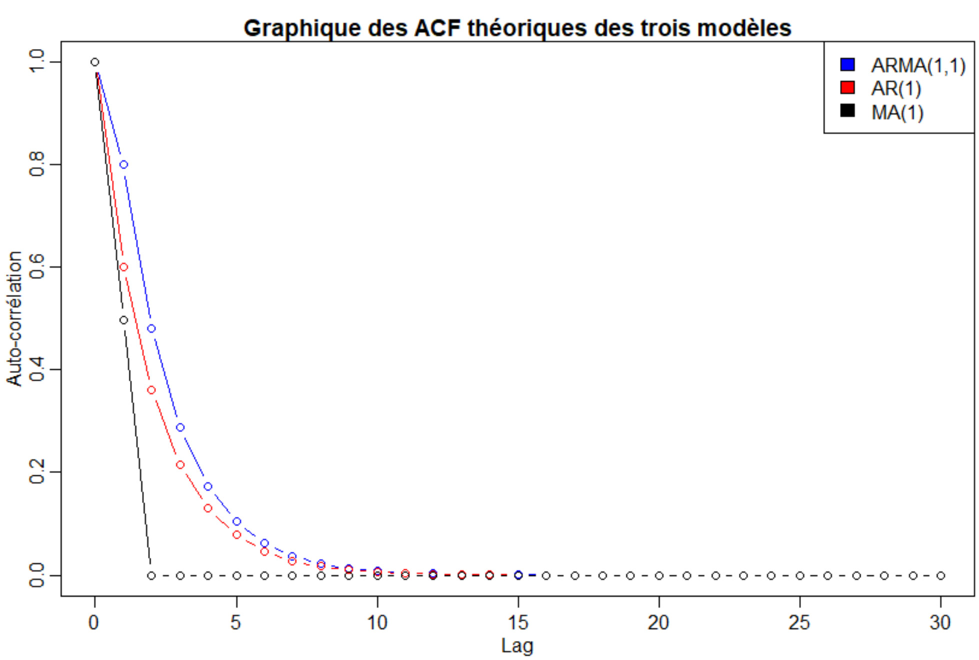
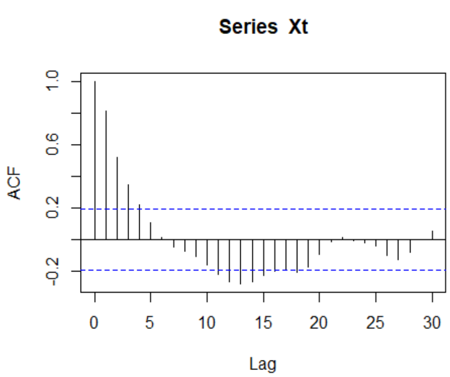
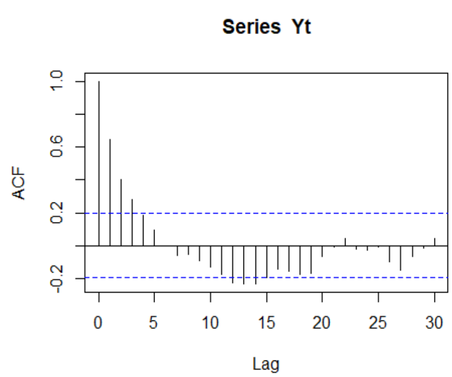
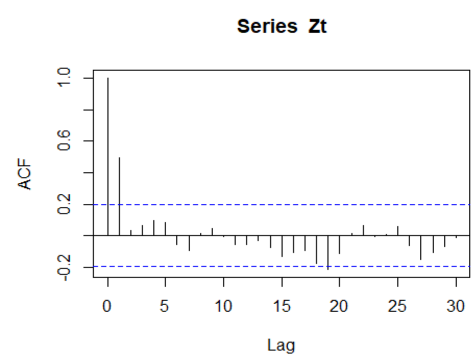
```
# PROJET FINAL
# Auteur: Ronnie Liu
# Date: Mercredi 27 juillet 2021
# Objet: Fichier R pour QUESTIONS 3 et 4
n <- 100
phi <- 0.6
theta <- 0.9

################################################################################
# a) Fonctions pour calculer les fonctions d'auto-corrélation

# Voir la démarche théorique de l'obtention des fonctions d'auto-corrélation
# pour chaque processus (p.13).

rho_1 <- ((phi+theta)*(1+phi*theta))/(1+2*phi*theta+theta^2)
# Pour le modèle ARMA(1,1):
#       | = 1                 si k = 0
# rho_k|
#       | = phi^(k-1)*rho_1 sinon

# Pour le modèle AR(1):
#       | = 1         si k = 0
# rho_k|
#       | = phi^|k|   sinon

# Pour le modèle MA(1):
#       | = 1                 si k = 0
# rho_k| = theta/(1+theta^2) si k = +/- 1
#       | = 0                 sinon

################################################################################
# b) Traçage des ACF des trois séries
```



