Forecasting Final Project

May 31, 2016

PHAN Trung Hoai Minh trunghoaiminh.phan@ieseg.fr

Table of Contents

I. **Executive Summary** 2

II. **Dataset 1: Crude Oil Spot Price FOB 1986-2016** 3

III. **Dataset 2: Temperature of Washington DC 1900-2016** 8

IV. **Appendix** 13

Executive Summary

The purpose of this project is to practice forecasting techniques by applying two main forecasting methods (i.e. ETS and ARIMA) to explore real-life data, getting experience, making forecasts and validating them with upcoming data of 2016.

There are two datasets having been selected for the project, one is related to economics topic (i.e. crude oil price FOB) and the other one is related to the global warming issue (i.e. temperature of Washington DC).

Working through the data, there some **key conclusions and learning experience** as follow:

- Oil price is much fluctuated and far more difficult to forecast than temperature, in other hand, is quite stable through decades.
- For highly fluctuated time series data like oil price, we can apply some transformation technique (e.g. logarithms) to shrink the variance before building the forecasting model.
- For stable time series data like temperature, we can choose a shorter period that will be easier to plot and to observe the trend of data.

Dataset 1: Crude Oil Spot Price FOB 1986-2016

Introduction

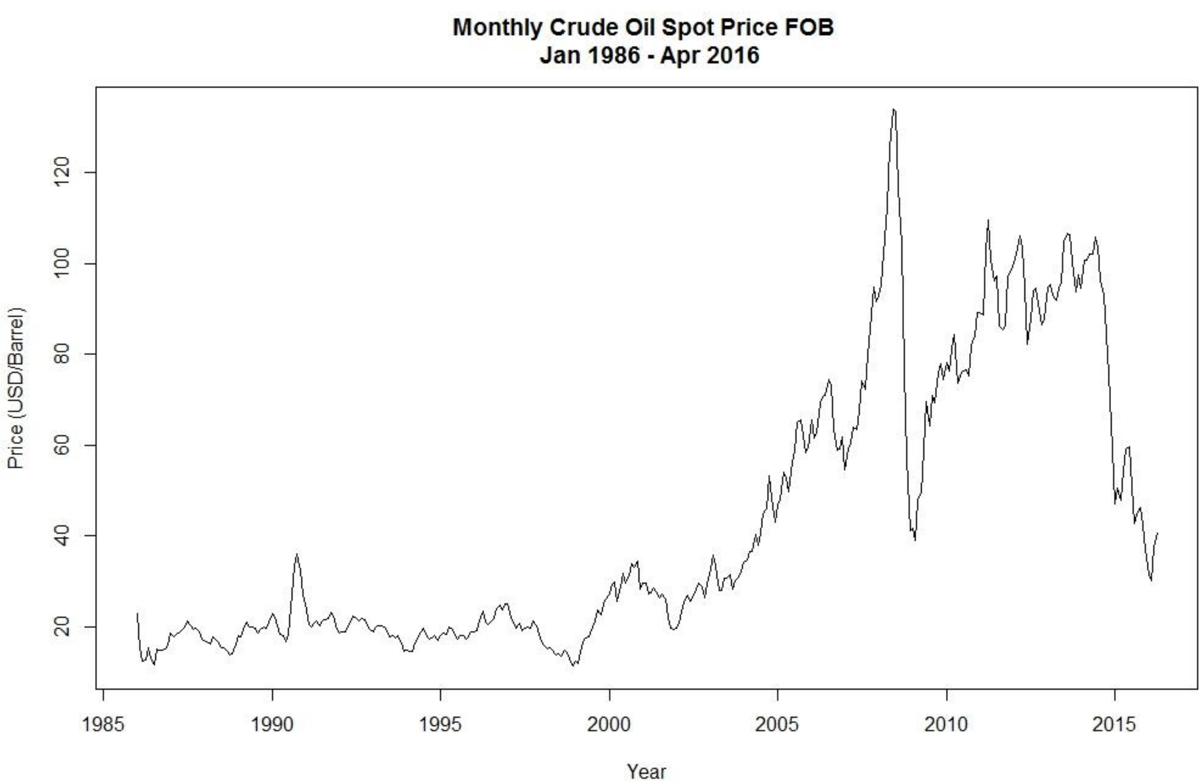
Oil is an indispensable and important commodities of the world, the economics of many countries (e.g. Saudi Arabia, Russia, Iran, Iraq, United Arab Emirates, etc.) depend heavily on oil exports and the changing in each oil barrel price can bring prosperity or misery to a country. As a result, finding out an efficient way to forecast the oil price is crucial and also very interesting to work on.

Data description and exploration

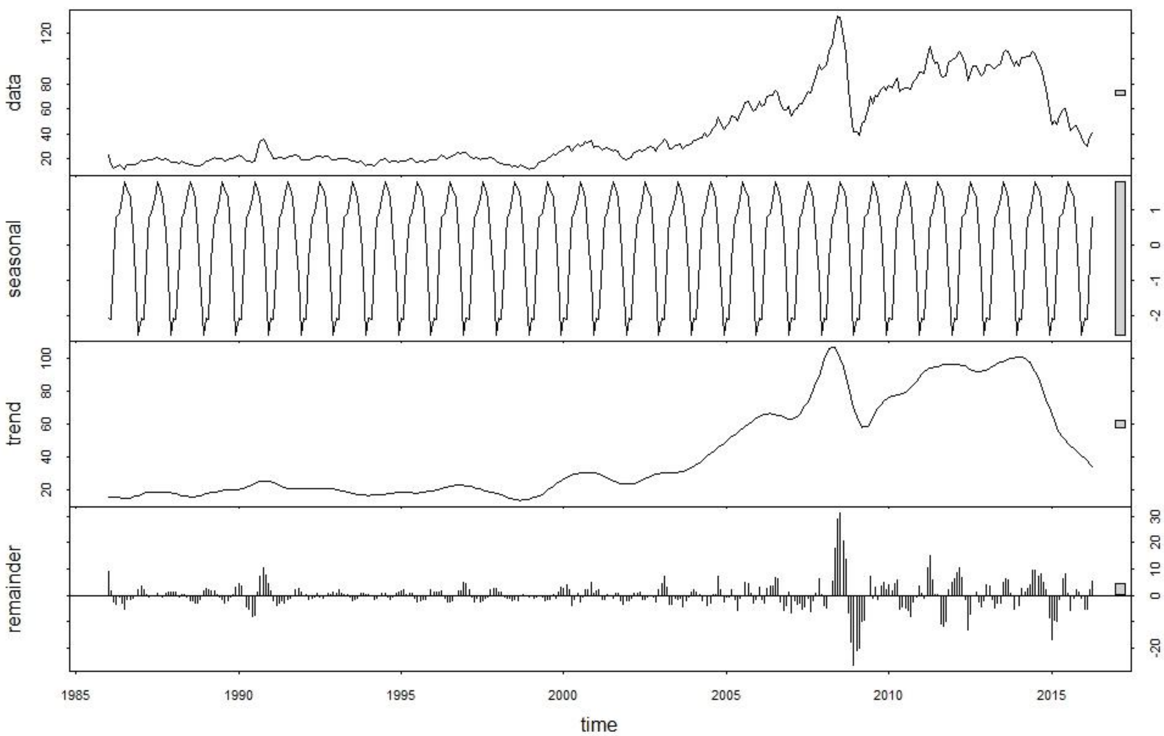
The data was provided by U.S. