STA442 Homework 3, Smoothing

student number: 1003942326

Yulin WANG

12/11/2019

Question 1 CO2

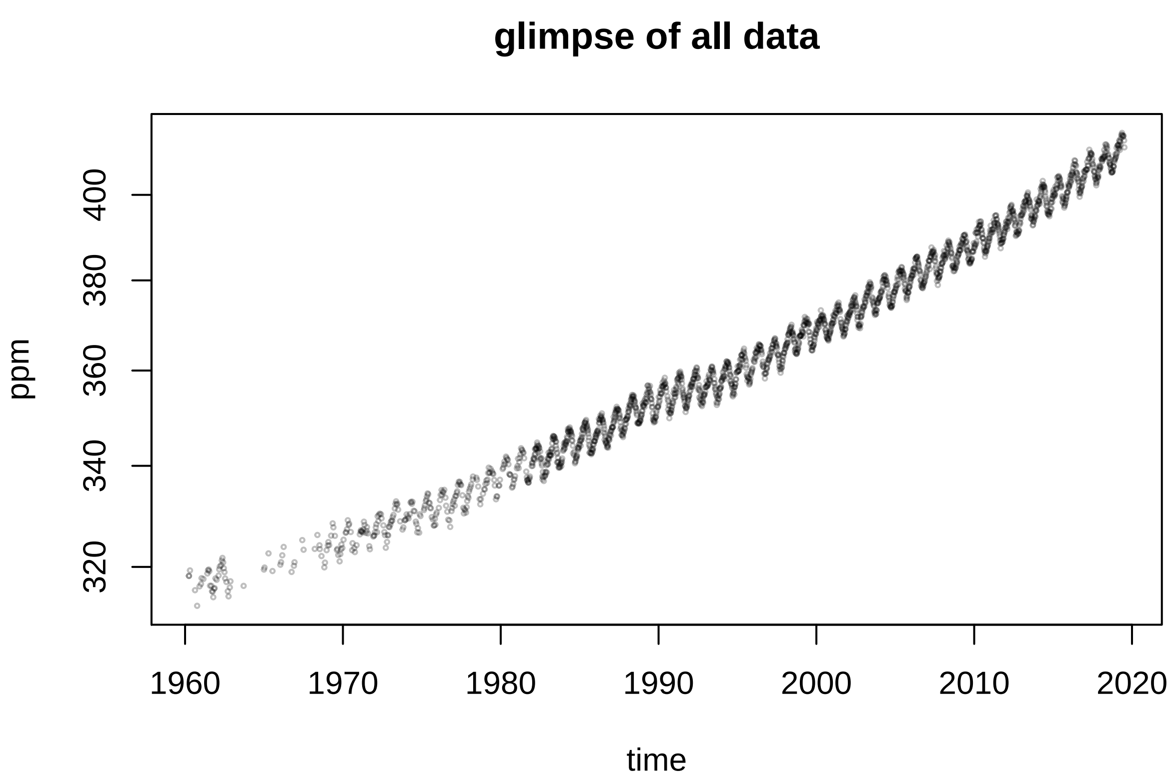
Method and Results

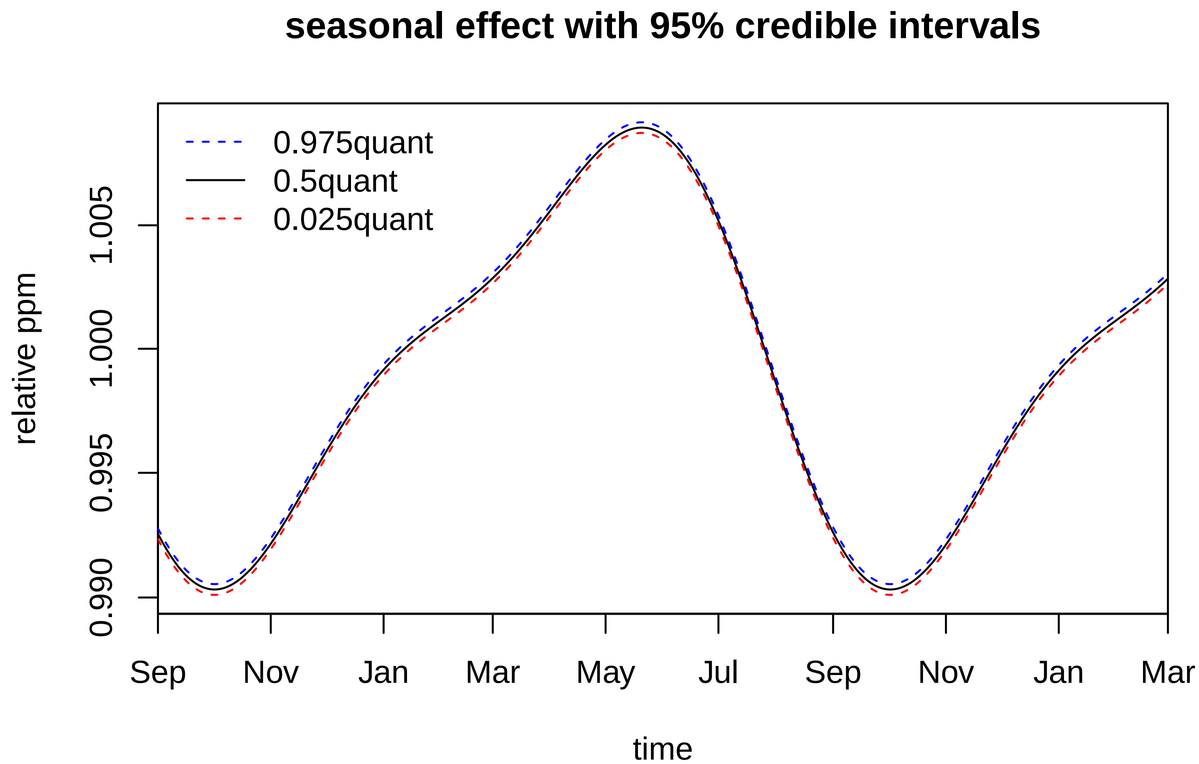
We investigated the changes in the atmospheric Carbon Dioxide concentrations from an observatory in Haiwaii, made available by the Scripps CO2 Program at scrippsco2.ucsd.edu. Since we noticed that there were much fewer data points before 1980 from the plot of all data, but the earlist event(the OPEC oil embargo) that we needed to analyze began in October 1973, so we chose to set 1970.1.1 as the baseline of date.