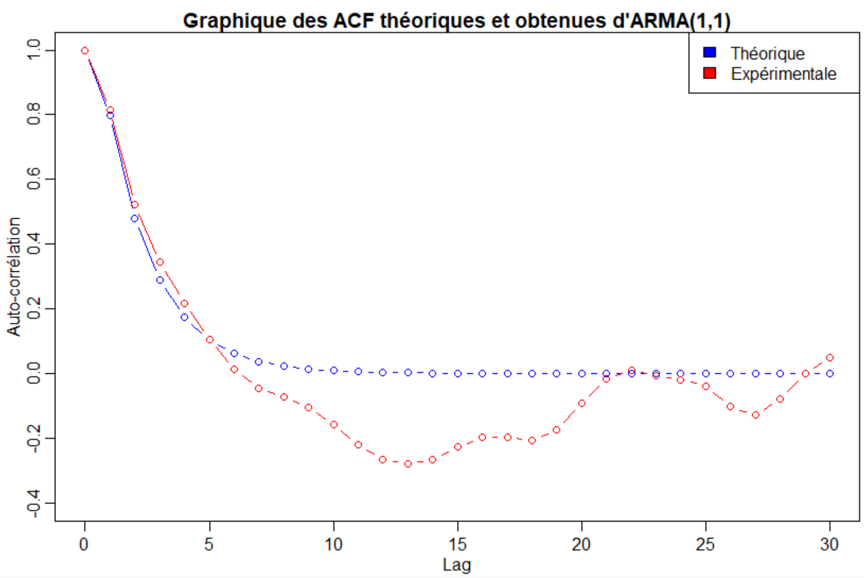
```
# Capacités de diagnostic pour chaque processus:
# Tous les processus sont inversibles et stationnaires dans cette question
# (voir 3a).

# Processus Xt: ARMA(1,1):
# 1. ACF: Décroissance exponentielle moins rapide que celle d'un modèle AR(1).
# 2. PACF: Décroissance exponentielle aussi, donc nombre infini de pics.
# Bref, détecter s'il y a un nombre infini de pics dans les deux graphiques
# d'auto-corrélation.

# Processus Yt: ARMA(1,0) ou AR(1):
# 1. ACF: Décroissance exponentielle plus rapide que celle d'un ARMA(1,1).
# 2. PACF: Présence d'un seul pic significatif.
# Bref, ACF avec nombre infini de pics qui décroît et PACF avec un seul pic.

# Processus Zt: ARMA(0,1) ou MA(1):
# 1. ACF: Présence d'Un seul pic significatif. On voit qu'après le 2e délai,
# la valeur d'auto-corrélation devient nulle.
# 2. PACF: Décroissance exponentielle quant à la longueur des pics.
# Bref, ACF avec un seul pic et PACF avec nombre infini de pics qui décroît.

# En conclusion, si on se fie uniquement des graphiques d'ACF:
# ARMA(1,1): décroissance exponentielle
# AR(1): décroissance exponentielle plus rapide qu'ARMA(1,1)
# MA(1): valeurs non nulles dans les délais k = {0,1} seulement.

################################################################################
# c) ACF Échantillonnale des séries générées et comparaison aux valeurs théo.
```



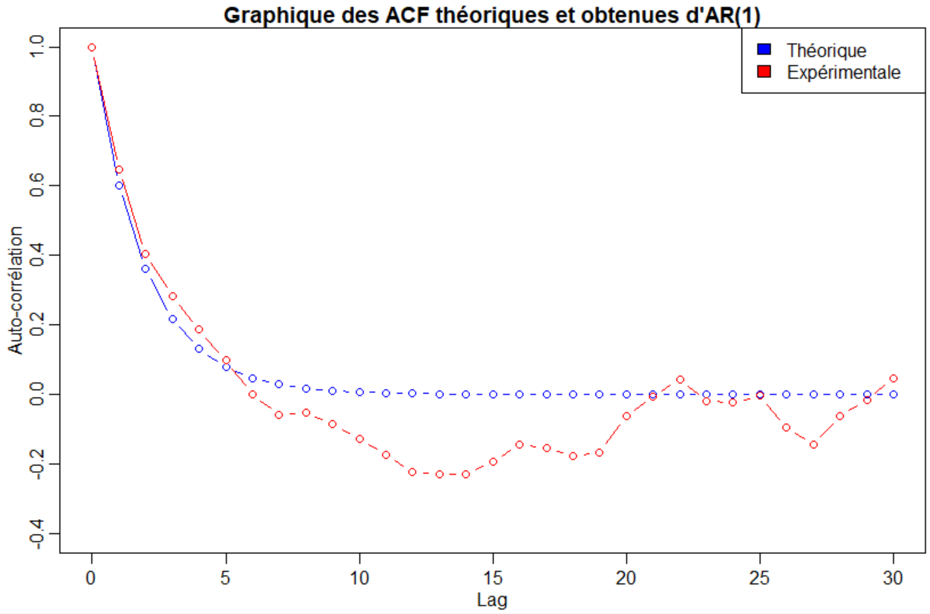




```
## [1] "RÉSULTATS POUR ARMA(1,1)"
EXPÉRIMENTALE
##
## Autocorrelations of series 'Xt', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000  0.814  0.522  0.343  0.218  0.106  0.011 -0.046 -0.072 -0.105 -0.159
##     11     12     13     14     15     16     17     18     19     20     21
## -0.219 -0.266 -0.280 -0.266 -0.228 -0.198 -0.197 -0.207 -0.174 -0.092 -0.017
##     22     23     24     25     26     27     28     29     30
##  0.011 -0.008 -0.019 -0.039 -0.104 -0.130 -0.078 -0.001  0.050

THÉORIQUE
##  [1] 1.000 0.799 0.480 0.288 0.173 0.104 0.062 0.037 0.022 0.013 0.008 0.005
## [13] 0.003 0.002 0.001 0.001 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [25] 0.000 0.000 0.000 0.000 0.000 0.000 0.000
```




```
## [1] "RÉSULTATS POUR AR(1)"
EXPÉRIMENTALE
##
## Autocorrelations of series 'Yt', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000  0.646  0.404  0.281  0.186  0.097  0.001 -0.060 -0.051 -0.087 -0.128
##     11     12     13     14     15     16     17     18     19     20     21
## -0.174 -0.224 -0.229 -0.231 -0.194 -0.144 -0.155 -0.177 -0.168 -0.064 -0.006
##     22     23     24     25     26     27     28     29     30
##  0.042 -0.019 -0.023 -0.004 -0.095 -0.146 -0.062 -0.015  0.046

THÉORIQUE
##  [1] 1.000 0.600 0.360 0.216 0.130 0.078 0.047 0.028 0.017 0.010 0.006 0.004
## [13] 0.002 0.001 0.001 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [25] 0.000 0.000 0.000 0.000 0.000 0.000 0.000
```



