

Problématique de notre sujet

Mathis Boutin - Romain Jegoux - Mattéo Rouanne - Maxime Tougard

24/10/2023

Contents

1. Import des données	1
1.1 Création de la série temporelle	1
2. Prédiction du nombre de voyageur sans la pandémie de Covid-19	2
2.1 Calcul de la tendance	4
2.2 Estimation de la saisonnalité	5
2.3 Prédiction à l'aide du modèle	7
3. Prédiction du nombre de voyageur à l'aide des données post-covid	8
3.1 Création de la série post covid	8
3.2 Calcul de la tendance	8
3.3 Estimation de la saisonnalité	10
3.4 Prédiction à l'aide du modèle	11
4. Comparaison des deux Prévisions	12

1. Import des données

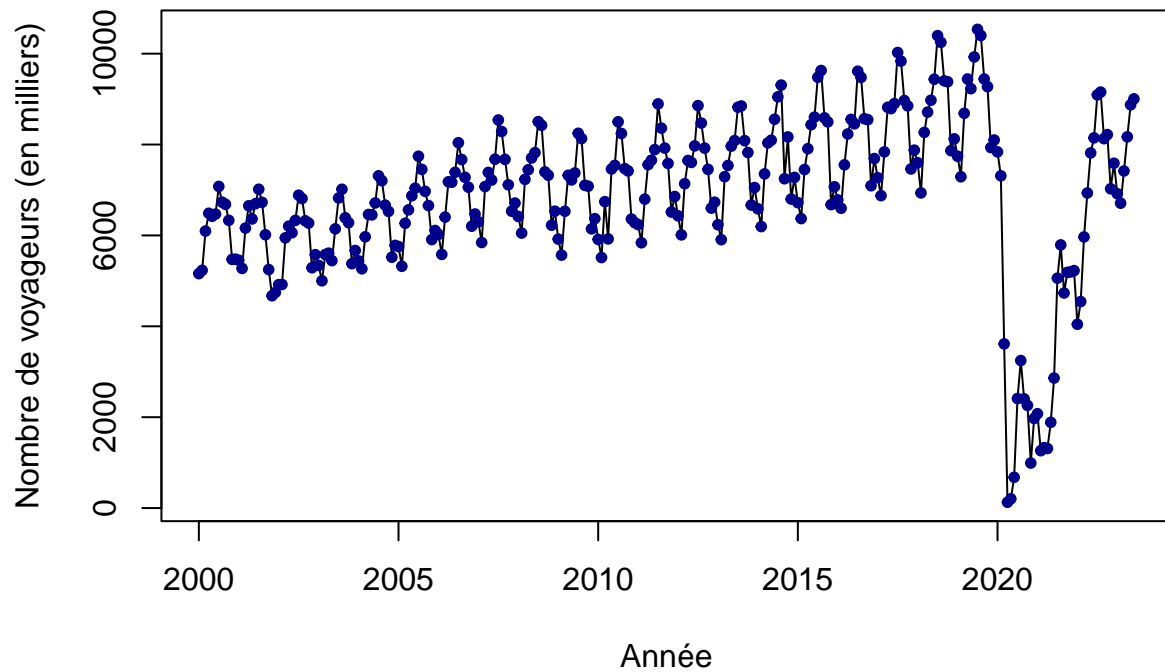
```
traffic <- read.csv("traffic_adp.csv", sep=";")
```

1.1 Création de la série temporelle

```
traffic$totpass = (traffic$passenger_CdG + traffic$passenger_only)/1000
ts_traffic = ts(traffic$totpass, start = c(2000, 1), frequency = 12)

plot(ts_traffic, xlab = "Année", ylab = "Nombre de voyageurs (en milliers)", main = "Traffic aérien des points", col = "darkblue", pch = 20)
```

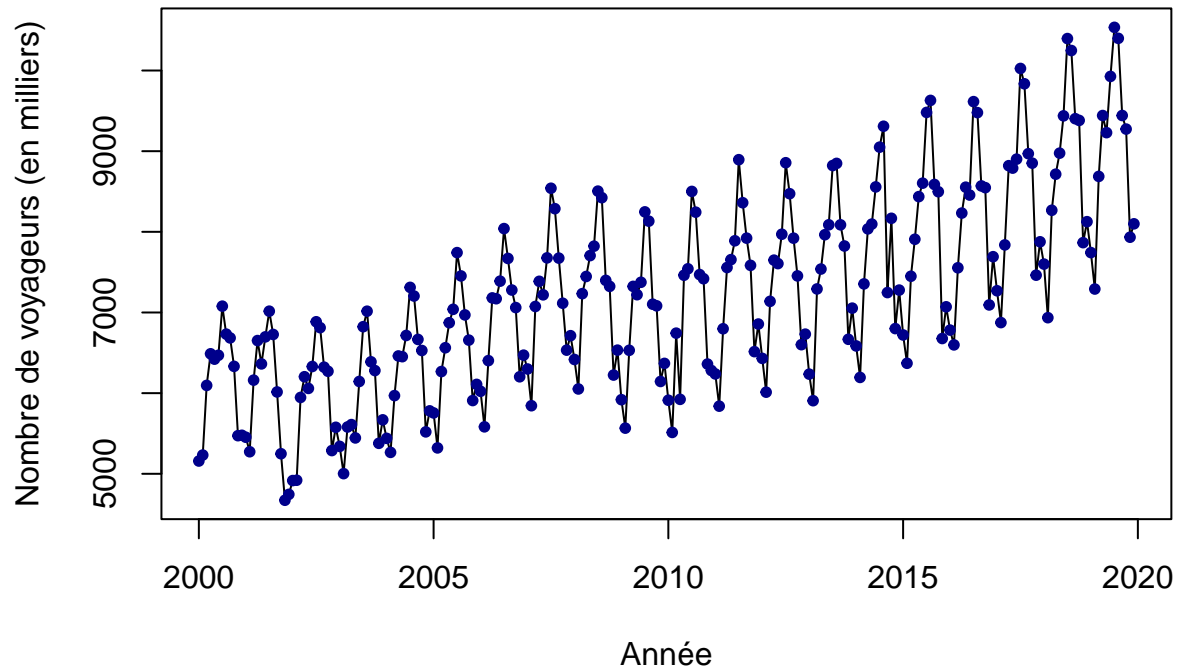
Traffic aérien des aéroports parisiens par années



