Buoy Data Aanalysis

Donghao Xue

09/25/2020

Instruction

For the problem of global warming, the most important thing is to see the fluctuation in temperature at a given time period. We have collected 20-years of data from NDBC station and in this program, I want to mainly focus on the changes of water temperature. So, I approach this problem in three ways. Firstly, I make a plot of the daily water temperature at a fixed time (12pm) during the 20 years. Then I fit a regression line to the data to see the trend. Second, I make a plot of average water temperature of December in each year with regression line. Finally I make a plot of maximum temperature of each year and also summarize the regression.

Analysis

Assemble the data into a single data frame

Comparing daily water temperature at 12pm

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------------- tidyverse 1.3.0 --
```

```
## √ ggplot2 3.3.2     √ purrr   0.3.4
## √ tibble  3.0.3     √ dplyr   1.0.2
## √ tidyr   1.1.2     √ stringr 1.4.0
## √ readr   1.3.1     √ forcats 0.5.0
```

```
## -- Conflicts ---------------------------------------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(rstanarm)
```

```
## Loading required package: Rcpp
```

```
## This is rstanarm version 2.21.1
```

```
## - See https://mc-stan.org/rstanarm/articles/priors for changes to default priors!
```

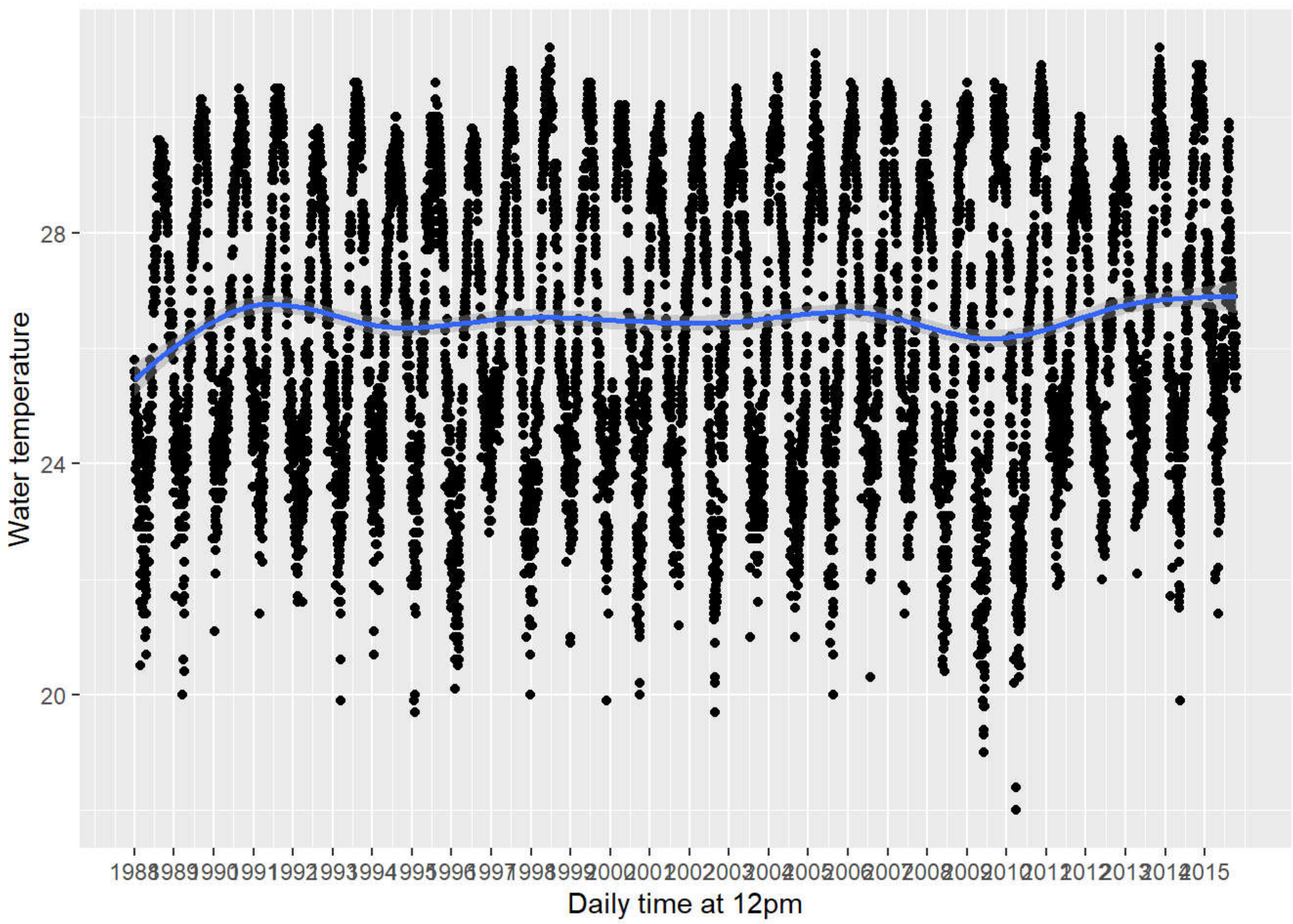
```
## - Default priors may change, so it's safest to specify priors, even if equivalent to the def
aults.
```

