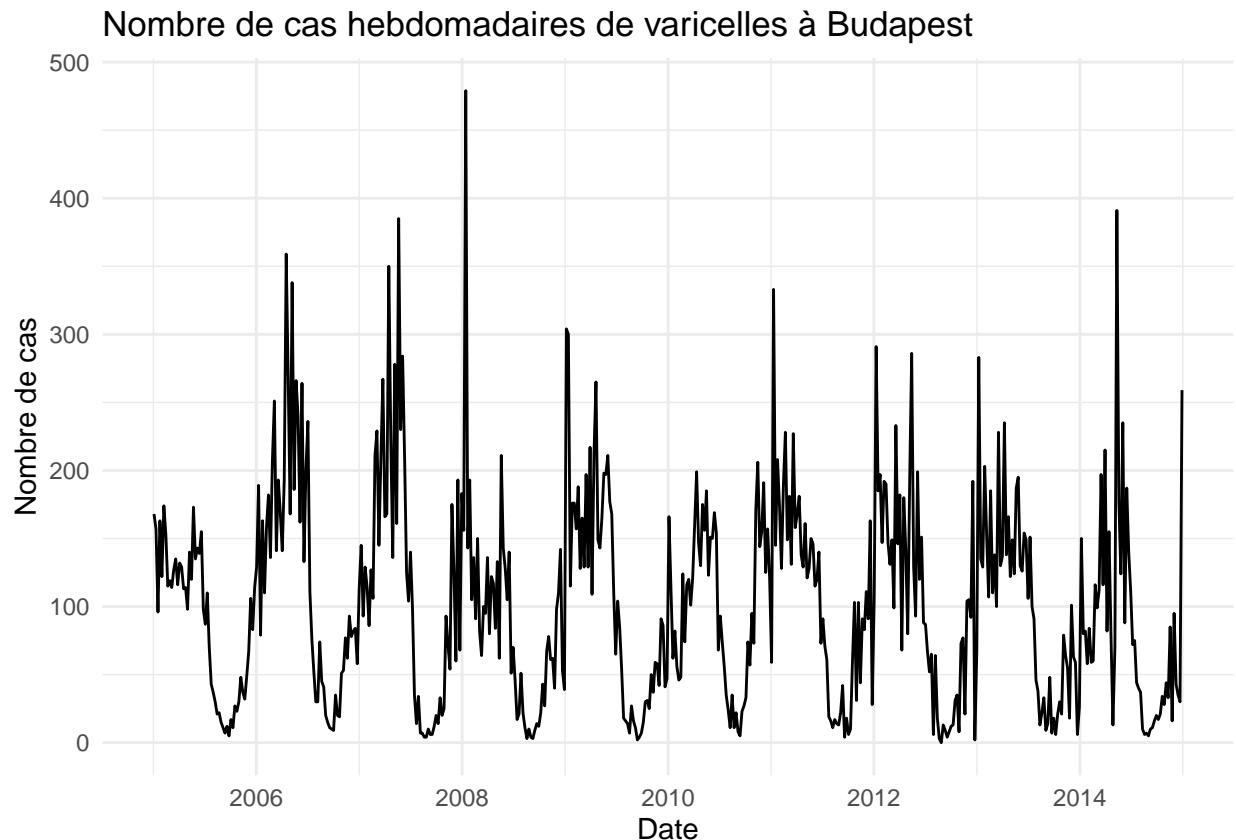
```

Choix de la ville

```

#On choisit BUDAPEST
budapest <- data[,1:2]
colnames(budapest) <- c("date", "nb")
budapest %>% ggplot() + aes(x=date, y=nb) + geom_line() + ggttitle("Nombre de cas hebdomadaires de varicelles à Budapest")

```



De cette première représentation, on ne voit pas de tendance, ou alors une légère décroissance. En revanche, on remarque bien une saisonnalité. De plus, on imagine un modèle additif puisque l'on voit une amplitude plutôt constante.

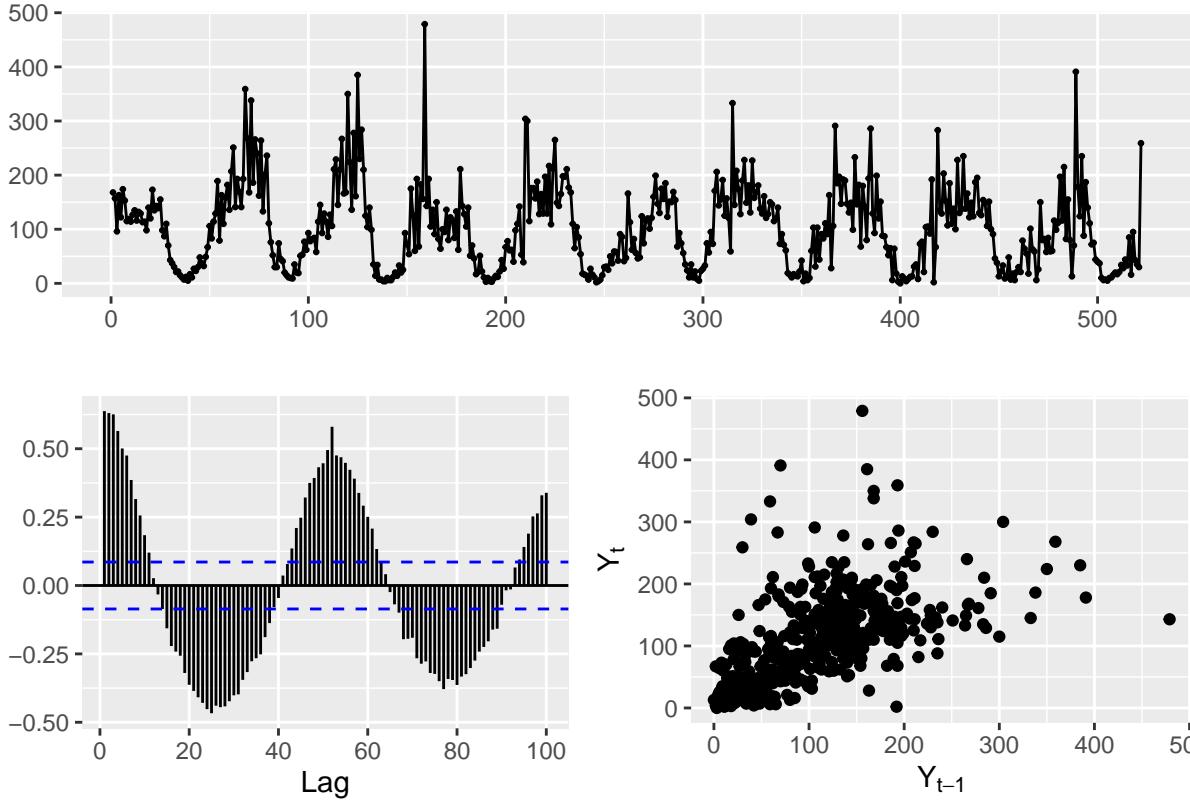
Première approche

