

Concentration de CO2 dans l'atmosphère depuis 1958

OS

12 mai 2023

L'objectif :

Votre mission si vous l'acceptez : 1. Réalisez un graphique qui vous montrera une oscillation périodique superposée à une évolution systématique plus lente. 2. Séparez ces deux phénomènes. Caractérissez l'oscillation périodique. Proposez un modèle simple de la contribution lente, estimez ses paramètres et tentez une extrapolation jusqu'à 2025 (dans le but de pouvoir valider le modèle par des observations futures). 3. Déposer dans FUN votre résultat

Import données

Les données utilisées sont collectées par C. Keeling. À ma connaissance ce sont ces données qui ont permis de détecter la tendance à la hausse du niveau du CO2 dans l'atmosphère. Cette tâche est assez difficile, car le niveau de CO2 est très variable en fonction de l'endroit où l'on capte. Il a fallu ainsi un jeu de données propre mesuré dans des conditions identiques.

J'ai pu récupérer les données sur le site web. Les données sont sous un format CSV, cependant elles comportent une grosse en-tête avec 57 lignes de la documentation. Les noms des colonnes sont étalés sur plusieurs lignes. Ainsi, il est le plus simple de simplement renseigner les noms de colonnes à la main.

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.0
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(lubridate)

## Loading required package: timechange
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

co2_file <- "monthly_in_situ_co2_mlo.csv"
co2 <- read_csv(
  co2_file,
  skip = 60,
  col_names = c(
```

```

    "Yr", "Mn", "Date...3", "Date...4", "CO2...5", "seasonally...6",
    "fit", "seasonally...8", "CO2...9", "seasonally...10", "Sta")
)

## Rows: 792 Columns: 11
## -- Column specification -----
## Delimiter: ","
## chr (2): Mn, Sta
## dbl (9): Yr, Date...3, Date...4, CO2...5, seasonally...6, fit, seasonally.....
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

```

En plus des données brutes les diverses valeurs ajustées sont inclus (par exemple valeurs lises). Je m'intéresse uniquement à la série brute, qui est donnée à la fréquence mensuelle, au 15 du mois. Les valeurs manquantes sont indiqués par "-100".

```

co2 <- co2 %>%
  select(yr = Yr, mn = Mn, co2 = CO2...5) %>%
  mutate(
    date = ym(paste0(yr, mn)) + days(14),
    co2 = replace(co2, co2 < 0, NA)
  )

```

Visualisation

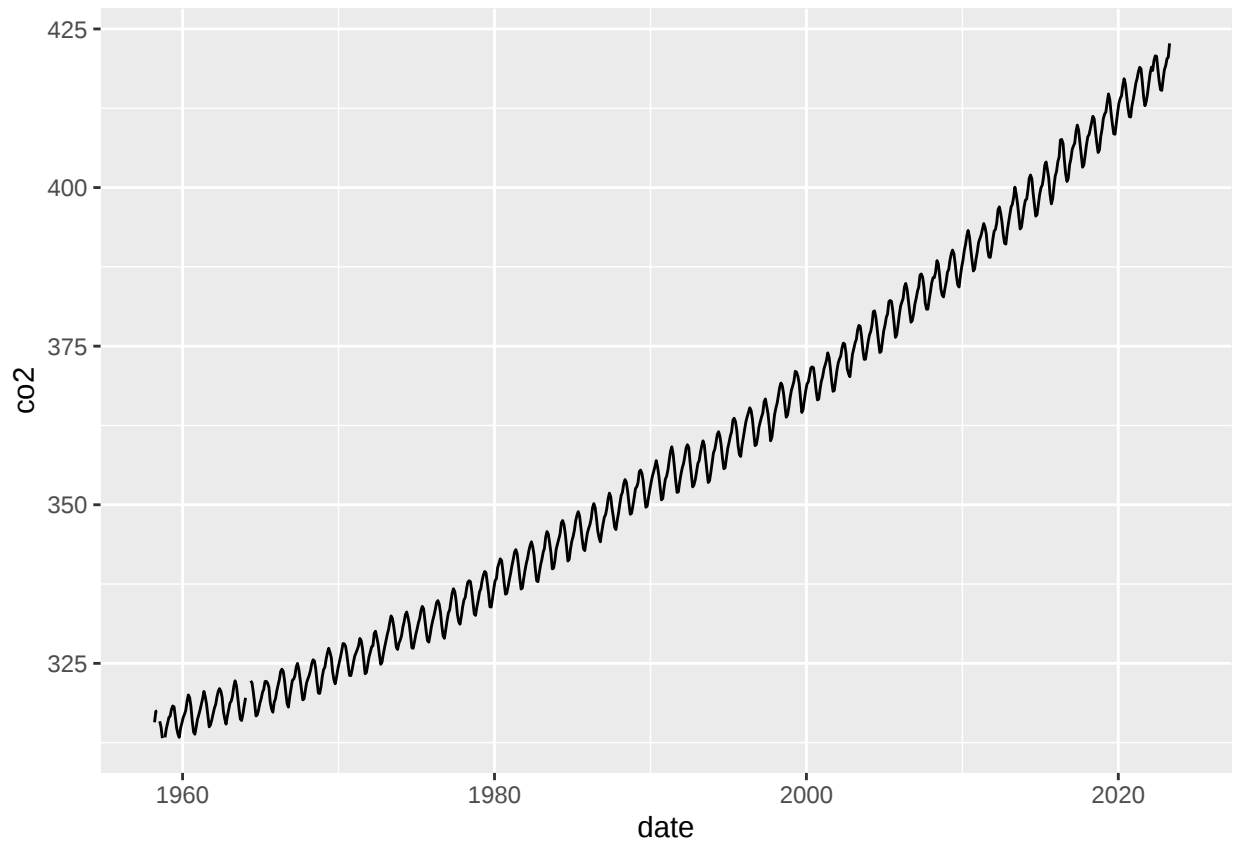
La tendance ainsi que le caractère périodique de la série sont assez évidents.

```

co2 %>%
  ggplot() +
  aes(x = date, y = co2) +
  geom_line()

```

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```

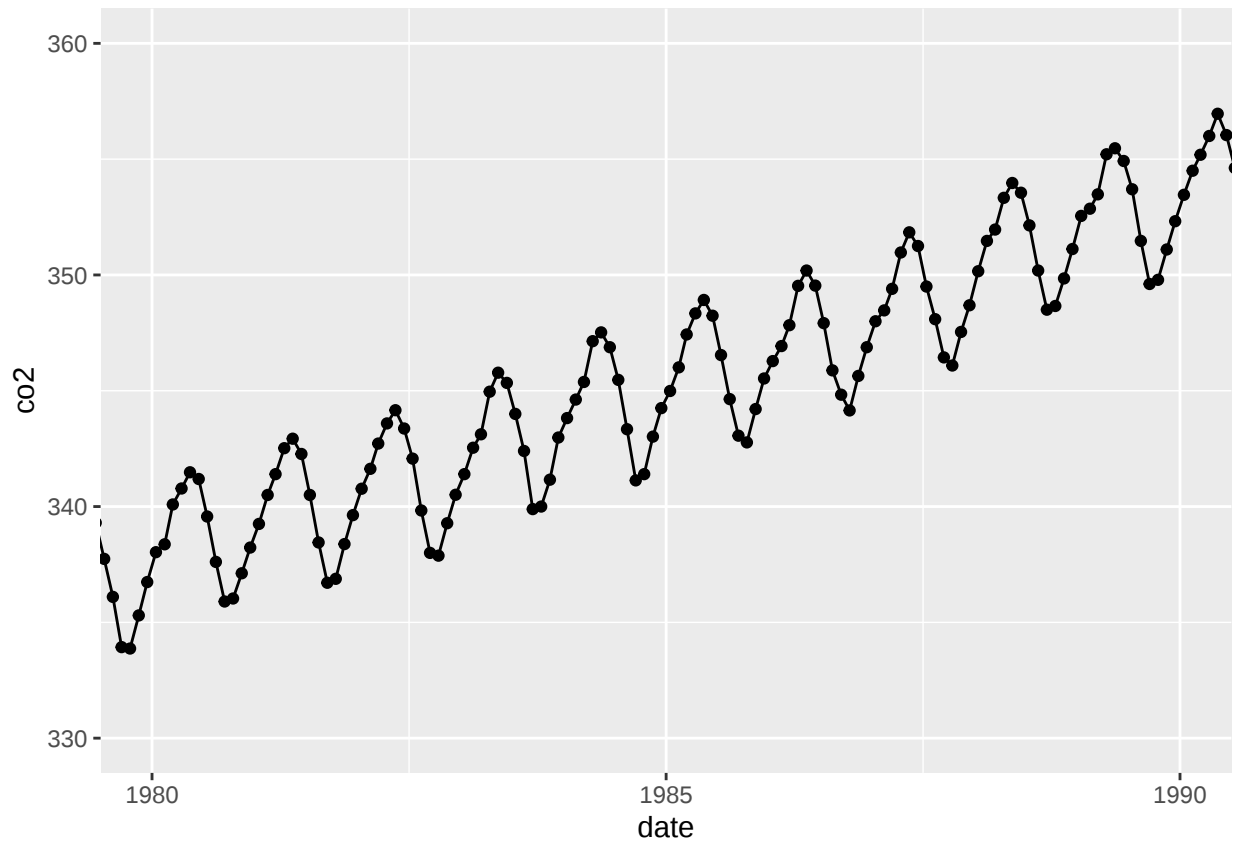


On peut néanmoins zoomer pour se donner une idée de la période de la série.

