Time Series Analysis - Energy Consumption

Bach Quang Minh - MS Candidate - John Von Neumann Institute

April 8, 2018

Contents

1   Data description

The data set shows different measures of a particular household, including temperature, humidity conditions, weather conditions and energy consumption. The data set is at 10 min for about 4.5 months, containing 19736 data points and 30 attributes in total. Therefore, no time-interval adjustments are needed. The outcome of interest is the Appliances column, which contains the energy consumption of the house. For the analysis task, a sample of 10 days from 1/12/2016 00:00 to 1/21/2016 23:50 is extracted from the data, keeping only the "date" and the "Appliances" columns. The sample contains 1440 data points which is used for the fitting process. The forecast is made for the next 2 hours, projecting 12 data points ahead.

The data can be downloaded here.

The following figure gives an overall look of the data set.

date	index	Appliances	lights	T1	RH_1	T2	RH_2	T3
1/11/2016 17:00	1	60	30	19.89	47.59667	19.2	44.79	19.79
1/11/2016 17:10	2	60	30	19.89	46.69333	19.2	44.7225	19.79
1/11/2016 17:20	3	50	30	19.89	46.3	19.2	44.62667	19.79
1/11/2016 17:30	4	50	40	19.89	46.06667	19.2	44.59	19.79
1/11/2016 17:40	5	60	40	19.89	46.33333	19.2	44.53	19.79
1/11/2016 17:50	6	50	40	19.89	46.02667	19.2	44.5	19.79
1/11/2016 18:00	7	60	50	19.89	45.76667	19.2	44.5	19.79
1/11/2016 18:10	8	60	50	19.85667	45.56	19.2	44.5	19.73
1/11/2016 18:20	9	60	40	19.79	45.5975	19.2	44.43333	19.73
1/11/2016 18:30	10	70	40	19.85667	46.09	19.23	44.4	19.79
1/11/2016 18:40	11	230	70	19.92667	45.86333	19.35667	44.4	19.79
1/11/2016 18:50	12	580	60	20.06667	46.39667	19.42667	44.4	19.79
1/11/2016 19:00	13	430	50	20.13333	48	19.56667	44.4	19.89
1/11/2016 19:10	14	250	40	20.26	52.72667	19.73	45.1	19.89
1/11/2016 19:20	15	100	10	20.42667	55.89333	19.85667	45.83333	20.03333
1/11/2016 19:30	16	100	10	20.56667	53.89333	20.03333	46.75667	20.1

Figure 1: This is a snapshot of the data. The figure only shows 9 out of 30 attributes.

2 Data exploration
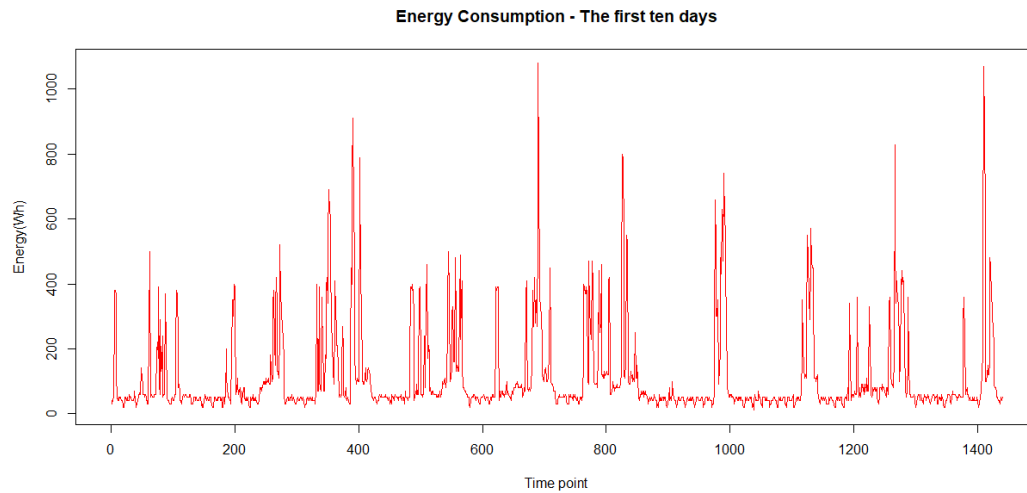
First, the "Appliances" is graphed as a time series.

Figure 2: First look at the time series.

As can be seen, there is a high volatility in the energy consumption over time. There are some points where the energy level gets very high, meanwhile some durations see significantly low levels of energy. This could be due to the family members' activities. There is a slight pattern of seasonality. The following graph is made to investigate seasonality in the time series (Figure 3).
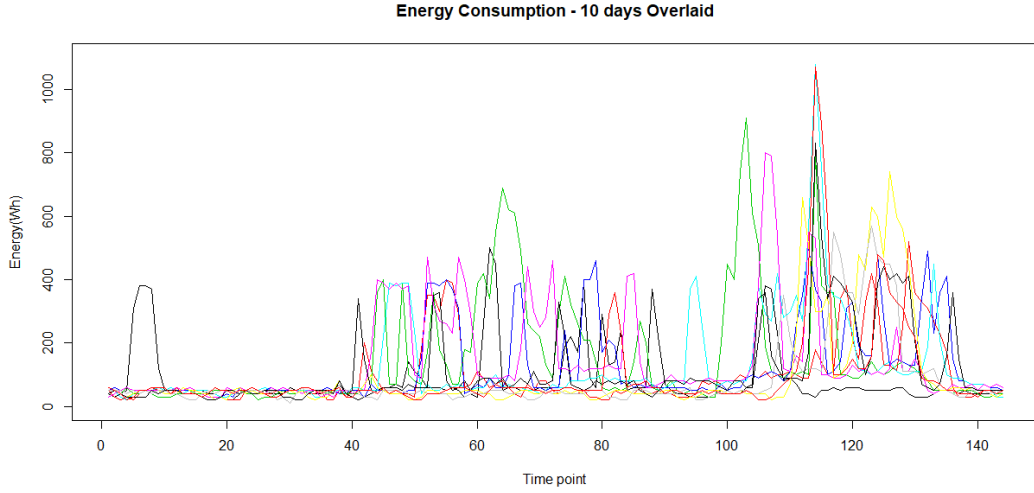
Figure 3: Each color represents one day's energy series.

Generally, a pattern in energy consumption can be seen from the graph. From the first 6 hours of a day, the consumption is quite low, fluctuating around 50Wh although an unusual event can be seen in day 1, creating a hump of 400Wh at around 1:30am (the black line). From 6:00am to around 3:30pm, a handful of days see high levels of energy consumption, meanwhile, some other days remain the same energy levels as those of the first 6 hours. After that energy levels back to around 50 Wh, then quickly jumps high, peaking at around 9:00pm. The consumption is quite high generally from 5:30pm to 11:00pm, then lowers back to around 50Wh.

To reduce the variability in the variance, the log10 transformation is applied to the data, producing a quite stationary-in-mean series (Figure 4).
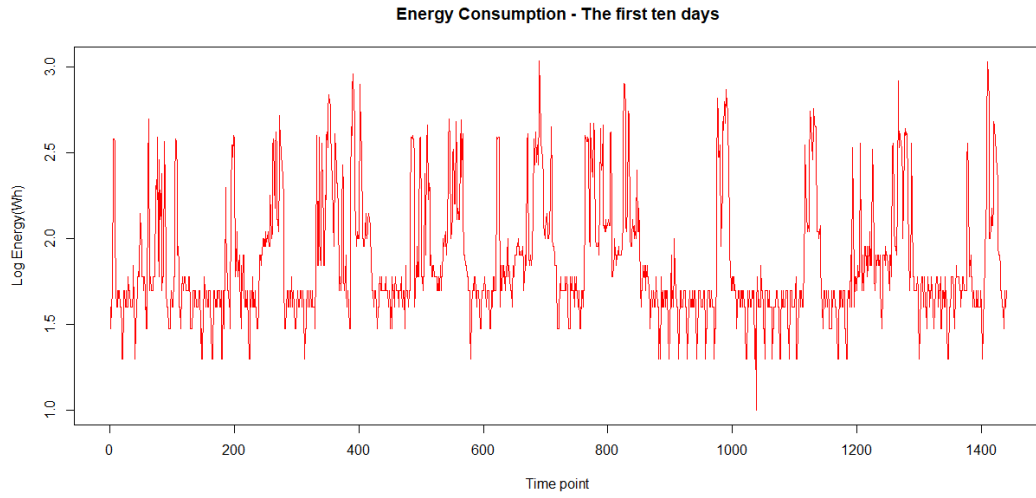
3

Figure 4: The time series in log base 10 scale. The series look quite stationary in mean and the variance is more stable

Further stationary analysis can be done by checking the acf and the pacf of the log series (Figure 5). Notice that the waving patterns in the acf plot, together with the fact that significant numbers of lags show high autocorrelation, give evidence against stationarity in the series, although the pacf show a quite good shape.
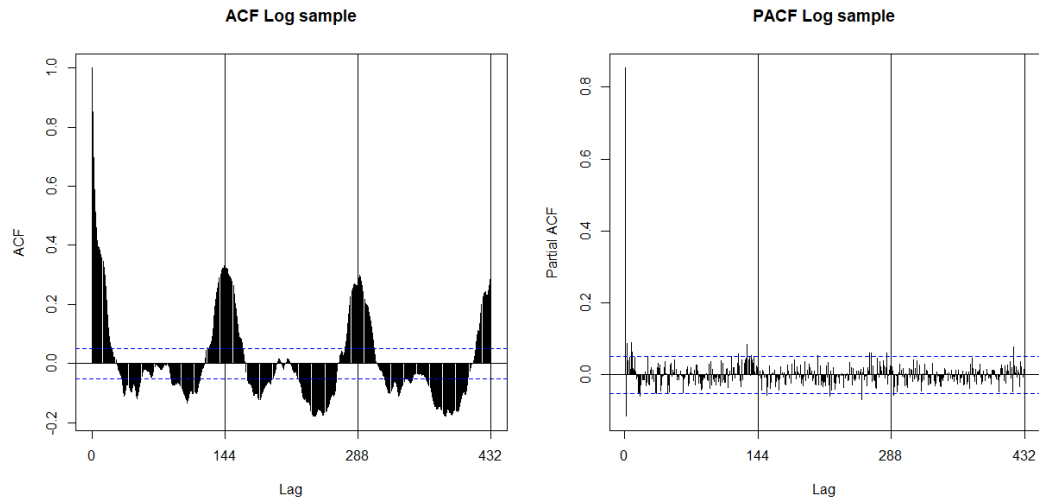

