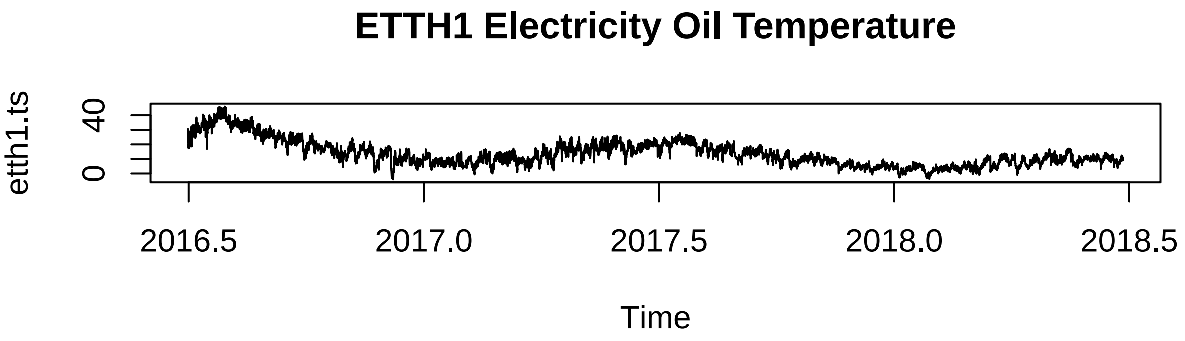
Traditional Methods 207 Final Project

Dylan Chou, Tim Yao

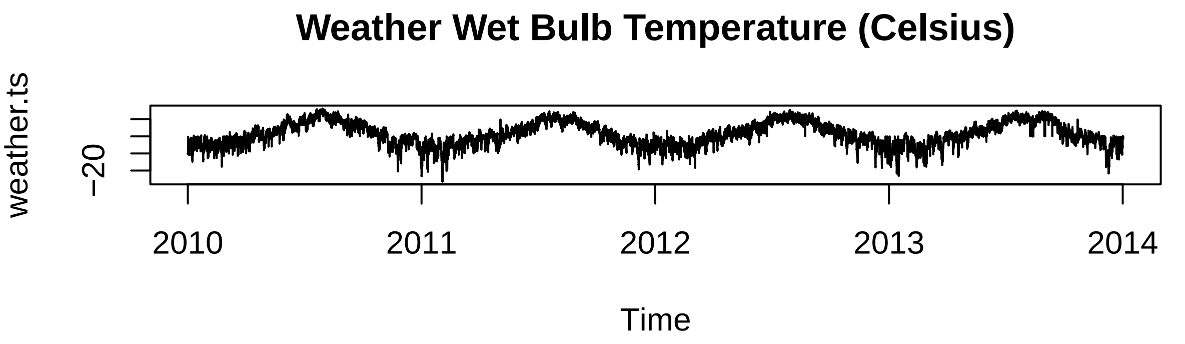
2024-05-30

Two Time Series Plots of Electricity and Weather

```r
par(mfrow=c(2,1))
etth1 = read.csv("./data/ETTh1.csv")
weather = read.csv("./data/WTH.csv")
# univariate time series
# https://stackoverflow.com/questions/33782218/how-to-create-a-time-series-of-hourly-data
first_hour_etth1 = 24*(as.Date("2016-07-01 00:00:00")-as.Date("2016-1-1 00:00:00"))
etth1.ts = ts(data=etth1$OT, start=c(2016, first_hour_etth1), freq=24*365)
weather.ts = ts(data=weather$WetBulbCelsius, start=c(2010, 0), freq=24*365)
plot(etth1.ts, main="ETTH1 Electricity Oil Temperature")
plot(weather.ts, main="Weather Wet Bulb Temperature (Celsius)")
```