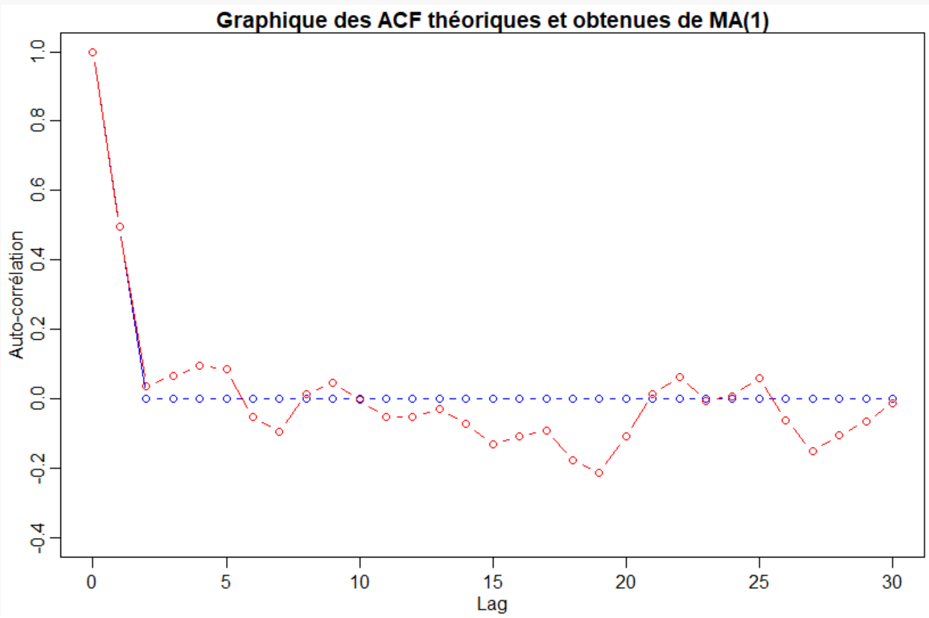

```
## [1] "RÉSULTATS POUR MA(1)"
```

EXPÉRIMENTALE

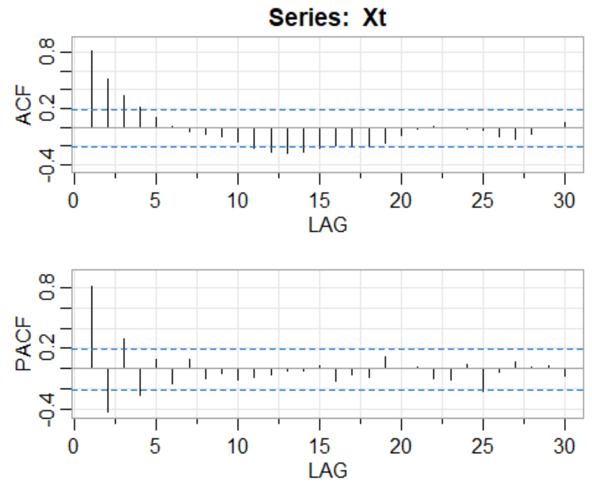
```
##
## Autocorrelations of series 'Zt', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000  0.495  0.035  0.066  0.094  0.085 -0.053 -0.094  0.013  0.046 -0.005
##     11     12     13     14     15     16     17     18     19     20     21
## -0.054 -0.054 -0.029 -0.073 -0.131 -0.108 -0.092 -0.177 -0.213 -0.110  0.013
##     22     23     24     25     26     27     28     29     30
##  0.063 -0.006  0.008  0.058 -0.061 -0.151 -0.105 -0.066 -0.014
```

THÉORIQUE

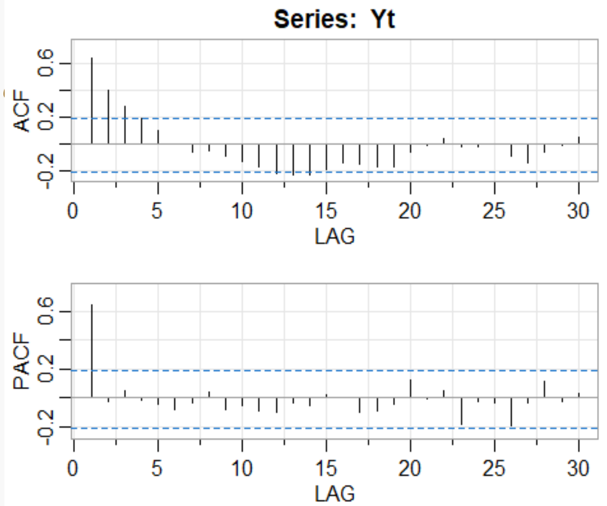
```
##  [1] 1.000 0.497 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [13] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [25] 0.000 0.000 0.000 0.000 0.000 0.000 0.000
```



