

Série Chronologique - Projet

M. Le Paumier, A. Maïtammar

2021-12-31

Contents

1	Introduction	3
2	Analyse visuelle	3
3	Décomposition	4
4	Différenciation	6
5	Modélisation	8
6	Prédiction	11
7	Conclusion	11
8	Annexe 1 : Différenciation	12
9	Annexe 2 : Décomposition STL	15
10	Annexe 3 : Transformation Box-Cox	16
11	Annexe 4 : xreg	19
12	Annexe 5 : auto.arima	20

1 Introduction

L'objet d'étude de ce projet est une série temporelle décrivant le nombre de commandes par jour d'une entreprise angevine de livraison de repas à domicile. Cette série est découpée en un jeu d'entraînement (train set) de 280 observations et un jeu de test (test set) de 21 observations. Notre objectif sera d'étudier cette série et de construire un modèle de la famille des modèles *ARMA* afin de prédire les valeurs du jeu de test soit 21 jours.

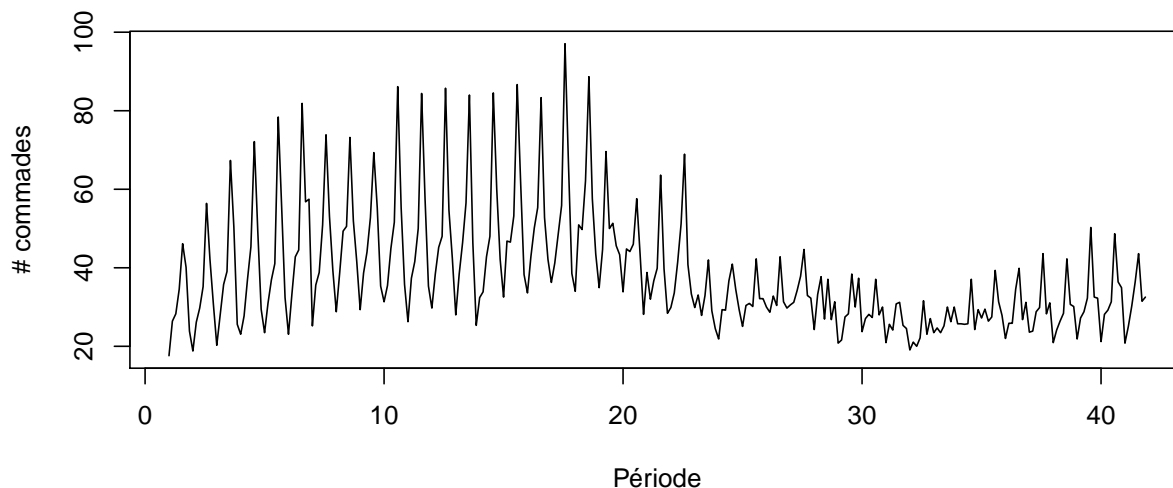


Figure 1: Série originale (entraînement)

2 Analyse visuelle

Cette série présente plusieurs caractéristiques remarquables.

Il y a une forte saisonnalité sur les jours de la semaine. En effet, le nombre de commandes croît du dimanche au vendredi puis décroît le samedi en moyenne.

La série semble composée de deux voire trois régimes. Un premier lors des 18 premières semaines avec une tendance haussière, une saisonnalité très marquée et une variance relativement grande. Puis vient une période de 16 semaines avec tendance baissière, une saisonnalité moins marquée et une plus faible variance. Enfin, les 5 dernières semaines semblent reprendre le schéma des premières semaines avec une tendance à la hausse, une saisonnalité marquée et une variance qui augmente.

```
## starting httpd help server ... done
```

Les modèles *ARMA* font l'hypothèse que la série temporelle est stationnaire (du second ordre), c'est à dire que sa moyenne et sa covariance sont invariantes par translation dans le temps. Pour rendre cette série stationnaire, nous allons utiliser plusieurs approches telles que la décomposition (partie 2) et la différenciation (partie 3). Puis nous établirons des modèles de la famille *ARIMA* à l'aide des fonctions `auto.arima` et `checkup_res` (partie 3) avant d'évaluer nos modèles sur le jeu de donnée test à l'aide de critères prédictifs (partie 4).

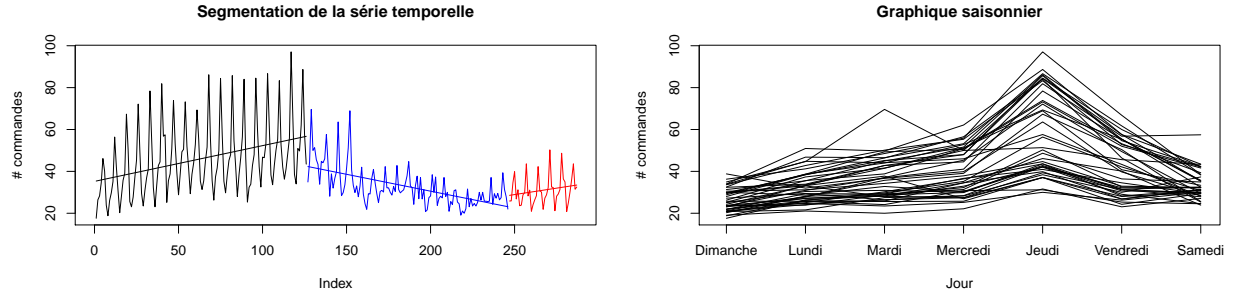
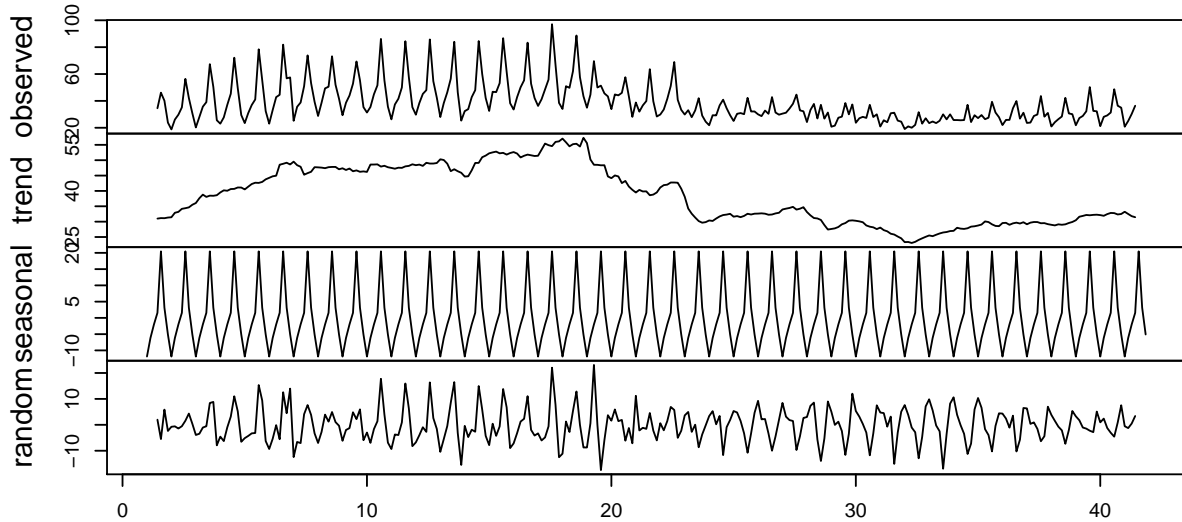