```
## - For execution on a local, multicore CPU with excess RAM we recommend calling
```

```
##   options(mc.cores = parallel::detectCores())
```

```
library(ggplot2)

Date = read.csv("MR_DATE.csv")

Date_1 = filter(Date, Date$WTMP<800, Date$ATMP<800)

x1 = grep("12:00:00", Date_1$DATETIME)
y1 = Date_1$WTMP[x1]
x2 = c(1:length(x1))
data1 = data.frame(x2,y1)

ggplot(data=data1, mapping=aes(x = x2, y = y1))+geom_point()+geom_smooth()+scale_x_continuous(b
reaks=seq(0,10146,365),labels=c(1988:2015))+
  xlab("Daily time at 12pm")+ylab("Water temperature")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

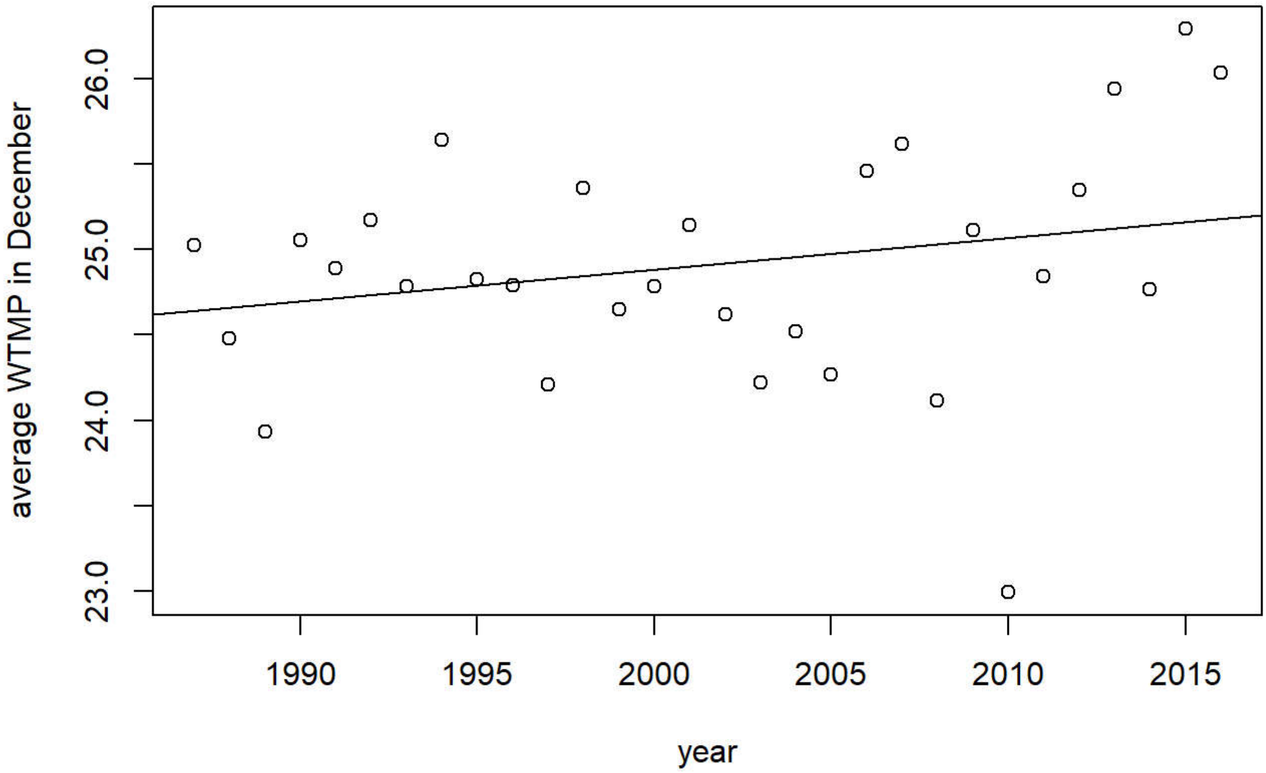


```
#fit a regression
fit_1 = stan_glm(y1~x2,data = Date_1, refresh=0)

summary(fit_1, digits=5)
```

```
##
## Model Info:
##  function:     stan_glm
##  family:       gaussian [identity]
##  formula:      y1 ~ x2
##  algorithm:    sampling
##  sample:       4000 (posterior sample size)
##  priors:       see help('prior_summary')
##  observations: 10146
##  predictors:   2
##
## Estimates:
##               mean     sd       10%      50%      90%
## (Intercept) 26.34618  0.05049 26.28105 26.34592 26.41042
## x2           0.00003  0.00001  0.00002  0.00003  0.00004
## sigma        2.50560  0.01822  2.48244  2.50554  2.52929
##
## Fit Diagnostics:
##            mean     sd       10%      50%      90%
## mean_PPD 26.48642  0.03545 26.44061 26.48641 26.53135
##
## The mean_ppd is the sample average posterior predictive distribution of the outcome variable
(for details see help('summary.stanreg')).
##
## MCMC diagnostics
##               mcse    Rhat    n_eff
## (Intercept)   0.00081 0.99979 3896
## x2            0.00000 0.99975 3995
## sigma         0.00029 0.99996 3923
## mean_PPD      0.00058 0.99929 3731
## log-posterior 0.03327 1.00095 1437
##
## For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effectiv
e sample size, and Rhat is the potential scale reduction factor on split chains (at convergence
Rhat=1).
```

