Analyses of the high frequency time series on example of Circadian rhythms in the Long-Tail Pocket mouse.

Todd Orrender, Zhichun Ke , Jiaen Zhang, Inna Williams

Abstract

The present study aimed to examine the forecasting performance of various univariate approaches to forecast high-frequency time series which contains more than one seasonal pattern that must be taken into consideration.

To achieve the purpose of the study, we performed the following procedures with major findings. First, we investigated different frequencies and produced traditional forecasts for it using ts object. The limitation of ts objects for only low frequency data and also for using multiple seasonalities led us to make additional sets of forecasts by applying non-benchmark approaches with double seasonality using msts object with Fourier terms for each seasonal pattern. This indicates that forecast accuracy was able to be improved by incorporating two different types of seasonal patterns simultaneously.

Seasonality is the main component of time series, and the consideration of seasonality has become more important with the increasing frequency of time series produced in industry. Circadian rhythms in the Long-Tail Pocket mouse time series is a representative example of business time series data that has been collected every two minutes for a long period of time. The high frequency data, with time series containing closely spaced time intervals creates a necessity of creating accurate modeling approaches for modeling seasonal patterns of time series.

As each seasonal pattern has distinct periods and effects, it is not trivial to design a model that can capture multiple seasonal patterns at once. We found that fourier regression gives more accurate results and can detect multiple seasonalities better then etc and tbats and sarima.

Introduction

The investigating dataset is a temperature recording made at a 2 minutes interval for 83 days on a nocturnal mammal. The Long-tailed pocket mouse live in the South West of the United States and Northern Mexico. The data represents interest because it does contain periodicities in the behaviour of the animals in regards to circadian rhythm. The Circadian is the reaction to light and darkness of their environment. The experiments used separate equipment for each mammal to monitor the environment and so data for each animal could be varied. The data given here are the telemeter frequency temperature recordings for one Long-tailed pocket mouse from the first experiment, with higher counts indicating higher temperature. Since the interest was in periodicity, no effort was made to relate the telemeter frequency to actual temperature.The file that was collected had 59616 observations that correspond to the 83 days. The proportion of outliers in the original data has been 8% and outliers were removed.

Typically, multiple seasonality patterns are more likely to occur whenever the series has data with high frequency (for example, daily, hourly, half-hourly, and so on), as there are more options to aggregate the series to a lower frequency. This data is a typical example of multiple seasonality , which could have multiple seasonal patterns, as the observations taken every two minutes of the hours of the day, the day of the week. On the other hand, as the frequency of the series is lower (for example, monthly, quarterly, and so on), it is more likely to have only one dominant seasonal pattern as opposed to a high frequency series, as there are fewer aggregation options for another type of frequencies.
Seven last days of data for investigation were selected.

Model Specification

7days of observations = 30(in one hour) * 24(hours) * 7(days) = 5040 observations

Having more observations per cycle unit, that is, high-frequency time series data, could potentially provide more insightful information about the series behavior as opposed
to a lower frequency time series data. However, this comes with the price of additional complexity, which therefore requires more effort in the analysis process.
Potentially, as mentioned previously, the series can have three different seasonal patterns.

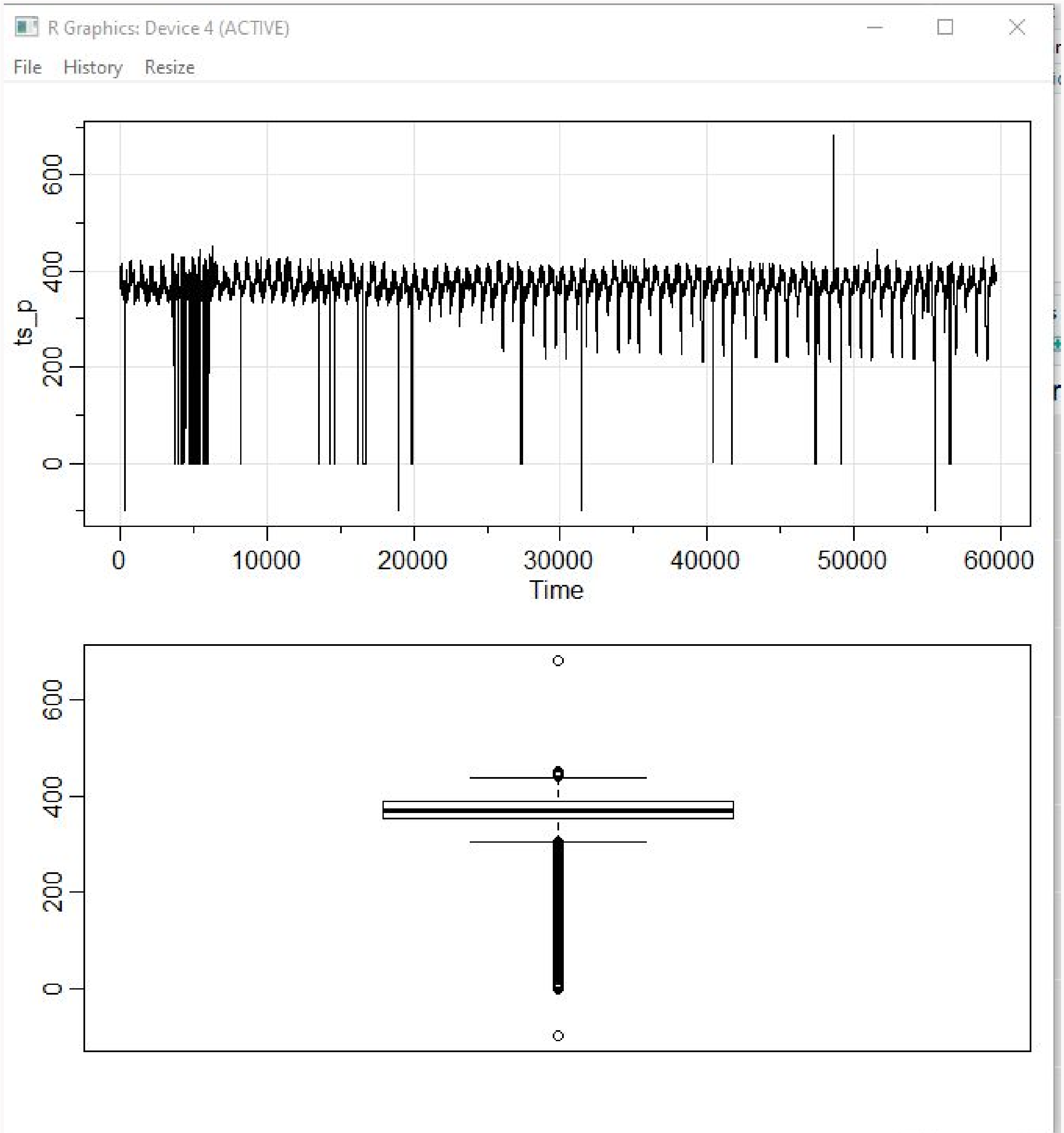
1 week = 30*24 *7 =5040
Daily 30 * 24 = 720
Hourly = 30
2 hourly = 30*2=60
6 hourly = 30*6 = 180
10 hourly = 30 *10 = 330

Time series objects allow maximum frequency =350 and therefore we will use hourly seasonality with frequency = 30 and 10 hourly with frequency = 300. For multiple seasonalities we have to use ms object instead of ts object.

Original data

```
library(astsa)
library("forecast")
library("lubridate")
examine.mod <- function(mod.fit.obj, p, d, q, P=0, D=0, Q=0, S=-1, lag.max=24) {
 dev.new(width=6, height=6)
 par(mfrow=c(2,1))
 pacf(mod.fit.obj$fit$residuals, main="PACF of Residuals", lag.max)
 if ((P==0)&(D==0)&(Q==0)) {
  title(paste("Model: (", p, ",", d, ",", q, ")", sep=""), adj=0, cex.main=0.75)
 }
 else {
  title(paste("Model: (", p, ",", d, ",", q, ") (", P, ",", D, ",", Q, ") [", S, "]", sep=""), adj=0, cex.main=0.75)
 }

 std.resid <- mod.fit.obj$fit$residuals/sqrt(mod.fit.obj$fit$sigma2)
 hist(std.resid, main="Histogram of Standardized Residuals", xlab="Standardized Residuals",
freq=FALSE)
 curve(expr=dnorm(x, mean=mean(std.resid), sd=sd(std.resid)), col="red", add=TRUE)
}
```

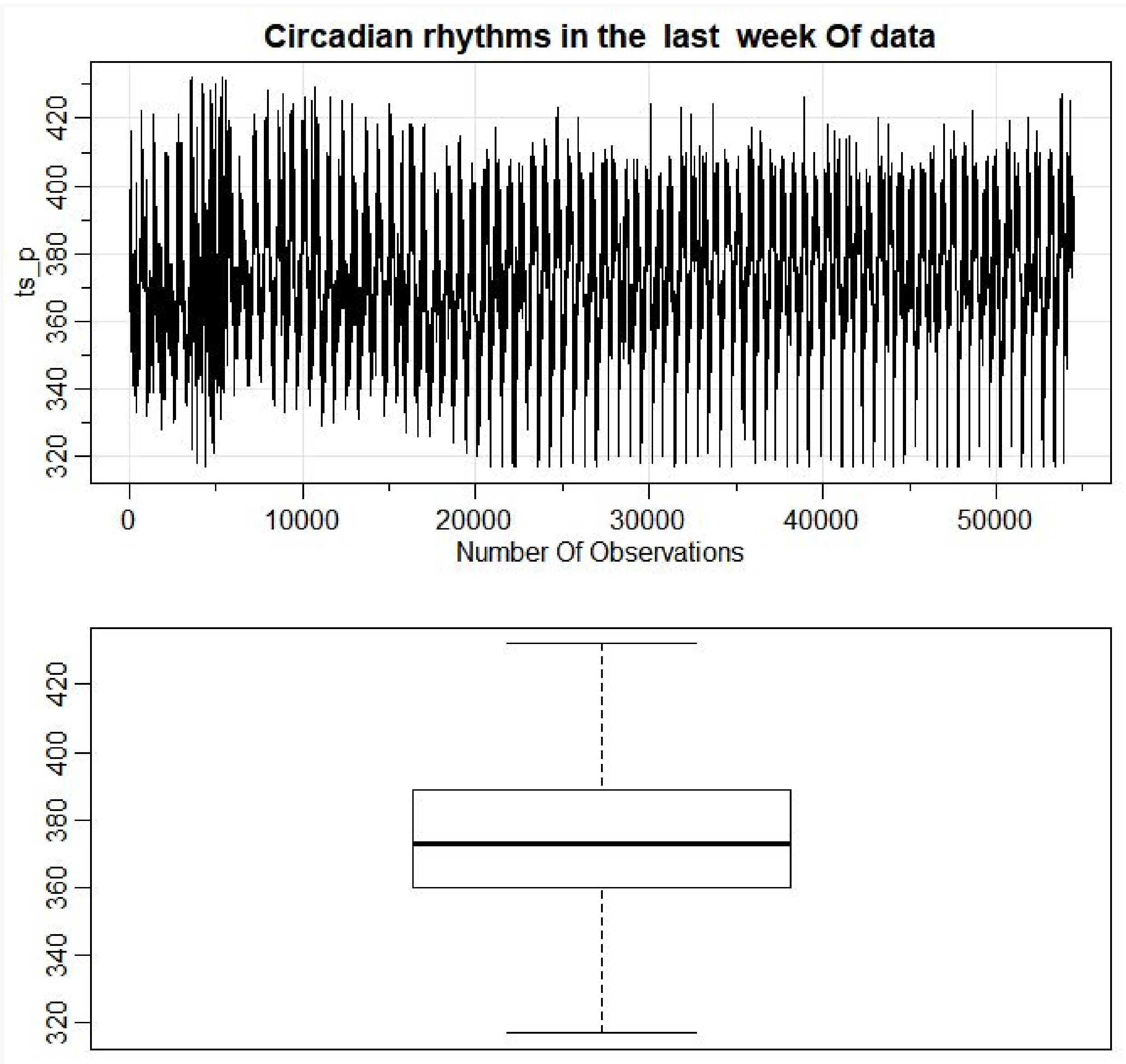
```
pformosu<-read.table(file =
"C:/Users/inna/Desktop/DepaulClasses/ApplyMathClasses/Time_Series/Final Project/pformosu.txt")

row =3726
col= 16
pformosu_dat = numeric(col*row)
count = 1
for(i in 1:row)
{
 for(j in 1:col)
 {
   pformosu_dat[count]=pformosu[i,j]
   count = count + 1
 }
}
ts_p<-ts(pformosu_dat)
ength(ts_p)
dev.new()
par(mfrow=c(2,1))
tsplot(ts_p)
boxplot(ts_p)$out
```



Removing outliers.

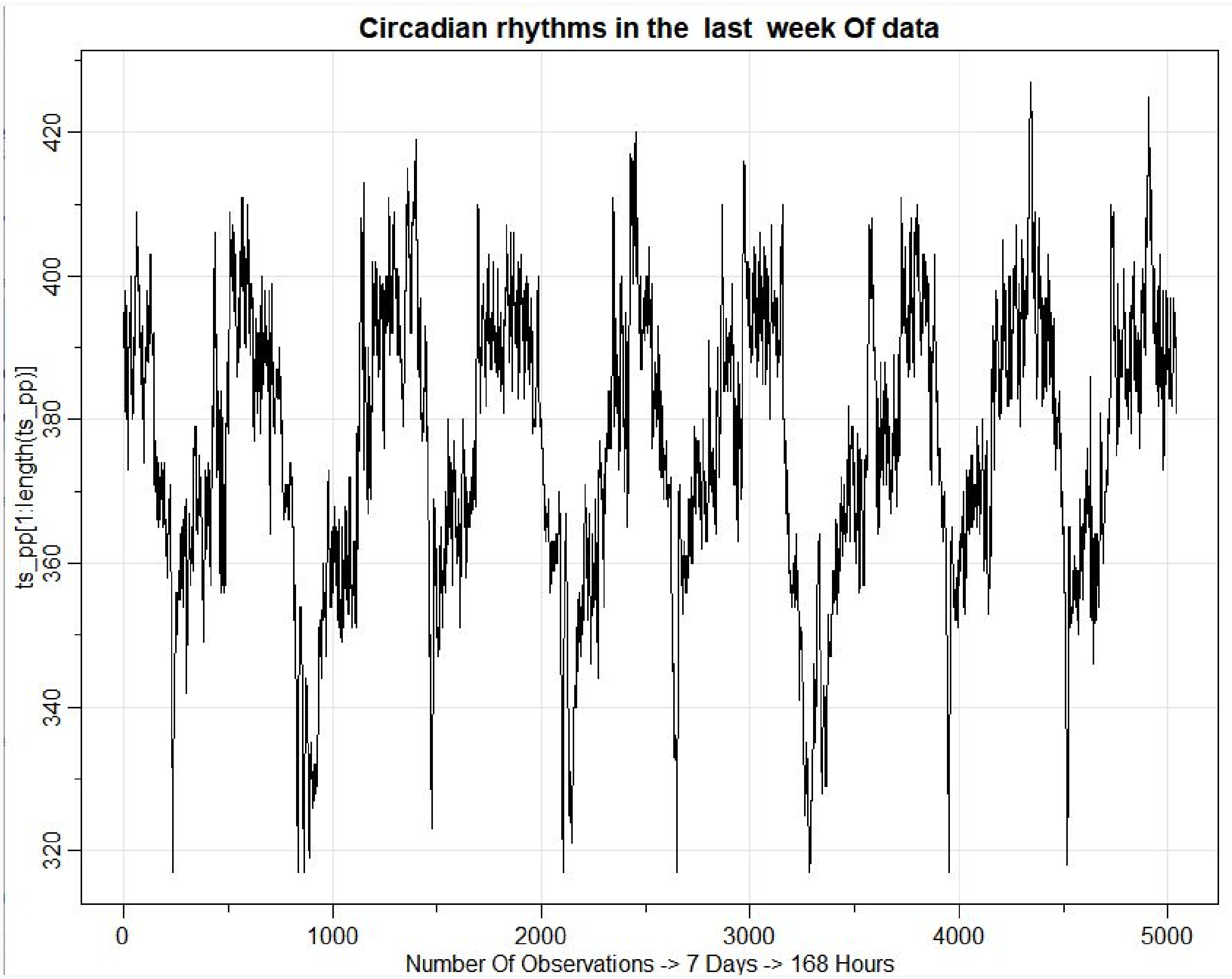
```
num_of_outliers = 0
repeat
{
 outliers <- boxplot(ts_p, plot=FALSE)$out
 v<-(which(ts_p %in% outliers))
 if(length(v)== 0)
  break
```

```
 else
 {
  ts_p<-ts_p[-v]
  num_of_outliers = num_of_outliers + length(v)
 }
}
print(paste("Number Of Outliers removed = ", num_of_outliers))
print(paste("Length Of Time Series",length(ts_p)))
dev.new()
par(mfrow=c(2,1))
tsplot(ts_p,xlab="Number Of Observations",main="Circadian rhythms in the  last  week Of data")
boxplot(ts_p)$out
```

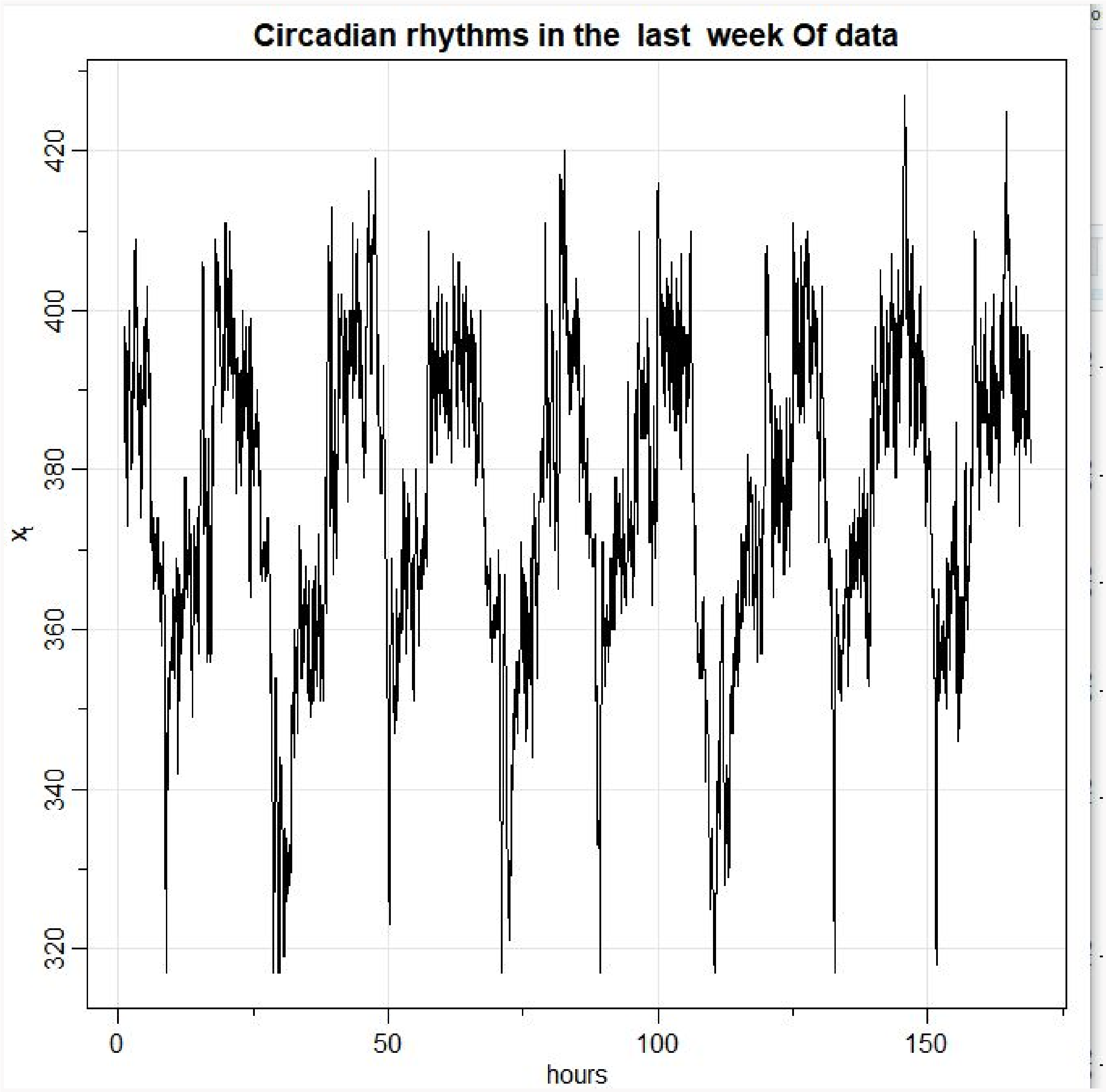


Getting one week of data and plotting it.

