Devoir 5. Séries temporelles

Heidi Pascagaza

31/7/2020

Les données du fichier hawai.csv comprennent les moyennes des mesures mensuelles de CO2 atmosphérique entre en ppm-volume collectées au Mauna Loa Observatory à Hawai de mars 1958 à décembre 2001, inclusivement.

library("tidyverse")

## -- Attaching packages ---------- tidyverse 1.3.0 --

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.1 v dplyr 0.8.5  
## v tidyr 1.0.3 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## Warning: package 'ggplot2' was built under R version 4.0.2

## -- Conflicts ------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library("forecast")

## Warning: package 'forecast' was built under R version 4.0.2

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

library ("cowplot")

## Warning: package 'cowplot' was built under R version 4.0.2

##   
## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

## Note: As of version 1.0.0, cowplot does not change the

## default ggplot2 theme anymore. To recover the previous

## behavior, execute:  
## theme\_set(theme\_cowplot())

## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

library ("lubridate")

## Warning: package 'lubridate' was built under R version 4.0.2

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:cowplot':  
##   
## stamp

## The following objects are masked from 'package:dplyr':  
##   
## intersect, setdiff, union

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

# 1. Exploration initiale des données

Les données de hawai.csv sont utilisées pour créer, modéliser et évaluer une série temporelle comme indiqué ci-dessous.

hawaiCO2\_data <- read\_csv("data/hawai.csv")

## Parsed with column specification:  
## cols(  
## time = col\_double(),  
## CO2 = col\_double()  
## )

hawai\_CO2 <- hawaiCO2\_data %>%  
 mutate(time = date\_decimal(hawaiCO2\_data$time, tz = "UTC"))  
glimpse(hawai\_CO2)

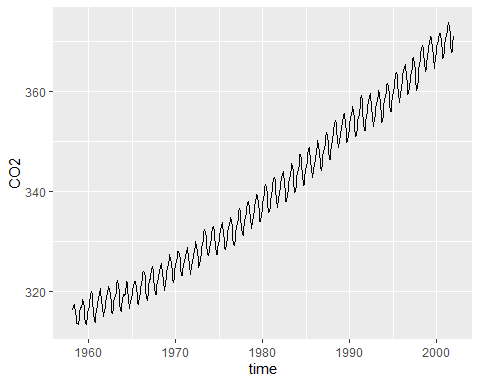
## Rows: 526  
## Columns: 2  
## $ time <dttm> 1958-03-02 20:00:01, 1958-04-02 06:00:00, 1958-05-02 16:00:00...  
## $ CO2 <dbl> 316.1000, 317.2000, 317.4333, 317.4333, 315.6250, 314.9500, 31...

Initialement, les données sont chargées. Il est possible d’observer à partir de l’ensemble des données que les valeurs de temps sont exprimées en décimales. Les données de temps sont transformées en données de type temporaires avec la fonction date\_decimal du paquet lubridate, indiquant le fuseau horaire au moyen de la fonction tz.

Ensuite avec la fonction glimpse, il est possible d’observer que les données sont composées de 526 lignes qui présentent le temps transformé en données de type temporelles et la concentration de CO2 (ppm-volume).

Les données sont ensuite explorées de manière graphique, en utilisant la fonction ggplot. On observe un comportement croissant de la concentration de CO2 en fonction du temps.

hawai\_CO2 %>%  
 ggplot(aes(x=time, y=CO2))+ geom\_line()



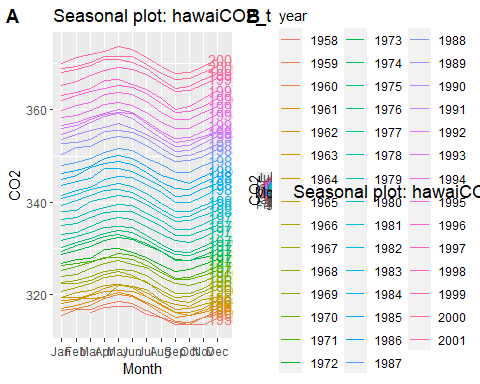
## 1.2. Création de la série temporelle du CO2

