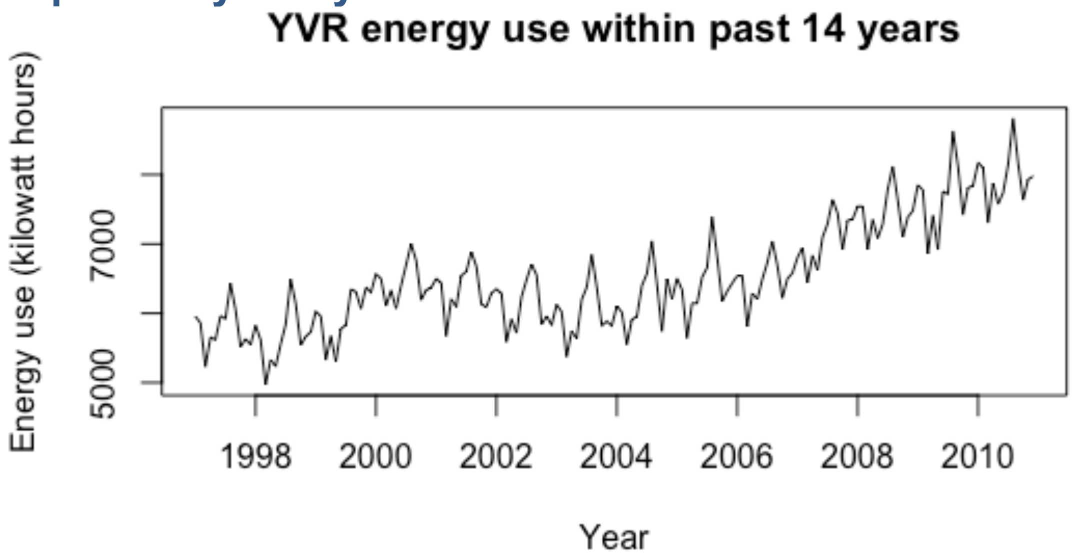
Forecasting Energy use for the Vancouver International Airport (YVR)

Jinye Lu

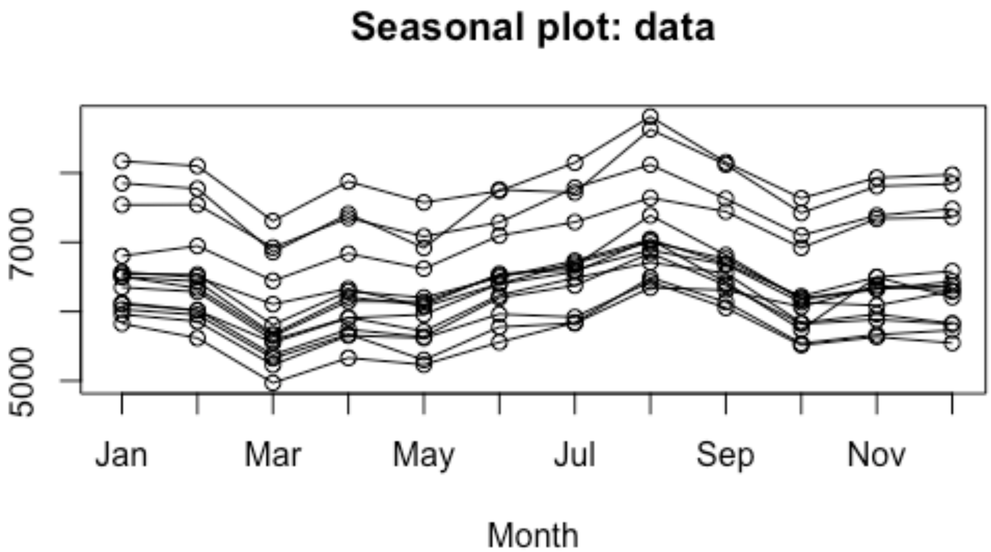
Introduction

Budget planning at YVR requires forecasting the cost of energy to operate the airport. An accurate forecast could help the YVR representatives negotiate more favorable contracts with energy suppliers.
So in order to forecast the cost of energy to operate the airport, the main goal of this project is to develop a model to forecast monthly amount of energy use for the Vancouver International Airport (YVR). This project will develop appropriate models, compare the models, and discuss their corresponding advantages and limitations of each model. Finally, this project will choose the best model and use it to forecast for the next three years.

Exploratory analysis



From the plot, I find that there is trend throughout the time series progresses. The use of energy increases from 1997 to 2000 and then decreases a bit. But after 2004 the use increases again until 2010.
Also, I find that in each year there are two peaks which one is higher than another. Also, there is a valley at the beginning month of the year. So there seems to be a seasonal pattern in the energy use.



From the seasonal plot, it is more clear that the data follow a seasonal plot that the bottom of the data is at March and the peak of the data is at August.

As the development of universal economics, more airlines routes have been launched and thus more people arrived at YVR. So in order to satisfy the demand, the airport expanded terminals to accommodate more aircrafts and passengers. As the increasing area of the airport, more electric energy was used to operate the larger airport.

As is known to us, summer is the most popular season for people to travel. So the number of the flights as well as the passengers landed on YVR achieves the peak. And the weather in summer is the hot which requires a lot of electricity to operate the air conditioner. Therefore, the August is the season with the highest energy use. Also, March is at the dull season of the flights. The number of passengers is smaller than other month. And weather in spring is cool which does not require a lot of electricity top operate the heater or air conditioner. So the March's energy use is the lowest.

I will not use any adjustments or transformations for the data.
Because the data does not show variation that increases or decreases through time. The variance within a year is approximately constant as the time series progresses.
Also, there is no variation due to calendar effects. After trying the calendar adjustment, the variation does not change much.

Developing Models (methods and results)

I will be using a variety of models to try to fit the data including the ETS model and the ARIMA model.

According to the pattern of the data, it has seasonal and trend pattern. So ETS and ARIMA models can handle this two patterns simultaneously and result in more accurate forecasts.

Divide the data into a training set and a test set. What are the time periods for each of these?
The test set is 3 years long and the training set is 11 years long.
The training set and test set data is shown in the table 1 and 2 below:

Table 1. The training set of the YVR energy use

The training set of the YVR energy use model												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1997	5958	5859	5230	5654	5618	5963	5920	6430	6053	5514	5633	5545
1998	5825	5617	4970	5331	5233	5557	5841	6489	6138	5539	5667	5730
1999	6030	5954	5326	5672	5301	5777	5831	6345	6309	6068	6372	6302
2000	6570	6497	6106	6330	6070	6401	6697	7003	6760	6192	6334	6371
2001	6501	6435	5674	6206	6094	6544	6602	6886	6677	6133	6088	6291
2002	6349	6286	5588	5914	5716	6235	6483	6708	6553	5837	5961	5828
2003	6126	6020	5375	5741	5635	6199	6376	6844	6379	5822	5884	5820
2004	6105	6014	5552	5908	5956	6399	6572	7043	6418	5741	6497	6207
2005	6504	6337	5644	6148	6141	6521	6661	7387	6815	6173	6319	6434
2006	6544	6534	5809	6290	6202	6492	6733	7040	6690	6220	6497	6582
2007	6802	6947	6443	6835	6619	7096	7291	7642	7443	6921	7338	7358

Table 2. The test set of the YVR energy use

The test set of the YVR energy use model												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2008	7540	7543	6923	7351	7084	7288	7788	8120	7628	7099	7385	7481
2009	7852	7773	6859	7410	6918	7759	7718	8630	8121	7423	7809	7841
2010	8171	8101	7306	7877	7575	7740	8149	8813	8154	7635	7932	7975