```

budapest %>% select(nb) %>%
  ggtsdisplay(
    plot.type = "scatter",

```

```
    lag.max=100
)
```



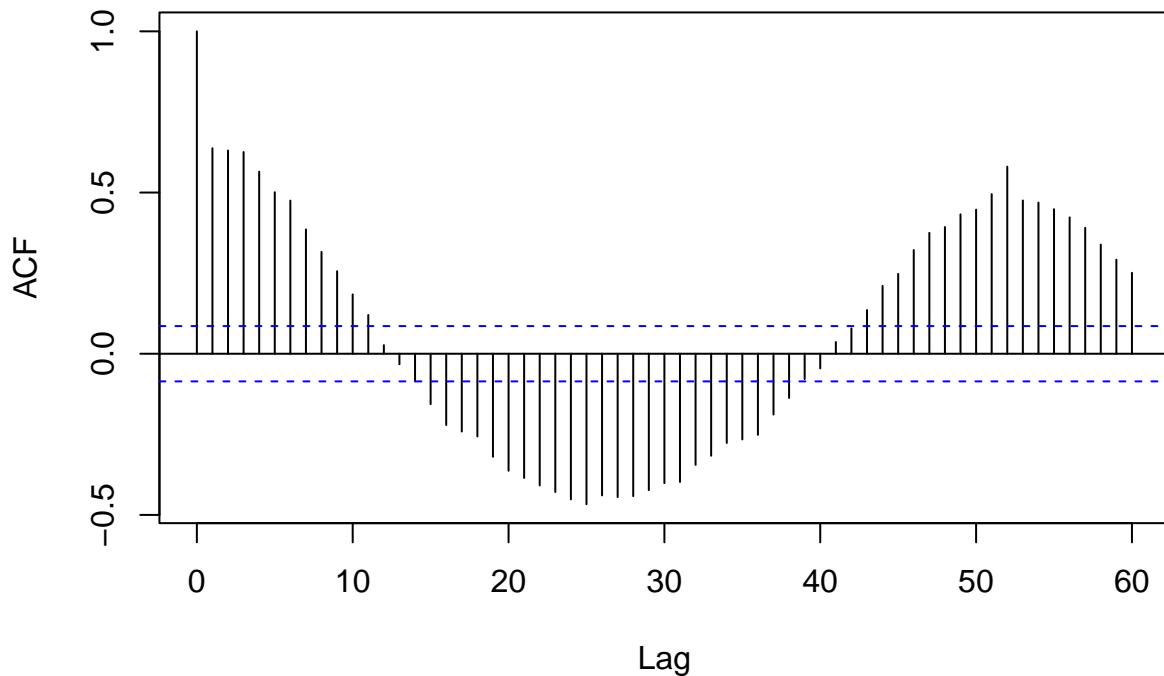
La fonction d'autocorrélation est périodique, ce qui indique une périodicité dans la série temporelle. La ligne pointillée bleue indique le niveau en-dessous duquel la corrélation n'est plus statistiquement significative.

Le nuage de point permet de visualiser l'auto-corrélation d'ordre 1, soit le quotient des covariances empiriques par la variance empirique. Plus le nuage de points est arrondis plus l'auto-corrélation est proche de 1. Ici on ne distingue rien de "remarquable".

Focus sur l'auto-corrélation :

