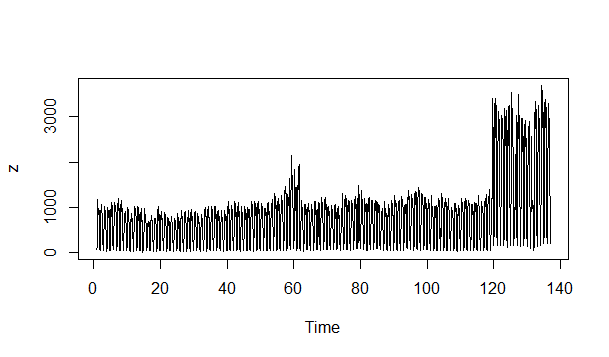
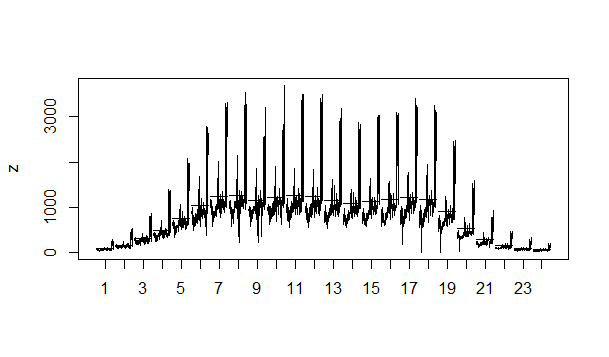
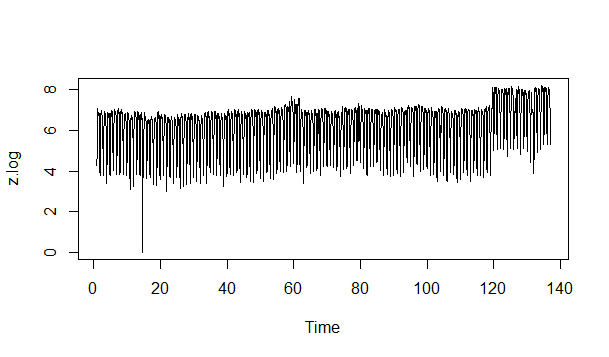
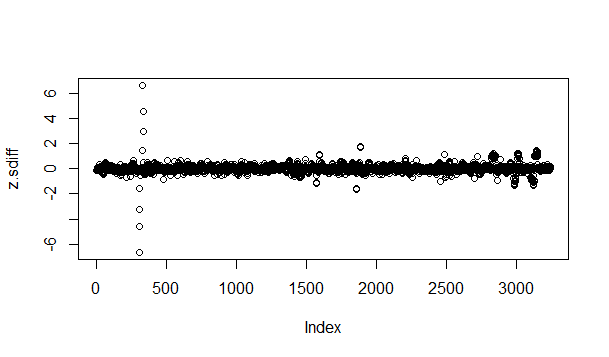
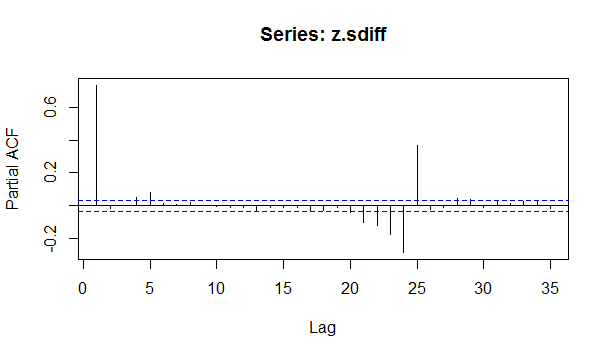
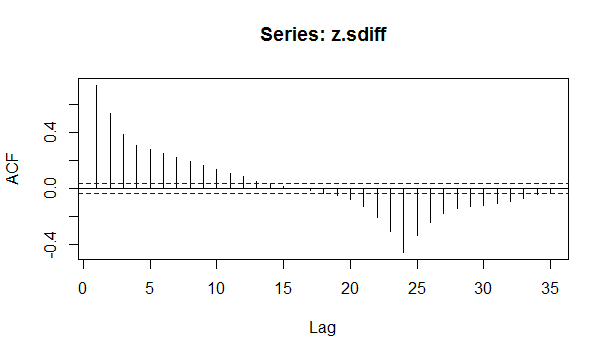
ARIMA(29,0,0)(2,0,0)24 model for Box-Cox transformed and seasonally conditioned data

1. The time plot shows sudden changes like at 120th day where it took a sharp jump, a smaller peak is at 60th day. 
2. The seasonal subseries plot also clearly indicated the presence of seasonality with the frequency of the 24 hours, which is also evident from the way data is indexed.
3. The time series is also clearly need some normalizing and seasonal differences. Taking the Box-Cox transformation with lambda=0 or rather taking log of the data we get another time series which we call z.log, with the time plot:
4. Next we take the seasonal difference, i.e. subtracting every data entry with the one taken 24hours before to get another time series which we name z.sdiff, with time plot:
5. This time series is now stationary enough for us to consider an Arima model. To explore the Arima model we look at the ACF and PACF plots:
6. As we can see the ACF is exponentially decreasing and it has a sinusoidal tendency, and PACF has a significant spike at lag 29, but none beyond lag p. This suggests an ARIMA(29,0,0) or AR(29) model.
7. An AR(29) model was generated using the R command

z.fitarima29<-Arima(z.sdiff, order=c(29L,0L,0L)) with following information:

Series: z.sdiff

ARIMA(29,0,0) with non-zero mean

Coefficients:

ar1 ar2 ar3 ar4 ar5 ar6 ar7 ar8 ar9

0.7590 -0.0228 -0.0592 0.0182 0.0806 -0.0044 0.0025 0.0232 -0.0048

s.e. 0.0176 0.0221 0.0221 0.0221 0.0210 0.0188 0.0188 0.0188 0.0188

ar10 ar11 ar12 ar13 ar14 ar15 ar16 ar17 ar18

-0.0031 -0.0024 0.0252 -0.0265 0.0008 0.0059 0.0184 -0.0088 -0.0055

s.e. 0.0188 0.0188 0.0189 0.0188 0.0188 0.0188 0.0188 0.0188 0.0188

ar19 ar20 ar21 ar22 ar23 ar24 ar25 ar26 ar27

0.0251 -0.0001 -0.0246 0.0162 0.0211 -0.5239 0.3860 -0.0122 -0.0464

s.e. 0.0188 0.0188 0.0188 0.0188 0.0188 0.0188 0.0209 0.0221 0.0221

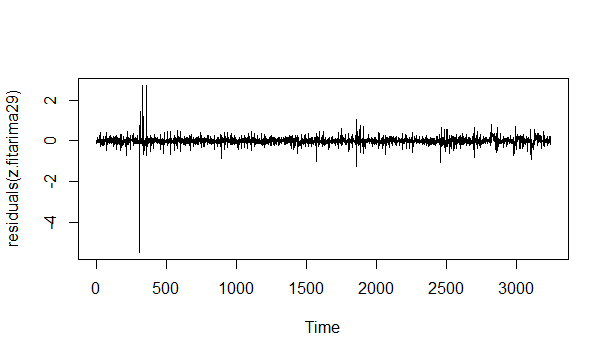
ar28 ar29 intercept

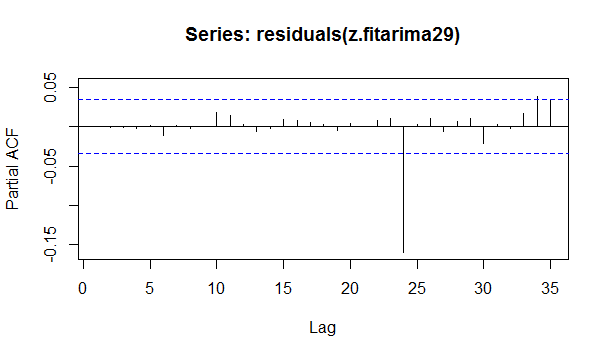
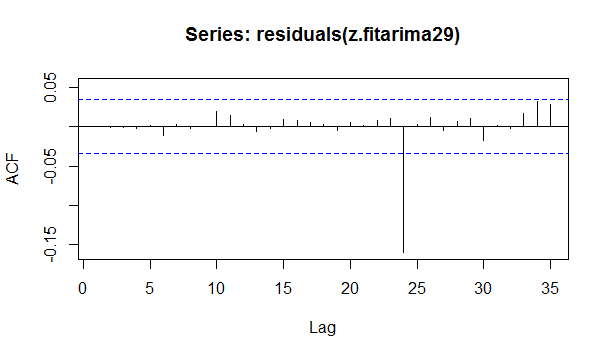
0.0179 0.0401 0.0090