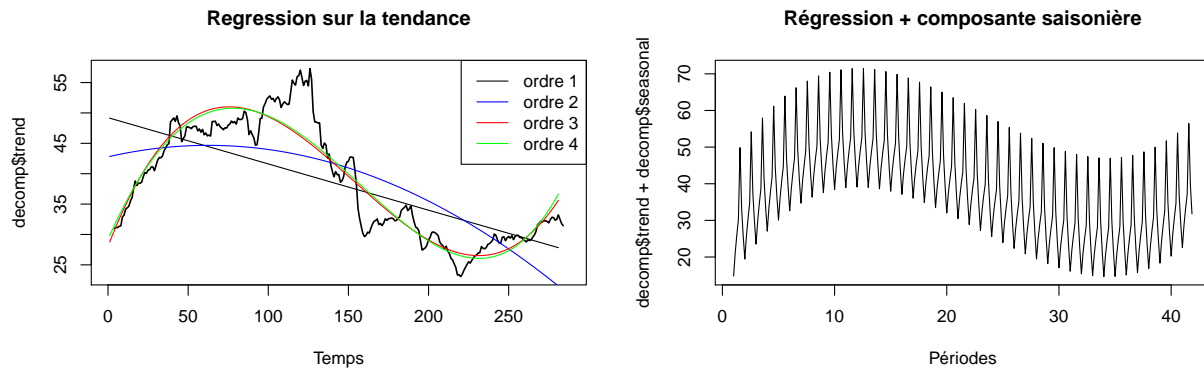
Figure 2: Analyse visuelle de la série

3 Décomposition

Afin de valider cette hypothèse de stationnarité, il est possible d'appliquer la procédure suivante. Dans un premier temps, on applique sur la série un opérateur moyenne mobile $M_m(B) = \frac{1}{2m+1} \sum_{n=1}^m B^{-k}$ qui élimine les tendances τ -périodiques avec $\tau = 2m + 1 = 7$ dans notre cas et B l'opérateur de retard tel que $B^k X_t = X_{t-k}$. On obtient alors la tendance. On soustrait cette tendance à la série. En suite, on estime la composante saisonnière en effectuant la moyenne des périodes et en dupliquant ce résultat autant de fois que nécessaire. Les résidus sont alors obtenus en soustrayant la tendance et la composante saisonnière de la série originale.

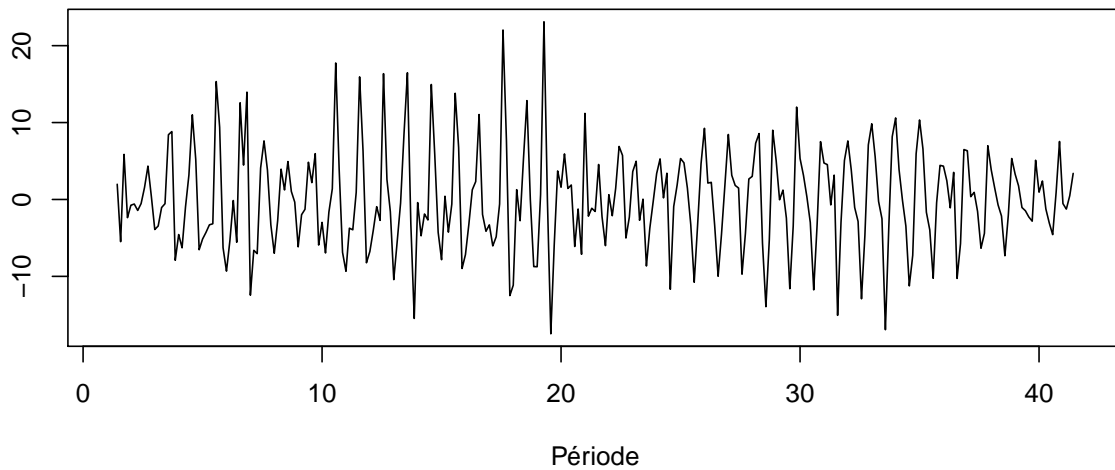


Pour pouvoir construire un modèle de prédiction sur la base de ces résultats, nous allons appliquer une régression linéaire sur la tendance puis y ajouter la composante saisonnière.



Visuellement, les régressions d'ordre 3 et 4 donnent de très bons résultats avec un R^2 ajusté de 0.88. Le principe de parcimonie nous pousse à garder la régression linéaire d'ordre 3. Analysons à présent les résidus de la décomposition.

Résidus de la décomposition



Il semble y avoir une légère saisonnalité dans les résidus.

En plus de l'aspect visuel, nous allons utiliser deux tests permettant d'évaluer l'hypothèse de stationarité. ADF teste \mathcal{H}_0 : "la trajectoire est issue d'un processus non stationnaire" contre $\mathcal{H}_1 = \bar{\mathcal{H}}_0$. KPSS teste \mathcal{H}_0 : "la trajectoire est issue d'un processus stationnaire" contre $\mathcal{H}_1 = \bar{\mathcal{H}}_0$.

Table 1: Stationnarité des résidus de la décomposition

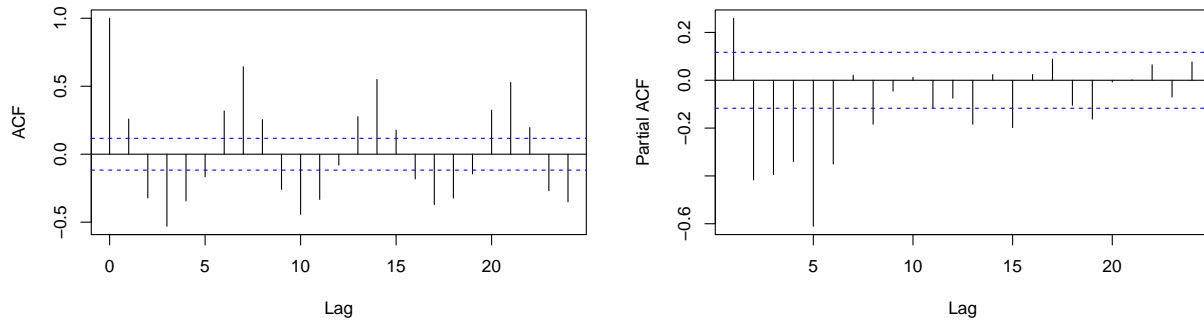
Test	Statistique	p-value
ADF	-12.369	<0.01
KPSS	0.022468	>0.1

Les tests s'accordent. Le test ADF rejette l'hypothèse de non stationnarité (p-value $\ll 0.05$) alors que le test KPSS ne rejette pas l'hypothèse de stationnarité (p-value > 0.05). Nous avons donc obtenu des résidus stationnaires d'après ces tests.

Présentons maintenant deux outils indispensables à l'étude des processus ARMA. L'ACF représente les autocorrélations entre deux valeurs distantes de h dans le temps. Une autocorrélation nulle à partir d'un

rang $q + 1$ est caractéristique d'un processus $MA(q)$. L'autocorrélogramme PAFC représente quand à lui les corrélations "pures" entre deux valeurs distantes de h dans le temps, c'est à dire entre lesquelles on a supprimé l'influence linéaire des valeurs intermédiaires. Une PACF avec des pics dans le couloir de non-significativité à partir du rang $p + 1$ est caractéristique d'un processus $AR(p)$.

De plus, on justifie une tentative de modélisation par un processus $ARMA$ lorsque la série est considérée comme stationnaire et que ses ACF et PACF empiriques montrent une décroissance rapide.



L'ACF présente des pics périodiques majoritairement significatifs avec une faible décroissance. On observe une corrélation élevée pour les lag multiples de 7. La PACF possède des corrélations partielles significatives aux lags 1 à 6. Ainsi, il est clair que nous avons encore une saisonnalité d'ordre 7.

En conclusion, les résidus du modèle de decomposition ne sont pas stationnaires et présentent une saisonnalité d'ordre 7.

4 Différenciation

Afin de rendre une série stationnaire, il est possible d'intégrer cette dernière en lui appliquant un filtre de la forme $(I - B)$. On regarde alors si $\Delta X_t = (I - B)X_t$ est stationnaire.