ETTH1 Electricity Oil Temperature



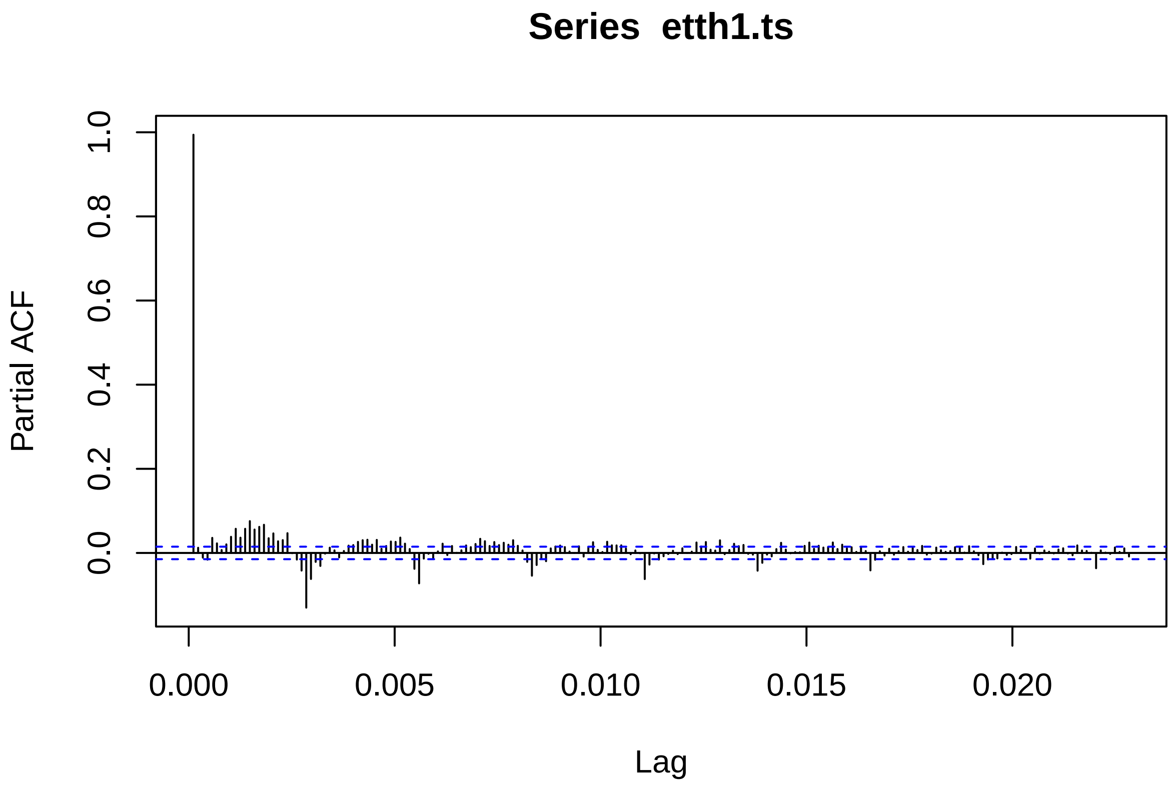
Weather Wet Bulb Temperature (Celsius)

Autocorrelation of current time series

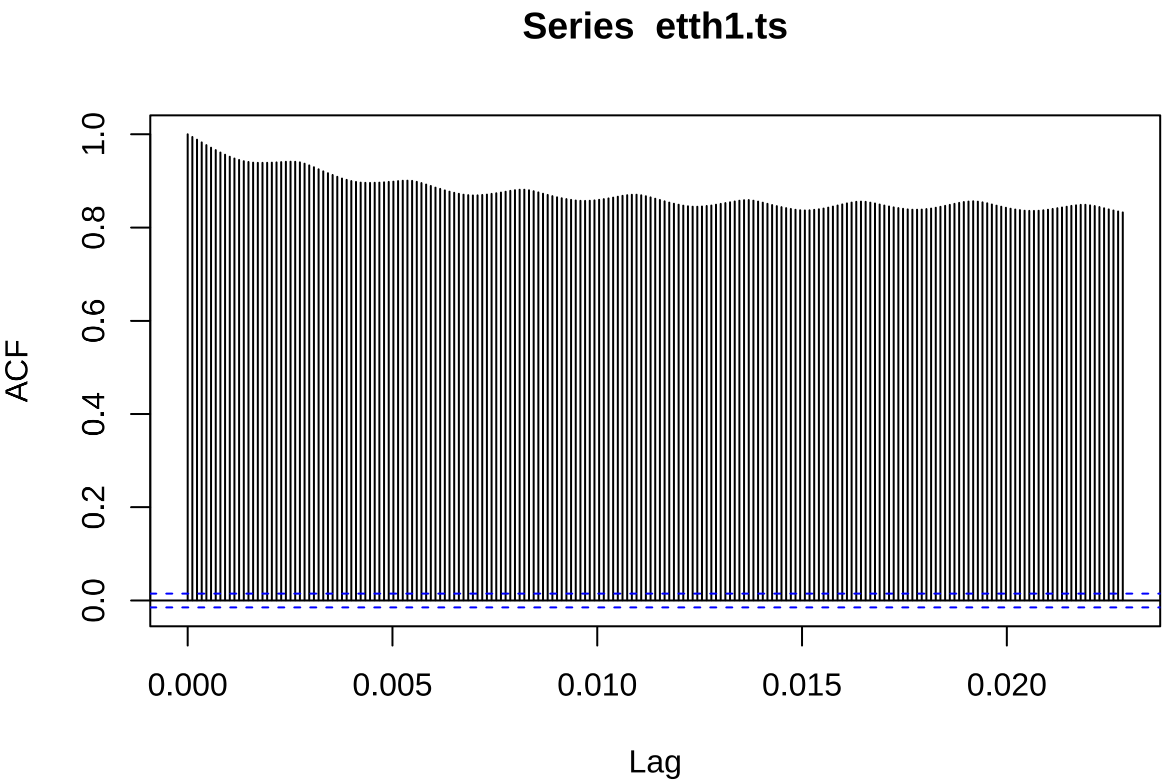
```
# first diagnose the autocorrelation of the existing time series
length(etth1.ts)
```

```
## [1] 17420
```

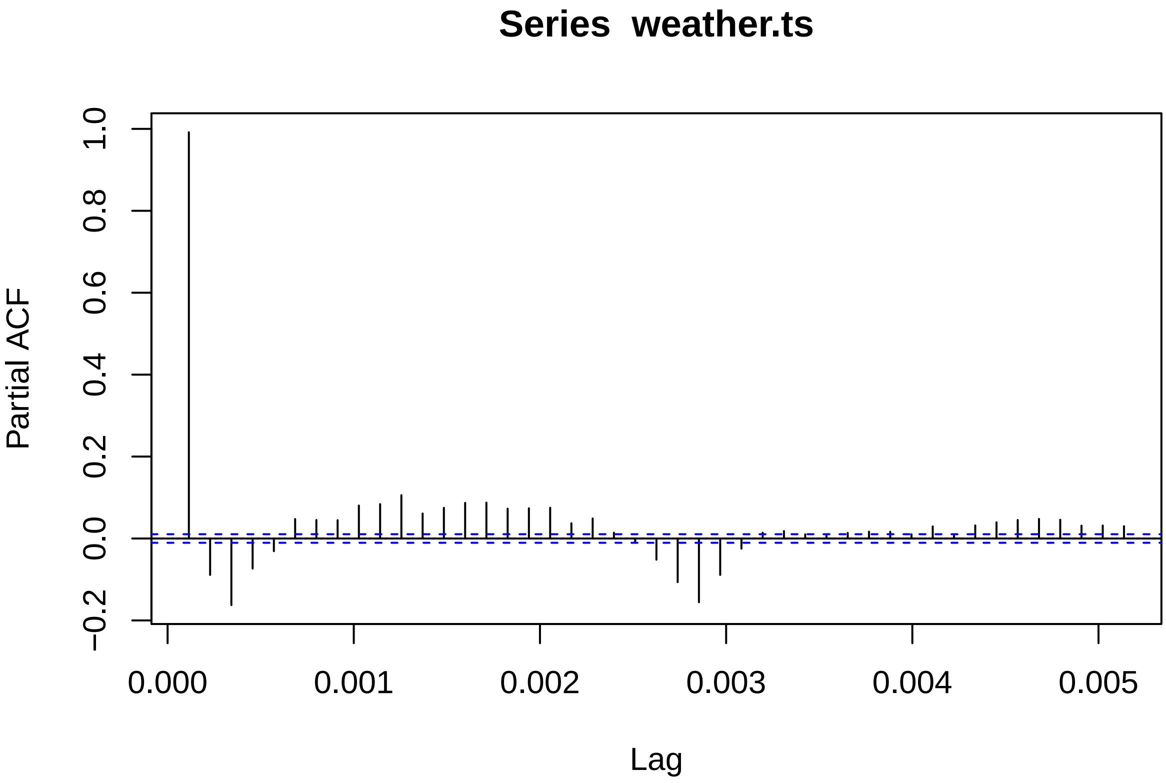
```
pacf(etth1.ts, lag.max=200)
```



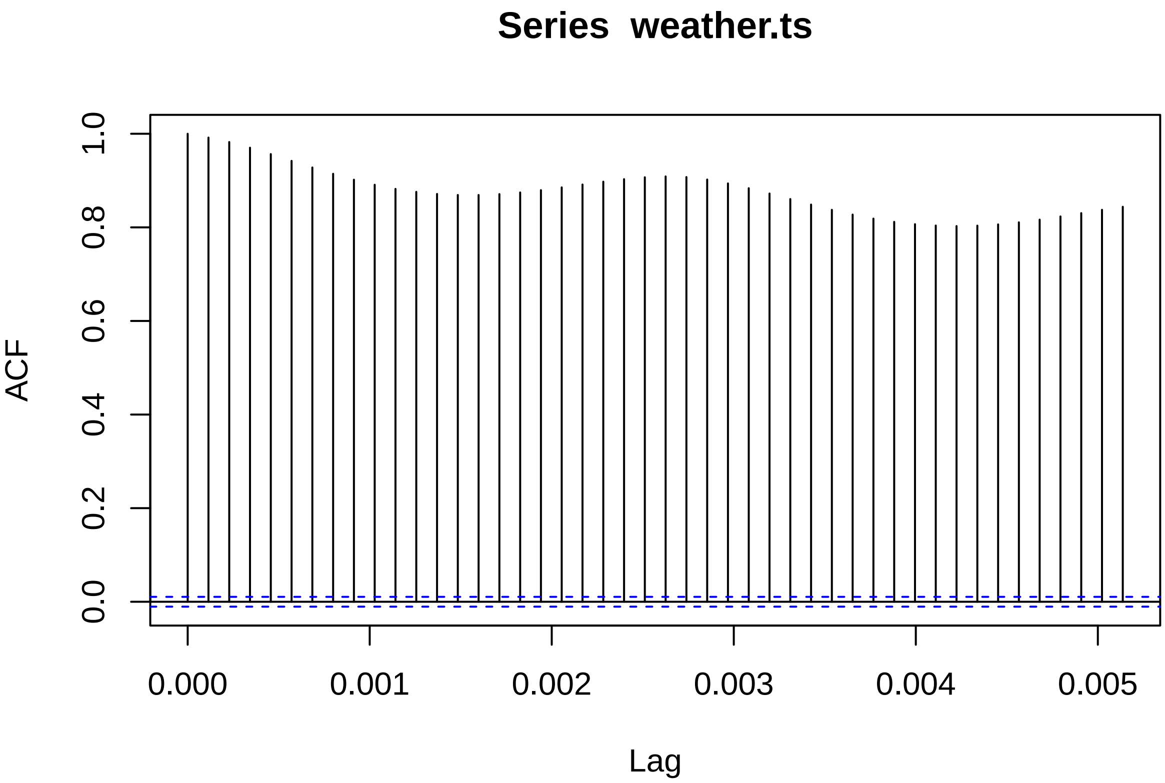
```
acf(etth1.ts, lag.max=200)
```



```
pacf(weather.ts)
```