Figure 5: The acf plot does not show strong evidence of stationarity

4

The first order difference of the log series is taken at lag 1 in attempting to stationarize the series. Figure 6 shows that the difference series looks stationary in mean 0. However, the variance varies unpredictably over time. Again, acf and pacf plots are checked for further conclusion.
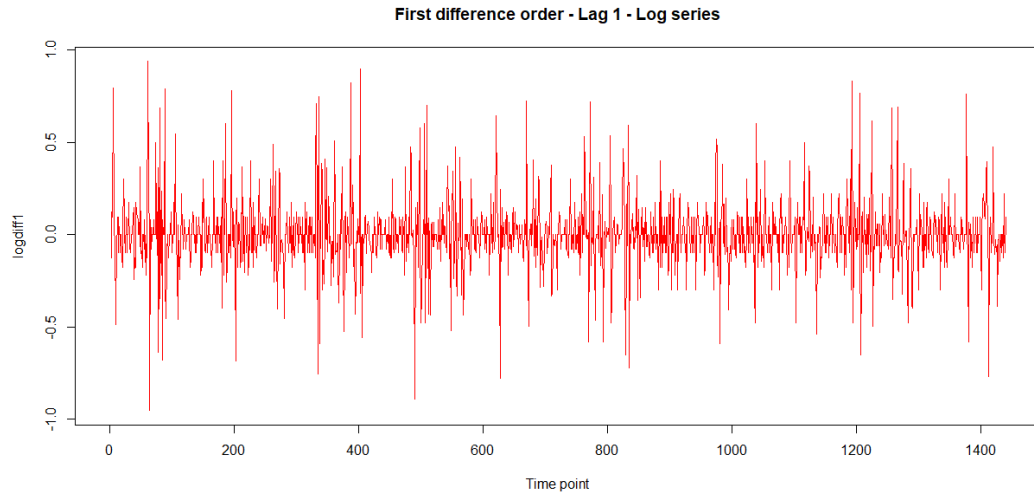

Figure 6: The difference series

Here, the acf and pacf seems to be in good shape. The acf spikes at the first few lags (1 to 3) and shows some autocorrelation at the first few lags, then quickly goes to zero afterwards. The pacf show signs of exponential decay and most pacf values are within the 95 confident interval. (Figure 7)
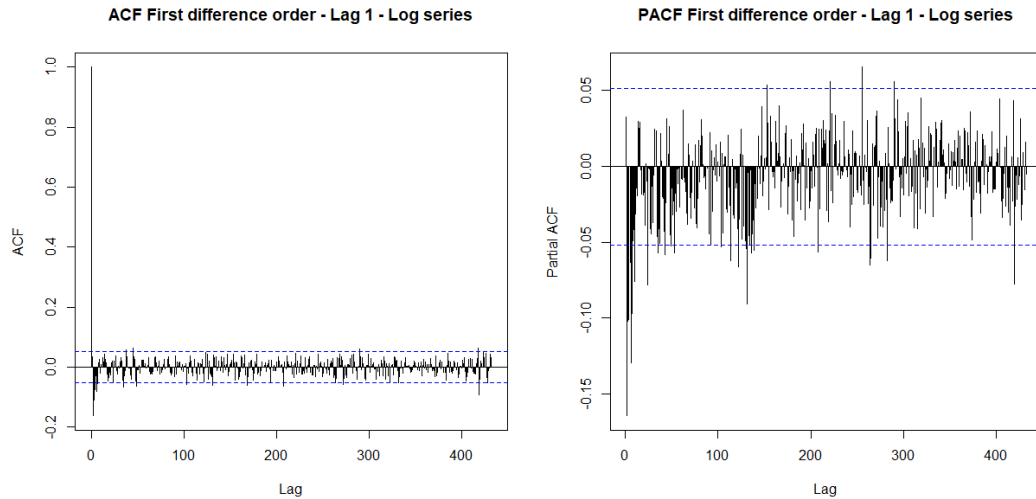
Figure 7: The acf and pacf plots of the difference series