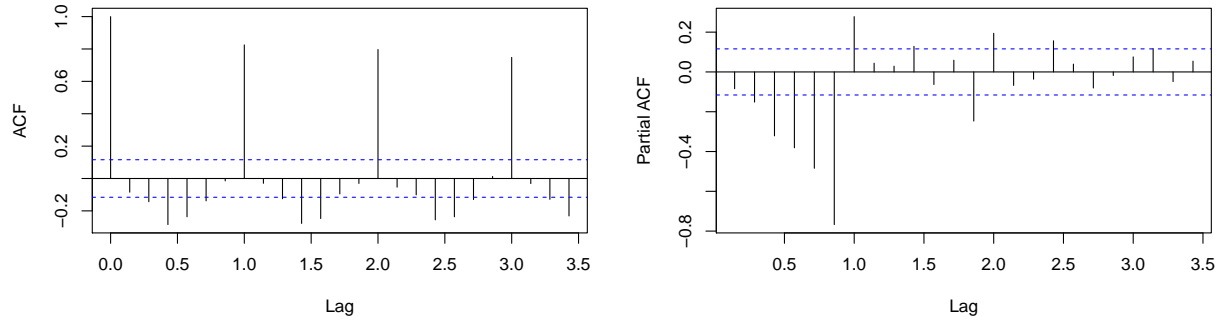
Cette méthode ne donne pas de résultat satisfaisant sur les résidus de la décomposition (cf. Annexe 1). Nous allons alors appliquer cette méthode sur la série originale.

Les tests de stationnarité nous incitent à penser que la série différenciée (ΔX_t) n'est pas stationnaire.

Table 2: Stationnarité des incréments ΔX_t

Test	Statistique	p-value
ADF	-11.875	<0.01
KPSS	0.06673	>0.1

Néanmoins, l'ACF et la PACF présentent des autocorrélations très significatives respectivement sur les lags 7, 14, 21, 28 et 1 à 7.



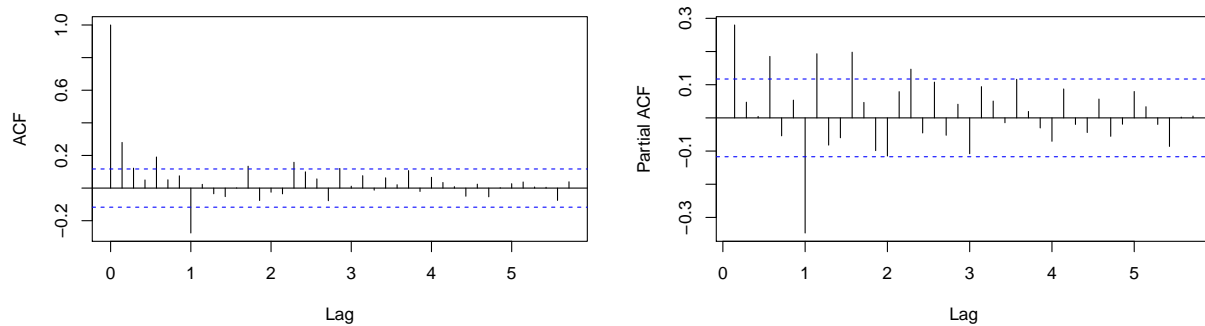
Nous en concluons que les incréments ne sont pas stationnaires car ils contiennent encore une composante saisonnière.

Tentons alors une différenciation saisonnière. Cela consiste à appliquer un filtre de la forme $(I - B^s)$ à notre série. On obtient alors les incréments saisonniers $(\Delta_s X_t = (I - B^s)X_t)$. L'application de ce filtre à la série semble tout indiquée car ce dernier élimine les tendances périodiques. Les tests ADF et KPSS s'accordent sur l'hypothèse de stationnarité.

Table 3: Stationnarité des incréments $\Delta_7 X_t$

Test	Statistique	p-value
ADF	-7.3894	<0.01
KPSS	0.32541	>0.1

L'ACF montre une décroissance rapide avec des pics significatifs en 1, 4 et 7. La PACF possède aussi une décroissance rapide et des pics significatifs en 1 et 7. On pourra alors tenter une modélisation *ARMA* sur la série $Y_t = \Delta_7 X_t = (I - B^7)X_t$.



Essayons de différencier à nouveau la série, c'est à dire d'appliquer le filtre $(I - B)(I - B^7)$ sur la série de départ.

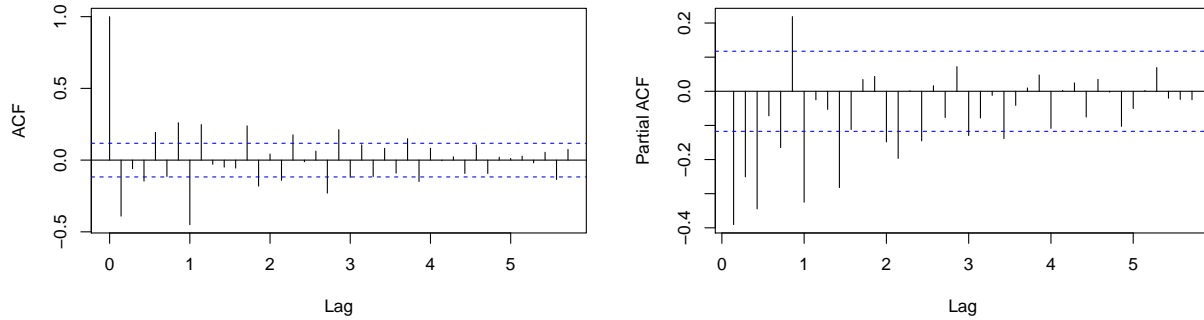
Les tests vont encore dans le sens de la stationnarité.

Table 4: Stationnarité des incréments $\Delta\Delta_7 X_t$

Test	Statistique	p-value
ADF	-9.6985	<0.01

Test	Statistique	p-value
KPSS	0.011688	>0.1

L'ACF et la PACF donnent de moins bons résultats dans le sens où l'ACF présente des résultats similaires mais la PACF décroît beaucoup moins vite et présente d'avantage de pics significatifs.



Conclusion, la série différenciée ($\Delta_7 X_t$) remplit le critère de stationnarité. Il est donc possible d'appliquer un modèle de type $SARIMA(p, d, q)(P, D, Q)_7$ avec $(d, D) = (0, 1)$.

5 Modélisation

Nous avons vu dans la partie précédente qu'une différenciation d'ordre 1 avec le filtre $(I - B^7)$ donne une série stationnaire. Cela motive l'idée d'appliquer un modèle $SARIMA(p, 0, q)(Q, 1, P)_7$. Nous déterminons les coefficients p, q, Q et P avec la méthode `auto.arima` (voir Annexe 5). Il est important d'effectuer la recherche du meilleur modèle avec les paramètres `allowdrift=TRUE` et `include.mean=TRUE` car la différenciation effectuée élimine totalement les tendances d'ordre 1. Le résultat de la recherche est présenté ci-dessous:

```
## Series: trainset
## ARIMA(3,0,4)(1,1,1)[7] with drift
##
## Coefficients:
##          ar1      ar2      ar3      ma1      ma2      ma3      ma4      sar1
##        -0.3957  0.7412  0.5702  0.8037 -0.4320 -0.7516 -0.0824  0.0974
## s.e.      0.2310  0.0760  0.2048  0.2420  0.1316  0.1792  0.1273  0.1018
##          sma1    drift
##        -0.6390  0.0206
## s.e.      0.0788  0.1125
##
## sigma^2 estimated as 28.58:  log likelihood=-863.39
## AIC=1748.78   AICc=1749.76   BIC=1788.76
```