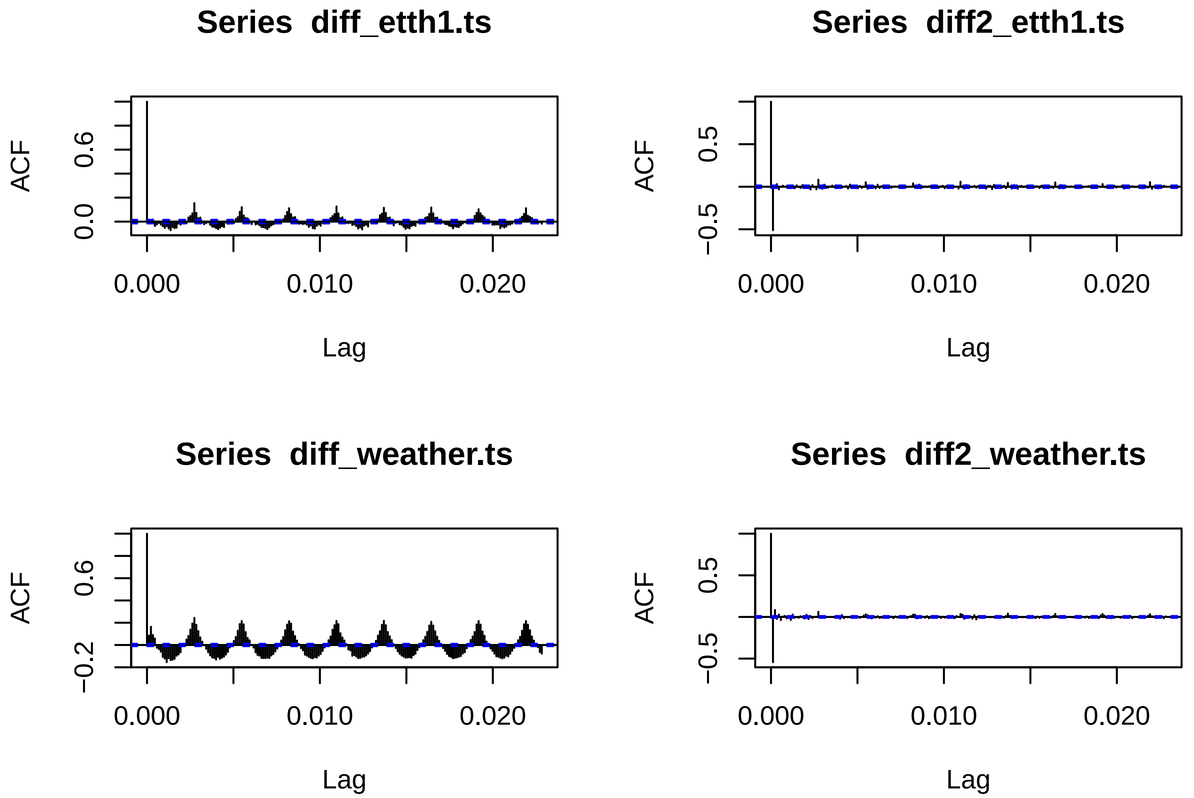


```
acf(weather.ts)
```

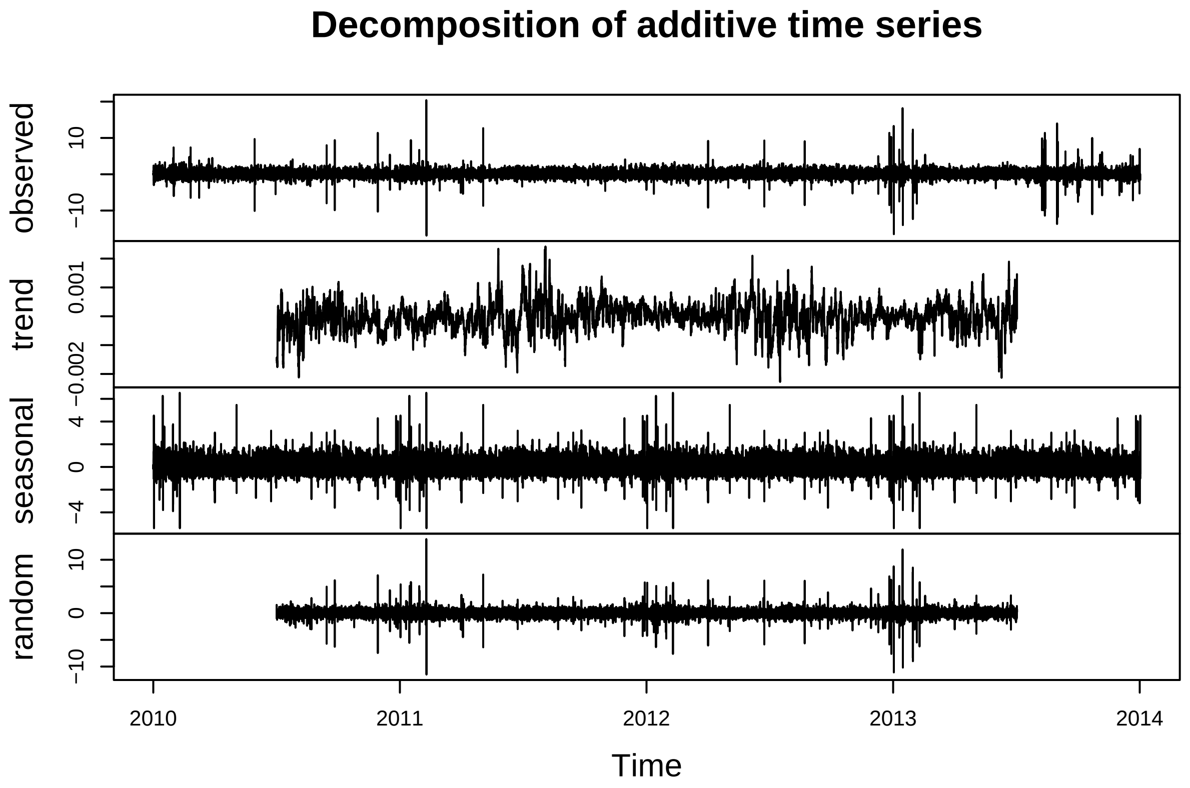


Differencing

```
par(mfrow=c(2,2))
# Electricity - trend elimination

diff_etth1.ts = diff(etth1.ts)
diff2_etth1.ts = diff(diff(etth1.ts))
acf(diff_etth1.ts, lag.max=200)
acf(diff2_etth1.ts,lag.max=200)


# Weather

diff_weather.ts = diff(weather.ts)
diff2_weather.ts = diff(diff(weather.ts))
acf(diff_weather.ts,lag.max=200)
acf(diff2_weather.ts,lag.max=200)
```



```
plot(decompose(diff_weather.ts))
```



```r
# Augmented Dickey-Fuller Test

# reject null that ts is non-stationary
adf.test(weather.ts)
```

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

```
##
##  Augmented Dickey-Fuller Test
##
## data:  weather.ts
## Dickey-Fuller = -8.1762, Lag order = 32, p-value = 0.01
## alternative hypothesis: stationary
```