```
acf(budapest$nb, lag.max = 60)
```

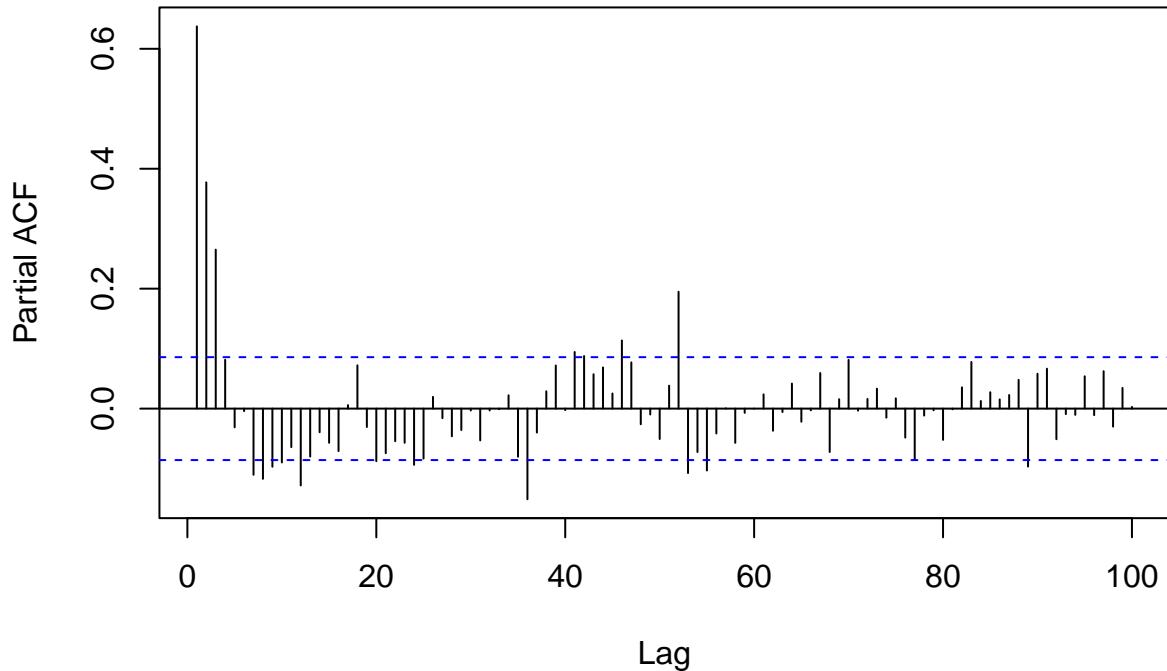
Series budapest\$nb



On voit ici que la périodicité est de 52 jours. Le corrélogramme indique donc des fortes autocorrélations qui se répètent de manière périodique, vérifions qu'il ne s'agit pas d'un effet résiduel avec la fonction *pacf* qui mesure l'autocorrélation partielle. Elle permet de mesurer l'autocorrélation d'un signal pour un décalage k “indépendamment” des autocorrélations pour les décalages inférieurs.

```
pacf(budapest$nb, lag.max = 100)
```

Series budapest\$nb

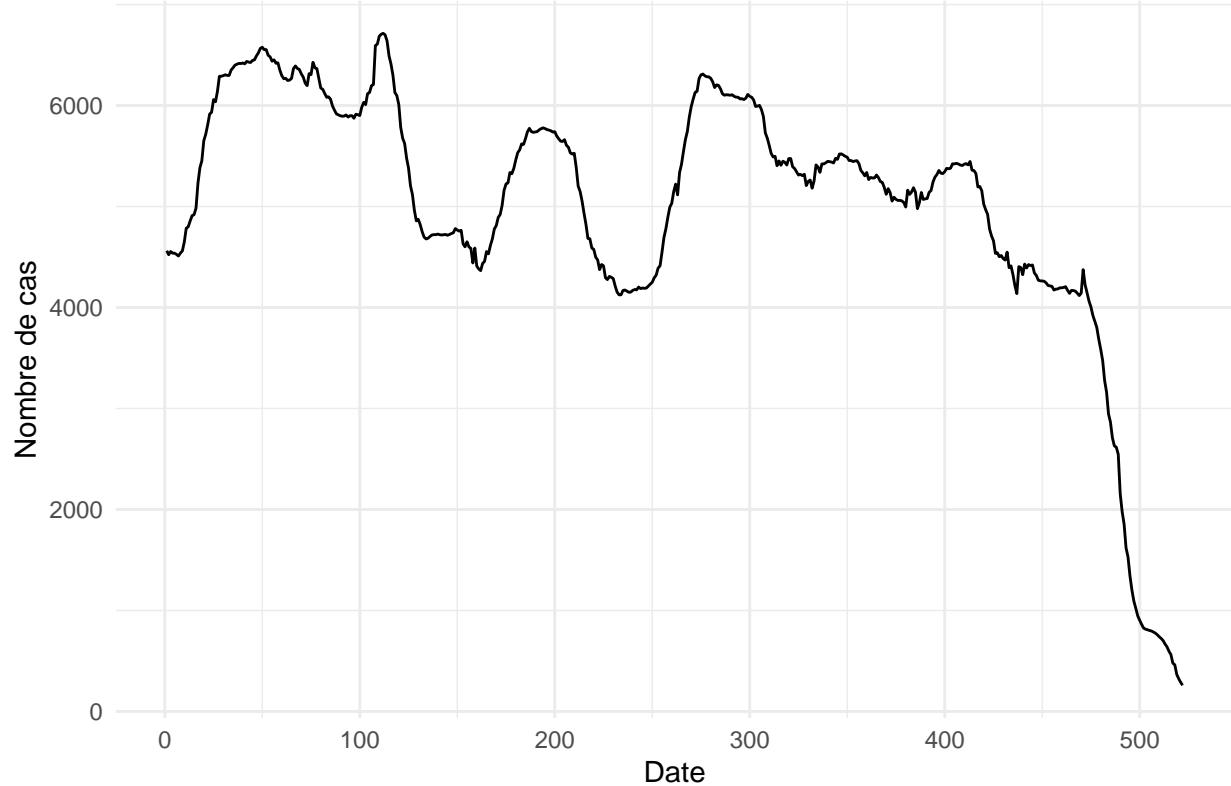


On voit bien que l'auto-corrélation atteint les mêmes valeurs, il n'y a donc pas d'effet résiduel.

On regarde l'évolution tout les 52 jours sur l'année :