Pour estimer la pertinence du modèle et en particulier la significativité d'un coefficient, on divise la valeur de ce derniers par son écart type. Le quotient doit se trouver hors de l'intervalle $[-1.96, 1.96]$ pour que le paramètre soit considéré comme significatif. Les coefficients `ma4`, `sar1` et `drift` ne sont pas significatifs. Testons donc le modèle $SARIMA(3, 0, 3)(0, 1, 1)_7$ sans tendance linéaire.

```
## Series: trainset
## ARIMA(3,0,3)(0,1,1)[7] with drift
##
## Coefficients:
##          ar1      ar2      ar3      ma1      ma2      ma3      sma1    drift
##        -0.3957  0.7412  0.5702  0.8037 -0.4320 -0.7516 -0.0824  0.0974
```



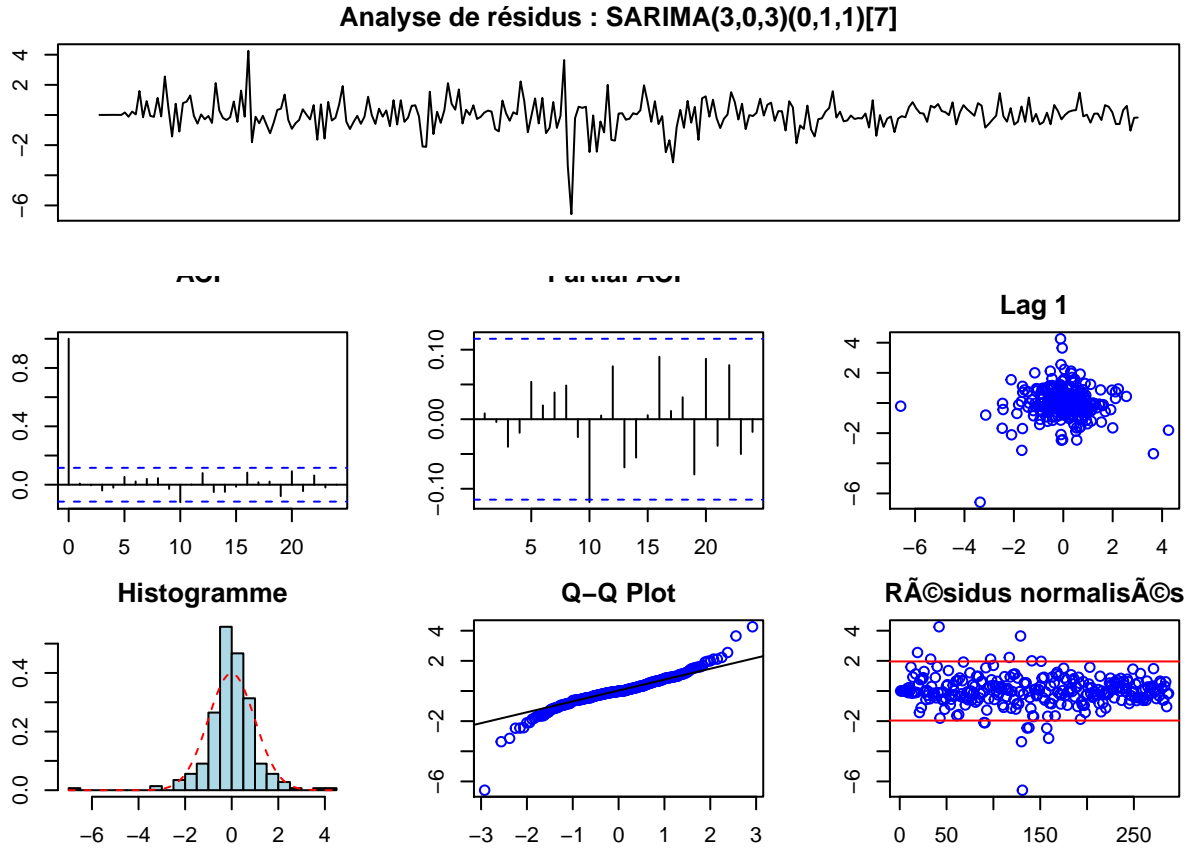
```
##          -0.2243  0.7073  0.4280  0.6240  -0.4796  -0.6160  -0.5851  0.0176
## s.e.      0.1738  0.0783  0.1367  0.1521   0.0953   0.0891   0.0631  0.1086
##
## sigma^2 estimated as 28.51:  log likelihood=-864.06
## AIC=1746.12  AICc=1746.78  BIC=1778.83
```

Les coefficients sont tous significatifs. Le score BIC obtenu pour ce modèle est meilleur que pour le modèle précédent avec 1778.83 contre 1788.76.

Pour considérer le modèle comme satisfaisant, les résidus doivent former si possible un bruit blanc. Autrement dit, les résidus réduits doivent être stationnaires, centrés, indépendants et suivre une loi normale. Pour ce faire nous utilisons différents outils statistiques tels que les tests ADF et KPSS présentés précédemment ainsi que les tests de Ljung-Box d'hypothèse nulle \mathcal{H}_0 : "La série temporelle ne possède pas d'autocorrélation." ou de Shapiro d'hypothèse nulle \mathcal{H}_0 : "L'échantillon suit une loi normale."

Table 5: Tests sur les résidus du modèle $SARIMA(3, 0, 3)(0, 1, 1)_7$.

Test	Statistique	p-value
ADF	-5.7081	<0.01
KPSS	0.18528	>0.1
Ljung-Box	1.9551	0.9623
Shapiro	0.9132	1.671e-11



Les résidus de ce modèle 1 semblent tout à fait correspondre à un bruit blanc. En effet, l'ACF et la PACF n'ont aucun pic hors du couloir de non-significativité, les tests ADF et KPSS estiment les résidus stationnaires et le test de Ljung-Box avec une p-value relativement élevée de 0.9628 ne rejette pas l'hypothèse d'indépendance.

Néanmoins, leur distribution est légèrement trop concentrée pour coller parfaitement à une distribution normale, comme peut en témoigner la p-value du test de Shapiro.

Ce modèle est assez satisfaisant. D'autres modèles ont été évalués en Annexe 2 et 3, notamment en appliquant des transformations de Box-Cox et logarithmique. L'annexe 4 contient les résultats (non concluants) des modélisations exploitant le paramètre `xreg` de la fonction `auto.arima`. Les modèles retenus sont les suivants :

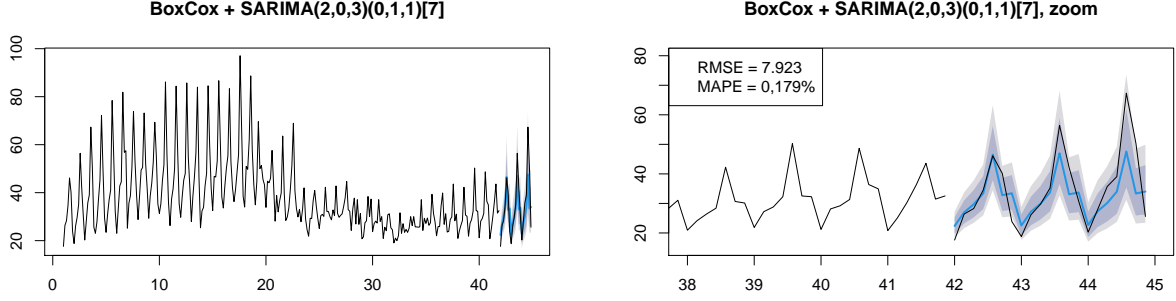
Table 6: Comparaison des modèles

Modèle	Transformation	BIC
SARIMA(5,0,0)(0,1,1)[7]	Décompositon	
SARIMA(3,0,3)(0,1,1)[7]	Aucune	1778.83
SARIMA(2,0,3)(0,1,1)[7]	Log	-344.91
SARIMA(2,0,3)(0,1,1)[7]	Box-Cox	-1250.18
Holt-Winters	Aucune	-1185.07
ETS	STL	

6 Prédiction

Précédemment, la comparaison des modèles était basée sur des critères intégrant la notion de parcimonie avec le score BIC et l'étude des coefficients significatifs. Nous portons aussi une grande importance à ce que les résidus du modèle forment un bruit blanc gaussien.