2. Prédiction du nombre de voyageur sans la pandémie de Covid-19

```
#On enlève les années après 2019
traffic_2000_2019 = window(ts_traffic,end= c(2019,12))
plot(traffic_2000_2019 ,xlab="Année", ylab= "Nombre de voyageurs (en milliers)", main = "Traffic aérien
points(traffic_2000_2019, col= "darkblue", pch = 20)
```

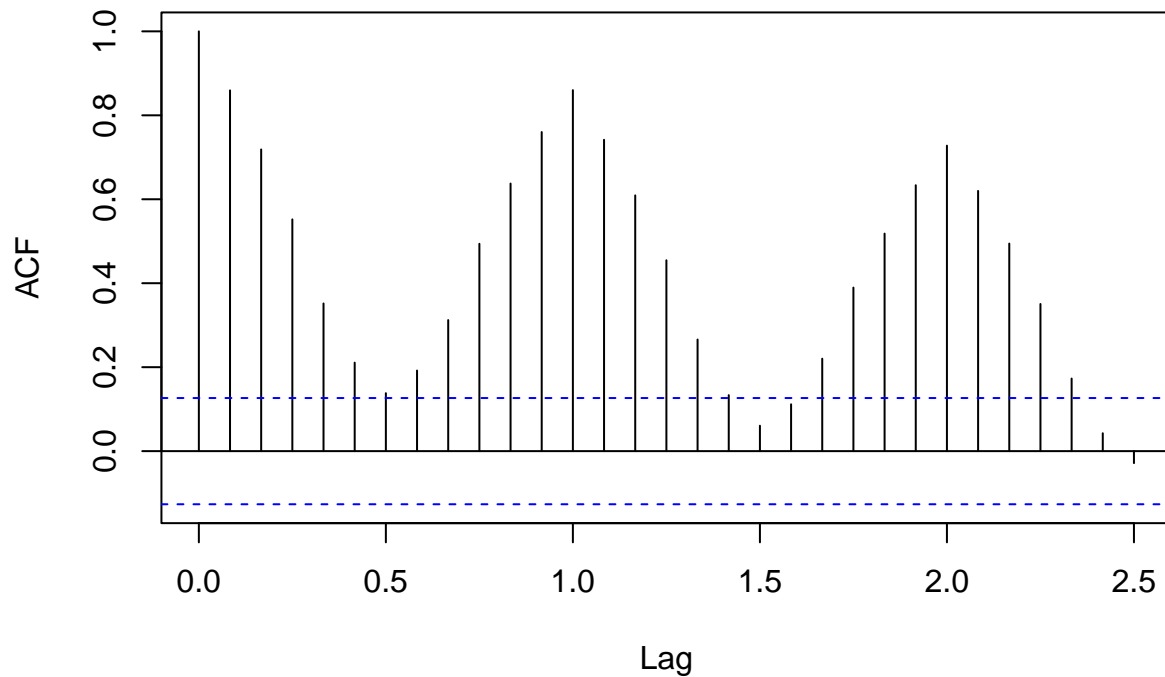
Traffic aérien des aéroports parisiens jusqu'en 2019



On retire les années après 2019 car la crise du covid-19 a fortement impactée le trafic aérien de cette période. On va donc estimer les voyageurs de cette période à l'aide des années précédentes afin de voir l'impact réel de cette crise

```
acf(traffic_2000_2019, lag.max = 30, main= "autocorrélogramme de la série" )
```

autocorrélogramme de la série



Sur cet autocorrélogramme on détecte une forte corrélation tous les 12 de lag soit toutes les années il y a donc une saisonnalité de période 12

2.1 Calcul de la tendance

2.1.1 Détermination du modèle de la tendance

```
p = 1:length(traffic_2000_2019) #longueur de la série
tend_traffic = lm(traffic_2000_2019~p) #modèle linéaire
tend_traffic2 = lm(traffic_2000_2019~p + I(p^2)) #modèle polynomial de degré 2
tend_traffic3 = lm(traffic_2000_2019~p + I(p^2) + I(p^3)) #modèle polynomial de degré 3

cat ("le R2 du modèle linéaire vaut",cor(traffic_2000_2019, p)^2, "\n") # R2 du modèle linéaire
```

```
## le R2 du modèle linéaire vaut 0.5246034
```

```
cat ("le R2 du modèle polynomial de degré 2 vaut",cor(traffic_2000_2019, p + I(p^2))^2, "\n")# R2 du m
```

```
## le R2 du modèle polynomial de degré 2 vaut 0.5223659
```

```
cat ("le R2 du modèle polynomial de degré 3 vaut",cor(traffic_2000_2019, p + I(p^2) + I(p^3))^2, "\n")
```

```
## le R2 du modèle polynomial de degré 3 vaut 0.4923603
```

