An Analysis of Atmospheric CO2 Concentrations

Emad Zadegan
emad.zadegan@mail.utoronto.ca
emadzadegan.github.io

2019-04-12

Abstract

We would like to investigate several hypotheses regarding the atmospheric carbon dioxide concetrations, using data [1] collected at an observatory in Hawaii. This data is summarized below in figures 1 and 2, which show the atmospheric carbon dioxide concentrations (ppm) respectively for all the recorded measurements (1960-present), and for recent times (2015-present), respectively. The overall long-term trend is clearly upward, and the concentration levels appear periodic. However, we are interested in making precise assessments of the following hypotheses:

I. The rate of increase of atmospheric CO2 concetrations is decreasing

II. The rate of increase of atmospheric CO2 concentrations has slowed down during the economic recessions of 1980-1982, 2008, as well as after the collapse of the Soviet Union in 1989

III. Atmospheric CO2 concentrations tend to be higher in October than in March

IV. There is a reasonable chance that carbon will exceed 430 parts per gallon by 2025

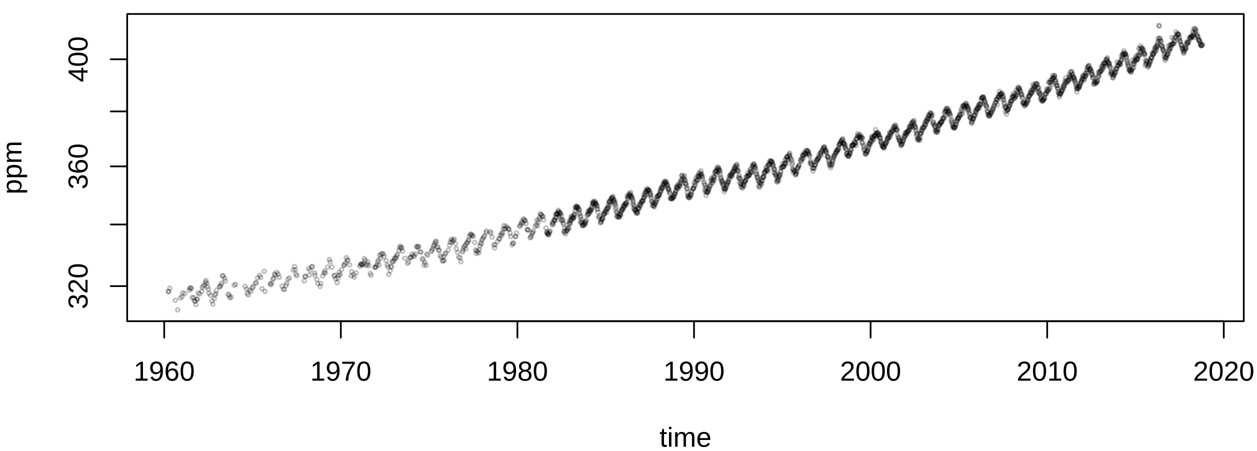

Figure 1: CO2 Concetrations

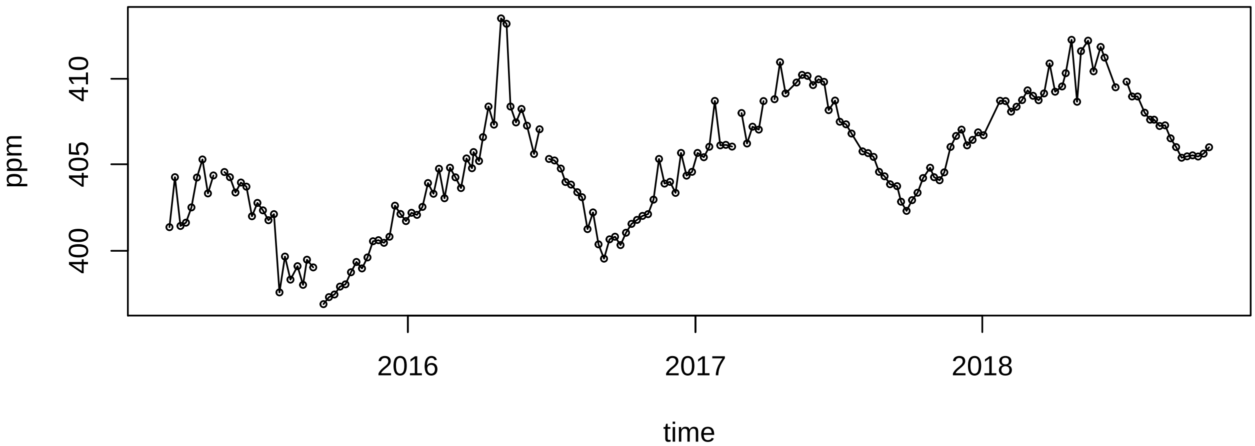

Figure 2: CO2 Concetrations in Recent Times

Methods

We use a GAM to model Y_i, the atmospheric CO2 concentration at time t_i, as follows:

$Y_i \sim N(\mu_i, \sigma^2)$ with

$$\mu_i = \beta_0 + \beta_1 \sin\left(2\pi(\frac{t_i}{365.25})\right) + \beta_2 \cos\left(2\pi(\frac{t_i}{365.25})\right) + f\left(t_i\right)$$

where $f(t)$ is a smooth function of time. Our use of the sine and cosine basis functions is motivated by the periodic nature of atmospheric CO2 concentrations. And furthermore, our model choice is validated by the QQ-plot of the residuals in figure 3, below.

