Travail 5 - GAA-7007

Karolane Bourdon, NI 111061527

21/04/2020

library(forecast)  
library(fpp2)  
library(tidyverse)  
library(lubridate)

## Créer une série temporelle

1- Importer les données

hawai <- read\_csv("data/hawai.csv")  
head(hawai)

## # A tibble: 6 x 2  
## time CO2  
## <dbl> <dbl>  
## 1 1958. 316.  
## 2 1958. 317.  
## 3 1958. 317.  
## 4 1958. 317.  
## 5 1958. 316.  
## 6 1959. 315.

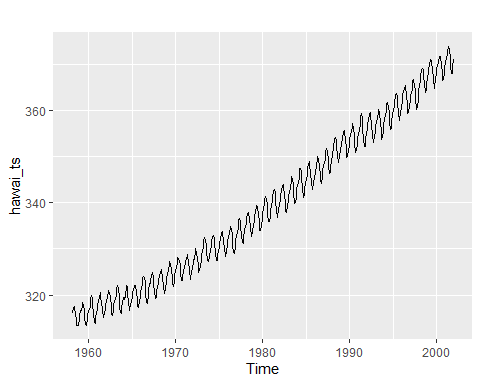
**Commentaire :** Les dates en format décimal peuvent être directement reconnues par la fonction de série temporelle. J’ai donc laissé tombé ma modification initiale qui avait été faite à l’aide de la fonction “date\_decimal()” du module Lubridate.

2- Créer la série temporelle

hawai\_ts <- ts(hawai %>% dplyr::select(-time), #Indique les données à utiliser  
 start = c(hawai$time[1]), #Spécifie la date de départ  
 frequency = 12) #Spécifie la fréquence mensuelle (12 observations/année)

3- Visualiser les données

autoplot(hawai\_ts)



#Graphique permettant de visualiser la série temporelle et de vérifier la conformité des données au fichier original.

**Commentaire :** La fonction ts() a reconnu les dates décimales. Les données débutent bien en 1958 et terminent à la fin de l’année 2001.

## Partitionner les données

hawai\_ts\_train <- window(hawai\_ts, start = 1958, end = 1988.9) #Données qui seront utilisées pour entraîner le modèle  
hawai\_ts\_test <- window(hawai\_ts, start = 1989, end = 2002) #Données qui seront utilisées pour tester le modèle

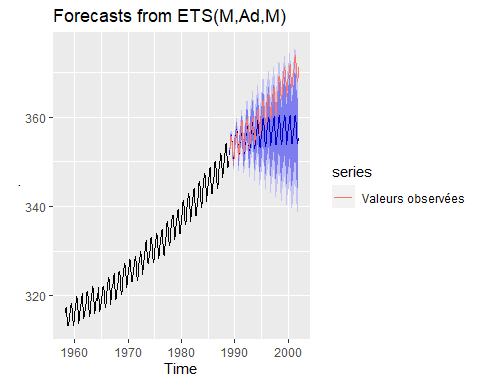
**Commentaire** : Puisque l’objectif est de prévoir le futur, la partie servant à tester le modèle doit se trouver à la fin pour travailler en extrapolation. Il est habituel d’utiliser 70% des données pour l’entrainement et 30% pour la vérification du modèle. Dans le cas présent, nous avons près de 44 années de suivi, donc 31 années seront utilisées pour l’entrainement et 13 années serviront à vérifier le modèle.

## Créer un modèle ETS

CO2\_model <- hawai\_ts\_train %>%  
 ets() %>%  
 forecast (h = 12\*13) #Commande une prédiction pour 13 années (12 mois/an\*13 ans)  
  
CO2\_model$model$par #Pour visualiser les paramètres du modèle

## alpha beta gamma phi l b   
## 0.61081947 0.03588844 0.00010599 0.97558847 314.86775308 0.04314888   
## s0 s1 s2 s3 s4 s5   
## 1.00174450 0.99985631 0.99723003 0.99393366 0.99073666 0.99128677   
## s6 s7 s8 s9 s10   
## 0.99639340 1.00228812 1.00687366 1.00866932 1.00707336

autoplot(CO2\_model) + #Pour tracer la prédiction  
 autolayer(hawai\_ts\_test, serie = "Valeurs observées") #Pour tracer les valeurs réellement observées

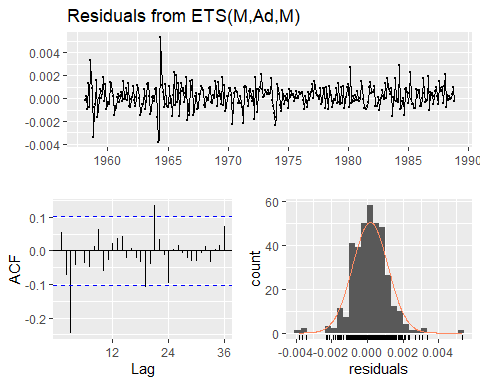


**Commentaires :**

* Bien que les valeurs réelles observées soient dans l’intervalle de confiance de 80% du modèle, ce dernier n’a pas permis de modéliser la tendance générale des données.
* Le modèle automatisé détecte une tendance “Additive damped”. En effet, un paramètre “phi” est inclus dans le modèle, ce qui cause une atténuation de la pente avec le temps. Toutefois, les valeurs réelles mesurées montrent une tendance additive sans atténuation dans le temps. Une méthode manuelle nous permettrait de spécifier la tendance additive stricte à l’aide de l’argument “damped = FALSE”.
* La saisonalité a été détectée comme multiplicative, ce qui permet de modéliser les cycles saisonniers qui sont en effet présents dans les données.

## Analyse des résidus

checkresiduals(CO2\_model)



##   
## Ljung-Box test  
##   
## data: Residuals from ETS(M,Ad,M)  
## Q\* = 50.166, df = 7, p-value = 1.34e-08  
##   
## Model df: 17. Total lags used: 24

**Commentaire :** Les résidus semblent respecter le postulat de normalité. Par contre, la valeur de p < à 0.05 pour le test de Ljung-Box indique qu’il y a des tendances dans les données qui ne sont pas collectées par le modèle. Ce qui pourrait expliquer que le modèle est détecté une tendance à l’aplatissement plutôt qu’une tendance additive stricte.

## Fiabilité et amélioration du modèle