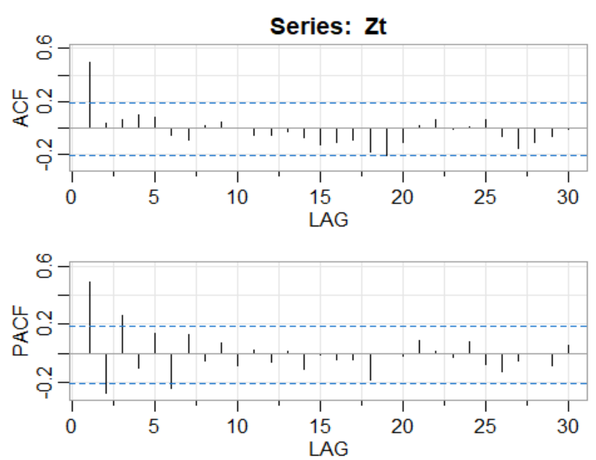
```r
# Les calculs théoriques représentent quand même les données expérimentales
# à l'exception des derniers délais. Cela est causé par le fait qu'il n'y a pas
# assez de données. Si par exemple n = 10000, on verra que les courbes
# théoriques et expérimentales sont exactement identiques.

################################################################################
# d) ACF, PACF, comparaisons et conclusion

# Modèle ARMA(1,1)
library(astsa)
acf2(Xt, max.lag=30)
# ACF: Nombre infini d'ACF
# PACF: Nombre infini de PACF
```




```r
# Modèle AR(1)
acf2(Yt, max.lag=30)
# ACF: Nombre infini d'ACF (décroiss
qu'ARMA(1,1))
# PACF: Nombre fini (un seul pic)
```

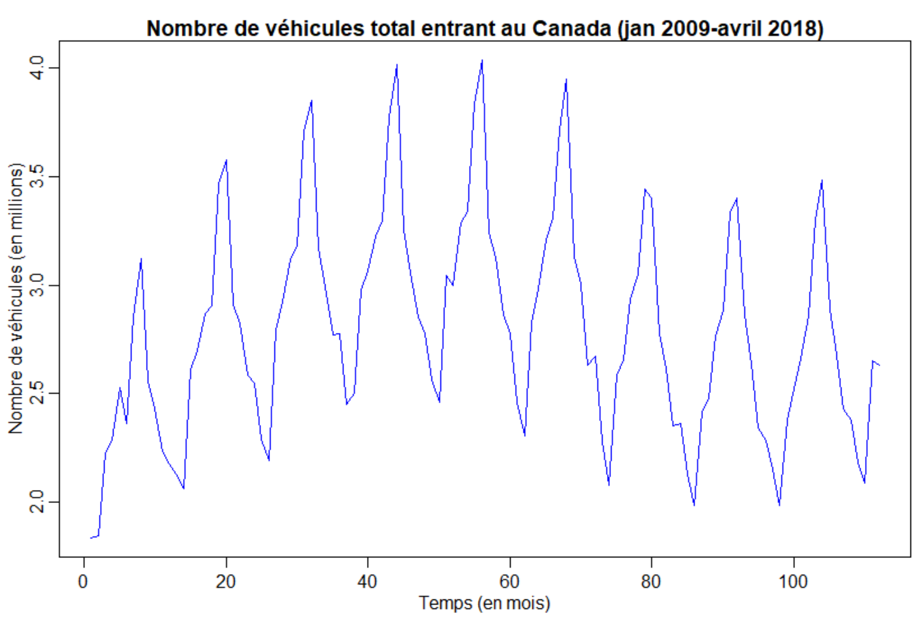



```r
# Modèle MA(1)
acf2(Zt, max.lag = 30)
# ACF: Nombre fini (un seul pic)
# PACF: Nombre infini de PACF

# En conclusion, toutes les observations
confirment les résultats
# d'identification des modèles ARMA vus en
classe.
```

