STA442 Assignment 3

Yiwen Yang (#1004244800)

Nov.13^{th}, 2019

1. Dilemma Between Industrialization and Emission of CO_2

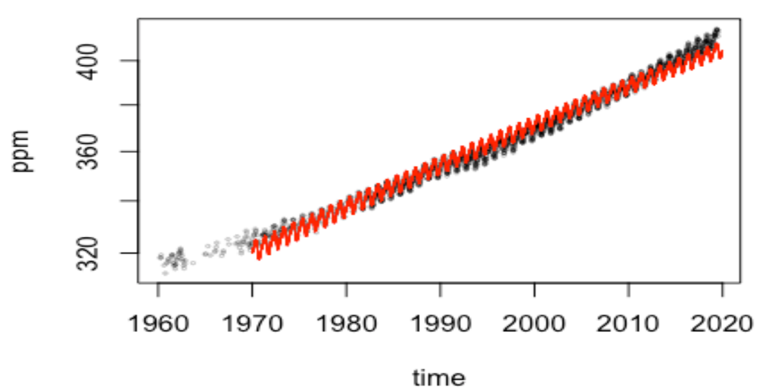
Setup of laboratory

We have measured the atmospheric Carbon Dioxide concentrations from an observatory in Hawaii, made available by the Scripps CO Program (http://scrippsco2.ucsd.edu/assets/data/atmospheric).

Our interest is to find out the impact of some well-known events from 1973 to 2015 to the level of CO2, its growth rate, trend, and make future predictions. The activities can be generalized by three categories; change in industrial production (fall of Berlin wall and China joining the WTO), economic recession or financial crisis (OPEC oil embargo, 1980 global economic recession, the bankruptcy of Lehman Brothers) and agreement on limiting the emission of CO2 (signing of Paris agreement).

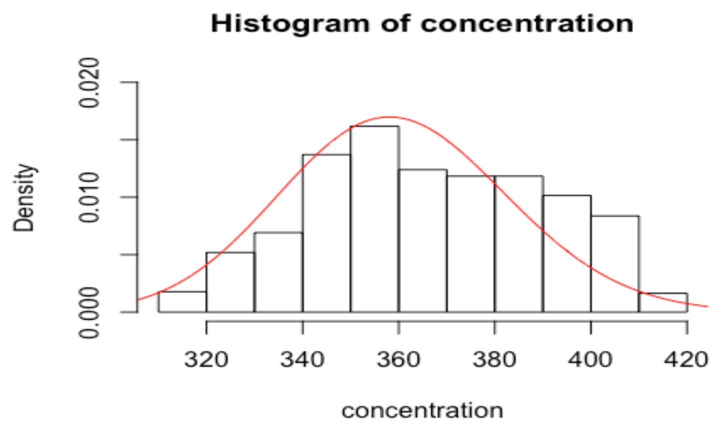
Model and Interpretation

Plot 1 (Assessing the fit of linear model)*



We first fitted a linear model with the covariate of some cyclical functions of year $\boldsymbol{x_i}$. As we can see from the graph below, the red line represents the predicted value from year 1970, the black dotted lines indicate observed values. The fit is not too bad.

Plot 2 (histogram of concentration of CO_2)*



We then fitted the non-parametric model using INLA from the gamma family with a log link to predict the trend of CO_2 from year 1970. From the above graph, the level of CO_2 roughly follows a gamma distribution. The reason why we use this model is that sin and cos functions are cyclical, they are seasonal predictors that can predict trend of level of CO_2 annually and semiannually *(please refer to the seasonal cycle in the appendix)*.

Down below is our model.

$$\boldsymbol{log(E(y_i)) = \sum_{j=1}^{4} \varphi_j(x_i)\beta_j + U(t_i) + V_i}$$

Model Assumptions:

$$\boldsymbol{y_i \sim \Gamma(\theta)}$$

$$\boldsymbol{[U_1 \ldots U_T]^T \sim RW2(0, \sigma_U^2)}$$

$$\boldsymbol{V_i \sim N(0, \sigma_V^2)}$$

$$\boldsymbol{\varphi_1(x_i) = cos\,(2\pi x_i),\ \varphi_2(x_i) = sin\,(2\pi x_i),\ \varphi_3(x_i) = cos\,(4\pi x_i),}$$

$$\boldsymbol{\varphi_4(x_i) = sin\,(4\pi x_i)}$$

y_i is the concentration of CO_2 at year x_i since 1970.

φ_1 *and* φ_2 are annual fluctuations; φ_3 *and* φ_4 are semiannual fluctuations.

$U(t_i)$ is a second-order random walk where its second derivative are *$N(0, \sigma_U^2)$*. We will treat it as a random slope effect.

V_i covers independent variation or over-dispersion.

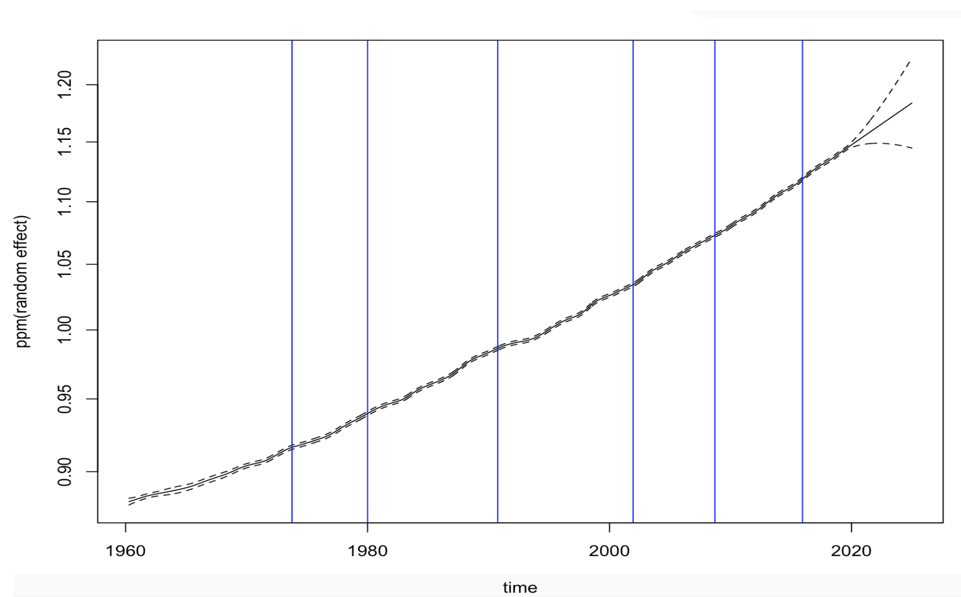
Priors:

We set two penalized complexity priors for the model, ***pc. prec (a, b)*** can be interpreted as $\boldsymbol{P(\sigma > a) = b}$. It is equivalent to $\boldsymbol{\sigma \sim exp(-log(b)/a)}$.

$$\mathbf{P}(\boldsymbol{\sigma_U} > \boldsymbol{log}\ (\mathbf{1.01})/\mathbf{26})) = \mathbf{0.5} \Rightarrow \boldsymbol{\sigma_U} \sim \boldsymbol{exp\ (1811)}$$ (from biweekly data over a year, slope change by 1%)

$$\mathbf{P}(\boldsymbol{\sigma_V} > \mathbf{2}) = \mathbf{0.5} \Rightarrow \boldsymbol{\sigma_V} \sim \boldsymbol{exp\ (0.35)}$$

Plot 3 (trend of random effect)*

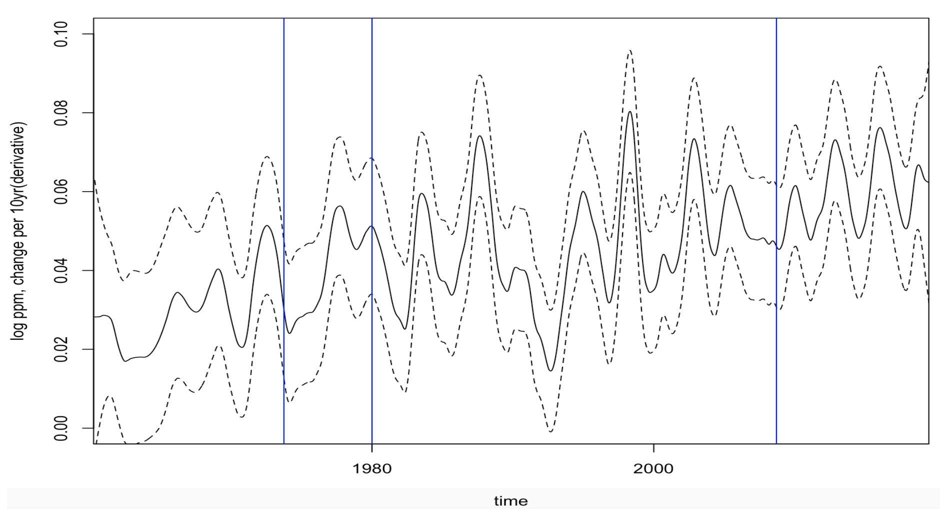


Since sum of sin, cos functions and V_i is still a cyclical line, the general trend of CO_2 is predominantly determined by the random slope factor, $U(t_i)$. We can then observe this random effect visually to figure out whether the slope becomes steeper or shallower when the six events happen. The blue lines indicate the year when six events occur. From this graph, the slope generally becomes steeper as time passes. The slope becomes shallower (concave down) in 1973,1980, 1989, 2008, and 2015 which corresponds to OPEC oil embargo, global economic recession, fall of Berlin Wall, bankruptcy of Lehman Brothers and signing of Paris Agreement respectively. When China joined the WTO in 2001, the slope appears to be steeper (concave up). However, the rate of growth is hard to observe. We need a more accurate plot to depict the actual change in increase in CO_2 level.

First - Order Derivative

We estimated the first-order derivative of the smoothed trend of CO_2 using finite difference approximation ($lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$). The blue line corresponds to the derivative at a specific year. We are going to analyze the events by group, in particular, change in industrial production, financial crisis or recession and signing agreement on the emission of CO_2.

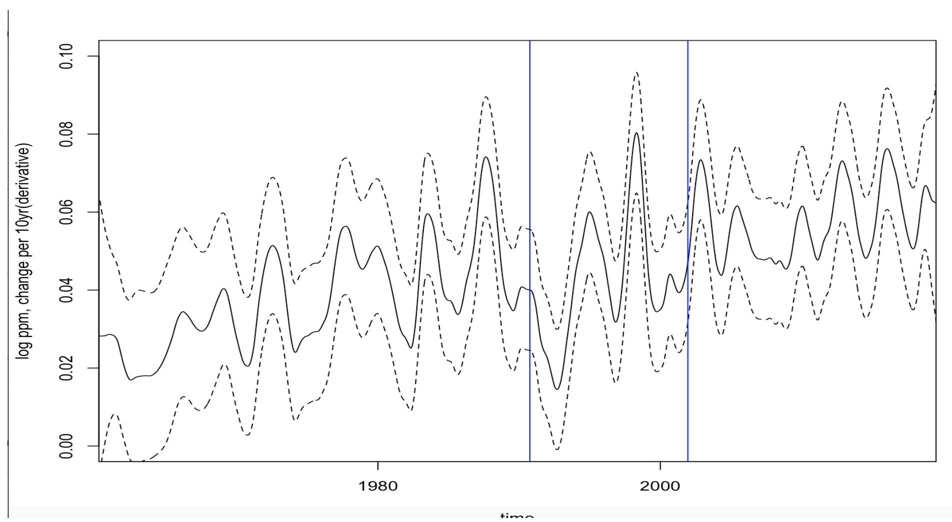
Plot 4 (first order derivative of CO_2 trend related to financial crisis or recession)*



The black line is the estimated slope for each consecutive year, and the dotted lines are credible intervals for every estimate. The blue lines in the above graph show the change in concentration of CO_2 for three events. From left to right, the **OPEC oil embargo in 1973**, the global **economic recession from 1980 to 1982**, and the **bankruptcy of Lehman Brothers in 2008**.