On choisi donc le modèle linéaire car il a le R^2 le plus grand

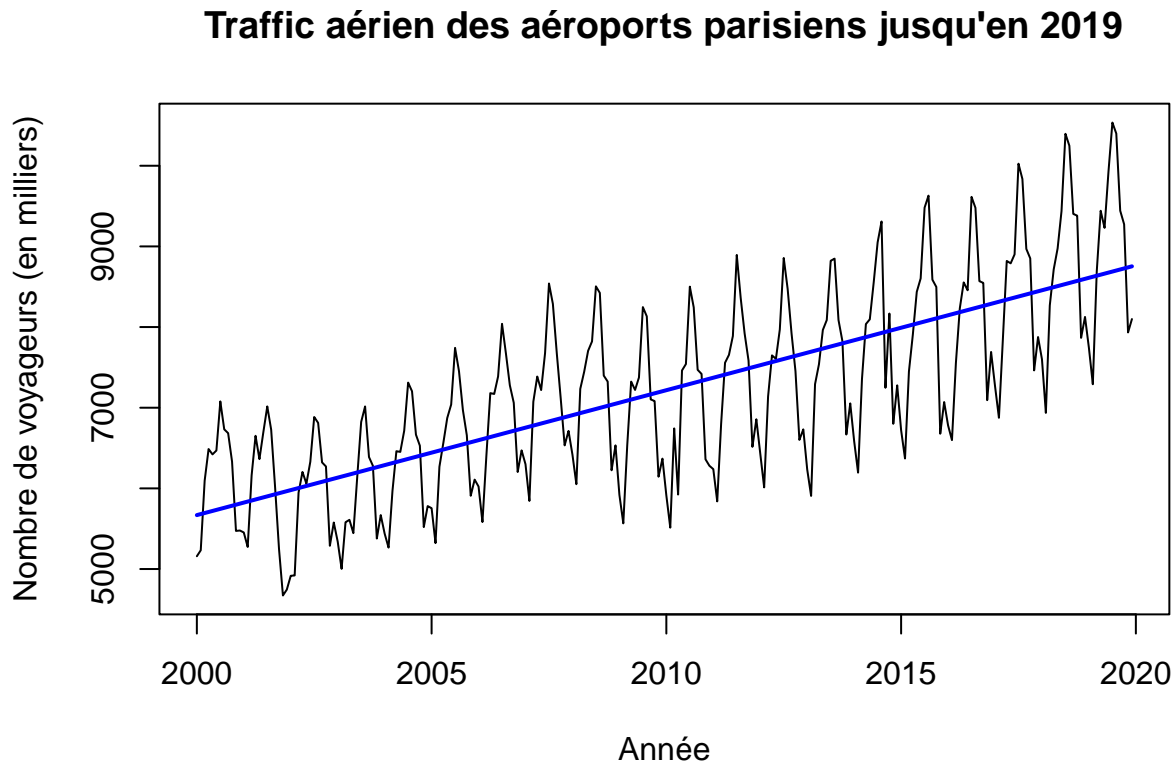
2.1.2 Création de la série temporelle de la tendance

```
fittedval = ts(data = tend_traffic$fitted.values, start = c(2000, 1), frequency = 12)  
tend_traffic$coefficients
```

```
## (Intercept)          p  
## 5655.79928      12.90419
```

On a choisi un modèle linéaire pour estimer la tendance et on récupère donc ses coefficients $\hat{\beta}_0$ et $\hat{\beta}_1$

```
plot(traffic_2000_2019, main = "Traffic aérien des aéroports parisiens jusqu'en 2019" , xlab= "Année", ylab= "Nombre de voyageurs (en milliers)",  
lines(fittedval , col="blue", lwd = 2))
```