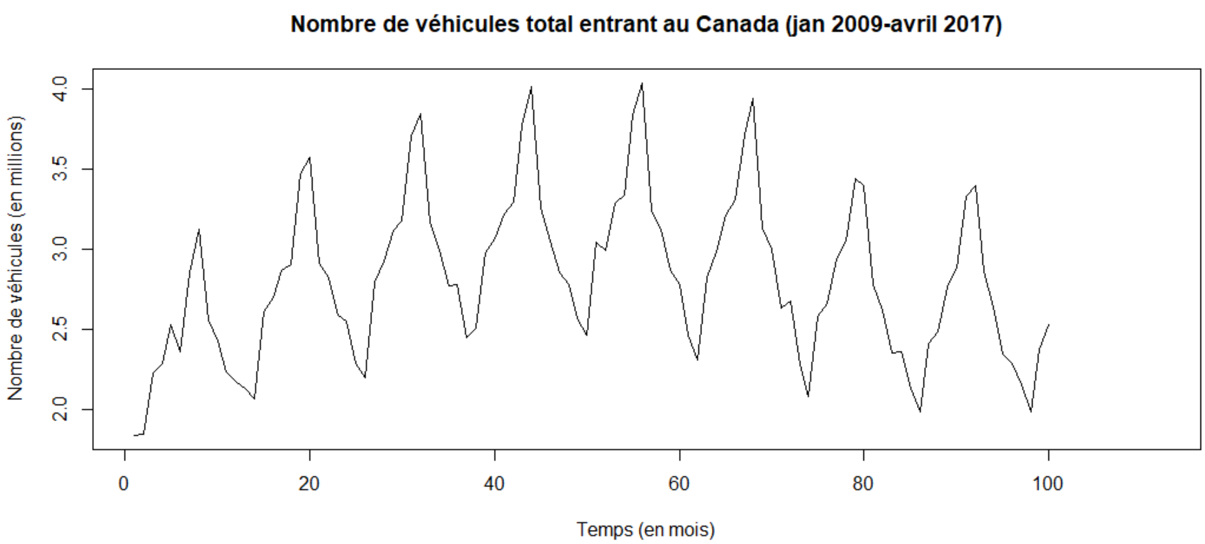


EXERCICE 4

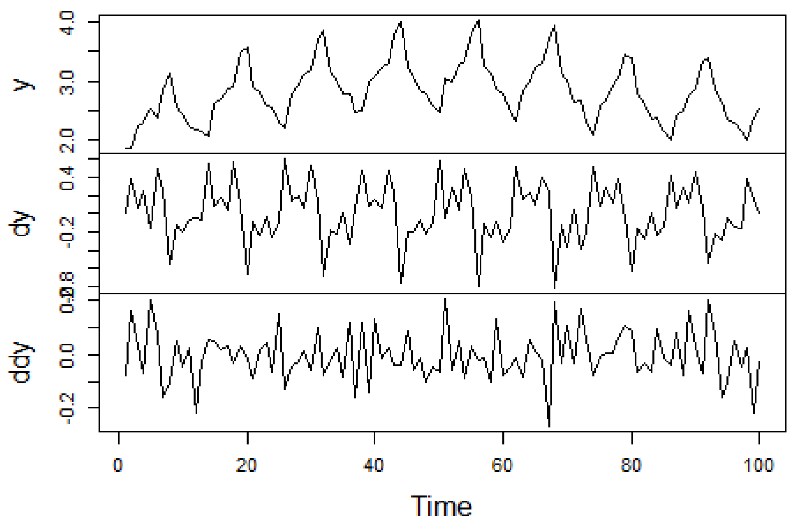
```r
#################################################################################
# ÉTAPES PRÉLIMINAIRES POUR L'ANALYSE DES DONNÉES
# 1) CUEILLETTE D'INFORMATIONS

# Les données sont tirées du site: https://www150.statcan.gc.ca/n1/fr/type/donnees
# Quelques informations pertinentes par rapport aux données cueillies:

# Numéro d'identification: 24-10-0002-01 (anciennement CANSIM 427-0002)
# Nom d'étude: Nombre de véhicules voyageant entre le Canada et les États-Unis
#               (section: Total des véhicules entrant au Canada)
# Période considérée: Janvier 2009 - Avril 2018

# On veut étudier le nombre total des véhicules entrant au Canada entre janvier
# 2009 et avril 2018. Étant donné qu'on reçoit des millions de véhicules par
# mois, l'analyse sera abordée selon les données recueillies, mais divisées par
# un million afin de travailler avec des nombres plus petits. Les conclusions
# demeurent identiques; le but est uniquement de faciliter la tâche à visualiser
# les résultats en réduisant le nombre de chiffres dans une donnée.

# Nous avons créé un fichier nommé Data_Q4.csv qui contient toutes les données
# nécessaires. L'annexe A montrera le contenu de ce dernier.

data <- read.csv('Data_Q4.csv', sep=';')
mois <- data$ï..Mois
y <- data$Qty
y_new <- data$Qty/1000000

write.csv(cbind(mois, y_new),'Data_Q4_modified.csv')
data <- read.csv('Data_Q4_modified.csv', sep=',')
y <- data$y_new

# Premier aperçu des données recueillies
```



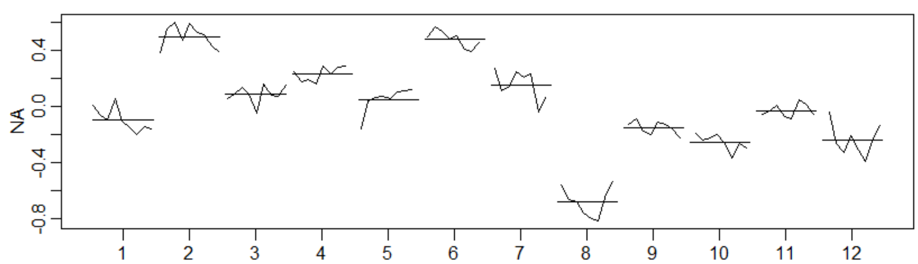
```r
# Commentaires: La série chronologique est saisonnière de fréquence mensuelle.
# On peut voir que chaque année, le mois qui a le plus de véhicules qui entre
# au Canada est le mois d'août, ce qui a du sens étant donné que c'est la
# période des vacances.

# Selon le graphique, nous n'étions pas certain si la tendance est linéaire
# dans cette série chronologique, mais il y a une tendance peu importe.
# Nous avons essayé plusieurs tentatives pour éliminer cette tendance:
# la racine carrée, l'inverse, le logarithmique, mais la tendance demeure
# présente. L'hypothèse sera de faire une différenciation deux fois puisqu'on
# voit une tendance qui ressemble à une fonction quadratique. Ensuite,
# on fait une différenciation saisonnière afin que la série soit plus
# stationnaire.


# 2) AJUSTEMENT DE CETTE SÉRIE AVEC LES MÉTHODES ARIMA

# Ajustement 1: Diviser les données par un million pour simplifier la tâche
#               d'analyse
# Ajustement 2: Différenciation première (et double?) pour éliminer la tendance
# Ajustement 3: Différenciation saisonnière où s = 12 mois


# 3) NOMBRE D'OBSERVATIONS

# Troncage des 12 dernières données, afin de faire l'analyse avec les 110
# premières observations

n <- 100
t <- seq(1,100,1)
y <- data$y_new[-seq(101, 112,1)]
```

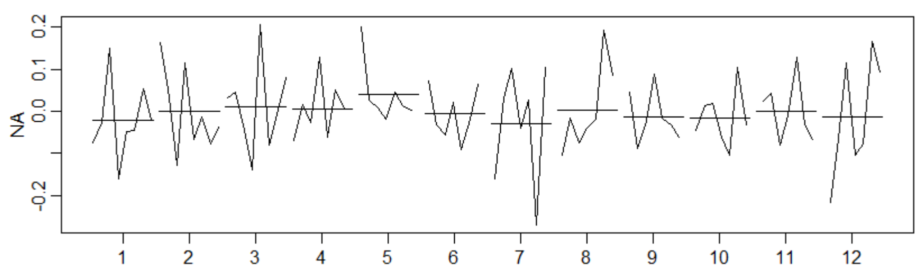


```
################################################################################
# a) Transformation de la série

# En faisant plusieurs expérimentations, on rejette notre hypothèse initiale,
# c'est-à-dire de faire deux différenciations non saisonnières.
# Avec une seule différenciation, on observe que les données semblent plus
# "stationnaires" et visuellement, nous avons éliminé la tendance.

# dy: données brutes différenciées une seule fois
# ddy: dy différenciées de façon saisonnière une seule fois
```