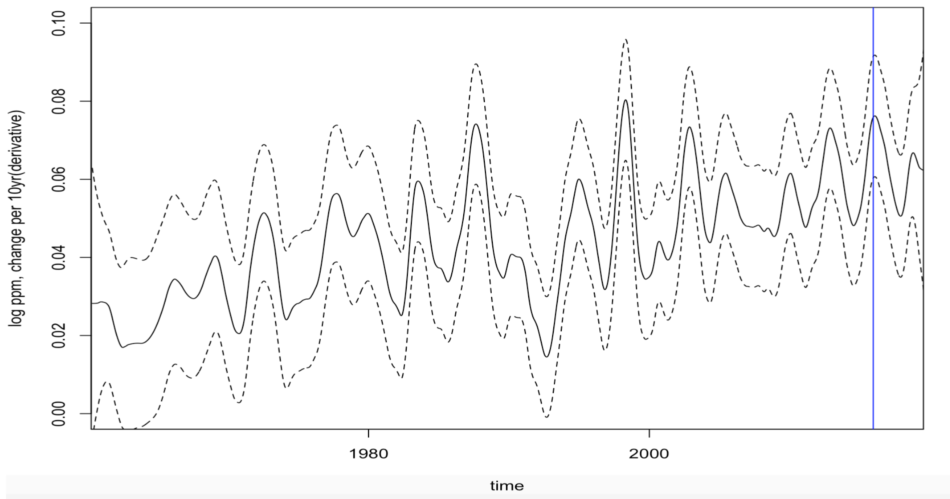
At the end of 1973, OPEC decided to stop exporting oil to the U.S and has an enormous negative effect on production in the states. The CO_2 level falls from 0.05, which expected to be between 0.03 and 0.07 to 0.03 in 1975. Around 1974, the rate of growth recovers. The 1980-1982 recession has a huge negative impact on global production; thus, emitting less CO_2. The level changes from 0.05 (***CI is (0.033,0.07)***) in 1980 to 0.025 (***CI is (0.01, 0.042)***) in 1982. Last but not least, Lehman Brother's bankruptcy in 2008 is a consequence of the recession at the time. The reason why the CO_2 level remained around 0.042 (***CI is (0.03,0.06)***) after the bankruptcy is that it mainly affects the financial market in the states. It hasn't interfered the production market significantly.

Plot 5 (first order derivative of CO_2 trend related to change in industrialization)*



The two events from left to right are the **fall of Berlin Wall happened in 1989**, and **China joins WTO on Dec.11th, 2001**. The fall of Berlin wall results in a dramatic fall in industrialization in the Soviet Union and Eastern Europe because eastern German no longer belongs to the Soviet Union, and the economy merely collapsed, so the production falls, emit less CO_2. The slope increases at a much lower rate between 1989 and 1994, roughly from 0.04 (***CI is (0.021,0.056)***) to 0.018 (***CI is (0.00, 0.03)***). In 2001, China opened its market after joining the WTO. As a large manufacturing country, China had rapid growth in industrial production, but the CO_2 emission grows tremendously as well. After 2001, the concentration had been increasing at a faster rate until approximately 2003. The estimated slope is from 0.045 (***CI is (0.03,0.06)***) to 0.07 (***CI is (0.058,0.09)***).

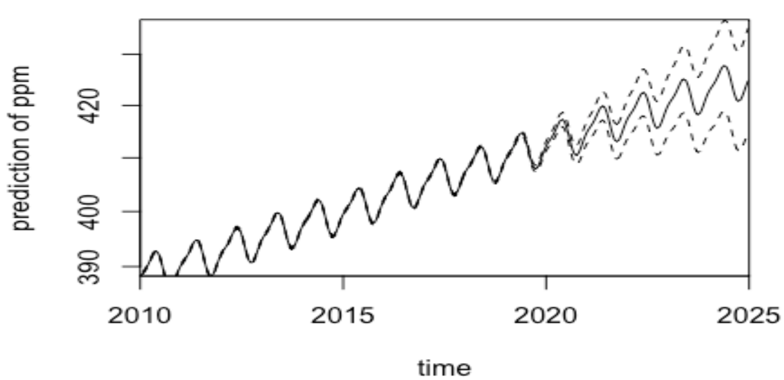
Plot 6 (first order derivative of CO_2 trend after signing the Paris Agreement)*



The blue lines indicate the **signing of the Paris Agreement on Dec.12th, 2015**, which intended to limit the emission of CO_2. This agreement is quite adequate, the growth rate is lower now that it has been in the recent past, the slope is decreased from 0.07 (***CI is (0.06,0.09)***) to approximately 0.05 (***CI is (0.037,0.062)***) within a year.

Predictions of CO_2

Plot 7 (Prediction of CO_2 level and associated credible interval)*



The graph above is the predicted CO_2 level until 2025. The concentration of carbon dioxide is increasing, and the credible interval associated with it becomes wider as time progresses. The expected value is between 410 and 435 parts per gallon in the year 2025.