To ensure stationarity, the Augmented Dickey-Fuller Test test is conducted. The null hypothesis is: 'The series is non-stationary'. The alternative hypothesis is: 'The series is stationary. Figure 8 shows that the p value is small enough (under 0.05) to reject the null hypothesis, accepting the alternative hypothesis.

```
          Augmented Dickey-Fuller Test

data:  logdiff1
Dickey-Fuller = -36.663, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

Figure 8: The p-value is smaller than the 0.05 α-level

After the first few steps of exploration, it could be concluded that analysis should be done on the difference series resulting from taking the first order difference of the log series at lag 1. Because the variance is now more stable

and the series is stationary in mean, therefore, is more suitable for model fitting than the original series.

3 Models

The general approach for applying models is:

- Fitting the models on a sample of 10 days

- Evaluating the goodness of fit of the models

- Evaluating the accuracy of the models using a test set of 12 data points ahead

Below is the list of models which will be applied:

1. AR(1)

2. AR(3)

3. MA(1)

4. MA(3)

5. ARMA(3,3)

6. ARIMA(1,1,1)(1,0,1)[144]

7. Exponential Smoothing ($\alpha = 0.9999339$)

8. Holt's linear ($\alpha = 1, \beta = 0.01253426$)

9. Holt Winter ($\alpha = 0.8, \beta = 0, \gamma = 1$)

10. Garch(1,1)-ARMA(1,1)

The choice for parameters will be explained in detailed later on. The models will be presented in three groups: models that require stationary, forecasting algorithms and models that can deal with changing variance over time.

3.1 Models that require stationary

Refer to Figure 7, the acf spikes at lag 1 to 3 and the pacf has an exponential decay shape like. This indicates that an MA(3) model could be a good fit for the first order difference series of the log sample at lag 1. However, other models such as AR(1), AR(3), MA(3) are also fitted. The reason for choosing lag 3 results from observing the acf and pacf, heuristics that energy consumption levels within 30 minutes are correlated and also for the sake of reducing computational cost, AR(144) could not be fitted. Figure 9 and Figure 10 show how the fitting looks like among the four models after undifferencing.
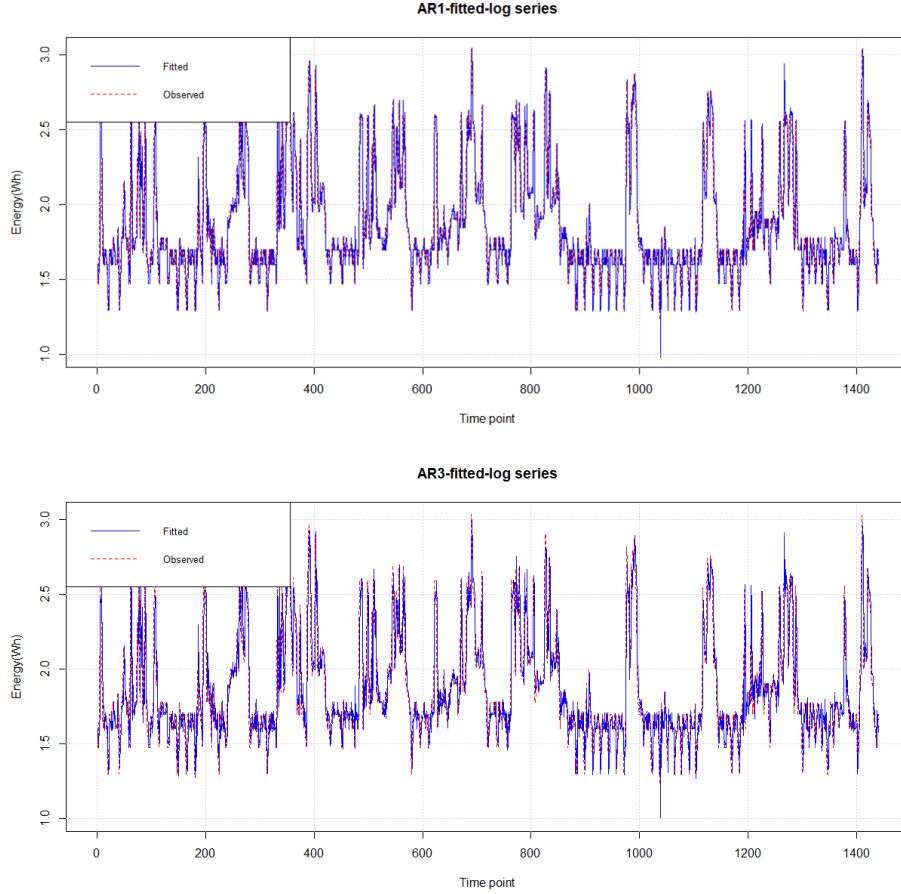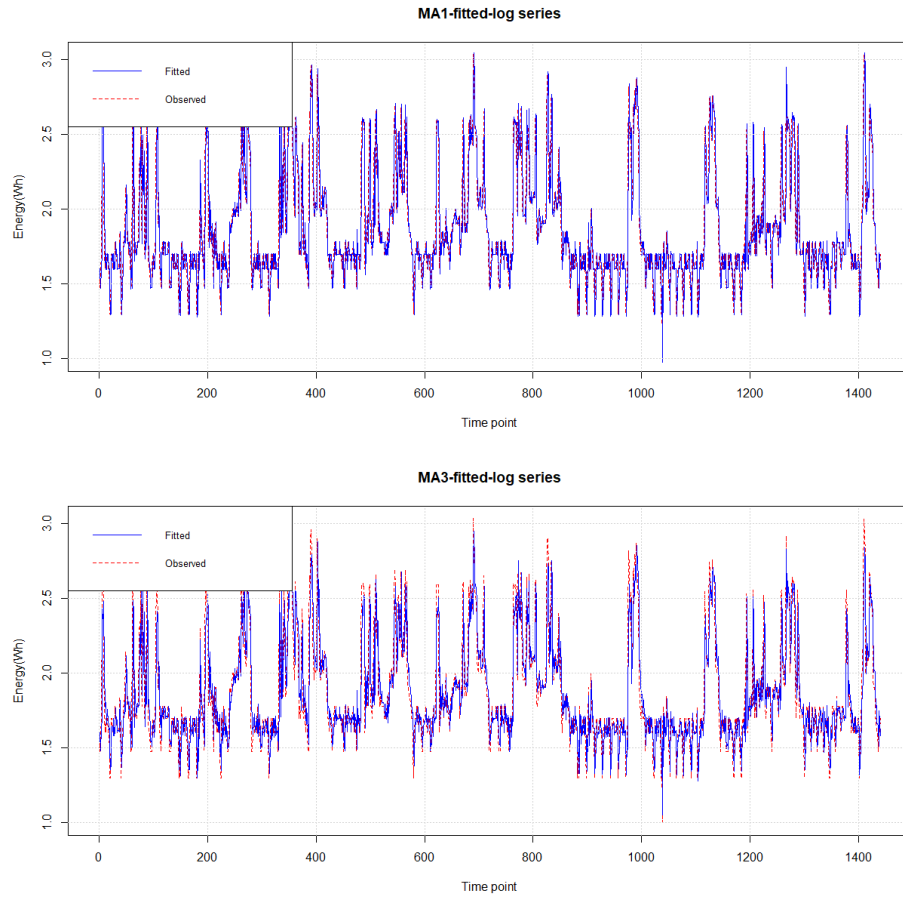

Figure 9: AR models

Figure 10: MA models

As can be seen, the fitting lines seem to be similar. It is hard to tell which model has better goodness of fit. The residual histograms, the QQ plots and the residual acf plots will then be compared.
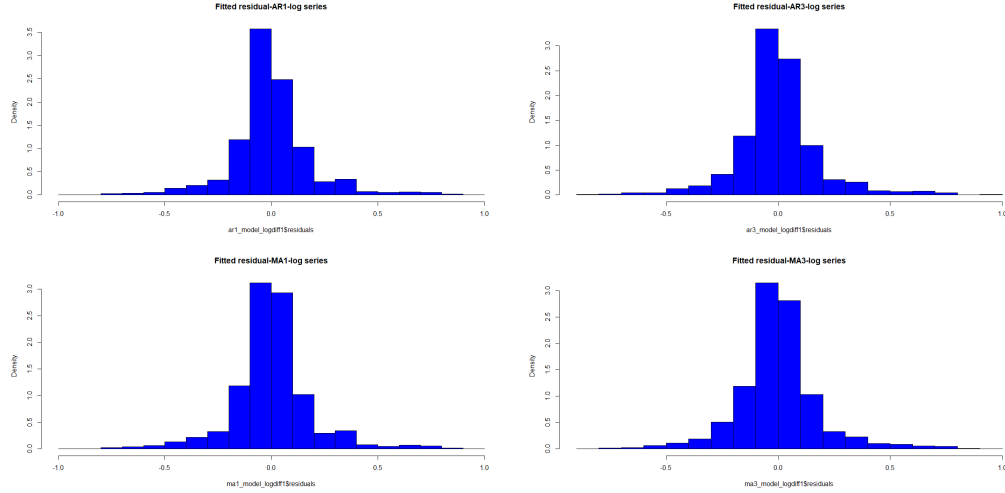
Figure 11: Residual histograms

As can be seen, the histograms from the MA models are a bit more symmetrical and normal than those of the AR models. This could be due to the characteristics described in Figure 7. Next QQ plots are observed.
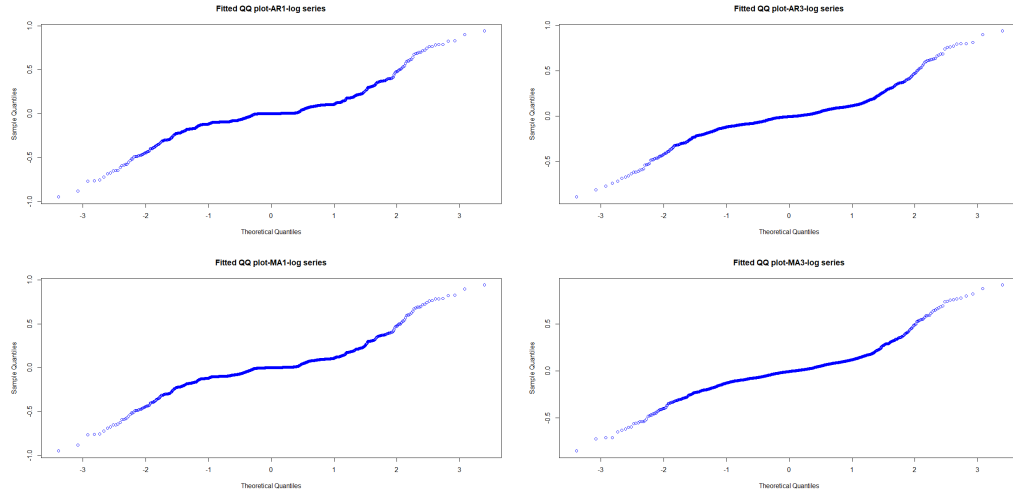

Figure 12: QQ plots

Overall, MA models also have better QQ plots. MA(3) seems to have the points form the 'straightest' line. To compare further goodness of fit, MAPE

and AIC of the models are computed. These two measures are chosen since MAPE does not depends on the range of outcome values, which could be different from samples to samples, and AIC penalize the models for being overfitting. The following tables show the results of the four models.

Table 1: MAPE and AIC

Model	MAPE	AIC
AR(1)	0.06328462	-664.2
AR(3)	0.06480294	-714.86
MA(1)	0.06356587	-664.92
MA(3)	0.06569768	-747.59

As can be seen, the models receive similar MAPE values. However, MA(3) gets the lowest AIC values, indicating it could be best to use MA(3) in comparison with the other three models. Next, forecasts are made for the models and the MAPE values are also computed given a test set of 12 data points on the log scale. Figure 13 shows the forecast of the models on the log scale. Figure 14 shows the forecast of the models on the original scale.
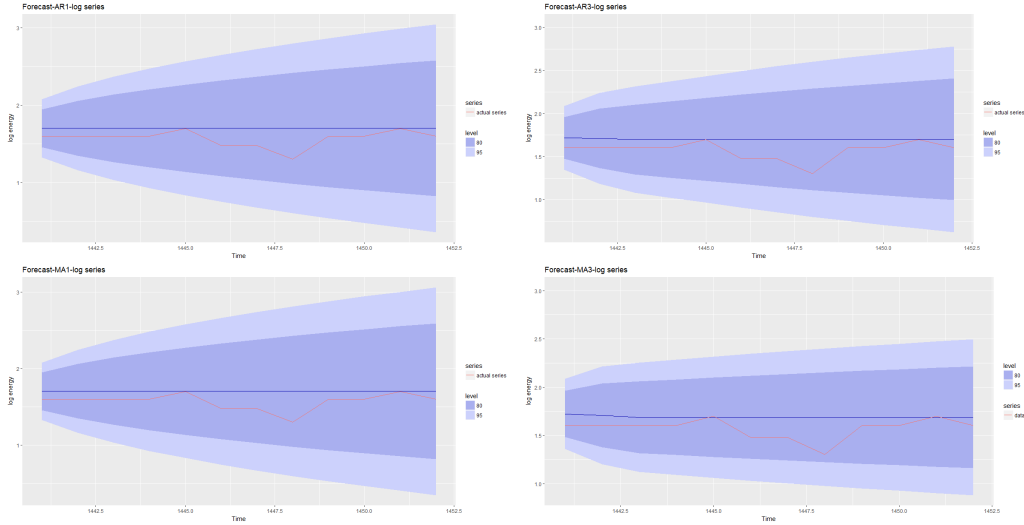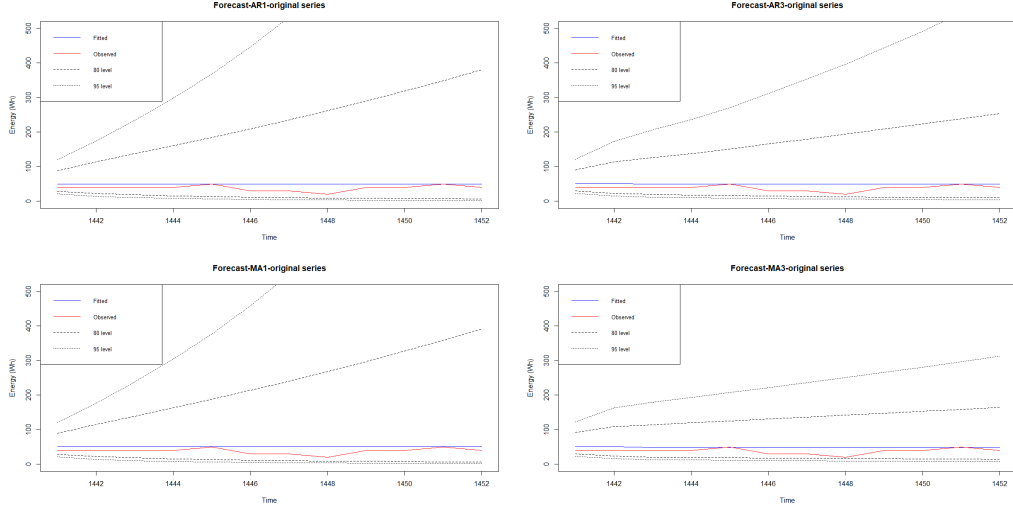

Figure 13: Forecast on log scale

Figure 14: Forecast on orginal scale

On both scales, MA(3) shows to have the smallest 95 confident interval. All the models give similar predictions. Table2 shows the MAPE computed on the given test set on the log scale. Again, MA(3) has the lowest test MAPE among the models. Other models receive similar test MAPE values.

Table 2: Test MAPE

Model	MAPE
AR(1)	0.08789718
AR(3)	0.08798882
MA(1)	0.08889993
MA(3)	0.08400015

In order to attempt decreasing the test MAPE and the AIC, ARIMA with a seasonal components is applied. For model specification, these values are set: $(p, d, q) = (1, 0, 1), (P, D, Q) = (1, 1, 1)$ and $s = 144$. This model is applied on the log series. The seasonal length is 144 because there are 144 10-minute intervals within a day. First, the log series is differenced by 1 season at first order, then 1 seasonal lag is applied to the series in both AR and MA parts. Next, the normal ARMA(1,1) process is applied. Figure 15
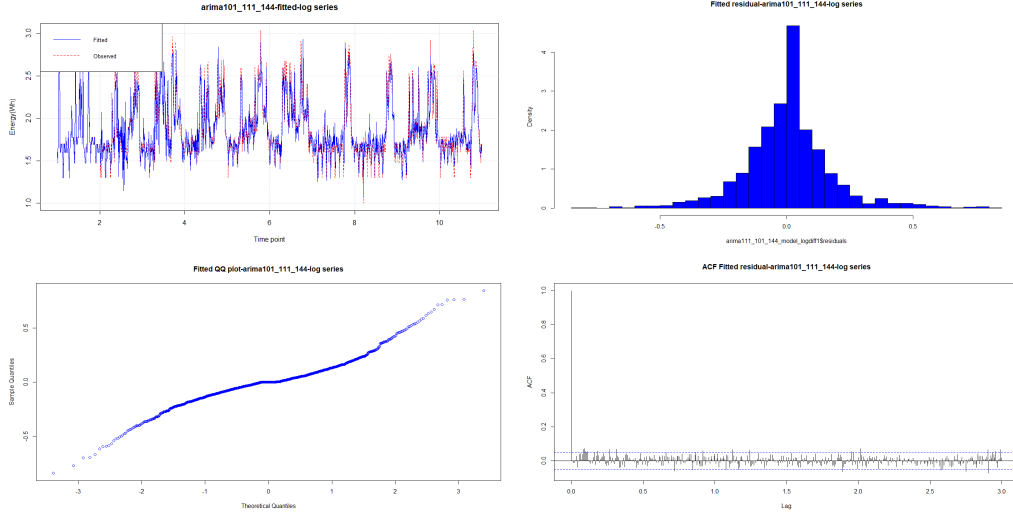
12

shows the process of fitting the model.

Figure 15: ARIMA(1,0,1)(1,1,1)[144].The upper left panel shows fit and the actual series. The upper right panel shows the histogram of the fitted residuals. The lower left panel shows the QQ plots and the lower right panel shows the acf of the residuals.

