Buoy Report

Yuxin Zeng

Abstract

In order to study whether there is evidence of global warming in the data collected by a single weather buoy in the NOAA National Data Buoy Center, I study the relationship between the air temperature measured by buoy and the year.First, I choose the data from 1997 to 2016 as the sample, and use the data at 12AM to represent the data every day. After preprocessing the data, I calculated the mean temperature of each month for each year, used it to regress the years, and visualized these regressions. Then I built a Seasonal ARIMA model to find the law and predict the next three years. Only the average temperature of July and August has a very slight upward trend over years. After removing the seasonal influence, there is no obvious upward trend in Seasonal ARIMA. Neither of these results provide a clear evidence of increase in air temperature.

1.Background Information

"Global warming" is an environmental issue that has been widely discussed. We hope to find out whether there is evidence of global warming in the data (esp. air temperature) collected by a single weather buoy in the NOAA National Data Buoy Center. The buoy we choose is from Station 44013 —— BOSTON 16 NM East of Boston, MA(42°20'44" N 70°39'4" W).

2.Data Collection

Make URLsand read the data from the website. Since data from different year has different columns, we seperate them and only take the columns we want. Combine all dataframes.

3.Data Preprocessing

Select 20 years of data from 1997 to 2016. Choose the data at 12:00AM to represent the data of each day. Use air temperature (ATMP) as a measure of global warming, deleting the rest variables except time variables. Air temperature greater than 99.0 Celsius are regarded as abnormal points, therefore are deleted. Use the unite function to change the date format.

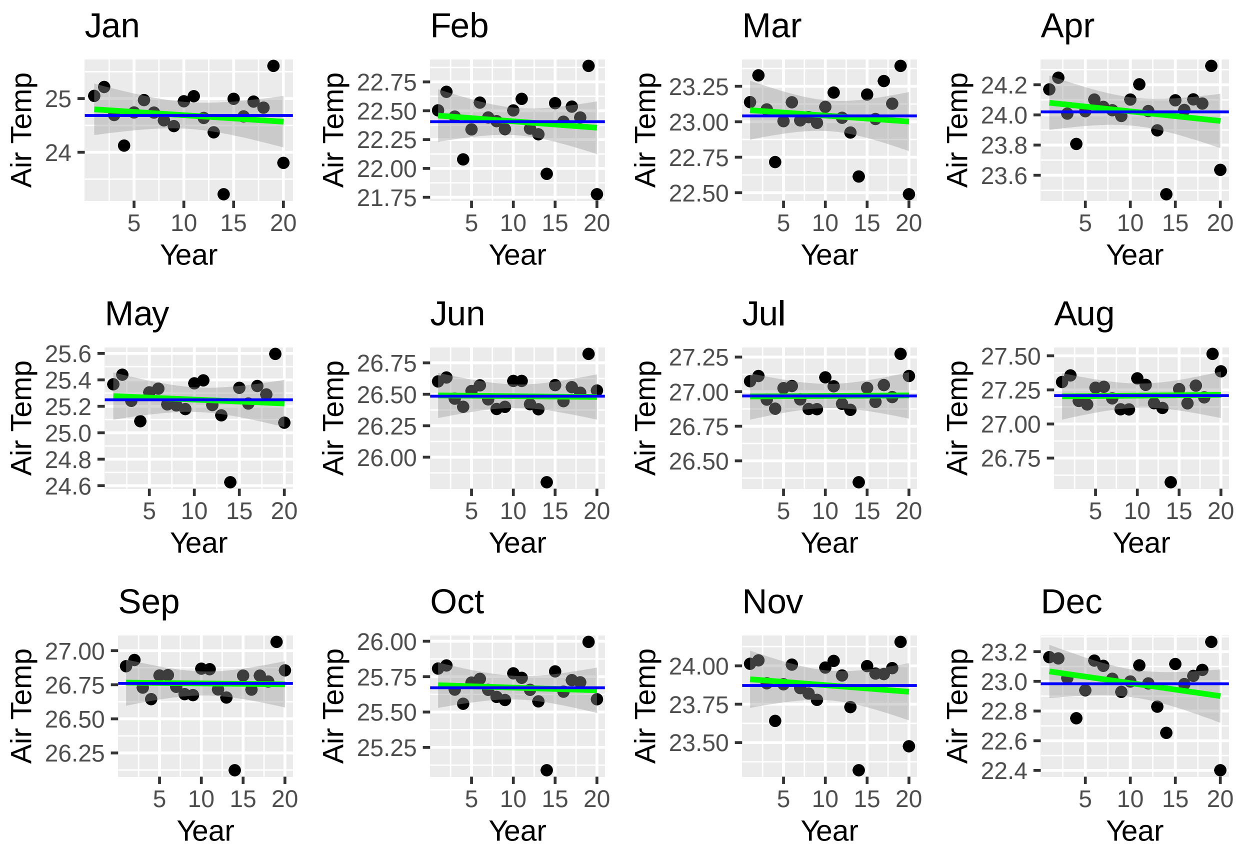
```r
load("C:/Users/lenovo/Desktop/615 R/buoy/MR.RData")
buoy = data.frame(MR)
buoy_1 = buoy %>% filter(YYYY >= "1997")
buoy_2 = buoy_1 %>% filter(hh == "12")
buoy_3 = buoy_2[,c(1,2,3,13)]
buoy_4 = buoy_3[buoy_3$ATMP<99.0,]
buoy_5 = buoy_4 %>% unite(date,YYYY,MM,DD,sep="-")

write.csv(buoy_4,file="C:/Users/lenovo/Desktop/615 R/buoy/Everyday.csv")
write.csv(buoy_5,file="C:/Users/lenovo/Desktop/615 R/buoy/Everyday_date.csv")
```