```
v=c()
for (i in 1:522){
  cp<-window(budapest$nb,start=c(i,1),end=c(i,52))
  v=c(v,sum(cp))
}
data.frame(x=1:522, y=v) %>% ggplot() + aes(x=x, y=y) + geom_line() + ggtitle("Nombre de cas sur 52 jours")
```

Nombre de cas sur 52 jours de varicelles à Budapest



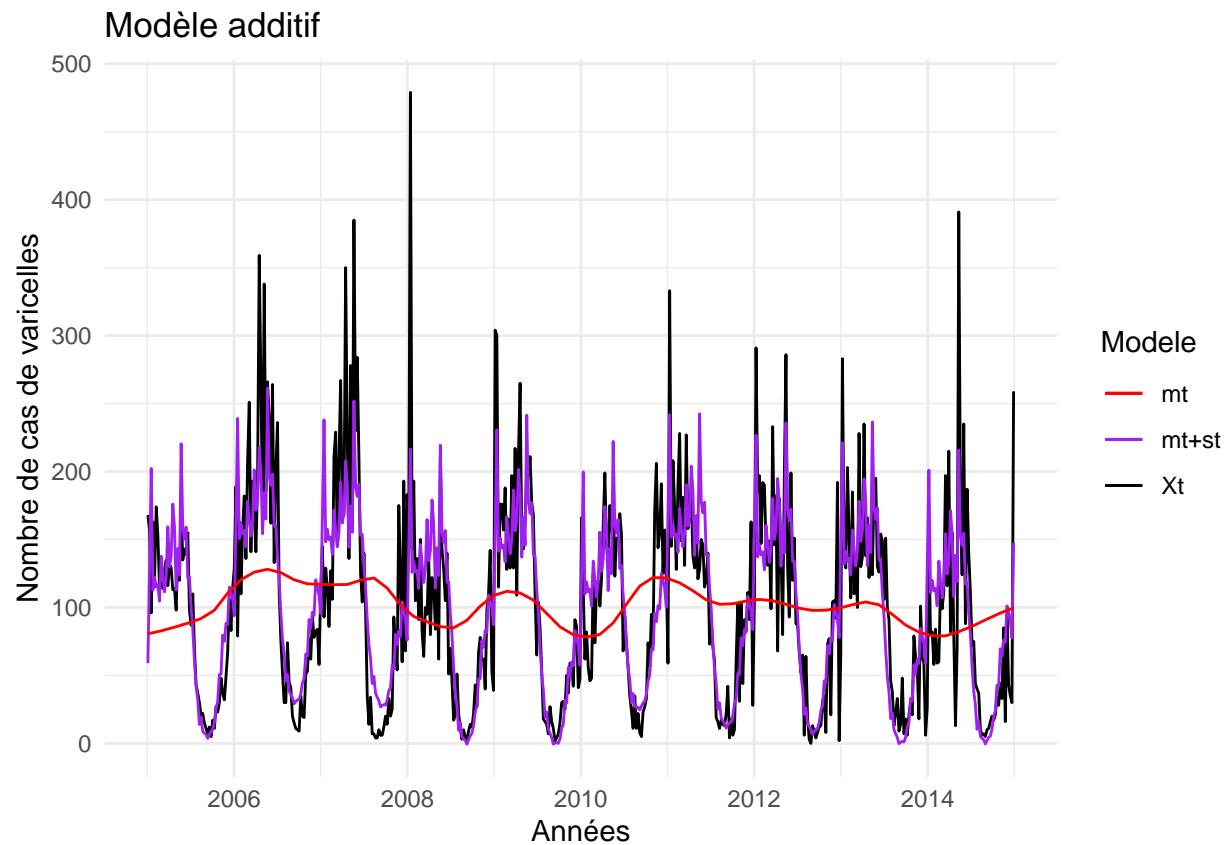
On ne décèle pas de tendance évidente si ce n'est qu'une décroissance vers la fin de la série.

Décomposition de la série

```
temp.ts <- ts(budapest$nb, start=c(2005,1,3), frequency=52)
mod_stl_add <- stl(temp.ts, s.window = "periodic")

budapest_decomp <- cbind(budapest,as.data.frame(mod_stl_add$time.series))

budapest_decomp %>% ggplot() +
  geom_line(aes(x = date, y=nb, color="Xt")) +
  geom_line(aes(x=date, y=trend+seasonal, color="mt+st")) + geom_line(aes(x=date, y=trend, color="mt"))
  scale_color_manual(values = c("red", "purple", "black")) +
  theme(legend.position = c(0.8, 0.08), legend.direction = "horizontal") +
  labs(colour = "Modèle") + ggtitle("Modèle additif") +
  xlab("Années") + ylab("Nombre de cas de varicelles") + theme_minimal()
```



Elimination de la saisonnalité

Methode avec opérateur différence