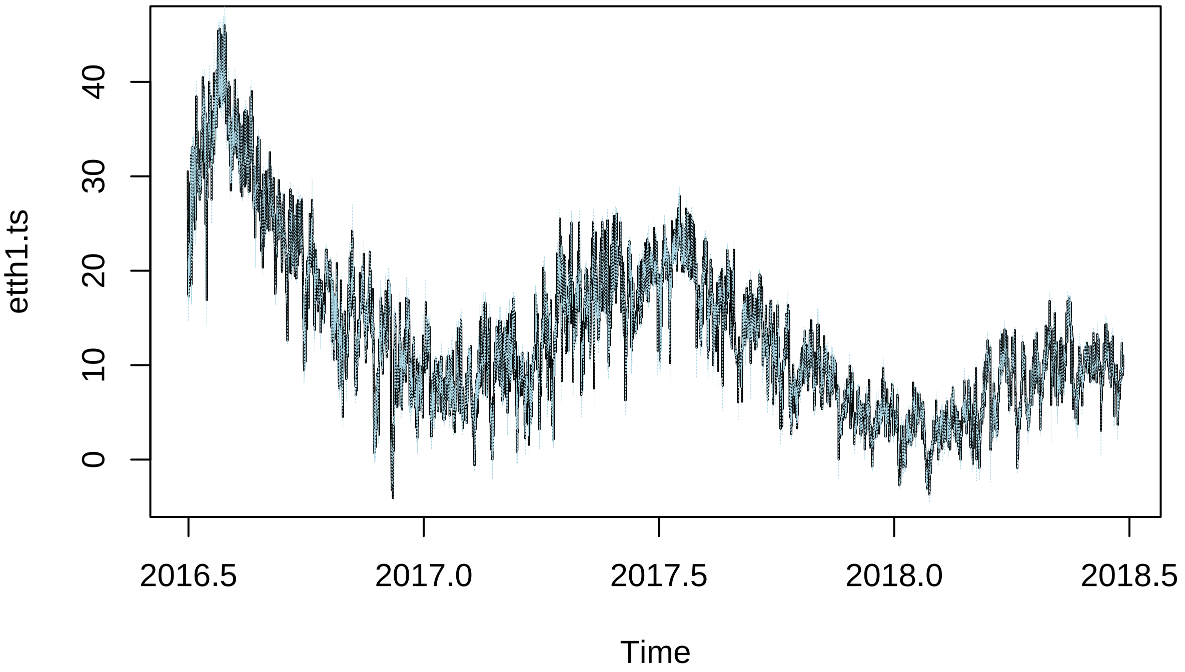
```r
adf.test(diff_weather.ts)
```

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

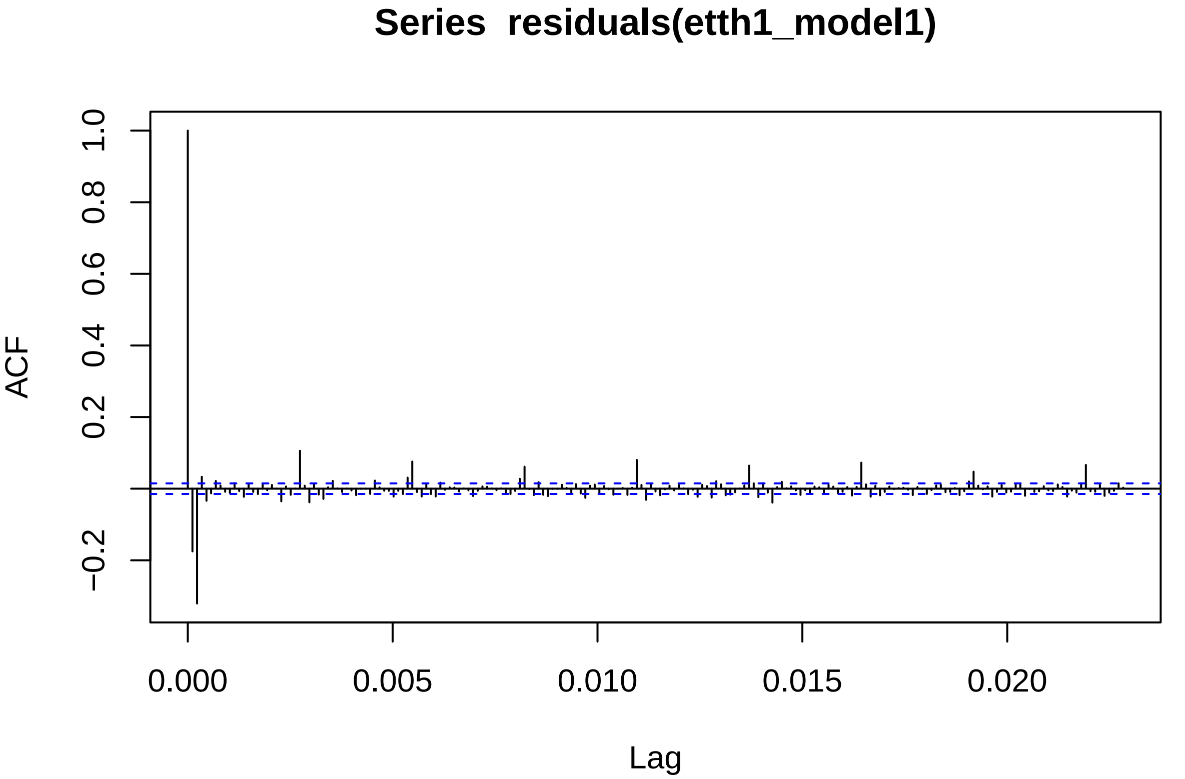
```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff_weather.ts
## Dickey-Fuller = -34.887, Lag order = 32, p-value = 0.01
## alternative hypothesis: stationary
```