4.Data Analysis & Visualization

Calculate the average temperature for each month from 1997 to 2016.

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



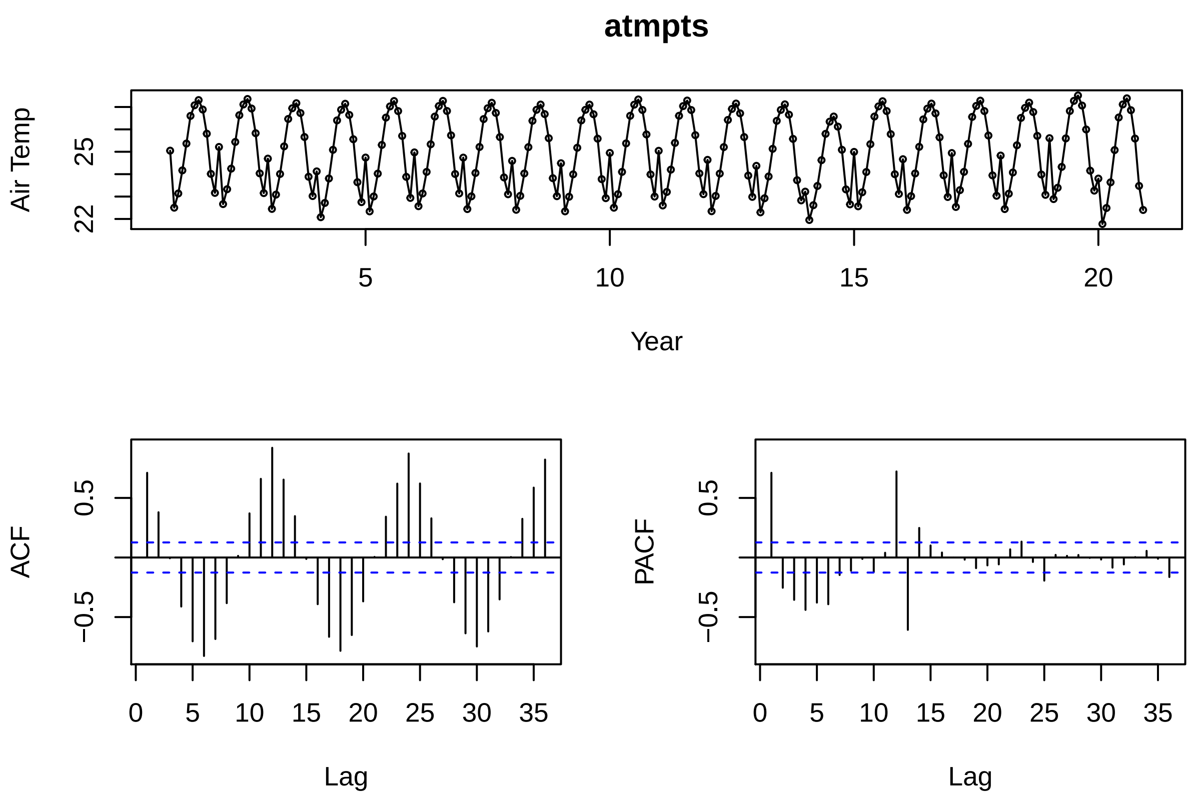
The above figures show the average air temperature of each month from 1997 to 2016.In each figure, the green line is the fitting result of the temperature regression to the year.It can be compared with the blue horizontal line, which is the average temperature of this month.From the figure, it is found that the monthly average air temperature in summer (Jul and Aug) shows a slight upward trend, while the monthly average air temperature in the rest of the year has a slight downward trend.