s.e. 0.0221 0.0175 0.0115

sigma^2 estimated as 0.03954: log likelihood=632.06

AIC=-1202.12 AICc=-1201.51 BIC=-1013.54

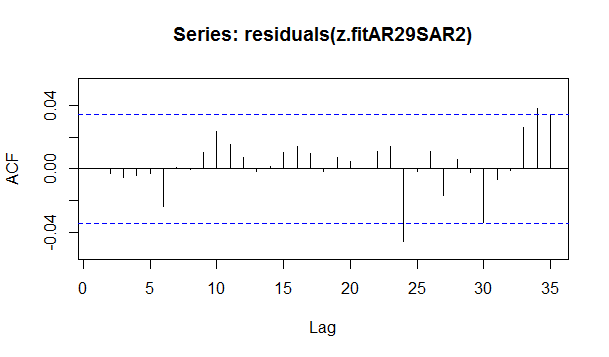
1. The AR(29) above has the following time plot for its residuals:
2. The ACF and PACF models for residuals(z.fitarima29) are as follows:

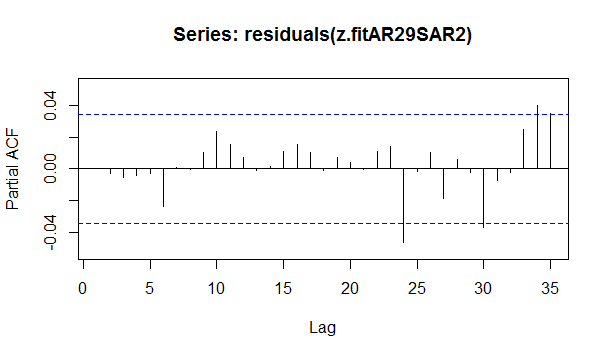


1. These graphs indicate that the model should include a seasonal autoregressive term.
2. An ARIMA(29,0,0)(2,0,0)24 or AR(29)SAR(2)24 model was explored and found to have a lesser AIC value, with following data:

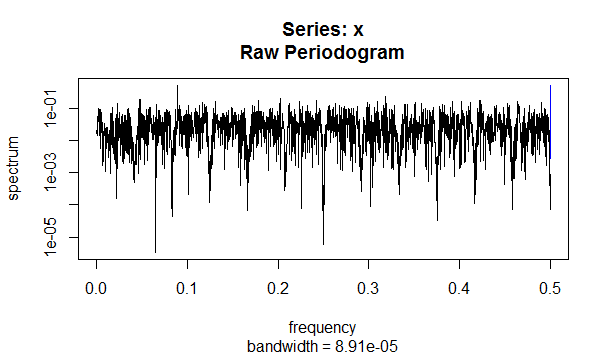
|  |
| --- |
| arima(x = z.sdiff, order = c(29L, 0L, 0L), seasonal = list(order = c(2L, 0L,  0L), period = 24))  Coefficients:  ar1 ar2 ar3 ar4 ar5 ar6 ar7 ar8 ar9  0.7699 -0.0113 -0.0619 0.0187 0.0884 -0.0096 0.0077 0.0190 0.0077  s.e. 0.0176 0.0222 0.0221 0.0221 0.0205 0.0174 0.0175 0.0175 0.0175  ar10 ar11 ar12 ar13 ar14 ar15 ar16 ar17 ar18  -0.0110 0.0000 0.0117 -0.0068 -0.0008 0.0055 0.0074 0.0011 -0.0155  s.e. 0.0175 0.0174 0.0175 0.0175 0.0174 0.0174 0.0175 0.0175 0.0174  ar19 ar20 ar21 ar22 ar23 ar24 ar25 ar26 ar27  0.0394 -0.0044 -0.0190 0.0099 0.0232 -0.6128 0.4680 -0.0067 -0.0621  s.e. 0.0174 0.0175 0.0175 0.0175 0.0175 0.0246 0.0246 0.0222 0.0222  ar28 ar29 sar1 sar2 intercept  0.0230 0.0597 -0.1459 -0.3834 0.0098  s.e. 0.0221 0.0177 0.0242 0.0187 0.0081  sigma^2 estimated as 0.0337: log likelihood = 886.38, aic = -1706.75 |
|  |
| |  | | --- | |  | |