```
T=52
x1=diff(budapest$nb,lag=T,difference=1)
x2=diff(budapest$nb,lag=T,difference=2)
x3=diff(budapest$nb,lag=T,difference=3)
print(mean(x1))
```

```
## [1] -0.3978723
```

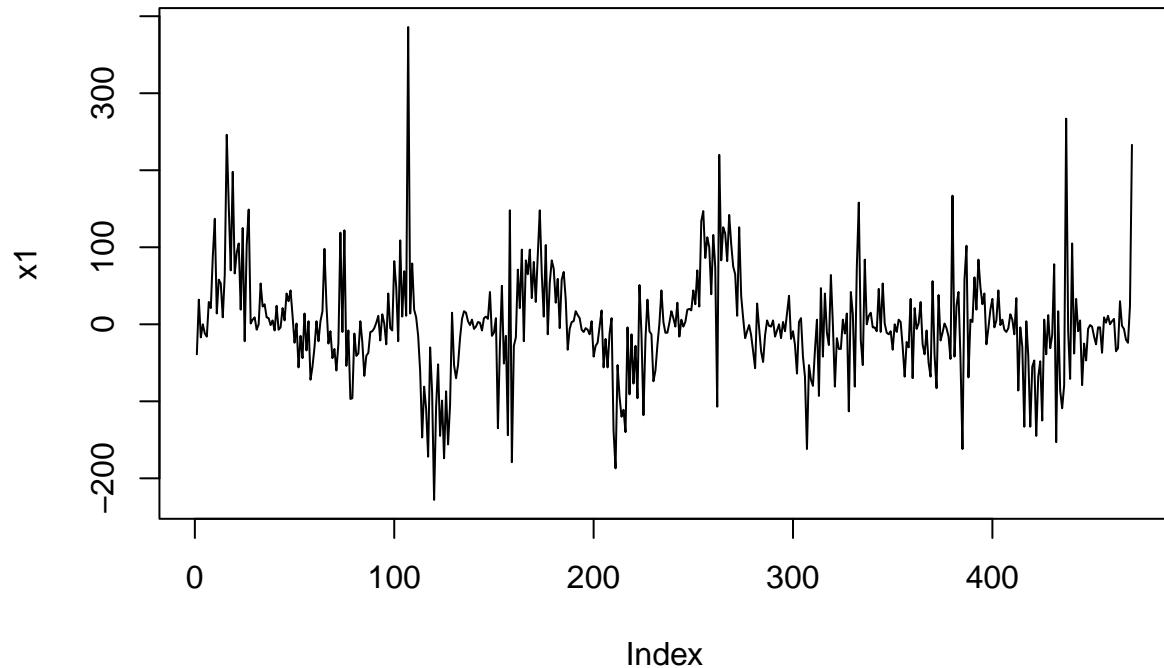
```
print(mean(x2))
```

```
## [1] -6.502392
```

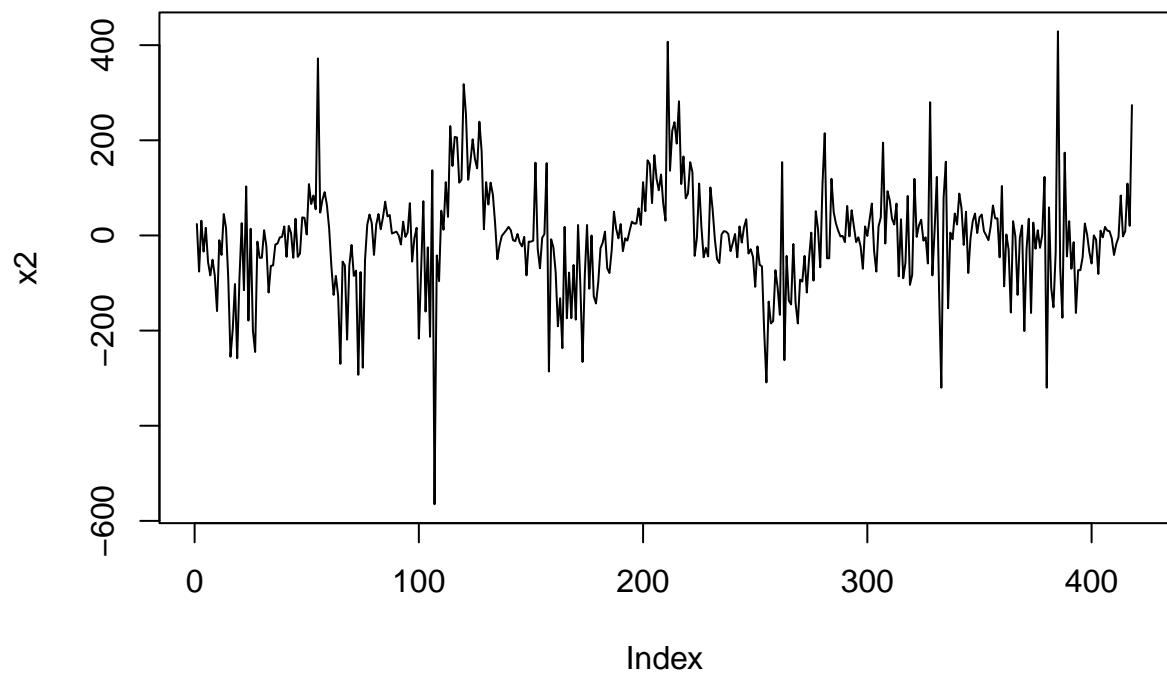
```
print(mean(x3))
```