5.Seasonal ARIMA

#Data preparation: Read the monthly average temperature for 20 years and process it into a chronological vector. The first element of the vector is the average temperature in January 1997, and the last element is the average temperature in December 2016. Convert it into time series format using ts function.

```
atmp=read.csv(file="atmp.csv",header=TRUE)
atmpt=t(atmp)
```

```
atmpt1=atmpt[-c(1,2),]
atmpv=as.vector(atmpt1)
atmpts=ts(atmpv,frequency=12)
tsdisplay(atmpts,xlab="Year",ylab="Air Temp")
```



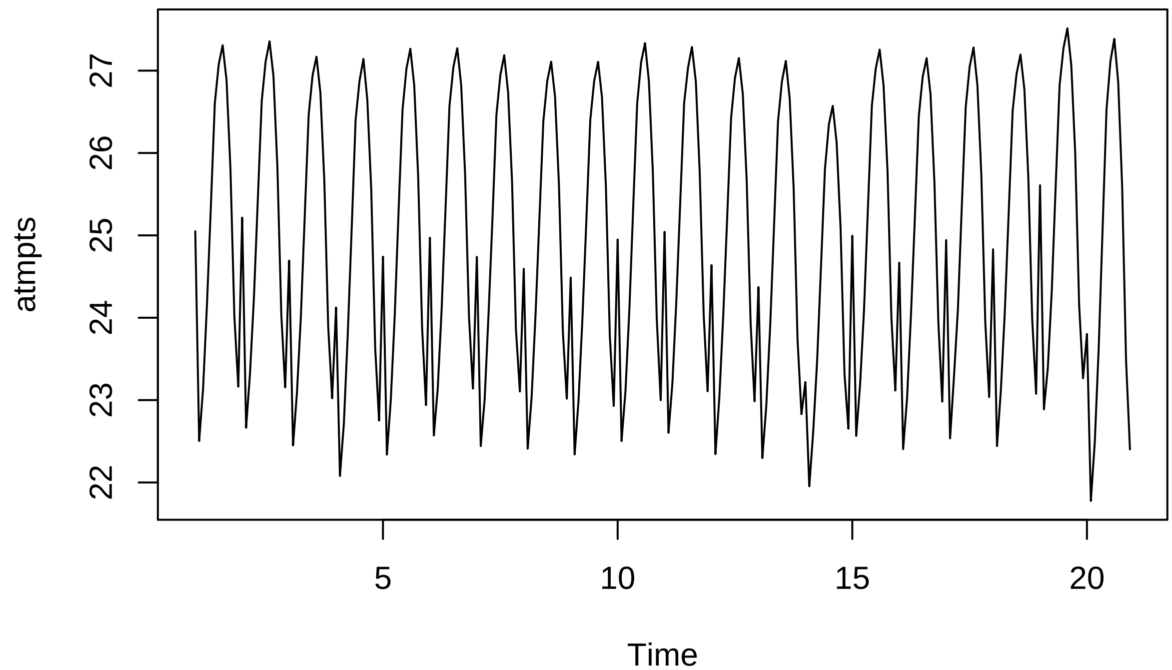

```
write.csv(atmpv,file="C:/Users/lenovo/Desktop/615 R/buoy/atmpv.csv")
```