Both the histogram and the QQ plot are in quite nice normal shape. The acf also shows low autocorrelation in the residuals. For this model, the fitting MAPE is 0.0636094, which is quite low compared to previous models. However, the AIC of this model sees a higher value of -458.34 (refer back to Table 1 for comparison). This could be showing that the model is more overfitting compared to the others. In fact, the test MAPE of this model is 0.09180982, which is also higher than previous models (refer back to Table 2 for comparison). This observation indicates that this complex model is more overfitting. The forecasts in the log scale and the original scale is shown in Figure 16.
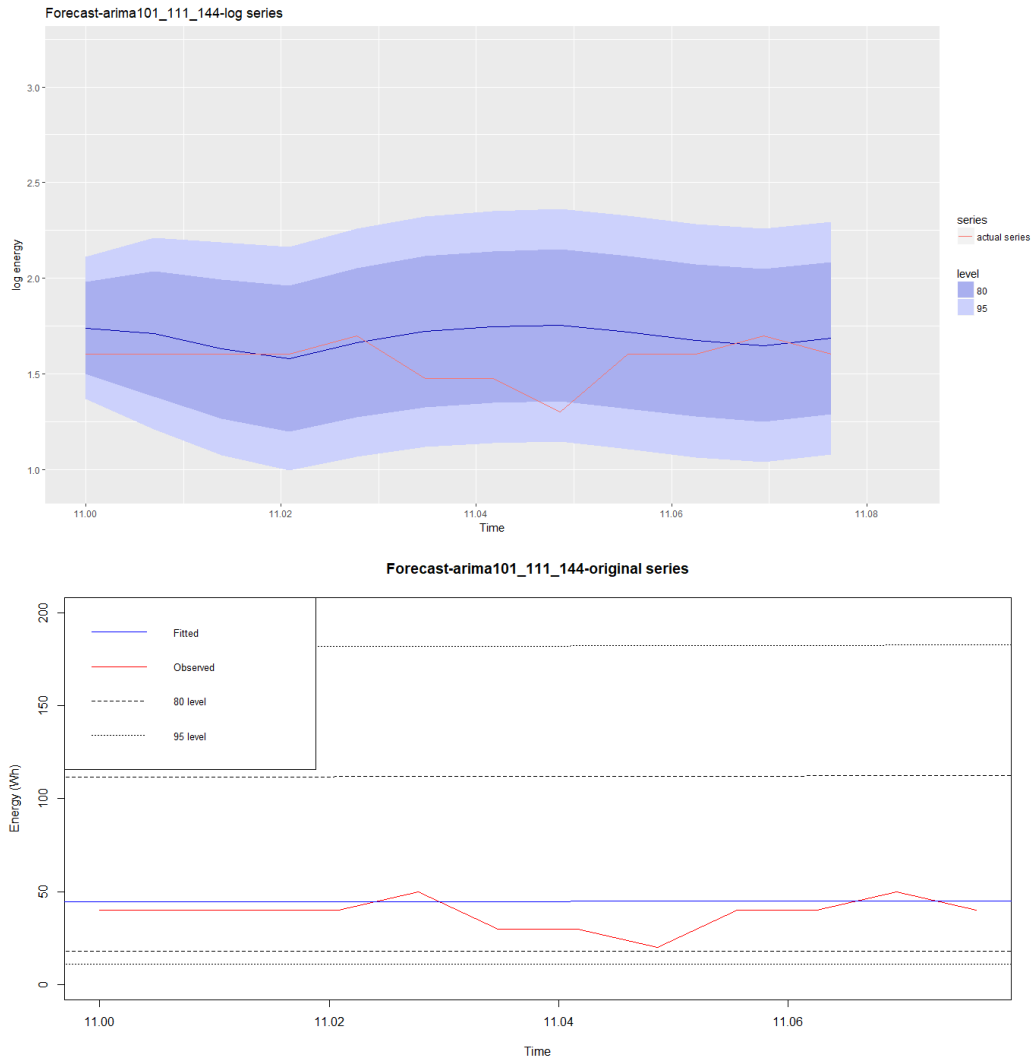
Figure 16: ARIMA(1,0,1)(1,1,1)[144] - forecast

Next, the function auto.arima() in R is applied to seek for the optimal ARIMA model for this log series. The function returns an ARIMA(3,0,3) model, which is also an ARIMA(3,3) model. This model receives the lowest AIC so far, hitting -803.36. However, the MAPE in both the fitting and the forecasting process are not among the lowest ones. Surprisingly, the test MAPE value is quite high compared to other models (0.1371812). Figure 17 shows all the common process for this model.
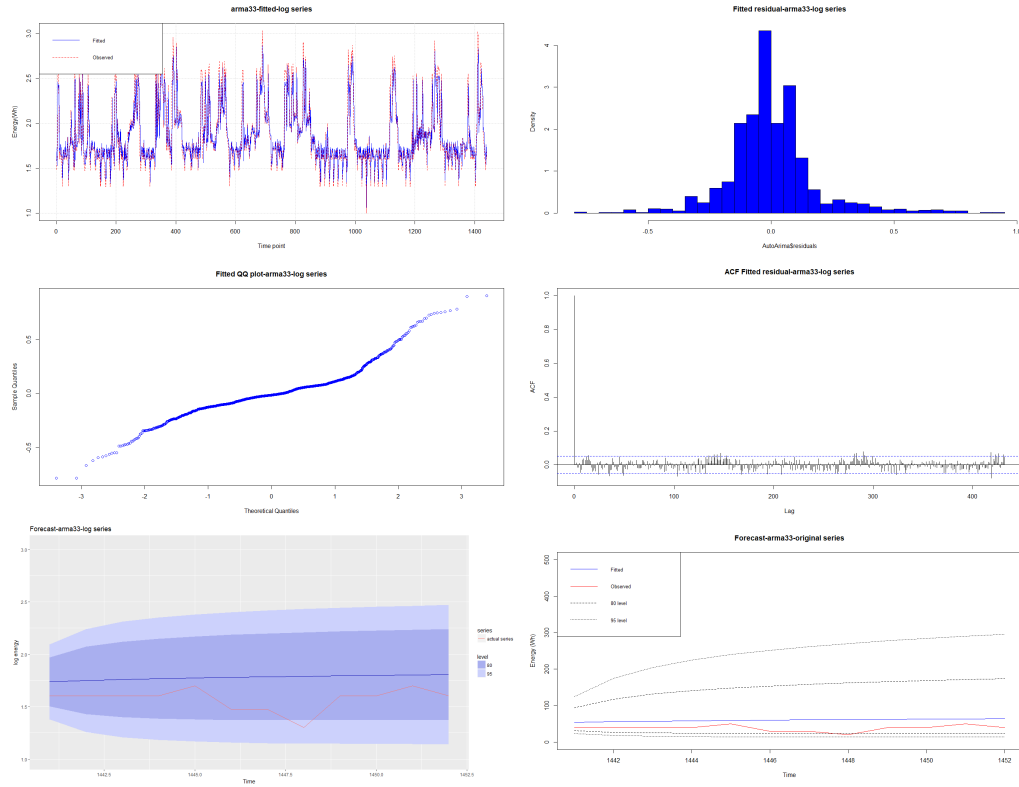
14

Figure 17: ARMA(3,3)

The histogram, QQ plot of the residuals are quite normal. However, the acf still shows some patterns and the forecasts in both the original scale and the log scale are obviously worse compared to other models. The following table summary the measures of fit for all the fitted models so far.

Table 3: MAPE and AIC

Model	MAPE	AIC	Test MAPE
AR(1)	0.063 284 62	−664.2	0.087 897 18
AR(3)	0.064 802 94	−714.86	0.087 988 82
MA(1)	0.063 565 87	−664.92	0.088 899 93
MA(3)	0.065 697 68	−747.59	0.084 000 15
ARIMA(1,0,1)(1,1,1)[144]	0.063 609 4	−458.34	0.091 809 82
ARMA(3,3)	0.064 489 67	−803.36	0.137 181 2

From analysis on different plots, together with the results presented in Table 3, MA(3) should be chosen over the ARMA(3,3). Although it does not have the lowest AIC, the test MAPE make up for that. Besides, its AIC value is the second lowest among the others.

3.2 Forecasting algorithms

The one of the common points of these algorithms is that they contains some parameter ranging from 0 to 1. One must decide which value of each parameter for the model to work well. The proposed approach is plotting the RMSE against several values of each parameter of the algorithm to investigate the effects of each parameter on the performance of the model. These are the models to be presented in this section:

1. Exponential Smoothing ($\alpha = 0.9999339$)

2. Holt's linear ($\alpha = 1, \beta = 0.01253426$)

3. Holt Winter ($\alpha = 0.8, \beta = 0, \gamma = 1$)