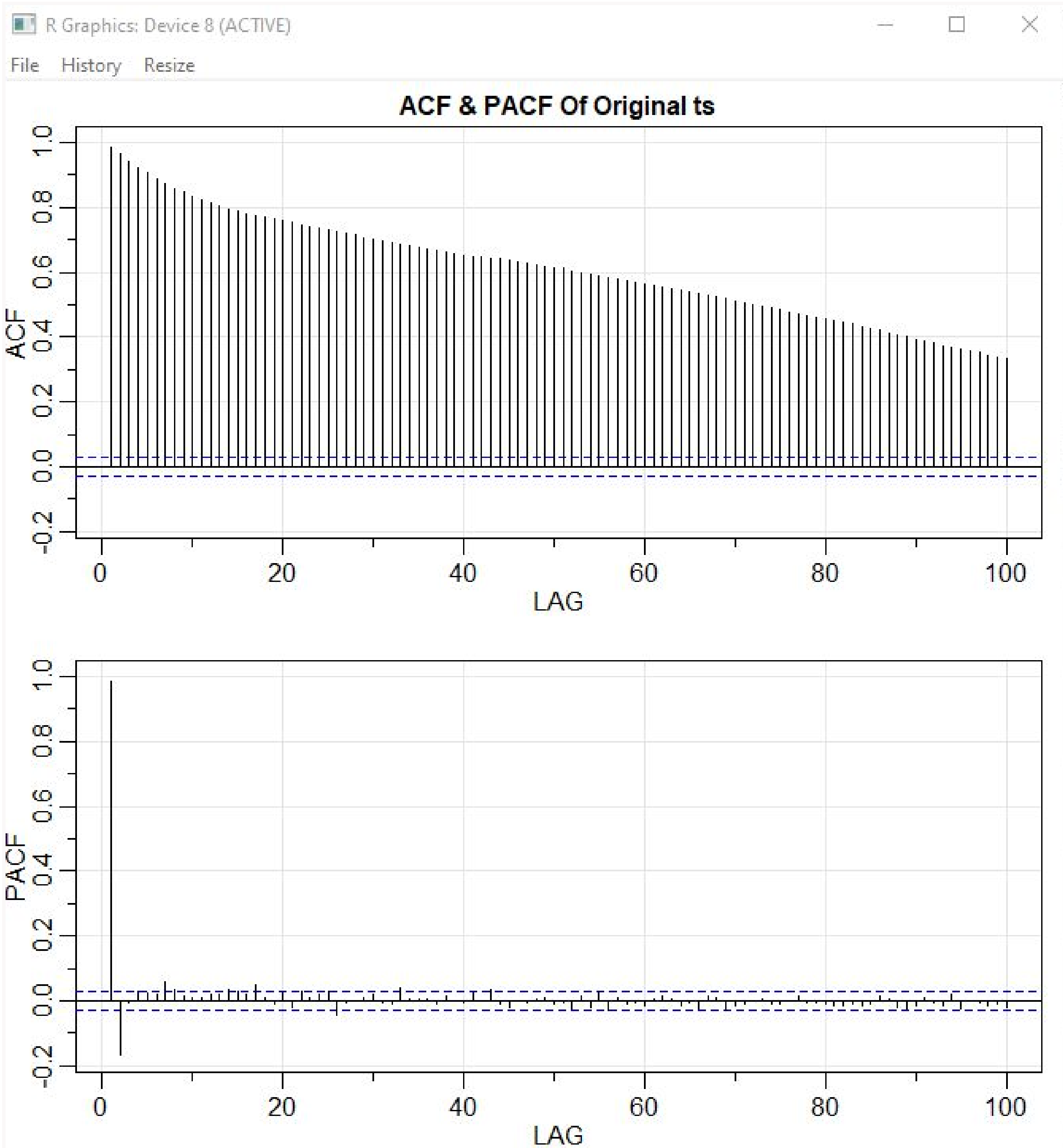
```
frequency <- 30
one_day<- 30*24
one_week<- 24*30*7
ts_pp  <- ts(ts_p[(length(ts_p)-(one_week) + 1 ):(length(ts_p))],frequency = frequency)
dev.new()
tsplot(ts_pp[1:length(ts_pp)],xlab="Number Of Observations -> 7 Days -> 168 Hours",
    main="Circadian rhythms in the  last  week Of data")
x<-as.numeric(ts_pp)
```



```
dev.new()
tsplot(ts_pp,  ylab=expression(x[t]),
     xlab="hours", main="Circadian rhythms in the  last  week Of data")
```

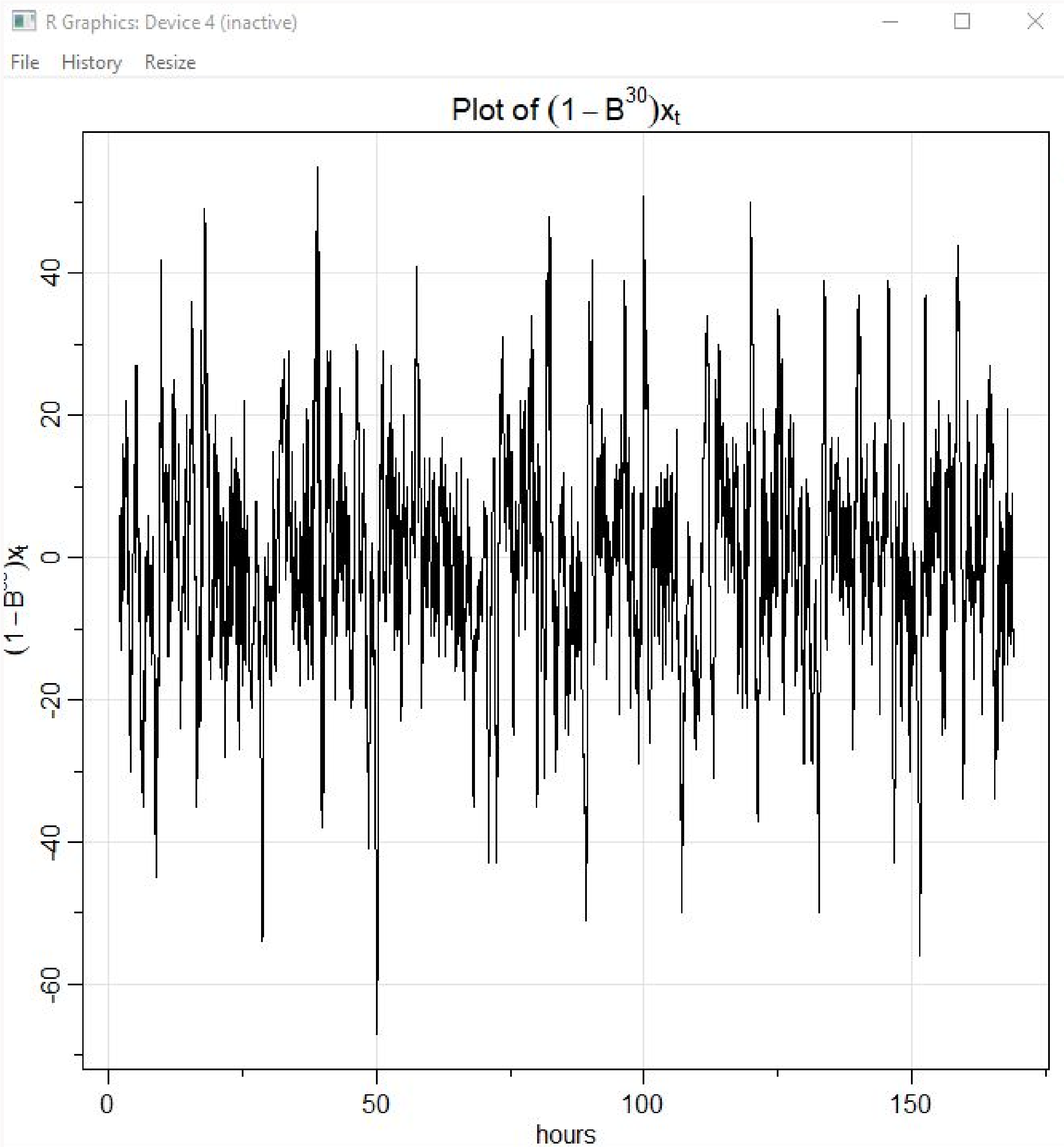


dev.new()
acf2(x, max.lag = length(ts_pp) - 1,main = "ACF & PACF Of Original ts")



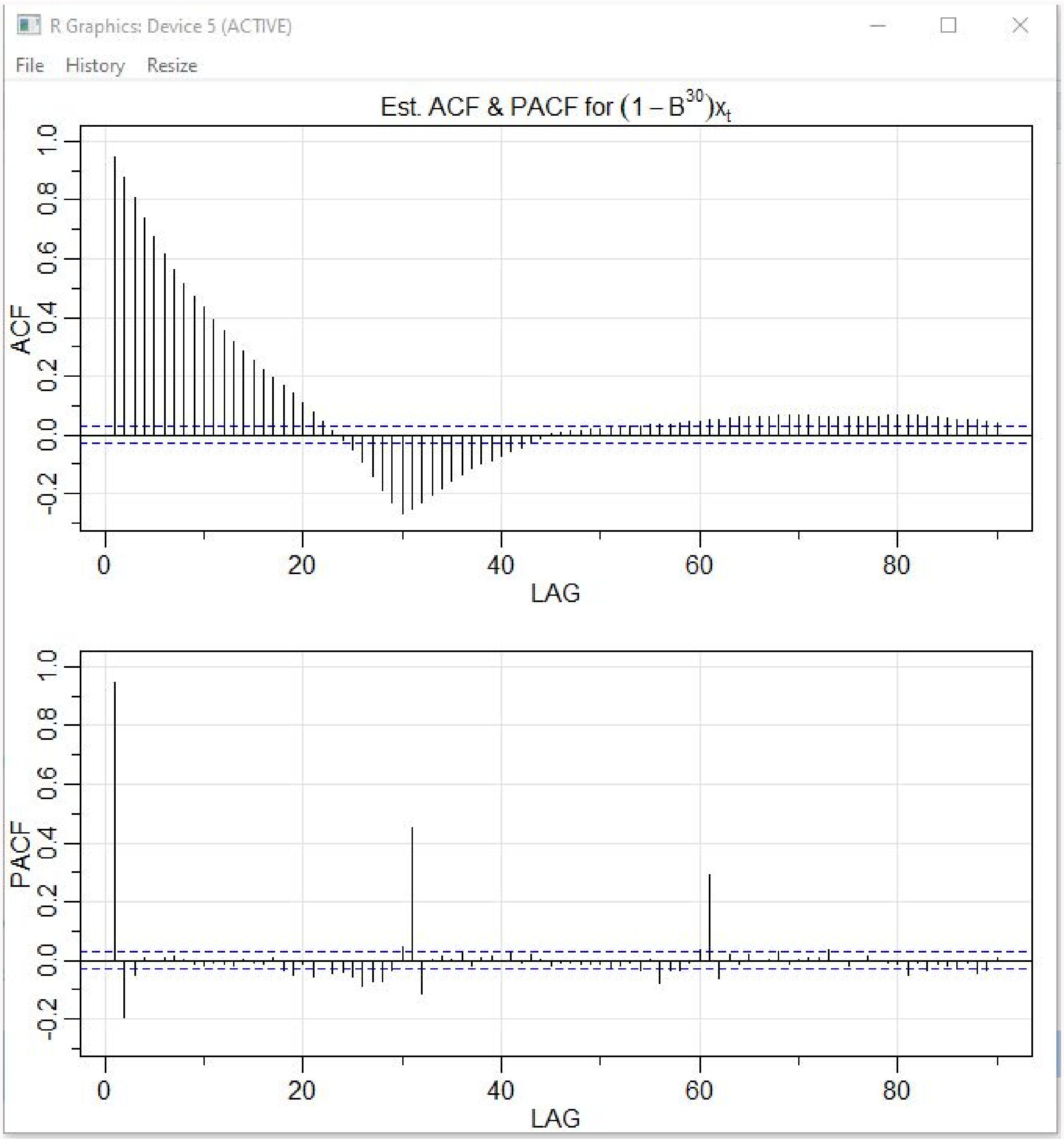
We can see from the plot that ACF is decaying exponentially and PACF has 2 lags but we have to differentiate to get rid of seasonalities (1- B^30)

Plot of (1-B^30)*x_t
```
dev.new()
tsplot(diff(ts_pp, lag=frequency, differences=1), ylab=expression((1-B^30)*x[t]),
    xlab="hours", main=expression(paste("Plot of ", (1-B^30)*x[t])))
```



ACF indicated cut of after 30 lag that suggests that ARIMA Q=1,
PACF exponentially decay that suggests ARIMA P=0
Diff =1 S=18 (1-B^30)*x_t

```
dev.new()
acf2(diff(x, lag=frequency, differences=1), max.lag=frequency*3,
   main=expression(paste("Est. ACF & PACF for ", (1-B^30)*x[t])))
```

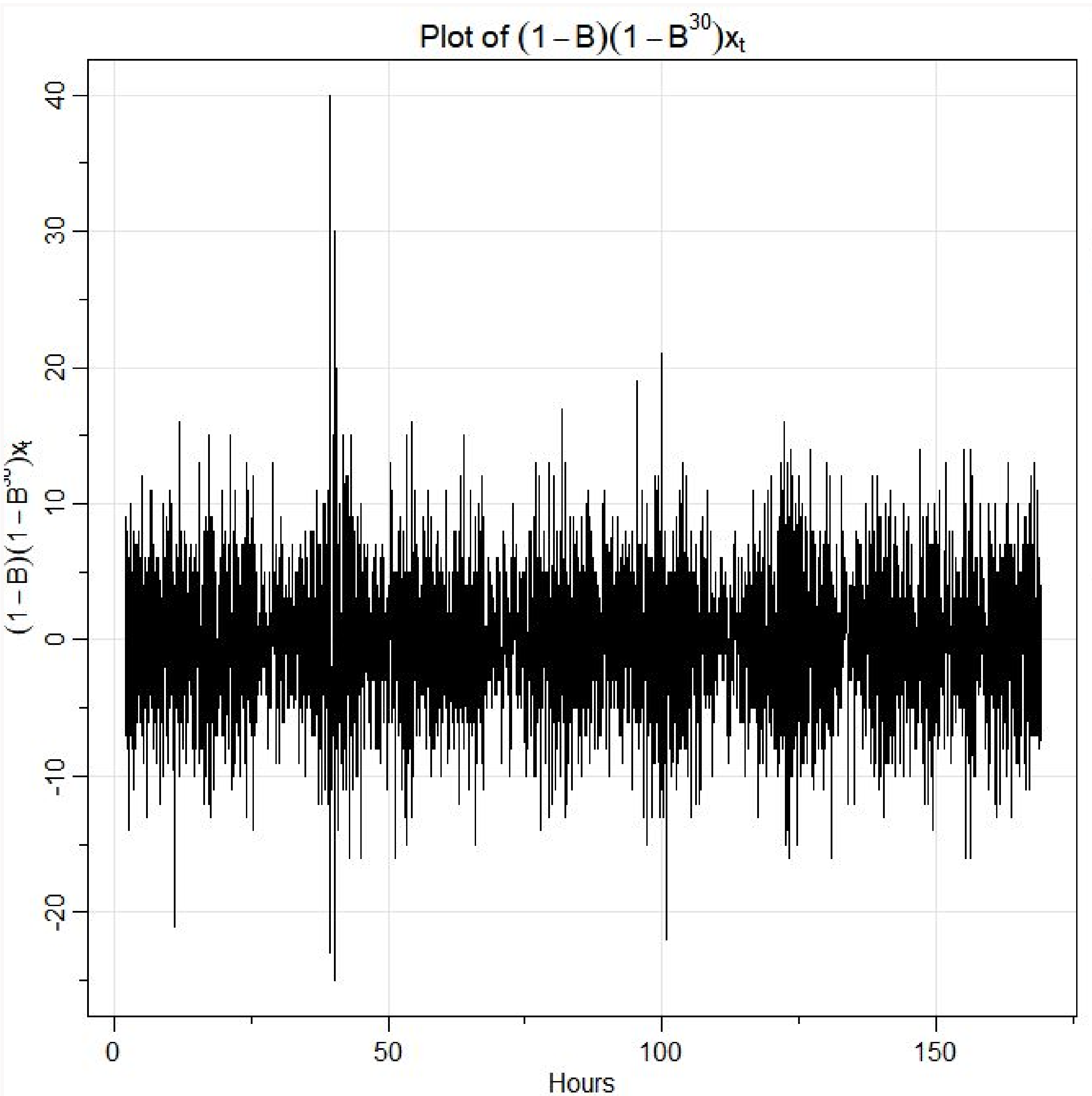


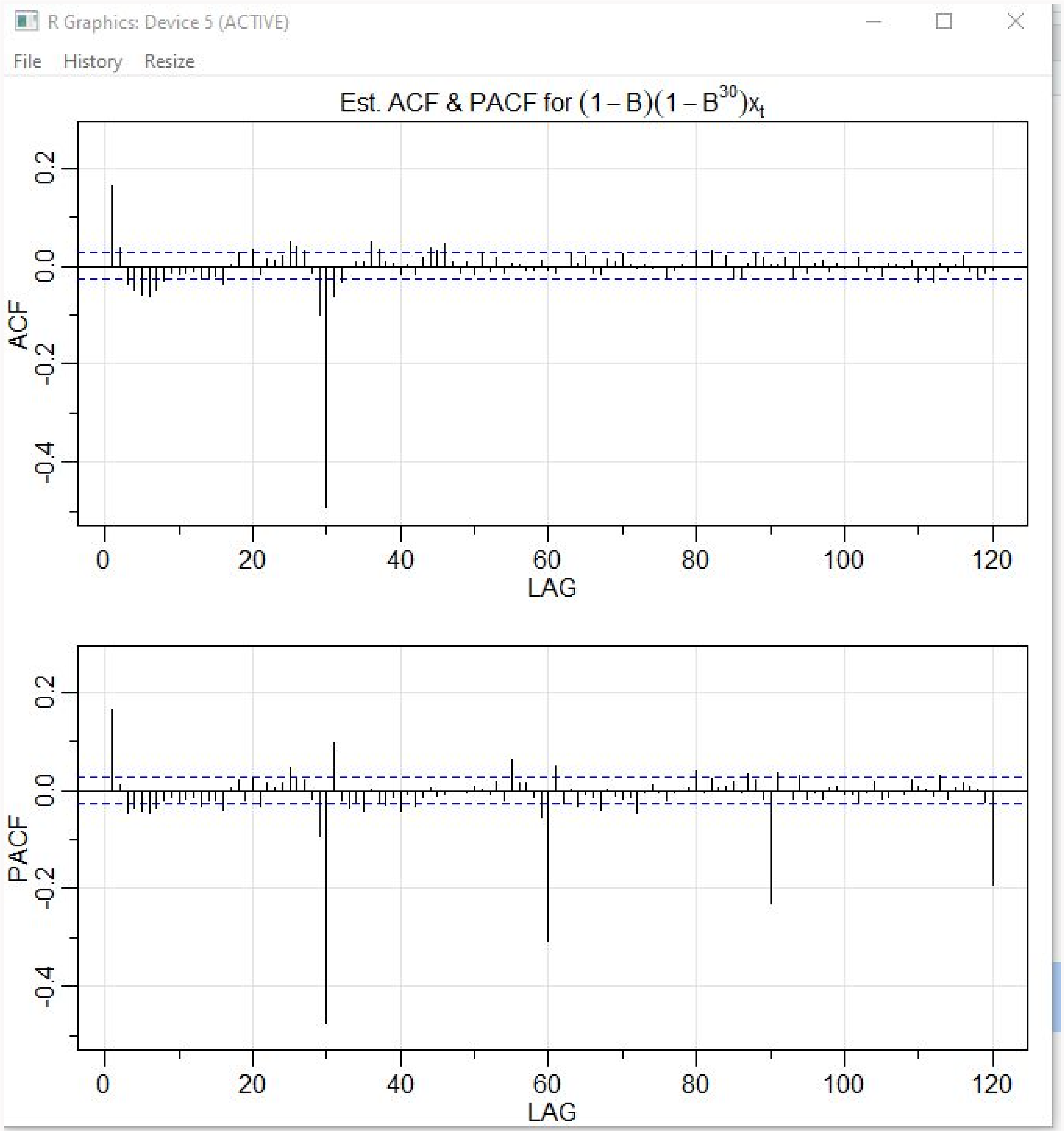
-

PACF exponentially decaying at lags 30,60 indicating that P = 0

ACF seasonal pick at Lag = 30 and no other seasonal picks at lags 60 , 90
Indication that Q = 1, giving us the initial estimation for ARIMA(),(0,1,1)30.
But graph has some little trend so we have to remove the trend using non seasonal difference (1-B)

```
ev.new()
tsplot(diff(diff(ts_pp, lag=frequency, differences=1)),
    ylab=expression((1-B)(1-B^30)*x[t]),
    xlab="Hours", main=expression(paste("Plot of ",
    (1-B)(1-B^30)*x[t])))
```



