

Traitement sujet 1 : Concentration de CO₂ dans l'atmosphère depuis 1958

David Pinaud

2023-09-12

Objectif

Il est de présenter un document reproductible réalisé selon les modes énoncés dans le MOOC. Le sujet choisi est le n°1 *Concentration de CO₂ dans l'atmosphère depuis 1958* avec des mesures de la concentration de CO₂ dans l'atmosphère effectuées à l'observatoire de Mauna Loa, Hawaï, États-Unis.

Les différents points à aborder pour l'exercice sont :

1. Réaliser un graphique qui montre une oscillation périodique superposée à une évolution systématique plus lente.
2. Séparer ces deux phénomènes, en caractérisant d'une part l'oscillation périodique et d'autre part en proposant un modèle simple de la contribution lente pour pouvoir ensuite proposer une extrapolation de cette dernière évolution jusqu'à 2025 (dans le but de pouvoir valider le modèle par des observations futures).

Données utilisées et importation

Les données sont disponibles sur le site internet de l'institut Scripps, on télécharge le fichier "weekly_in_situ_co2_mlo.csv" avec les observations hebdomadaires ici, jusqu'à la date du 2023-09-12 (date du téléchargement). Ce fichier est en copie locale et sert maintenant de base aux analyses.

```
data_url <-  
  "https://scrippsco2.ucsd.edu/assets/data/atmospheric/stations/in_situ_co2/weekly/weekly_in_situ_co2_mlo.csv"  
if(file.exists("weekly_in_situ_co2_mlo.csv"))  
{  
  data <- read.csv("weekly_in_situ_co2_mlo.csv")  
} else {  
  data <- read.csv(data_url, skip=44, header=FALSE, # on n'importe pas les 44emes premières lignes  
                  col.names=c("WeekDate", "CO2")) # colonnes "WeekDate" et "CO2".  
  write.csv(data, "weekly_in_situ_co2_mlo.csv", row.names=FALSE)  
}  
str(data)
```

```
## 'data.frame': 3332 obs. of 2 variables:  
## $ WeekDate: chr "1958-03-29" "1958-04-05" "1958-04-12" "1958-04-19" ...  
## $ CO2 : num 316 317 318 318 316 ...
```

```
range(data$WeekDate)
```

```
## [1] "1958-03-29" "2023-07-01"
```

Le fichier contient deux colonnes :

* WeekDate: la date de la semaine, ajustée à 12:00 pour le jour du milieu de chaque semaine

* CO2: la concentration en CO₂ en μ mole CO₂ par mole (ppm).

Mise en forme des données et vérification

La colonne `WeekDate` est en format `character`, nous allons la transformer en format `Date`, puis classer le `data.frame` selon cette variable par ordre croissant.

```
library(parsedate)
data$WeekDate <- as.Date(data$WeekDate)
data <- data[order(data$WeekDate),]
class(data$WeekDate)
```

```
## [1] "Date"
```

La colonne `WeekDate` est bien en format `Date`.

```
table(is.na(data$C02))
```

```
##
## FALSE
## 3332
```

Il n'y a pas de donnée manquante dans le tableau.

Représentation graphique

On représente sur un graphique la série en entier et un extrait sur 3 ans.