Conclusion

We have analyzed the trend, growth rate of level of CO_2 and have made predictions for it in the future. From the above analysis, the development of a country and its emission of CO_2 is directly proportional. Furthermore, economists say there will be a recession every five years, in these cases, CO_2 concentration has a slight reduction, but soon recovers to the growth rate as it has before. Though every country needs development, develop in an eco-friendly way is essential.

Appendix

Please refer to the R-code at the very last several pages of this document.

2. Analysis of Heat

Summary of Result

We have analyzed Sable Island's temperature data from 2016 to present. We are keen to determine whether IPCC's statement about global warming is true or not. From the analysis, after the pre-industrial period, human activity caused 0.9°C of global warming, with a likely range of 0.7°C to 1.4°C; while from 2030-2052, global warning reach approximately with a range of 1.1°C to 5°C if we assume the rate of increase is unchanged. The result is roughly consistent with IPCC's statement.

Introduction

We interpreted the temperature data recorded on Sable Island. The R version of the data is named as 'sableIsland.rds', which can be downloaded from http://pbrown.ca/teaching/appliedstats/data/sableIsland.rds page. Our primary focus is to find out the accuracy of IPCC's statement, which can be found on http://www.ipcc.ch/sr15/resources/headline-statements. It can be summarized as human activities have caused 1°C of global warming above the pre-industrial period level. Global warming can reach 1.5°C between 2030 and 2052 if it keeps increasing at the current rate.

Methods and Interpretation

We fitted a non-parametric model using INLA from the t family with identity link to predict the trend of global warming. The reason why we include sin and cos functions as covariates is that they are cyclical, they can predict the temperature of the island annually and semiannually.

Model Assumptions:

$$\sqrt{s\,\tau}\,(y - \eta) \sim T_{\nu}$$

for response y that is continuous where

τ: is the precision parameter

s: is a fixed scaling, s > 0

η: is the linear predictor,

$$\boldsymbol{\eta = \sum_{j=1}^{4} \varphi_j(x_i)\beta_j + U(t_k) + V_k + Z_l}$$

$$[\boldsymbol{U_1 \dots U_T}]^T \sim RW2(0, \sigma_U^2)$$

$$\boldsymbol{V_k} \sim N\,(0,\, \sigma_V^2)$$

$$\boldsymbol{Z_l} \sim N\,(0,\, \sigma_Z^2)$$

$$\boldsymbol{\varphi_1(x_i) = cos\,(2\pi x_i),\ \varphi_2(x_i) = sin\,(2\pi x_i),\ \varphi_3(x_i) = cos\,(4\pi x_i),}$$

$$\boldsymbol{\varphi_4(x_i) = sin\,(4\pi x_i)}$$

y_{ikl}: is the highest temperature in Sable Island at i^{th} day, k^{th} week, l^{th} year since 1880.

φ_1 *and* φ_2 are annual fluctuations; φ_3 *and* φ_4 are semiannual fluctuations.

$U(t_k)$ is a second-order random walk where its second derivative are $N\,(0, \sigma_U^2)$, we treat it as a random slope. Since it follows random walk 2, we have no intercept.

V_k is the random effect for k^{th} week, Z_l is the random effect for l^{th} year.

T_v: is a reparametrized standard Student - *t* with v > 2 degree of freedom with unit variance for all values of v.

Priors:

$$\mathbf{P}\left(\boldsymbol{\sigma_U > \frac{0.1}{52*100}}\right) = \mathbf{0.05} \Rightarrow \boldsymbol{\sigma_U \sim exp\ (36043)}$$ (weekly data over a hundred years, the slope should change by 10%)

$$\mathbf{P}(\boldsymbol{\sigma_V} > \mathbf{1}) = \mathbf{0.5} \Rightarrow \boldsymbol{\sigma_V \sim exp\ (0.693)}$$

$$\mathbf{P}(\boldsymbol{\sigma_Z} > \mathbf{1}) = \mathbf{0.5} \Rightarrow \boldsymbol{\sigma_Z \sim exp\ (0.693)}$$

$$\mathbf{P}(\boldsymbol{\tau} > \mathbf{1}) = \mathbf{0.5} \Rightarrow \boldsymbol{\tau \sim exp\ (0.693)}$$

$$\boldsymbol{P\ (v < 10) = 0.5}$$

Hyperparameters:

This likelihood has two hyperparameters

$$\theta_1 = \log(\tau)$$

$$\theta_2 = \log(\nu - 2)$$

$$\theta_3 = \log\left(1/\sigma_U^2\right)$$

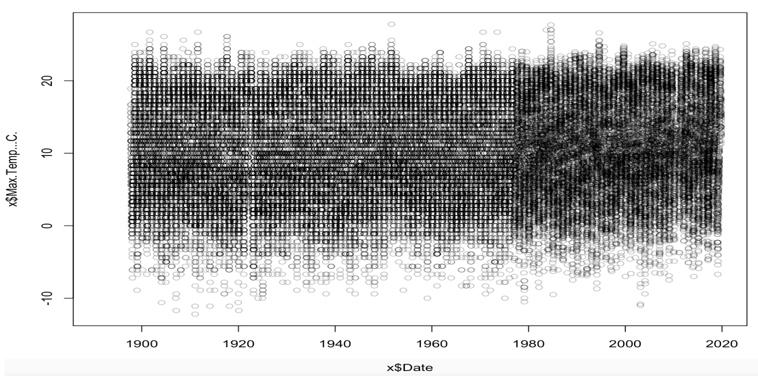
$$\theta_4 = \log\left(1/\sigma_V^2\right)$$

$$\theta_5 = \log\left(1/\sigma_Z^2\right)$$

and the prior is defined on $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$