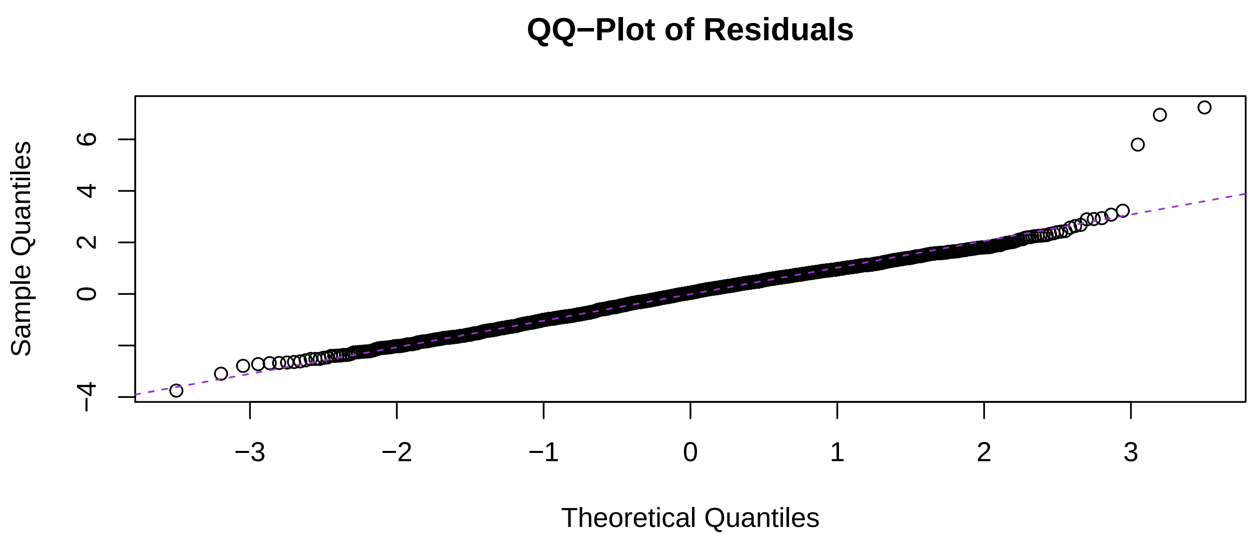

Figure 3: QQ-Plot of residuals from the GAM

Results

To examine hypotheses **I** and **II**, we estimate the derivative of atmospheric CO2 concentration. The results are summarized in figure 4. It is clear that the derivative's upward trend ended sometime in 2016 and has remained flat until today; thus the rate of increase of CO2 concentrations has slowed somewhat recently. Furthermore, we see that the derivative's upward trend remained intact through both economic recessions of 1980-1982, and the more recent one, which started in December 2007 and ended in June 2009; thus the rate of increase of CO2 concentrations did not slow during these economic recessions. However, we see that the collapse of the Soviet Union began in 1989, while the CO2 derivative was in a downtrend, and the official dissolution of the Soviet Union (the day on which the Soviet Union granted self-governing independence to the Republics of the USSR), December 26th, 1991, coincides with a local bottom in the derivative of the atmospheric CO2 concentrations. Therefore, the rate of increase of CO2 concentrations was decreasing at the beginning of the collapse, but it began to increase immediately after the conclusion of the collapse.

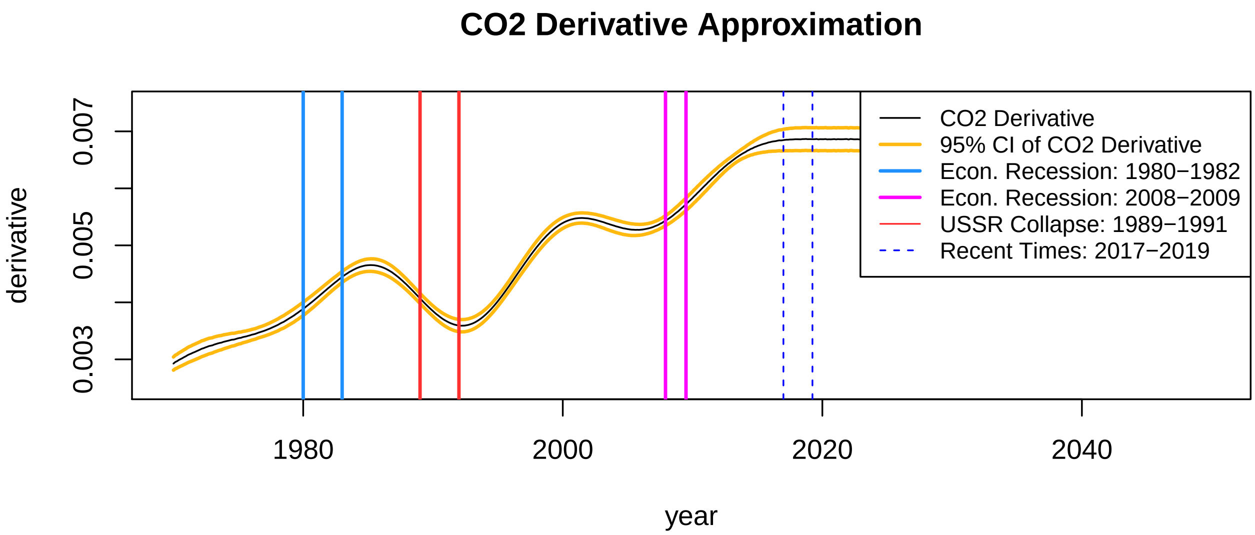

Figure 4: Approximation of atmospheric CO2 concentration using finite differences

To examine hypothesis **III** we look at the monthly average of CO2 concentrations. The results are summarized in figure 5. The lower bound of the confidence intervals for March is higher than the upper bound of the confidence interval for October. Therefore, it is true that atmospheric CO2 concentrations tend to be higher in March than October.