1. Let us look at the ACF and PACF graphs of the new residuals:





1. Although there is spike at lag 24 in both of these graphs, but it is now negligible. We do have a white noise now:



1. Thus we settle for the ARIMA(29,0,0)(2,0,0)24 or AR(29)SAR(2)24 model for the seasonal differences for log(z), where z is the given time series. Thus our model looks like:



1. In this model we take the intercept c, the non-seasonal coefficients ari and seasonal coefficients sarj from the following table:

ar1 7.699166e-01

ar2 -1.128915e-02

ar3 -6.193436e-02

ar4 1.871771e-02

ar5 8.835534e-02

ar6 -9.550217e-03

ar7 7.700764e-03

ar8 1.902851e-02

ar9 7.744997e-03

ar10 -1.100078e-02

ar11 -2.751673e-05

ar12 1.166803e-02

ar13 -6.778000e-03

ar14 -7.594131e-04

ar15 5.480208e-03

ar16 7.403947e-03

ar17 1.103122e-03

ar18 -1.550102e-02

ar19 3.944067e-02

ar20 -4.437393e-03

ar21 -1.903567e-02

ar22 9.933290e-03

ar23 2.323067e-02

ar24 -6.128026e-01

ar25 4.679633e-01

ar26 -6.742042e-03

ar27 -6.208018e-02

ar28 2.298195e-02

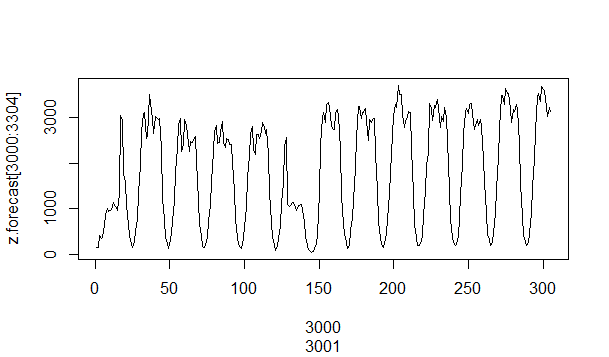
ar29 5.970577e-02

sar1 -1.458680e-01

sar2 -3.834068e-01

c 9.821938e-03

1. This model is forecasting the right end of the following graph:



1. A few predictions based on the sample from the given data are as follows where the z[n+1] is predicted using the sample z[1:n]:

Date Hour Original Predicted Relative Error

234 2014-05-10 17 627 700 -0.11782066

459 2014-05-20 2 238 254 -0.06760526

623 2014-05-26 22 51 42 0.17070400

835 2014-06-04 18 664 675 -0.01661476

1042 2014-06-13 9 871 1007 -0.15690965

1223 2014-06-20 22 68 71 -0.05174137

1536 2014-07-03 23 54 35 0.33589113

1827 2014-07-16 2 234 223 0.04625502

2122 2014-07-28 9 1103 966 0.12397450

2333 2014-08-06 4 713 762 -0.06875276

2530 2014-08-14 9 1129 1286 -0.13912457

2984 2014-09-02 7 3546 3288 0.07252162

3100 2014-09-07 3 1019 947 0.06979563

3260 2014-09-13 19 1585 1263 0.20264578