Il est temps de comparer nos modèles avec des critères prédictifs. La procédure d'évaluation consiste en une validation simple sur le jeu d'entraînement. Les scores utilisés pour comparer les modèles sont la RMSE (Root Mean Square Error) et la MAPE (Mean Absolute Percentage Error).



On obtient de bons résultats avec une erreur MAPE de seulement 0.179% et une RMSE de 7,923 sur le jeu test.

7 Conclusion

Grâce aux des études de la série réalisées dans les parties 1 et 2 nous avons pu construire un modèle $SARIMA(2, 0, 3)(0, 1, 1)_7$ ayant obtenu un score RMSE de 7,923 sur le jeu de données test. Ce modèle a été appliquée à la série corrigée par la transformation de Box-Cox $B(x, \lambda) = \frac{x^\lambda - 1}{\lambda}$ avec un $\lambda = -1$.

L'écriture explicite de ce modèle est $(I - \phi_1 B - \phi_2 B^2)Y_t = (I + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)(I + \beta_1 B^7)\epsilon_t$ avec $Y_t = (I - B^7)X_t$.

On réécrit cela sous la forme :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \theta_3 \epsilon_{t-3} + \beta_1 (\epsilon_{t-7} + \theta_1 \epsilon_{t-8} + \theta_2 \epsilon_{t-9} + \theta_3 \epsilon_{t-10}).$$

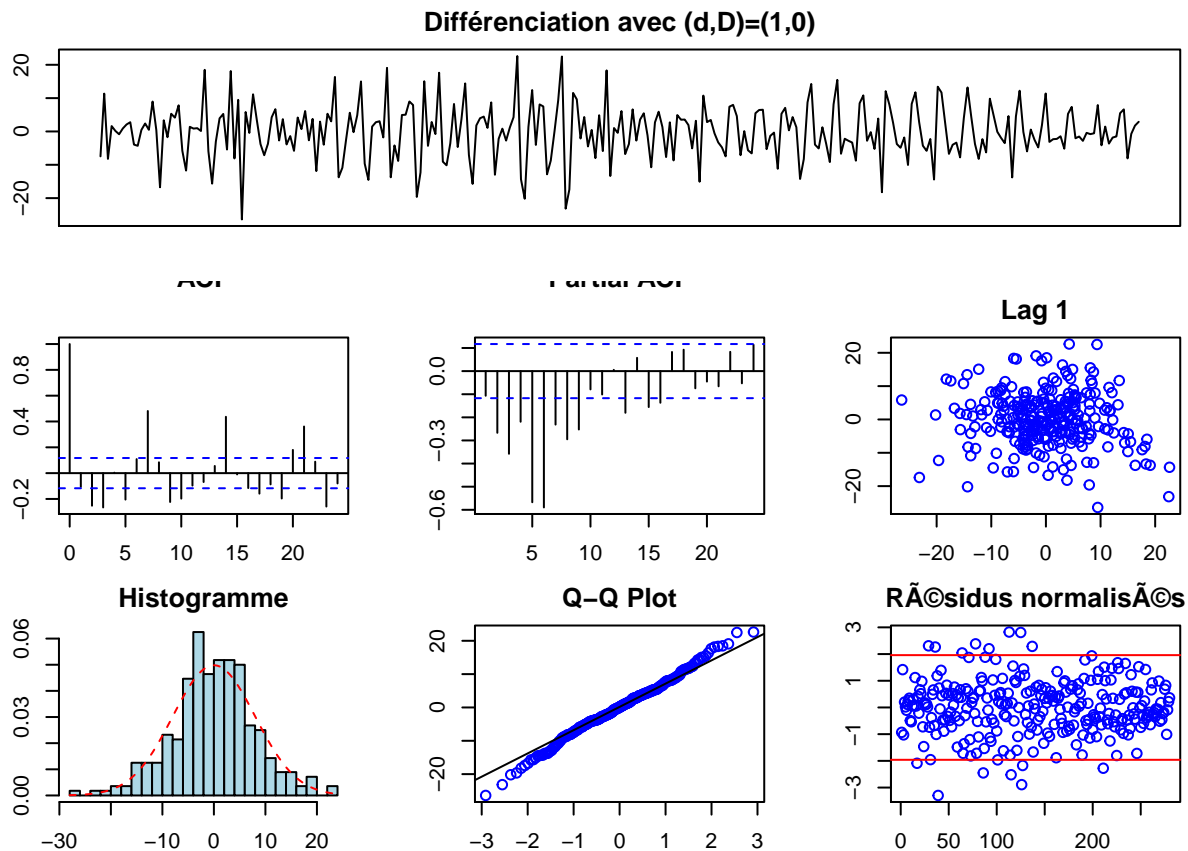
Ce modèle pourrait être challengé par une approche basée sur des techniques d'apprentissage automatique (par exemple en entraînant une forêt aléatoire sur les lags de la série) ou des méthodes d'apprentissage profond (cf. réseaux de neurones LSTM).

8 Annexe 1 : Différenciation

L'objectif de cette section est de rendre les résidus (X_t) de la décomposition simple stationnaires.

Analyse :

La série $\Delta X_t = (I - B)X_t$ est stationnaire pour les tests ADF et KPSS mais on observe des pics significatifs aux lags 7, 14 et 21 sur l'ACF ainsi que des pics significatifs jusqu'aux lags 9 de la PACF. La distribution semble suivre une loi normale.

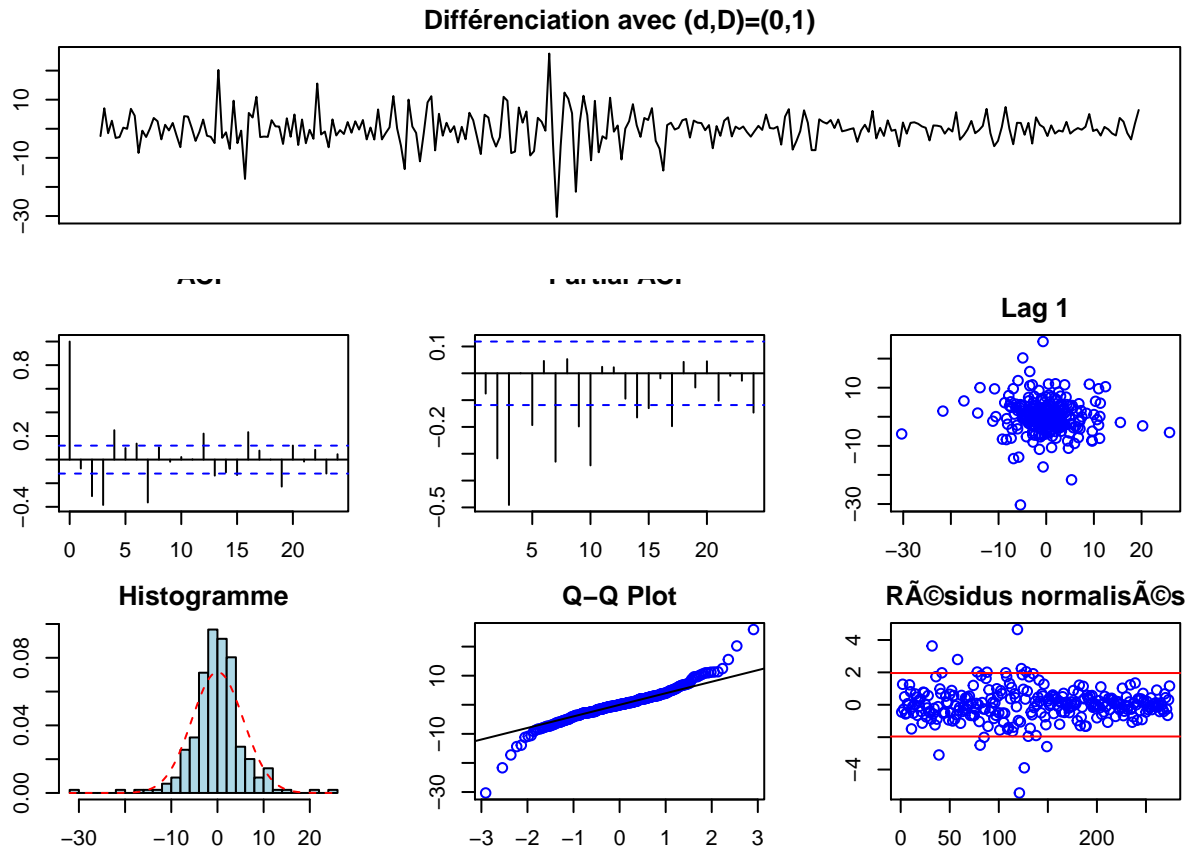


```
##
## Augmented Dickey-Fuller Test
##
## data: res
## Dickey-Fuller = -17.419, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
##
## KPSS Test for Level Stationarity
##
## data: res
## KPSS Level = 0.018332, Truncation lag parameter = 5, p-value = 0.1
##
## Box-Ljung test
##
## data: res
## X-squared = 123.94, df = 7, p-value < 2.2e-16
```

```
##
##
## Shapiro-Wilk normality test
##
## data:  res
## W = 0.99412, p-value = 0.3505
```