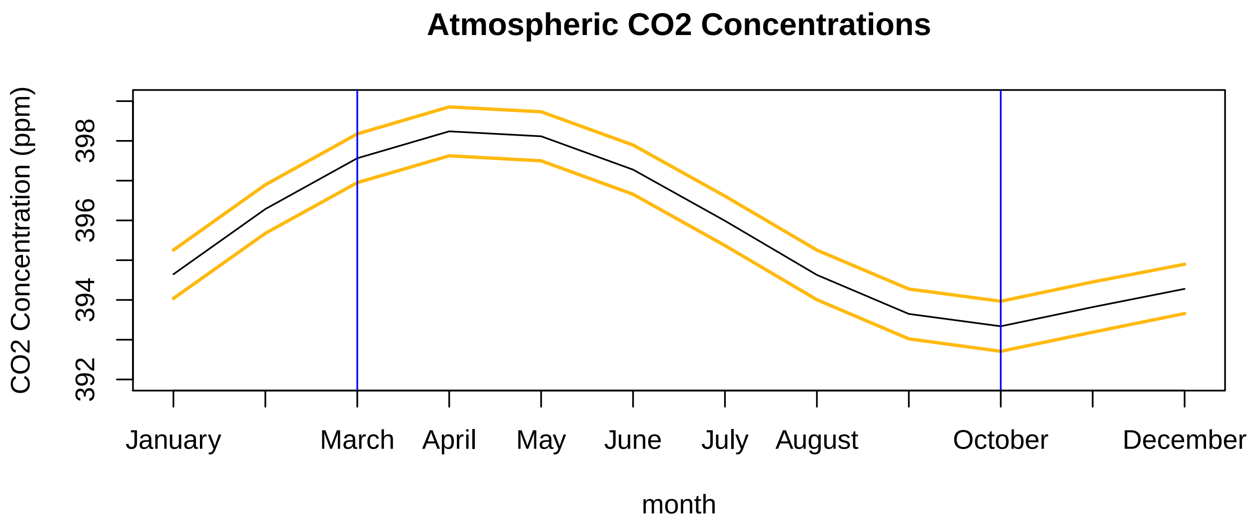

Figure 5: Monthly average of atmospheric CO2 concentrations (black line) with confidence intervals (yellow lines)

Examining hypothesis **IV** invovles forcasting atmospheric CO2 concentration for year 2025. Using the GAM, we make forecasts for up to the year 2030. The results are summarized below in figure 5. The black line represents past measurements and future forecasts, the yellow lines are the 95% prediction intervals. The vertical blue line is drawn on January 1st, 2025, and the horizontal blue line is drawn at 430 ppm. We see that at the year 2025, the upper bound of the CI is below the 430ppm mark. Therefore, the atmospheric CO2 concentrations will not exceed 430 ppm by the year 2025.

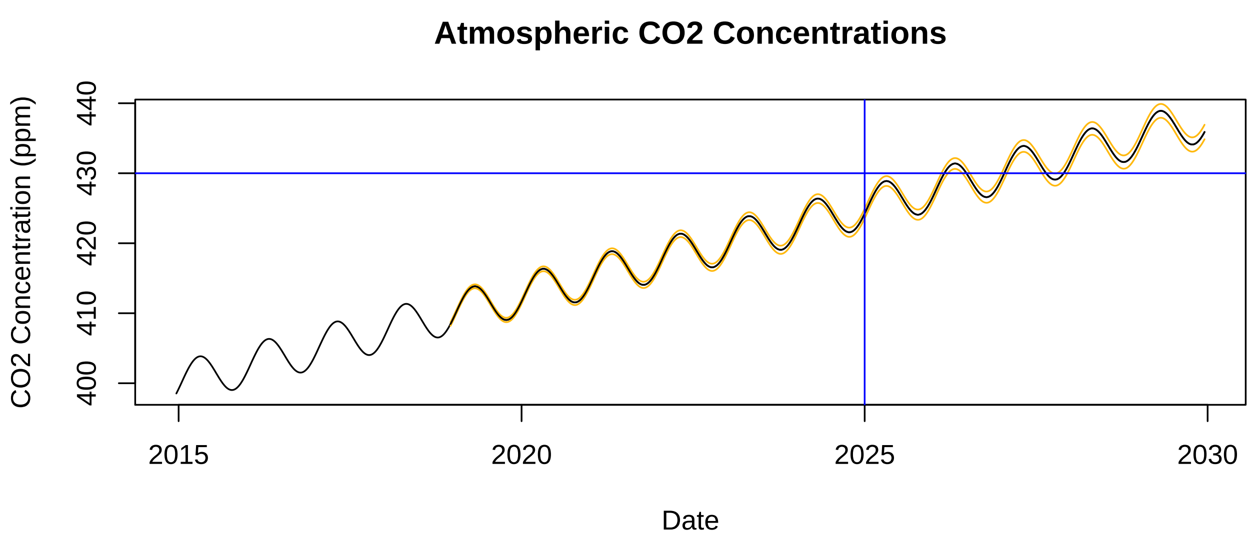

Figure 6: Past measurements and future forecasts of atmospheric CO2 concentration (black line) with confidence intervals (yellow lines)

