Boston Buoy Data Analysis

Chenxun Li

9/21/2020

1.Introduction

The project provides evidence of golbal warning in the data collected by a single weather buoy in the NOAA National Data Buoy Center. In this project, assembling data, transforming data and sampling frequency, time series, etc are included.

2.Meterials and Methods

The whole work is organized as a RStudio(R version 4.0.2) project.And the code is shown below.

(1)Reading data

In this step, I read the data in the folder with loop function. At first, download the data(.txt) from the NOAA National Data center and save them in the folder.

```
##Set the certain file as Working Directory
setwd("~/Desktop/615/assignment")
##Enter "a" for all files in the Input folder
a <- list.files("data")
##Build path variable "dir"
dir <- paste("./data/",a,sep="")
##Read the length of "dir", which is the number of files in the folder.
n <- length(dir)
##Read the contents of the first file
##We don't have to read one first, but to keep things simple and save time defining "data.frame",
##I chose to read a file first.
merge.data <- read.table(file=dir[1],header=TRUE,fill=TRUE)

##The loop starts with the second file and reads in all the files and combines them into
##the "merge.data" variable
for(i in 2:n){
  new.data <- read.table(file=dir[i],header=TRUE,fill=TRUE)
  merge.data <- merge(merge.data,new.data,all=T,sort = T)
}

##Now,we can see 168800 obs. of 20 variables.
str(merge.data)
```

```
## 'data.frame':    168800 obs. of  20 variables:
##  $ YY  : int  2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 ...
##  $ MM  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ DD  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ hh  : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ WSPD: num  0.8 0.7 2.3 2.9 4.1 5.4 4.6 4.8 5 4 ...
```

```
##  $ GST : num  1.5 1.9 3.1 3.8 4.9 6.5 5.5 5.5 6.2 4.8 ...
##  $ WVHT: num  0.54 0.53 0.52 0.52 0.5 0.46 0.5 0.48 0.44 0.44 ...
##  $ DPD : num  10 4.55 4.76 10 10 4.55 10 10 10 10 ...
##  $ APD : num  4.55 4.61 4.78 4.86 5 4.94 5.23 4.65 5.18 5.08 ...
##  $ MWD : int  999 999 999 999 999 999 999 999 999 999 ...
##  $ ATMP: num  1.1 1.2 1.5 1.3 0.9 0.7 0.5 0.4 0.5 0.9 ...
##  $ WTMP: num  5.9 5.9 5.8 5.9 5.9 5.9 5.9 5.9 5.9 5.9 ...
##  $ DEWP: num  -3.6 -3.6 -3.3 -3.6 -4 -6.5 -5.2 -4.9 -4.6 -4.5 ...
##  $ VIS : num  99 99 99 99 99 99 99 99 99 99 ...
##  $ TIDE: num  NA NA NA NA NA NA NA NA NA NA ...
##  $ mm  : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ WDIR: int  NA NA NA NA NA NA NA NA NA NA ...
##  $ PRES: num  NA NA NA NA NA NA NA NA NA NA ...
##  $ WD  : int  315 271 232 236 232 228 240 235 242 244 ...
##  $ BAR : num  1019 1020 1020 1021 1021 ...
```