Une série temporelle de la concentration en CO2 (ppm - volume) en fonction du temps est alors créée. Ceci est fait en utilisant la fonction stats::ts(). La date de début ( 03 mars - 1958) est identifiée comme la première donnée de l’ensemble de données à l’aide de la fonction start au mois de mars (3), avec une fréquence mensuelle (12 mois dans une année).

hawaiCO2\_ts <- ts(hawai\_CO2 %>% dplyr::select(-time),  
 start = c(hawai\_CO2$time[1]%>% year(),3),  
 frequency = 12)

Ensuite, on peut avoir un aperçu du comportement de la série en utilisant le paquet forecast et la fonction ggseasonplot.

ggA <- ggseasonplot(hawaiCO2\_ts, year.labels=TRUE) +  
 ylab("CO2")   
ggB <- ggseasonplot(hawaiCO2\_ts, polar=TRUE) + ylab("CO2")  
plot\_grid(ggA,ggB, ncol=2, labels=c("A","B"))



Les graphiques *A* et *B* permettent de constater qu’il existe une tendance à l’augmentation de la concentration de CO2 (ppm-volume) au fil des ans. De plus, que le graphique *A* permet aussi de voir une fluctuation saisonnière, qui montre une tendance générale de façon régulière dans les mois, avec une diminution des concentrations au mois de juillet - août.

# 3. Série temporelle.

## 3.1. Création de la série d’entraînement et la série de test.

Postérieurement, la série est divisée en deux parties en utilisant la fonction window: la série d’entraînement qui représente environ le 70 % des données et une autre la série de test qui représente environ le 30 % des données. La série d’entraînement est séparée des observations enregistrées depuis le mois 3 (début) de 1958 au mois 12 (fin) de 1988. D’ailleurs la série de test des données est comprise entre les mois 1 de 1989 au mois 12 de 2001 (30%).

CO2\_ts\_train <- window(hawaiCO2\_ts, start = c(1958, 3), end = c(1988, 12))  
CO2\_ts\_test <- window(hawaiCO2\_ts, start = c(1989, 1), end = c(2001, 12))

## 3.2.Modélisation de la série temporelle - Méthode ETS

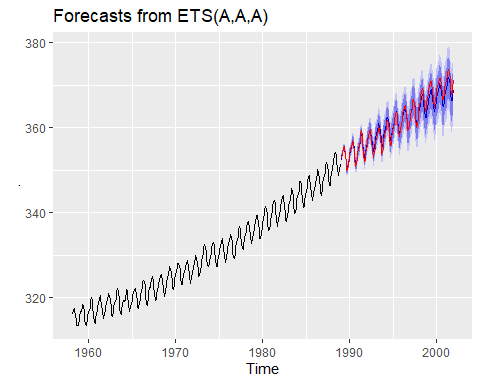
La méthode ETA est choisie pour réaliser le modèle de prévision pour les raisons suivantes :

* La série temporelle présente un comportement non stationnaire, ce qui signifie qu’il n’y a pas de moyenne constante dans le temps. Dans ce cas-là la moyenne augmente en fonction du temps.
* Il y a une tendance de fluctuations saisonnières des données qui peut expliquer les données futures de la modélisation.
* La méthode donne plus de poids aux valeurs précédentes proches.

Le modèle ETA (A,A,A) permet la description du modèle de projection au moyen d’équations de prévision, de niveau, de tendance et de session par modélisation additive.

La modélisation est effectuée à l’aide de la méthode ETS et la comparaison de la série de test et la série d’origine est effectuée pour visualiser l’ajustement du modèle.

ets\_CO2 <- CO2\_ts\_train %>% ets()  
hawai\_CO2\_ets <- ets\_CO2 %>% forecast(h = length(CO2\_ts\_test))  
autoplot(hawai\_CO2\_ets) + autolayer(CO2\_ts\_test, color = "red")



Le représentation graphique de la série confirme une tendance qui décrit l’augmentation à long terme de la variable de concentration en CO2. De plus, il est possible de voir une fluctuation saisonnière avec une structure périodique qui oscille dans une tendance régulière selon le calendrier.