Appendix

Data

[1] http://scrippsco2.ucsd.edu/assets/data/atmospheric/stations/flask_co2/daily/daily_flask_co2_mlo.csv

Code

```
cUrl = paste0("http://scrippsco2.ucsd.edu/assets/data/atmospheric/",
              "stations/flask_co2/daily/daily_flask_co2_mlo.csv")
cFile = basename(cUrl)
if (!file.exists(cFile)) download.file(cUrl, cFile)
co2s = read.table(cFile, header = FALSE, sep = ",", skip = 69,
                  stringsAsFactors = FALSE, col.names =
                    c("day", "time","junk1", "junk2", "Nflasks", "quality", "co2"))
co2s$date = strptime(paste(co2s$day, co2s$time), format = "%Y-%m-%d %H:%M",
                     tz = "UTC")
# remove low-quality measurements
co2s[co2s$quality > 2, "co2"] = NA

plot(co2s$date, co2s$co2, log = "y", cex = 0.3, col = "#00000040",
     xlab = "time", ylab = "ppm")

plot(co2s[co2s$date > ISOdate(2015, 3, 1, tz = "UTC"),
          c("date","co2")], log = "y", type = "o", xlab = "time", ylab = "ppm",
     cex = 0.5)

timeOrigin = ISOdate(1980, 1, 1, 0, 0, 0, tz = "UTC")
co2s$days = as.numeric(difftime(co2s$date, timeOrigin, units = "days"))
co2s$cos12 = cos(2 * pi * co2s$days/365.25)
co2s$sin12 = sin(2 * pi * co2s$days/365.25)
# co2s$cos6 = cos(2 * 2 * pi * co2s$days/365.25)
# co2s$sin6 = sin(2 * 2 * pi * co2s$days/365.25)

library(mgcv)
gam_fitted <- gam(co2 ~ s(days) + cos12 + sin12 , data = co2s) ; # summary(cLm)

qqnorm(gam_fitted$res, main="QQ-Plot of Residuals")
qqline(gam_fitted$res, lty=2, col="darkorchid")


total_number_of_predictions=365*80

x.mesh <- seq(ISOdate(1970, 1, 1, 0, 0, 0, tz = "UTC"),
              by = "1 days",length.out = total_number_of_predictions) ## where to evaluate

days.mesh=as.numeric(difftime(x.mesh, timeOrigin, units = "days"))

newd=data.frame(date=x.mesh,
                days=days.mesh,
                cos12=cos(2*pi*days.mesh/365.25),
                sin12=sin(2*pi*days.mesh/365.25))

coPred <- predict(gam_fitted, newd, se.fit = TRUE) ;
coPred <- data.frame(est = coPred$fit,
                     lower = coPred$fit - 2*coPred$se.fit, upper = coPred$fit + 2*coPred$se.fit)

##find the derivatives
eps=1e-7 ## finite difference interval
```

```
p0 = predict(gam_fitted, newd, type="lpmatrix")
p1 = predict(gam_fitted, newd+eps, type="lpmatrix")
Xp = (p1-p0)/eps ## maps coefficients to (fd approx.) derivatives
Xi <- Xp*0
Xi[,4:12] <- Xp[,4:12] ## Xi%*%coef(b) = smooth derivative
df <- Xi%*%coef(gam_fitted) ## ith smooth derivative
df.sd <- rowSums(Xi%*%gam_fitted$Vp*Xi)^.5 ## cheap diag(Xi%*%b$Vp%*%t(Xi))^.5


par(mfrow=c(1,1))
split=50 # plot every 50 for stability
plot(newd$date[seq(1,total_number_of_predictions,split)],
     df[seq(1,total_number_of_predictions,split)], xlab="year", ylab="derivative",
     main="CO2 Derivative Approximation", type="l", ylim=c(0.0025,0.0075))
lines(newd$date[seq(1,total_number_of_predictions,split)],