Analyse :

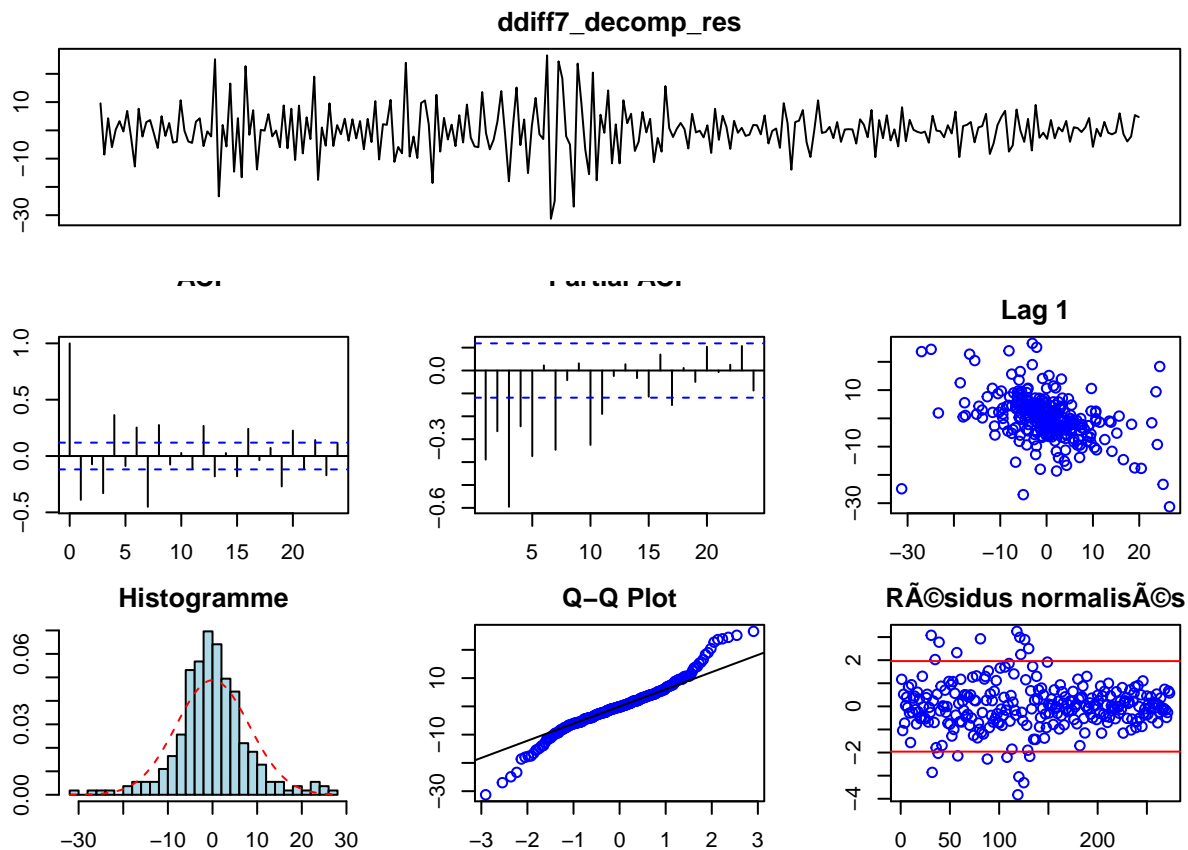
L'ACF contient des pics significatifs aux lags 2, 3 et 7 et la PACF aux lags 2, 3, 5, 7, 9 et 10. La distribution ne semble pas normale car trop centrée.



```
##
## Augmented Dickey-Fuller Test
##
## data:  res
## Dickey-Fuller = -11.311, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
##
##
## KPSS Test for Level Stationarity
##
## data:  res
## KPSS Level = 0.014093, Truncation lag parameter = 5, p-value = 0.1
##
##
## Box-Ljung test
##
```

```
## data: res
## X-squared = 132.05, df = 7, p-value < 2.2e-16
##
##
## Shapiro-Wilk normality test
##
## data: res
## W = 0.92686, p-value = 2.327e-10
```

L'ACF contient des pics légèrement hors du couloir de significativité aux lags 1, 3, 4, 6 et 8 notamment. La PACF obtient des pics significatifs aux lags 1 à 5 et aux lag 10. La distribution ne semble pas normale car trop centrée.



```
##
## Augmented Dickey-Fuller Test
##
## data: res
## Dickey-Fuller = -13.973, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
##
##
## KPSS Test for Level Stationarity
##
## data: res
## KPSS Level = 0.01681, Truncation lag parameter = 5, p-value = 0.1
##
##
```

```
## Box-Ljung test
##
## data: res
## X-squared = 187.89, df = 7, p-value < 2.2e-16
##
##
## Shapiro-Wilk normality test
##
## data: res
## W = 0.95942, p-value = 6.327e-07
```

Malgré ces résultats peu concluant, essayons d'appliquer un modèle sur ces résidus.

```
auto.arima(decomp_res)
```