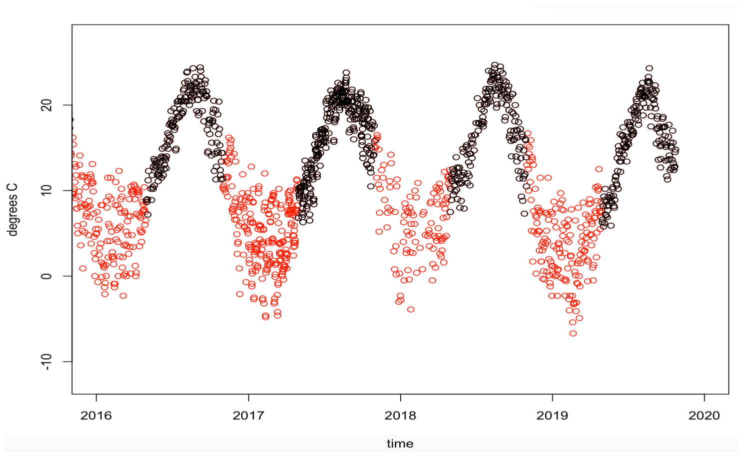
Analyzing Plots

Plot 1 (daily maximum temperature data recorded on Sable Island)*



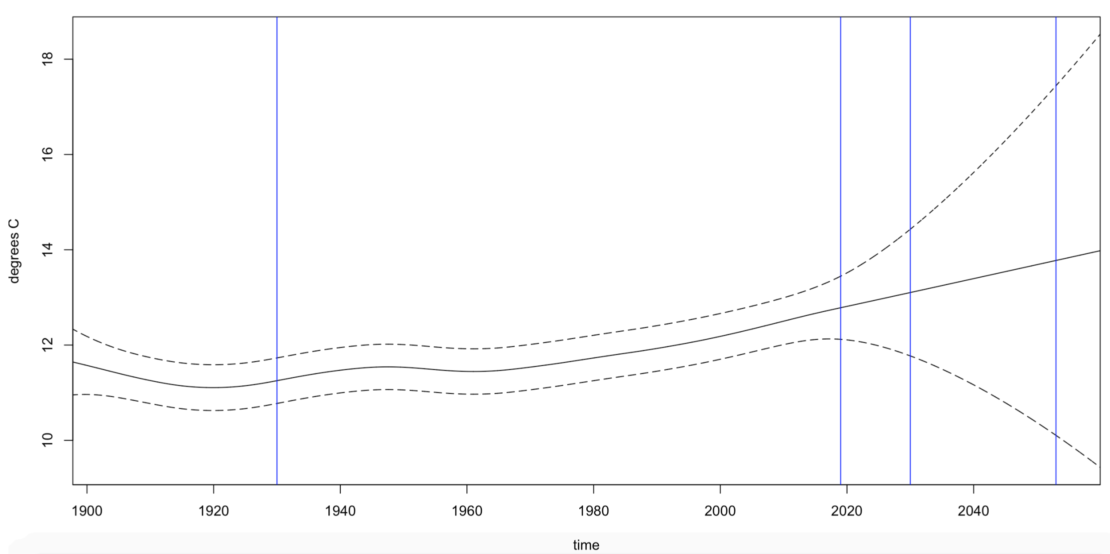
The graph shows the daily maximum temperature data for Sable Island from 1880 to 2019. But since the data are recorded daily for almost 40 years, there are nearly 15000 data points. As we can see, they are all crowded together, which is not efficient for the prediction of global warming. The points are generally ranged between -5°C and 25°C. However, from around 1976, there are more points located above 0°C. Visually, the right portion of the graph seems to be darker, which might be a sign of global warming.

Plot 2 (daily maximum temperature data from 2016 to present)*



Plot 2 is a zoomed version of Plot 1 where it shows the period from 2016 to the present, with summer in black and winter in red. Summer temperature is higher, so the points are clustered at the top of the graph. The summer temperature is much more uniform compared to winter temperature, but a reliable environmental scientist has advised us that it is sufficient to consider only summer temperatures when modelling historical temperature time series.

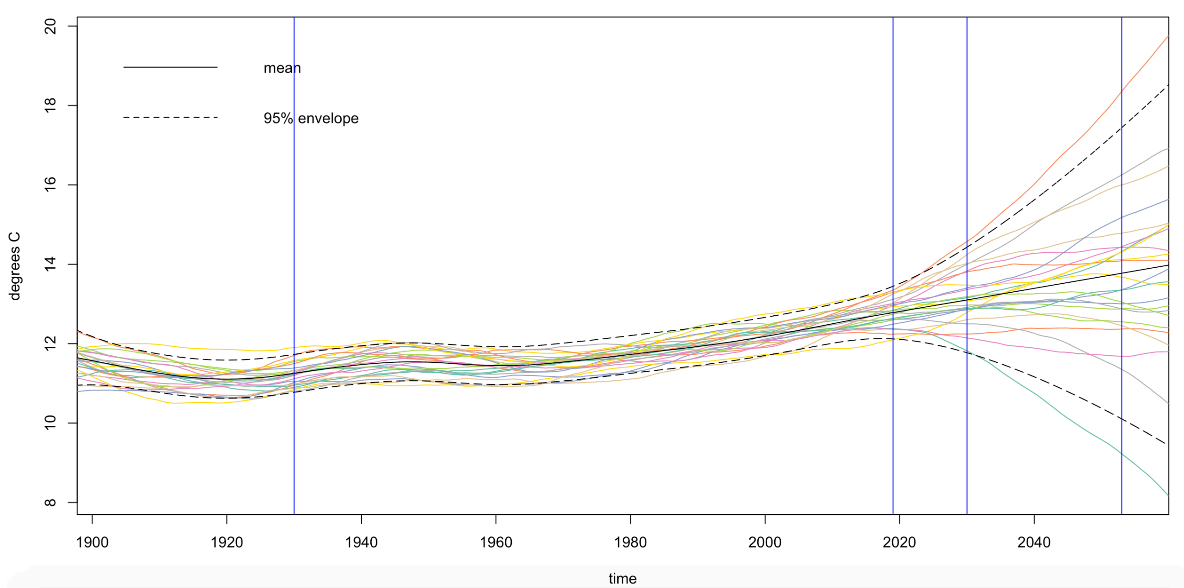
Plot 3 (fitted model and predictions)*



Above is the model fit. The black line represents the fitted trend of temperature, whereas the dotted line is the confidence interval for each estimate.