(2)Data processing

In this step, I use several functions to make the data in smaller quantity and easy for analyzing. We can find some problems in the data read just now, such as repetitive variables which include "WDIR"and "WD", "BAR" and"PRES". Then, we solve them and make data into appropriate form.

```
##Process missing values
merge.data[is.na(merge.data)] <- 0
##Merge repetitive values and delete repetitive variables
merge.data$PRES <- merge.data$PRES+merge.data$BAR
merge.data$WDIR <- merge.data$WDIR+merge.data$WD
merge.data <- merge.data[,-18]
merge.data <- merge.data[,-17]
merge.data <- merge.data[,-16]
dt <- merge.data

##Use zeros to pad out the values of "MM', "DD", "hh"
dt$MM <- formatC(dt$MM,flag = '0',width = 2)
dt$DD <- formatC(dt$DD,flag = '0',width = 2)
dt$hh <- as.numeric(formatC(dt$hh,flag = '0',width = 2))
##unite the time
dt_1 <- dt%>%unite('time','YY','MM','DD',sep='-')
dt_1$time <- as.Date(dt_1$time)
##remove outlier
dt_1 <- filter(dt_1,ATMP<40&WTMP<40)
##select the max/mean WTMP(sea surface temperature) of each day
dtWTMP_max <- tapply(dt_1$WTMP,dt_1$time,max)
dtWTMP_mean <- tapply(dt_1$WTMP,dt_1$time,mean)
##select the max/mean ATMP(air temperature) of each day
dtATMP_max <- tapply(dt_1$ATMP,dt_1$time,max)
dtATMP_mean <- tapply(dt_1$ATMP,dt_1$time,mean)

##Now, let's take look at an example:we can see the mean temperature of each day.
head(dtATMP_mean, 30)
```

```
##  2000-01-01  2000-01-02  2000-01-03  2000-01-04  2000-01-05  2000-01-06
##   2.6500000   6.3166667   9.7291667   7.6583333   4.7208333  -2.1958333
##  2000-01-07  2000-01-08  2000-01-09  2000-01-10  2000-01-11  2000-01-12
##   4.1625000   1.5791667   3.7333333   5.9750000   8.2583333   4.6500000
```

```
## 2000-01-13  2000-01-14  2000-01-15  2000-01-16  2000-01-17  2000-01-18 
## -1.2291667  -9.2000000 -11.2208333  -2.0500000  -9.5791667 -11.4041667 
## 2000-01-19  2000-01-20  2000-01-21  2000-01-22  2000-01-23  2000-01-24 
## -8.1916667  -5.2333333  -5.9500000 -11.9750000  -9.5913043  -2.4208333 
## 2000-01-25  2000-01-26  2000-01-27  2000-01-28  2000-01-29  2000-01-30 
##  1.3125000  -0.3333333  -5.8708333 -10.9000000  -6.6083333  -0.3333333
```