First, Exponential Smoothing is attempted for the original series. This algorithm contains only one parameter. Figure 18 shows the effect of α on the RMSE of the model.
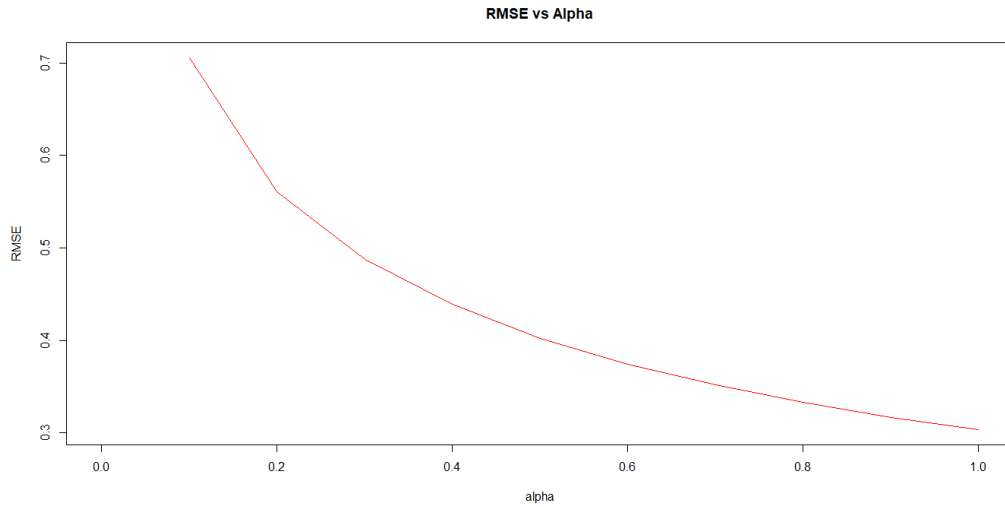
Figure 18: The effect of α

As α increases, the RMSE tends to decrease. One can expect that the optimal value of alpha for this algorithm should be close to 1. The function HoltWinters() is used to help search for the optimal value of α. After running, the function returns the optimal model as Exponential Smoothing ($\alpha = 0.9999339$). Figure 19 show the quality of fit of the model
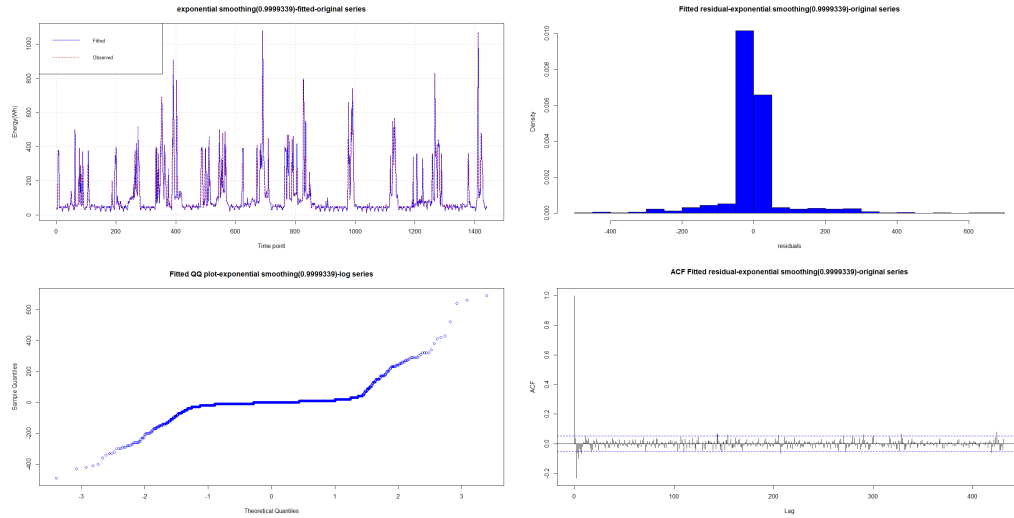
Figure 19: Exponential Smoothing ($\alpha = 0.9999339$)

The histogram does not show the normal shape. Interestingly, it shows that the residual values quite centralized around zero, indicating that the algorithmm fits quite well the observed data. The acf of the residuals also looks quite nice since there are no clear patterns. Here is the forecast of the algorithm.
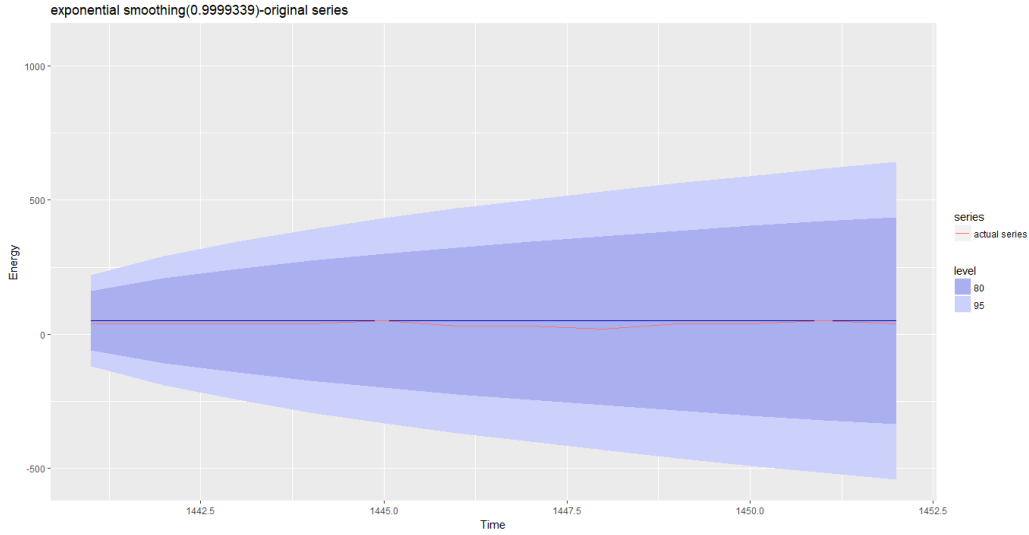
Figure 20: Exponential Smoothing ($\alpha = 0.9999339$) - Forecast

The forecast seems to be naive, failing to catch no variation in the series, although it shows to be quite closed to the original values. This could be because from the point when there is no observed value left, the algorithm takes on the previously predicted value to applied for the forecast, meaning $F_{t+1} = F_t + \alpha(y_t - F_t) = F_t + \alpha(F_t - F_t) = F_t$ in which t is the out of sample point of time. Fortunately in this case, the series didn't have any large variation.

Next, the same approach is used for the Holt's linear algorithm. This algorithm is an extension of the exponential smoothing algorithm. It incorperates a term for linear trends. There are two parameters to be observed. The following plots show how each of the parameters affect the performance of the model.
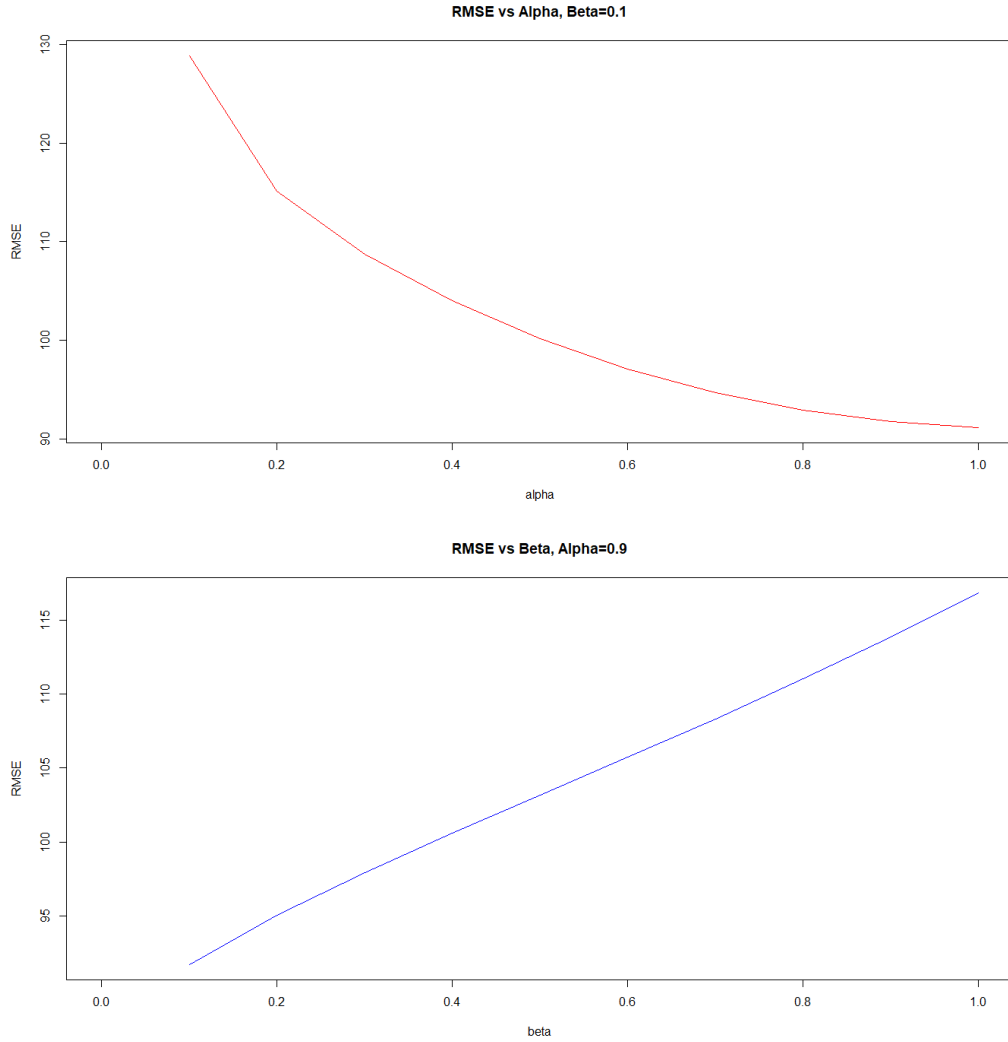
Figure 21: The effects of α and β

As can be seen, increasing α could reduce the RMSE, meanwhile the opposite happens to β. The optimal value of α is expected to be close to 1 and that of β is expected to be close to 0. Using R package, the optimal model is obtained. The model is Holt's linear($\alpha = 1$, $\beta = 0.01253426$). This model is applied on the **log series**, since R gives a warning message regarding the difficulty in optimization when using the original series. Figure 21 shows the results of the model on the log series
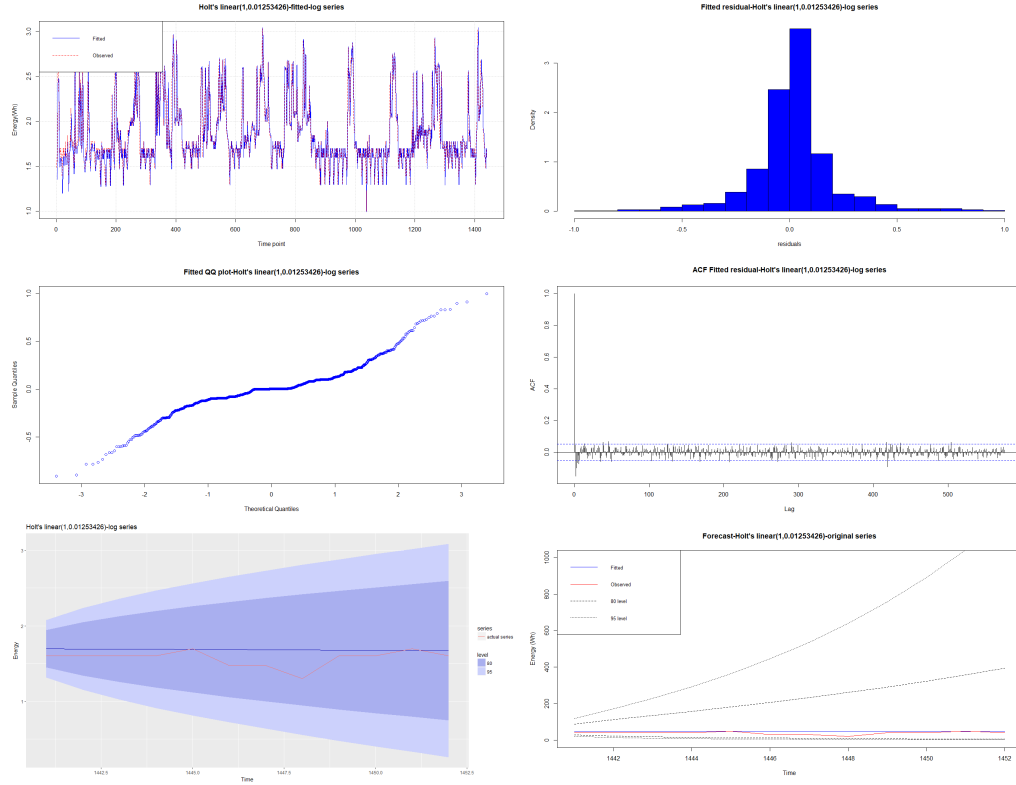
Figure 22: Holt's linear($\alpha = 1$, $\beta = 0.01253426$)

The model shows to have similar results as previous ones. No significant improvements are obvious. However, there is a slight trend in the forecast in opposed to projecting a straight line into the future.