## 3.3. Analyse de la précision du modèle.

accuracy(hawai\_CO2\_ets, hawaiCO2\_ts)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.009127538 0.3329635 0.2496935 0.002272597 0.07593717 0.2034533  
## Test set 0.258792487 1.1125984 0.8835110 0.067546737 0.24278880 0.7198956  
## ACF1 Theil's U  
## Training set 0.06670592 NA  
## Test set 0.94372350 0.8673732

L’évaluation de la précision du modèle est effectuée par la fonction accuracy et la valeur de Mean Absolute Scaled Error (MASE) résultant ainsi de la modélisation :

MASE - training set ETS | 0.2034 MASE - test set ETS | 0.7198

À partir des valeurs MASE, il est possible de prouver que les valeurs sont proches de zéro, ce qui indique une bonne précision. Il est important de remarquer qu’il y a une meilleure précision dans les données d’entraînement par rapport aux données de test.

En plus du graphique du modèle, on observe que les données originales mises en évidence par la ligne rouge sont situées dans les limites inférieure et supérieure projetées par le modèle.

Dans la parte suivant les paramètres des équations du modèle sont décrits et il est mis en évidence les valeurs des intervalles inférieurs et supérieurs entre lesquels les données de prédiction du modèle pourraient être situées.

summary(hawai\_CO2\_ets)