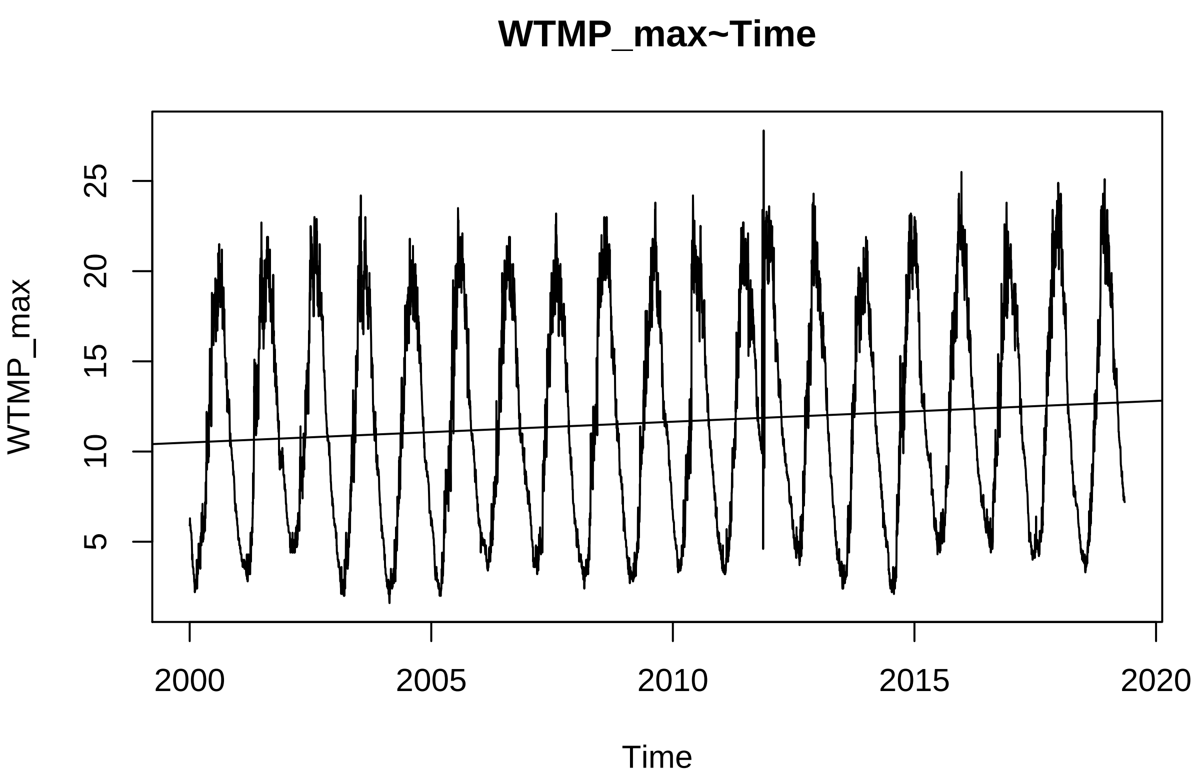
(3)Data analysis

In this step, I use time series to remove the seasonal effect.

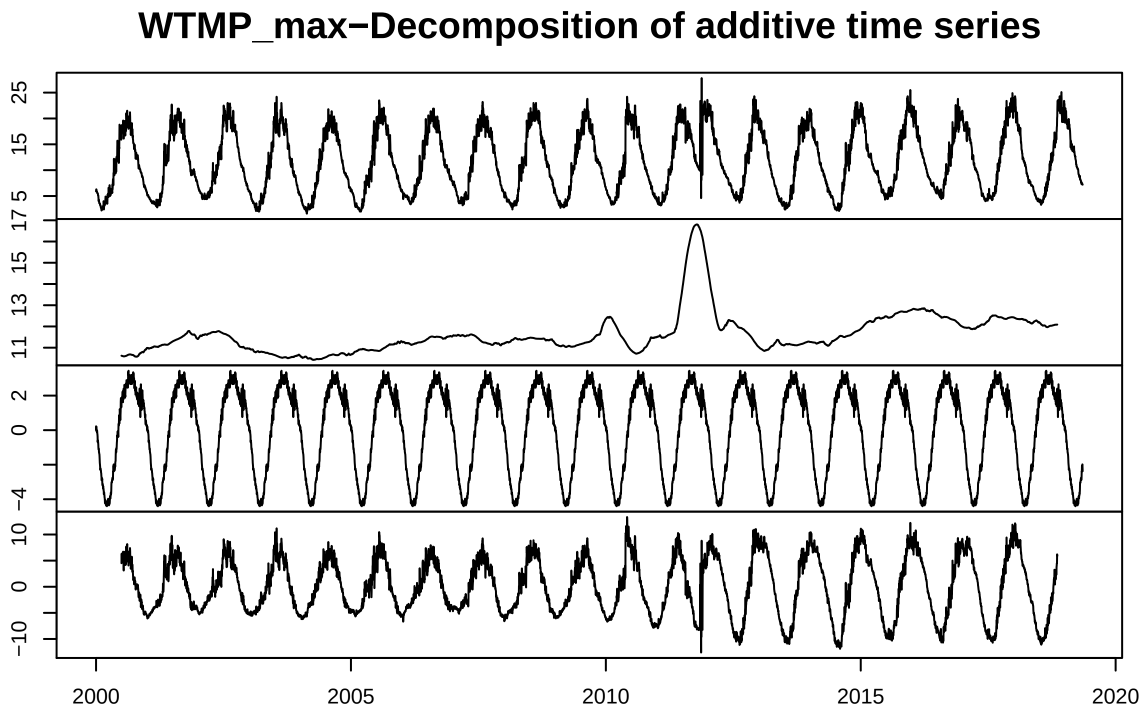
[1]max and mean of WTMP(sea surface temperature)

At first, I research in the sea surface temperature.