The blue lines from left to right represent year 1930,2019,2030 and 2052. The reason why I choose the time interval from 1930 to present is that pre-industrial ended in 1850; 1930's temperature level is indifferent to pre-industrial's level and is about the time when temperature started to increase gradually. The temperature increases from 11.5(***CI (11,11.8)***) to 12.4 (***CI (11.7,13.2)***). There is about 0.9°C of global warming, with a range of 0.7°C to 1.4°C. If we assume the temperature keeps rising at the present rate, temperature is likely to reach 13.5°C in 2050, and global warming is about 1.1°C. The rising-rate is fairly constant, probably because of the signing of Paris Agreement that limits the emission of carbon dioxide. Surprisingly, high confidence indicates global warming will reach about 1.5°C in 2030 and 5°C in 2052! For low confidence, there is a sign of dramatic decrease which is not really realistic. This is broadly consistent with IPCC's statement.

Plot 4 (24 posterior samples in different times with associated interpretation)*



These lines with 24 different colours are posteriors samples. The black line is the fitted trend of temperature, and the dotted line is the confidence interval. From 1930 to the present, there are only a handful of samples that is consistent to the result where human activities caused around 0.8°C to 1.2°C of global warming. 95% of samples are in the confidence interval. It means that whatever the 'true' trend is, there is a 95% chance it is entirely contained in this envelope.

Conclusion

Our predominant interest is to discover the extent of global warming caused by human activities in different periods and make predictions. After the pre-industrial period, the temperature oscillates between 11°C and 11.5°C. From 1960, global warming kept rising at an increasing rate until now, and increased by 1.0°C. Between 2030 and 2052, global warming is likely to reach a range of 1.5°C to 5°C.

Dear Maxim Burningier,

I have finished the analysis about global warming effect to Sable island. I am very thankful that you would compensate me 100 barrels of bitumen. However, it is a pity that I could not accept it because IPCC's statement is generally right. Though CO2 is beneficial for agriculture, but global warming is a non-ignorable fact. I do look forward for our next cooperation.