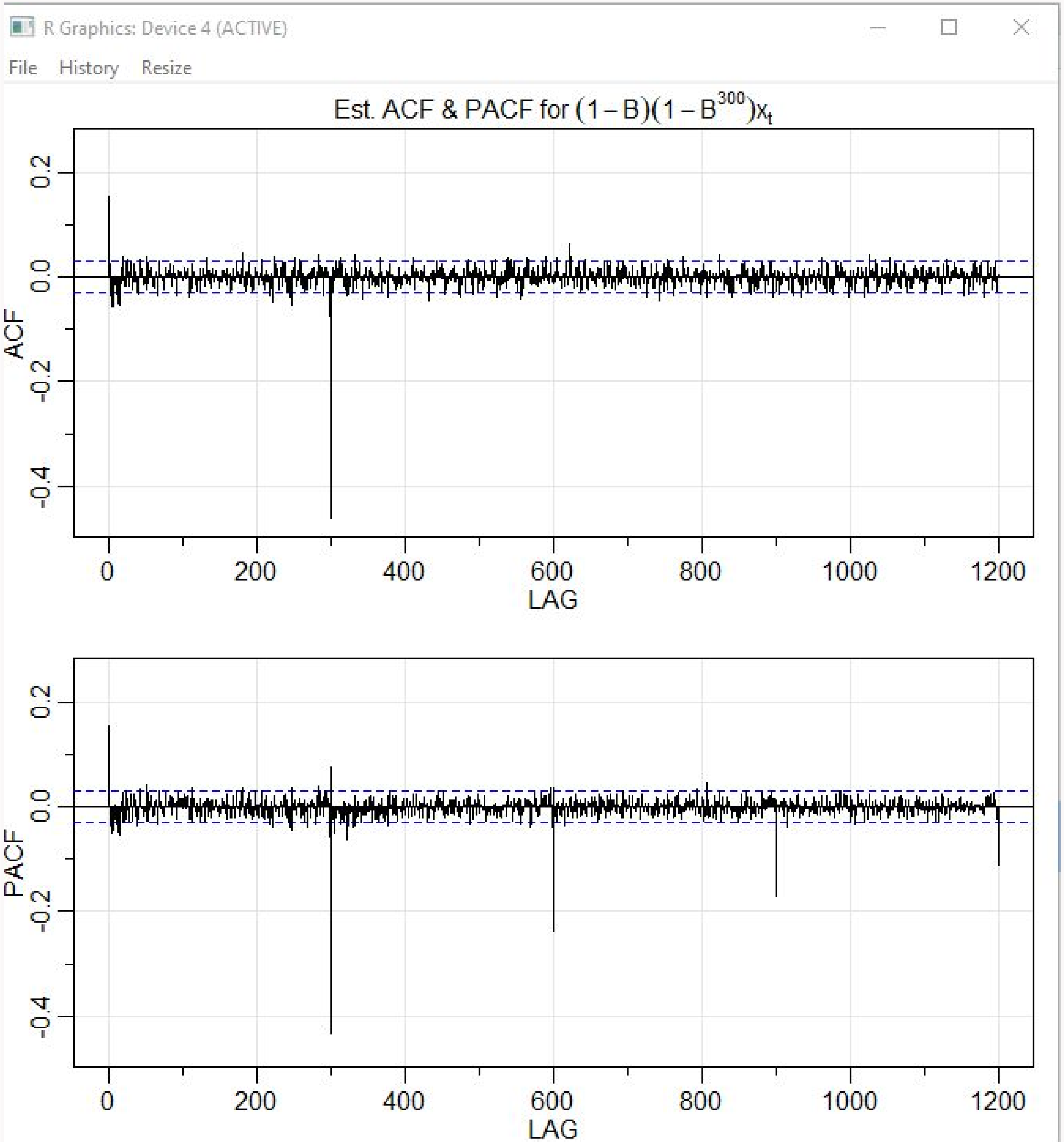


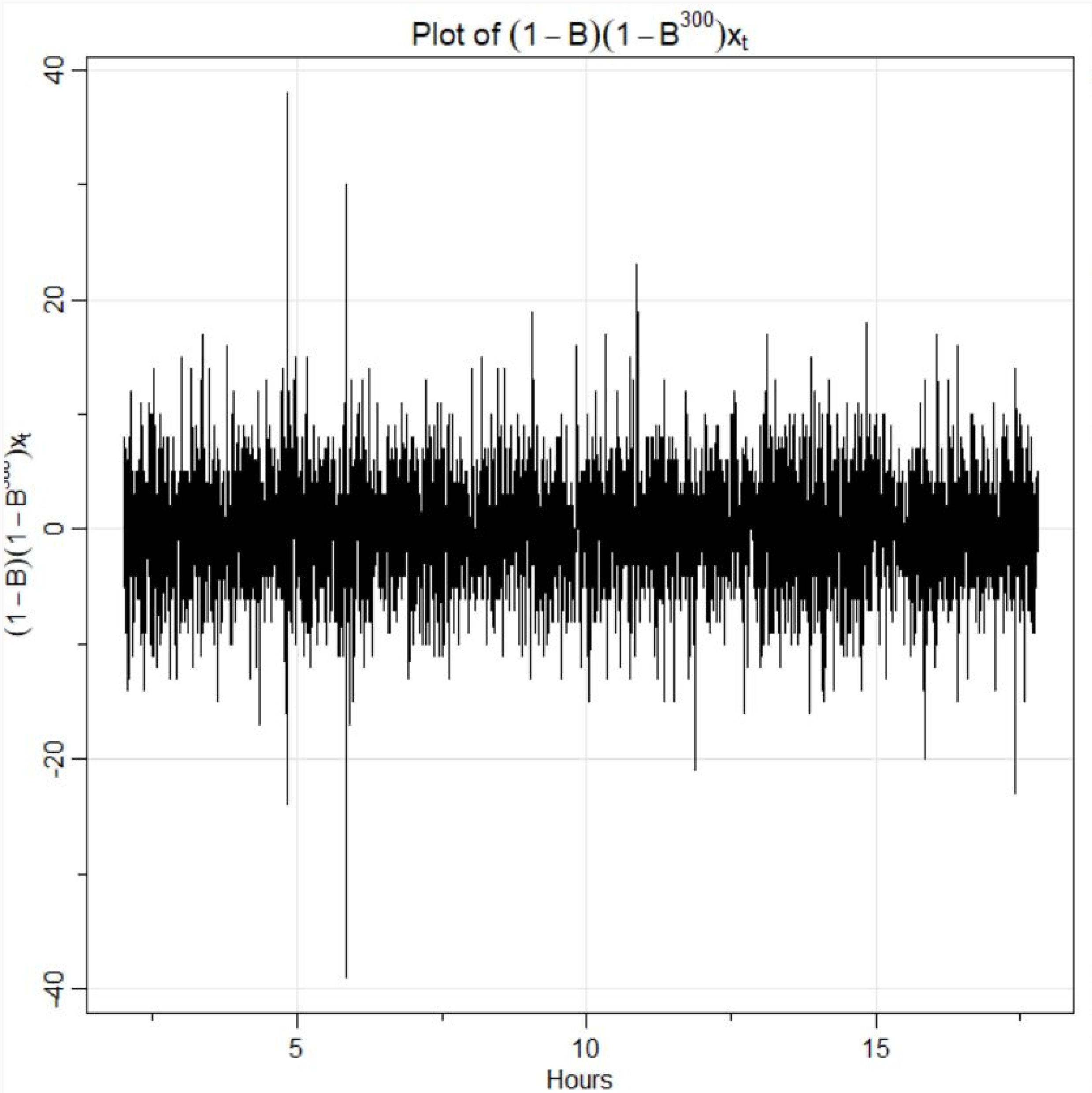
ACF seasonal Q = 1, non-seasonal q = 1-5

PACF decaying at lags 30,60,90 , and clusters around each lag indication that seasonal P = 0, nonseasonal p = 1 -5

ARIMA(p=(0-5),d=1,q=(0-5)),(P=0,D=1,Q=1)[30]

Trying ACF frequency = 300


R Graphics: Device 4 (ACTIVE)
File History Resize
Est. ACF & PACF for $(1-B)(1-B^{300})x_t$
ACF
0.2
0.0
-0.2
-0.4
0
200
400
600
800
1000
1200
LAG
PACF
0.2
0.0
-0.2
-0.4
0
200
400
600
800
1000
1200
LAG

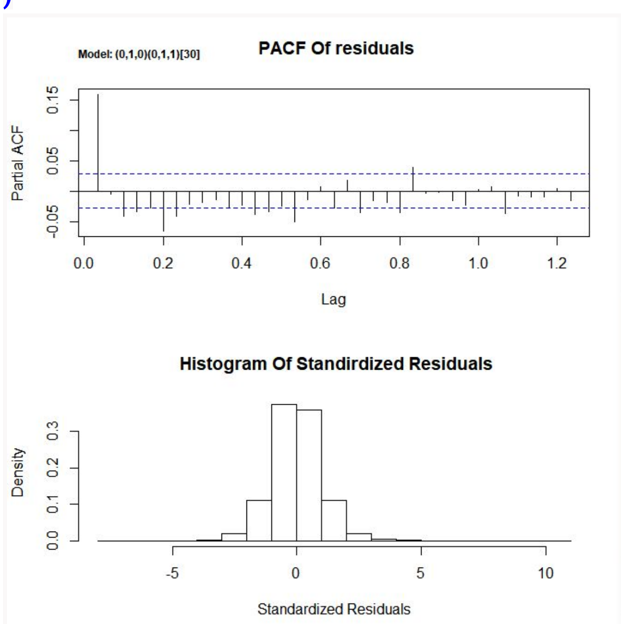


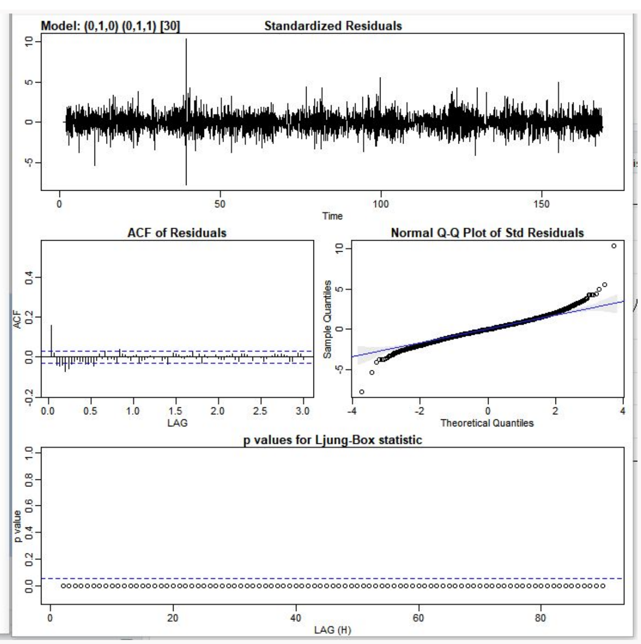
ACF shows seasonal spike at lag 300 and non-seasonal at some cluster around leg 1
PACF shows seasonal exponential decay at 300,600 .. and also cluster around lag one
Also indicating **ARIMA(p=(0-5),d=1,q=(0-5)),(P=0,D=1,Q=1)[300]**
ARIMA(p=(0-5),d=1,q=(0-5),P=0,D=1,Q=1,S=30)

Fitting and Diagnostics

Trying ARIMA(p=(0),d=1,q=(0),P=0,D=1,Q=1,S=30)

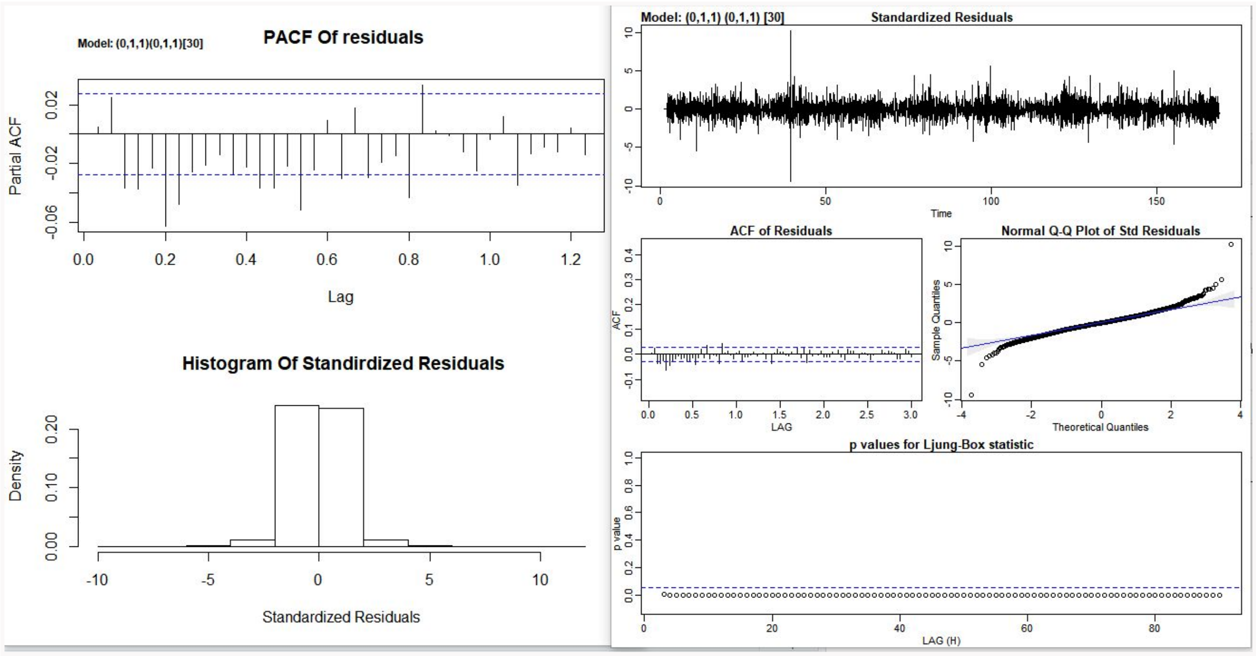
```
dev.new()
mod.fit1<- sarima(ts_pp,p=0,d=1,q=0,P=0,D=1,Q=1,S=30)
mod.fit1

dev.new()
par(mfrow=c(2,1))
pacf(mod.fit1$fit$residuals,main = " PACF Of residuals")
title(paste("Model: (",
        p = 0, ",",
        d = 1, ",",
        q = 0, ")(",
        P = 0, ",",
        D = 1, ",",
        Q = 1, ")[30]",
        sep=""),adj=0,cex.main=0.75)
std.resid1 <- mod.fit1$fit$residuals / sqrt(mod.fit1$fit$sigma2)
hist(std.resid1,main = " Histogram Of Standardized Residuals",
   xlab='Standardized Residuals',
   freq = FALSE
)
```





```
$`fit`

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
         sma1
      -1.0000
s.e.   0.0152

sigma^2 estimated as 12.15:  log likelihood = -13438.07,  aic = 26880.13

$degrees_of_freedom
[1] 5008

$ttable
     Estimate     SE  t.value p.value
sma1       -1 0.0152 -65.6435       0

$AIC
[1] 5.335477

$AICc
[1] 5.335477

$BIC
[1] 5.338065
```

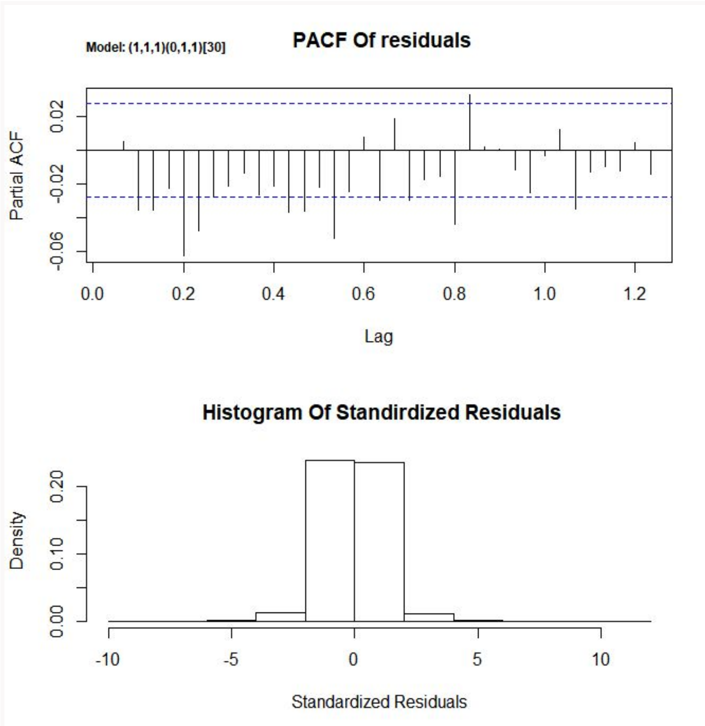
Trying ARIMA(p=0,d=1,q=1,P=0,D=1,Q=1,S=30)