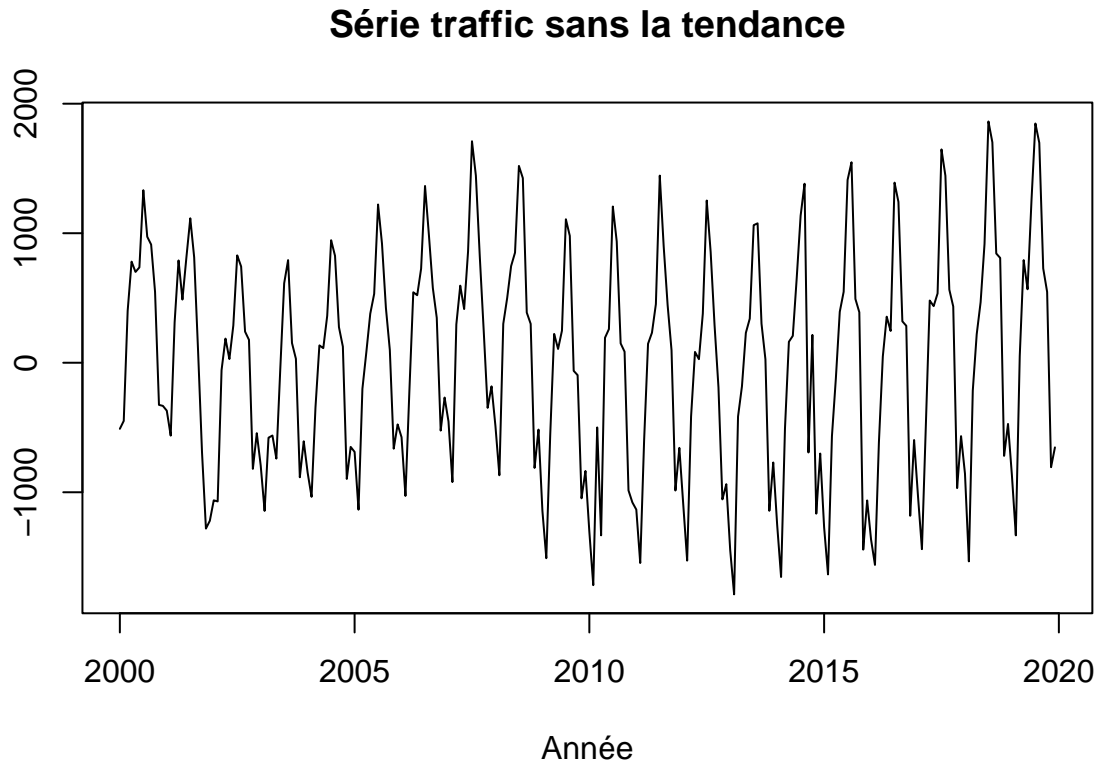
On voit ici la tendance linéaire que l'on a calculé, affichée sur la serie réelle

2.2 Estimation de la saisonnalité

2.2.1 Élimination de la tendance

On élimine la tendance par différenciation

```
notend = traffic_2000_2019 - fittedval
plot(notend , main= "Série traffic sans la tendance", xlab="Année",ylab = "")
```



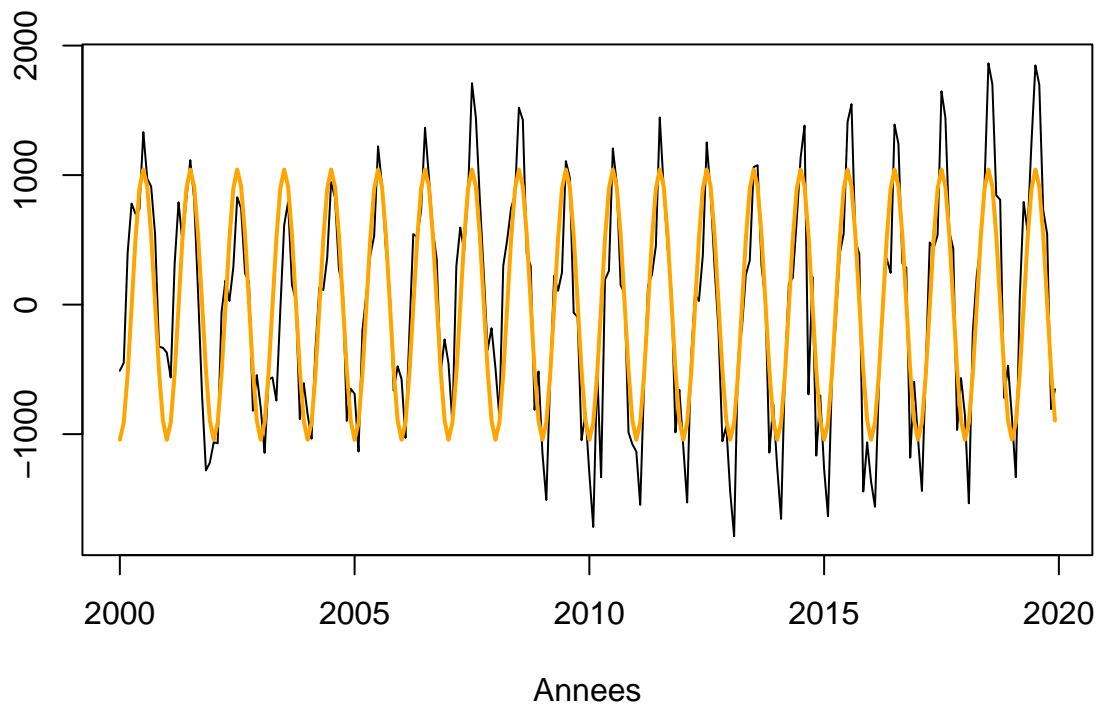
2.2.2 Calcul de la saisonnalité

```
T = 12
nb_NA = sum(is.na(tend_traffic))/2
t = (nb_NA + 1):(length(notend) - nb_NA)

sint <-sin((2*pi*t)/T)
cost <-cos((2*pi*t)/T)

plot (notend, xlab="Annees", ylab = " ", main =" Saisonnalite estimée")
sais_1 = lm(notend[t]~sint + cost )
sais_2 <- ts(sais_1$fit, start = 2000 , freq = 12)
lines(sais_2, col="orange", lwd= 2)
```