Best,
Yiwen

Reference

1. Patrick Brown. http://pbrown.ca/teaching/appliedstats/slides/nonparametric.pdf

Appendix

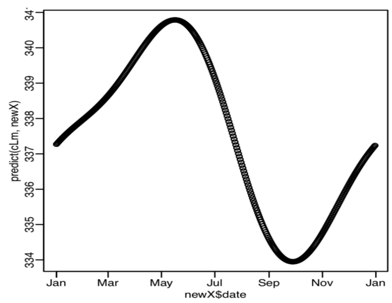
Code for Analyzing CO_2

```
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 >= 1, "co2"] = NA
plot(co2s$date, co2s$co2, log = "y", cex = 0.3, col = "#00000040", xlab = "time", ylab = "ppm")
matlines(as.numeric(newX$date), coPred[, c("lower",
                                  "upper", "est")], lty = 1, col = c("red", "red",
                                                             "red"))
plot(co2s[co2s$date > ISOdate(2015, 3, 1, tz = "UTC"), c("date", "co2")],
    log = "y", type = "o", xlab = "time", ylab = "ppm", cex = 0.5)

timeOrigin = ISOdate(1970, 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)
cLm = lm(co2 ~ days + cos12 + sin12 + cos6 + sin6,data = co2s)
summary(cLm)$coef[, 1:2]

newX = data.frame(date = seq(ISOdate(1970, 1, 1, 0, 0, 0, tz = "UTC"), by = "1 days", length.out = 365 * 50))
newX$days = as.numeric(difftime(newX$date, timeOrigin, units = "days"))
newX$cos12 = cos(2 * pi * newX$days/365.25)
newX$sin12 = sin(2 * pi * newX$days/365.25)
newX$cos6 = cos(2 * 2 * pi * newX$days/365.25)
newX$sin6 = sin(2 * 2 * pi * newX$days/365.25)
coPred = predict(cLm, newX, se.fit = TRUE)
coPred = data.frame(est = coPred$fit, lower = coPred$fit -
                    2 * coPred$se.fit, upper = coPred$fit + 2 * coPred$se.fit)

plot(newX$date, coPred$est, type = "l")
matlines(as.numeric(newX$date), coPred[, c("lower",
                                    "upper", "est")], lty = 1, col = c("yellow", "yellow",  "black"))
newX = newX[1:365, ]
newX$days = 0

plot(newX$date, predict(cLm, newX))
library("INLA")
# time random effect
timeBreaks = seq(min(co2s$date), ISOdate(2025, 1, 1,
                                  tz = "UTC"), by = "14 days")
timePoints = timeBreaks[-1]
co2s$timeRw2 = as.numeric(cut(co2s$date, timeBreaks))
# derivatives of time random effect
D = Diagonal(length(timePoints)) - bandSparse(length(timePoints),
                                       k = -1)
derivLincomb = inla.make.lincombs(timeRw2 = D[-1, ])
names(derivLincomb) = gsub("^lc", "time", names(derivLincomb)) # seasonal effect
StimeSeason = seq(ISOdate(2009, 9, 1, tz = "UTC"),
                ISOdate(2011, 3, 1, tz = "UTC"), len = 1001)
StimeYear = as.numeric(difftime(StimeSeason, timeOrigin,
                             "days"))/365.35
seasonLincomb = inla.make.lincombs(sin12 = sin(2 * pi * StimeYear),
                              cos12 = cos(2 * pi * StimeYear),
                              sin6 = sin(2 * 2 * pi * StimeYear), cos6 = cos(2 *
                                  2 * pi * StimeYear))
names(seasonLincomb) = gsub("^lc", "season", names(seasonLincomb)) # predictions
StimePred = as.numeric(difftime(timePoints, timeOrigin,
                             units = "days"))/365.35
predLincomb = inla.make.lincombs(timeRw2 = Diagonal(length(timePoints)),
                              `(Intercept)` = rep(1, length(timePoints)), sin12 = sin(2 *
                               pi * StimePred), cos12 = cos(2 * pi * StimePred), sin6 = sin(2 * 2 * pi * StimePred),
                              cos6 = cos(2 * 2 * pi * StimePred))
names(predLincomb) = gsub("^lc", "pred", names(predLincomb))
StimeIndex = seq(1, length(timePoints))
timeOriginIndex = which.min(abs(difftime(timePoints, timeOrigin)))
# disable some error checking in INLA
library("INLA")
mm = get("inla.models", INLA:::inla.get.inlaEnv())
if(class(mm) == 'function') mm = mm()
mm$latent$rw2$min.diff = NULL
assign("inla.models", mm, INLA:::inla.get.inlaEnv())
co2res = inla(co2 ~ sin12 + cos12 + sin6 + cos6 + f(timeRw2, model = 'rw2',
                                          values = StimeIndex,
                                          prior='pc.prec', param = c(log(1.01)/26, 0.5)),
             data = co2s, family='gamma', lincomb = c(derivLincomb, seasonLincomb, predLincomb),
```

```
        control.family = list(hyper=list(prec=list(prior='pc.prec', param=c(2, 0.5)))),
        control.inla = list(strategy='gaussian', int.strategy='eb'),
        verbose=TRUE)
matplot(timePoints, exp(co2res$summary.random$timeRw2[, c("0.5quant", "0.025quant", "0.975quant")]), type = "l", col =
"black",
    lty = c(1, 2, 2), log = "y", xaxt = "n", xlab = "time", ylab = "ppm(random effect)")
xax = pretty(timePoints)
axis(1, xax, format(xax, "%Y"))
derivPred = co2res$summary.lincomb.derived[grep("time",
                                rownames(co2res$summary.lincomb.derived)), c("0.5quant",
                                                        "0.025quant", "0.975quant")]
scaleTo10Years = (10 * 365.25/as.numeric(diff(timePoints,
                                units = "days")))
abline(v = ISOdate(2008, 9, 15, tz = "UTC"), col = "blue")
abline(v = ISOdate(1973, 10, 1, tz = "UTC"), col = "blue")
abline(v = ISOdate(1980, 1, 1, tz = "UTC"), col = "blue")
abline(v = ISOdate(1990, 10, 3, tz = "UTC"), col = "blue")
abline(v = ISOdate(2001, 12, 11, tz = "UTC"), col = "blue")
abline(v = ISOdate(2015, 12, 12, tz = "UTC"), col = "blue")
matplot(timePoints[-1], scaleTo10Years * derivPred,
    type = "l", col = "black", lty = c(1, 2, 2), ylim = c(0, 0.1), xlim = range(as.numeric(co2s$date)),
    xaxs = "i", xaxt = "n", xlab = "time", ylab = "log ppm, change per 10yr(derivative)")
axis(1, xax, format(xax, "%Y"))
abline(v = ISOdate(2008, 9, 15, tz = "UTC"), col = "blue")
abline(v = ISOdate(1973, 10, 1, tz = "UTC"), col = "blue")
abline(v = ISOdate(1980, 1, 1, tz = "UTC"), col = "blue")
abline(v = ISOdate(1989, 10, 3, tz = "UTC"), col = "blue")
abline(v = ISOdate(2001, 12, 11, tz = "UTC"), col = "blue")
abline(v = ISOdate(2015, 12, 12, tz = "UTC"), col = "blue")
matplot(StimeSeason, exp(co2res$summary.lincomb.derived[grep("season",
                rownames(co2res$summary.lincomb.derived)),
                c("0.5quant", "0.025quant", "0.975quant")]), type = "l", col = "black",
            lty = c(1, 2, 2), log = "y", xaxs = "i", xaxt = "n", xlab = "time", ylab = "seasonal relative ppm")
xaxSeason = seq(ISOdate(2009, 9, 1, tz = "UTC"), by = "2 months", len = 20)
axis(1, xaxSeason, format(xaxSeason, "%b"))
timePred = co2res$summary.lincomb.derived[grep("pred",
                                rownames(co2res$summary.lincomb.derived)), c("0.5quant",
                                                        "0.025quant", "0.975quant")]
matplot(timePoints, exp(timePred), type = "l", col = "black",
    lty = c(1, 2, 2), log = "y", xlim = ISOdate(c(2010,
   2025), 1, 1, tz = "UTC"), ylim = c(390, 435),
   xaxs = "i", xaxt = "n", xlab = "time", ylab = "prediction of ppm")
xaxPred = seq(ISOdate(2010, 1, 1, tz = "UTC"), by = "5 years", len = 20)
axis(1, xaxPred, format(xaxPred, "%Y"))

concentration = na.omit(co2s$co2)
mean(concentration)
var(concentration)
shape = (mean(a))^2/var(a)
scale = exp(5.885)/shape
hist(concentration,prob=TRUE,ylim=c(0,0.020))
xseq = seq(280,450,len=1000)
dgamma(xseq,shape= shape, scale= scale)
lines(xseq,dgamma(xseq,shape= shape, scale= scale),col = 'red')
```