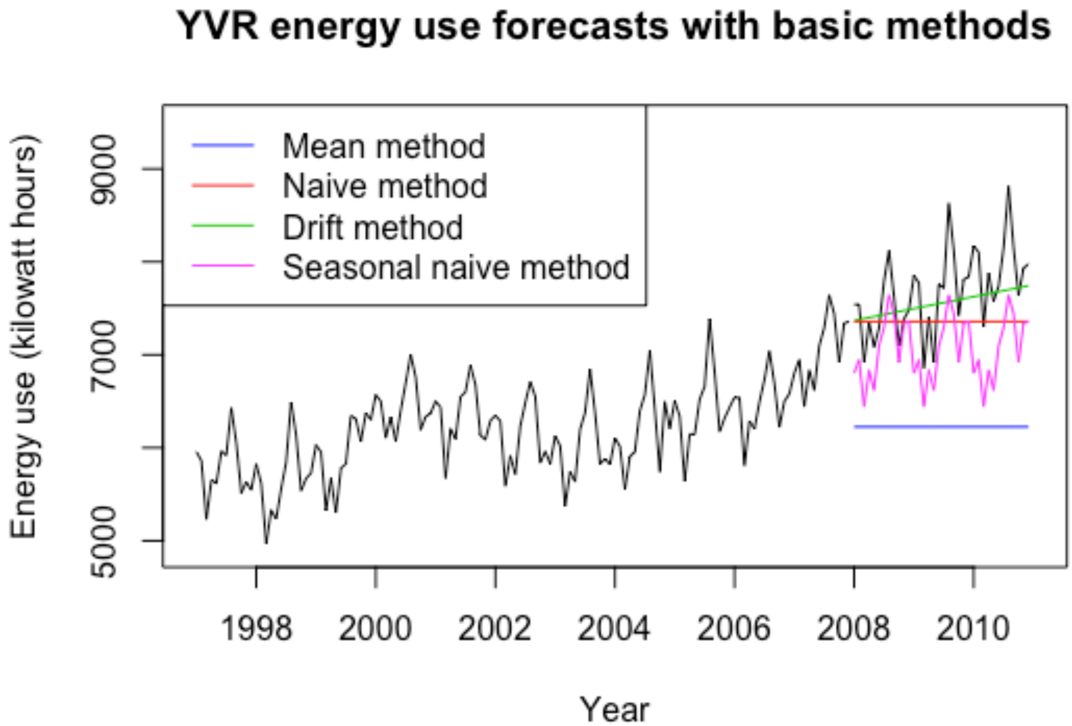
Because the size of the test set is typically about 20% of the total sample. And it should ideally be at least a large as the maximum forecast horizon required which is the next 3 years. So the test set length is 3 years and the left 11 years are training set.

Basic methods

Use the basic methods we have learned to develop forecasts for the test set

1. mean method
2. drift method
3. naïve method
4. seasonal naïve method

Plot the training set and test set data. Include the forecasts of the test set for each of the basic forecasting methods in a different color with a legend to explain.



From the time plot, the mean method does not fit the test set well because it does not follow the increasing trend cycle and seasonality pattern.
The naive method fits the test set better than the mean method but it still does not fit the test set well because it does not follow the increasing trend cycle and seasonality pattern.
The drift method fits the test set better than the mean and naive method but it still does not fit the test set well because it does not follow seasonality pattern of the data.
The drift method fits the test set better than the mean and naive method but it still does not fit the test set well because it does not follow the increasing trend cycle pattern of the data.

Calculate the accuracy measures (RMSE, MAE, MAPE, MASE) to show how well the model forecasts for the test set. Present these in a table.

Table 3. The accuracy measures of the basic forecasting methods

	RMSE	MAE	MAPE	MASE
Mean	1528.4982	1463.7955	18.775576	5.524104
Drift	415.979	328.4824	4.213286	1.239634
Naïve	550.0588	443.1944	5.603453	1.672537
Seasonal Naïve	698.6573	626.8333	8.038624	2.365558

Which of these will provide the best benchmark to compare other models to? Why?
Drift method provide the best benchmark to compare others.

Since its accuracy measures are all the smallest, which means that it fits the original data best. So it can be used as the benchmark for other advanced models to compare and evaluate the following model whether is better than the basic method or not.

Exponential smoothing/ETS model

1. What is the model? Explain the type of model and any parameters.
The model is ETS (A, A, A).
The model is additive Holt-Winters method with additive errors.

The forecast equation is shown below:

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t-m+h_m^+}$$

The level equation is shown below:

$$l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$$

The trend equation is shown below:

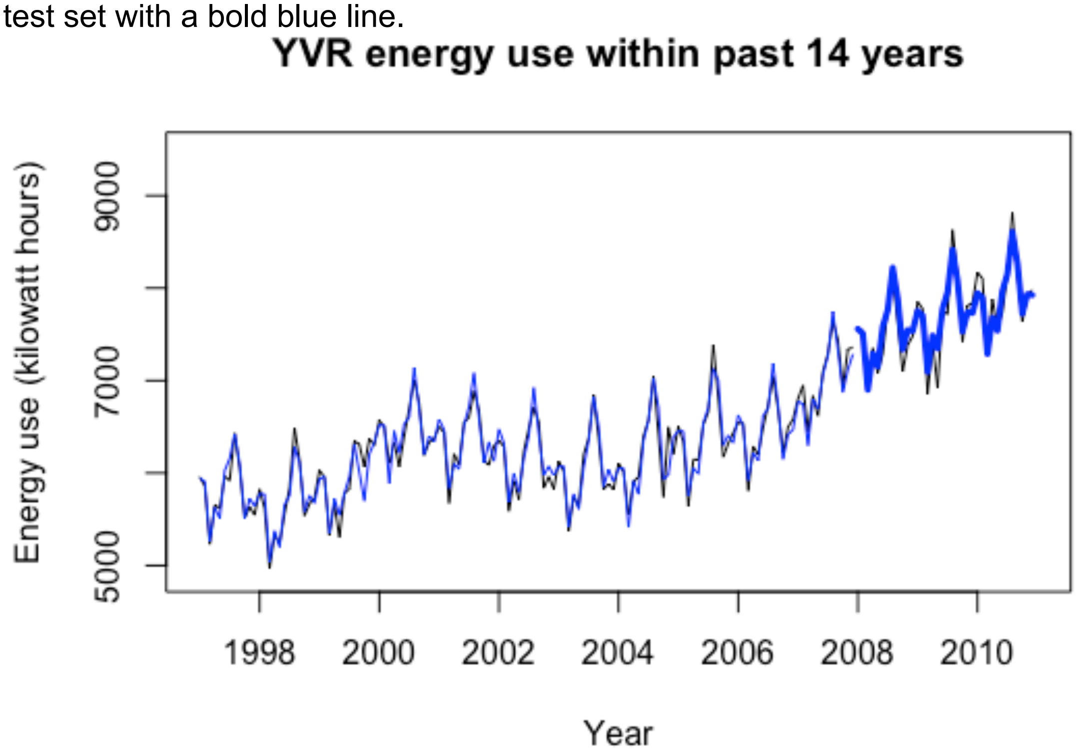
$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$$

The seasonal equation is shown below:

$$s_t = \gamma^*(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$$

The parameter α is a level smoothing parameter and β^* is a trend smoothing parameter. l_t denotes an estimate of the level of the series at time t, while b_t is an estimate of the trend at time t. s_t is the seasonal adjustment component and is a weighted average of the current seasonal index and the seasonal index of the same season last year.