Trying different models based on preliminary plots

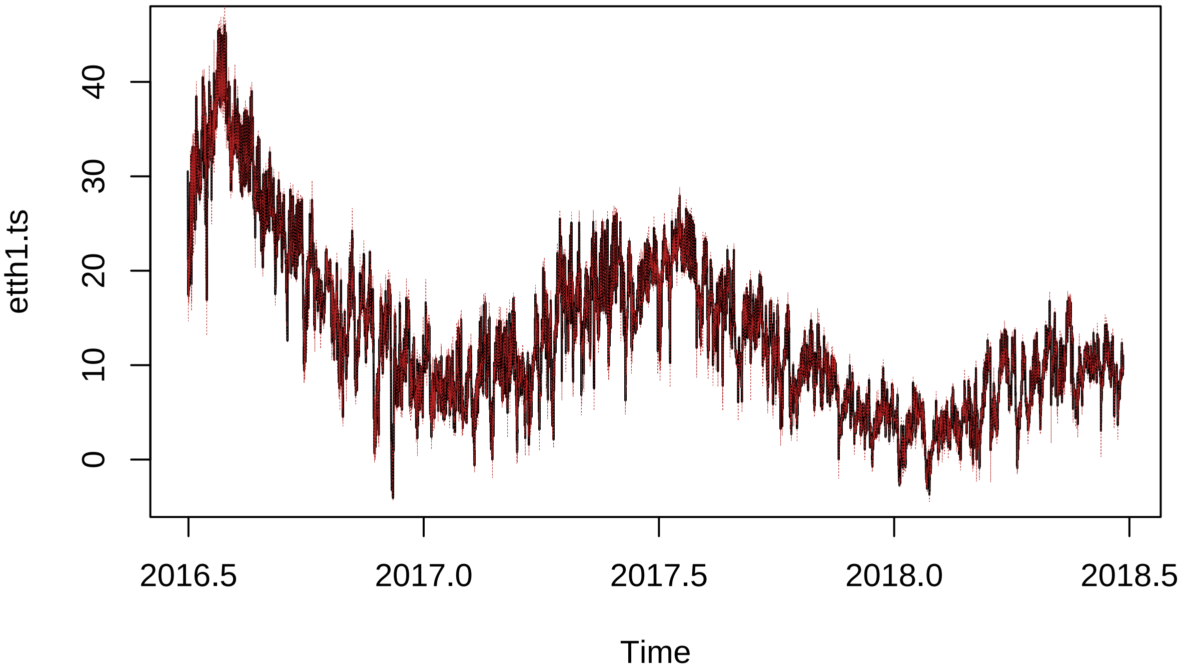
```r
# Ideal to difference twice to remove trend. Estimates using CSS-ML
etth1_model1 = arima(etth1.ts, order = c(1,2,0))
plot(etth1.ts)
fit1 = etth1.ts - residuals(etth1_model1)
points(fit1, type = "l", col = "lightblue", lty = 2, lwd=0.2)
```