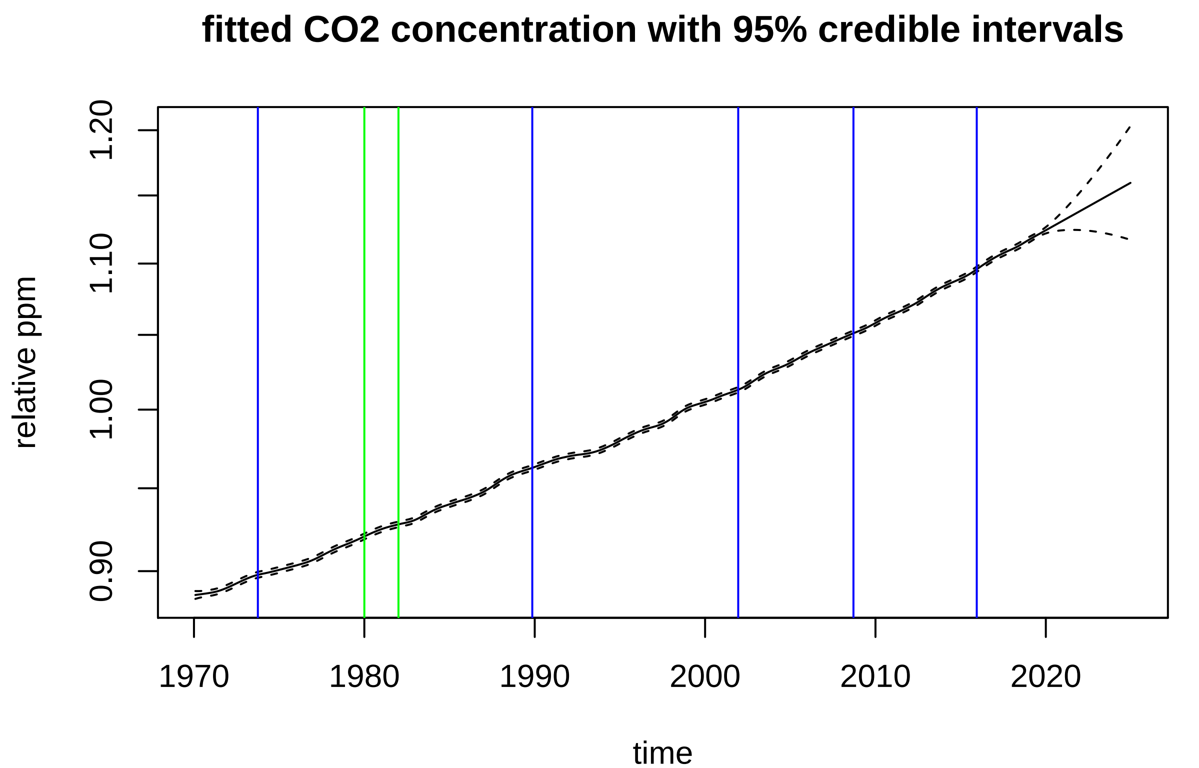
Since the CO2 concentration variation is periodic (i.e. there exists seasonal effect for CO2 consentration), so our model involved four seasonality predictors: $sin12 = sin(2 * \pi * days/365.25)$ and $cos12 = cos(2 * \pi * days/365.25)$ for 1-year period as annual cyclical, while $sin6 = sin(2 * 2 * \pi * days/365.25)$ and $cos6 = cos(2 * 2 * \pi * days/365.25)$ for 0.5-year period as semiannual cyclical respectively, where "days" means how many days away from the original date(1970.1.1).

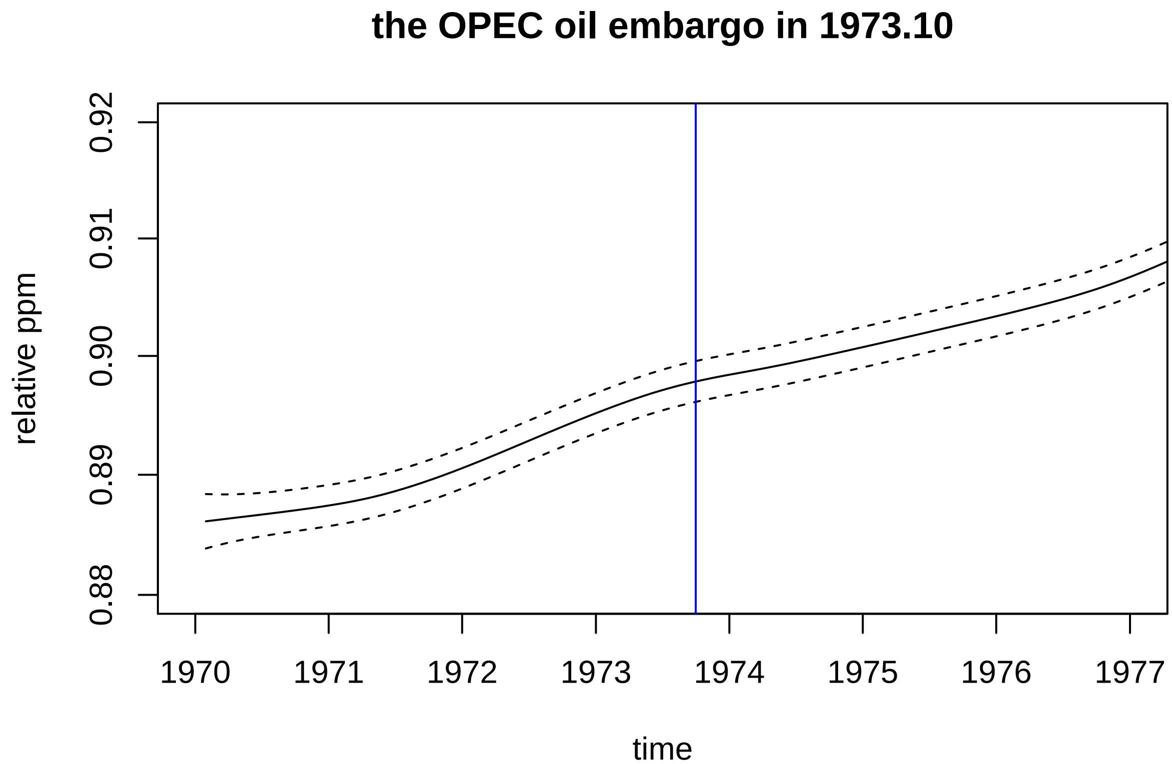
We assumed that the differences of time(days) followed a Normal distribution that is $(V_{t+1day} - V_t) - (V_t - V_{t-1day}) \sim N(0, \tau^2)$, and CO2 increase rate was 1% per year, so we set a PC prior for τ that it folllowed a $P(\tau > log(1.01)/26) = 0.5$ distribution, since there are 26 two-weeks in each year. Thus, we treated time as a random walk 2 effect(i.e. RW(2)) while other four seasonality variables as fixed effects and fitted a generalized linear mixed gamma model to model the atmospheric CO2 concentration as a function of time and four seasonality variables by using INLA.

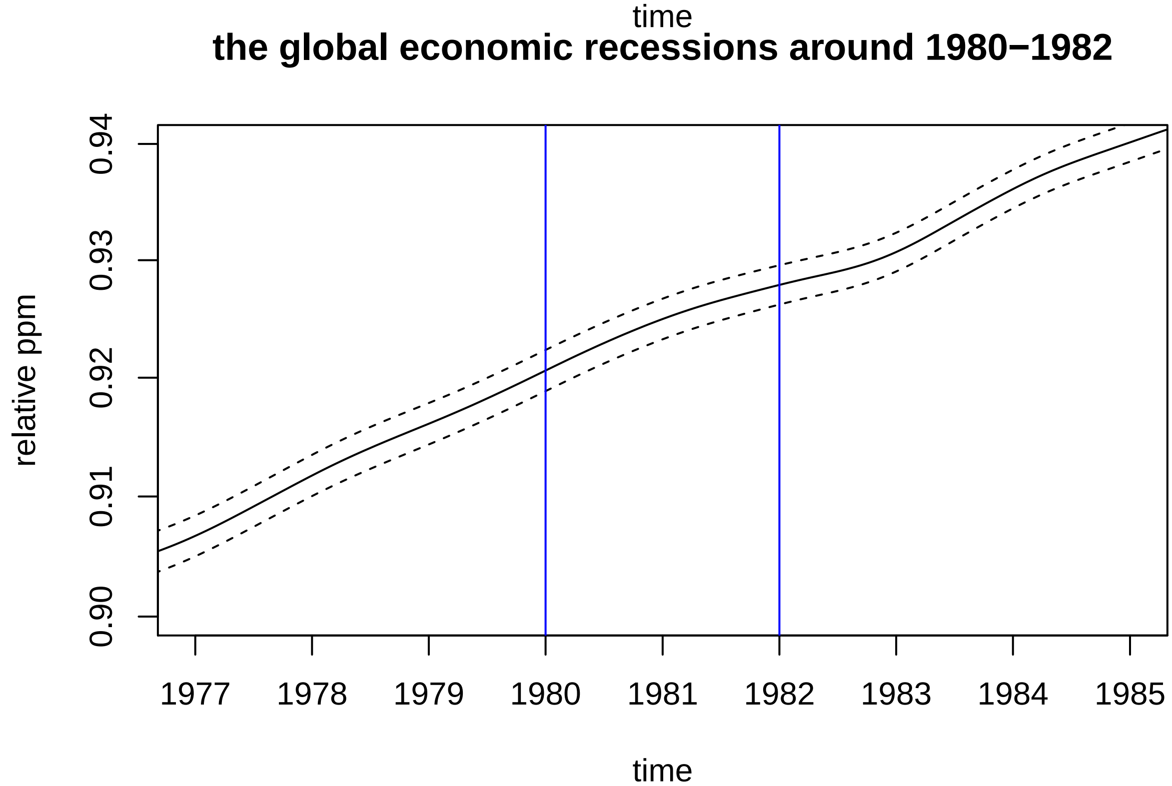
With the fitted gamma model, we created the predicted CO2 concentration plot with 95% credible intervals and zoomed in to it seperately for each event. And then we plotted the approximation to the first derivatives of the trend for CO2 concentration changes with 95% credible intervals. Thus, we could combine the results of the slope changes of zoomed-in fitted CO2 concentration plot and the first derivatives plot to analyze if the CO2 concentration appeared to be impacted by the events, and whether the trend of CO2 appeared shallower or steeper after each event happened.