Saisonnalité estimée



```
N = 120
n = (length(traffic_2000_2019) + 1):(length(traffic_2000_2019) + N)
df_sais = data.frame(matrix(c(sin(2*pi*n/T), cos(2*pi*n/T)), ncol = 2))
colnames(df_sais) = c("sint", "cost")
pred_tend = predict(sais_1, newdata = df_sais)

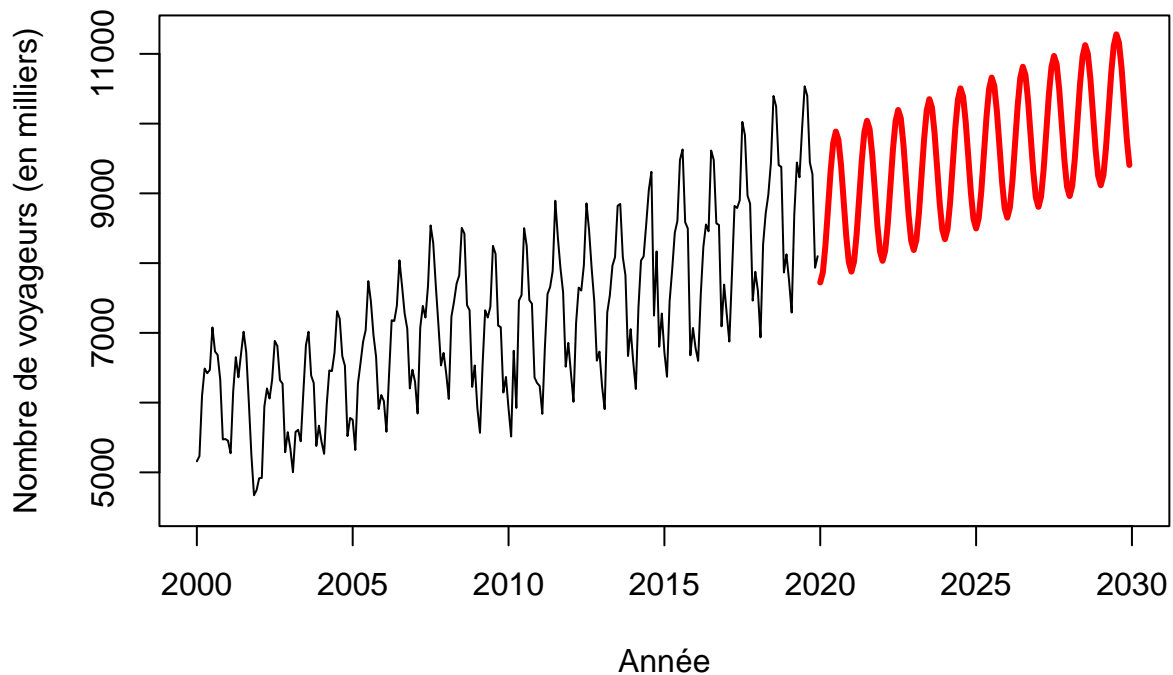
i = (length(traffic_2000_2019) + 1):(length(traffic_2000_2019) + N)
df_tend = data.frame(i)
colnames(df_tend) = c("p")
pred_sais = predict(tend_traffic, newdata = df_tend)

pred_tend_sais = ts(data = pred_tend+pred_sais, start = c(2020,1), frequency = 12)
ajust= ts(data =tend_traffic$fitted.values+sais_1$fitted.values, start = c(2000,1), frequency = 12)
```

2.3 Prédiction à l'aide du modèle

```
plot(traffic_2000_2019, xlim = c(2000,2020 + N/12),ylim=c(4500,max(pred_tend_sais)),main = "Prévisions a
lines(pred_tend_sais, lwd = 3,col = "red")
```

Prévisions après 2020



3. Prédiction du nombre de voyageur à l'aide des données post-covid

On choisi de faire cette étude sur les données post covid. On crée donc un série temporelle sur les données a partir d'avril 2022 soit 2 ans après le confinement. On a pu remarquer sur les graphiques précédents qu'il semblait y avoir une reprise a la normale du trafic aérien a cette période. Il faut toutefois relativiser puisque les prédictions se basent sur les données entre avril 2022 à juin 2023 il y a donc peu d'observations.

3.1 Création de la série post covid

```
traffic_post_cov = window(ts_traffic, start= c(2022,4))
```

3.2 Calcul de la tendance

3.2.1 Choix du modèle pour représenter la tendance

```
p = 1:length(traffic_post_cov) #longueur de la série
tend_traffic_c = lm(traffic_post_cov~p) #modèle linéaire
tend_traffic2_c = lm(traffic_post_cov~p + I(p^2)) #modèle polynomial de degré 2
```



```
cat ("le R2 du modèle linéaire vaut",cor(traffic_post_cov, p)^2, "\n") # R2 du modèle linéaire
```

```
## le R2 du modèle linéaire vaut 0.005476107
```

```
cat ("le R2 du modèle polynomial de degré 2 vaut",cor(traffic_post_cov, p + I(p^2))^2 , "\n")# R2 du mo
```

```
## le R2 du modèle polynomial de degré 2 vaut 0.01544017
```