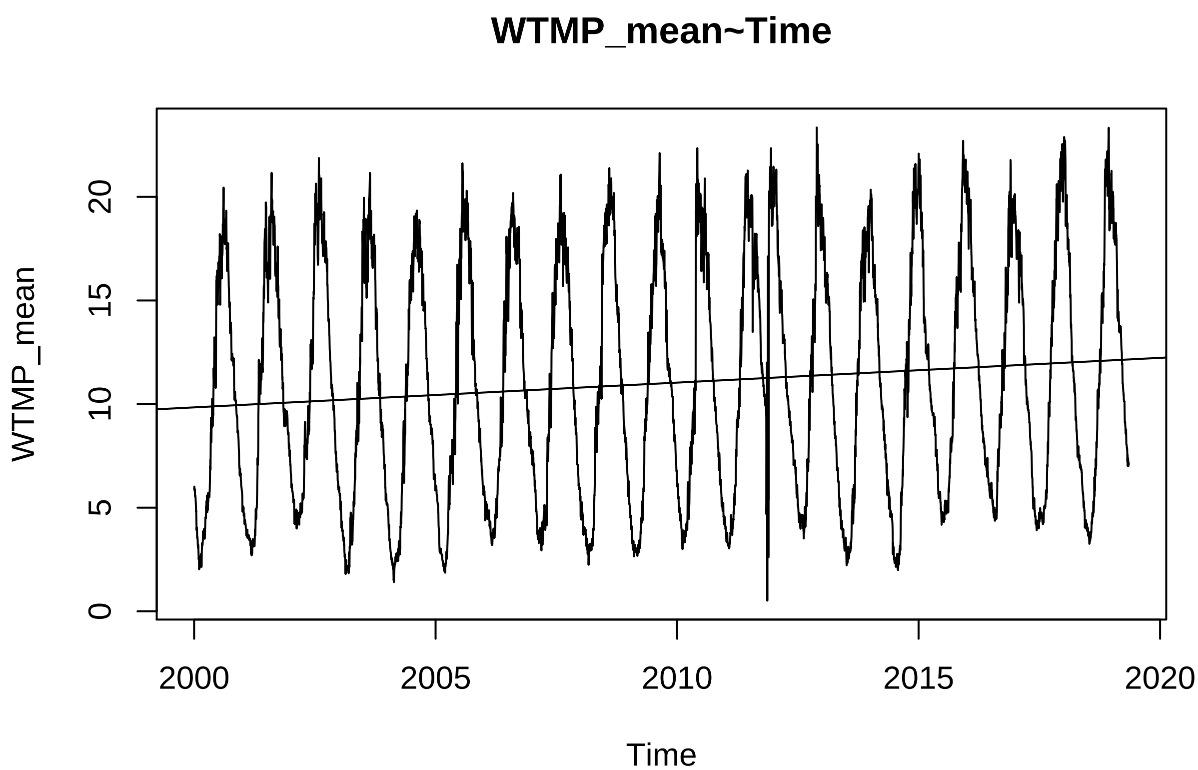
```r
##Change "dtWTMP_max" into a vector form
dtWTMP_max <- as.vector(dtWTMP_max)
##Read in the time series
WTMP_max <-  ts(dtWTMP_max,frequency=365,start = c(2000,1))
##Draw a time series diagram
plot(WTMP_max)
abline(lm(WTMP_max~time(WTMP_max)))
title(main="WTMP_max~Time")
```