comparing average water temperature of December every year

```
x1 = grep("-12-",Date_1$DATETIME)

y1 = Date_1$WTMP[x1]

y2 =c()
for (i in 1987:2016){
  y2= append(y2, mean(y1[grep(i, Date_1$DATETIME[x1])]))
}


x = c(1987:2016)
plot(x , y2 , xlab = "year", ylab = "average WTMP in December")
fit_2 = stan_glm(y2 ~ x, data = data.frame(x,y2), refresh=0)
abline(fit_2)
```




```
print(fit_2,digits=5)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      y2 ~ x
##  observations: 30
##  predictors:   2
## ------
##             Median    MAD_SD
## (Intercept) -12.22625  29.67821
## x             0.01855   0.01483
##
## Auxiliary parameter(s):
##       Median  MAD_SD
## sigma 0.67814 0.09336
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

comparing maximum temperature every year

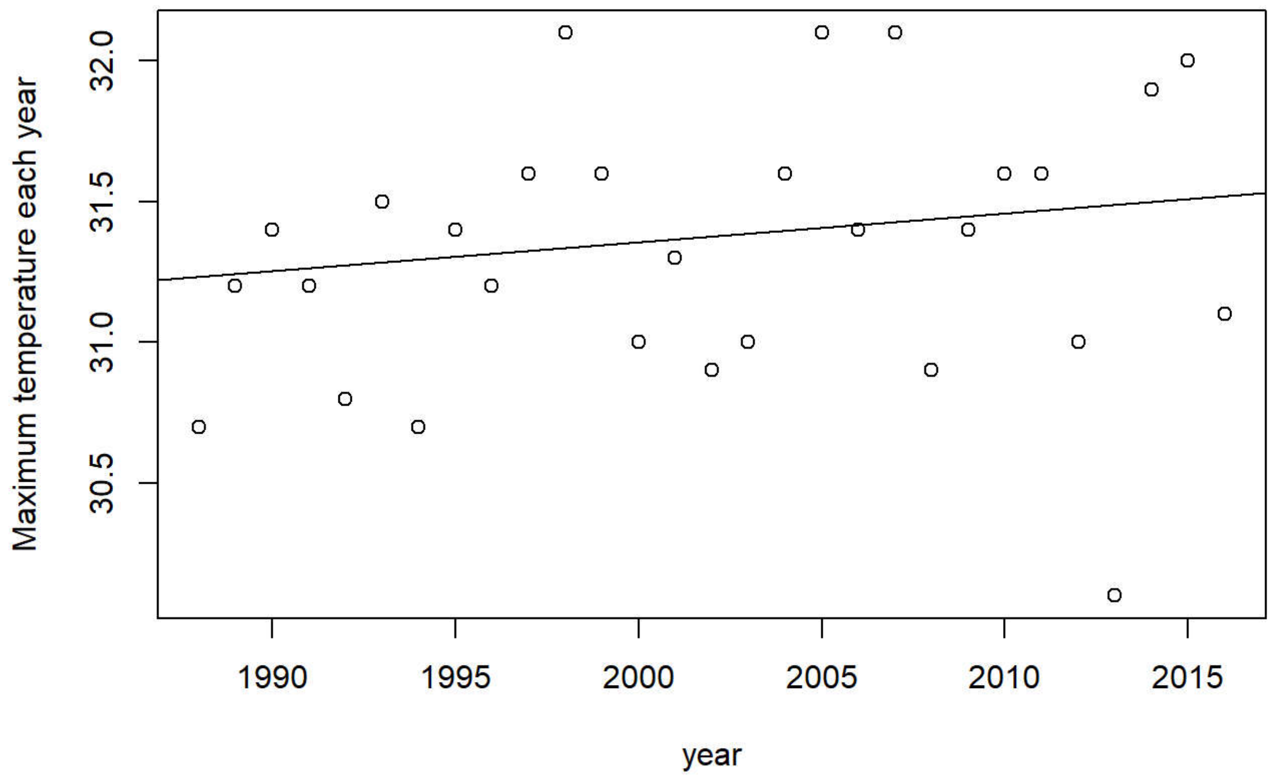
```
y1 = c()
for(i in 1988:2016){
  y1 = append(y1, max(Date_1$WTMP[grep(i, Date_1$DATETIME)]))