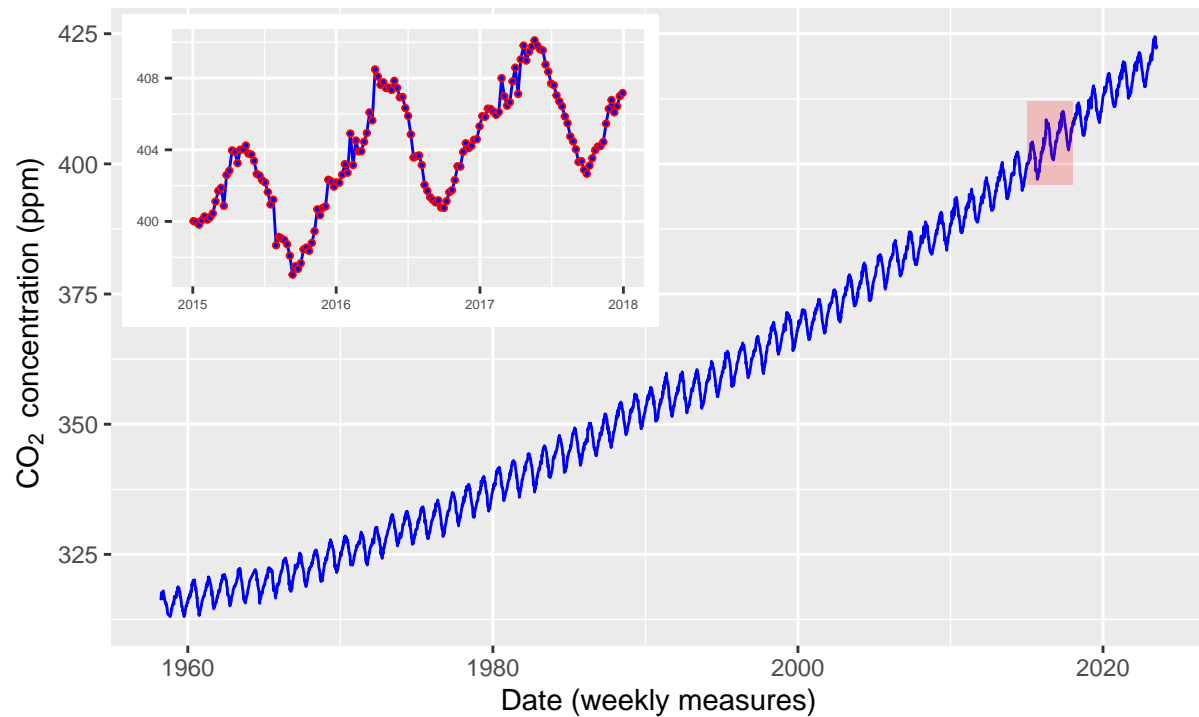
```
library(ggplot2)
library(patchwork)
p1 <- ggplot(data) +
  geom_line(aes(WeekDate, C02), col="blue") +
  labs(title=bquote("Atmospheric " ~CO[2]~" concentration"),
        subtitle = "Mauna Loa Observatory, Hawaii ; data from 1958 to 2023",
        x="Date (weekly measures)",
        y = bquote(~CO[2]~" concentration (ppm)"))
p1 <- p1 + annotate("rect", xmin=as.Date("2015-01-01"), xmax=as.Date("2017-12-31"),
                  ymin=396, ymax=412, fill="red", alpha=0.2)

subd <- data[data$WeekDate >= "2015-01-01" & data$WeekDate <= "2017-12-31",] # extrait sur 3 ans
p2 <- ggplot(subd) +
  geom_line(aes(WeekDate, C02), col="blue") +
  geom_point(aes(WeekDate, C02), size=0.8, fill="blue", shape=21, col="red") +
  labs(x=NULL, y = NULL) +
  theme(axis.text=element_text(size=5), axis.title=element_text(size=3))

p1 + inset_element(p2, 0.01, 0.5, 0.5, 0.99)
```

Atmospheric CO₂ concentration

Mauna Loa Observatory, Hawaii ; data from 1958 to 2023



Décomposition de la série

La série montre une oscillation périodique superposée à une évolution systématique plus lente. Nous allons caractériser ces deux phénomènes.