#White Noise Test & Stationary Test

```
Box.test(atmpts)
```

```
##
##  Box-Pierce test
##
## data:  atmpts
## X-squared = 121.3, df = 1, p-value < 2.2e-16
```

```
plot.ts(atmpts)
```




```r
adf.test(atmpts)
```

```
## Warning in adf.test(atmpts): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  atmpts
## Dickey-Fuller = -13.572, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

Use Box.test function to complete the white noise test. The result (p-value is less than 0.05) shows that the sequence is correlated, therefore it is a non-white noise sequence. This time series can be used for modeling. According to the time series chart, the series has no obvious trend. Use adf.test function to complete the stationary test. The result (p-value is less than 0.05) shows the sequence is stationary.

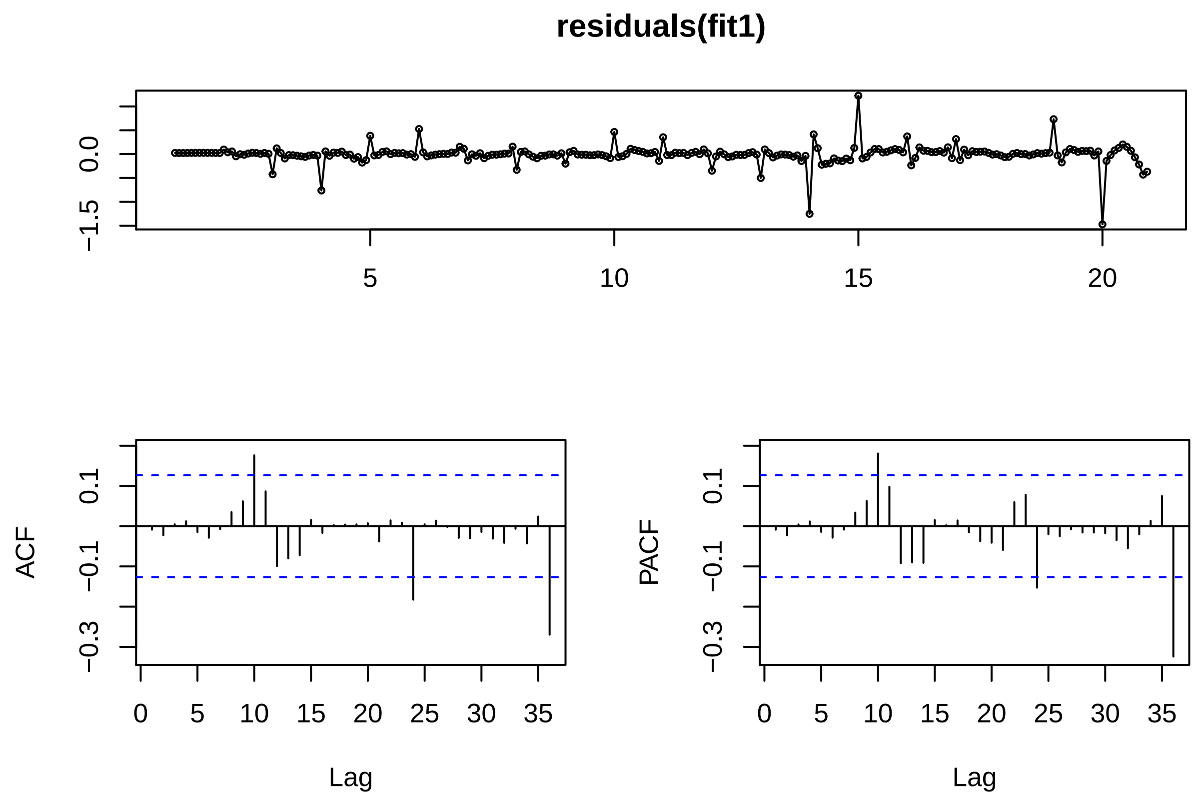
#Seasonal ARIMA