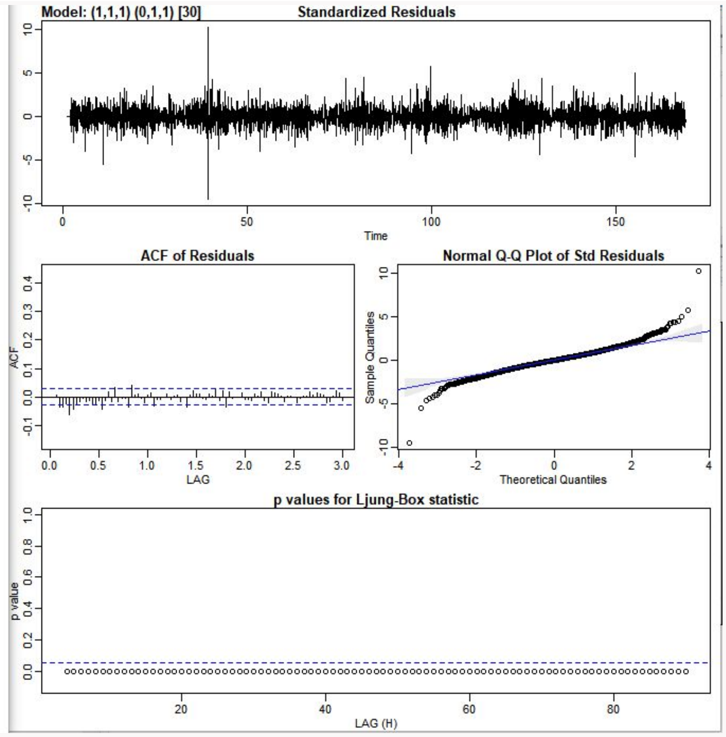



```
> mod.fit1
$`fit`

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
         ma1     sma1
      0.1538  -1.0000
s.e.  0.0135   0.0126

sigma^2 estimated as 11.85:  log likelihood = -13376.12,  aic = 26758.24

$degrees_of_freedom
[1] 5007

$ttable
     Estimate     SE  t.value p.value
ma1    0.1538 0.0135  11.3692       0
sma1  -1.0000 0.0126 -79.6657       0

$AIC
[1] 5.311282

$AICc
[1] 5.311282

$BIC
[1] 5.315164

>
```

ARIMA(p=0,d=1,q=1,P=0,D=1,Q=1,S=30)
model seems like better that the previous because **BIC = 5.315164** compare to
ARIMA(p=0,d=1,q=0,P=0,D=1,Q=1,S=30) BIC = 5.338065

Trying ARIMA(p=1,d=1,q=1,P=0,D=1,Q=1,S=30)

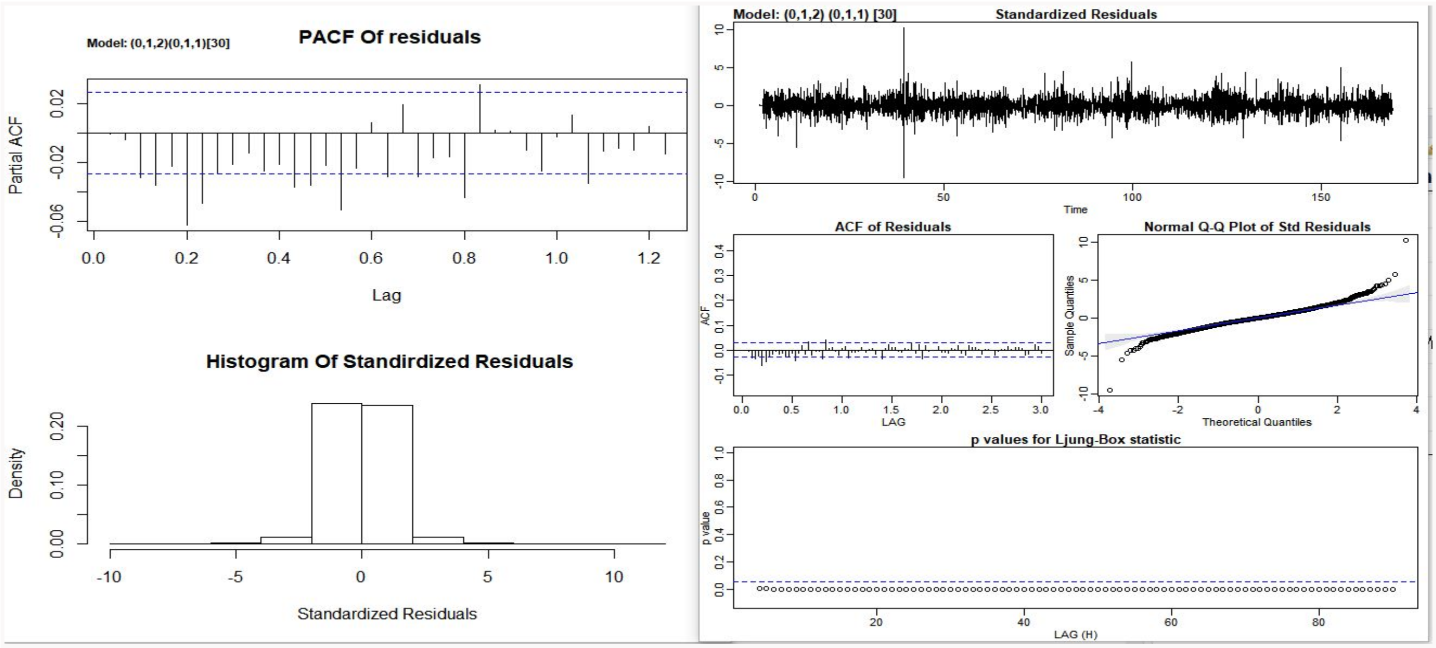





```
> mod.fit1
$`fit`

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
         ar1     ma1     sma1
      0.1370  0.0219  -1.0000
s.e.  0.0727  0.0730   0.0129

sigma^2 estimated as 11.84:  log likelihood = -13374.49,  aic = 26756.98

$degrees_of_freedom
[1] 5006

$ttable
     Estimate     SE  t.value p.value
ar1    0.1370 0.0727   1.8833  0.0597
ma1    0.0219 0.0730   0.2995  0.7646
sma1  -1.0000 0.0129 -77.7387  0.0000

$AIC
[1] 5.311032

$AICc
[1] 5.311033

$BIC
[1] 5.316208
```

The ARIMA(p=1,d=1,q=1,P=0,D=1,Q=1,S=30) model has p-value for ma1 > 0.05 and this makes ARIMA(p=0,d=1,q=1,P=0,D=1,Q=1,S=30) is better out of all previous models

Trying ARIMA(p=0,d=1,q=2,P=0,D=1,Q=1,S=30)

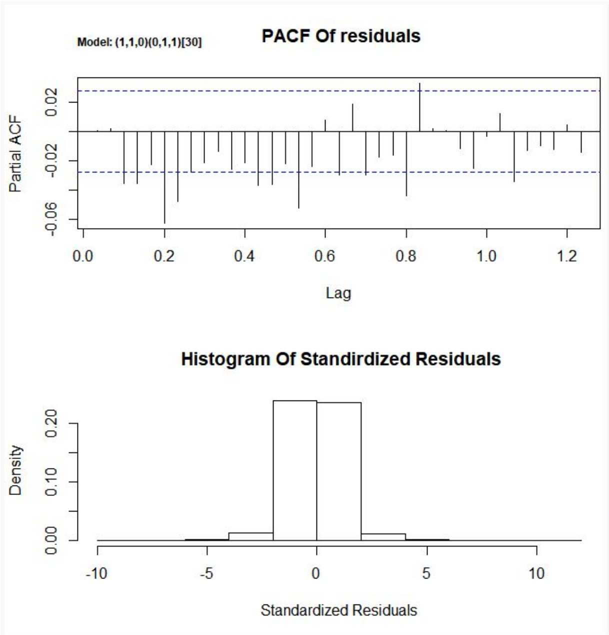


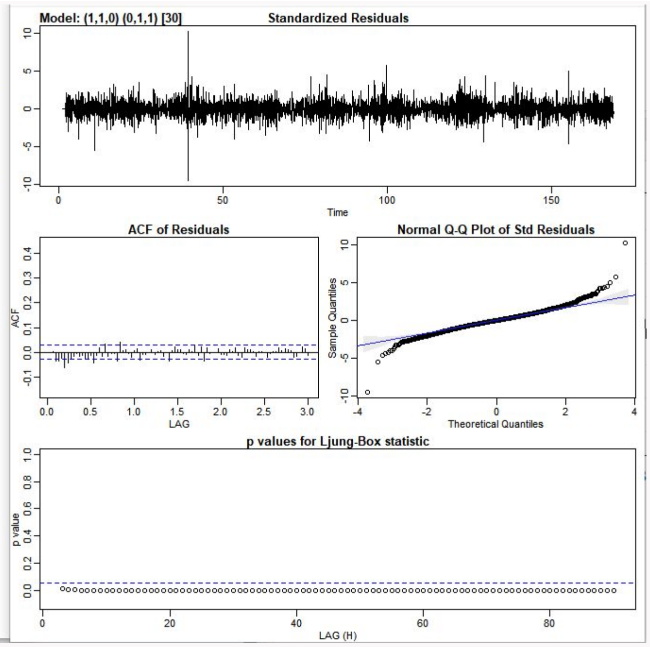

```
Coefficients:
         ma1     ma2    sma1
      0.1599  0.0320  -1.000
s.e.  0.0142  0.0147   0.013

sigma^2 estimated as 11.84:  log likelihood = -13373.75,  aic = 26755.5

$degrees_of_freedom
[1] 5006

$ttable
     Estimate     SE  t.value p.value
ma1    0.1599 0.0142  11.2785  0.0000
ma2    0.0320 0.0147   2.1813  0.0292
sma1  -1.0000 0.0130 -76.7035  0.0000

$AIC
[1] 5.310738

$AICc
[1] 5.310739

$BIC
[1] 5.315914
```

The ARIMA(p=0,d=1,q=2,P=0,D=1,Q=1,S=30) model has aal p-value for ma1 < 0.05 and this makes And BIC=5.315914
ARIMA(p=0,d=1,q=1,P=0,D=1,Q=1,S=30) BIC = 5.315164 < 5.315914 = BIC
ARIMA(p=0,d=1,q=2,P=0,D=1,Q=1,S=30) which makes the best out of all previous models
ARIMA(p=0,d=1,q=1,P=0,D=1,Q=1,S=30)

When Trying ARIMA(p=0,d=1,q=(3 or 4),P=0,D=1,Q=1,S=30) values of BIC is increasing and therefore the best model now is still ARIMA(p=0,d=1,q=1,P=0,D=1,Q=1,S=30)

Trying ARIMA(p=1,d=1,q=0,P=0,D=1,Q=1,S=30)

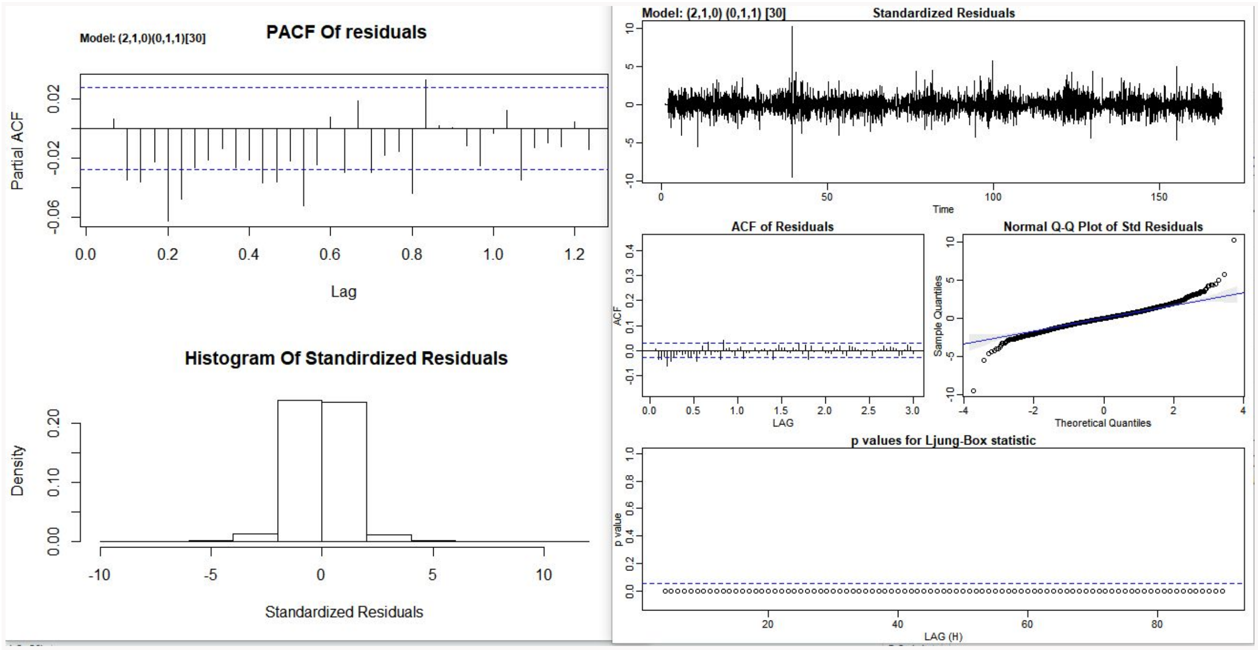





```
Coefficients:
         ar1     sma1
      0.1583  -1.0000
s.e.  0.0139   0.0129

sigma^2 estimated as 11.84:  log likelihood = -13374.53,  aic = 26755.07

$degrees_of_freedom
[1] 5007

$ttable
     Estimate     SE  t.value p.value
ar1    0.1583 0.0139  11.3460       0
sma1  -1.0000 0.0129 -77.3007       0

$AIC
[1] 5.310653

$AICc
[1] 5.310653

$BIC
[1] 5.314535
```

The ARIMA(p=1,d=1,q=0,P=0,D=1,Q=1,S=30) model has aal p-value for ma1 < 0.05 and
And BIC=5.314535 < BIC = 5.315164 of ARIMA(p=0,d=1,q=1,P=0,D=1,Q=1,S=30)
This makes ARIMA(p=1,d=1,q=0,P=0,D=1,Q=1,S=30) better model

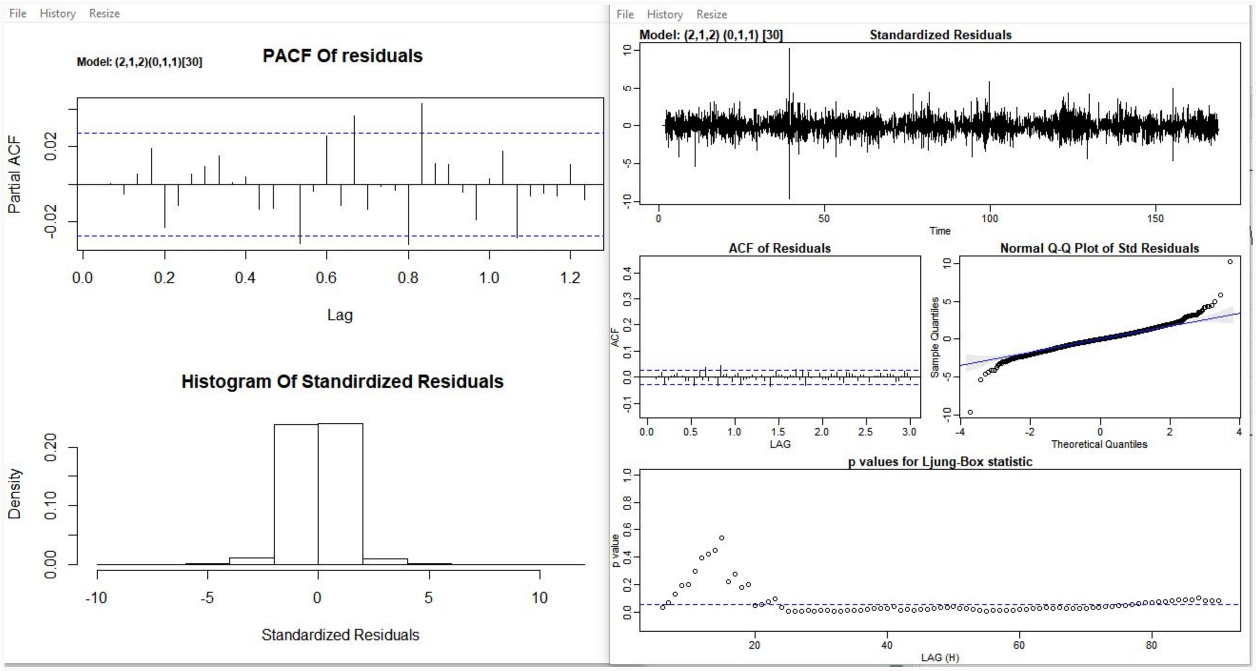
Trying ARIMA(p=2,d=1,q=0,P=0,D=1,Q=1,S=30)




```
$ttable
     Estimate     SE  t.value p.value
ar1    0.1591 0.0141  11.2615  0.0000
ar2   -0.0052 0.0141  -0.3683  0.7127
sma1  -1.0000 0.0128 -77.9575  0.0000

$AIC
[1] 5.311023

$AICc
[1] 5.311024

$BIC
[1] 5.316199

>
> dev.new()
```

Ar2 pvalue >0.05 which makes the ARIMA(p=1,d=1,q=0,P=0,D=1,Q=1,S=30) better model

Trying ARIMA(p=2,d=1,q=2,P=0,D=1,Q=1,S=30)




```
Coefficients:
         ar1      ar2      ma1     ma2     sma1
      1.3022  -0.3833  -1.1594  0.2056  -1.0000
s.e.  0.0682   0.0654   0.0722  0.0707   0.0102

sigma^2 estimated as 11.64:  log likelihood = -13332.36,  aic = 26676.72

$degrees_of_freedom
[1] 5004

$ttable
     Estimate     SE  t.value p.value
ar1    1.3022 0.0682  19.0957  0.0000
ar2   -0.3833 0.0654  -5.8623  0.0000
ma1   -1.1594 0.0722 -16.0541  0.0000
ma2    0.2056 0.0707   2.9097  0.0036
sma1  -1.0000 0.0102 -97.6950  0.0000

$AIC
[1] 5.295102

$AICc
[1] 5.295104

$BIC
[1] 5.302866
```

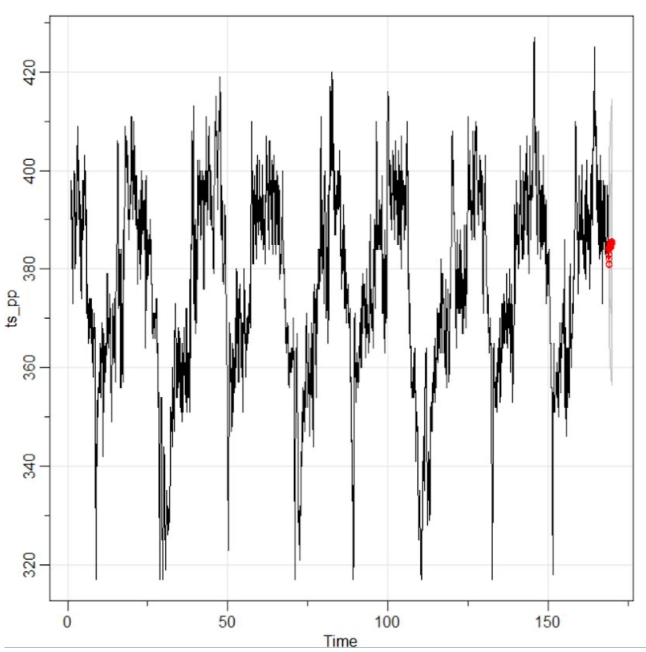
p values <0.05 which makes the ARIMA(p=2,d=1,q=2,P=0,D=1,Q=1,S=30) BIC =5.302866 < BIC=5.314535 ARIMA(p=1,d=1,q=0,P=0,D=1,Q=1,S=30) which makes ARIMA(p=2,d=1,q=2,P=0,D=1,Q=1,S=30) better than all above models