Tendance à long terme

On teste d'abord un ajustement linéaire (modèle le plus simple).

```
lm1 <- lm(CO2 ~ as.numeric(WeekDate), data)
plot(lm1, which=1)
```



```
knitr::kable(AIC(lm1, lm2))
```

	df	AIC
lm1	3	19898.32
lm2	4	14875.52

Le modèle avec une tendance quadratique s'ajuste bien mieux aux données.

```
knitr::kable(summary(lm2)$coefficients)
```

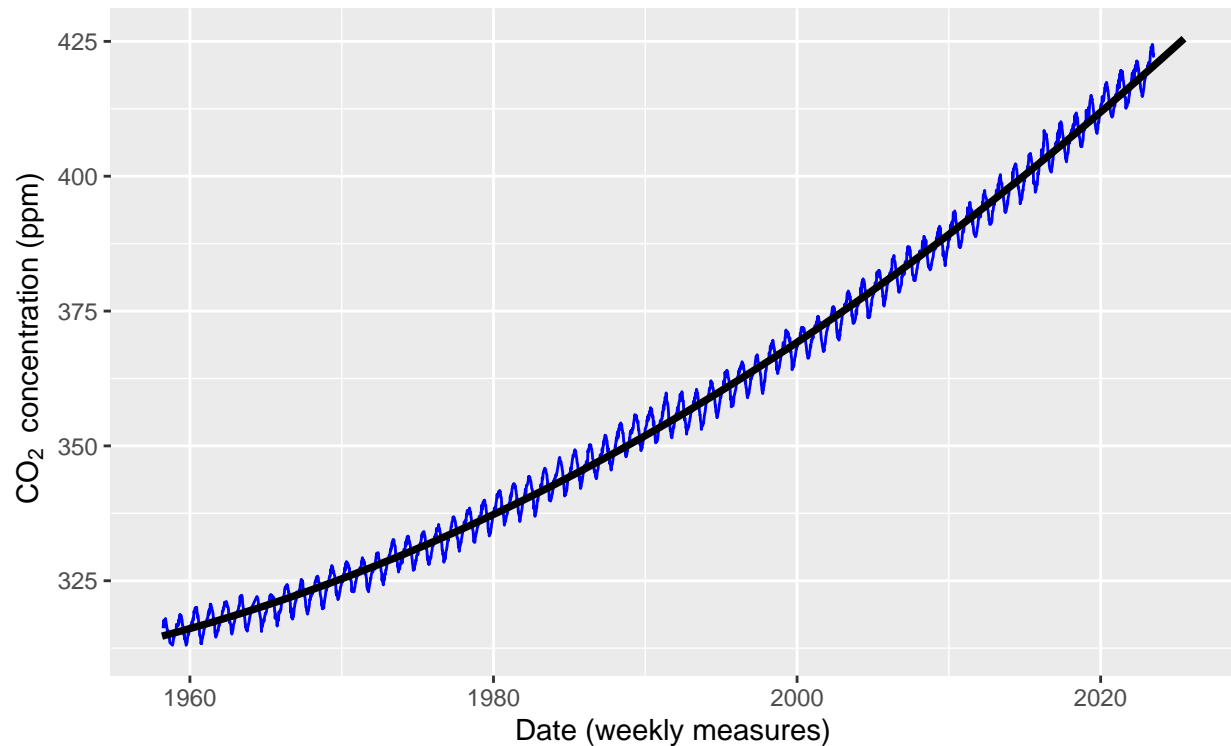
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	325.3651015	0.0604680	5380.7823	0
I(as.numeric(WeekDate)^2)	0.0000001	0.0000000	108.2175	0
as.numeric(WeekDate)	0.0028956	0.0000155	186.6996	0

On représente alors ce modèle avec une prédiction jusqu'en milieu d'année 2025.

```
newdata <- data.frame(WeekDate=seq.Date(from = as.Date("1958-03-01"),
                                         to = as.Date("2025-07-01"), by = "weeks"))
pred <- predict(lm2, newdata = newdata, se=TRUE)
pred <- cbind(newdata, as.data.frame(pred)[,1:2])
ggplot() +
  geom_line(data=data, aes(WeekDate, CO2), col="blue") +
  geom_line(data=pred, aes(WeekDate, fit), col="black", linewidth=1.3) +
  labs(title=bquote("Atmospheric " ~CO[2]~" concentration"),
       subtitle = "Mauna Loa Observatory, Hawaii ; data from 1958 to 2023",
       x="Date (weekly measures)",
       y = bquote(~CO[2]~" concentration (ppm)"))
```