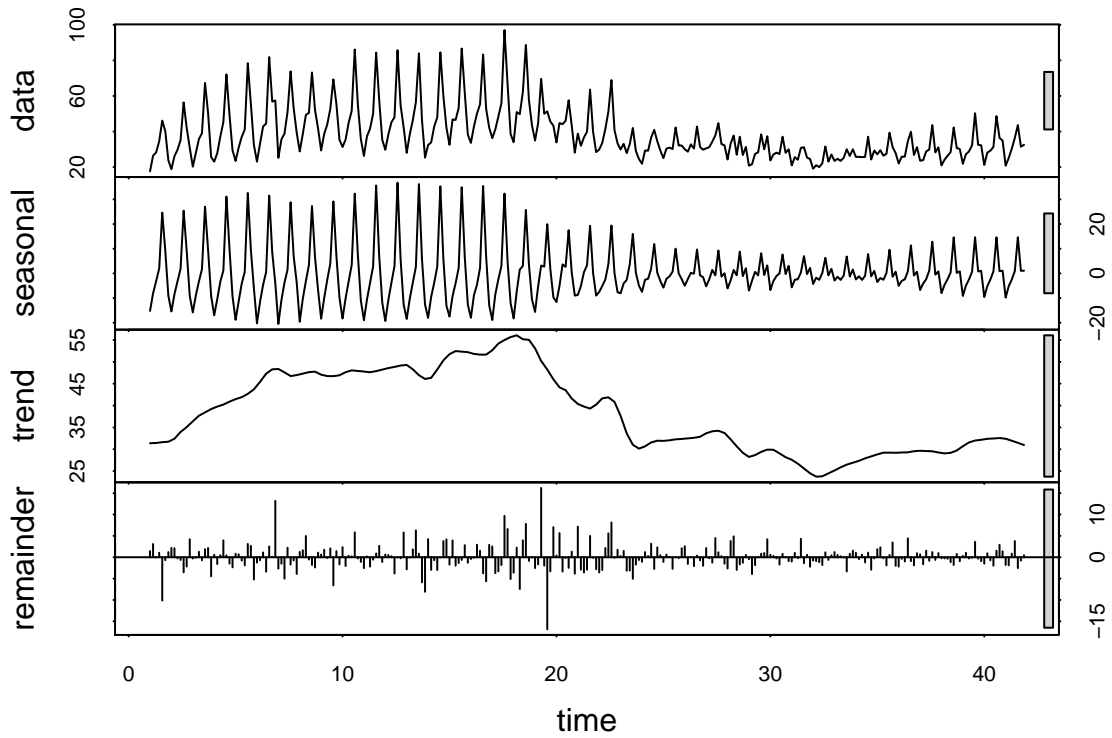
```
## Series: decomp_res
## ARIMA(5,0,0)(0,1,1)[7]
##
## Coefficients:
##          ar1      ar2      ar3      ar4      ar5      sma1
##      -0.1801  -0.5311  -0.5806  -0.1842  -0.4193  -0.7169
## s.e.   0.0554   0.0557   0.0531   0.0577   0.0624   0.0525
##
## sigma^2 estimated as 14.55: log likelihood=-755.41
## AIC=1524.82  AICc=1525.24  BIC=1550.11
```

Tous les coefficients sont significatifs. Le fait d'obtenir $(d, D) = (0, 1)$ est cohérent avec l'analyse car la différenciation saisonnière est la transformation donnant les ACF et PACF les plus convaincantes.

9 Annexe 2 : Décomposition STL

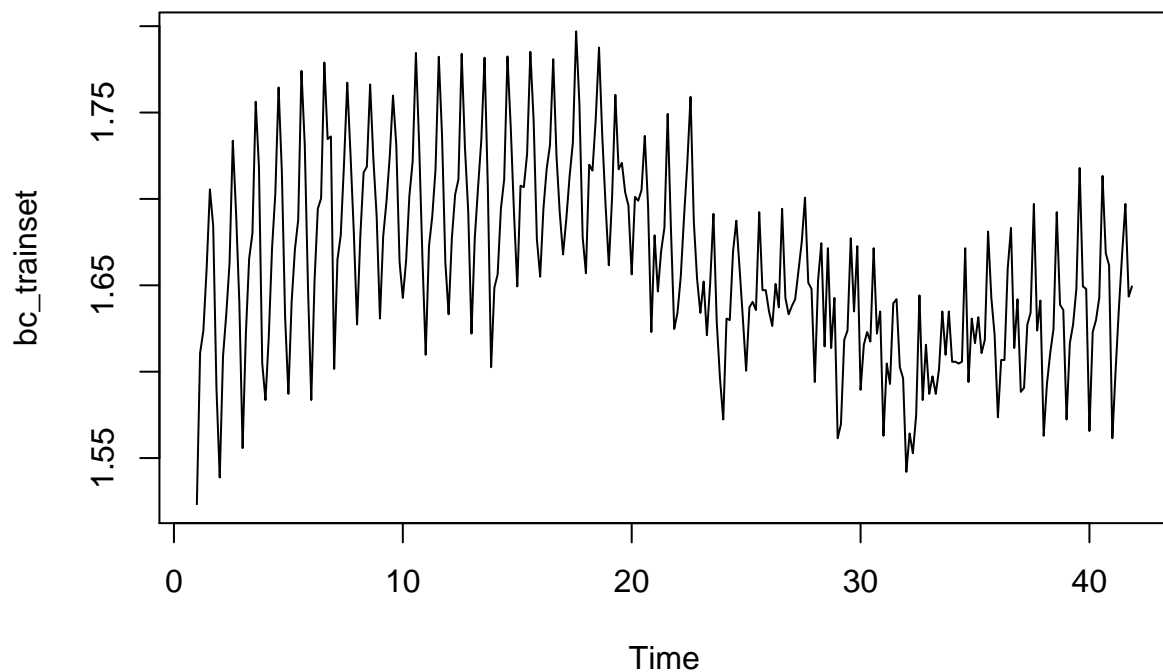
Une présentation de cette méthode est disponible au lien suivant : <https://otexts.com/fpp2/stl.html>.

```
decomposition_stl = stl(trainset, s.window=7)
plot(decomposition_stl)
```



10 Annexe 3 : Transformation Box-Cox

Dans cette annexe, nous détaillons de façon succincte notre procédure de choix de modèles pour une transformation BoxCox des données. Nous utilisons la fonction `BoxCox.lambda` pour déterminer le meilleur paramètre de transformation BoxCox. Nous utilisons ici cette fonction avec la méthode de maximisation de la log-vraisemblance.



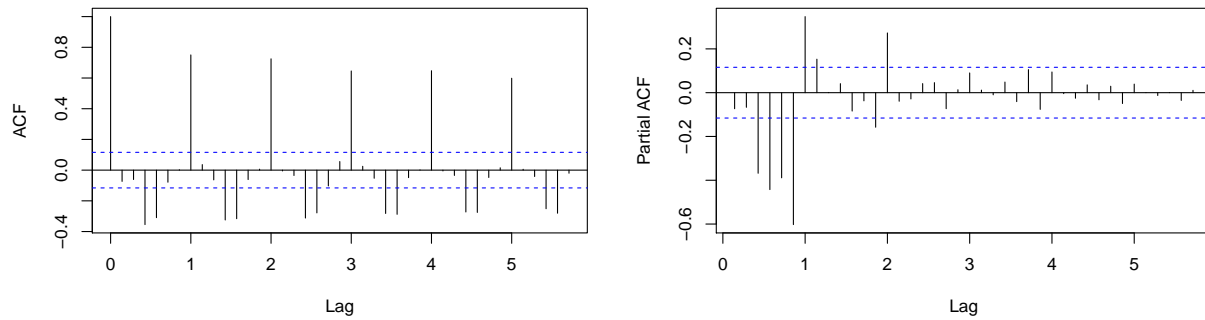
Une telle transformation n'est pas satisfaisante, on va tenter plusieurs différenciations jusqu'à ce que la série soit stationnaire.

Ainsi, sur les différents incréments de cette série, les tests KPSS et ADF donnent :

Filtre	Test	Statistique	p-value
$(I - B)$	ADF	-11.621	<0.01
$(I - B)$	KPSS	0.10789	>0.1
$(I - B^7)$	ADF	-8.6009	<0.01
$(I - B^7)$	KPSS	0.27522	>0.1
$(I - B^7)(I - B)$	ADF	-10.188	<0.01
$(I - B^7)(I - B)$	KPSS	0.011297	>0.1

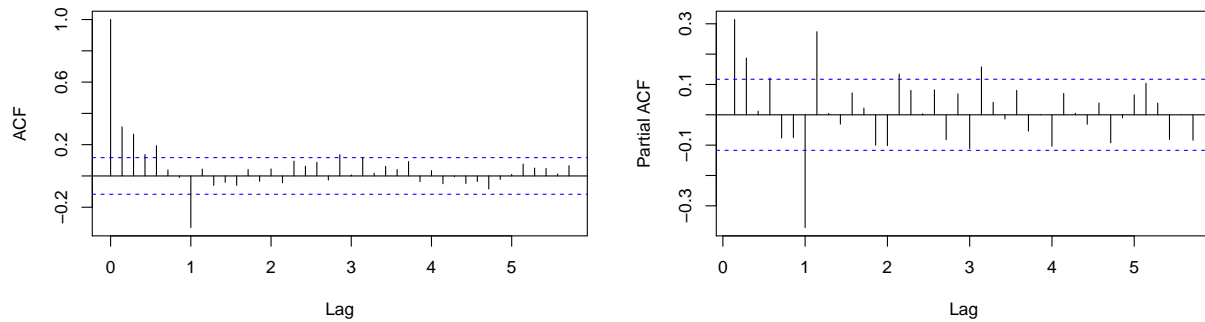
Ces tests vont tous dans le sens de la stationnarité. Regardons à présents les ACF et les PACF.