##   
## Forecast method: ETS(A,A,A)  
##   
## Model Information:  
## ETS(A,A,A)   
##   
## Call:  
## ets(y = .)   
##   
## Smoothing parameters:  
## alpha = 0.6698   
## beta = 0.003   
## gamma = 2e-04   
##   
## Initial states:  
## l = 315.2483   
## b = 0.102   
## s = 0.6164 0.0116 -0.9384 -2.0207 -3.0899 -2.8853  
## -1.2381 0.7047 2.2379 2.8344 2.416 1.3513  
##   
## sigma: 0.3404  
##   
## AIC AICc BIC   
## 1408.202 1409.940 1474.731   
##   
## Error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.009127538 0.3329635 0.2496935 0.002272597 0.07593717 0.2034533  
## ACF1  
## Training set 0.06670592  
##   
## Forecasts:  
## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## Jan 1989 352.3719 351.9356 352.8081 351.7047 353.0391  
## Feb 1989 353.0892 352.5634 353.6150 352.2851 353.8933  
## Mar 1989 353.9360 353.3332 354.5388 353.0141 354.8579  
## Apr 1989 355.1129 354.4413 355.7845 354.0857 356.1400  
## May 1989 355.6436 354.9091 356.3781 354.5203 356.7670  
## Jun 1989 355.1591 354.3661 355.9520 353.9464 356.3718  
## Jul 1989 353.7379 352.8901 354.5857 352.4413 355.0345  
## Aug 1989 351.9070 351.0072 352.8067 350.5309 353.2830  
## Sep 1989 350.3716 349.4224 351.3209 348.9199 351.8234  
## Oct 1989 350.2792 349.2825 351.2760 348.7549 351.8036  
## Nov 1989 351.4604 350.4180 352.5029 349.8662 353.0547  
## Dec 1989 352.6548 351.5683 353.7414 350.9931 354.3165  
## Jan 1990 353.7165 352.5871 354.8458 351.9893 355.4436  
## Feb 1990 354.4338 353.2629 355.6046 352.6431 356.2245  
## Mar 1990 355.2806 354.0693 356.4919 353.4281 357.1332  
## Apr 1990 356.4574 355.2067 357.7082 354.5445 358.3704  
## May 1990 356.9882 355.6988 358.2776 355.0163 358.9601  
## Jun 1990 356.5037 355.1765 357.8308 354.4740 358.5333  
## Jul 1990 355.0825 353.7183 356.4466 352.9962 357.1687  
## Aug 1990 353.2516 351.8511 354.6520 351.1097 355.3934  
## Sep 1990 351.7162 350.2801 353.1524 349.5198 353.9126  
## Oct 1990 351.6238 350.1526 353.0951 349.3737 353.8739  
## Nov 1990 352.8050 351.2992 354.3109 350.5020 355.1080  
## Dec 1990 353.9994 352.4595 355.5393 351.6443 356.3545  
## Jan 1991 355.0611 353.4875 356.6346 352.6546 357.4676  
## Feb 1991 355.7784 354.1717 357.3850 353.3212 358.2356  
## Mar 1991 356.6252 354.9858 358.2646 354.1179 359.1325  
## Apr 1991 357.8020 356.1303 359.4738 355.2453 360.3588  
## May 1991 358.3328 356.6290 360.0366 355.7271 360.9385  
## Jun 1991 357.8482 356.1128 359.5837 355.1942 360.5023  
## Jul 1991 356.4271 354.6603 358.1938 353.7251 359.1291  
## Aug 1991 354.5961 352.7984 356.3939 351.8467 357.3456  
## Sep 1991 353.0608 351.2323 354.8893 350.2644 355.8573  
## Oct 1991 352.9684 351.1095 354.8274 350.1254 355.8114  
## Nov 1991 354.1496 352.2605 356.0387 351.2604 357.0388  
## Dec 1991 355.3440 353.4249 357.2631 352.4090 358.2790  
## Jan 1992 356.4056 354.4568 358.3545 353.4252 359.3861  
## Feb 1992 357.1229 355.1446 359.1013 354.0974 360.1485  
## Mar 1992 357.9698 355.9622 359.9774 354.8994 361.0401  
## Apr 1992 359.1466 357.1099 361.1833 356.0318 362.2614  
## May 1992 359.6774 357.6118 361.7429 356.5184 362.8364  
## Jun 1992 359.1928 357.0986 361.2871 355.9899 362.3957  
## Jul 1992 357.7716 355.6489 359.8944 354.5251 361.0181  
## Aug 1992 355.9407 353.7896 358.0919 352.6509 359.2306  
## Sep 1992 354.4054 352.2261 356.5847 351.0724 357.7384  
## Oct 1992 354.3130 352.1056 356.5204 350.9371 357.6889  
## Nov 1992 355.4942 353.2589 357.7294 352.0756 358.9127  
## Dec 1992 356.6886 354.4255 358.9516 353.2276 360.1495  
## Jan 1993 357.7502 355.4596 360.0409 354.2470 361.2535  
## Feb 1993 358.4675 356.1494 360.7856 354.9223 362.0128  
## Mar 1993 359.3144 356.9689 361.6599 355.7273 362.9015  
## Apr 1993 360.4912 358.1185 362.8639 356.8624 364.1200  
## May 1993 361.0220 358.6221 363.4218 357.3517 364.6922  
## Jun 1993 360.5374 358.1105 362.9643 356.8258 364.2490  
## Jul 1993 359.1162 356.6624 361.5700 355.3635 362.8690  
## Aug 1993 357.2853 354.8047 359.7659 353.4916 361.0791  
## Sep 1993 355.7500 353.2427 358.2573 351.9154 359.5846  
## Oct 1993 355.6576 353.1236 358.1915 351.7823 359.5329  
## Nov 1993 356.8388 354.2783 359.3992 352.9229 360.7547  
## Dec 1993 358.0331 355.4462 360.6200 354.0768 361.9895  
## Jan 1994 359.0948 356.4815 361.7081 355.0982 363.0915  
## Feb 1994 359.8121 357.1726 362.4517 355.7753 363.8489  
## Mar 1994 360.6590 357.9932 363.3247 356.5821 364.7359  
## Apr 1994 361.8358 359.1439 364.5277 357.7189 365.9526  
## May 1994 362.3665 359.6486 365.0845 358.2098 366.5232  
## Jun 1994 361.8820 359.1381 364.6259 357.6856 366.0784  
## Jul 1994 360.4608 357.6910 363.2306 356.2247 364.6969  
## Aug 1994 358.6299 355.8342 361.4256 354.3543 362.9055  
## Sep 1994 357.0946 354.2731 359.9160 352.7795 361.4096  
## Oct 1994 357.0022 354.1550 359.8494 352.6477 361.3566  
## Nov 1994 358.1834 355.3105 361.0562 353.7897 362.5770  
## Dec 1994 359.3777 356.4792 362.2762 354.9449 363.8106  
## Jan 1995 360.4394 357.5153 363.3635 355.9674 364.9114  
## Feb 1995 361.1567 358.2071 364.1063 356.6457 365.6677  
## Mar 1995 362.0035 359.0285 364.9786 357.4536 366.5535  
## Apr 1995 363.1804 360.1799 366.1808 358.5915 367.7692  
## May 1995 363.7111 360.6853 366.7370 359.0835 368.3388  
## Jun 1995 363.2266 360.1754 366.2778 358.5602 367.8929  
## Jul 1995 361.8054 358.7289 364.8818 357.1004 366.5104  
## Aug 1995 359.9745 356.8728 363.0762 355.2308 364.7181  
## Sep 1995 358.4392 355.3122 361.5661 353.6570 363.2213  
## Oct 1995 358.3467 355.1947 361.4988 353.5261 363.1674  
## Nov 1995 359.5279 356.3507 362.7051 354.6688 364.3870  
## Dec 1995 360.7223 357.5200 363.9246 355.8248 365.6198  
## Jan 1996 361.7840 358.5566 365.0114 356.8481 366.7198  
## Feb 1996 362.5013 359.2489 365.7537 357.5272 367.4754  
## Mar 1996 363.3481 360.0707 366.6255 358.3358 368.3605  
## Apr 1996 364.5250 361.2226 367.8273 359.4744 369.5755  
## May 1996 365.0557 361.7284 368.3830 359.9670 370.1444  
## Jun 1996 364.5712 361.2189 367.9234 359.4444 369.6979  
## Jul 1996 363.1500 359.7729 366.5271 357.9851 368.3148  
## Aug 1996 361.3191 357.9171 364.7210 356.1162 366.5219  
## Sep 1996 359.7837 356.3569 363.2105 354.5429 365.0246  
## Oct 1996 359.6913 356.2397 363.1429 354.4125 364.9701  
## Nov 1996 360.8725 357.3961 364.3489 355.5558 366.1892  
## Dec 1996 362.0669 358.5657 365.5681 356.7123 367.4215  
## Jan 1997 363.1286 359.6026 366.6545 357.7361 368.5210  
## Feb 1997 363.8459 360.2952 367.3965 358.4156 369.2761  
## Mar 1997 364.6927 361.1173 368.2681 359.2246 370.1608  
## Apr 1997 365.8695 362.2695 369.4696 360.3637 371.3754  