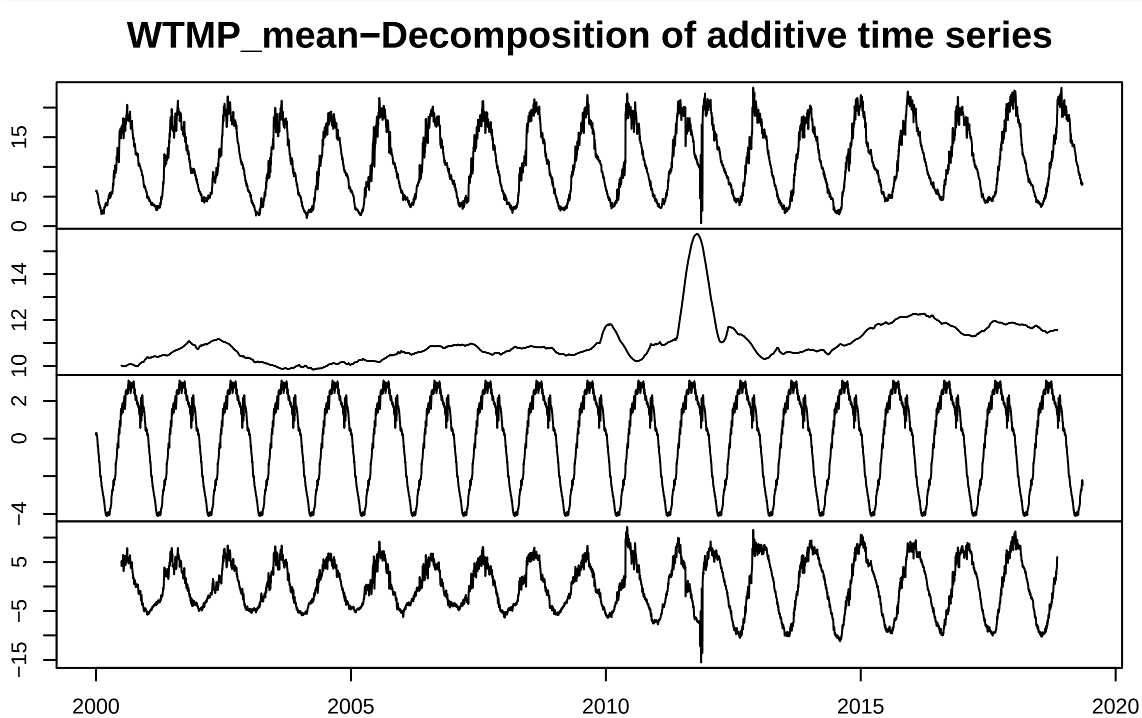

```r
##Estimate the trending, seasonal, and irregular parts of the time series.
WTMP_max_components <- decompose(WTMP_max)
plot(WTMP_max_components,ann=F)
title(main ="WTMP_max-Decomposition of additive time series")
```




```r
##The analysis of "dtWTMP_mean" is same as the "dtWTMP_max"
##Change "dtWTMP_max" into a vector form
dtWTMP_mean <- as.vector(dtWTMP_mean)
##Read in the time series
WTMP_mean <-  ts(dtWTMP_mean,frequency=365,start = c(2000,1))
##Draw a time series diagram
plot(WTMP_mean)
abline(lm(WTMP_mean~time(WTMP_mean)))
title(main="WTMP_mean~Time")
```



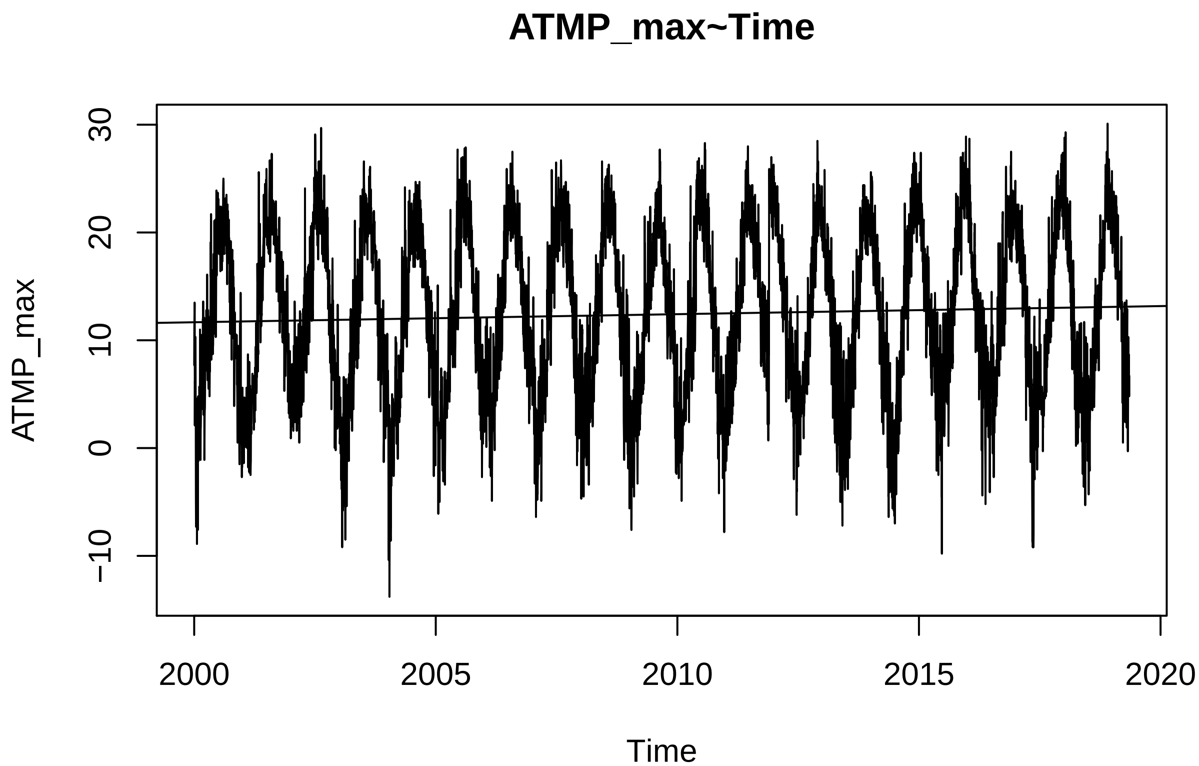

```r
##Estimate the trending, seasonal, and irregular parts of the time series.
WTMP_mean_components <- decompose(WTMP_mean)
plot(WTMP_mean_components,ann=F)
title(main ="WTMP_mean-Decomposition of additive time series")
```