2. Include a time plot of the data in black with a gap between the training set and test set data.
Show the fitted values of the model graphed in blue. Show the forecasts for the test set with a bold blue line.

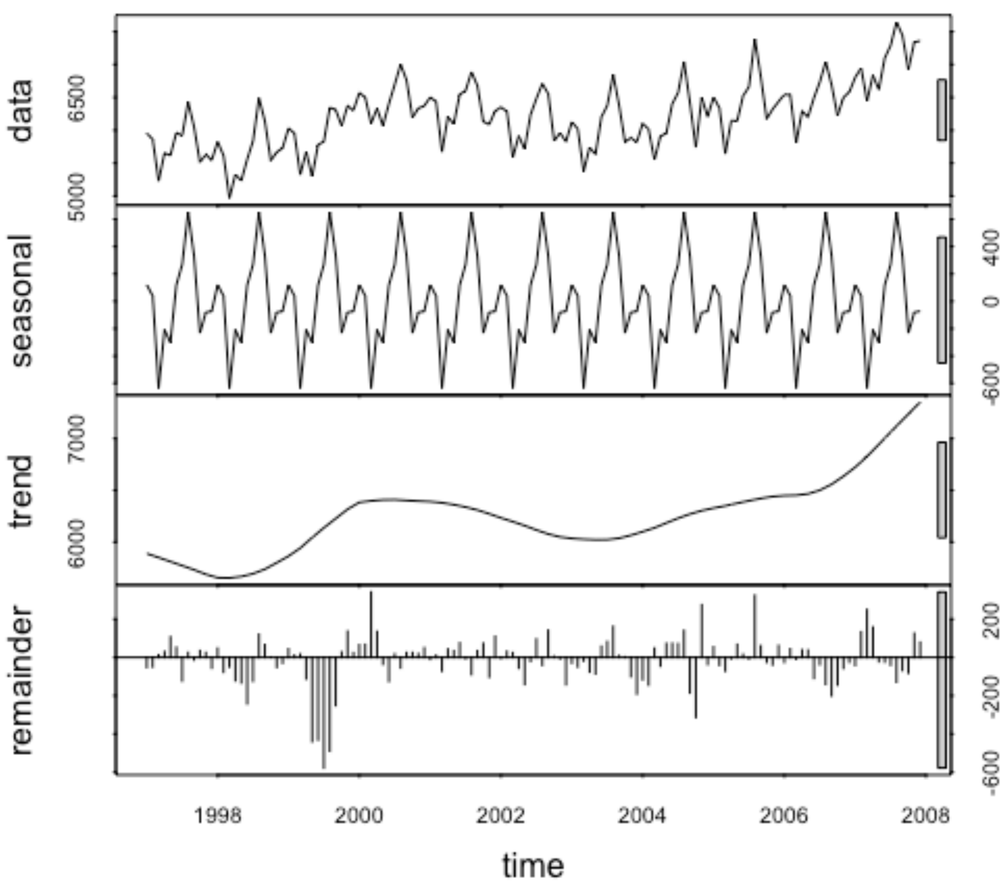


From the time plot with the fitted value and forecasts, the ETS model fit the training set data well. Also, the forecasts for test set are moderately fit. It follows the test set's trend and seasonal pattern in general. However, the peaks of 2009 and 2010 does not predict as high enough as the actual data. Also, the valleys of 2008 and 2009 does not predict as low enough as the actual test set. So the forecasts are not perfectly fit with the data.

3. Explain why this model is appropriate based on the features of the data

In order to get a clear look of the feature of the data, I use the STL decomposition to analyze it.

STL decomposition of the training set



According to the decomposition of the data, it has an increasing trend in general with some rise and fall in the whole process. And the increasing rate does not speed up as the time series progresses, the additive trend component is appropriate.
And according to the seasonal component, it is clear that there is a seasonality in the original data and the variance of seasonal component is constant through the time series. So the additive seasonal component is appropriate.
And the remainder component shows that the errors are randomly distributed, I set the error component as the additive.
Overall, the Holt-Winter additive method is appropriate to handle the data and

forecasting.

4. Quantify and discuss the goodness of fit of the model to the training set

The goodness of fit of the model to the training set can be quantifies by the accuracy measures. The accuracy measures are shown below:

Table 4. The accuracy measures of the ETS model for the training set

RMSE	MAE	MAPE	MASE
124.1839	95.84571	1.5394	0.3617047

According the accuracy measures, the mean absolute scaled errors(MASE) is less than 1. It indicates that the forecast is better than the average naïve forecast computed using the training data. The model fits better than the basic methods.

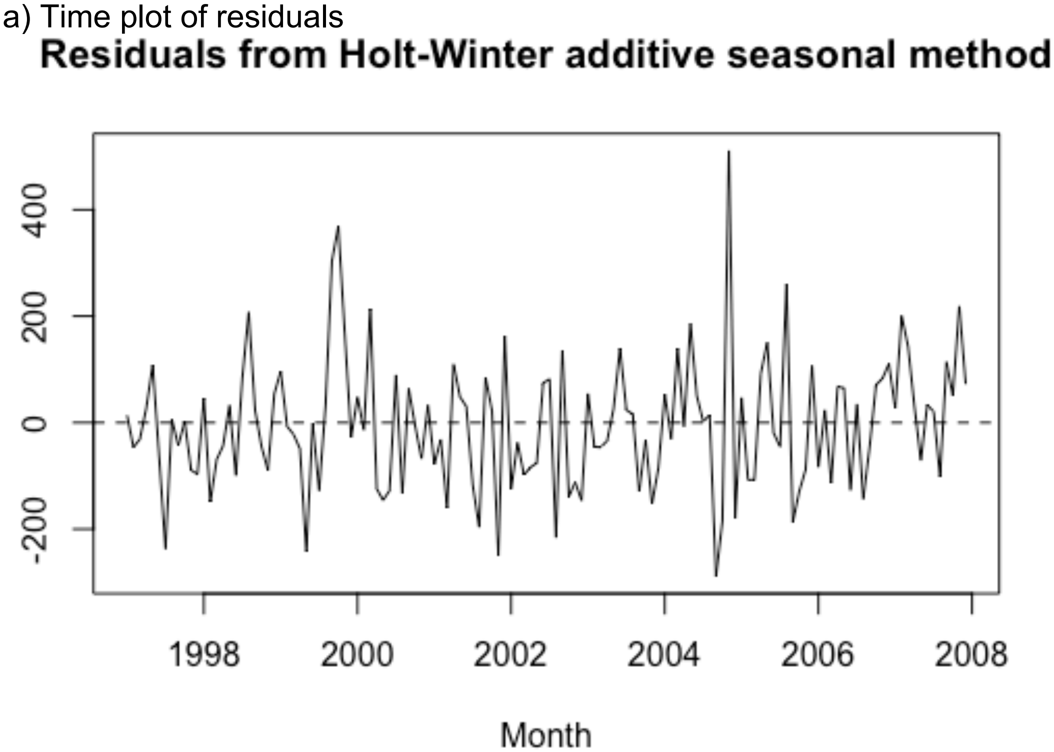
5. Calculate the accuracy measures (RMSE, MAE, MAPE, MASE) to show how well the model forecasts for the test set