```
co2 %>%  
  ggplot() +  
  aes(x = date, y = co2) +  
  geom_line() +  
  geom_point() +  
  coord_cartesian(xlim = c(ymd("1980-01-01"), ymd("1990-01-01")), ylim = c(330, 360))
```

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```

```
## Warning: Removed 15 rows containing missing values (`geom_point()`).
```



Il apparait que la période est 12 mois, une année.

Analyse

Pour analyser ce phénomène, je propose une régression linéaire, avec deux effets :

1. Tendence linéaire dans le temps ;
2. effet de saisonnalité.

```
m1 <- lm(co2 ~ date + factor(month(date)), data = co2)
summary(m1)
```

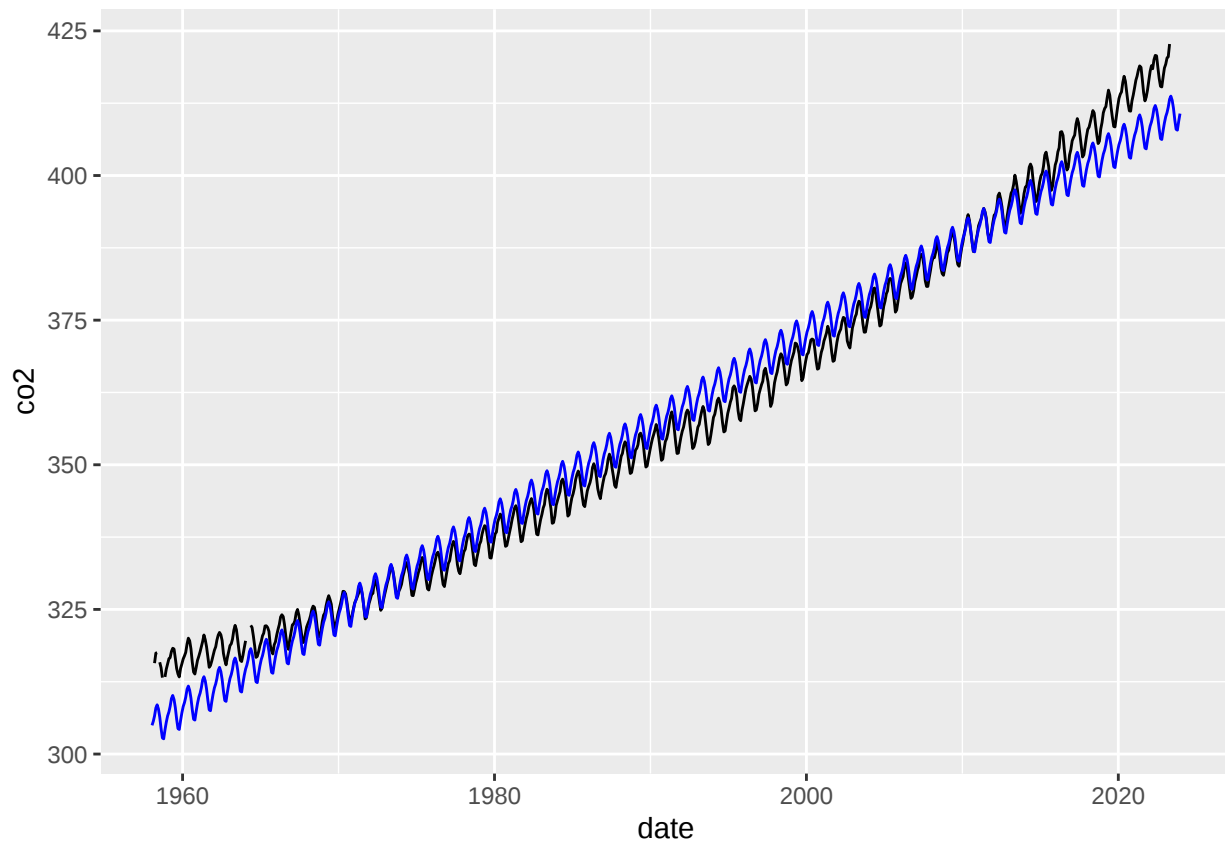
```
##
## Call:
## lm(formula = co2 ~ date + factor(month(date)), data = co2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.932 -3.296 -1.444  2.746 10.504
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.243e+02  5.621e-01  576.971  < 2e-16 ***
## date          4.432e-03  2.261e-05  196.027  < 2e-16 ***
## factor(month(date))2  5.743e-01  7.590e-01   0.757  0.44953
## factor(month(date))3  1.365e+00  7.560e-01   1.806  0.07132 .
## factor(month(date))4  2.609e+00  7.560e-01   3.451  0.00059 ***
```