```
## [1] 4.409836
```

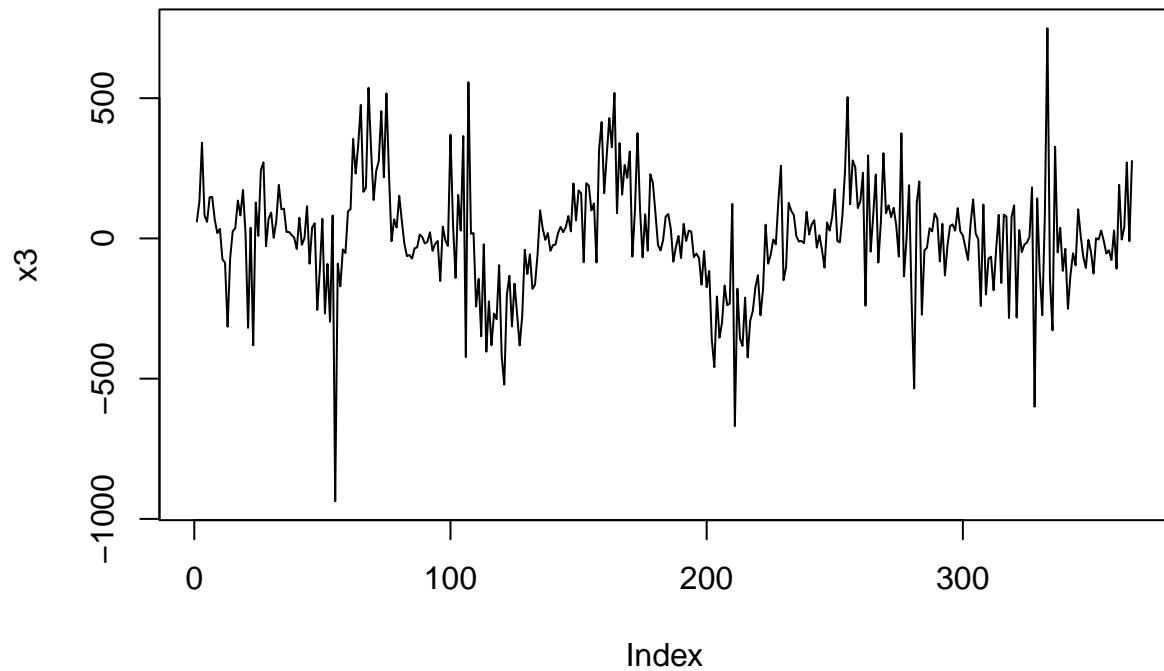
```
# mean(x2) et mean(x3) sont du même ordre donc on en déduit que x2 est de moyenne nulle, donc la tendan  
plot(x1, type='l')
```



```
plot(x2, type='l')
```



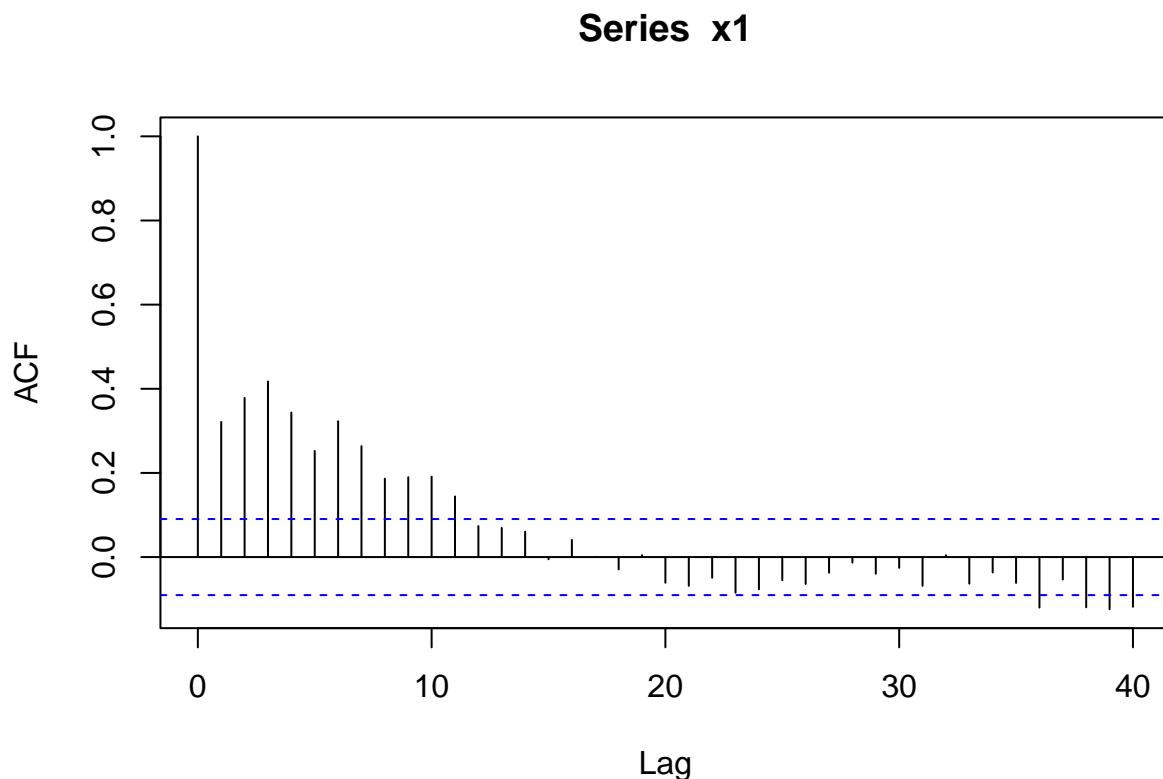
```
plot(x3, type='l')
```