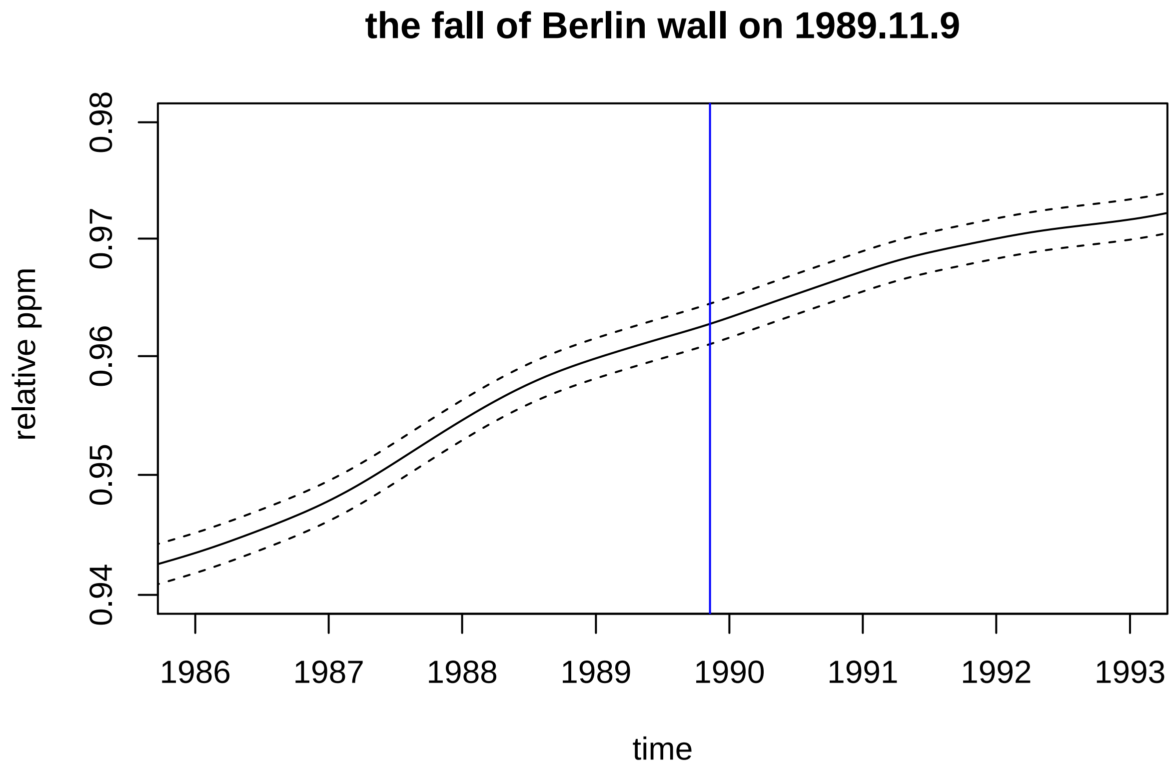


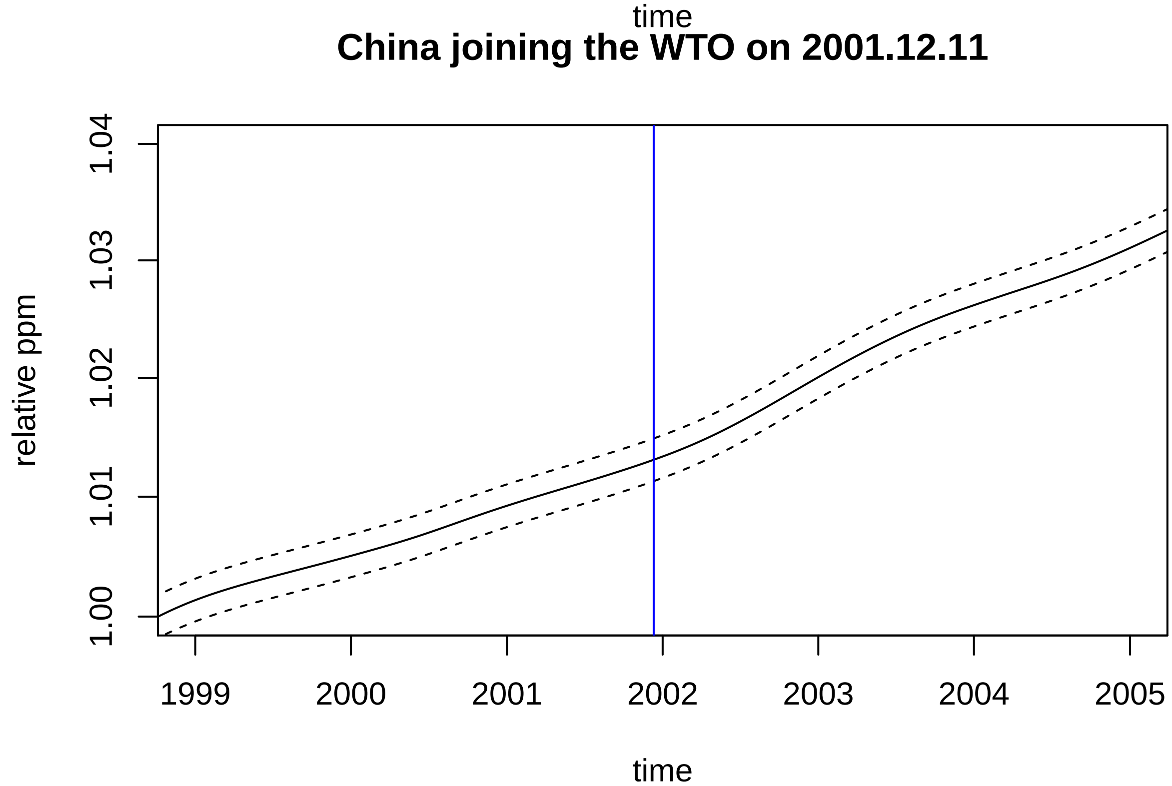

seasonal effect with 95% credible intervals
0.975quant
0.5quant
0.025quant
relative ppm
0.990
0.995
1.000
1.005
Sep
Nov
Jan
Mar
May
Jul
Sep
Nov
Jan
Mar
time


fitted CO2 concentration with 95% credible intervals
relative ppm
0.90
1.00
1.10
1.20
1970
1980
1990
2000
2010
2020
time