Table 5. The accuracy measures of the ETS model for the test set

RMSE	MAE	MAPE	MASE
158.9079	123.0302	1.614566	0.4642941

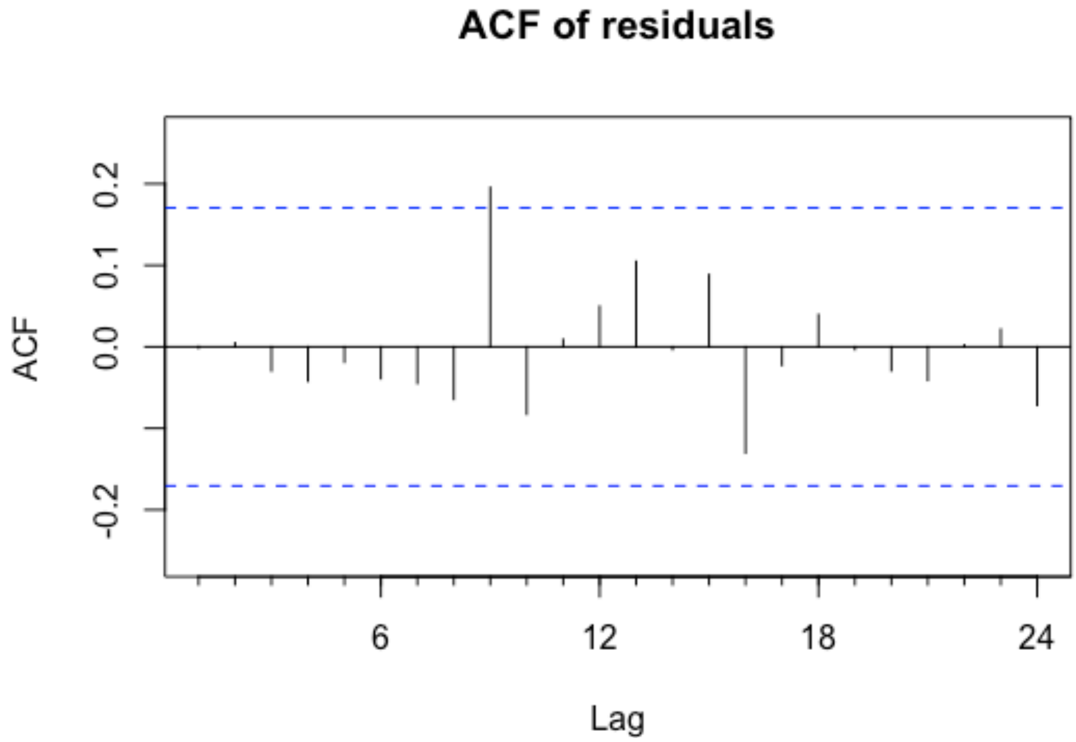
All the accuracy measures are smaller than the benchmark, which means that the model forecasts better than the basic method.

6. a) Time plot of residuals



From the residual time plot, the mean of the residuals is approximately equal to zero. And they have a constant variance in [-200,200] throughout the time series although there are few spikes beyond this range.

6. b) ACF plot of residuals



From the ACF plot, it is clear that 95% of the peaks to occur between between ± 2/√T for a white noise series, where T is the length of the whole series. So there is no correlation in the residuals.

6. c) Ljung-Box test and/or Box-Pierce test of autocorrelations of residuals

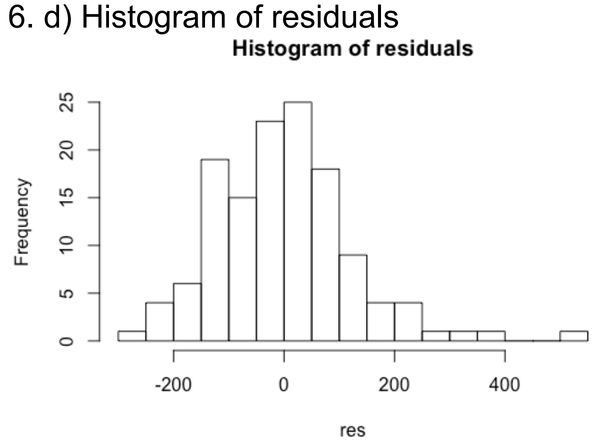
H_0: the first h autocorrelations are not significantly different from a white noise process.
H_1: the first h autocorrelations are significantly different from a white noise process.

Ljung-Box statistic:
X-squared = 15.422, df = 24,

p-value = 0.9078>0.05

So it fails to reject to the null hypothesis.
It means that the residuals are from a white noise process without autocorrelations.

6. d) Histogram of residuals



Based on the histogram, it seems like a right-skewed distribution. Because the residuals on the left side seem more concentrated on the middle than the right side. So the histogram does not look like a perfect normal distribution.

6. e) Mean of residuals
-4.86032

6. f) What properties do the residuals have? What information can you tell about your model from the residual diagnostics?

The properties the residuals have are uncorrelated and its variance are constant through the time series. But the mean is not equal to zero but very close to zero. Since the mean of residual diagnostics is not exact equal to zero, the forecasts are a little biased but moderately suitable.

ARIMA model

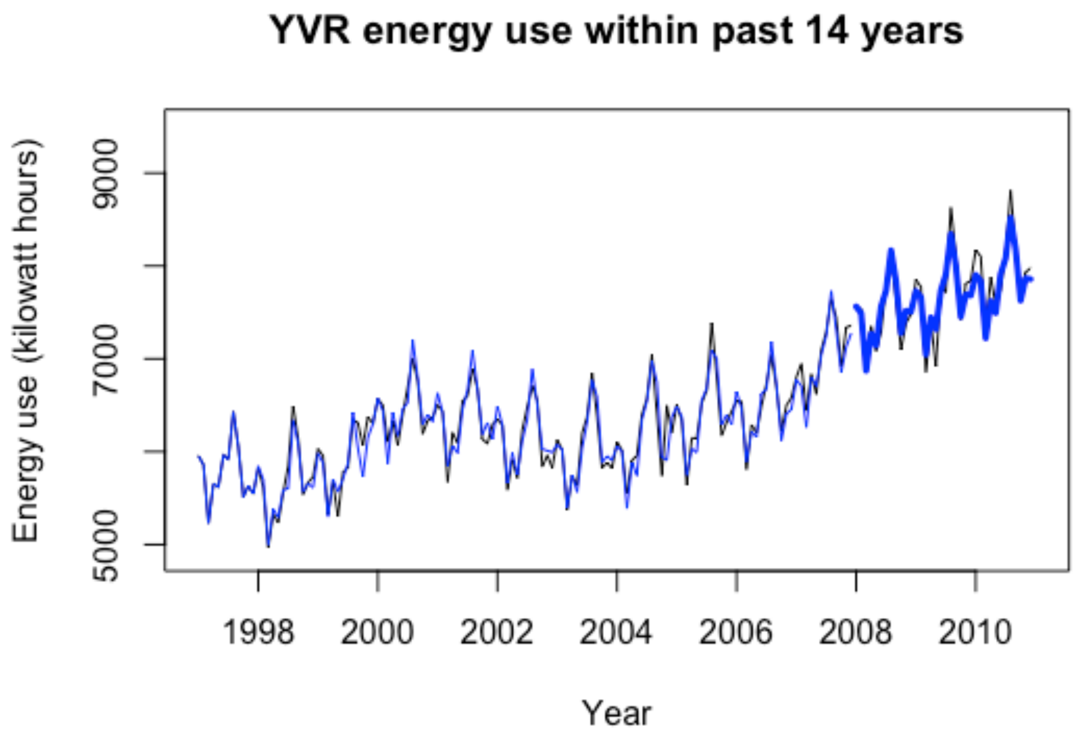
1. What is the model? Explain the type of model and any parameters.
The model is ARIMA(0,1,3)(0,1,1)$_{[12]}$

ARIMA model is the autoregressive integrated moving average model.