```
## factor(month(date))5    3.015e+00  7.561e-01    3.987 7.33e-05 ***
## factor(month(date))6    2.157e+00  7.590e-01    2.842 0.00460 **
## factor(month(date))7    6.505e-01  7.561e-01    0.860 0.38985
## factor(month(date))8   -1.479e+00  7.561e-01   -1.956 0.05087 .
## factor(month(date))9   -3.264e+00  7.560e-01   -4.317 1.79e-05 ***
## factor(month(date))10  -3.545e+00  7.590e-01   -4.671 3.55e-06 ***
## factor(month(date))11  -2.125e+00  7.560e-01   -2.811 0.00507 **
## factor(month(date))12  -9.552e-01  7.560e-01   -1.263 0.20681
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.31 on 764 degrees of freedom
## (15 observations deleted due to missingness)
## Multiple R-squared:  0.9806, Adjusted R-squared:  0.9803
## F-statistic: 3218 on 12 and 764 DF, p-value: < 2.2e-16
```

Rien qu'avec ça on obtient un excellent fit.

```
co2 %>%
  mutate(prediction = predict(m1, .)) %>%
  ggplot() +
  aes(x = date, y = co2) +
  geom_line() +
  geom_line(aes(y = prediction), color = "blue")
```

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```



Le graphique met en évidence que la vraie tendance n'est pas tout à fait linéaire. Je propose de l'ajuster avec

un terme quadratique.

```
m2 <- lm(co2 ~ poly(date, 2) + factor(month(date))), data = co2)
summary(m2)
```

```
##
## Call:
## lm(formula = co2 ~ poly(date, 2) + factor(month(date))), data = co2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.74140 -0.54351 -0.09451  0.52804  2.06917
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    358.45624    0.09141 3921.544 < 2e-16 ***
## poly(date, 2)1    869.58437    0.75704 1148.659 < 2e-16 ***
## poly(date, 2)2    122.21499    0.76725  159.290 < 2e-16 ***
## factor(month(date))2    0.64982    0.12977   5.007 6.85e-07 ***
## factor(month(date))3    1.29102    0.12926   9.988 < 2e-16 ***
## factor(month(date))4    2.53394    0.12926  19.603 < 2e-16 ***
## factor(month(date))5    3.01139    0.12927  23.296 < 2e-16 ***
## factor(month(date))6    2.30138    0.12977  17.734 < 2e-16 ***
## factor(month(date))7    0.64927    0.12926   5.023 6.34e-07 ***
## factor(month(date))8   -1.47912    0.12926 -11.443 < 2e-16 ***
## factor(month(date))9   -3.26386    0.12926 -25.250 < 2e-16 ***
## factor(month(date))10  -3.40301    0.12977 -26.223 < 2e-16 ***
## factor(month(date))11  -2.12486    0.12926 -16.438 < 2e-16 ***
## factor(month(date))12  -0.95497    0.12926  -7.388 3.92e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7369 on 763 degrees of freedom
## (15 observations deleted due to missingness)
## Multiple R-squared:  0.9994, Adjusted R-squared:  0.9994
## F-statistic: 1.036e+05 on 13 and 763 DF, p-value: < 2.2e-16
```

```
co2 %>%
  mutate(prediction = predict(m2, .)) %>%
  ggplot() +
  aes(x = date, y = co2) +
  geom_line() +
  geom_line(aes(y = prediction), color = "blue")
```

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```