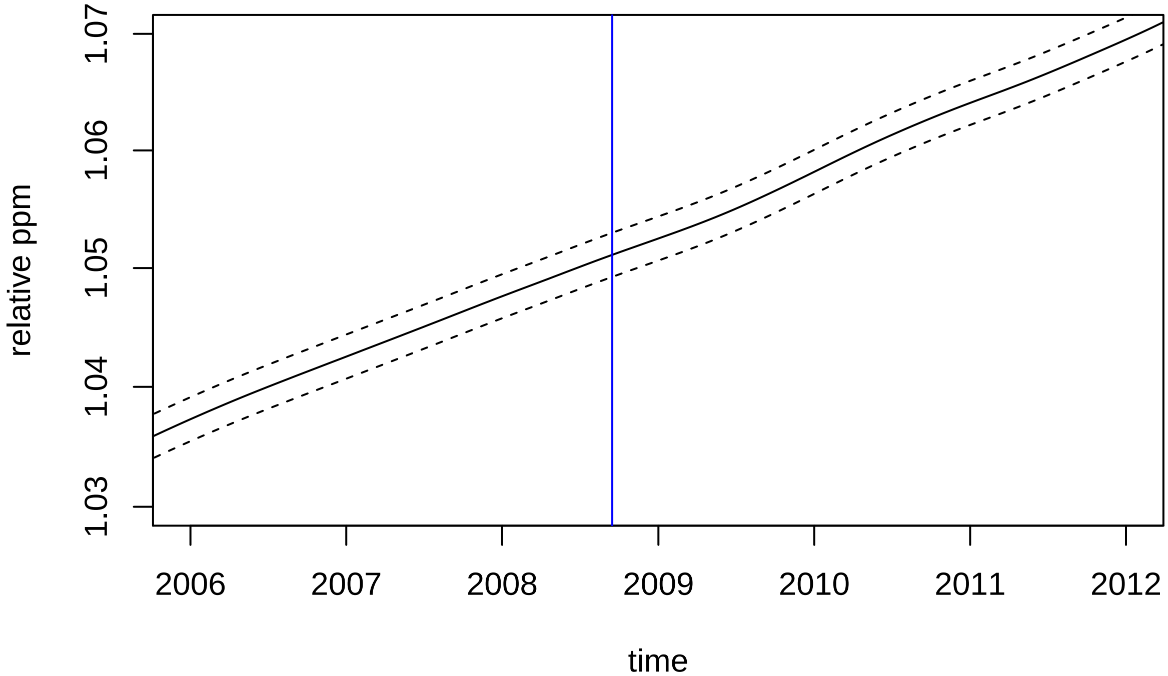

the OPEC oil embargo in 1973.10
relative ppm
0.88
0.89
0.90
0.91
0.92
1970
1971
1972
1973
1974
1975
1976
1977
time


the global economic recessions around 1980−1982
relative ppm
0.90
0.91
0.92
0.93
0.94
1977
1978
1979
1980
1981
1982
1983
1984
1985
time

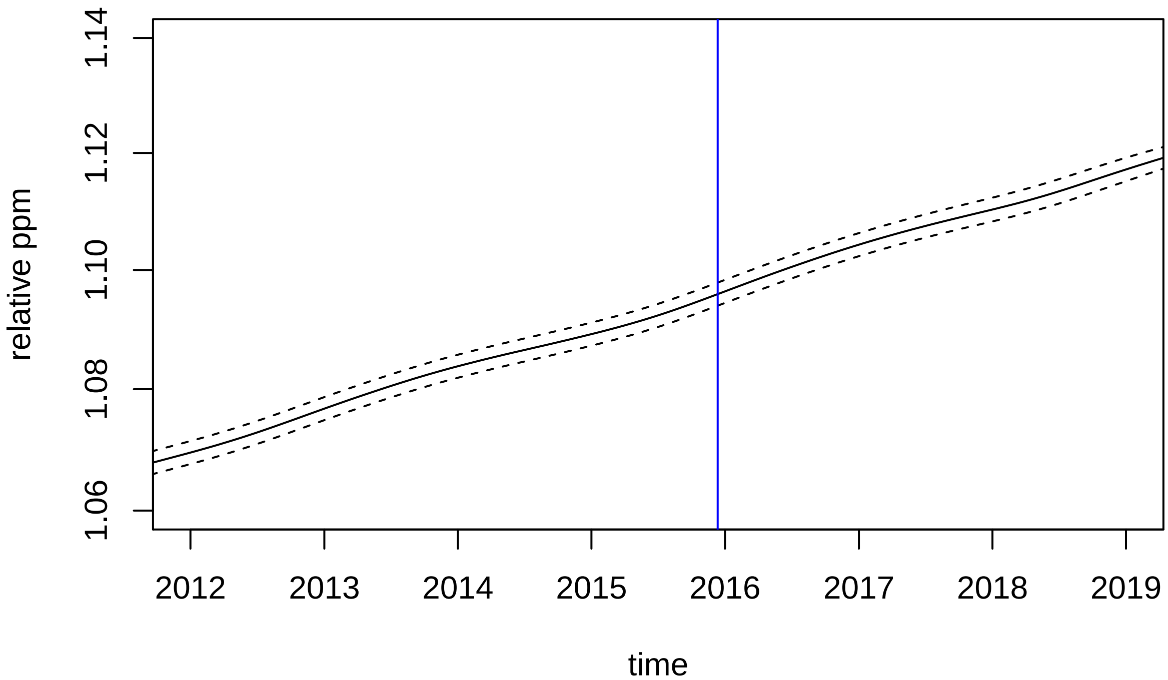

the fall of Berlin wall on 1989.11.9
relative ppm
0.94
0.95
0.96
0.97
0.98
1986
1987
1988
1989
1990
1991
1992
1993
time

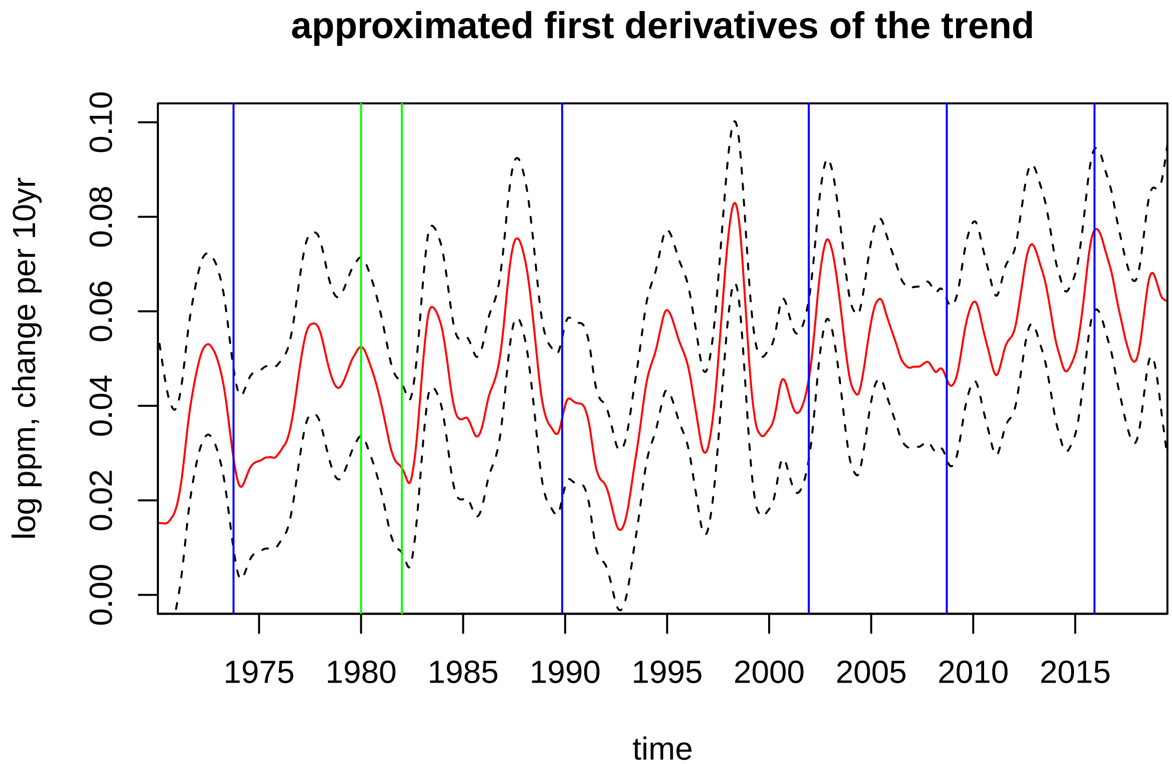

China joining the WTO on 2001.12.11
relative ppm
1.00
1.01
1.02
1.03
1.04
1999
2000
2001
2002
2003
2004
2005
time

the bankruptcy of Lehman Brothers on 2008.9.15



the signing of the Paris Agreement on 2015.12.12





Conclusion

1.the OPEC oil embargo which began in October 1973 - small impact

The increasing trend of CO2 concentration appeared slightly shallower after the OPEC oil embargo. Since it caused the 1973-1975 recession and aggravated inflation by raising oil prices, which reduced the industrial production, so that limited the CO2 emissions.

2.the global economic recessions around 1980-1982 - great impact

The increasing trend of CO2 concentration appeared greatly shallower during the global economic recessions, which caused a rapid decrease in industrial production.

3.the fall of the Berlin wall almost exactly 30 years ago(1989.11.9) - great impact

The increasing trend of CO2 concentration appeared greatly shallower after the fall of the Berlin wall. Since after that there was a dramatic fall in industrial production in the Soviet Union and Eastern Europe.

4.China joining the WTO on 11 December 2001 - great impact

The increasing trend of CO2 concentration appeared greatly steeper after China joining the WTO, which was followed by rapid growth in industrial production.

5.the bankruptcy of Lehman Brothers on 15 September 2008 - small impact

The increasing trend of CO2 concentration appeared slightly shallower and then steeper after the bankruptcy of Lehman Brothers, which regarded as the symbolic start of the most recent global financial crisis.

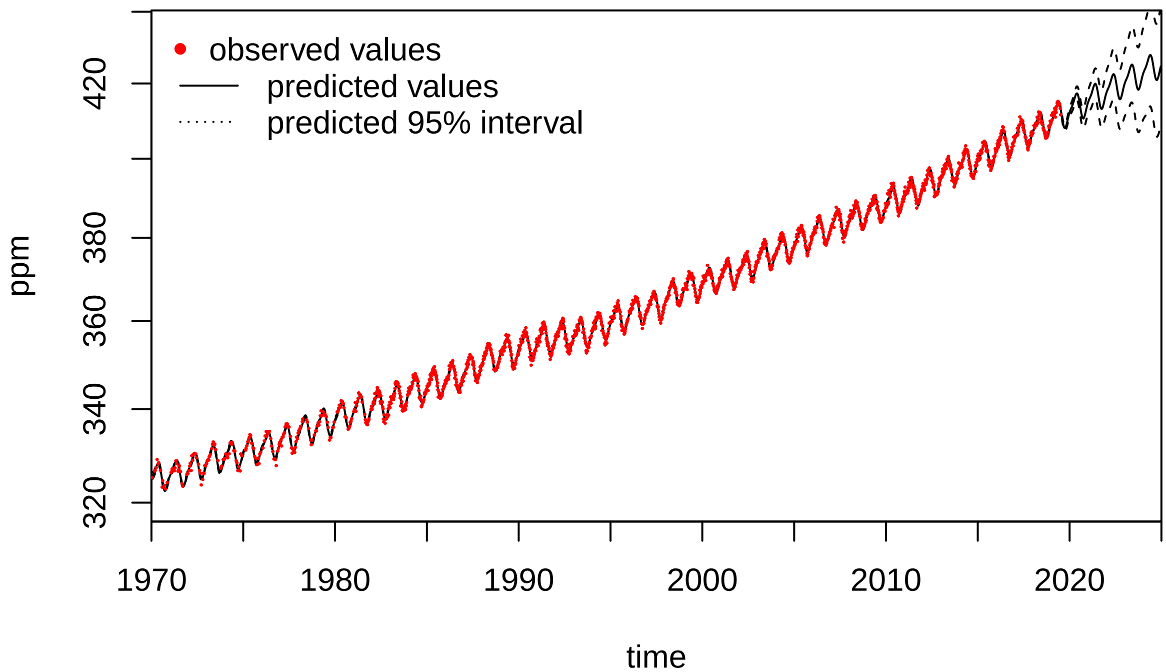
6.the signing of the Paris Agreement on 12 December 2015 - great impact

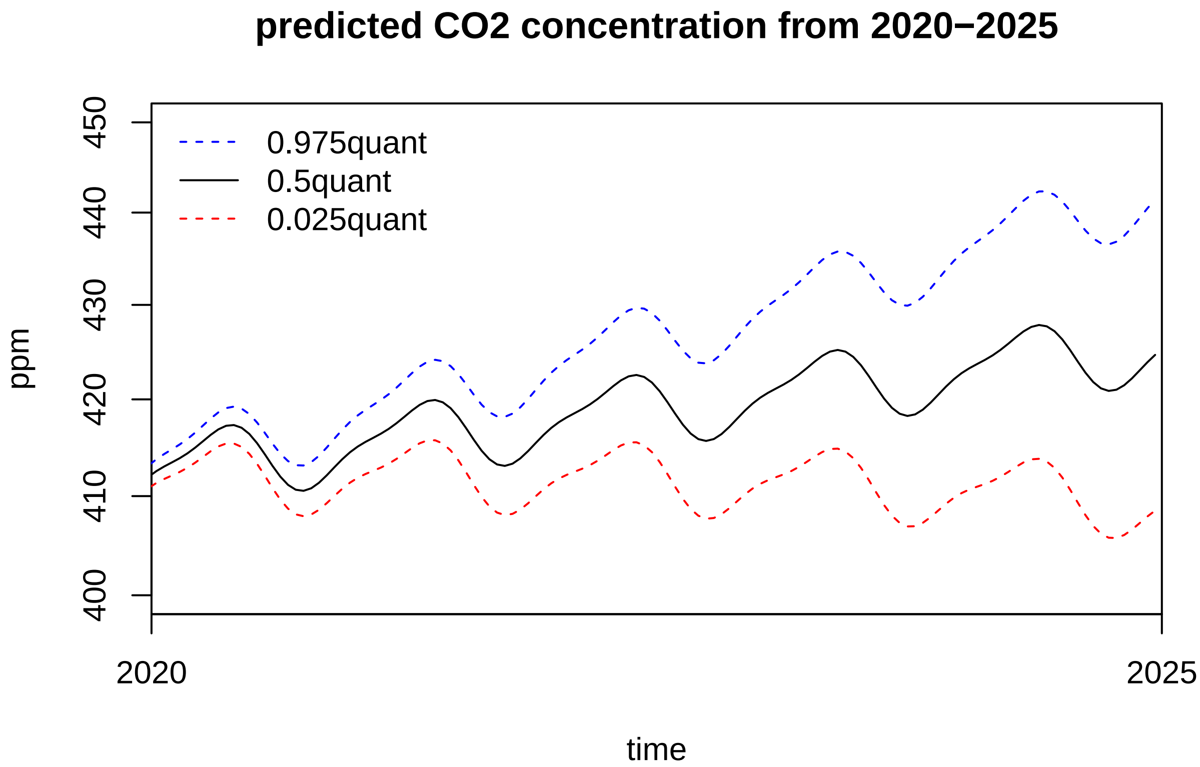
The increasing trend of CO2 concentration appeared greatly shallower after the signing of the Paris Agreement by 195 UNFCCC members, which intended to limit the global CO2 emissions.