Level of CO_2 within each year



(b) cycle

Code for Analyzing Heat

```
heatUrl = "http://pbrown.ca/teaching/appliedstats/data/sableIsland.rds"
heatFile = tempfile(basename(heatUrl))
download.file(heatUrl, heatFile)
x = readRDS(heatFile)
x$month = as.numeric(format(x$Date, "%m"))
xSub = x[x$month %in% 5:10 & !is.na(x$Max.Temp...C.),]
weekValues = seq(min(xSub$Date), ISOdate(2060, 1, 1,0, 0, 0, tz = "UTC"), by = "7 days")
xSub$week = cut(xSub$Date, weekValues)
xSub$weekIid = xSub$week
xSub$day = as.numeric(difftime(xSub$Date, min(weekValues),
                    units = "days"))
xSub$cos12 = cos(xSub$day * 2 * pi/365.25)
xSub$sin12 = sin(xSub$day * 2 * pi/365.25)
xSub$cos6 = cos(xSub$day * 2 * 2 * pi/365.25)
xSub$sin6 = sin(xSub$day * 2 * 2 * pi/365.25)
xSub$yearFac = factor(format(xSub$Date, "%Y"))
lmStart = lm(Max.Temp...C. ~ sin12 + cos12 + sin6 + cos6, data = xSub)
startingValues = c(lmStart$fitted.values, rep(lmStart$coef[1], nlevels(xSub$week)), rep(0, nlevels(xSub$weekIid) +
nlevels(xSub$yearFac)), lmStart$coef[-1])
INLA::inla.doc('^t$')
library("INLA")
mm = get("inla.models", INLA:::inla.get.inlaEnv())
if(class(mm) == 'function') mm = mm()
mm$latent$rw2$min.diff = NULL
assign("inla.models", mm, INLA:::inla.get.inlaEnv())
sableRes = INLA::inla(
  Max.Temp...C. ~ 0 + sin12 + cos12 + sin6 + cos6 + f(week, model='rw2',
                              constr=FALSE,
                              prior='pc.prec',
                              param = c(0.1/(52*100), 0.05)) +
    f(weekIid, model='iid', prior='pc.prec', param = c(1, 0.5)) +
    f(yearFac, model='iid', prior='pc.prec', param = c(1, 0.5)),
  family='T', control.family = list(
    hyper = list(
      prec = list(prior='pc.prec', param=c(1, 0.5)), dof = list(prior='pc.dof', param=c(10, 0.5)))),
  control.mode = list(theta = c(-1,2,20,0,1), x = startingValues, restart=TRUE),
  control.compute=list(config = TRUE),
  control.inla = list(strategy='gaussian', int.strategy='eb'),
  data = xSub, verbose=TRUE)
sableRes$summary.hyper[, c(4, 3, 5)]
Pmisc::priorPost(sableRes)$summary[, c(1, 3, 5)]
mySample = inla.posterior.sample(n = 24, result = sableRes, num.threads = 8, selection = list(week = seq(1,
                    nrow(sableRes$summary.random$week))))
length(mySample)
```

```
names(mySample[[1]])
weekSample = do.call(cbind, lapply(mySample, function(xx) xx$latent))
dim(weekSample)
head(weekSample)
plot(x$Date, x$Max.Temp...C., col = mapmisc::col2html("black", 0.3))
forAxis = ISOdate(2016:2020, 1, 1, tz = "UTC")
plot(x$Date, x$Max.Temp...C., xlim = range(forAxis),
                                    xlab = "time", ylab = "degrees C", col = "red",
                                    xaxt = "n")
points(xSub$Date, xSub$Max.Temp...C.)
axis(1, forAxis, format(forAxis, "%Y"))
matplot(weekValues[-1], sableRes$summary.random$week[,
   paste0(c(0.5, 0.025, 0.975), "quant")], type = "l", lty = c(1, 2, 2),
   xlab = "time", ylab = "degrees C", xaxt = "n", col = "black", xaxs = "i")
forXaxis2 = ISOdate(seq(1880, 2060, by = 20), 1, 1, tz = "UTC")
axis(1, forXaxis2, format(forXaxis2, "%Y"))
abline(v = ISOdate(2030, 1, 1, tz = "UTC"), col = "blue")
abline(v = ISOdate(2052, 12, 31, tz = "UTC"), col = "blue")
abline(v = ISOdate(2019, 1,1 , tz = "UTC"), col = "blue")
abline(v = ISOdate(1930, 1, 1, tz = "UTC"), col = "blue")
myCol = mapmisc::colourScale(NA, breaks = 1:8, style = "unique",
                      col = "Set2", opacity = 0.3)$col
matplot(weekValues[-1], weekSample, type = "l", lty = 1,
      col = myCol, xlab = "time", ylab = "degrees C",
      xaxt = "n", xaxs = "i")
axis(1, forXaxis2, format(forXaxis2, "%Y"))
matlines(weekValues[-1], sableRes$summary.random$week[,
                                    paste0(c(0.5, 0.025, 0.975), "quant")], type = "l", lty = c(1, 2, 2),
       xlab = "time", ylab = "degrees C", xaxt = "n", col = "black", xaxs = "i")
abline(v = ISOdate(2030, 1, 1, tz = "UTC"), col = "blue")
abline(v = ISOdate(2052, 12, 31, tz = "UTC"), col = "blue")
legend("topleft", bty = "n", lty = c(1, 2),col = c("black", "black"), legend = c("mean", "95% envelope"))


abline(v = ISOdate(2019, 1,1 , tz = "UTC"), col = "blue")
abline(v = ISOdate(1930, 1, 1, tz = "UTC"), col = "blue")
```