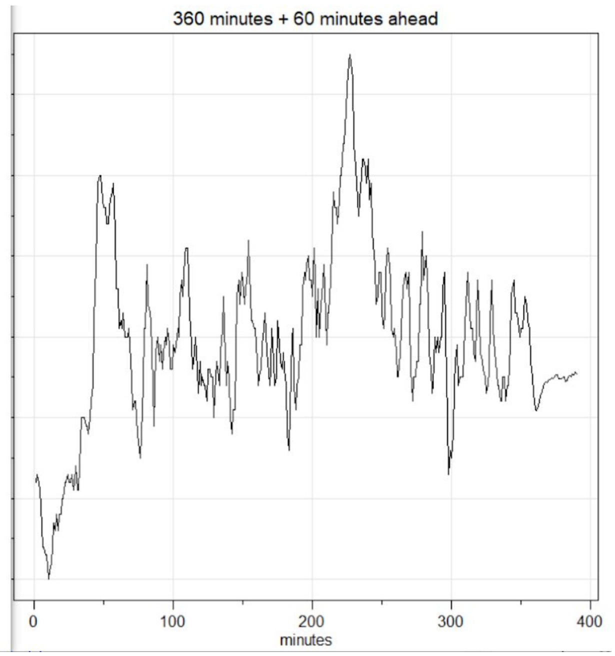
Trying ARIMA(p=3,d=1,q=3,P=0,D=1,Q=1,S=30)
Trying ARIMA(p=3,d=1,q=2,P=0,D=1,Q=1,S=30)
ARIMA(p=2,d=1,q=3,P=0,D=1,Q=1,S=30)
P-values >0.05 and therefore the best selected model is

ARIMA(p=2,d=1,q=2,P=0,D=1,Q=1,S=30)

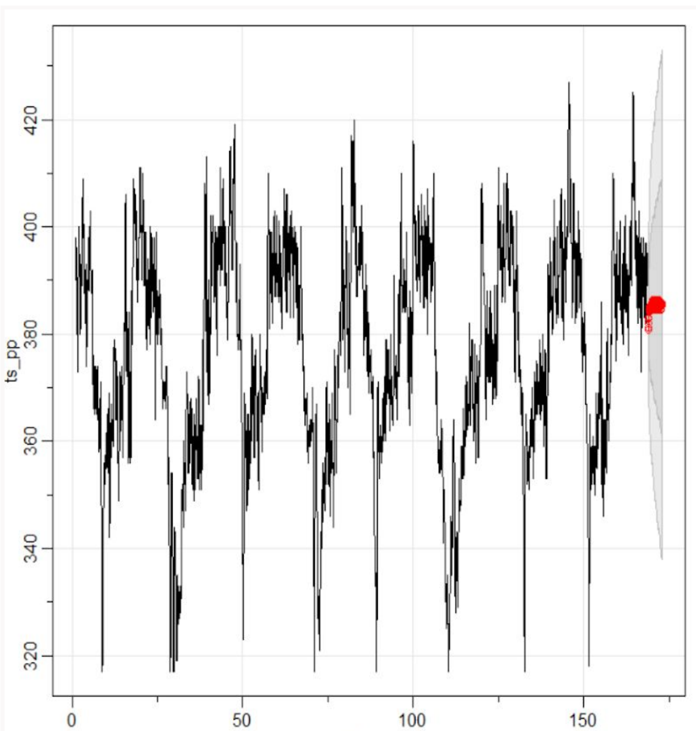
Forecasting

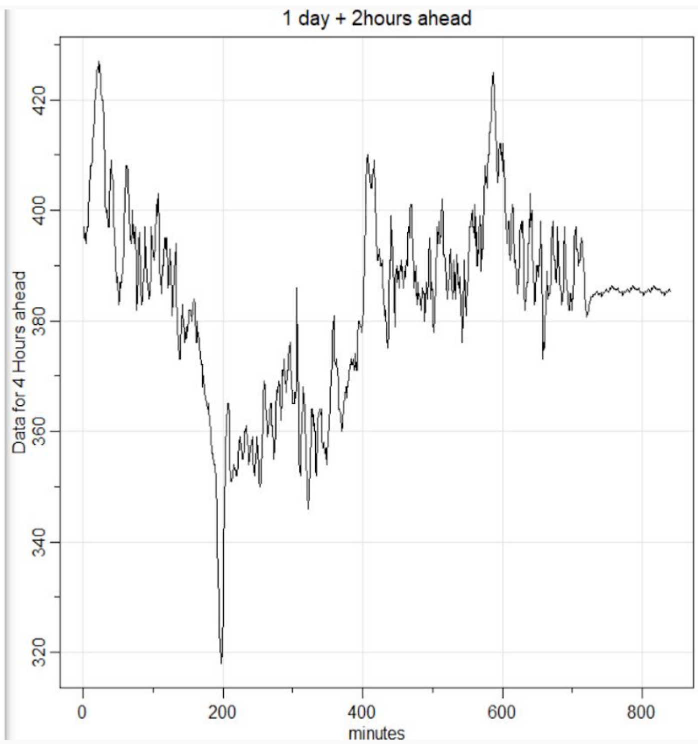
```
ahead_1_hours <-30
dev.new()
fore.mod <- sarima.for(ts_pp, n.ahead=ahead_1_hours,
      p=2, d=1, q=2, P=0, D=1, Q=1, S=30, plot.all=TRUE)
```



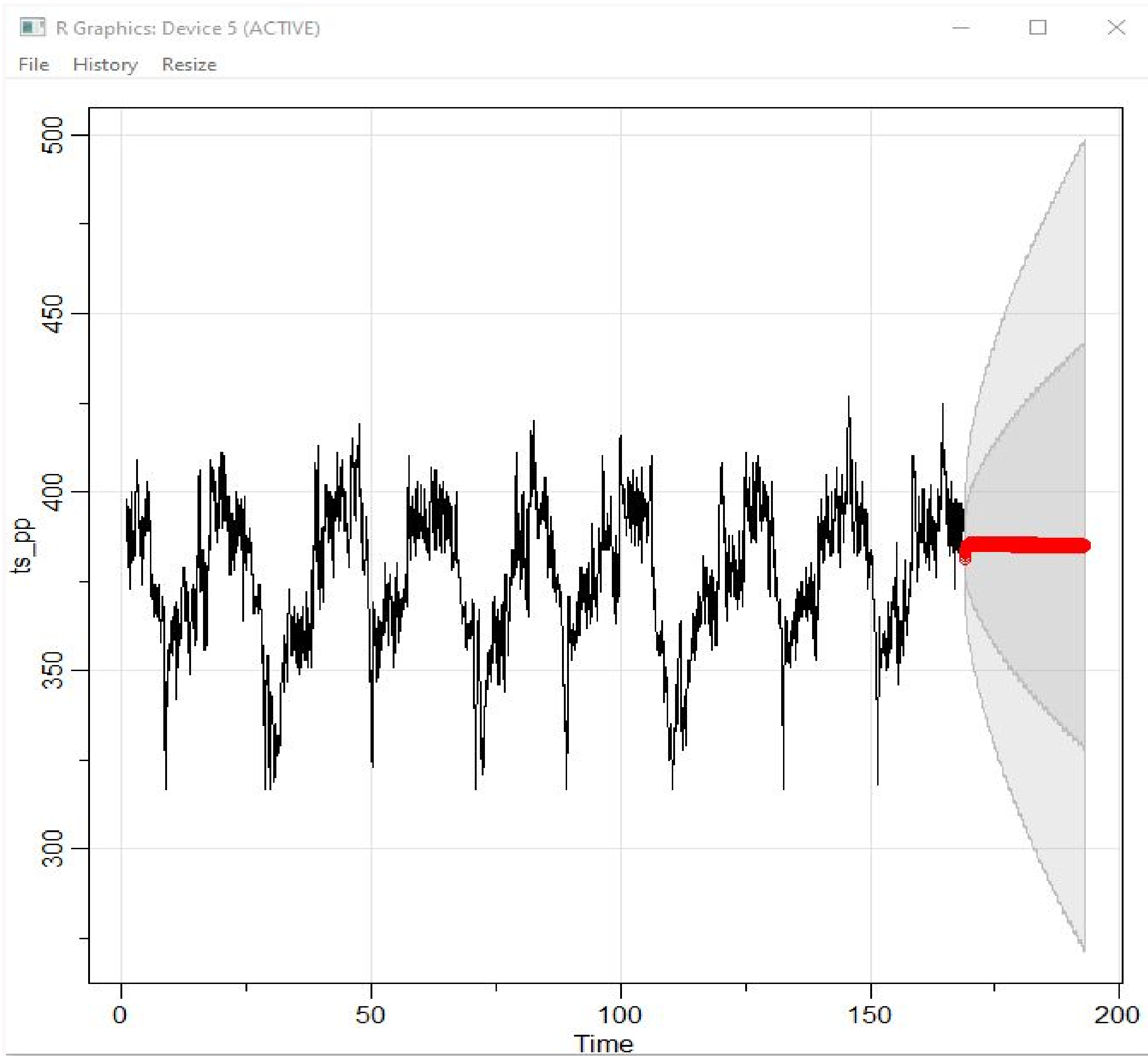


```
ahead_4_hours <- 4*30
dev.new()
fore.mod <- sarima.for(ts_pp, n.ahead=ahead_4_hours,
                p=2, d=1, q=2, P=0, D=1, Q=1, S=30, plot.all=TRUE)

dev.new()
ts_pp_4_hour <- c(ts_pp[(length(ts_pp)-6*ahead_4_hours + 1):(length(ts_pp))],
            fore.mod$pred[1:length(fore.mod$pred)])
tsplot(ts_pp_4_hour,
    ylab="Data for 4 Hours ahead",
    xlab="minutes", main=expression("1 day + 2hours ahead"))
```

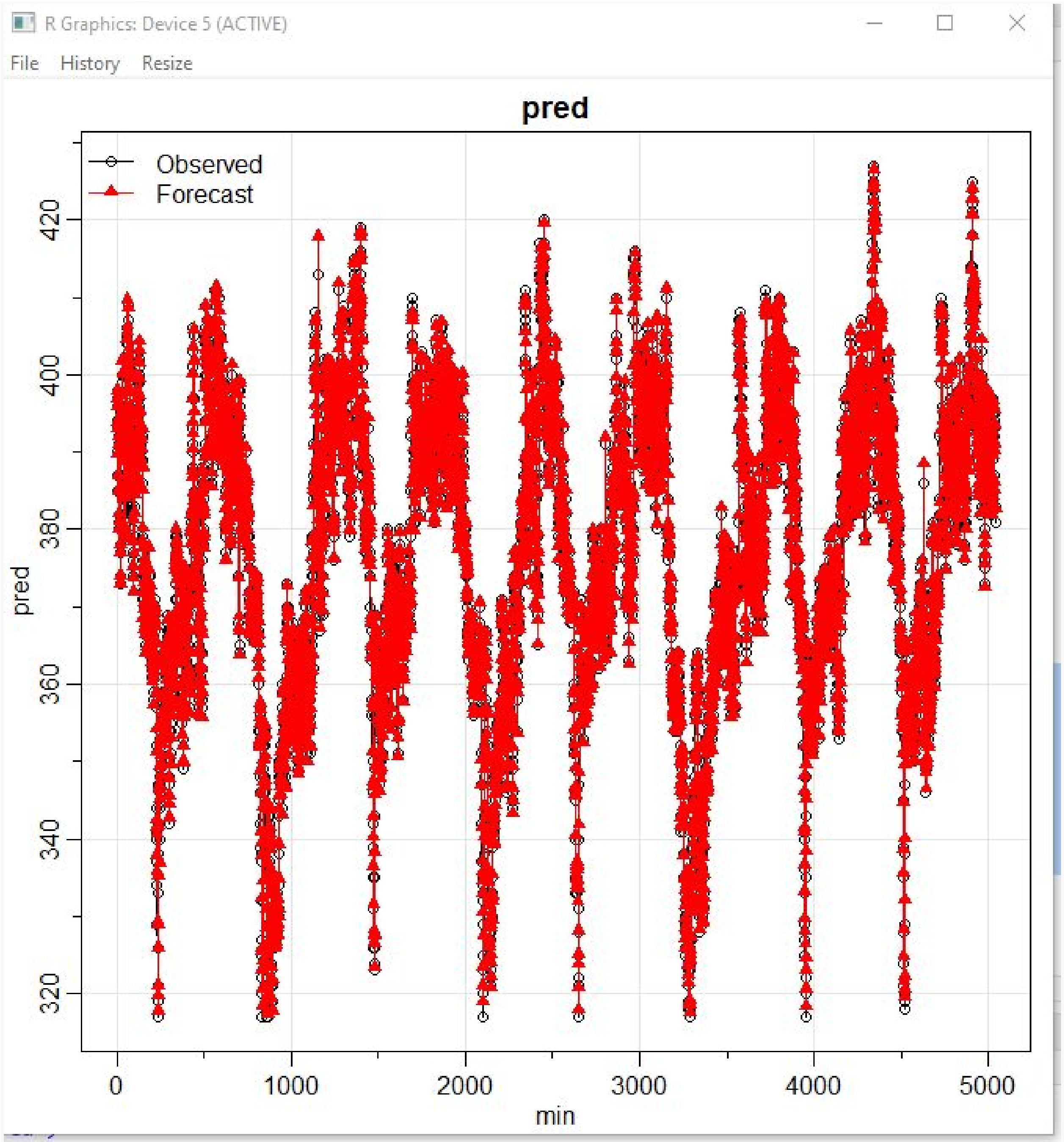




```
ahead_24_hours <- 24*30
dev.new()
fore.mod <- sarima.for(ts_pp, n.ahead=ahead_4_hours,
                p=2, d=1, q=2, P=0, D=1, Q=1, S=30, plot.all=TRUE)
```

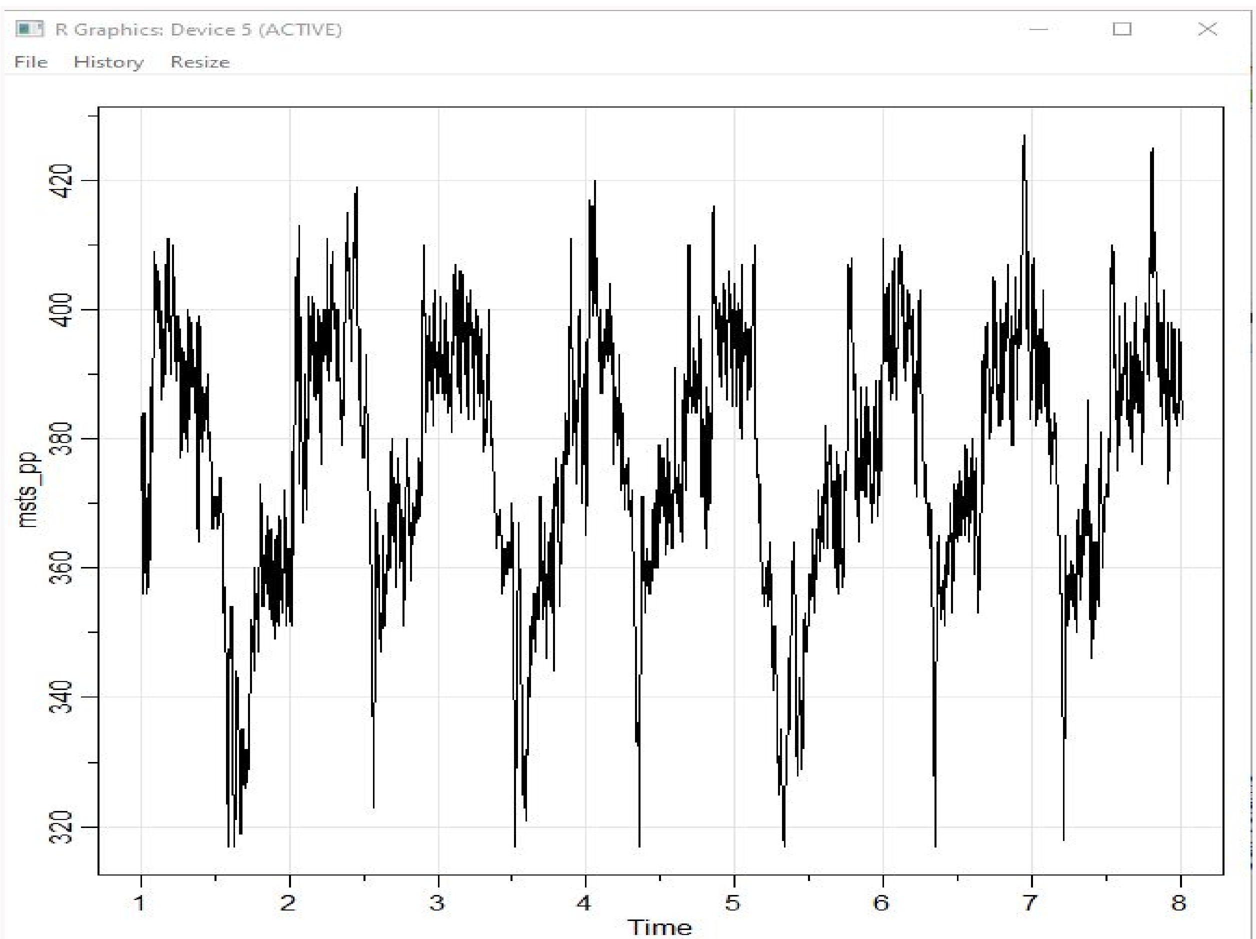


```
pred.mod <- ts(x - mod.fit.110.011$fit$residuals)
dev.new()
tsplot(ts_pp[1:length(ts_pp)], ylab="pred", xlab="min", type="o", main="pred")
lines(pred.mod, col="red", type="o", pch=17)
legend("topleft", legend=c("Observed", "Forecast"), lty=c("solid", "solid"), col=c("black",
"red"), pch=c(1, 17), bty="n")
```