```r
#Fit the Time Series
fit1=auto.arima(atmpts)
print(fit1)
```

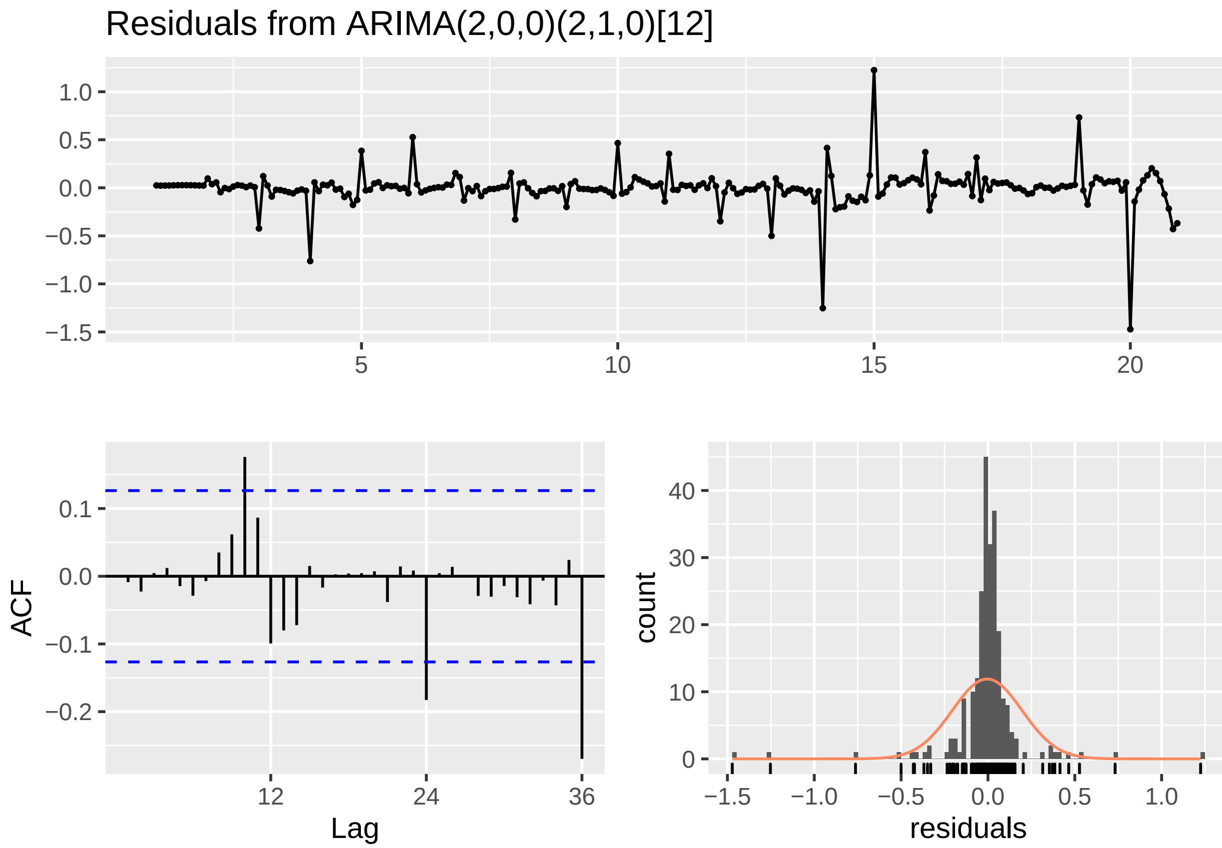
```
## Series: atmpts
## ARIMA(2,0,0)(2,1,0)[12]
##
## Coefficients:
##          ar1     ar2     sar1     sar2
##       0.5639  0.2100  -0.6818  -0.2303
## s.e.  0.0652  0.0664   0.0759   0.0755
```

```
##
## sigma^2 estimated as 0.04313:  log likelihood=33.77
## AIC=-57.54   AICc=-57.27   BIC=-40.4
```