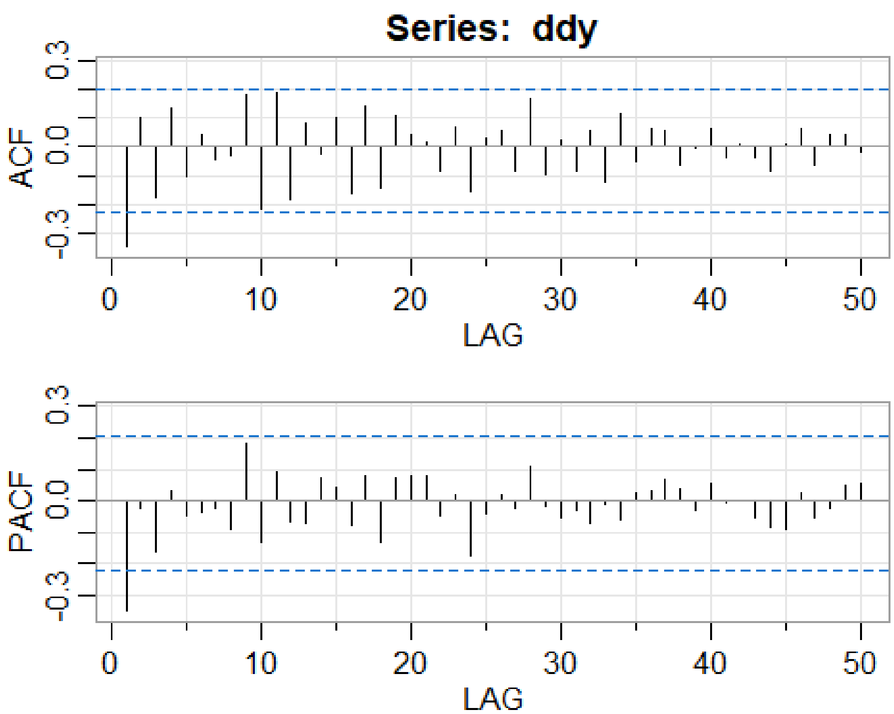

```
# On observe que la présence d'un motif pour chaque mois, donc il faut
# l'éliminer à l'aide de la différenciation saisonnière afin que les données
# semblent être stationnaires.
```





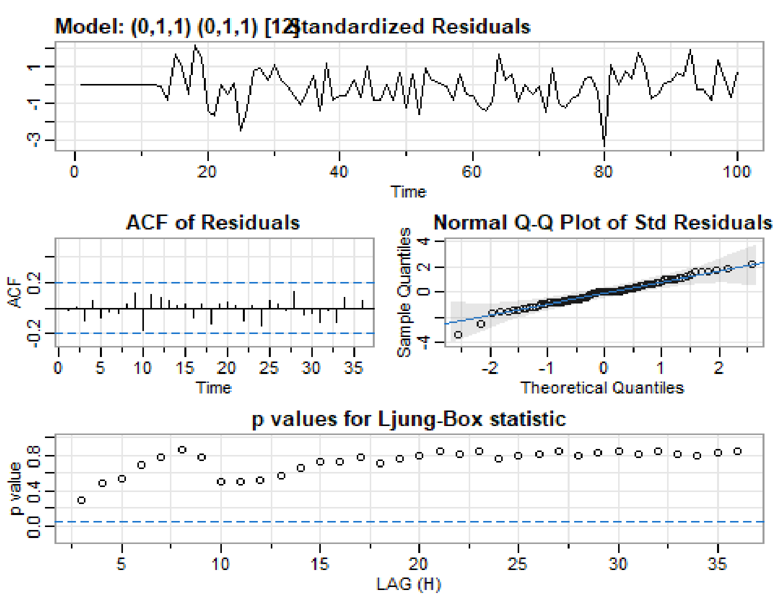

```
# Conclusion, on prend les données brutes qui sont différenciées
# de façon linéaire et saisonnière une fois chacune.
```

```
################################################################################
# b) Les ordres du modèle choisi

# Commençons par analyser les ACF et les PACF du modèle
acf2(ddy, max.lag=50)
```