The final algorithm in this group to be applied is the Additive Holt Winter algorithm, which contains a seasonal component. The similar approach is done and below shows the effect of each parameter.

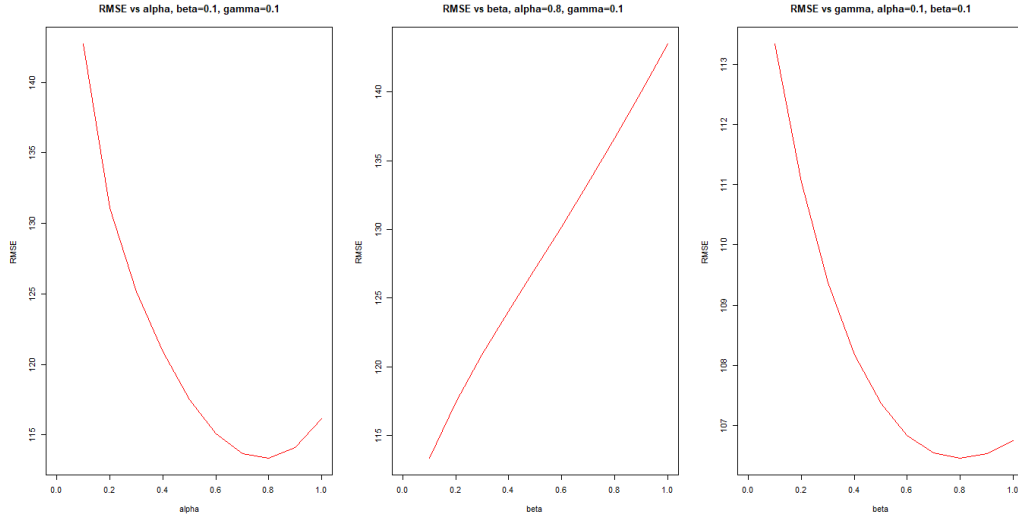Figure 23: The effects of α, β and γ

The figure show how each parameter impacts on the RMSE. As can be seen, the optimal values for α, β, γ could be around 0.8, 0.1 and 0.8 respectively. Using R package, the returned parameters are $\alpha = 0.8, \beta = 0, \gamma = 1$, which is close to the expectation.Figure 23 shows the results of the model on the log series.

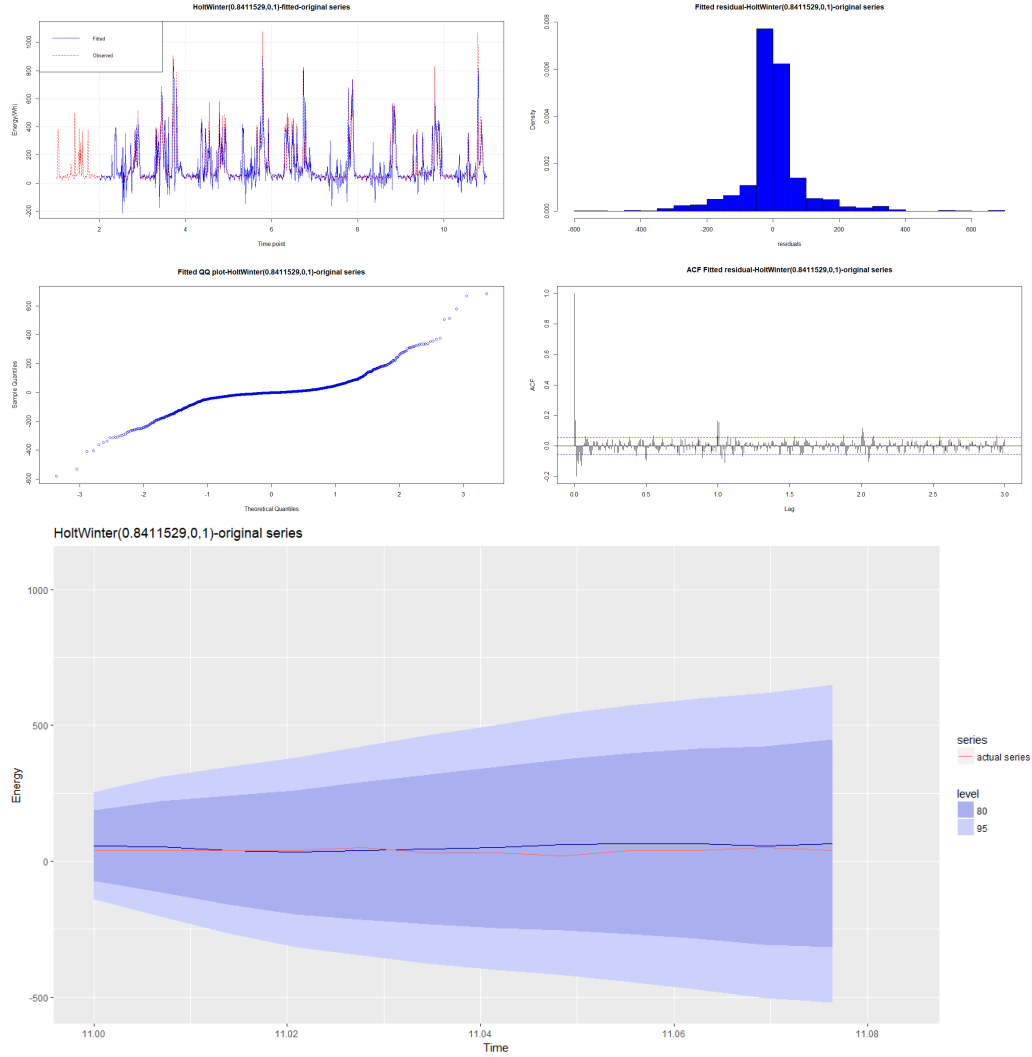Figure 24: Holt Winter($\alpha = 0.8, \beta = 0, \gamma = 1$)

Although the forecast of given by the parameters are quite good, the fit is not adequate. The acf plot show quite high autocorrelation in the residuals. Carefully check the panel on the top left, one can see that at some points the evergy receives negative values. The following table summarizes the MAPE values of the algorithms computed in the original scale.

Table 4: MAPE

Model	MAPE	Test MAPE
Exp. Smoothing(0.9999339)	0.303 678 8	0.381 930 6
Holt's linear(1,0.01253426)	0.307 854 1	0.350 228 1
HoltWinter(0.8,0,1)	0.621 766 1	0.537 838 7

As the table suggests, one should opt to Holt's linear algorithm in this case since it gives the lowest test MAPE. However, the Holt Winter is also worth considering because the algorithm takes into the seasonality components.

3.3 Models that can deal with changing variance over time

As the original series shows to have unstable variance over time, it is worth considering a model that can cope with changing variance. In this section, we attempt to see if a GARCH model could do a better job for the problem. We first conduct some analysis to make sure that the method fits the series. The series is first transformed into log scale, then the first order difference at lag 1 is taken. Let the resulting series be called x_t. Here is how x_t looks like.
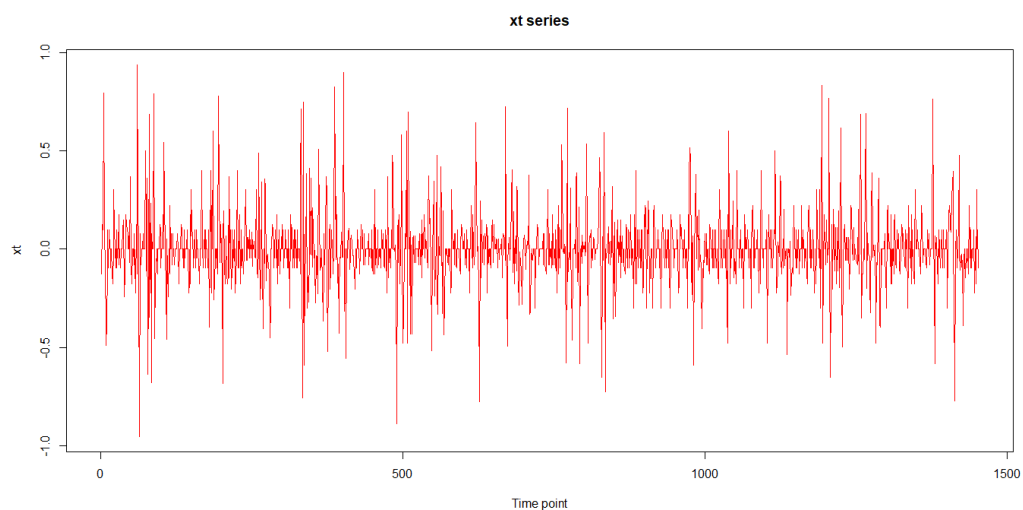
Figure 25: x_t series

The series looks like it is stationary in mean, the variance, however, is not. This can be checked by the acf plots of the x_t series and the x_t^2 series, which represents the variance.
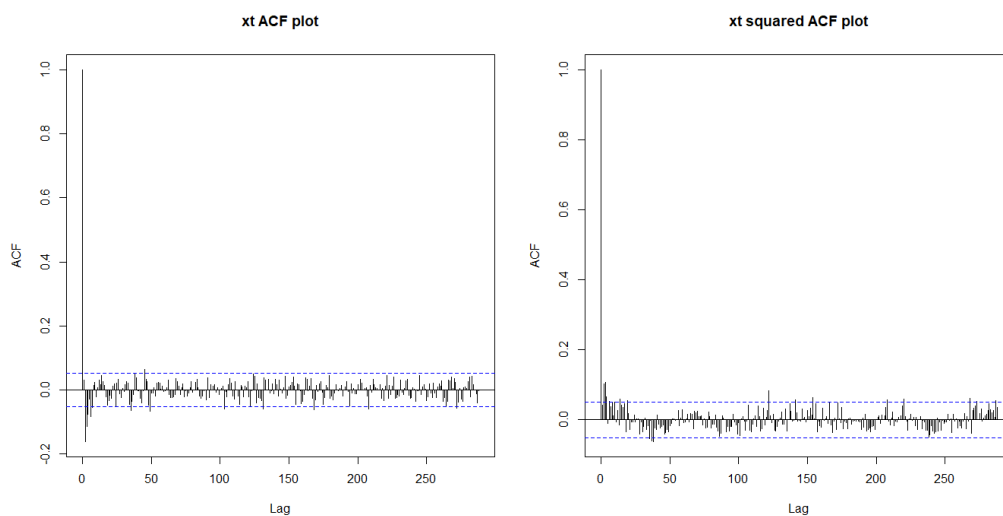

Figure 26: x_t series

The acf of the x_t series goes to zero quickly, meanwhile, the acf of the x_t^2 series show somes patterns, showing evidence against stationary. This fits the description of a series where a GARCH model may be suitable. Alternatively, GARCH(1,1)-ARMA(1,1) is applied to the original series to avoid having to transform and undifferencing the series. Figure 27 shows the learned model.