```r
#Check the Residuals
tsdisplay(residuals(fit1))
```

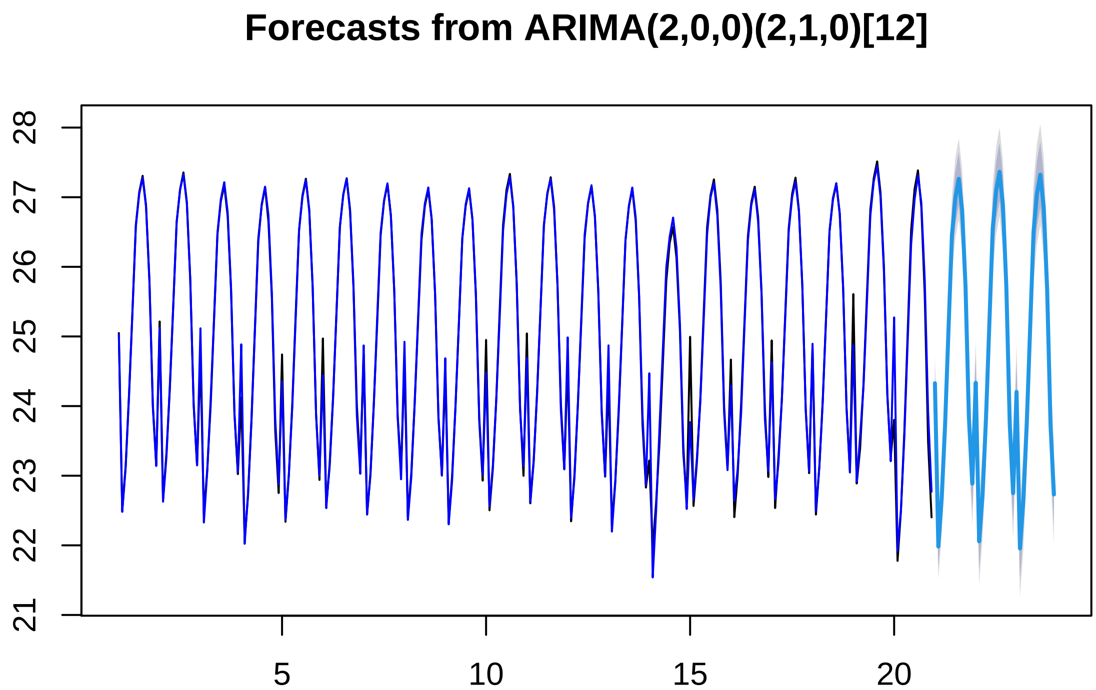



```r
checkresiduals(fit1)
```

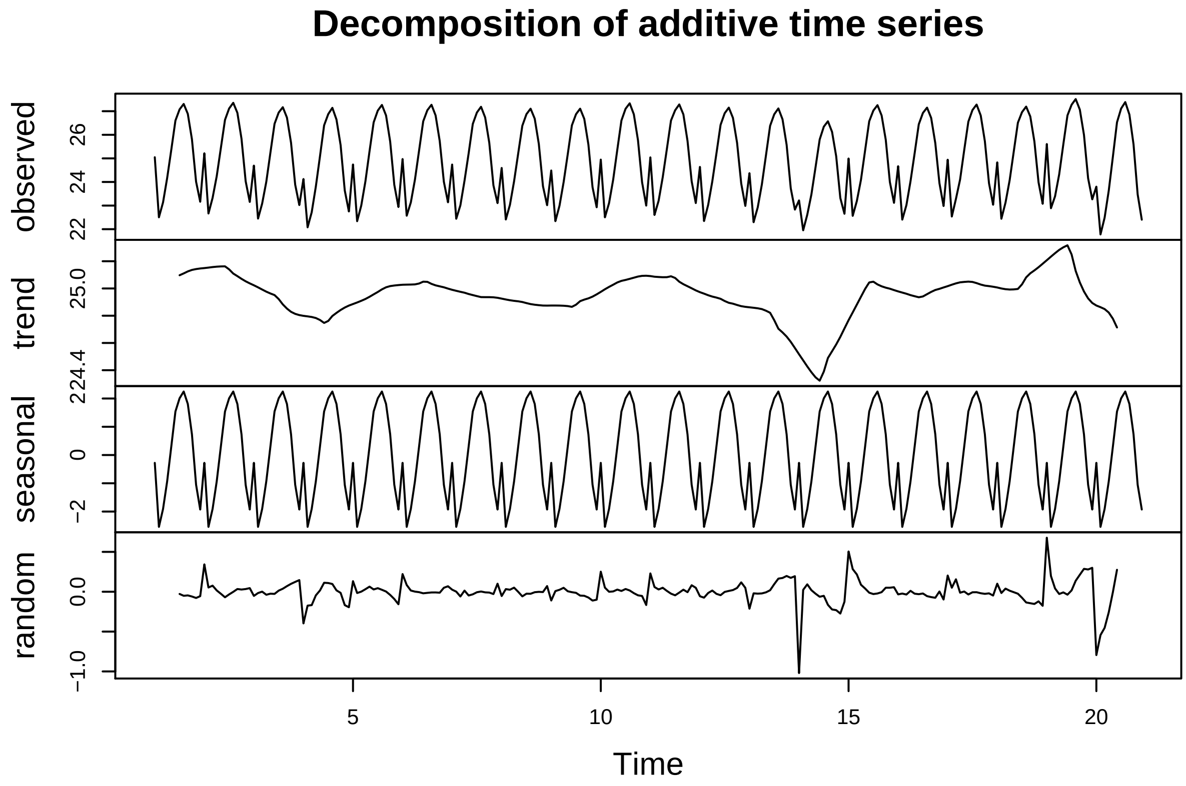



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,0,0)(2,1,0)[12]
## Q* = 26.551, df = 20, p-value = 0.1484
##
## Model df: 4.   Total lags used: 24
```