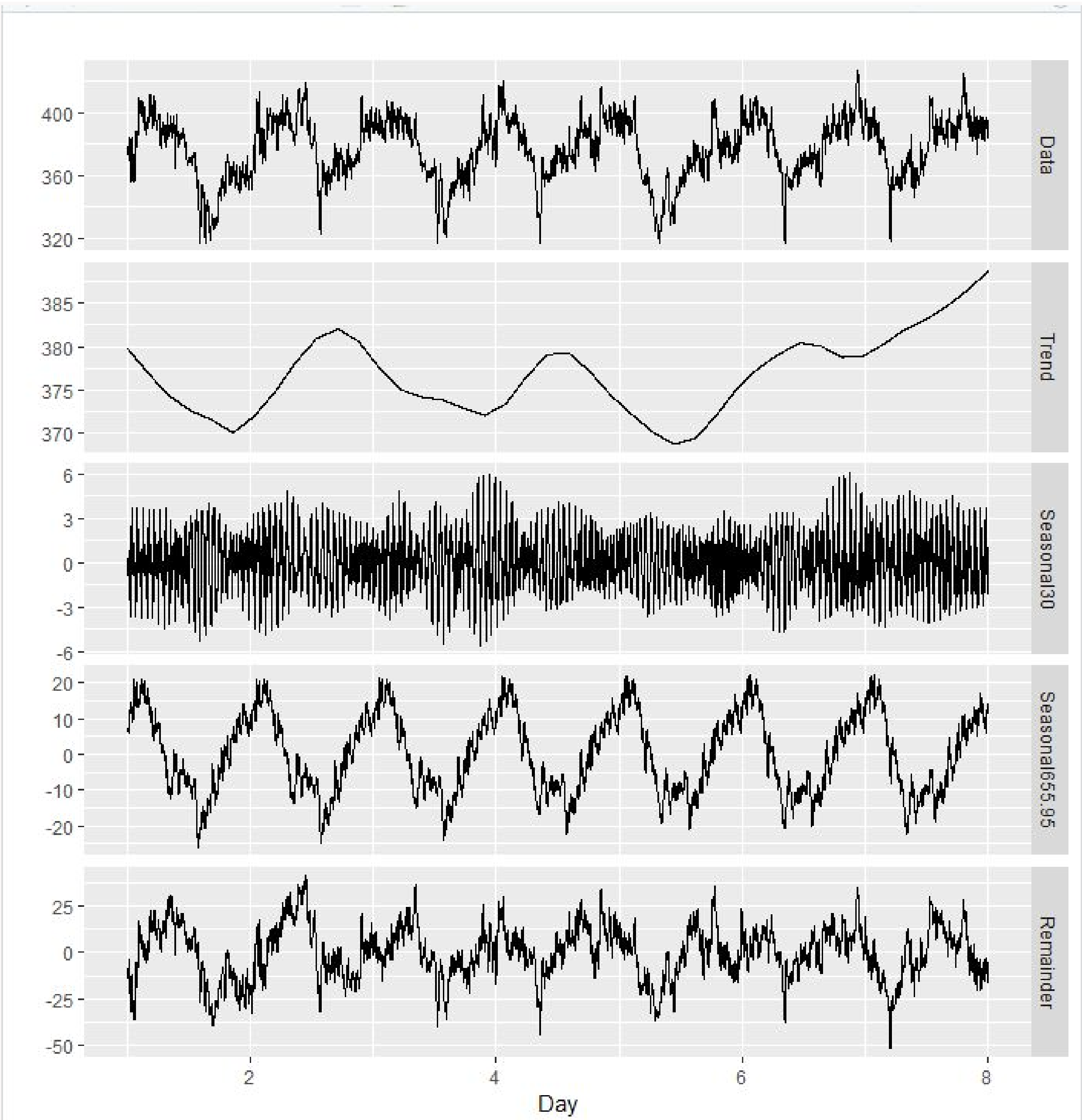
Discussion

```
On_hour <- 30
one_day <- 30*24
one_week <- 30*24*7
frequency <- c(30,one_day)
msts_pp <- msts(ts_p[(length(ts_p)-(one_week) + 1 ):(length(ts_p))],seasonal.periods =
frequency)
dev.new()tsplot(msts_pp)
```



The `mstl()` function is a variation on `stl()` designed to deal with multiple seasonality. It will return multiple seasonal components, as well as a trend and remainder component.

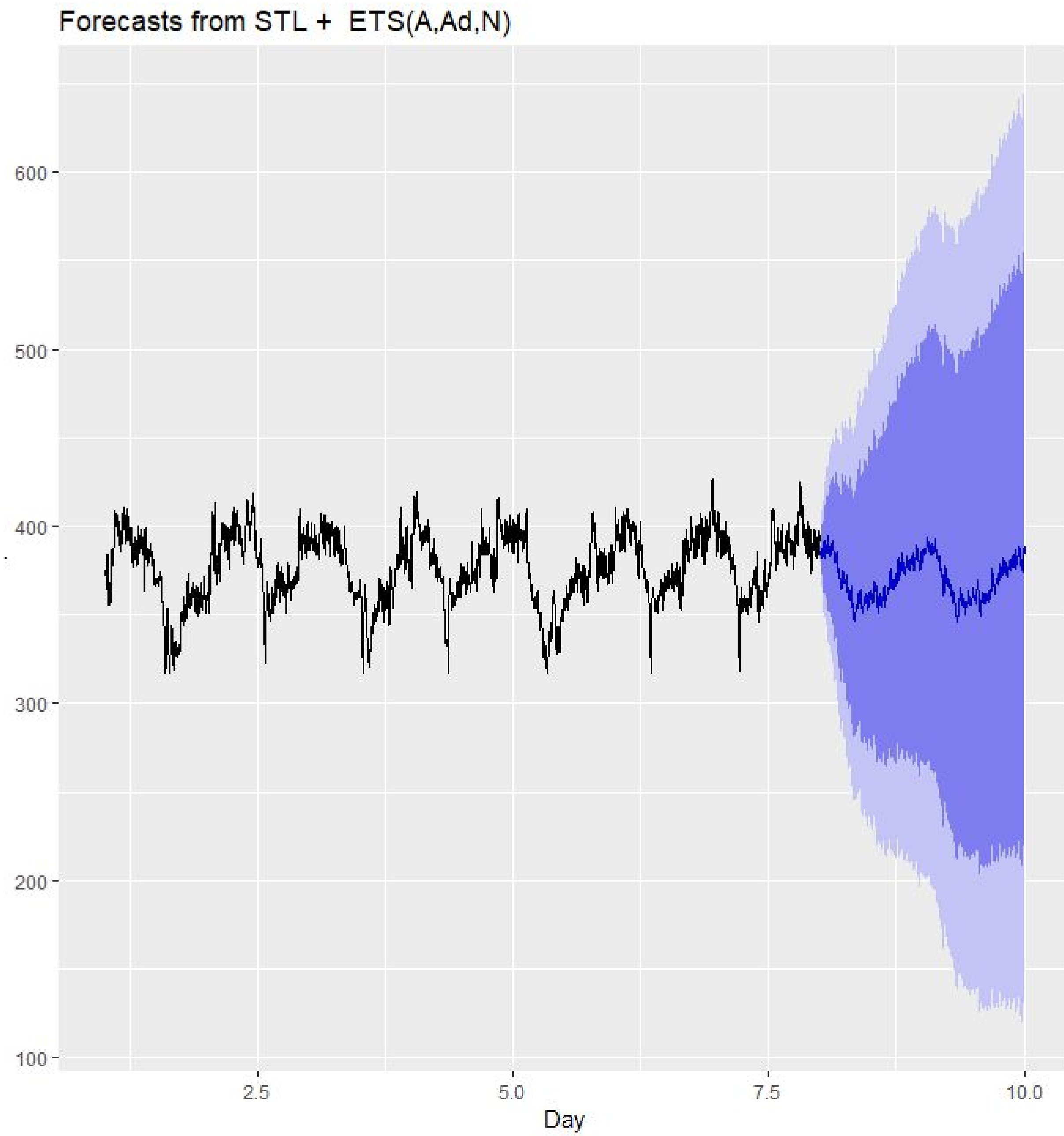
```
msts_pp %>% mstl() %>%autoplot() + xlab("Day")
```



There are two seasonal patterns shown, one for the time of minutes in hour (the third panel), and one for the time of day (the fourth panel). To properly interpret this graph, it is important to notice the vertical scales. In this case,

the trend and the daily seasonality have relatively narrow ranges compared to the other components, because there is little trend seen in the data, and the daily seasonality is weak.
The decomposition can also be used in forecasting, with each of the seasonal components forecast using a seasonal naïve method, and the seasonally adjusted data forecasting using ETS (or some other user-specified method). The `stlf()` function will do this automatically.

```
msts_pp %>%  stlf() %>%
 autoplot() + xlab("Day")
```



Dynamic harmonic regression with multiple seasonal periods

Because there are multiple seasonalities, we need to add Fourier terms for each seasonal period. In this case, the seasonal periods are 30 and 656, so the Fourier terms are of the form

$\sin(2\pi kt/30)$, $\cos(2\pi kt/30)$, $\sin(2\pi kt/656)$, $\cos(2\pi kt/656)$

For $k = 1,2, \ldots$

The `fourier()` function can generate these for you.

We will fit a dynamic harmonic regression model with an ARMA error structure. The total number of Fourier terms for each seasonal period have been chosen to minimise the AICc. We will use a log transformation (`lambda=0`) to ensure the forecasts and prediction intervals remain positive.

```
fit <- auto.arima(msts_pp, seasonal=FALSE, lambda=0,xreg=fourier(msts_pp, K=c(14,65)))
fit
```

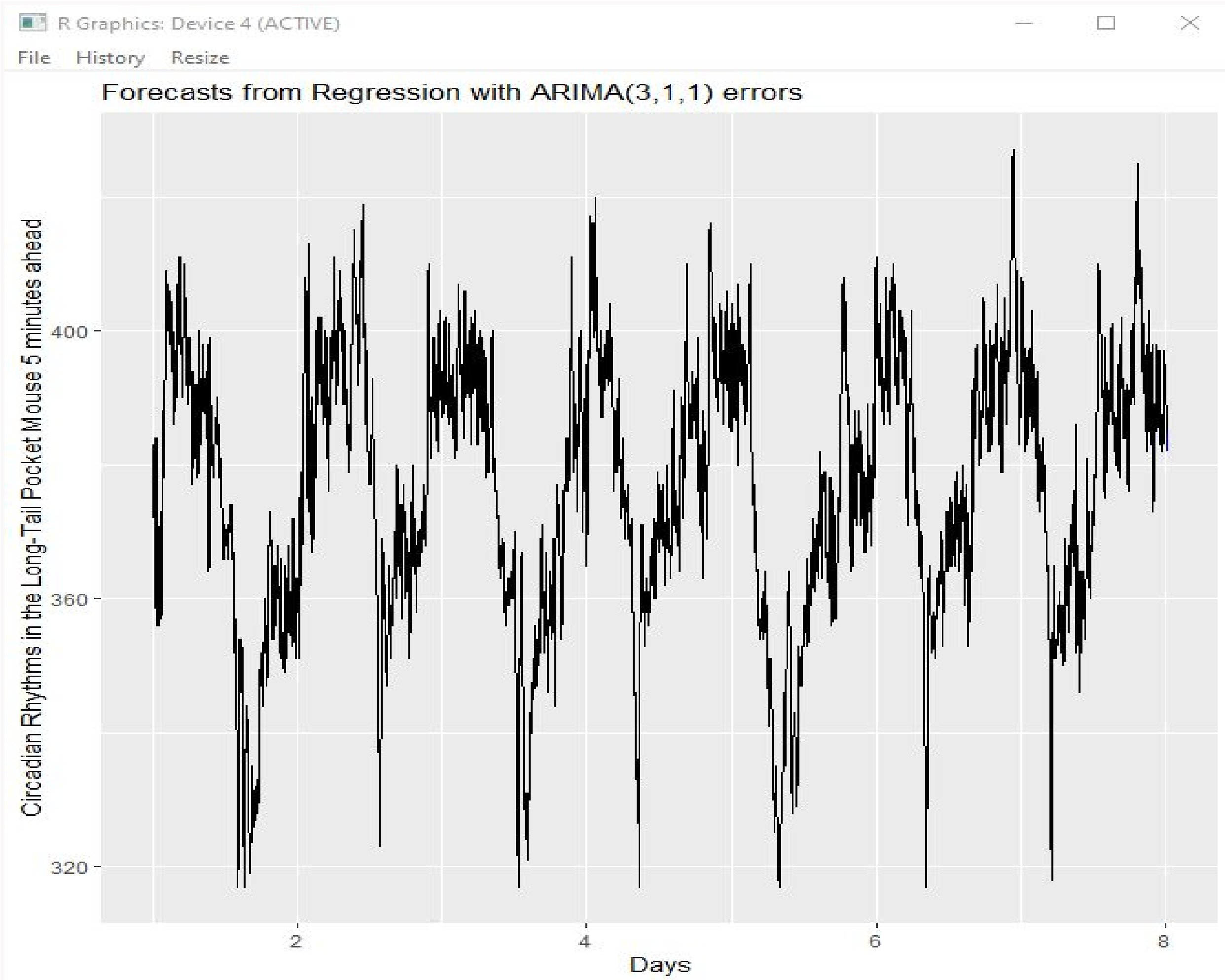
```
> fit <- auto.arima(msts_pp, seasonal=FALSE, lambda=0,
+                   xreg=fourier(msts_pp, K=c(14,65)))
> fit
Series: msts_pp
Regression with ARIMA(3,1,1) errors
Box Cox transformation: lambda= 0
```

```
Coefficients:
         ar1      ar2      ar3      ma1  S1-30    C1-30    S2-30  C2-30    S3-30  C3-30  S4-30  C4-30  S5-30  C5-30  S6-30    C6-30  S7-30  C7-30
      1.0682  -0.1358  -0.0397  -0.9440  9e-04   -1e-04   -1e-04  2e-04   -3e-04  6e-04  2e-04  2e-04  1e-04  0e+00  2e-04   -3e-04  3e-04  1e-04
s.e.  0.0187   0.0216   0.0154   0.0117  1e-03    1e-03    6e-04  6e-04    4e-04  4e-04  3e-04  3e-04  2e-04  2e-04  2e-04    2e-04  1e-04  1e-04
       S8-30    C8-30    S9-30  C9-30  S10-30  C10-30  S11-30  C11-30  S12-30  C12-30  S13-30  C13-30  S14-30  C14-30  S1-656  C1-656  S2-656
       0e+00   -1e-04   -1e-04  0e+00  -2e-04   0e+00   0e+00  -1e-04   1e-04   0e+00   0e+00  -1e-04  -1e-04  -2e-04  0.0069  0.0343  0.0034
```

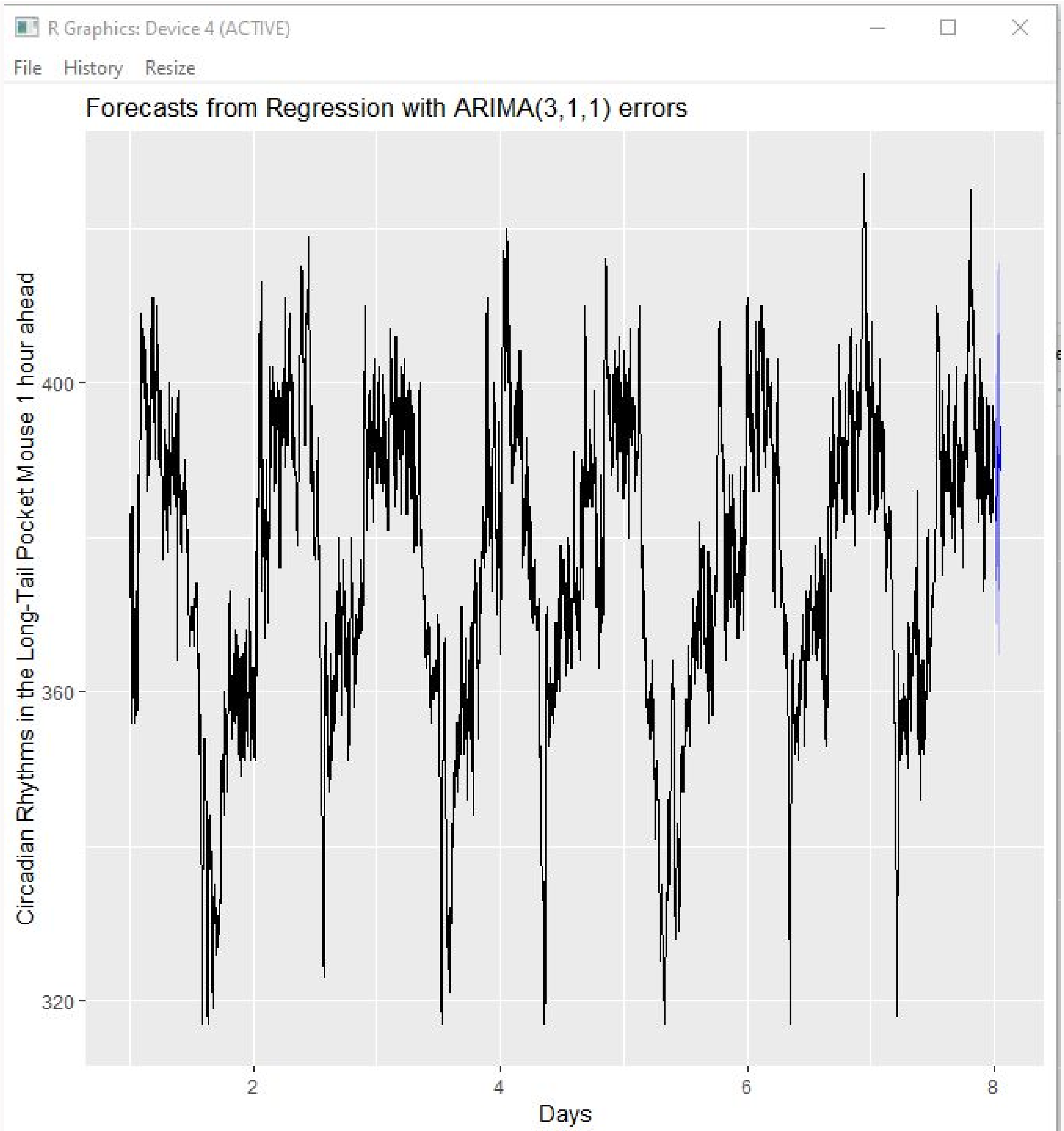
```
s.e.   1e-04   1e-04   1e-04  1e-04    1e-04   1e-04    1e-04   1e-04    1e-04    1e-04    1e-04   1e-04   1e-04   1e-04  0.0102   0.0102   0.0053
      C2-656  S3-656   C3-656   S4-656  C4-656  S5-656   C5-656  S6-656   C6-656   S7-656   C7-656  S8-656  C8-656  S9-656   C9-656  S10-656
      0.0039  0.0059  -0.0076  -0.0012  0.0018  0.0009  -0.0032  0.0000  -0.0012  -0.0033  -0.0013  0.0038  0.0017  0.0013  -0.0007  -0.0005
s.e.  0.0053  0.0037   0.0037   0.0029  0.0029  0.0025   0.0025  0.0022   0.0022   0.0020   0.0020  0.0019  0.0019  0.0018   0.0018   0.0017
      C10-656  S11-656  C11-656  S12-656  C12-656  S13-656  C13-656  S14-656  C14-656  S15-656  C15-656  S16-656  C16-656  S17-656  C17-656
      -0.0010   0.0000  -0.0006   0.0015  -0.0022  -0.0009   0.0035   0.0024  -0.0023   0.0003   0.0003  -0.0025   0.0017   0.0005   0.0006
s.e.   0.0017   0.0016   0.0016   0.0015   0.0015   0.0014   0.0014   0.0014   0.0014   0.0013   0.0013   0.0013   0.0013   0.0012   0.0012
      S18-656  C18-656  S19-656  C19-656  S20-656  C20-656  S21-656  C21-656  S22-656  C22-656  S23-656  C23-656  S24-656  C24-656  S25-656
      -0.0002   0.0008   0.0024   0.0005  -0.0010   0.0024  -0.0009  -0.0016   -6e-04   -1e-04    5e-04   -8e-04    6e-04  -0.0011    6e-04
s.e.   0.0012   0.0012   0.0011   0.0011   0.0011   0.0011   0.0011   0.0011    1e-03    1e-03    1e-03    1e-03    1e-03   0.0010    9e-04
      C25-656  S26-656  C26-656  S27-656  C27-656  S28-656  C28-656  S29-656  C29-656  S30-656  C30-656  S31-656  C31-656  S32-656  C32-656
      -0.0011   -9e-04    8e-04   -4e-04   -6e-04   0.0019   -3e-04   -8e-04    0e+00   -7e-04  -0.0013    1e-04    9e-04   -3e-04    8e-04
s.e.   0.0009    9e-04    9e-04    9e-04    9e-04   0.0008    8e-04    8e-04    8e-04    8e-04   0.0008    8e-04    8e-04    8e-04    8e-04
      S33-656  C33-656  S34-656  C34-656  S35-656  C35-656  S36-656  C36-656  S37-656  C37-656  S38-656  C38-656  S39-656  C39-656  S40-656
        4e-04   -2e-04   -7e-04   -8e-04   -1e-03   -4e-04   -6e-04   -2e-04    0e+00    5e-04    1e-04    1e-03  -0.0013  -0.0015    0e+00
```

```
s.e.     7e-04     7e-04     7e-04     7e-04     7e-04     7e-04     7e-04     7e-04     7e-04     7e-04     6e-04     6e-04    0.0006    0.0006     6e-04
       C40-656   S41-656   C41-656   S42-656   C42-656   S43-656   C43-656   S44-656   C44-656   S45-656   C45-656   S46-656   C46-656   S47-656   C47-656
         1e-04     0e+00    -2e-04    -7e-04     6e-04    -4e-04    -1e-04     3e-04    -8e-04    -5e-04    -1e-04    -2e-04     4e-04     2e-04     5e-04
s.e.     6e-04     6e-04     6e-04     6e-04     6e-04     6e-04     6e-04     6e-04     6e-04     5e-04     5e-04     5e-04     5e-04     5e-04     5e-04
       S48-656   C48-656   S49-656   C49-656   S50-656   C50-656   S51-656   C51-656   S52-656   C52-656   S53-656   C53-656   S54-656   C54-656   S55-656
         0e+00    -4e-04     2e-04    -5e-04     0e+00     2e-04    -4e-04    -6e-04    -1e-04    -4e-04     1e-04    -8e-04    -2e-04    -3e-04    -5e-04
s.e.     5e-04     5e-04     5e-04     5e-04     5e-04     5e-04     5e-04     5e-04     5e-04     5e-04     5e-04     5e-04     5e-04     5e-04     4e-04
       C55-656   S56-656   C56-656   S57-656   C57-656   S58-656   C58-656   S59-656   C59-656   S60-656   C60-656   S61-656   C61-656   S62-656   C62-656
        -1e-04    -2e-04    -1e-04     1e-04     3e-04    -2e-04     3e-04     5e-04     0e+00     2e-04    -1e-04    -1e-04     1e-04    -5e-04    -1e-04
s.e.     4e-04     4e-04     4e-04     4e-04     4e-04     4e-04     4e-04     4e-04     4e-04     4e-04     4e-04     4e-04     4e-04     4e-04     4e-04
       S63-656   C63-656   S64-656   C64-656   S65-656   C65-656
         6e-04     1e-04     4e-04     6e-04    -1e-04     4e-04
s.e.     4e-04     4e-04     4e-04     4e-04     4e-04     4e-04

sigma^2 estimated as 8.103e-05:  log likelihood=15190.22
AIC=-30054.44   AICc=-30042.36   BIC=-29006.08
```