```
*---------------------------------*
*            GARCH Model Fit       *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,1)
Mean Model      : ARFIMA(1,0,1)
Distribution    : norm

Optimal Parameters
-----------------------------------
        Estimate  Std. Error  t value  Pr(>|t|)
mu      57.71754   6.336618    9.1086        0
ar1      0.76766   0.034850   22.0274        0
ma1      0.37536   0.040105    9.3593        0
omega  219.70377  19.848584   11.0690        0
alpha1   0.26703   0.018629   14.3340        0
beta1    0.73197   0.011232   65.1654        0
```

Figure 27: acf of x_t and x_t^2

Notice that the p-values of the coefficients are all very small, indicating they are statistically significant. Figure 28 shows the results.
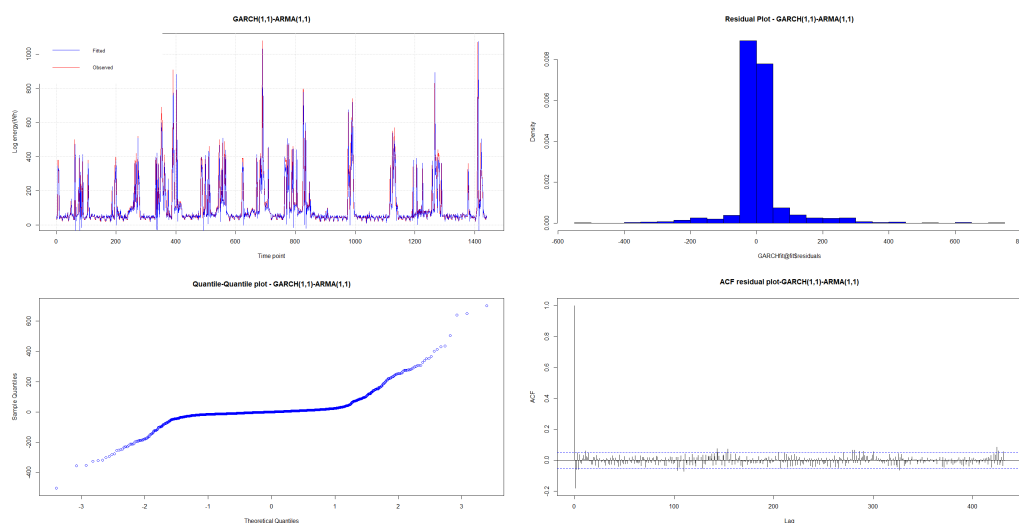

Figure 28: GARCH(1,1)-ARMA(1,1)

The results are similar to previous models. The errors are assumed to follow a normal distribution with mean 0 and variance changing over time, therefore the histogram of the errors and the QQ plot should looks different than those of a sample drawn from a normal distributtion.The acf plot of the residuals is quite nice, although some lag still has high autocorrelation. When using a GARCH component to model the error, the forecasts for the series and the sigma series are of interests. The R package for forecasting provide two different modes: the k-step ahead standard forecast and the rolling forecast. The standard forecast will be based on the unconditional expectation of the model. The rolling forecast give the model the ability to update the forecast once a new data point is observed. This requires the out-of-sample setting. The number of rolling forecast must not be greater than the number of out-of-sample data points. Figure 28 shows forecasts of the model.
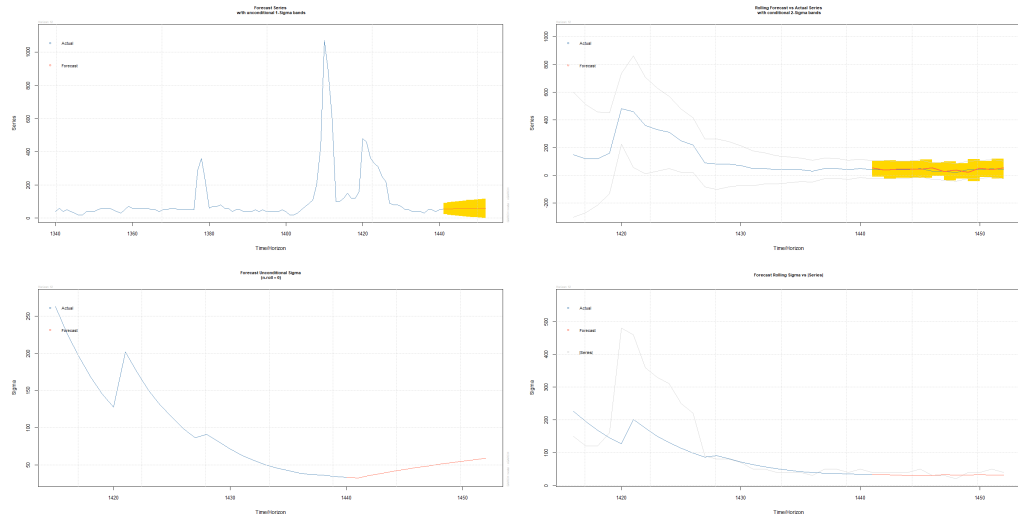

Figure 29: The top panels show the standard forecasts and the rolling forecasts for the series. The lower panels show the standard forecasts and the rolling forecasts for the sigma series.

The rolling forecasts clearly does a better job, since the forecasts does not seem to be naive, it could be because they get updated on each rolling basis. In the lower left panel, the volatility is expected to increase in the next 12 steps, meanwhile it is shown to be level in the lower left panel, which is made

by the rolling forecast approach and is actually more correct.

Notice that this is not the best order sets for this approach. To choose the best fit order of the models, optimization must be done against some measures such as RMSE or AIC. Within the scope of this project, the optimization is not presented here. The fitted MAPE and the test MAPE of the model is 0.3157198 and 0.5724684 respectively, calculated based on the rolling forecasts. The AIC of the model is 10.954. However, we cannot compare this value with previous model since the AIC given by the package is defined differently as follow:

$$AIC = \frac{1}{N}(-2LogLikelihood + 2m)$$

- N is the number of data points

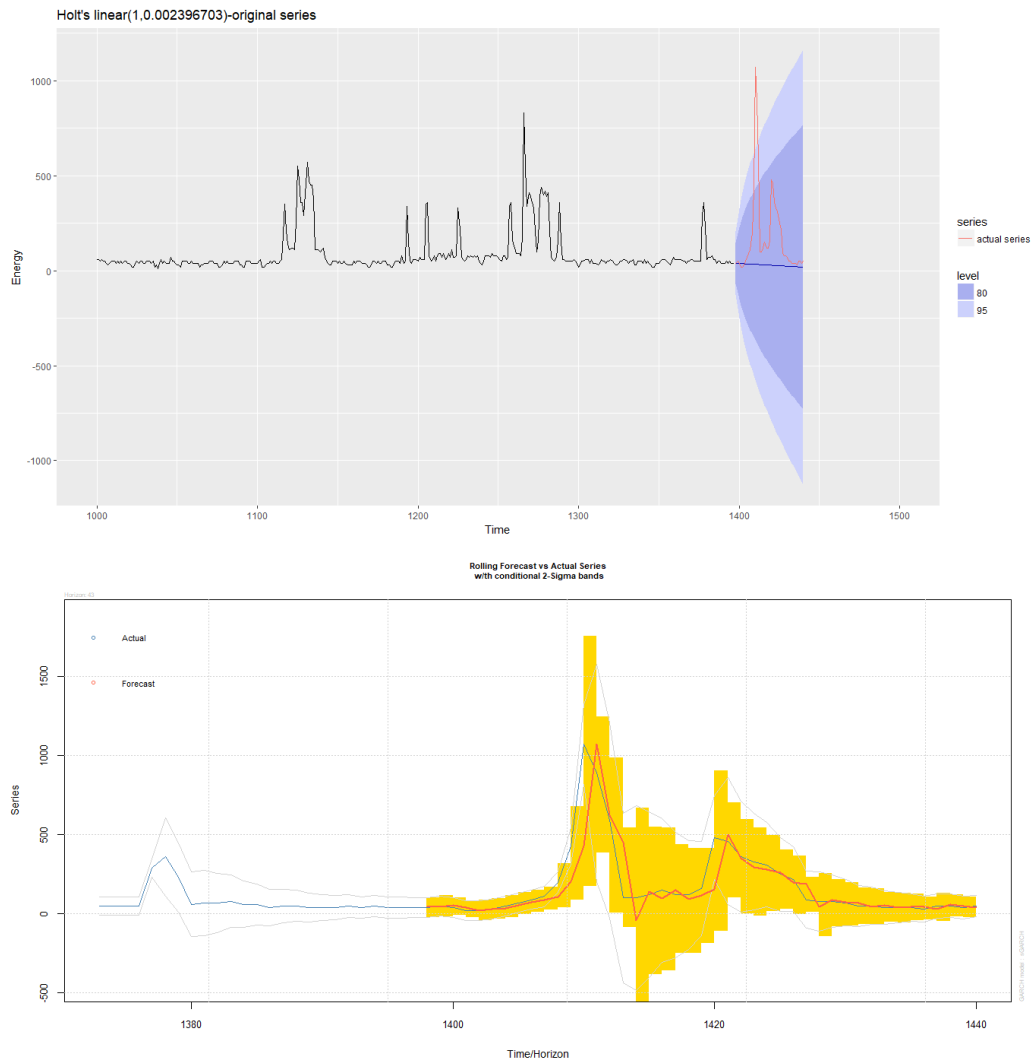- m is the number of fitted parameters

4   Results and conclusion

The following table summarizes the results of the fitted models.

Table 5: Summary

Model	Test MAPE	AIC
AR(1)	0.392 393 9	−664.2
AR(3)	0.392 632 6	−714.86
MA(1)	0.397 431 3	−664.92
MA(3)	0.370 116 6	−747.59
ARIMA(1,0,1)(1,1,1)[144]	0.810 895 3	−458.34
ARMA(3,3)	0.666 924 2	−803.36
Exp. Smoothing (0.9999339)	0.381 930 6	x
Holt's linear(1,0.01253426)	0.350 228 1	x
HoltWinter(0.8,0,1)	0.537 838 7	x
GARCH(1,1)-ARMA(1,1)	0.572 468 4	10.954*

The MAPE in these section is computed on the original series for consistency in comparison. However, the AIC in the GARCH model is computed by a different formula as mentioned before. If using the test MAPE as the criteria of choosing the model, one would opt to Holt's linear since the model gain the lowest test MAPE value. However, observing the figure 3 again we notice one important thing that we have not taken into account: The energy consumption at the end of a day and at the beginning of a day are usually low and stable, slightly fluctuating around 40 to 50. Since we take a sample of 10 days of data to fit the model, and then try to make forecast for the next two hours (12 data points), therefore, it does not show a big difference between models which do not take into account seasonality and variation and models which do so. As one can see here, some models such as MA(3), Holt's linear or AR(3) recieve quite good results, meanwhile GARCH, which can deal with changing variance, seems not to have an ideal test MAPE. However, the next question is: If we had sampled at different periods where variance is high, how would these models have performed? To answer this question, further investigation needs to be done. Here, we show a small experiment to illustrate this idea.

The following figure shows the forecasts of the GARCH model and the Holt's linear algorithm on a different period. The fitting procedure is the same, however, the period is intendedly chosen to have high variation in the forecast set.

Holt's linear(1,0.002396703)-original series

Rolling Forecast vs Actual Series
w/th conditional 2-Sigma bands

Clearly that the Holt's linear completely failed to capture the variation in the series, on the other hand the GARCH shows to do a better job. Due to this reason, we would lean towards models that could capture seasonality and variation such as GARCH and try to optimize the parameters of the model to get the best result in the near future experiments.