## May 1997 366.4003 362.7756 370.0250 360.8567 371.9439  
## Jun 1997 365.9157 362.2663 369.5652 360.3345 371.4970  
## Jul 1997 364.4946 360.8205 368.1686 358.8756 370.1135  
## Aug 1997 362.6637 358.9650 366.3623 357.0070 368.3203  
## Sep 1997 361.1283 357.4050 364.8516 355.4340 366.8226  
## Oct 1997 361.0359 357.2880 364.7838 355.3040 366.7678  
## Nov 1997 362.2171 358.4446 365.9896 356.4476 367.9866  
## Dec 1997 363.4115 359.6144 367.2085 357.6044 369.2186  
## Jan 1998 364.4731 360.6515 368.2948 358.6285 370.3178  
## Feb 1998 365.1904 361.3443 369.0366 359.3082 371.0727  
## Mar 1998 366.0373 362.1666 369.9080 360.1175 371.9571  
## Apr 1998 367.2141 363.3188 371.1094 361.2568 373.1714  
## May 1998 367.7449 363.8251 371.6647 361.7501 373.7397  
## Jun 1998 367.2603 363.3160 371.2047 361.2280 373.2927  
## Jul 1998 365.8391 361.8703 369.8080 359.7693 371.9090  
## Aug 1998 364.0082 360.0149 368.0016 357.9009 370.1155  
## Sep 1998 362.4729 358.4551 366.4907 356.3281 368.6177  
## Oct 1998 362.3805 358.3382 366.4228 356.1983 368.5627  
## Nov 1998 363.5617 359.4949 367.6285 357.3420 369.7814  
## Dec 1998 364.7561 360.6648 368.8474 358.4989 371.0132  
## Jan 1999 365.8177 361.7019 369.9335 359.5232 372.1123  
## Feb 1999 366.5350 362.3948 370.6753 360.2030 372.8670  
## Mar 1999 367.3819 363.2171 371.5466 361.0124 373.7513  
## Apr 1999 368.5587 364.3695 372.7479 362.1518 374.9656  
## May 1999 369.0895 364.8758 373.3032 362.6452 375.5337  
## Jun 1999 368.6049 364.3668 372.8431 362.1232 375.0866  
## Jul 1999 367.1837 362.9211 371.4463 360.6646 373.7028  
## Aug 1999 365.3528 361.0657 369.6399 358.7963 371.9093  
## Sep 1999 363.8175 359.5059 368.1290 357.2236 370.4114  
## Oct 1999 363.7251 359.3891 368.0611 357.0937 370.3564  
## Nov 1999 364.9063 360.5458 369.2667 358.2375 371.5750  
## Dec 1999 366.1006 361.7157 370.4856 359.3945 372.8068  
## Jan 2000 367.1623 362.7529 371.5717 360.4187 373.9059  
## Feb 2000 367.8796 363.4457 372.3135 361.0986 374.6606  
## Mar 2000 368.7265 364.2681 373.1848 361.9080 375.5449  
## Apr 2000 369.9033 365.4205 374.3861 363.0474 376.7591  
## May 2000 370.4340 365.9268 374.9413 363.5408 377.3273  
## Jun 2000 369.9495 365.4178 374.4812 363.0188 376.8802  
## Jul 2000 368.5283 363.9721 373.0845 361.5602 375.4965  
## Aug 2000 366.6974 362.1167 371.2781 359.6918 373.7030  
## Sep 2000 365.1621 360.5569 369.7673 358.1191 372.2051  
## Oct 2000 365.0697 360.4400 369.6993 357.9892 372.1501  
## Nov 2000 366.2509 361.5967 370.9050 359.1329 373.3688  
## Dec 2000 367.4452 362.7666 372.1239 360.2898 374.6006  
## Jan 2001 368.5069 363.8037 373.2101 361.3140 375.6998  
## Feb 2001 369.2242 364.4965 373.9519 361.9939 376.4545  
## Mar 2001 370.0710 365.3189 374.8232 362.8032 377.3389  
## Apr 2001 371.2479 366.4712 376.0246 363.9425 378.5532  
## May 2001 371.7786 366.9774 376.5798 364.4358 379.1215  
## Jun 2001 371.2941 366.4683 376.1198 363.9137 378.6744  
## Jul 2001 369.8729 365.0226 374.7232 362.4550 377.2908  
## Aug 2001 368.0420 363.1672 372.9168 360.5866 375.4974  
## Sep 2001 366.5067 361.6073 371.4060 359.0137 373.9996  
## Oct 2001 366.4143 361.4903 371.3382 358.8838 373.9447  
## Nov 2001 367.5954 362.6470 372.5439 360.0274 375.1635  
## Dec 2001 368.7898 363.8168 373.7629 361.1842 376.3955