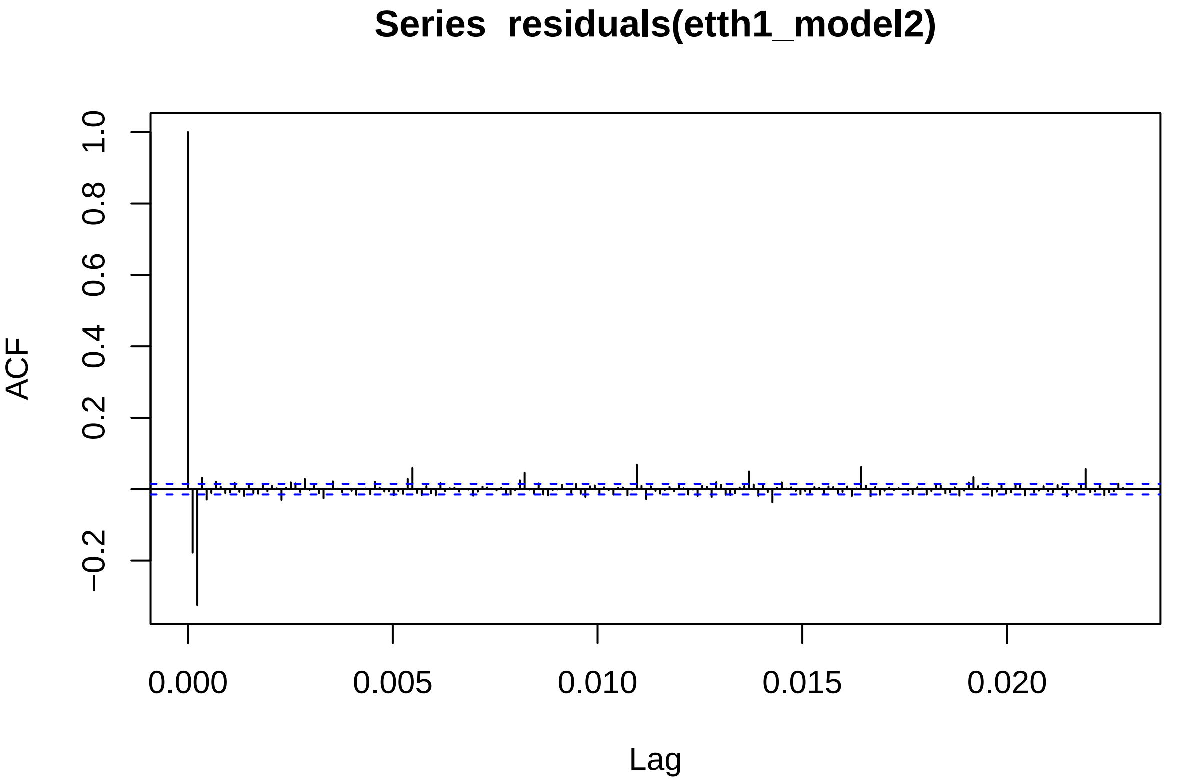
```
acf(residuals(etth1_model1), lag.max=200)
```