      df[seq(1,total_number_of_predictions,split)]+
        2*df.sd[seq(1,total_number_of_predictions,split)],lty=1, lwd=2, col="darkgoldenrod1")
lines(newd$date[seq(1,total_number_of_predictions,split)],
      df[seq(1,total_number_of_predictions,split)]-
        2*df.sd[seq(1,total_number_of_predictions,split)],lty=1, lwd=2, col="darkgoldenrod1")

# abline(v=ISOdate(2019, 3, 29, 0, 0, 0, tz = "UTC"), lty=2, col="green") # current
abline(v=ISOdate(2019, 3, 29, 0, 0, 0, tz = "UTC"),
       lty=2, lwd=1, col="blue") # present
abline(v=ISOdate(2017, 1, 1, 0, 0, 0, tz = "UTC"),
       lty=2, lwd=1, col="blue") # when the flattening began

abline(v=ISOdate(1980, 1, 1, 0, 0, 0, tz = "UTC"),
       lty=1, lwd=2, col="dodgerblue") # economic recession start: 1980
abline(v=ISOdate(1983, 1, 1, 0, 0, 0, tz = "UTC"),
       lty=1, lwd=2, col="dodgerblue") # economic recessions end: 1982

abline(v=ISOdate(2007, 12, 1, 0, 0, 0, tz = "UTC"),
       lty=1,lwd=2,  col="magenta") # economic recession start: December 2007
abline(v=ISOdate(2009, 6, 30, 0, 0, 0, tz = "UTC"),
       lty=1,lwd=2,  col="magenta") # economic recession end: June 2009

abline(v=ISOdate(1989, 1, 1, 0, 0, 0, tz = "UTC"),
       lty=1, lwd=2, col="firebrick1") # collapse of soviet union: start 1989
abline(v=ISOdate(1992, 1, 1, 0, 0, 0, tz = "UTC"),
       lty=1, lwd=2, col="firebrick1") # collapse of soviet union: December 26, 1991


legend("topright", legend=c("CO2 Derivative",
                            "95% CI of CO2 Derivative",
                            "Econ. Recession: 1980-1982",
                            "Econ. Recession: 2008-2009",
                            "USSR Collapse: 1989-1991",
                            "Recent Times: 2017-2019"),

       col=c("black", "darkgoldenrod1","dodgerblue",
             "magenta" ,"firebrick1","blue"), lty=c(rep(1,5),2,2),
       lwd=c(1,rep(2,3),1,1), cex=0.8,bg="white")
```

```
agg=aggregate(coPred, list(month=month(x.mesh,label=TRUE,abbr=FALSE)),
              mean)

plot(1:12,agg$est,type='l',xaxt='n', ylim=c(392,399), xlab='month',
     ylab="CO2 Concentration (ppm)")

title("Atmospheric CO2 Concentrations")

lines(1:12, agg$lower,col='darkgoldenrod1', lwd=2)
lines(1:12, agg$upper, col='darkgoldenrod1', lwd=2)
abline(v=3,col='blue')
abline(v=10,col='blue')

axis(side=1,labels=agg$month,at=1:12)


plot(newd$date[seq(365*45,365*60)], coPred$est[seq(365*45,365*60)],
     type = "l", xlab="Date", ylab="CO2 Concentration (ppm)",
     main="Atmospheric CO2 Concentrations")
matlines(as.numeric(newd$date[seq(365*49,365*60)]),
         coPred[seq(365*49,365*60), c("lower", "upper","est")],
         lty = c(1, 1, 1), col = c("darkgoldenrod1", "darkgoldenrod1", "black"))
abline(v=ISOdate(2025, 1, 1, 0, 0, 0, tz = "UTC"), lty=1, col="blue")
abline(h=430, lty=, col="blue")
```