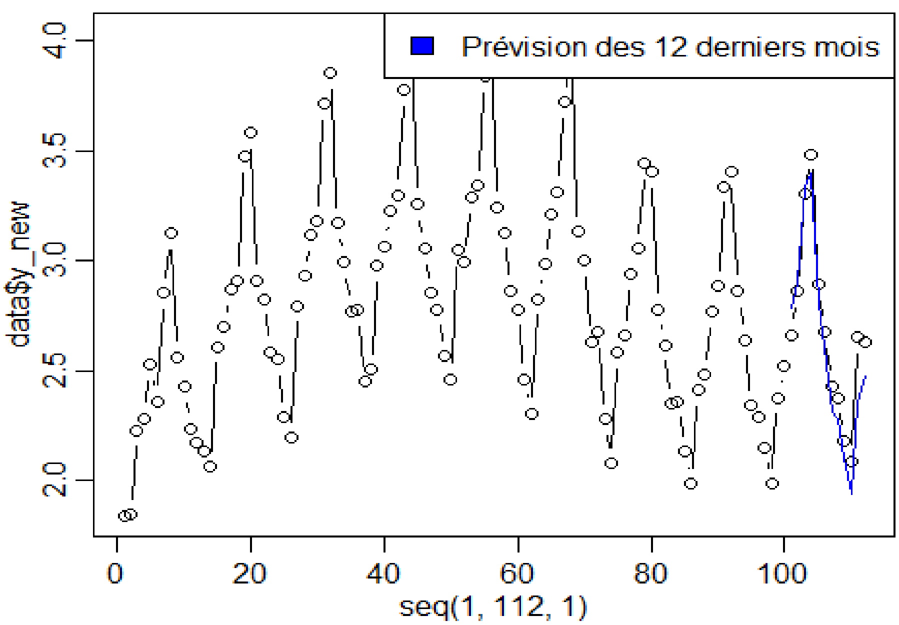

```
# COMPOSANTE NON SAISONNIÈRE (délais faibles entre 1 et 11):
# ACF: Observation d'un seul pic (ou) décroissance exponentielle des auto-
#      corrélations
# PACF: Observation d'un seul pic aussi (ou) décroissance exponentielle des
#       auto-corrélations

# Nous sommes certains la valeur de p ou de q ne dépassera pas de 1, donc nous
# avons trois modèles possibles à tester:
# ARIMA(1,1,0) - ARIMA(0,1,1) - ARIMA(1,1,1)

# COMPOSANTE SAISONNIÈRE (délais 12, 24, 36...):
# Dans le graphique des ACF et des PACF, c'est la partie la plus compliquée
# à remarquer puisque plusieurs phénomènes peuvent y arriver.

# Voici quelques hypothèses par rapport à la composante saisonnière du modèle:
# Le délai 12 dans ACF n'atteint pas vraiment la ligne "bleue", mais est proche,
# tandis que ce dernier n'est pas du tout atteint dans PACF.

# 1. Il se peut que le délai 12 est ignoré dans les deux graphiques, donc ça va
#       être ARIMA(0,1,0)_12.
# 2. Sinon, puisque le délai 12 a une auto-corrélation plus élevée dans ACF, il
#       se peut que ça soit un pic à considérer, donc ARIMA(0,1,1)_12 ou
#       ARIMA(1,1,1)_12 (peut considérer comme une décroissance exponentielle
#       dans le graphique ACF).
```

```
# Noter que les valeurs à côté de la ligne "sarima" sont les valeurs
# de log(vraisemblance) correspondante.

# Test 1: Avec ARMA(1,1) non saisonnier (modèles 1 à 3)
1. sarima(y, 1,1,1, 0,1,0, 12)              #92.00  -- non-significatif
2. sarima(y, 1,1,1, 0,1,1, 12)              #93.88  -- non-significatif
3. sarima(y, 1,1,1, 1,1,1, 12)              #94.82  -- non-significatif

# Test 2: Avec AR(1) non saisonnier (modèles 4 à 6)
4. sarima(y, 1,1,0, 0,1,0, 12)              #91.93  -- significatif
5. sarima(y, 1,1,0, 0,1,1, 12)              #93.76  -- significatif
6. sarima(y, 1,1,0, 1,1,1, 12)              #94.76  -- non significatif

# Test 3: Avec MA(1) non saisonnier (modèles 7 à 9)
7. sarima(y, 0,1,1, 0,1,0, 12)              #91.96  -- significatif
8. sarima(y, 0,1,1, 0,1,1, 12)              #93.86  -- significatif
9. sarima(y, 0,1,1, 1,1,1, 12)              #94.80  -- non-significatif

# Dès le départ les modèles 1, 2, 3, 6 et 9 ne sont pas considérées puisqu'il
# existe des paramètres dont la p-valeur n'est pas du tout significative.

# Il reste à choisir un modèle parmi 4, 5, 7 et 8.
# La valeur log(vraisemblance) la plus élevée (AIC la moins élevée) est le
# modèle 8, donc on choisira ce dernier et l'analyser pour confirmer si
# c'est un bon choix.

# Modèle choisi: ARIMA(0,1,1) x (0,1,1)_12

###############################################################################
# c) Statistiques diagnostics et autres diagnostics
# Observations des tests et des résidus

# MODÈLE ARIMA(0,1,1) X (0,1,1)_12
sarima(y, 0,1,1, 0,1,1, 12)
## converged
```