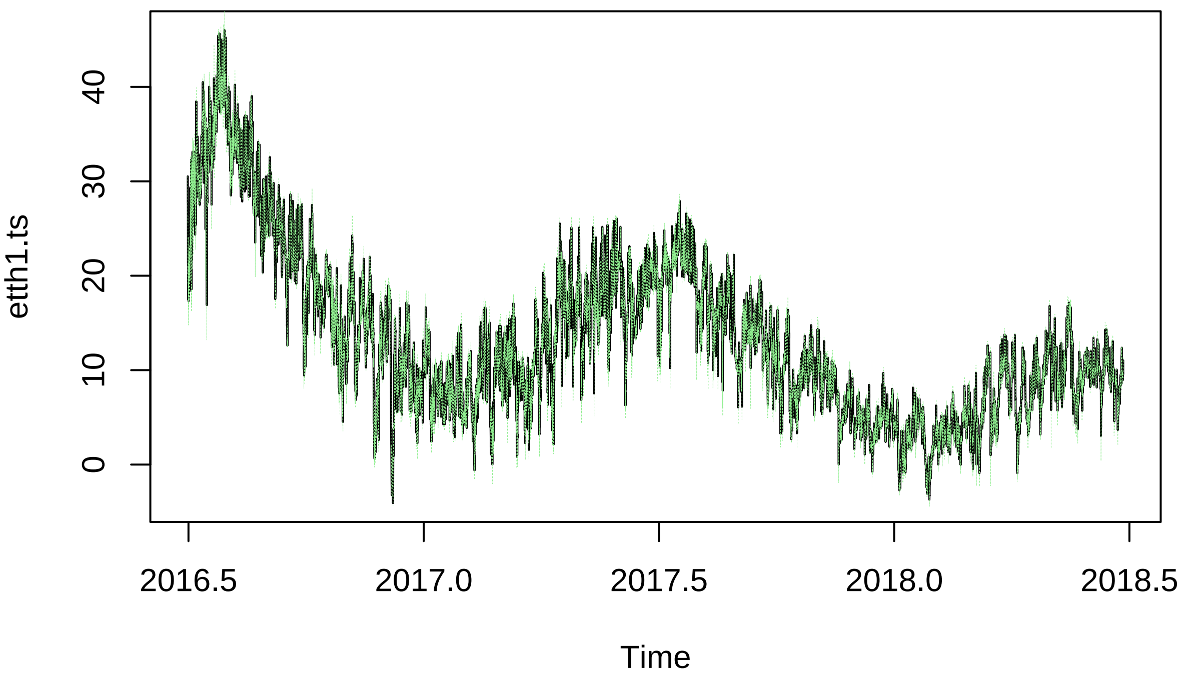
```
# seasonal component every 24 time steps (every 24 hours).
etth1_model2 = arima(etth1.ts, order = c(1,2,0), seasonal=list(order=c(1,0,0), period=24))
plot(etth1.ts)
fit2 = etth1.ts - residuals(etth1_model2)
points(fit2, type = "l", col = "firebrick", lty = 2, lwd=0.2)
```



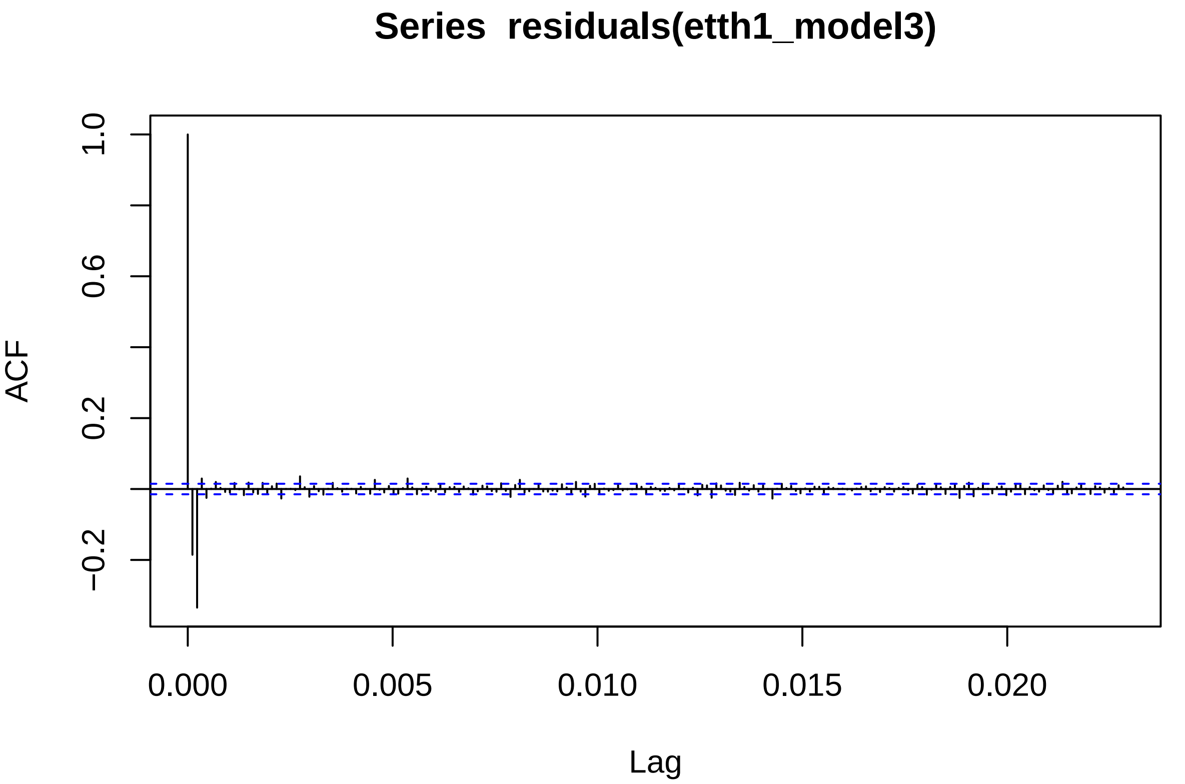
```
acf(residuals(etth1_model2), lag.max=200)
```



```
# MA component
etth1_model3 = arima(etth1.ts, order = c(1,2,0), seasonal=list(order=c(1,0,1), period=24))
plot(etth1.ts)
fit3 = etth1.ts - residuals(etth1_model3)
points(fit3, type = "l", col = "lightgreen", lty = 2, lwd=0.2)
```



```
acf(residuals(etth1_model3), lag.max=200)
```


Series residuals(etth1_model3)
ACF
Lag