}

x = c(1988:2016)

x1 = grep(3898, Date_1$DATETIME)
y = max(Date_1$WTMP[x1])
```

```
## Warning in max(Date_1$WTMP[x1]): max里所有的参数都不存在；回覆-Inf
```

```
y1[11]=32.1

plot(x , y1 , xlab = "year", ylab = "Maximum temperature each year")

fit_3 = stan_glm(y1 ~ x, refresh=0)
```

```
## Warning: Omitting the 'data' argument is not recommended and may not be allowed
## in future versions of rstanarm. Some post-estimation functions (in particular
## 'update', 'loo', 'kfold') are not guaranteed to work properly unless 'data' is
## specified as a data frame.
```

```
abline(fit_3)
```




```
print(fit_3,digits=5)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      y1 ~ x
##  observations: 29
##  predictors:   2
## ------
##             Median   MAD_SD
## (Intercept) 10.98682 21.81504
## x            0.01018  0.01090
##
## Auxiliary parameter(s):
##       Median  MAD_SD
## sigma 0.48196 0.06680
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

Discussion

From the plot of daily temperature, it shows that the trend of temperature is fluctuant, and from the summary of regression, the slope of the regression line is very close to zero. Therefore it is not evident to conclude that the temperature is increasing.

From the plot of average temperature in December for each year, it clearly shows that the there is a positive slope of regression line. From the summary of regression, the slope of regression line is 0.01 which is more evident than the slope of regression line in the last plot.

From the plot of maximum temperature in each year, the slope of the regression line is also about 0.01. This result can show that there is a slightly increasing tendency of maximum temperature in each year.

Conclusion

In general, the second and third plot give us evidence that there is a slightly increasing trend in water temperature with each passing year. However, due to the sample size is not very large, the subtle change in temperature cannot prove the existence of global warming directly. I think we need more data to analysis and include more factors so that we can get a more accurate result to show the problem of global warming.