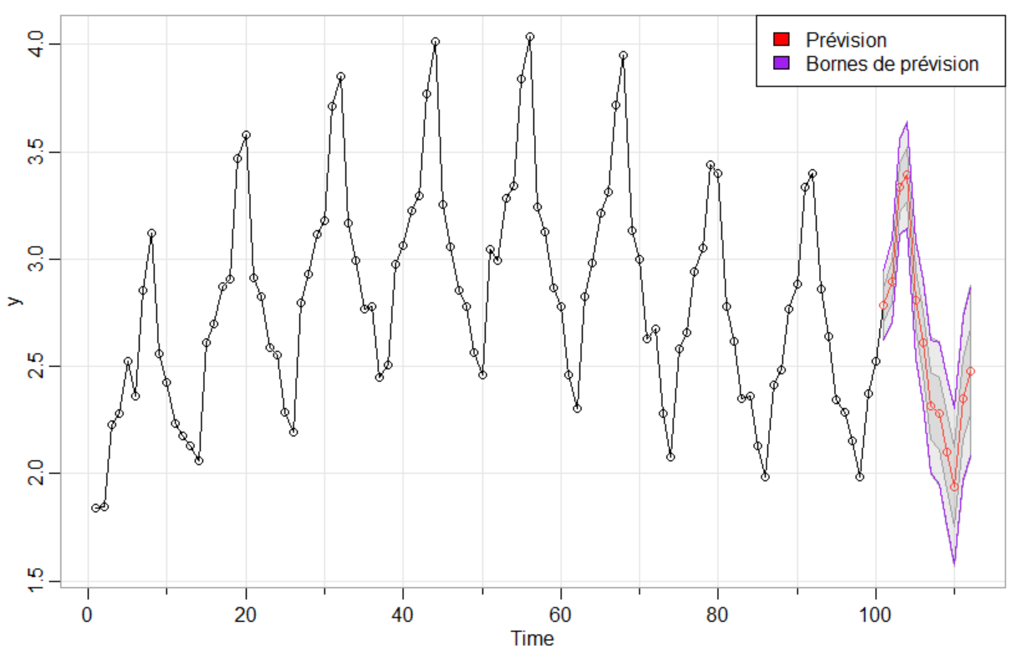
```
## $fit
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period =
S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed,
optim.control = list(trace = trc,
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##           ma1     sma1
##       -0.3226  -0.3164
## s.e.   0.1037   0.1554
##
## sigma^2 estimated as 0.006662:  log likelihood = 93.86,  aic = -181.72
##
## $degrees_of_freedom
## [1] 85
##
## $ttable
##      Estimate     SE t.value p.value
## ma1   -0.3226 0.1037 -3.1100  0.0025
## sma1  -0.3164 0.1554 -2.0362  0.0448
##
## $AIC
## [1] -1.854237
##
## $AICc
## [1] -1.852948
##
## $BIC
## [1] -1.77875

# AIC de -181.72 qui est le plus petit parmi les autres modèles testés.
# Log likelihood de 93.86.

# Standardized Residuals: Les résidus sont majoritairement entre -3 et 3 comme
# dans la loi normale centrée réduite (sauf au temps t=80 où ça dépasse un peu
# -3). On peut dire que les résidus se comportent selon la loi normale.

# ACF of Residuals: Aucune présence d'auto-corrélation entre les résidus, donc
# le modèle est accepté.

# Ici, les résidus sont selon la loi normale de façon majoritaire. On peut
# voir un peu de déviations lors des extrémités, mais la plupart des données
# demeurent "normales".

# Ljung-Box: les p-valeurs se trouvent au-dessus de la ligne bleue, donc on ne
# rejette pas l'hypothèse nulle pour tous les délais. Donc, on peut prendre
# ce modèle pour faire des prévisions des 12 prochains mois.
```

```r
# DISCUSSION DES P-VALEURS DES ESTIMATEURS ET DE LA VALEUR ALPHA:
# Il est vrai que les estimateurs ont des p-valeurs significatives, mais en
# ce qui concerne de sma1: la p-valeur est 0.0448, ce qui est très proche de 5%.

# Par conséquent, le modèle choisi va dépendre de l'erreur du type I tolérée.
# Si par exemple on choisit un modèle avec un alpha de 2.5%, ce modèle ne sera
# plus valide. Dans ce cas-ci, on prendra le modèle 7 (car le modèle 5 a aussi
# un estimateur avec une p-valeur proche de 5%) qui est ARIMA(0,1,1)x(0,1,0)_12.

# Bref, ici, alpha = 5% et c'est pour cela qu'on prend le modèle choisi en b).
# Cependant, le modèle 5 doit être réservé si jamais on veut des prévisions plus
# précises (alpha < 5%).

###############################################################################
# d) Performance prévisionnelle

# Vraies valeurs des 12 derniers mois
last_y <- data$y_new[seq(101,112,1)]

# Valeurs prédites des 12 derniers mois
predict_model <- sarima.for(y, 12, 0,1,1, 0,1,1, 12)
predict_y <- predict_model$pred

# EQMP
EQMP <- mean((predict_y - last_y)^2)
EQMP
## [1] 0.01698084

# Ligne bleue: les prévisions
# Ligne noire: les données brutes
```




```r
# Nous avons une EQMP de 0.01698084, ce qui est proche de zéro. Cela montre
# que le modèle choisi prédit assez bien les 12 prochains mois.
```

```r
###############################################################################
# e) Intervalles de prévision

# En se servant de la formule y_hat +/- 1.95*se, on obtient les intervalles
# de prévision suivants pour les 12 dernières données
borneInf <- predict_y - 1.96*predict_model$se
borneSup <- predict_y + 1.96*predict_model$se

cbind(borneInf, predict_y, borneSup)
## Time Series:
## Start = 101
## End = 112
## Frequency = 1
##     borneInf predict_y borneSup
## 101 2.624813  2.784791 2.944768
## 102 2.704049  2.897273 3.090497
## 103 3.112857  3.334393 3.555929
## 104 3.145944  3.392562 3.639181
## 105 2.540047  2.809423 3.078799
## 106 2.318592  2.608947 2.899302
## 107 2.005358  2.315275 2.625193
## 108 1.951728  2.280045 2.608361
## 109 1.756602  2.102340 2.448077
## 110 1.577291  1.939613 2.301934
## 111 1.972231  2.350411 2.728590
## 112 2.081813  2.475211 2.868609
###############################################################################
# f) Représentation graphique des prévisions

# Les lignes violettes représentent les bornes inférieures et supérieures
# des intervalles de prévisions.
sarima.for(y, 12, 0,1,1, 0,1,1, 12)
```



```
# Modèle choisi: ARIMA(0,1,1) X (0,1,1)12
# (1 - B12)(1 - B) Yt = (1 - 0.3226B)(1 - 0.3164B12) et
# ou
# Yt = Yt-1 + Yt-12 - Yt-13 + et - 0.3226 et-1 - 0.3164 et-12 + 0.10207064 et-13

# où e = epsilon
```