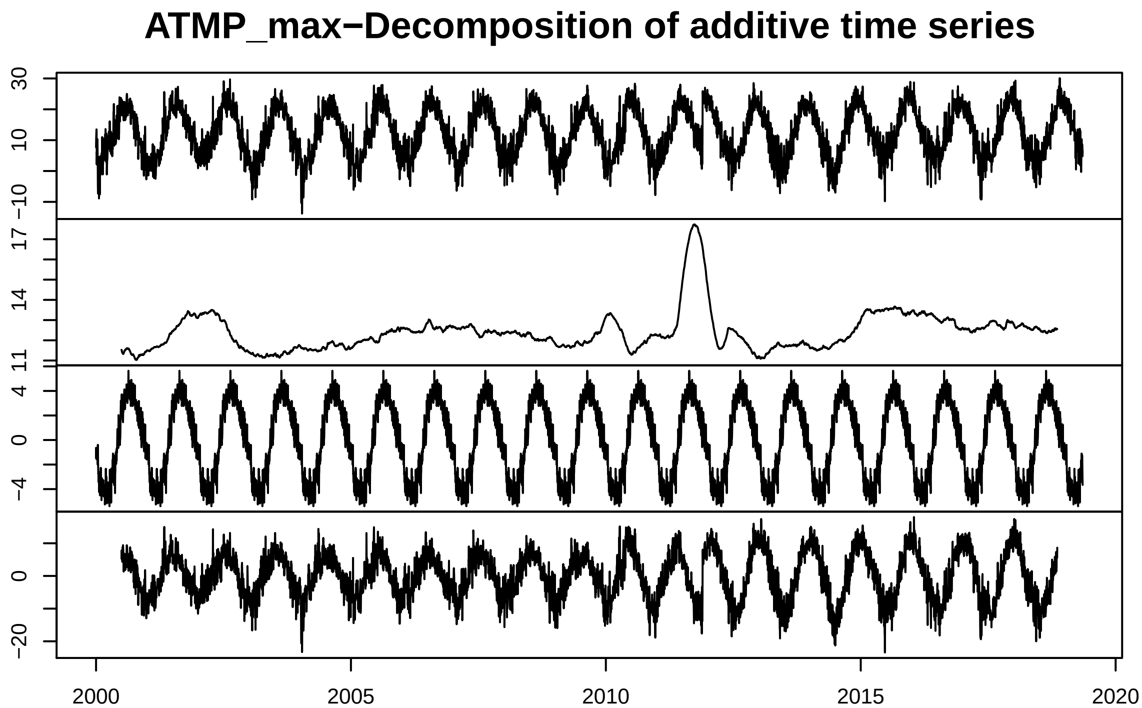
[2]max and mean of ATMP(air temperature)

Then, I research in the air temperature.