Limitation and furthur research

From the figure assessing fit of data, we could find that most of the observed values(red points) fell on the predicted lines, so that our model fitted well for the CO2 data. If we want to furthur focast the CO2 concentration in next five years, we could zoom in the fitted values plot from 2020-2050, and we can see that the CO2 concentration in 2025 is expected to be approximately between 408 ppm and 438ppm with 95% credible interbvals. Additionally, we can fit a GAM model for prediction since it has advantages of stability and prediction with Big Data.

Assessing the fit of CO2 concentration with our model





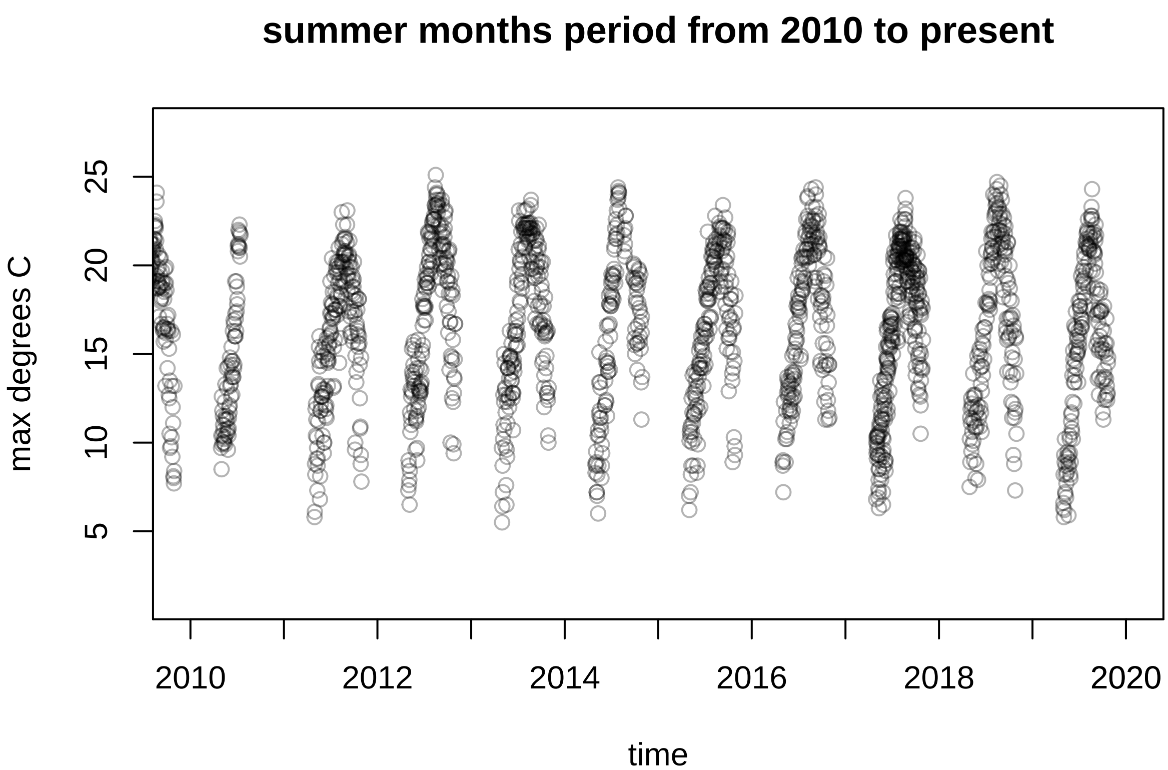
predicted CO2 concentration from 2020−2025
0.975quant
0.5quant
0.025quant
ppm
400
410
420
430
440
450
2020
2025
time

Question2 Heat

Introduction

The IPCC states human activities are estimated to have caused approximately 1.0°C of global warming above preindustrial levels, with a likely range of 0.8°C to 1.2°C. Global warming is likely to reach 1.5°C between 2030 and 2052 if it continues to increase at the current rate. (high confidence)
We analyzed the daily maximum temperature data recorded on Sable Island, off the coast of Nova Scotia. And we wanted to investigate whether the data from Sable Island was broadly supportive of this statement from the IPCC.

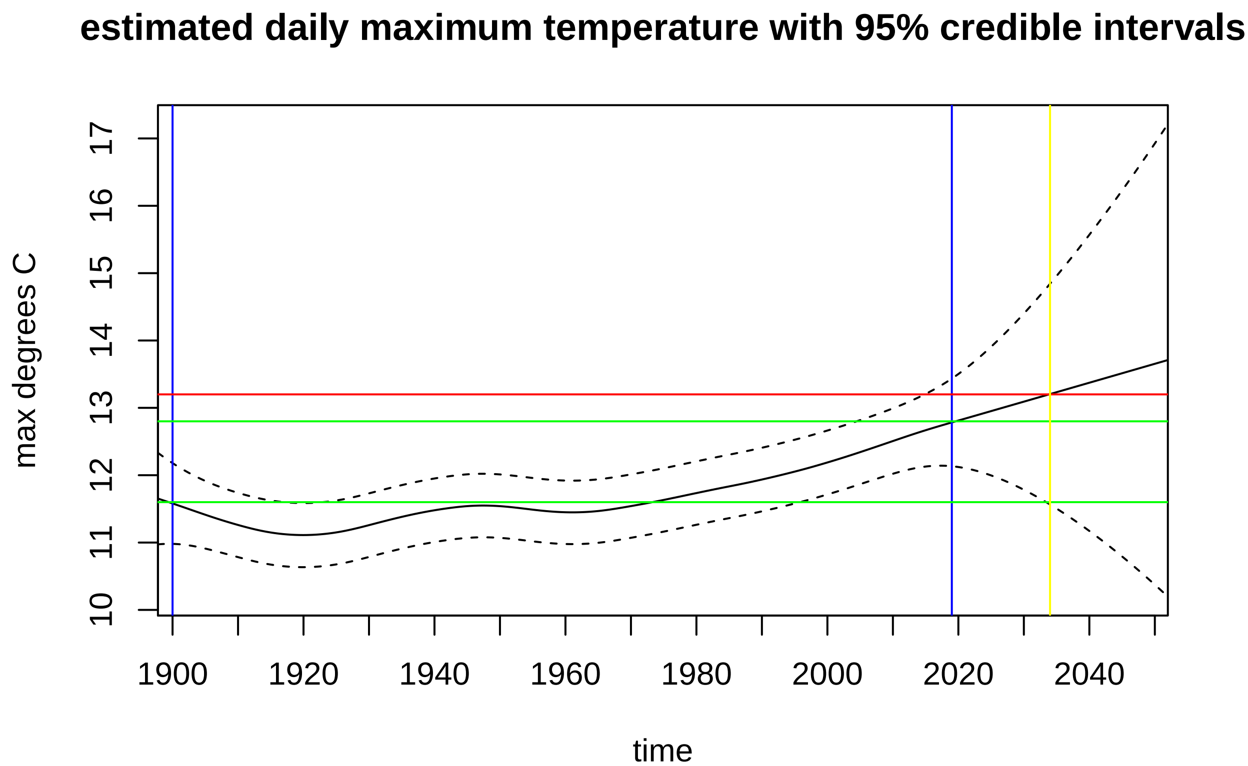
Methods



Given that the winter temperatures are more variable than summer temperatures, and advised by a reliable environmental scientist, we decided to consider only summer temperatures when modelling historical temperature time series. From the above figure about the daily maximum temperature in summer months from 2010 to present, we noticed that the changes of the daily maximum temperatures were periodic. Thus, our model involved four periodic predictors: $sin12 = sin(2 * \pi * days/365.25)$ and $cos12 = cos(2 * \pi * days/365.25)$ for 1-year period as annual cyclical, while $sin6 = sin(2*2*\pi*days/365.25)$ and $cos6 = cos(2*2*\pi*days/365.25)$ for 0.5-year period as semiannual cyclical respectively, where "days" means how many days away from the original date(1897.1.1).
We firstly fitted a simple linear regression model and then used its fitted values as the initial values for INLA. We created variables "week" and "weekIid" which both represented how many weeks away from the original date and a "yearFac" which represented how many years away from the original date.
Since the changes of the daily maximum temperatures were periodic, in order to fit a better model, we set up a few assumptions as below:
First assumption: We assumed that each "weekIid" identically independently followed a Normal distribution, that is $W_t \overset{iid}{\sim} N(0, \sigma_1^2)$ and we set a PC prior for σ_1 that it followed a $P(\sigma_1 > 1) = 0.5$ distribution.
Second assumption: We assumed that each "yearFac" identically independently followed a Normal distribution, that is $Y_t \overset{iid}{\sim} N(0, \sigma_2^2)$ and we set a PC prior for σ_2 that it followed a $P(\sigma_2 > 1) = 0.5$ distribution.