Atmospheric CO₂ concentration

Mauna Loa Observatory, Hawaii ; data from 1958 to 2023



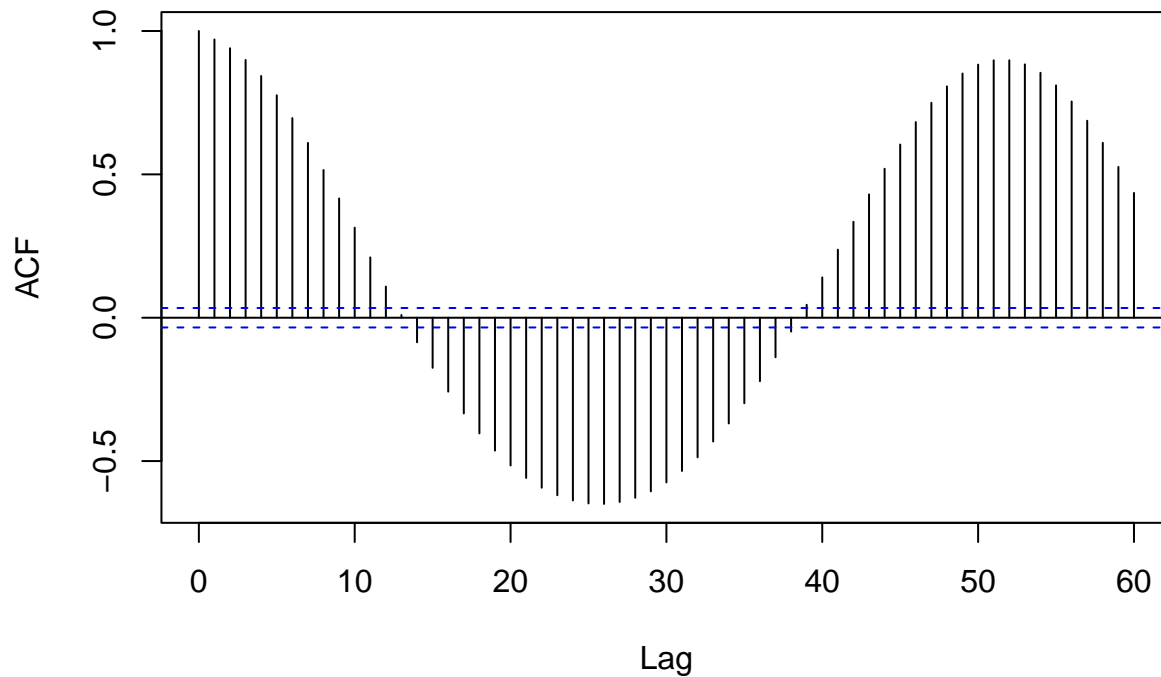
La concentration de CO₂ dans l'atmosphère prédite par ce modèle pour la semaine du 2025-06-28 est de 425.5 ppm \pm 0.26 (CI 95%).

Variations périodiques

Nous allons travailler sur les résidus du modèle quadratique pour analyser la périodicité des variations à plus fine échelle temporelle à l'aide d'une fonction d'autocorrelation.

```
data$resid.lm2 <- residuals(lm2, arg = "pearson")
acf(data$resid.lm2, lag.max=60, main = "Auto-correlation plot for residuals")
```

Auto-correlation plot for residuals



On observe une autocorrelation positive maximale autour du lag 52, soit 1 année (52 semaines). La périodicité de concentrations de CO_2 dans l'atmosphère est donc d'environ 1 an.

Conclusion

Les données de concentrations en CO_2 dans l'atmosphère mesurées à l'observatoire de Mauna Loa (Hawaii) de 1958 à 2023 montrent des variations importantes, avec une superposition de deux phénomènes : * une augmentation continue positive à long terme, avec une accélération constante. En 2025, la concentration devrait atteindre 425.5 ppm ;

* une périodicité de l'ordre d'un an, probablement expliquée par des variations saisonnières dans les émissions de CO_2 .