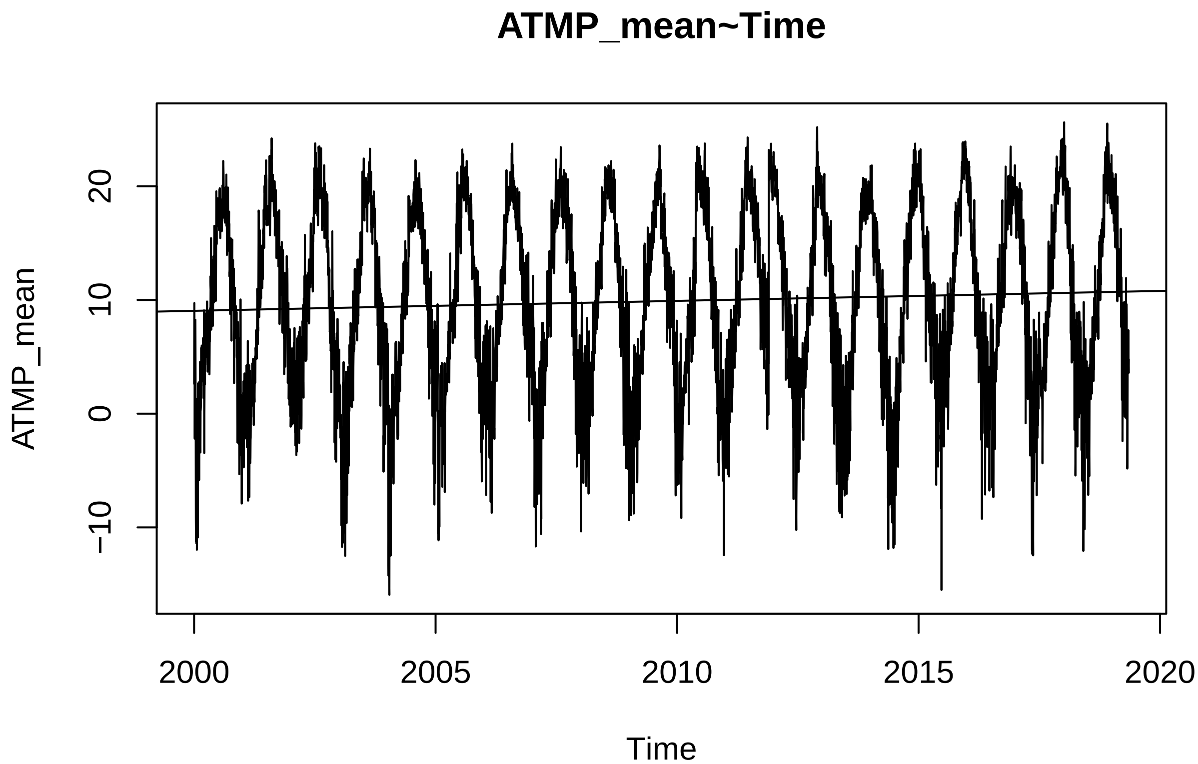
```
##Change "dtATMP_max" into a vector form
dtATMP_max <- as.vector(dtATMP_max)
##Read in the time series
ATMP_max <-  ts(dtATMP_max,frequency=365,start = c(2000,1))
##Draw a time series diagram
plot(ATMP_max)
abline(lm(ATMP_max~time(ATMP_max)))
title(main="ATMP_max~Time")
```