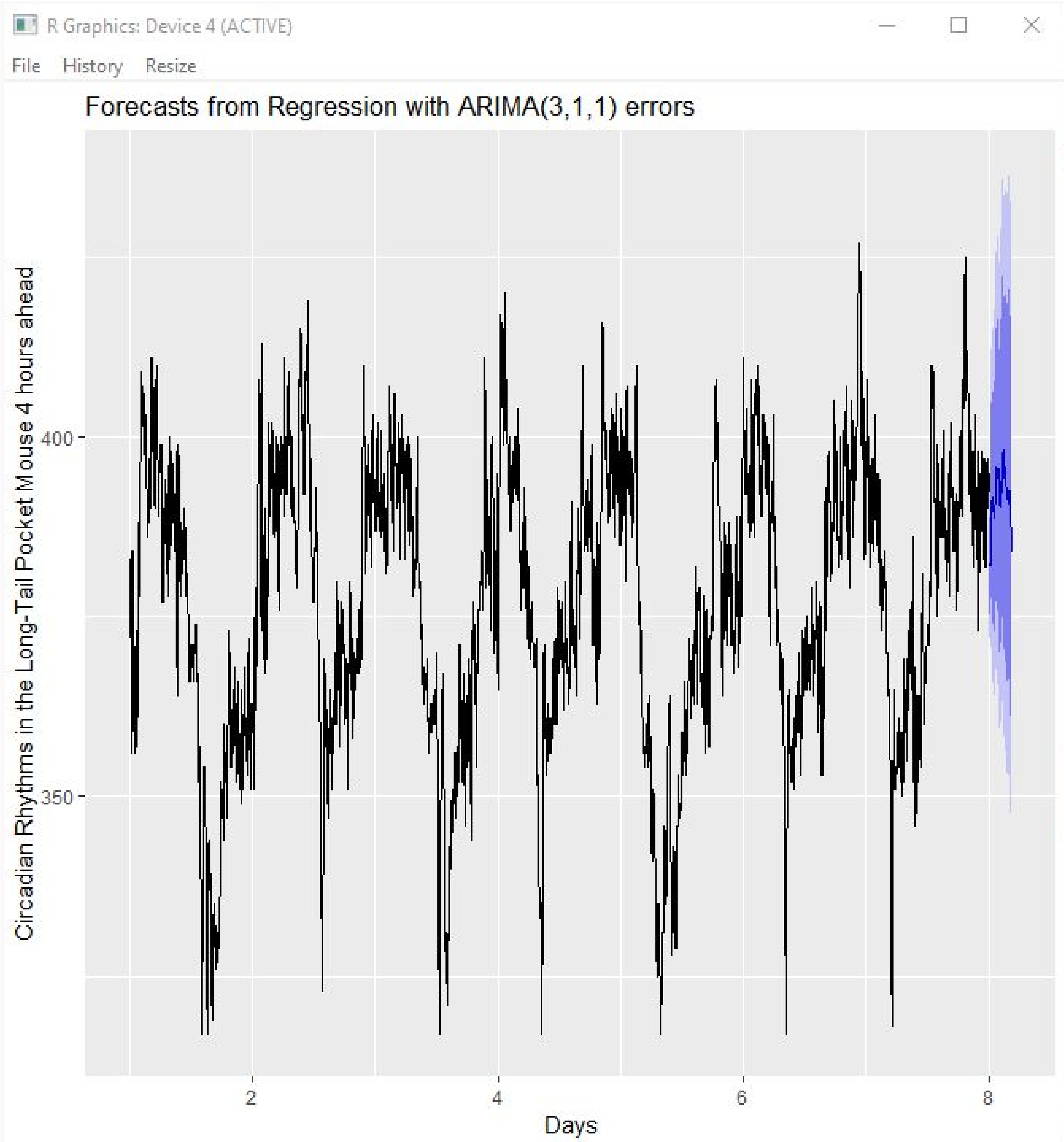
```
dev.new()
minutes_10<- 5
fit %>%
 forecast(xreg=fourier(msts_pp, K=c(14,65), minutes_10)) %>%
 autoplot(include=one_week) + ylab("Circadian Rhythms in the Long-Tail Pocket Mouse 10 minutes ahead") + xlab("Days")
```



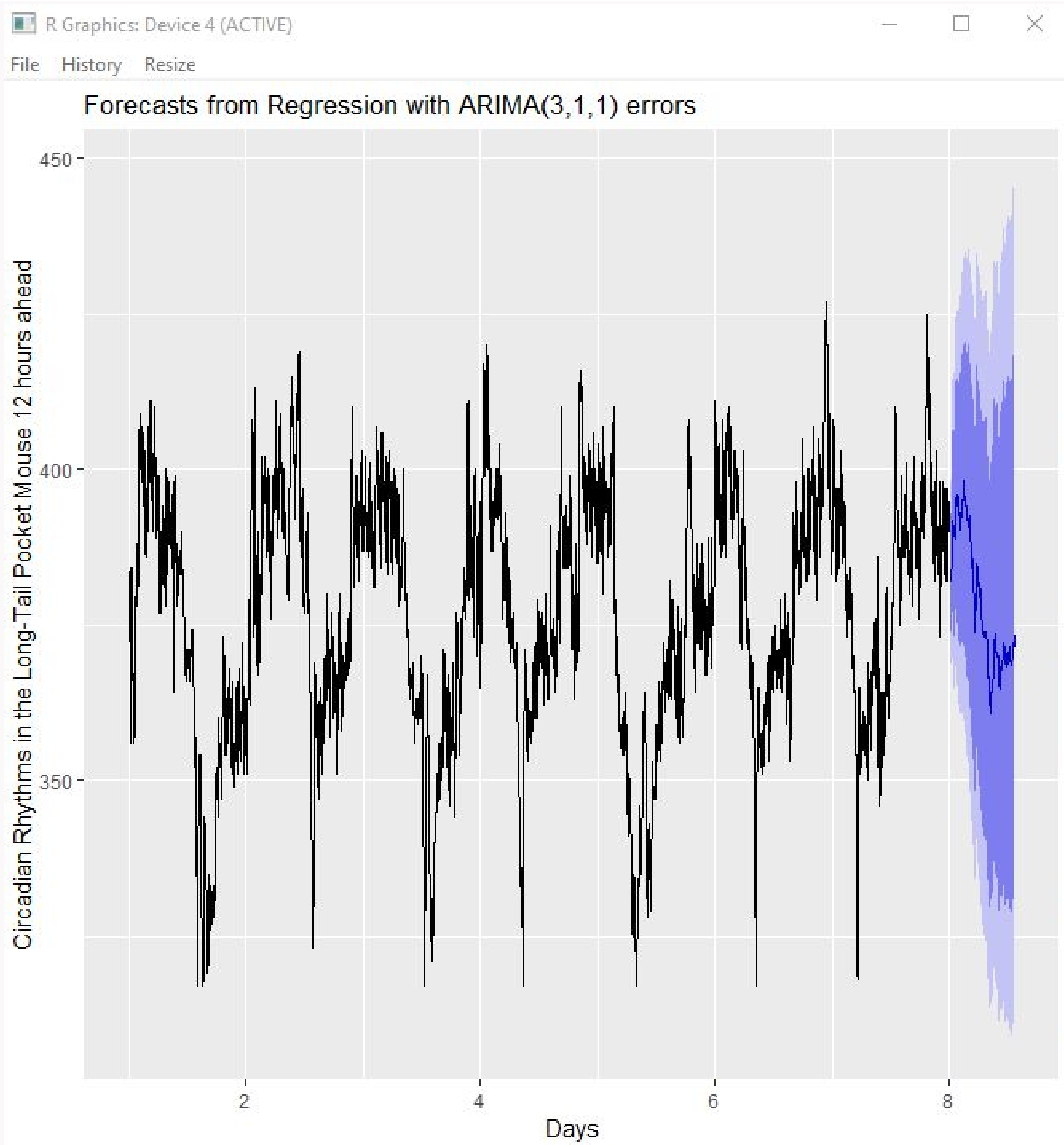
```
dev.new()
hour_1<- 30
fit %>%
 forecast(xreg=fourier(msts_pp, K=c(14,65), h=hour_1)) %>%
 autoplot(include=one_week) + ylab("Circadian Rhythms in the Long-Tail Pocket Mouse 1
hour ahead") + xlab("Days")
```



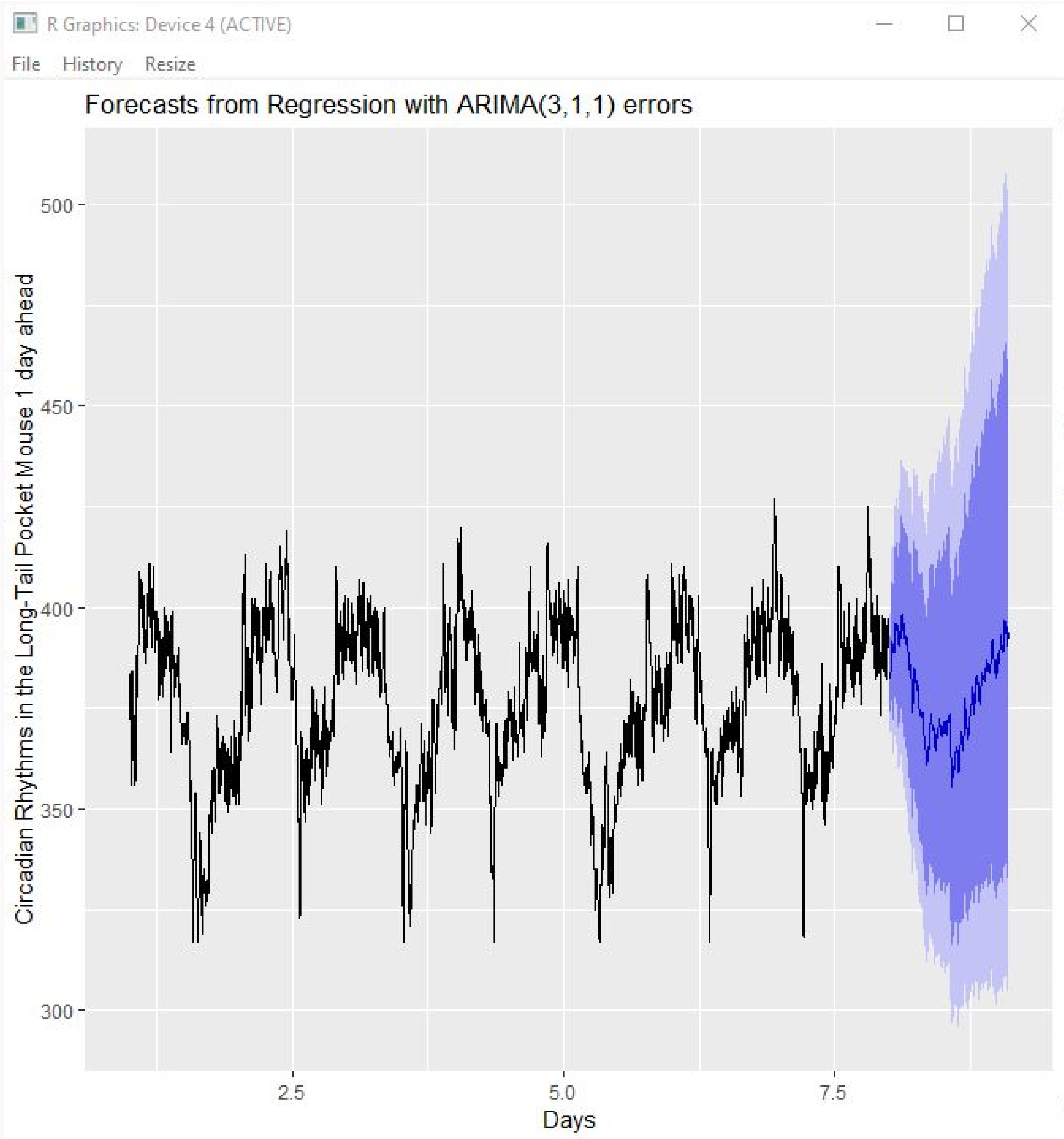
```
dev.new()
hours_4<- 30 *4
fit %>%
 forecast(xreg=fourier(msts_pp, K=c(14,65), h=hours_4)) %>%
 autoplot(include=one_week) + ylab("Circadian Rhythms in the Long-Tail Pocket Mouse 4
hours ahead") + xlab("Days")
```



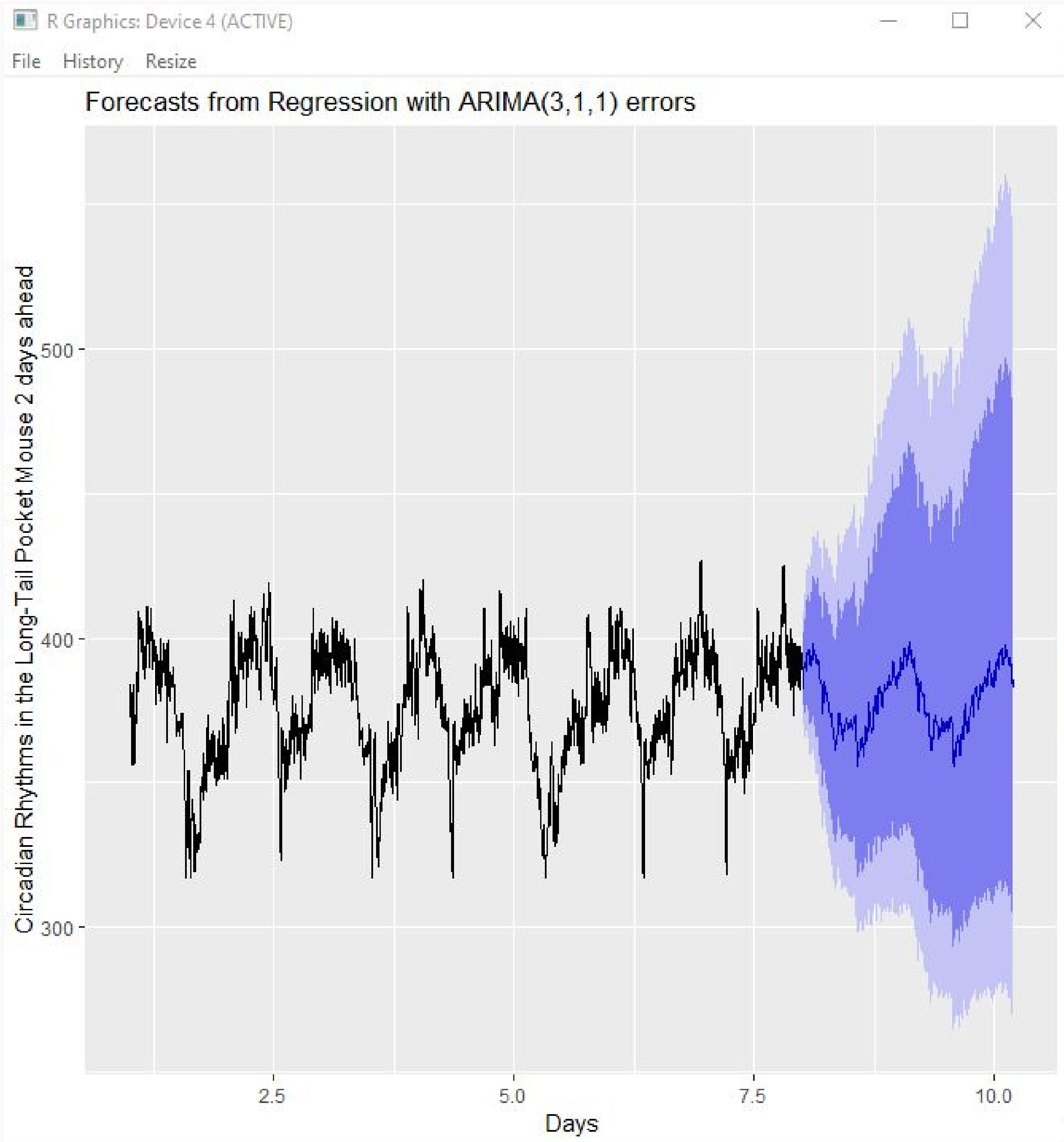
```
dev.new()
hours_12<- 30 * 12
fit %>%
 forecast(xreg=fourier(msts_pp, K=c(14,65), h=hours_12)) %>%
 autoplot(include=one_week) + ylab("Circadian Rhythms in the Long-Tail Pocket Mouse 12 hours ahead") + xlab("Days")
```