Filtre $(I - B)$:



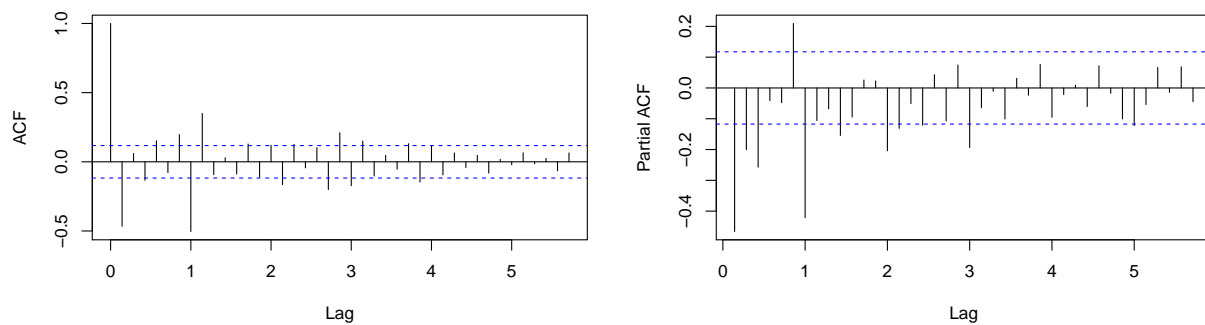
Un motif périodique se dessine clairement sur l'ACF, cette différenciation n'est satisfaisante.

Filtre $(I - B^7)$:



Le résultat est bien meilleur. Il y a un pics significatif au lag 7 sur l'ACF et aux lags 1, 2, 7 et 8 sur la PACF.

Filtre $(I - B)(I - B^7)$:



Des pics significatif sont présents aux lags 1,7 et 8 sur l'ACF et aux lags 1, 2, 3, 6, 7, 21 et 28 sur la PACF.

En conclusion, il est judicieux de considérer les incréments correspondant à la différenciations $I - B^7$ donnant les meilleurs résultats pour les tests ADF et KPSS ainsi que pour l'analyse des ACF et PACF. On notera aussi que la tendance et la saisonnalité sont éliminées sur cette série.

On utilise la fonction `auto.arima` pour déterminer le modèle SARIMA expliquant le mieux nos données.

`## Series: trainset`

```
## ARIMA(2,0,3)(0,1,1)[7]
## Box Cox transformation: lambda= -0.45
##
## Coefficients:
##          ar1      ar2      ma1      ma2      ma3      sma1
##        -0.0043  0.9378  0.3667 -0.685 -0.1494 -0.6869
## s.e.    0.0411  0.0391  0.0717  0.074  0.0617  0.0504
##
## sigma^2 estimated as 0.0005885:  log likelihood=644.81
## AIC=-1275.62  AICc=-1275.21  BIC=-1250.18
```

Tous les coefficients sont significatifs, on choisit de garder ce modèle.

En suivant une procédure similaire, il est possible de construire un modèle avec la transformation logarithmique.

```
auto.arima(trainset, d=0, D=1, max.P=1, max.Q=1, allowmean=TRUE, allowdrift=TRUE, lambda=0, biasadj=TRUE)
```

```
## Series: trainset
## ARIMA(2,0,3)(0,1,1)[7]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##          ar1      ar2      ma1      ma2      ma3      sma1
##         0.0034  0.9409  0.3736 -0.7083 -0.1875 -0.6856
## s.e.    0.0409  0.0390  0.0710  0.0678  0.0594  0.0492
##
## sigma^2 estimated as 0.01492:  log likelihood=192.18
## AIC=-370.35  AICc=-369.94  BIC=-344.91
```

11 Annexe 4 : xreg

Cette section a pour but de tester l'ajout d'une régression au sein d'un modèle *SARIMA*.

```
## Series: trainset
## Regression with ARIMA(3,0,3)(0,1,1)[7] errors
##
## Coefficients:
##          ar1      ar2      ar3      ma1      ma2      ma3      sma1      t
##        -0.3033  0.6101  0.3961  0.6749 -0.3983 -0.5746 -0.5914  0.5882
## s.e.    0.2127  0.1109  0.1298  0.1964  0.1637  0.0978  0.0648  0.1325
##          t2  t3
##        -0.0050  0
## s.e.    0.0012  0
##
## sigma^2 estimated as 27.9:  log likelihood=-860.05
## AIC=1742.1  AICc=1743.09  BIC=1782.09

## Series: trainset
## Regression with ARIMA(3,0,3)(0,1,1)[7] errors
##
## Coefficients:
##          ar1      ar2      ar3      ma1      ma2      ma3      sma1      t
##        -0.2312  0.7013  0.4275  0.6280 -0.4756 -0.6147 -0.5838  0.1638
## s.e.    0.1762  0.0786  0.1363  0.1546  0.0982  0.0892  0.0636  0.1864
##          t2
##        -5e-04
```

```

## s.e.    5e-04
##
## sigma^2 estimated as 28.53:  log likelihood=-863.64
## AIC=1747.27  AICc=1748.09  BIC=1783.62

## Series: trainset
## Regression with ARIMA(3,0,3)(0,1,1)[7] errors
##
## Coefficients:
##          ar1      ar2      ar3      ma1      ma2      ma3      sma1      t
##          -0.2243  0.7073  0.4280  0.6240 -0.4796 -0.6160 -0.5851  0.0176
## s.e.      0.1738  0.0783  0.1367  0.1521  0.0953  0.0891  0.0631  0.1086
##
## sigma^2 estimated as 28.51:  log likelihood=-864.06
## AIC=1746.12  AICc=1746.78  BIC=1778.83

## Series: trainset
## Regression with ARIMA(1,0,0)(2,0,0)[7] errors
##
## Coefficients:
##          ar1      sar1      sar2  intercept      t      t2  t3
##          0.3198  0.5486  0.3405    28.5148  0.6114 -0.0051  0
## s.e.      0.0567  0.0554  0.0560     5.9332  0.1483  0.0011  0
##
## sigma^2 estimated as 30.85:  log likelihood=-900.81
## AIC=1817.62  AICc=1818.14  BIC=1846.89

## Series: trainset
## Regression with ARIMA(1,0,0)(2,0,0)[7] errors
##
## Coefficients:
##          ar1      sar1      sar2  intercept      t      t2
##          0.3769  0.5667  0.3322    33.4567  0.1345 -5e-04
## s.e.      0.0562  0.0565  0.0568     6.8957  0.0939  3e-04
##
## sigma^2 estimated as 32.08:  log likelihood=-907.24
## AIC=1828.49  AICc=1828.89  BIC=1854.1

## Series: trainset
## Regression with ARIMA(1,0,0)(2,0,0)[7] errors
##
## Coefficients:
##          ar1      sar1      sar2  intercept      t
##          0.3899  0.5689  0.3333    38.2762 -0.0170
## s.e.      0.0561  0.0567  0.0569     6.6452  0.0341
##
## sigma^2 estimated as 32.27:  log likelihood=-908.72
## AIC=1829.44  AICc=1829.74  BIC=1851.4

```

On conclut que les coefficients de régression ne sont pas significatifs pour les régressions d'ordre 1, 2 et 3. On ne retient donc pas cette option.

12 Annexe 5 : auto.arima

How does auto.arima() work?