Third assumption: We assumed that the differences of “weeks” followed a Normal distribution, that is $(V_{t+1week} - V_t) - (V_t - V_{t-1week}) \sim N(0, \tau^2)$ and the maximum temperature increased by 10% for every 100 years, so we set a PC prior for τ that it follwed a $P(\tau > 0.1/(52 * 100)) = 0.05$ distribution, since there are 52 weeks in each year.
Thus, we used these three reasonable priors and treated “weekIid” as a random walk 0 effect(i.e. RW(0)), “yearFac” as a random walk 0 effect(i.e. RW(0)) and “week” as a random walk 2 effect(i.e. RW(2)), while other four periodic variables as fixed effects. Then we fitted a mixed T model to model the daily maximum temperature as a function of “week”, “weekIid”, “yearFac” and four periodic variables by using INLA. And then we created the plot of estimated daily maximum temperature with 95% credible intervals.

Results



Two blue lines represent the year 1900(pre-industrial level) and the year 2019 respectively.
Two green lines represent the median of the estimated daily maximum temperatures according to these two years respectively.
The red line represents the daily maximum temperature which is approximately 1.5°C above the pre-industrial level and the yellow line shows the according year.

Note that we chose the period 1850-1900 as “pre-industrial”, since this period is often described as “pre-industrial” according to the report on the conversation website. (https://theconversation.com/what-is-a-pre-industrial-climate-and-why-does-it-matter-78601)

From the above plot of estimated daily maximum temperature with 95% credible intervals, we found that the estimated daily maximum temperature appeared to increase by about 1°C from pre-industrial to 2019 and the increase amount tended to reach 1.5°C between 2030 and 2040.

Conclusion

We analyzed the daily maximum temperature data recorded on Sable Island, off the coast of Nova Scotia. We found that if given 1900 as our pre-industrial level baseline, the daily maximum temperature on Sable Island appeared to increase by about 1°C from 1900 to 2019 and the increase amount tended to reach 1.5°C between 2030 and 2040. Thus, there is no statistically significant evidence to refute the statement from the IPCC.

Limitation and furthur research

1.The 'starting line' of the pre-industrial era is not defined by the UN agreements, or by the Intergovernmental Panel on Climate Change (IPCC), so it is hard to determine the specified increase amount of the maximum temperature above the pre-industrial level.
2. Our conclusion is based on the predicted plot we created, if we want a more accurate estimated change interval of maximum temperatures, we should do some furthur research such as Bayesian inference methods for the credible intervals of the change.
3. In our report, the data set we used only recorded the daily maximum temperature data recorded on Sable Island, off the coast of Nova Scotia, which is a small part of the whole world. Thus if we want to reduce the bias, our analysis should include more data sets.