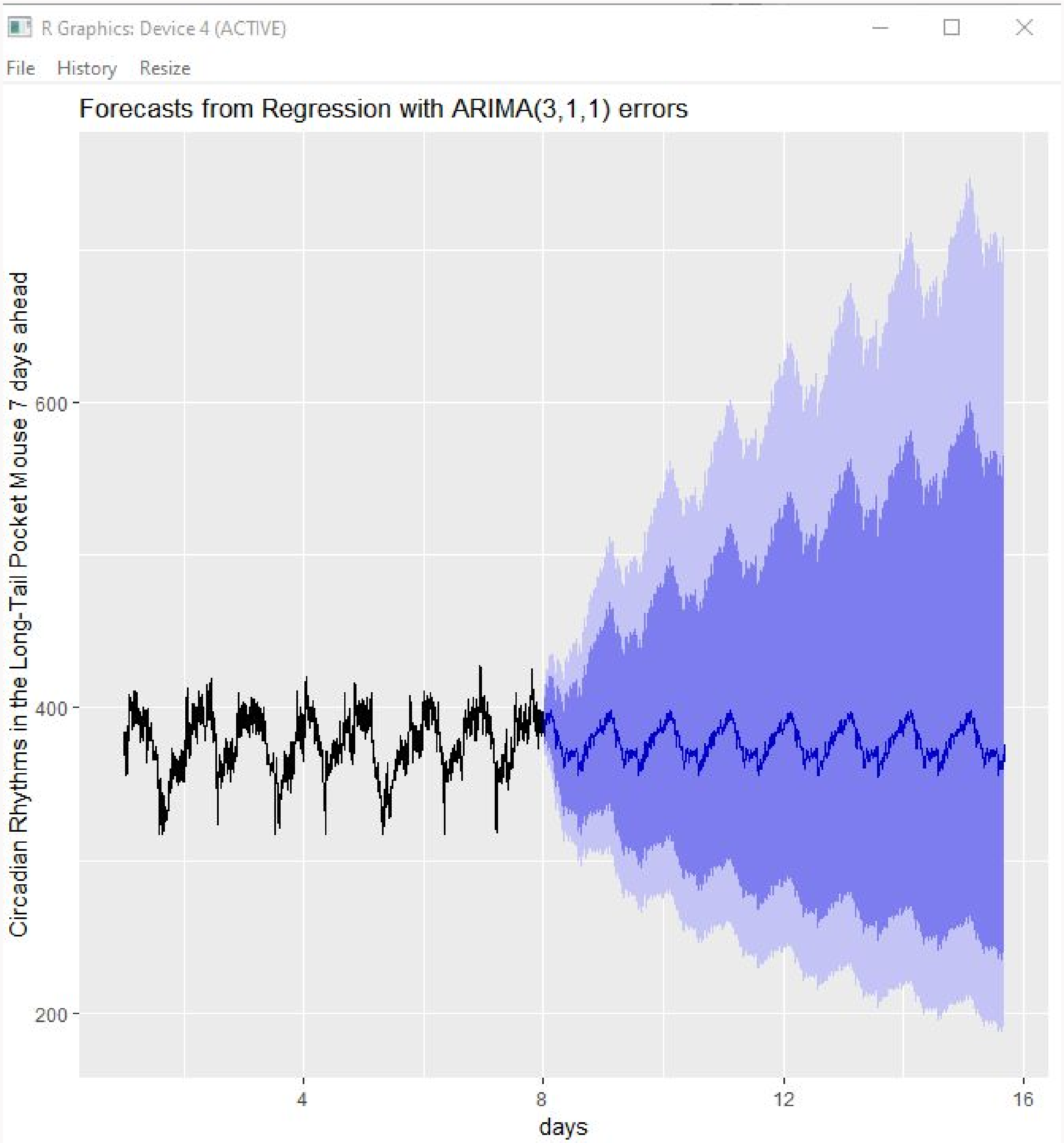
```
dev.new()
day_1<- 30 * 24
fit %>%
 forecast(xreg=fourier(msts_pp, K=c(14,65), h=day_1)) %>%
 autoplot(include=one_week) + ylab("Circadian Rhythms in the Long-Tail Pocket Mouse 1 day ahead") + xlab("Days")
```



```
dev.new()
day_2<- 30 * 24 * 2
fit %>%
 forecast(xreg=fourier(msts_pp, K=c(14,65), h=day_2)) %>%
 autoplot(include=one_week) + ylab("Circadian Rhythms in the Long-Tail Pocket Mouse 2 days ahead") + xlab("Days")
```



```
dev.new()
week_1<- 30 * 24 * 7
fit %>%
 forecast(xreg=fourier(msts_pp, K=c(14,65), h=week_1)) %>%
 autoplot(include=one_week) + ylab("Circadian Rhythms in the Long-Tail Pocket Mouse 7 days ahead") + xlab("days")
```



Bibliography

David Benson, Forecasting Daily Data with Multiple Seasonality in R
http://www.dbenson.co.uk/Rparts/subpages/forecastR/

E. E. Holmes, M.D. Scheuerell, and E. J. Ward 2020-02-03,
Analysis for Fisheries and Environmental Sciences

Rob J Hyndman, George Athanasopoulos,
Forecasting Principles And Practice
DataSet http://www.statsci.org/data/general/pformos5.html

Appendix

```
library(astsa)
library("forecast")
library("lubridate")
examine.mod <- function(mod.fit.obj, p, d, q, P=0, D=0, Q=0, S=-1, lag.max=24) {
 dev.new(width=6, height=6)
 par(mfrow=c(2,1))
 pacf(mod.fit.obj$fit$residuals, main="PACF of Residuals", lag.max)
 if ((P==0)&(D==0)&(Q==0)) {
  title(paste("Model: (", p, ",", d, ",", q, ")", sep=""), adj=0, cex.main=0.75)
 }
 else {
  title(paste("Model: (", p, ",", d, ",", q, ") (", P, ",", D, ",", Q, ") [", S, "]", sep=""), adj=0,
cex.main=0.75)
 }

 std.resid <- mod.fit.obj$fit$residuals/sqrt(mod.fit.obj$fit$sigma2)
 hist(std.resid, main="Histogram of Standardized Residuals", xlab="Standardized
Residuals", freq=FALSE)
 curve(expr=dnorm(x, mean=mean(std.resid), sd=sd(std.resid)), col="red", add=TRUE)
}

pformosu<-read.table(file =
"C:/Users/inna/Desktop/DepaulClasses/ApplyMathClasses/Time_Series/Final
Project/pformosu.txt")

row =3726
col= 16
pformosu_dat = numeric(col*row)
count = 1
for(i in 1:row)
{
 for(j in 1:col)
 {
  pformosu_dat[count]=pformosu[i,j]
  count = count + 1
 }
}
```

```
ts_p<-ts(pformosu_dat)
dev.new()
par(mfrow=c(2,1))
tsplot(ts_p)
boxplot(ts_p)$out
#################################################################
num_of_outliers = 0
repeat
{
 outliers <- boxplot(ts_p, plot=FALSE)$out
 v<-(which(ts_p %in% outliers))
 if(length(v)== 0)
  break
 else
 {
  ts_p<-ts_p[-v]
  num_of_outliers = num_of_outliers + length(v)
 }
}
print(paste("Number Of Outliers removed = ", num_of_outliers))
print(paste("Length Of Time Series",length(ts_p)))
dev.new()
par(mfrow=c(2,1))
tsplot(ts_p,xlab="Number Of Observations",main="Circadian rhythms in the  last  week Of
data")
boxplot(ts_p)$out

################################################################
frequency <- 30
one_day<- 30*24
one_week<- 24*30*7
ts_pp  <- ts(ts_p[(length(ts_p)-(one_week) + 1 ):(length(ts_p))],frequency = frequency)
dev.new()
tsplot(ts_pp,xlab="Number Of Observations -> 7 Days -> 168 Hours",
    main="Circadian rhythms in the  last  week Of data")
x<-as.numeric(ts_pp)

###############################################################
dev.new()
tsplot(ts_pp,  ylab=expression(x[t]),
    xlab="hours", main="Circadian rhythms in the  last  week Of data")

dev.new()
acf2(x, max.lag = length(ts_pp) - 1,main = "ACF & PACF Of Original ts")
dev.new()
acf2(x, max.lag = 100,main = "ACF & PACF Of Original ts")
```

```
##################################################################
# Plot of (1-B^30)*x_t
dev.new()
tsplot(diff(ts_pp, lag=frequency, differences=1), ylab=expression((1-B^30)*x[t]),
    xlab="hours", main=expression(paste("Plot of ", (1-B^30)*x[t])))

# ACF indicated cut of after 330 lag that sugests that ARIMA Q=1,
# PACF exponencially decay that sugests ARIMA P=0
# Diff =1 S=18 (1-B^30)*x_t
dev.new()
acf2(diff(x, lag=frequency, differences=1), max.lag=frequency*5,
   main=expression(paste("Est. ACF & PACF for ", (1-B^30)*x[t])))


################################################################
# Plot of (1-B)(1-B^30)*x_t
dev.new()
tsplot(diff(diff(ts_pp, lag=frequency, differences=1)),
    ylab=expression((1-B)(1-B^30)*x[t]),
    xlab="Hours", main=expression(paste("Plot of ",
                        (1-B)(1-B^30)*x[t])))

# ACF and PACF of (1-B)(1-B^30)*x_t
dev.new()
acf2(diff(  diff(x[1:length(x)], lag=frequency, differences=1)),
   max.lag=frequency,
   main=expression(paste("Est. ACF & PACF for ", (1-B)(1-B^30)*x[t])))
##################################################################
#########
# First Estimate ARIMA(p=0,d=1,q=0,P=0,D=1,Q=1,S=30) #######

dev.new()
mod.fit1<- sarima(ts_pp,p=0,d=1,q=0,P=0,D=1,Q=1,S=30)
mod.fit1

dev.new()
par(mfrow=c(2,1))
pacf(mod.fit1$fit$residuals,main = " PACF Of residuals")
title(paste("Model: (",
        p = 0, ",",
        d = 1, ",",
        q = 0, ")(",
        P = 0, ",",
        D = 1, ",",
        Q = 1, ")[30]",
        sep=""),adj=0,cex.main=0.75)
std.resid1 <- mod.fit1$fit$residuals / sqrt(mod.fit1$fit$sigma2)
hist(std.resid1,main = " Histogram Of Standirdized Residuals",
```

```
   xlab='Standardized Residuals',
   freq = FALSE
)
### # First Estimate ARIMA(p=0,d=1,q=1,P=0,D=1,Q=1,S=30) #######

dev.new()
mod.fit1<- sarima(ts__pp,p=0,d=1,q=1,P=0,D=1,Q=1,S=30)
mod.fit1

dev.new()
par(mfrow=c(2,1))
pacf(mod.fit1$fit$residuals,main = " PACF Of residuals")
title(paste("Model: (",
        p = 0, ",",
        d = 1, ",",
        q = 1, ")(",
        P = 0, ",",
        D = 1, ",",
        Q = 1, ")[30]",
        sep=""),adj=0,cex.main=0.75)
std.resid1 <- mod.fit1$fit$residuals / sqrt(mod.fit1$fit$sigma2)
hist(std.resid1,main = " Histogram Of Standardized Residuals",
   xlab='Standardized Residuals',
   freq = FALSE
)
### ARIMA(p=1,d=1,q=1,P=0,D=1,Q=1,S=30)
dev.new()
mod.fit1<- sarima(ts__pp,p=1,d=1,q=1,P=0,D=1,Q=1,S=30)
mod.fit1

dev.new()
par(mfrow=c(2,1))
pacf(mod.fit1$fit$residuals,main = " PACF Of residuals")
title(paste("Model: (",
        p = 1, ",",
        d = 1, ",",
        q = 1, ")(",
        P = 0, ",",
        D = 1, ",",
        Q = 1, ")[30]",
        sep=""),adj=0,cex.main=0.75)
std.resid1 <- mod.fit1$fit$residuals / sqrt(mod.fit1$fit$sigma2)
hist(std.resid1,main = " Histogram Of Standirdized Residuals",
   xlab='Standardized Residuals',
   freq = FALSE
)
#### ARIMA(p=0,d=1,q=2,P=0,D=1,Q=1,S=30)
dev.new()
```

```
mod.fit1<- sarima(ts_pp,p=0,d=1,q=2,P=0,D=1,Q=1,S=30)
mod.fit1

dev.new()
par(mfrow=c(2,1))
pacf(mod.fit1$fit$residuals,main = " PACF Of residuals")
title(paste("Model: (",
        p = 0, ",",
        d = 1, ",",
        q = 2, ")(",
        P = 0, ",",
        D = 1, ",",
        Q = 1, ")[30]",
        sep=""),adj=0,cex.main=0.75)
std.resid1 <- mod.fit1$fit$residuals / sqrt(mod.fit1$fit$sigma2)
hist(std.resid1,main = " Histogram Of Standardized Residuals",
   xlab='Standardized Residuals',
   freq = FALSE
)
#### ARIMA(p=0,d=1,q=3,P=0,D=1,Q=1,S=30)
dev.new()
mod.fit1<- sarima(ts_pp,p=0,d=1,q=3,P=0,D=1,Q=1,S=30)
mod.fit1

dev.new()
par(mfrow=c(2,1))
pacf(mod.fit1$fit$residuals,main = " PACF Of residuals")
title(paste("Model: (",
        p = 0, ",",
        d = 1, ",",
        q = 3, ")(",
        P = 0, ",",
        D = 1, ",",
        Q = 1, ")[30]",
        sep=""),adj=0,cex.main=0.75)
std.resid1 <- mod.fit1$fit$residuals / sqrt(mod.fit1$fit$sigma2)
hist(std.resid1,main = " Histogram Of Standirdized Residuals",
   xlab='Standardized Residuals',
   freq = FALSE
)
#### ARIMA(p=1,d=1,q=0,P=0,D=1,Q=1,S=30)
dev.new()
mod.fit1<- sarima(ts_pp,p=1,d=1,q=0,P=0,D=1,Q=1,S=30)
mod.fit1

dev.new()
par(mfrow=c(2,1))
pacf(mod.fit1$fit$residuals,main = " PACF Of residuals")
```

```
title(paste("Model: (",
        p = 1, ",",
        d = 1, ",",
        q = 0, ")(",
        P = 0, ",",
        D = 1, ",",
        Q = 1, ")[30]",
        sep=""),adj=0,cex.main=0.75)
std.resid1 <- mod.fit1$fit$residuals / sqrt(mod.fit1$fit$sigma2)
hist(std.resid1,main = " Histogram Of Standirdized Residuals",
   xlab='Standardized Residuals',
   freq = FALSE
)
#### ARIMA(p=2,d=1,q=0,P=0,D=1,Q=1,S=30)
dev.new()
mod.fit1<- sarima(ts__pp,p=2,d=1,q=0,P=0,D=1,Q=1,S=30)
mod.fit1

dev.new()
par(mfrow=c(2,1))
pacf(mod.fit1$fit$residuals,main = " PACF Of residuals")
title(paste("Model: (",
        p = 2, ",",
        d = 1, ",",
        q = 0, ")(",
        P = 0, ",",
        D = 1, ",",
        Q = 1, ")[30]",
        sep=""),adj=0,cex.main=0.75)
std.resid1 <- mod.fit1$fit$residuals / sqrt(mod.fit1$fit$sigma2)
hist(std.resid1,main = " Histogram Of Standardized Residuals",
   xlab='Standardized Residuals',
   freq = FALSE
)


#### ARIMA(p=2,d=1,q=2,P=0,D=1,Q=1,S=30)
dev.new()
mod.fit1<- sarima(ts__pp,p=2,d=1,q=2,P=0,D=1,Q=1,S=30)
mod.fit1

dev.new()
par(mfrow=c(2,1))
pacf(mod.fit1$fit$residuals,main = " PACF Of residuals")
title(paste("Model: (",
        p = 2, ",",
        d = 1, ",",
        q = 2, ")(",
```

```
        P = 0, ",",
        D = 1, ",",
        Q = 1, ")[30]",
        sep=""),adj=0,cex.main=0.75)
std.resid1 <- mod.fit1$fit$residuals / sqrt(mod.fit1$fit$sigma2)
hist(std.resid1,main = " Histogram Of Standardized Residuals",
   xlab='Standardized Residuals',
   freq = FALSE
)

#### ARIMA(p=2,d=1,q=3,P=0,D=1,Q=1,S=30)
dev.new()
mod.fit1<- sarima(ts_pp,p=2,d=1,q=3,P=0,D=1,Q=1,S=30)
mod.fit1

dev.new()
par(mfrow=c(2,1))
pacf(mod.fit1$fit$residuals,main = " PACF Of residuals")
title(paste("Model: (",
        p = 3, ",",
        d = 1, ",",
        q = 2, ")(",
        P = 0, ",",
        D = 1, ",",
        Q = 1, ")[30]",
        sep=""),adj=0,cex.main=0.75)
std.resid1 <- mod.fit1$fit$residuals / sqrt(mod.fit1$fit$sigma2)
hist(std.resid1,main = " Histogram Of Standardized Residuals",
   xlab='Standardized Residuals',
   freq = FALSE
)
######################### Forecasting #####################

dev.new()
mod.fit.110.011 <- sarima(x, p=2,d=1,q=2,P=0,D=1,Q=1,S=30)
mod.fit.110.011
examine.mod(mod.fit.110.011, 2,1,2, 0,1,1, 30)

ahead_1_hours <-30
dev.new()
fore.mod <- sarima.for(ts_pp, n.ahead=ahead_1_hours,
     p=2, d=1, q=2, P=0, D=1, Q=1, S=30, plot.all=TRUE)

dev.new()
ts_pp_1_hour <- c(ts_pp[(length(ts_pp)-12*ahead_1_hours + 1):(length(ts_pp))],
            fore.mod$pred[1:length(fore.mod$pred)])
tsplot(ts_pp_1_hour,
      ylab="Data for 1 Hour ahead",
```

```
     xlab="minutes", main=expression("360 minutes + 60 minutes ahead"))
###########################################################
ahead_4_hours <- 4*30
dev.new()
fore.mod <- sarima.for(ts_pp, n.ahead=ahead_4_hours,
     p=2, d=1, q=2, P=0, D=1, Q=1, S=30, plot.all=TRUE)

dev.new()
ts_pp_4_hour <- c(ts_pp[(length(ts_pp)-6*ahead_4_hours + 1):(length(ts_pp))],
           fore.mod$pred[1:length(fore.mod$pred)])
tsplot(ts_pp_4_hour,
    ylab="Data for 4 Hours ahead",
    xlab="minutes", main=expression("1 day + 2hours ahead"))
#########################################################
ahead_24_hours <- 24*30
dev.new()
fore.mod <- sarima.for(ts_pp, n.ahead=ahead_24_hours,
              p=2, d=1, q=2, P=0, D=1, Q=1, S=30, plot.all=TRUE)

dev.new()
ts_pp_24_hour <- c(ts_pp[(length(ts_pp)-2*ahead_24_hours + 1):(length(ts_pp))],
           fore.mod$pred[1:length(fore.mod$pred)])
tsplot(ts_pp_24_hour,
    ylab="Data for 24 Hours ahead",
    xlab="minutes", main=expression("4 day + 24 hours ahead"))
########################################################
ahead_24_hours <- 24*30
dev.new()
fore.mod <- sarima.for(ts_pp, n.ahead=ahead_24_hours, p=2, d=1, q=2, P=0, D=1, Q=1,
S=30, plot.all=TRUE)

pred.mod <- ts(x - mod.fit.110.011$fit$residuals)

dev.new()
tsplot(ts_pp, ylab="pred", xlab="min", type="o", main="Predicted Model fitting with
Data")
lines(pred.mod, col="red", type="o", pch=17)
legend("topleft", legend=c("Observed", "Forecast"), lty=c("solid", "solid"), col=c("black",
"red"),
    pch=c(1, 17), bty="n")
```