La construction des équations est donnée par :

avec un

Le paramètre alpha décrit la distribution du poids, cette valeur s’éloignant de 0 et se rapprochant de 1 indique le poids le plus élevé des événements les plus récents.

On a , étant le paramètre qui décrit la vitesse à laquelle la tendance peut changer. Cette faible valeur indique que le changement est lent au cours du temps.

$Saison =γ(y\_t−l\_{t−1}−b\_{t−1})+(1−γ)S\_{t−m}; $

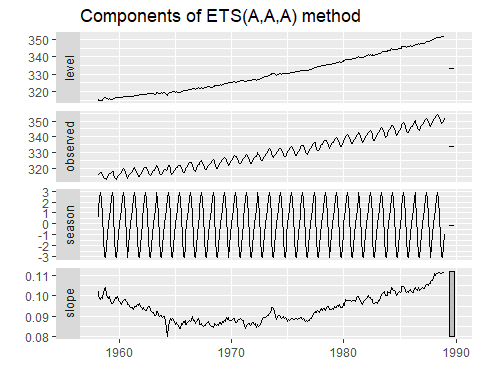
avec un

est un paramètre de la portion saisonnière qui fluctue autour de zéro.

L’absence de la valeur phi indique que l’adoucissement n’est probablement pas nécessaire.

Les composantes du modèle par la méthode ETS sont présentées ci-dessous:

autoplot (ets\_CO2)

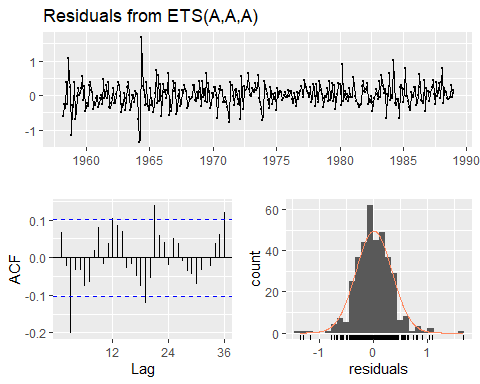


En observant les graphiques de level et observed, il est possible de constater que la concentration de CO2 augmente en fonction du temp.

En autre, le graphique season montre la variabilité des données au cours du temps, il est possible de mettre en évidence la fluctuation autour de zéro, présentant un comportement oscillent similaire au cours des années.

# 4. Analyse des résidus du modéle.

checkresiduals(ets\_CO2)



##   
## Ljung-Box test  
##   
## data: Residuals from ETS(A,A,A)  
## Q\* = 54.438, df = 8, p-value = 5.676e-09  
##   
## Model df: 16. Total lags used: 24

Le premier graphique permet de constater que les résidus sont principalement répartis autour de la valeur zéro. Cela peut donner une indication que la moyenne des résidus est proche de zéro, mais on observe deux pics qui peuvent être liés à des événements extrêmes au cours de ces deux années et qui peuvent influencer l’augmentation anormale des concentrations de CO2.

À partir du deuxième graphique de l’autocorrélogramme, il est possible de prouver que la plupart des résidus restent dans les intervalles de confiance, de sorte que le modèle peut être considéré comme valide.

À partir du troisième graphique, il est possible d’apprécier une distribution des données apparemment de type normal. Cependant, la présence d’une paire de données qui s’écartent de la distribution peut faire que la distribution des résidus ne soient pas de façon normale. C’est pourquoi il est important de vérifier cette indication au moyen d’un test de normalité.

La valeur de p = 5.676e-09 indique la faible probabilité que les résidus soient générés par un bruit blanc.

Le test de Shapiro est fait pour évaluer la normalité.

shapiro.test(residuals(ets\_CO2))

##   
## Shapiro-Wilk normality test  
##   
## data: residuals(ets\_CO2)  
## W = 0.96983, p-value = 6.061e-07

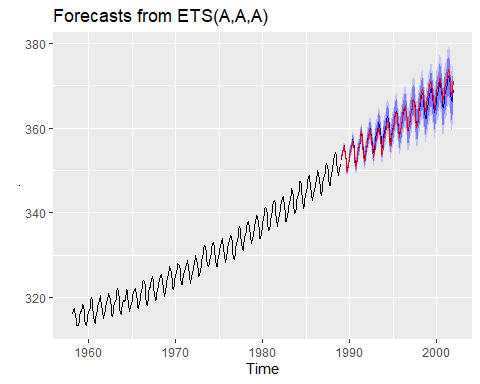
La faible valeur p < 0,05 indique que les résidus ne sont pas distribués de manière normale, peut-être parce que les intervalles du modèle sont trop restreints. Cependant, la prédiction est valable telle qu’elle a été évaluée précédemment par la précision du modèle.