The model contains three components in non-seasonal part:
p=0, which refers to none of past time periods to be included in the model.
d=1, which refers to 1 time the series needs to be differenced before it is stationary.
q=3, which refers to 3 lags for the error component.
The seasonal part parameters are:
P=0, which refers to none of past seasonal time periods to be included in the model.
D=1, which refers to 1 time the series needs to be seasonal differenced before it is stationary.
Q=1, which refers to 1 lag for the seasonal error component.

m is the number of season which is equal to 12 in this case.

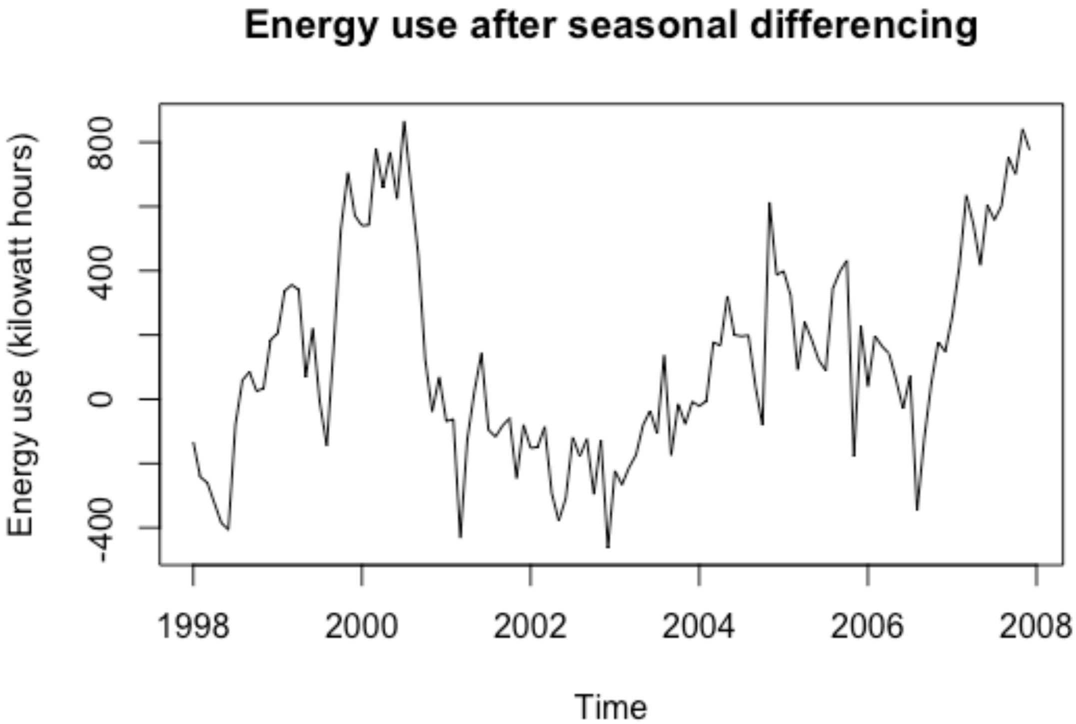
2. Include a time plot of the data in black with a gap between the training set and test set data. Show the fitted values of the model graphed in blue. Show the forecasts for the test set with a bold blue line.



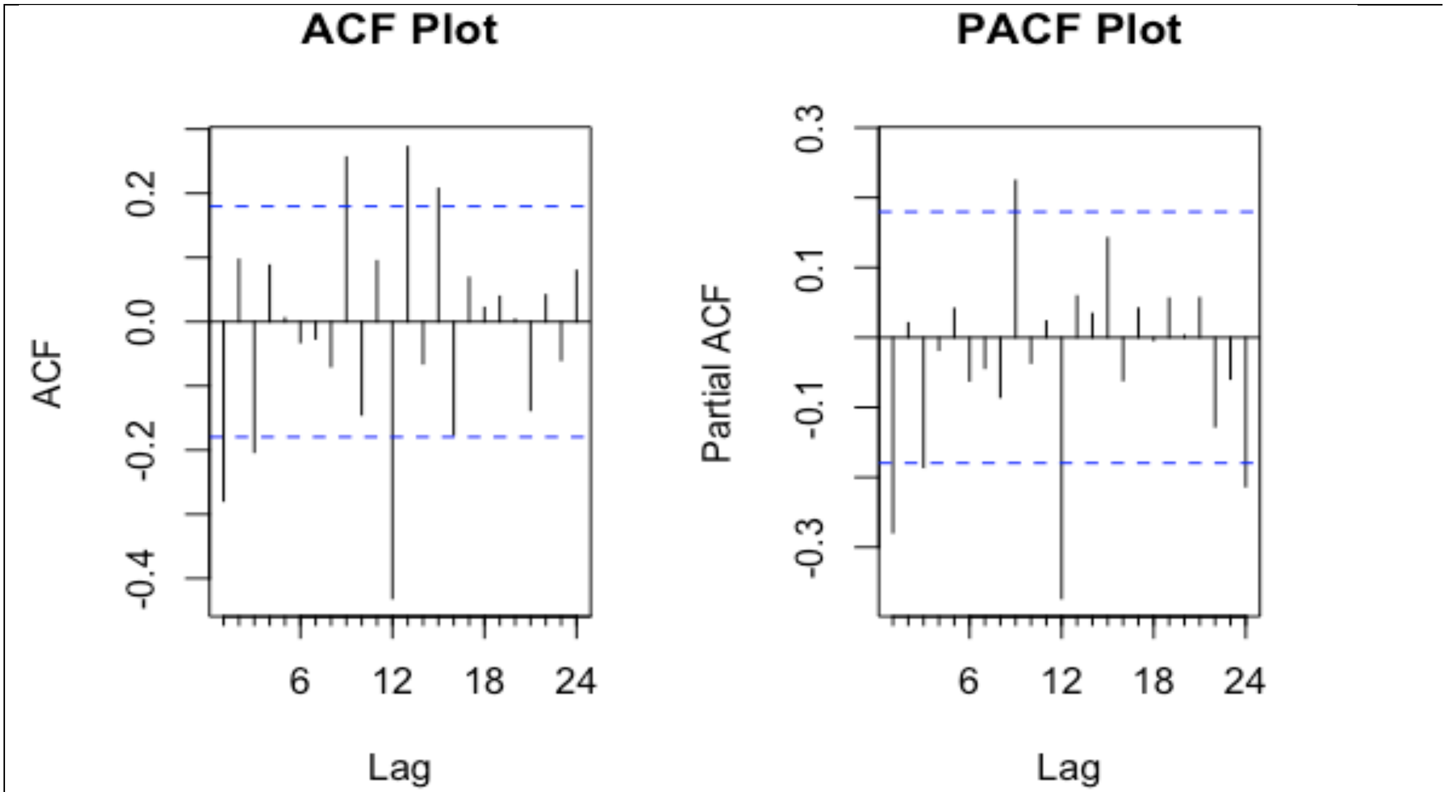
From the time plot with the fitted value and forecasts, the ARIMA model fits the training set data well. Also, the forecasts for test set are moderately fit. It follows the test set's trend and seasonal pattern in general. However, the peaks of 2009 and 2010 does not predict as high enough as the actual data. Also, the valleys of 2008 and 2009 does not predict as low enough as the actual test set. So the forecasts are not perfectly fit with the data.