```
Box.test(x1,lag=20)
```

```
##  
## Box-Pierce test  
##  
## data: x1  
## X-squared = 434.29, df = 20, p-value < 2.2e-16
```

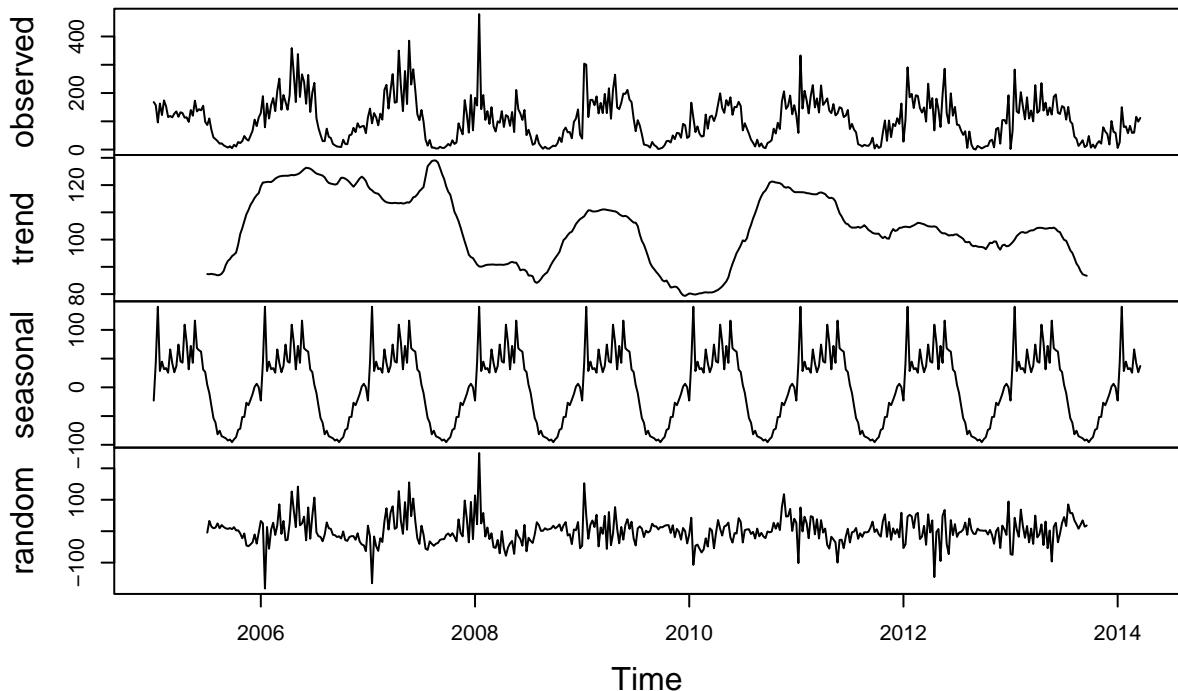
#On voit que la valeur de la statistique est en dessous du seuil (0,05) donc on garde l'hypothèse "bruite"



Fonction *decompose* : moyennes mobiles

```
budapest.ts <- ts(budapest$nb, start=c(2005,1,3), end =c(2014,12,29), frequency=52)
budapest_dec <- decompose(budapest.ts, type = "additive")
plot(budapest_dec)
```

Decomposition of additive time series



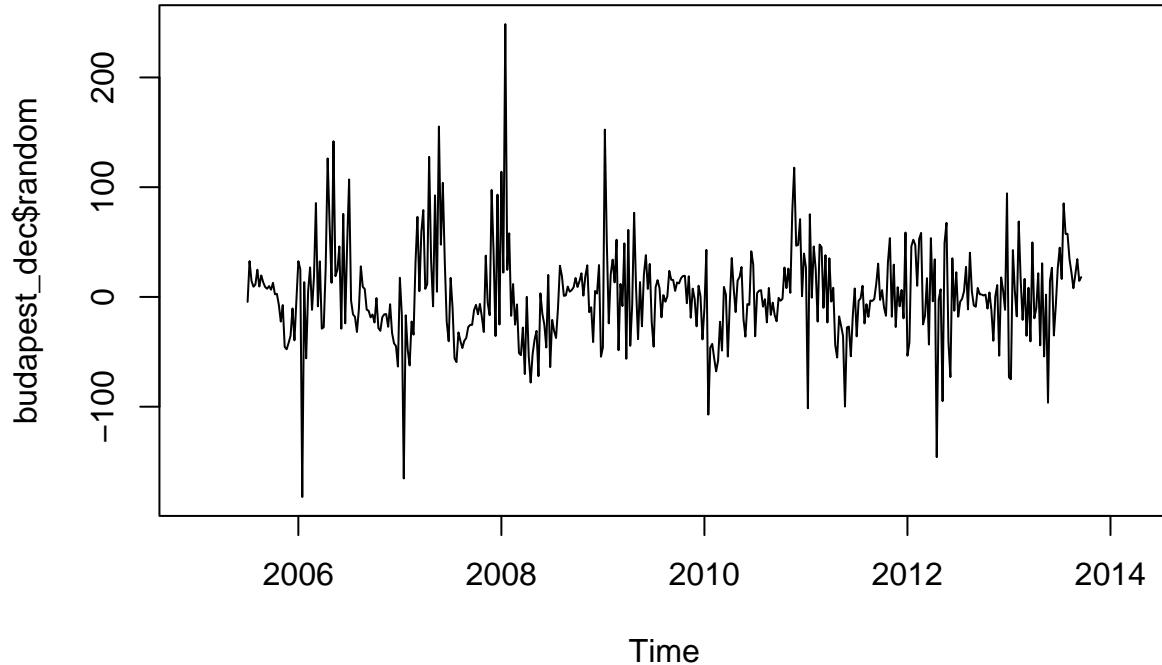
Test sur les résidus