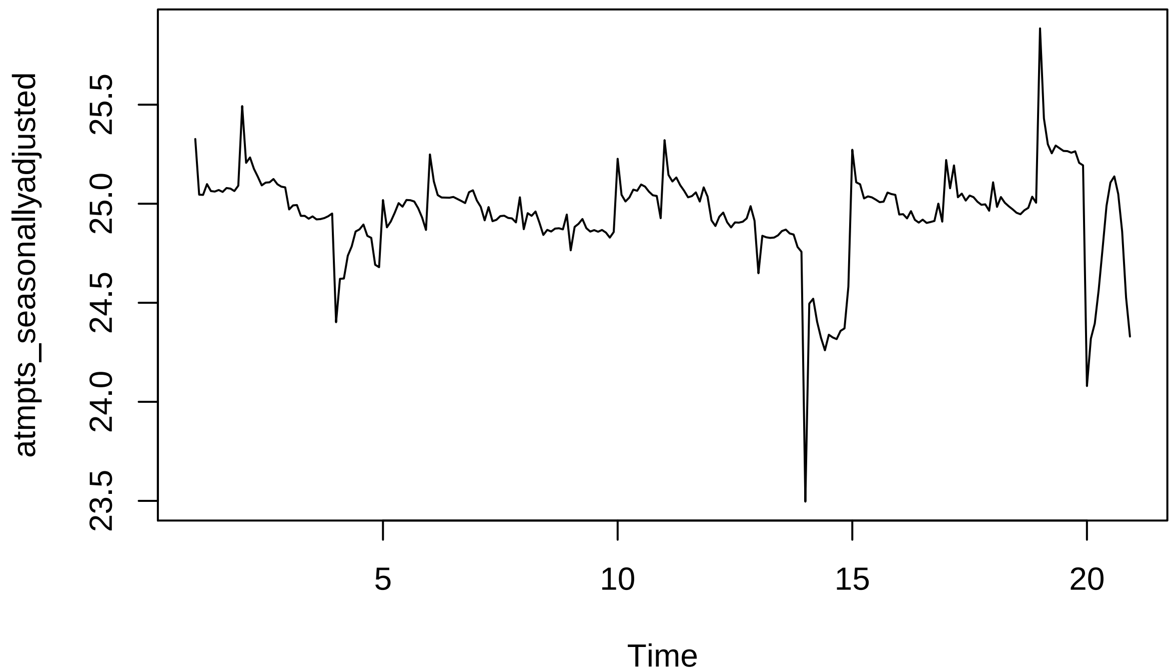
```
#Predict Air Temperature of the Next Three Years
fore1=forecast(fit1,h=3*12)
plot(fore1)
lines(fore1$fitted,col="blue")
lines(atmp,col="yellow")
```




```
#Eliminate Seasonal Effects
atmpts_components=decompose(atmpts)
plot(atmpts_components)
```




```
atmpts_seasonallyadjusted=atmpts-atmpts_components$seasonal
plot(atmpts_seasonallyadjusted)
```

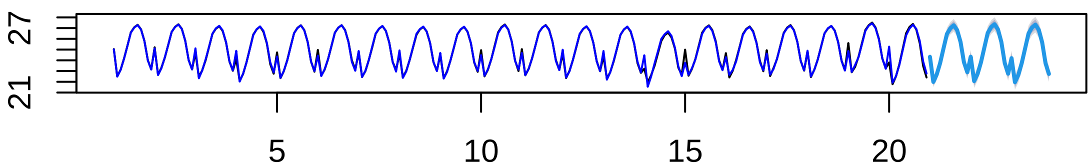



```
fit2=auto.arima(atmpts_seasonallyadjusted)
print(fit2)
```