3. Explain why this model is appropriate based on the features of the data

According to the original data plot, the data have a strong seasonal pattern, so it is best to do a seasonal differencing first. And after the seasonal differencing, the mean of the time series is not constant as the year increases. So it needs a further differencing to make it stationary.



After seasonal and non-seasonal differencing, the ACF and PACF is shown below:



From the plots, I find that ACF lag 1 is negative and ACF lag 1 and 3 is significant spikes and then drop sharply. And the PACF shows slow decay in non-seasonal lag. So I add a moving average of order 3 in the non-seasonal component of ARIMA model.
Also, in the seasonal part, the lag 12 in ACF is negative and ACF has only 1 significant lag before dropping sharply. PACF shows slow decay of lag 12, 24.
So I add a moving average of order 1 in the seasonal component of ARIMA model.

4. Quantify and discuss the goodness of fit of the model to the training set

The goodness of fit of the model to the training set can be quantifies by the accuracy measures. The accuracy measures are shown below:

Table 6. The accuracy measures of the ARIMA model for the training set

RMSE	MAE	MAPE	MASE
128.9141	94.07808	1.496286	0.3550339

According the accuracy measures, the mean absolute scaled errors (MASE) is less than 1. It indicates that the forecast is better than the average naïve forecast computed using the training data. The model fits better than the basic methods.

Table 7. The AICc value of several ARIMA models

ARIMA model	AICc value
ARIMA(0,1,0)(0,0,1)[12]	1850.88
ARIMA(0,1,0)(1,0,0)[12]	1745.92
ARIMA(0,1,1)(0,0,1)[12]	1840.78
ARIMA(0,1,1)(1,0,0)[12]	1737.93
ARIMA(0,1,2)(0,0,1)[12]	1837.76
ARIMA(0,1,2)(1,0,0)[12]	1740.00
ARIMA(0,1,3)(0,0,1)[12]	1826.82
ARIMA(0,1,3)(0,0,2)[12]	1781.70
ARIMA(0,1,3)(1,0,0)[12]	1737.17
ARIMA(0,1,3)(1,1,0)[12]	1554.27
ARIMA(0,1,3)(0,1,1)[12]	1535.34

And from the table 7, the AICc value of the ARIMA model is the smallest, which means that the goodness-of-fit of the model is the better than other models.

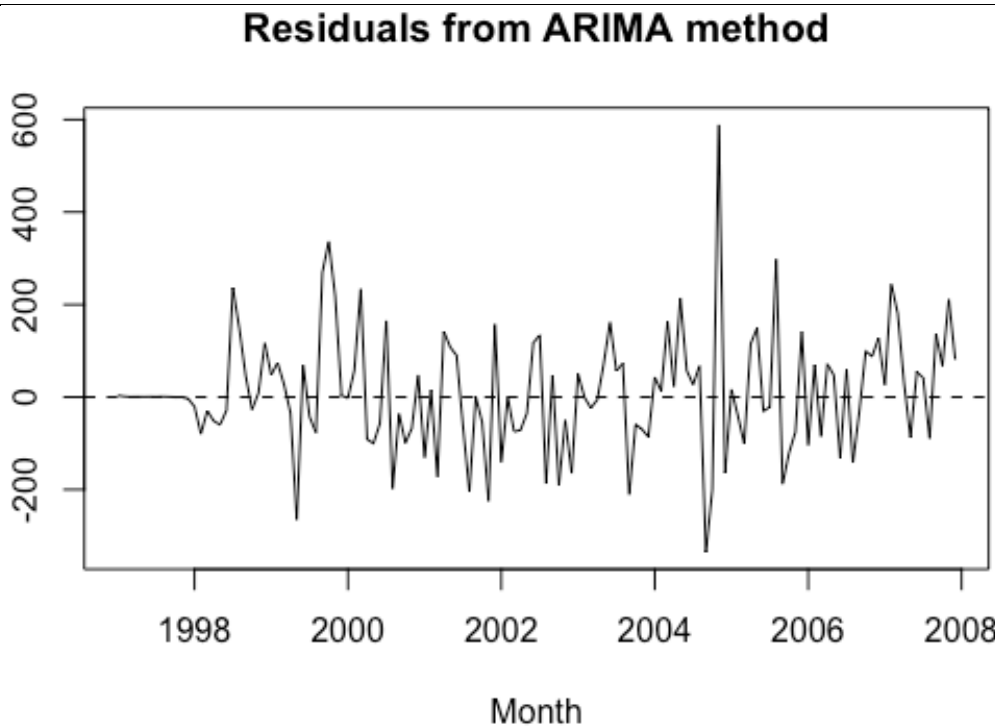
5. Calculate the accuracy measures (RMSE, MAE, MAPE, MASE) to show how well the model forecasts for the test set

Table 8. The accuracy measures of the ARIMA model for the test set

RMSE	MAE	MAPE	MASE
162.1557	129.6366	1.683507	0.4892253

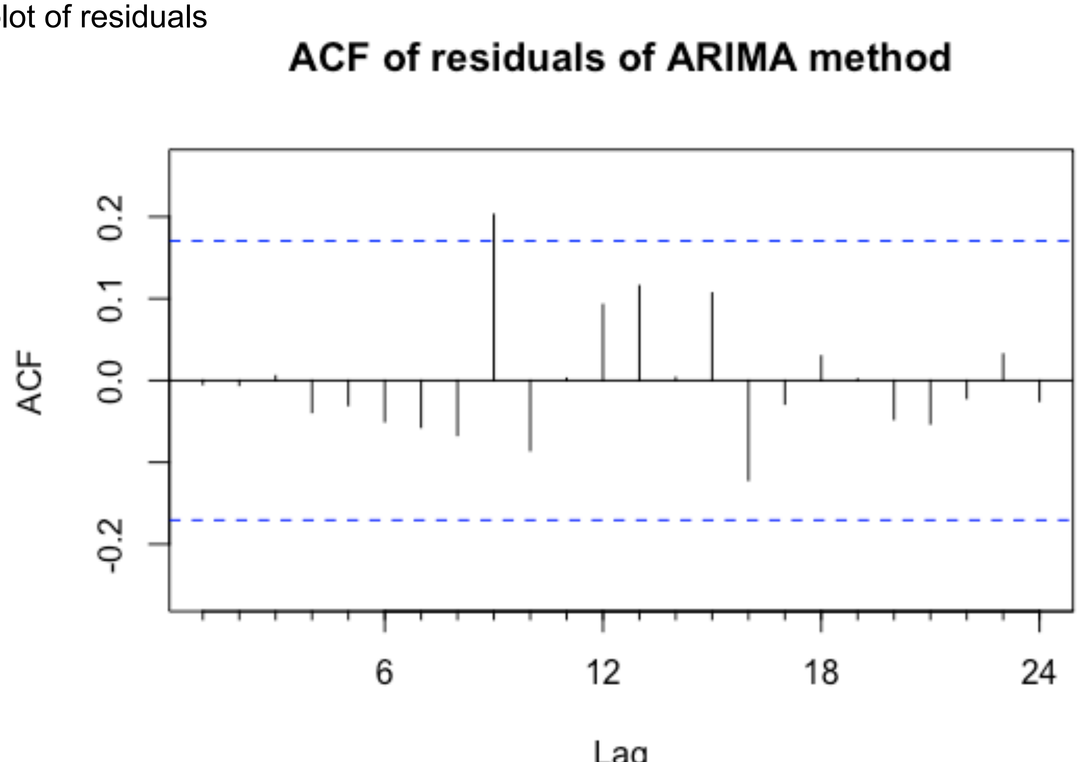
All the accuracy measures are smaller than the benchmark, which means that the model forecasts better than the basic method.