Appendix

Question 1 CO2

```
co2s = read.table("/Users/mac/Desktop/2019\ Fall/STA442/assignments/A3/daily_flask_co2_mlo.csv",
                  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", main= "glimpse of all data")
# there is fewer data before 1980

#remove NA
co2s <- na.omit(co2s)
#set baseline of date: start from 1970,
#since there is few data before 1980 but the question asks for 1973.
timeoriginal = ISOdate(1970, 1, 1, 0, 0, 0, tz = "UTC")
#normalized date as numeric values
co2s$days = as.numeric(difftime(co2s$date, timeoriginal, units = "days"))
#days away from original date

#periodic settings
co2s$cos12 = cos(2 * pi * co2s$days/365.25) #period = 1 yr
co2s$sin12 = sin(2 * pi * co2s$days/365.25) #period = 1 yr
co2s$cos6 = cos(2 * 2 * pi * co2s$days/365.25) #period = 0.5 yr
co2s$sin6 = sin(2 * 2 * pi * co2s$days/365.25) #period = 0.5 yr


#using INLA
library("INLA")
# time random effect
co2s <- co2s[co2s$date > ISOdate(1970, 1, 1, tz = "UTC"),]
#select data starts from 1970.1.1
timeBreaks = seq(min(co2s$date), ISOdate(2025, 1, 1, tz = "UTC"), by = "14 days")
#one factor level for each two weeks (14 days)
timePoints = timeBreaks[-1] #get rid of first time point as baseline
co2s$timeRw2 = as.numeric(cut(co2s$date, timeBreaks))
# first 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, timeoriginal, "days"))/365.35
#how many years away from the original
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
```

```r
StimePred = as.numeric(difftime(timePoints, timeoriginal, 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))
timeoriginalIndex = which.min(abs(difftime(timePoints, timeoriginal)))

# 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())

#fit the gamma model by inla
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)))),verbose=TRUE)


#seasonal effect
matplot(StimeSeason, exp(co2res$summary.lincomb.
                          derived[grep("season", rownames(co2res$summary.lincomb.derived)),
                                                               c("0.5quant", "0.025quant", "0.975quant")]),
        type = "l", col = c("black", "red", "blue"), lty = c(1, 2, 2), log = "y",
        xaxs = "i", xaxt = "n", xlab = "time", ylab = "relative ppm"
        , main = "seasonal effect with 95% credible intervals")
xaxSeason = seq(ISOdate(2009, 9, 1, tz = "UTC"), by = "2 months", len = 20)
axis(1, xaxSeason, format(xaxSeason, "%b"))
legend("topleft", bty = "n", lty = c(2, 1, 2), col = c("blue", "black", "red"),
       legend = c("0.975quant", "0.5quant", "0.025quant"))


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 = "relative ppm",
        main = "fitted CO2 concentration with 95% credible intervals")
xax = pretty(timePoints)
axis(1, xax, format(xax, "%Y"))
abline(v = ISOdate(1973, 10, 1, tz = "UTC"), col = "blue")
abline(v = ISOdate(1980, 1, 1, tz = "UTC"), col = "green")
abline(v = ISOdate(1982, 1, 1, tz = "UTC"), col = "green")
abline(v = ISOdate(1989, 11, 9, tz = "UTC"), col = "blue")
abline(v = ISOdate(2001, 12, 11, tz = "UTC"), col = "blue")
abline(v = ISOdate(2008, 9, 15, tz = "UTC"), col = "blue")
```

```r
abline(v = ISOdate(2015, 12, 12, tz = "UTC"), col = "blue")


#zoom in for a close look from 1970-1977 for the OPEC oil embargo in 1973.10
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 = "relative ppm", ylim = c(0.88, 0.92),
        xlim = c(ISOdate(1970, 1, 1, tz = "UTC"),ISOdate(1977, 1, 1, tz = "UTC"))
        ,main = "the OPEC oil embargo in 1973.10")
xax = seq(ISOdate(1970, 1, 1, tz = "UTC"), by = "1 year", len = 20)
axis(1, xax, format(xax, "%Y"))
abline(v = ISOdate(1973, 10, 1, tz = "UTC"), col = "blue") #shallower

#zoom in for a close look from 1977-1985 for the global economic recessions around 1980-1982
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 = "relative ppm", ylim = c(0.9, 0.94),
        xlim = c(ISOdate(1977, 1, 1, tz = "UTC"),ISOdate(1985, 1, 1, tz = "UTC"))
        , main = "the global economic recessions around 1980-1982")
xax = seq(ISOdate(1977, 1, 1, tz = "UTC"), by = "1 year", len = 20)
axis(1, xax, format(xax, "%Y"))
abline(v = ISOdate(1980, 1, 1, tz = "UTC"), col = "blue")
abline(v = ISOdate(1982, 1, 1, tz = "UTC"), col = "blue")

#zoom in for a close look from 1986-1993 for the fall of Berlin wall on 1989.11.9
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 = " relative ppm",ylim = c(0.94, 0.98),
        xlim = c(ISOdate(1986, 1, 1, tz = "UTC"),ISOdate(1993, 1, 1, tz = "UTC"))
        , main = "the fall of Berlin wall on 1989.11.9")
xax = seq(ISOdate(1986, 1, 1, tz = "UTC"), by = "1 year", len = 20)
axis(1, xax, format(xax, "%Y"))
abline(v = ISOdate(1989, 11, 9, tz = "UTC"), col = "blue")

#zoom in for a close look from 1998-2005 for China joining the WTO on 2001.12.11
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 = "relative ppm", ylim = c(1.0, 1.04),
        xlim = c(ISOdate(1999, 1, 1, tz = "UTC"),ISOdate(2005, 1, 1, tz = "UTC"))
        , main = "China joining the WTO on 2001.12.11")
xax = seq(ISOdate(1999, 1, 1, tz = "UTC"), by = "1 year", len = 20)
axis(1, xax, format(xax, "%Y"))
abline(v = ISOdate(2001, 12, 11, tz = "UTC"), col = "blue")

#zoom in for a close look from 2005-2012 for the bankruptcy of Lehman Brothers on 2008.9.15
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 = "relative ppm", ylim = c(1.03, 1.07),
        xlim = c(ISOdate(2006, 1, 1, tz = "UTC"),ISOdate(2012, 1, 1, tz = "UTC"))
        , main = "the bankruptcy of Lehman Brothers on 2008.9.15")
xax = seq(ISOdate(2006, 1, 1, tz = "UTC"), by = "1 year", len = 20)
axis(1, xax, format(xax, "%Y"))
abline(v = ISOdate(2008, 9, 15, tz = "UTC"), col = "blue")
```

```r
#zoom in for a close look from 2012-2019 for the signing of the Paris Agreement on 2015.12.12
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 = "relative ppm", ylim = c(1.06, 1.14),
        xlim = c(ISOdate(2012, 1, 1, tz = "UTC"),ISOdate(2019, 1, 1, tz = "UTC"))
        , main = "the signing of the Paris Agreement on 2015.12.12")
xax = seq(ISOdate(2012, 1, 1, tz = "UTC"), by = "1 year", len = 20)
axis(1, xax, format(xax, "%Y"))
abline(v = ISOdate(2015, 12, 12, tz = "UTC"), col = "blue")

# derivative of time effect
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")))
matplot(timePoints[-1], scaleTo10Years * derivPred,
        type = "l", col = c("red", "black", "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"
        , main = "approximated first derivatives of the trend")
xax = seq(ISOdate(1970, 1, 1, tz = "UTC"), by = "5 years", len = 20)
axis(1, xax, format(xax, "%Y"))
abline(v = ISOdate(1973, 10, 1, tz = "UTC"), col = "blue") #shallower
abline(v = ISOdate(1980, 1, 1, tz = "UTC"), col = "green") #shallower
abline(v = ISOdate(1982, 1, 1, tz = "UTC"), col = "green") #shallower
abline(v = ISOdate(1989, 11, 9, tz = "UTC"), col = "blue") #steeper then shallower
abline(v = ISOdate(2001, 12, 11, tz = "UTC"), col = "blue") #steeper
abline(v = ISOdate(2008, 9, 15, tz = "UTC"), col = "blue") #shallower then steeper
abline(v = ISOdate(2015, 12, 12, tz = "UTC"), col = "blue") #shallower


#predcit till 2025
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(1970,2025), 1, 1, tz = "UTC"),
        ylim = c(320, 435), xaxs = "i", xaxt = "n", xlab = "time", ylab = "ppm"
        , main = "Assessing the fit of CO2 concentration with our model")
xaxPred = seq(ISOdate(1970, 1, 1, tz = "UTC"), by = "5 years", len = 20)
axis(1, xaxPred, format(xaxPred, "%Y"))
points(co2s$date, co2s$co2, cex = 0.1, col = "red") #observed values
legend("topleft", bty = "n", pch = 20, col = "red", legend = "observed values")
legend(x = 1972, y = 430, bty = "n", lty= 1, col = "black", legend = "predicted values")
legend(x = 1972, y = 420, bty = "n", lty= 3, col = "black", legend = "predicted 95% interval")


#predict from 2020-2025
matplot(timePoints, exp(timePred), type = "l", col = c("black", "red", "blue"),
        lty = c(1, 2, 2), log = "y", xlim = ISOdate(c(2020,2025), 1, 1, tz = "UTC"),
        ylim = c(400, 450), xaxs = "i", xaxt = "n", xlab = "time", ylab = "ppm"
        , main = "predicted CO2 concentration from 2020-2025")
xaxPred = seq(ISOdate(2020, 1, 1, tz = "UTC"), by = "5 years", len = 20)
axis(1, xaxPred, format(xaxPred, "%Y"))
legend("topleft", bty = "n", lty = c(2, 1, 2), col = c("blue", "black", "red"),
```

```
      legend = c("0.975quant", "0.5quant", "0.025quant"))
```

Question 2 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")) #make month numeric
xSub = x[x$month %in% 5:10 & !is.na(x$Max.Temp...C.),]   #only summer time
weekValues = seq(min(xSub$Date), ISOdate(2052, 1, 1, 0, 0, 0, tz = "UTC"), by = "7 days")
#change 2030 to 2052
xSub$week = cut(xSub$Date, weekValues)
xSub$weekIid = xSub$week
xSub$day = as.numeric(difftime(xSub$Date, min(weekValues), units = "days"))
#min(weekValues) = "1897-10-01 UTC" the initial date
xSub$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])
#as intial value for INLA

# 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())


#fit the model by inla
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),
        data = xSub, verbose=TRUE)


# randomly choose 100 samples from posterior samples
mySample = inla.posterior.sample(n = 100, result = sableRes, #change n from 24 to 100
                                 num.threads = 8,
                                 selection = list(week = seq(1,nrow(sableRes$summary.random$week))))


plot(xSub$Date, xSub$Max.Temp...C., col = mapmisc::col2html("black",0.3),
     xlim = c(ISOdate(2010, 1, 1, tz = "UTC"), ISOdate(2020, 1, 1, tz = "UTC"))
     , xlab = "time", ylab = "max degrees C", main = "summer months period from 2010 to present")
xaxPred = seq(ISOdate(2010, 1, 1, tz = "UTC"), by = "1 year", len = 20)
axis(1, xaxPred, format(xaxPred, "%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 = "max degrees C",
        xaxt = "n", col = "black", xaxs = "i",
        main = "estimated daily maximum temperature with 95% credible intervals") #time effect
forXaxis2 = ISOdate(seq(1880, 2052, by = 10), 1, 1,tz = "UTC")
axis(1, forXaxis2, format(forXaxis2, "%Y"))
abline(v = ISOdate(1900, 1, 1, tz = "UTC"), col = "blue") #pre-industrial level
abline(v = ISOdate(2019, 1, 1, tz = "UTC"), col = "blue")
abline(h = 11.6, col = "green") #max temp in 1990 (median)
abline(h = 12.8, col = "green")
abline(h = 13.2, col = "red") # about 1.5 above pre-industrial level
abline(v = ISOdate(2034, 1, 1, tz = "UTC"), col = "yellow")
```