# 5.Transformation de Box-Cox

La transformation de Box-Cox largement utilisée pour sa polyvalence est effectuée avec la série temporelle appliquée à la série d’entraînement, en utilisant le paramètre lambda qui est estimé comme la valeur optimale par la fonction Forecast::Box.Cox.lambda

Le modèle d’ajustement est présenté ci-dessous.

ets\_CO2\_B <- CO2\_ts\_train %>% ets(lambda = BoxCox.lambda(CO2\_ts\_train))  
hawai\_CO2\_B <- ets\_CO2\_B %>%  
 forecast(h = length(CO2\_ts\_test))  
  
autoplot(hawai\_CO2\_B) + autolayer(CO2\_ts\_test, color = "red")

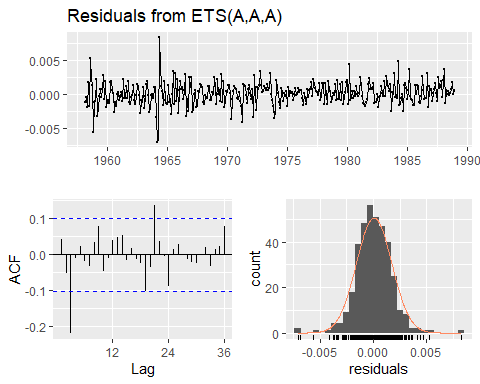


accuracy(hawai\_CO2\_B, hawaiCO2\_ts)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.01317536 0.3201071 0.2414950 0.003562256 0.07340163 0.1967731  
## Test set 0.17310078 1.0530018 0.8432859 0.044216546 0.23196266 0.6871197  
## ACF1 Theil's U  
## Training set 0.03928175 NA  
## Test set 0.94442197 0.8221877

À partir des valeurs MASE, on peut observer que le modèle présente une légère amélioration étant donné que ces valeurs sont plus proches de zéro. Pour la série d’entraînement la valeur MASE a un pourcentage de différence de 3,3% par rapport au modèle initial, alors que pour la valeur MASE de la série de test il y a un pourcentage de différence de 4,5%.

checkresiduals(hawai\_CO2\_B)



##   
## Ljung-Box test  
##   
## data: Residuals from ETS(A,A,A)  
## Q\* = 42.618, df = 8, p-value = 1.037e-06  
##   
## Model df: 16. Total lags used: 24

shapiro.test(residuals(hawai\_CO2\_B))

##   
## Shapiro-Wilk normality test  
##   
## data: residuals(hawai\_CO2\_B)  
## W = 0.96958, p-value = 5.491e-07

En ce qui concerne l’analyse des résidus, il est possible d’affirmer qu’il n’y a pas de changements significatifs. Dans le premier graphique, on observe que les résidus sont constamment répartis autour de la valeur zéro. Alors que le graphique de normalité montre la possibilité que les résidus ne soient pas distribués de manière normale.

La valeur p du test Ljung-Box et le test de normalité des résidus ne montrent pas de changements significatifs. D’après le test de normalité, la valeur p = 5,491e-07 < 0,05 indique qu’il n’y a pas de distribution normale des résidus.

L’application de la transformation Box-Cox montre une légère amélioration de l’efficacité du modèle mais n’est pas significative, de sorte que le modèle initial pourrait être conservé pour être utilisé dans la projection.

Le modèle ETS s’adapte bien à la projection comme l’a prouvé la fonction de précision. Graphiquement il est possible de voir que les données réelles se trouvent dans la zone de projection du modèle, ce qui indique que la projection s’adapte bien aux données de formation pour projeter 30% des données avec une grande fiabilité.