6. a) Time plot of residuals



From the residual time plot, the mean of the residuals is approximately equal to zero. And they have a constant variance in [-200,200] throughout the time series although there are few spikes beyond this range.

6. b) ACF plot of residuals



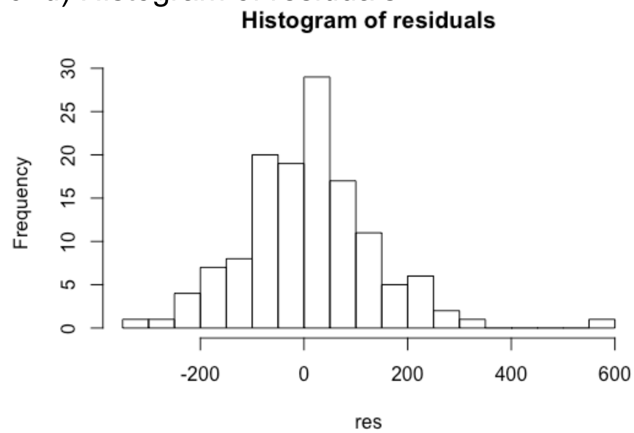
From the ACF plot, it is clear that all of the peaks to occur between between ± 2/√T except one spike, where T is the length of the whole series. So there is no correlation in the residuals. And it is

<table>
<tr><td colspan="2">a white noise time series error.</td></tr>
<tr><td>6. c) Ljung-Box test and/or Box-Pierce test of autocorrelations of residuals

H_0: the first h autocorrelations are not significantly different from a white noise process.
H_1: the first h autocorrelations are significantly different from a white noise process.

Ljung-Box statistic:
X-squared = 17.413, df = 24,

p-value = 0.8305>0.05

So it fails to reject to the null hypothesis.
It means that the residuals are from a white noise process without autocorrelations.</td><td>6. d) Histogram of residuals


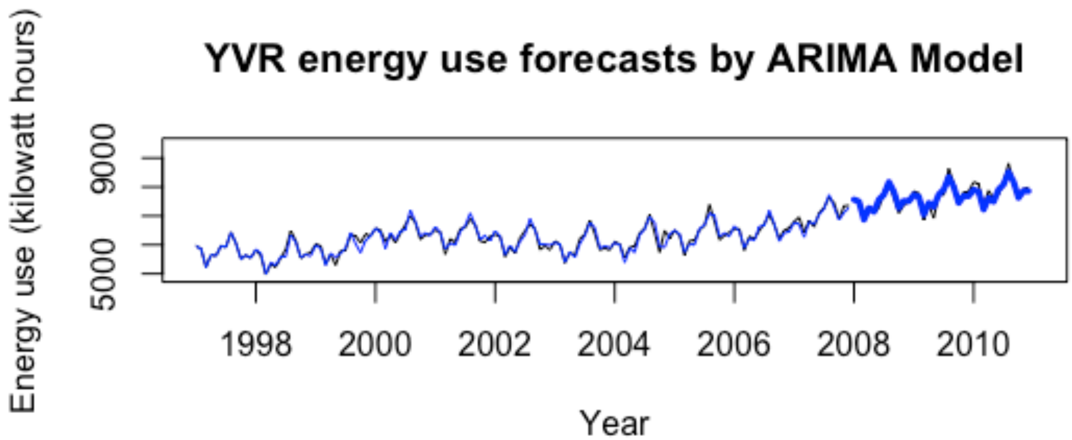
Based on the histogram, it seems like a left-skewed distribution. Because the mode and center of the residuals are on the right side. So the histogram does not look like a perfect normal distribution.

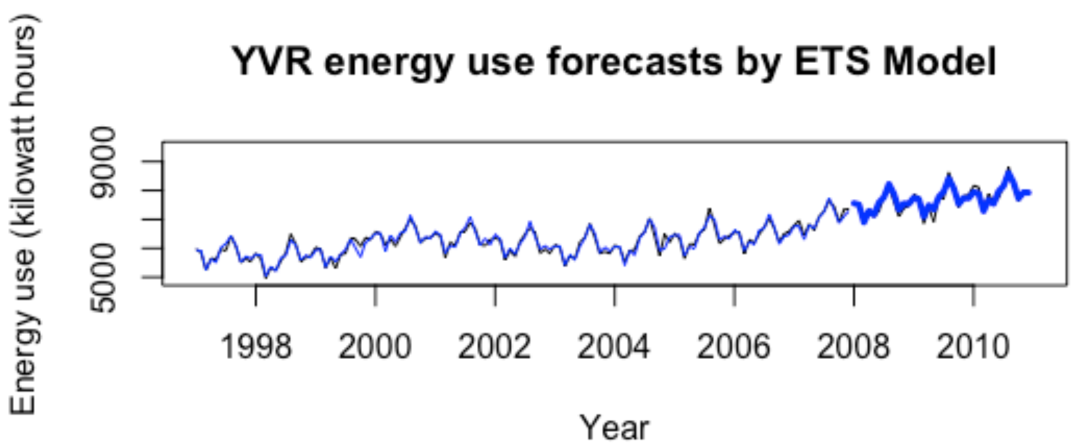
6. e) Mean of residuals
10.29461</td></tr>
<tr><td colspan="2">6. f) What properties do the residuals have? What information can you tell about your model from the residual diagnostics?
The properties the residuals have are uncorrelated and its variance are constant through the time series. But the mean is not equal to zero but close to zero.
Since the mean of residual diagnostics is not zero but close to zero, the forecasts are a little biased but moderately suitable.</td></tr>
</table>

Comparison of models

Compare your exponential smoothing/ETS model, and your ARIMA model to the basic methods. Choose a final model, and give the point forecasts for the next three years (January 2011 through December 2013) in a table.

The forecasts of test set from ARIMA model and ETS model are shown below:





Based on the time plots, the forecasts in the ETS model are almost the same as the ARIMA model. They both follow the same seasonal cycle and trend.
By comparing their accuracy measures for the forecasts of the test set, it is a better idea to make a decision.