On choisi le modèle linéaire car c'est celui avec le R^2 le plus grand

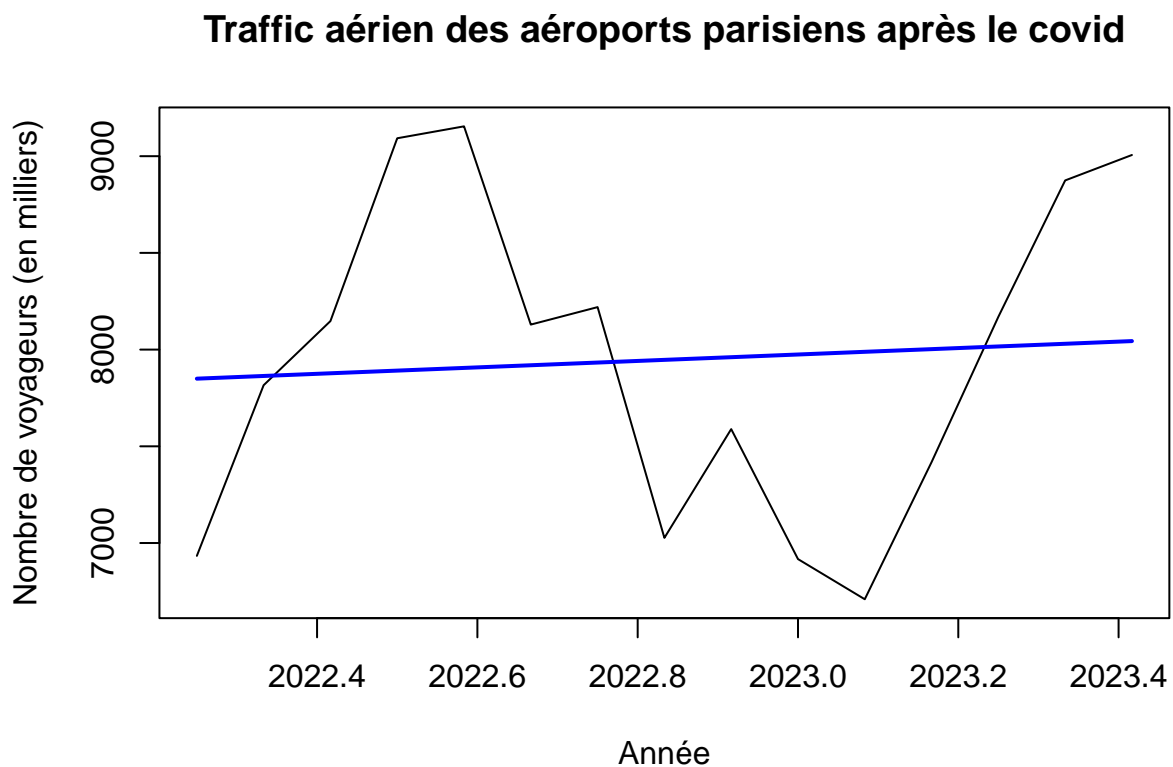
3.2.2 Création série temporelle de la tendance

```
fittedval_c = ts(data = tend_traffic_c$fitted.values, start = c(2022, 4), frequency = 12)
tend_traffic_c$coefficients
```

```
## (Intercept)          p
## 7835.89736    13.87357
```

On a choisi un modèle linéaire pour estimer la tendance et on récupere donc ses coefficients $\hat{\beta}_0$ et $\hat{\beta}_1$

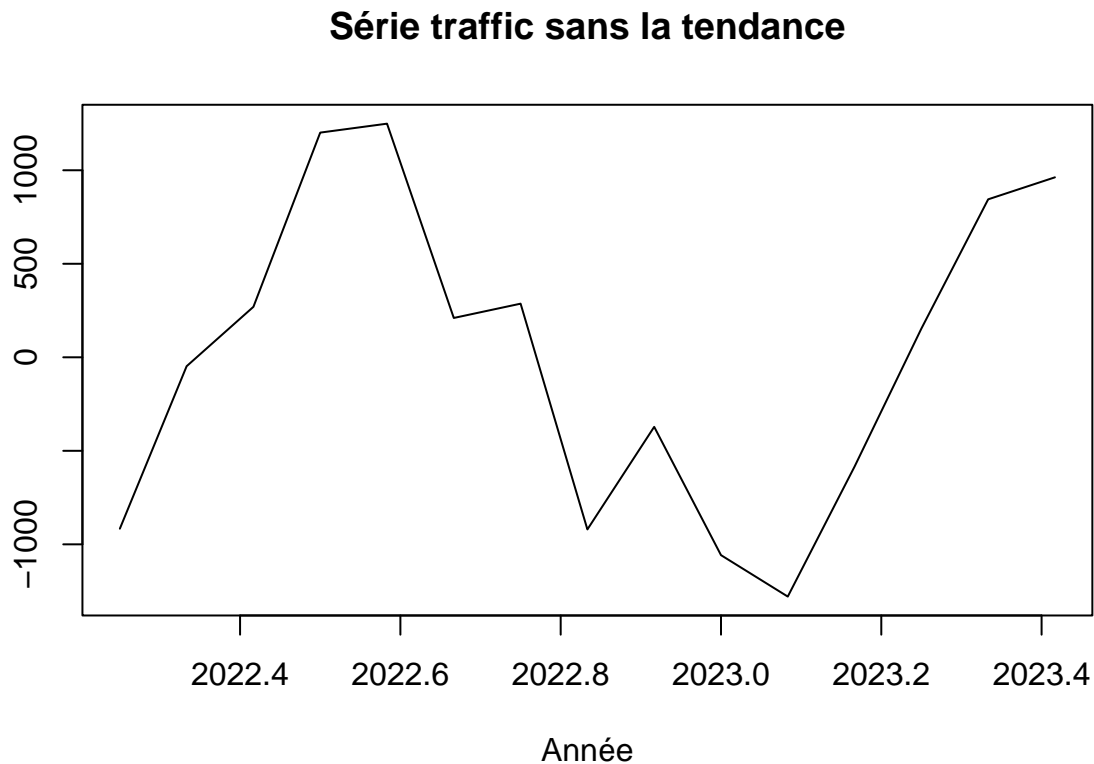
```
plot(traffic_post_cov, main = "Traffic aérien des aéroports parisiens après le covid" , xlab= "Année", ylab= "Nombre de voyageurs (en milliers)",
lines(fittedval_c , col="blue", lwd = 2))
```