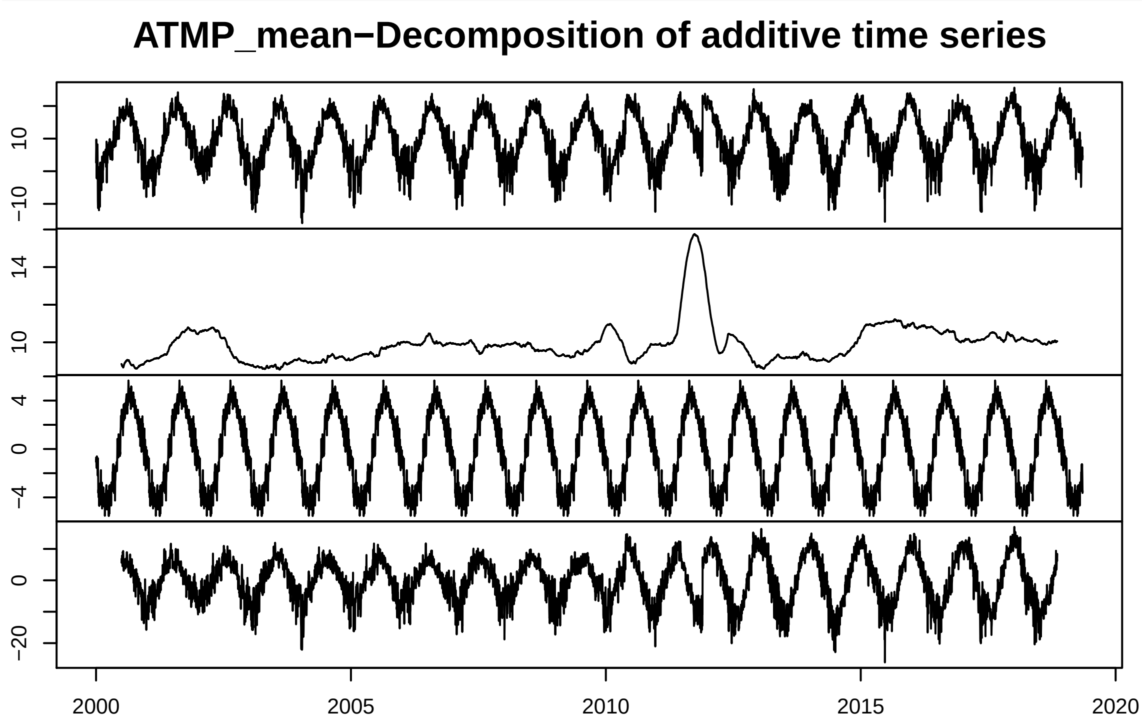

```
##Estimate the trending, seasonal, and irregular parts of the time series.
ATMP_max_components <- decompose(ATMP_max)
plot(ATMP_max_components,ann=F)
title(main ="ATMP_max-Decomposition of additive time series")
```




```
##The analysis of "dtATMP_mean" is same as the "dtATMP_max"
##Change "dtATMP_max" into a vector form
dtATMP_mean <- as.vector(dtATMP_mean)
##Read in the time series
ATMP_mean <-  ts(dtATMP_mean,frequency=365,start = c(2000,1))
##Draw a time series diagram
plot(ATMP_mean)
abline(lm(ATMP_mean~time(ATMP_mean)))
title(main="ATMP_mean~Time")
```




```r
##Estimate the trending, seasonal, and irregular parts of the time series.
ATMP_mean_components <- decompose(ATMP_mean)
plot(ATMP_mean_components,ann=F)
title(main ="ATMP_mean-Decomposition of additive time series")
```



3.Result and Discussion

(1)Result analysis

At first , from the four time series diagrams, we can find that the line' slope is above zero, which means that the max and the mean of the sea surface temperature increases with time passing. So as the air temperature. Then, from the four "decomposition of additive time series" diagrams, I seperate the seasonal effect from the data and we can also find max and the mean of the sea surface temperature increases with time passing with smaller seasonal influence. And so as the air temperature.

(2)Considerations and limitations

[1]There is a peak in every "decomposition of additive time series" diagram, and I think probably one year is warmer than others.

[2]The function "tapply()" produce a multidimensional array, in this research, I change it to a vector, but I wonder how to change it into a matrix.

[3]I read the data from the folder this time, and next time I should try to read the data from the url.

4.Conlusion

From the research above, it is indeed that global is warmer from the data collected by a single weather buoy in the NOAA National Data Buoy Center.

5.Reference

(1)R packages:tidyverse, tidyr, forecast, tseries, lubridate.

(2)Hadley Wickham & Garrett Grolemund(2017). R for data science