Table 9. The accuracy measures of the forecasts in ARIMA model

RMSE	MAE	MAPE	MASE
162.1557	129.6366	1.683507	0.4892253

Table 10. The accuracy measures of the forecasts in ETS model

RMSE	MAE	MAPE	MASE
158.9079	123.0302	1.614566	0.4642941

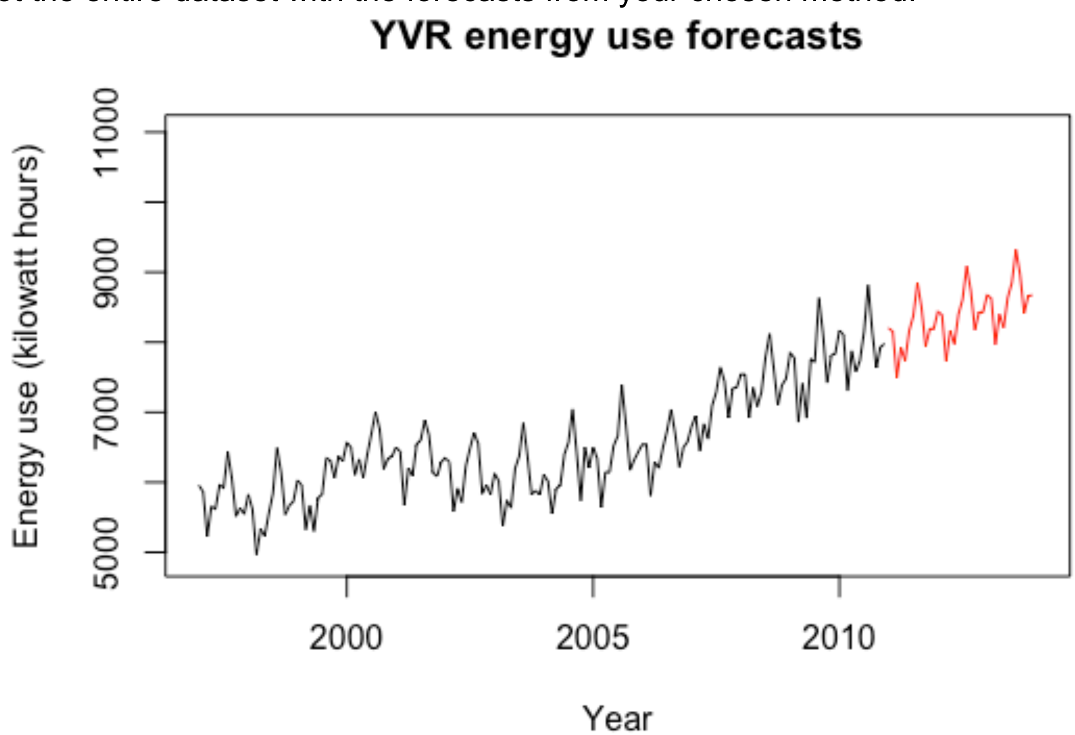
From the table 9 and 10, it is clear that all the accuracy measures of the forecasts in ETS is smaller than ARIMA model. So the ETS model has a better forecast.
Although the mean of residuals in both models are not equal to zero, the mean of ETS (A, A, A) residuals is closer to 0 than ARIMA model. So the forecast of the ETS model is more appropriate and accurate.

The point forecasts for the next three years (January 2011 through December 2013) of the ETS final model is shown in the table below:

Table 11. The point forecasts for the next three years of final model

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2011	8206.274	8145.781	7498.273	7923.28	7738.295	8163.699	8374.555	8845.559	8498.293	7943.21	8187.798	8196.194
2012	8443.102	8382.609	7735.101	8160.108	7975.123	8400.527	8611.383	9082.387	8735.121	8180.037	8424.626	8433.022
2013	8679.93	8619.437	7971.929	8396.936	8211.951	8637.355	8848.211	9319.215	8971.949	8416.865	8661.454	8669.85

Plot the entire dataset with the forecasts from your chosen method:



Discuss any limitations of this final model, and recommendations you have to improve it.

And the model does not take into account other factors that will influence the energy use of YVR. For example, the temperature at YVR may affect the amount of use in air conditioner thus affect the energy use. Also, the area of the airport also affects the use of energy since larger airport would use more electricity. And the operating time can also affect the energy use. Previously YVR was only open from 8AM to 12AM. But as more international airline routes have been launched, the airport began to operate 24/7. So this would increase the use of energy greatly.

The recommendations for solving the issue is that it is also better to include more variables into the model to forecast the energy use in the future. It is more reasonable to use the explanatory forecasting model which the explanatory variables include time and other factors such as the mean of temperature, total area of YVR, etc. The forecasts might be more believable and accurate.

Describe three other possible models to try (you do not have to create these models).

Model Ideas	**Explain why this model is appropriate based on the features of the data**
ETS(A, M, A)	The trend in the recent years seems to speed up increasing. So the exponential trend method with additive season is appropriate for handle the increasing rate of increasing trend time series.
ETS(A, M, N)	The seasonal pattern in recent years seems to be ambiguous where the month of valley is uncertain. So the simple exponential trend method can be used to forecast this non-seasonal data.
ARIMA(0,1,3)(0,0,1)$_{[12]}$	The seasonal pattern in recent years seems to be ambiguous where the month of valley is uncertain. So the ARIMA model without a seasonal differencing is appropriate.

Judgmental forecasting

YVR is planning to expand one of their terminals. They will use a new material that could have very different energy use compared to other areas of the airport. In order to forecast the energy use after the expansion, I will use a judgmental forecasting method since the historical data is not applicable to predicting the energy use in this new area.

Judgmental forecasting methods incorporate intuitive judgment, opinions and subjective probability estimates. Judgmental forecasting is used in cases during

completely new and unique market conditions[1]. I will use the scenario analysis to forecast it.
Scenario analysis is a process of analyzing possible future events by considering alternative possible outcomes. Thus, scenario analysis does not try to show one exact picture of the future. Instead, it presents several alternative future developments. Consequently, a scope of possible future outcomes is observable. Several scenarios are fleshed out in a scenario analysis to show possible future outcomes[2].

Steps in scenario analysis

- The first step is the determination of which factors the scenarios will be built around. These factors can range from the total area of the new materials used to build, to the effect of the new material per unit on the energy use, to the temperature outside the new terminals. I will focus on the most critical factors (the new area & the effect on the energy use) that will determine the value of the energy use and build scenarios around these factors.

- The second step is determining the number of scenarios to analyze for each factor. While more scenarios may be more realistic than fewer, it becomes more difficult to collect information and differentiate between the scenarios in terms of energy use. Thus, estimating the energy use under each scenario will be easier if I lays out 3 scenarios, for instance, than if I lays out 20 scenarios. The question of how many scenarios to consider will depend upon how different the scenarios are, and how well I can forecast energy use under each scenario.
  Since the area of the new section is certain, I would lay out 3 scenarios which the first one is the amount of energy use per unit will decreases greater than 10%, the second one is the amount will decrease within 10%, and last scenario is the amount will not decrease.

- The third component is the estimation of energy use for the whole airport under each scenario. It is to ease the estimation at this step that we focus on only three critical factors and build relatively few scenarios for each factor.

- The final component is the assignment of probabilities to each scenario. The output from a scenario analysis can be presented as values under each scenario and as an expected value across scenarios[3].

Conclusion

This project has developed several forecasting models to forecast monthly energy use for YVR and chose the ETS (A, A, A) model as the final model based on comprehensive comparison and analysis. This model has some limitation for use of forecasts and thus some recommendation has been proposed to improve the model. At last, a judgmental forecasting has been conducted to help develop a forecast in the context of the terminal expansion using new materials in YVR.

References

[1] Forecasting. (n.d.). In Wikipedia. Retrieved February 9, 2018, https://en.wikipedia.org/wiki/Forecasting#Judgmental_methods

[2] Scenario analysis. (n.d.). In Wikipedia. Retrieved February 9, 2018, https://en.wikipedia.org/wiki/Scenario_analysis

[3] Probabilistic approaches: scenario analysis, decision trees and simulations. Retrieved February 9, 2018, http://people.stern.nyu.edu/adamodar/pdfiles/valrisk/ch6.pdf