3.3 Estimation de la saisonnalité

3.3.1 Élimination de la tendance

```
notend_c = traffic_post_cov - fittedval_c
plot(notend_c , main= "Série trafic sans la tendance", xlab="Année",ylab = "")
```



On élimine la tendance par différenciation

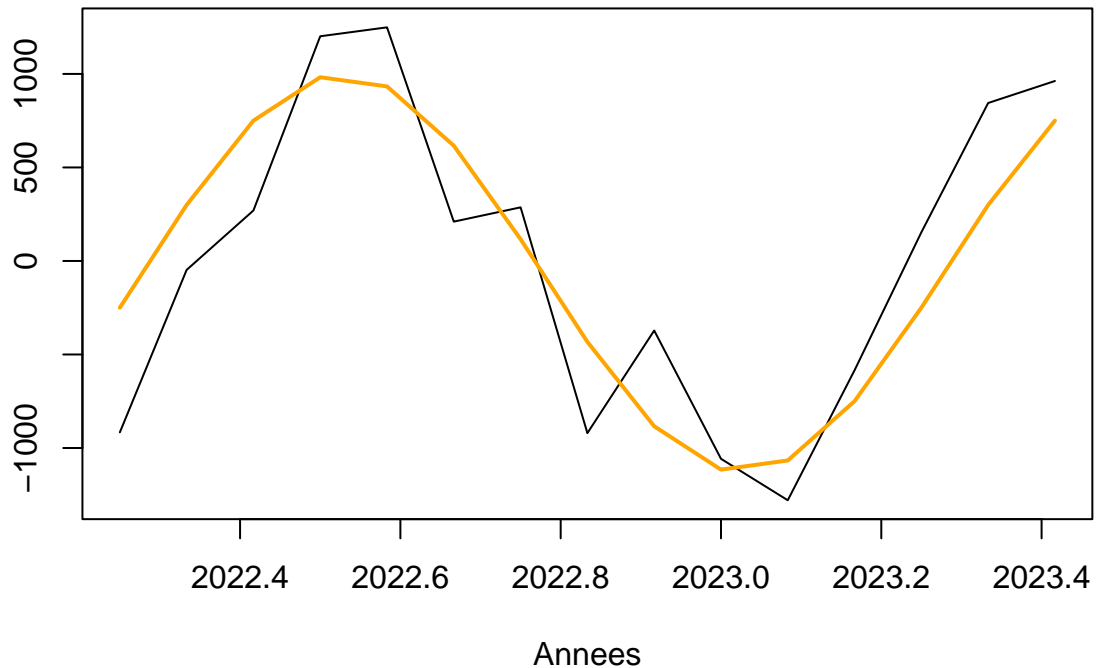
3.3.2 Estimation de la saisonnalité

```
T = 12
nb_NA = sum(is.na(tend_traffic_c))/2
t = (nb_NA + 1):(length(notend_c) - nb_NA)

sint <-sin((2*pi*t)/T)
cost <-cos((2*pi*t)/T)

plot (notend_c, xlab ="Annees", ylab =" ", main =" Saisonnalite estimée")
sais_1 = lm(notend_c[t]~sint + cost )
sais_2 <- ts(sais_1$fit, start = c(2022,4) , freq = 12)
lines(sais_2, col="orange", lwd= 2)
```

Saisonnalité estimée



3.4 Prédiction à l'aide du modèle

3.4.1 Prédiction de la tendance

```
C19 = 76
n = (length(traffic_post_cov) + 1):(length(traffic_post_cov) + C19)
df_sais = data.frame(matrix(c(sin(2*pi*n/T), cos(2*pi*n/T)), ncol = 2))
colnames(df_sais) = c("sint", "cost")
pred_tend = predict(sais_1, newdata = df_sais)
```

3.4.2 Prédiction de la saisonnalité

```
i = (length(traffic_post_cov) + 1):(length(traffic_post_cov) + C19)
df_tend = data.frame(i)
colnames(df_tend) = c("p")
pred_sais = predict(tend_traffic_c, newdata = df_tend)
```