```
#Effectuons un test de niveau 0,95  
Box.test(budapest_dec$random, lag=20)
```

```
##  
## Box-Pierce test  
##  
## data: budapest_dec$random  
## X-squared = 193.29, df = 20, p-value < 2.2e-16
```

La p-valeur est supérieure au seuil de signification, on peut alors conclure que les valeurs résiduelles sont indépendantes et donc décorrélées.

```
plot(budapest_dec$random)
```



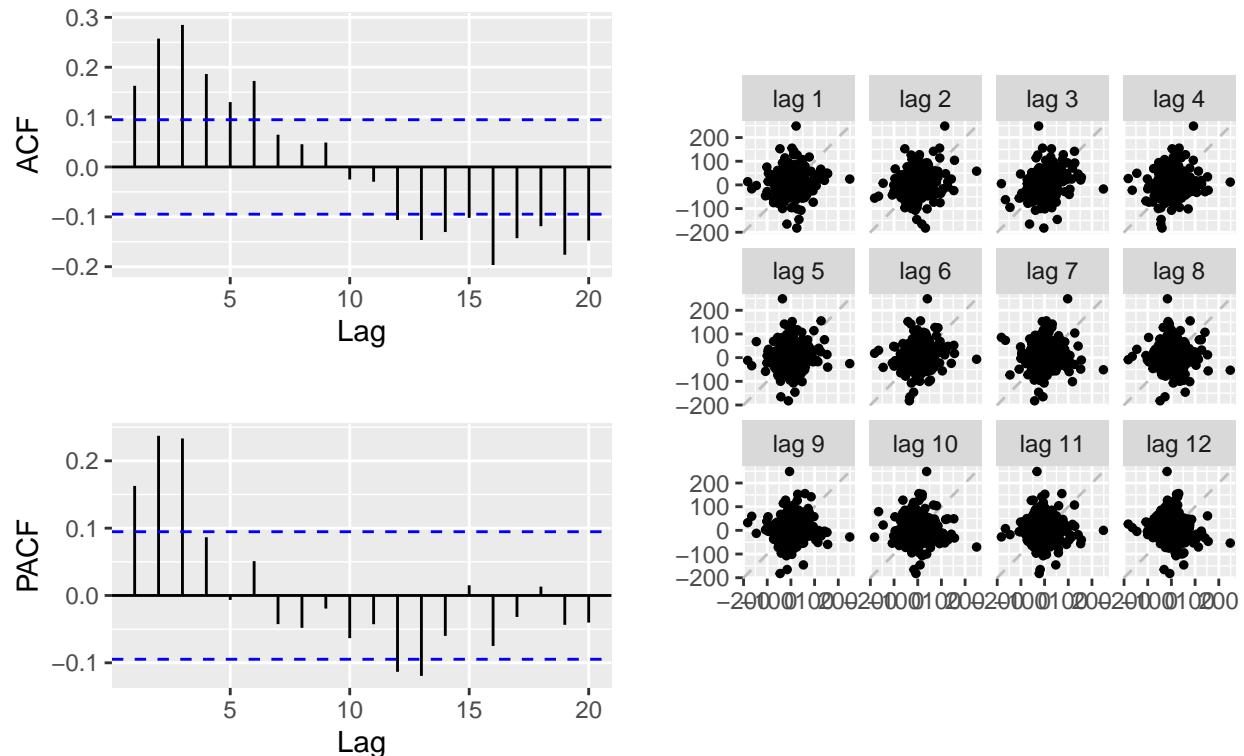
Le résidus semblent bien centrés en 0. Ainsi on peut dire qu'il s'agit de bruits blanc, ils sont centrés et décorrélates.

```
p1 <- gglagplot(budapest_dec$random, do.lines = FALSE, set.lags = 1:12, colour = FALSE)
p2 <- ggAcf(budapest_dec$random, lag.max = 20) + ggtitle(" ")
p3 <- ggPacf(budapest_dec$random, lag.max = 20) + ggtitle(" ")

grid.arrange(top = "Etude des résidus", p2,p3, p1, layout_matrix = rbind(c(1,3),c(2,3)))

## Warning: Removed 624 rows containing missing values (geom_point).
```

Etude des résidus



Le lag-plot nous laisse penser que l'auto-corrélation est proche de 1 pcq nuage arrondi PACF ACF ?

Modélisation de la série

Choix modèles ArR(p) ou MA(q) ou ARMA(p,q) ?

Modèle retenus