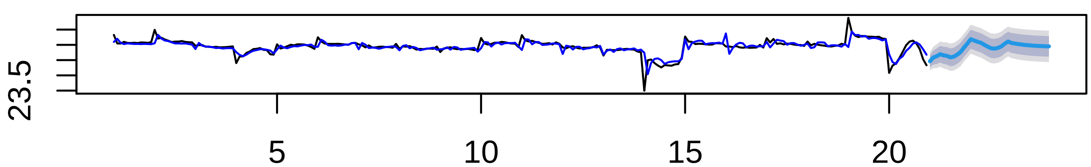
```
## Series: atmpts_seasonallyadjusted
## ARIMA(2,0,0)(0,0,2)[12] with non-zero mean
##
## Coefficients:
##          ar1     ar2     sma1     sma2     mean
##       0.6096  0.1882  -0.4492  -0.3152  24.9321
## s.e.  0.0650  0.0657   0.0698   0.0666   0.0149
##
## sigma^2 estimated as 0.02364:  log likelihood=106.31
## AIC=-200.62   AICc=-200.26   BIC=-179.74
```

```
fore2=forecast(fit2,h=3*12)
par(mfrow=c(2,1))
plot(fore1)
lines(fore1$fitted,col="blue")
lines(atmp,col="yellow")
plot(fore2)
lines(fore2$fitted,col="blue")
lines(atmp,col="yellow")
```

Forecasts from ARIMA(2,0,0)(2,1,0)[12]



Forecasts from ARIMA(2,0,0)(0,0,2)[12] with non−zero mean



Use the auto.arima function to establish a Seasonal ARIMA model for the temperature time series, and get the parameter (2,0,0)(2,1,0)[12]. Observe the autocorrelation coefficient of the residual,except for the periodic (12) spike, the value of acf is basically within the range of the dotted line, which shows that the collinearity in the model generated by Seasonal ARIMA is not serious.The residuals are close to a normal distribution. Using this model to predict the monthly average temperature of the next three years, the fluctuation range has not changed much compared with the past, and three cycles are clear.

Use decompose function (type is additive) to seperate temperature into "random", "seasonal", "trend", and "observed". After removing seasonal factors, the periodicity of the time series is no longer obvious. At the same time, the "trend" in the decompose figure can also confirm that the time series is stationary. Use the auto.arima function to establish a Seasonal ARIMA model for new sequence.The parameter this time is (2,0,0)(0,0,2)[12]. Use new sequence to predict the next three years.

6.Conclusion

According to the result of data analysis & visualization and Seasoanl ARIMA model,my conslusions are as follows: (1)Only the average temperature of July and August has a very slight upward trend over years (the positive slope is very close to zero); (2)Integrating all the monthly average temperatures of 20 years, there is a strong periodicity. After removing the seasonal influence, there is no obvious upward trend.

Therefore I would like to draw the conclusion that there may not be an evidence of global warming in the data collected by a single weather buoy in the NOAA National Data Buoy Center.

7.Consideration

This study uses air temperature (atmp) as a measure of global warming, which is the most intuitive. However, other factors can be comprehensively considered, such as the weighted average of air temperature and water surface temperature (wtmp).And for the years with "TIDE" records, we can study the changes of relative water level. I only use 20 years of data (from 1997 to 2016) according to the project requirement, and there is no clear upward pattern. If 30 years of data were used, the trend might be a little more obvious.

8.Refernce

#Website Refernce:

National Data Buoy Center

#Package reference:

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Goodrich B, Gabry J, Ali I & Brilleman S. (2020). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.21.1 https://mc-stan.org/rstanarm.

Brilleman SL, Crowther MJ, Moreno-Betancur M, Buros Novik J & Wolfe R. Joint longitudinal and time-to-event models via Stan. StanCon 2018. 10-12 Jan 2018. Pacific Grove, CA, USA. https://github.com/stan-dev/stancon_talks/

Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL http://www.jstatsoft.org/v40/i03/.

David Stoffer (2020). astsa: Applied Statistical Time Series Analysis. R package version 1.10. https://CRAN.R-project.org/package=astsa

Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.29.

Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963