3.4.3 Prédiction selon le modèle

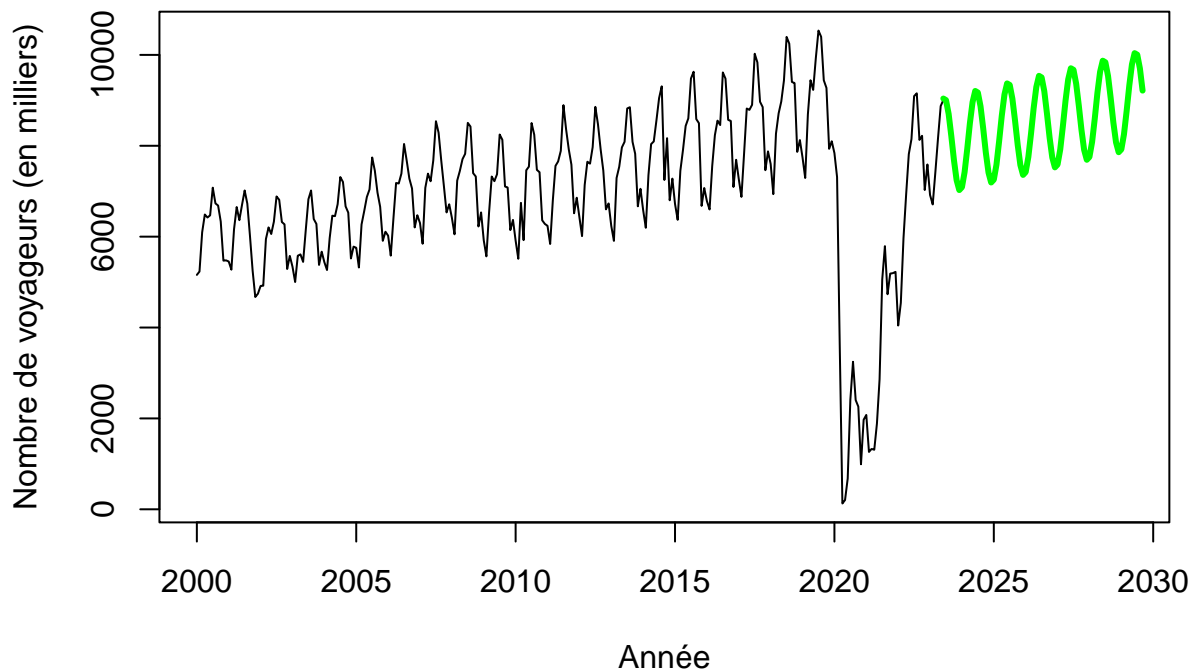
```

pred_tend_sais_c = ts(data = pred_tend+pred_sais, start = c(2023,6), frequency = 12)
ajust= ts(data =tend_traffic$fitted.values+sais_1$fitted.values, start = c(2000,1), frequency = 12)

plot(ts_traffic, xlim = c(2000,2023 + C19/12),main = "Prévisions après 2020",xlab="Année" ,ylab = "Nombre de voyageurs (en milliers)")
lines(pred_tend_sais_c, lwd = 3,col = "green")

```

Prévisions après 2020



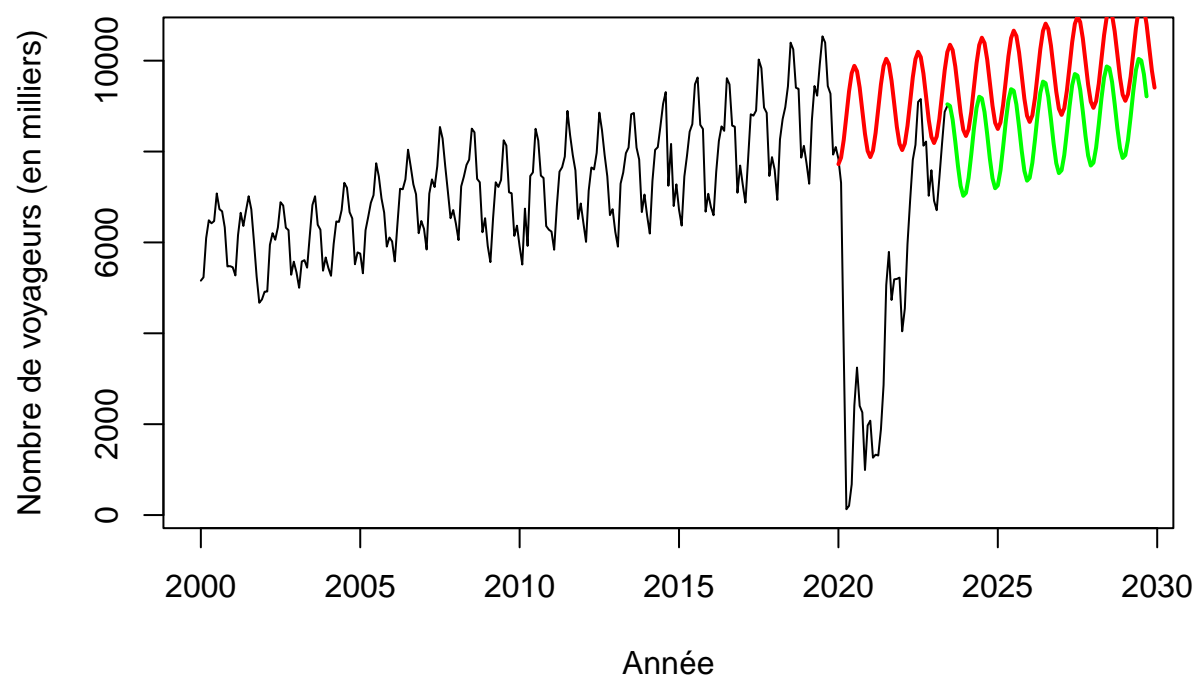
4. Comparaison des deux Prévisions

```

plot(ts_traffic, xlim = c(2000,2023 + C19/12),main = "Comparaison des modèles",xlab="Année" ,ylab = "Nombre de voyageurs (en milliers)")
lines(pred_tend_sais_c, lwd = 2,col = "green")
lines(pred_tend_sais, lwd = 2,col = "red")

```

Comparaison des modèles



Il alors est possible de voir que, si les tendances se poursuivent, d'ici 2030 le trafic aérien a CDG et Orly aura quasiment repris un rythme normal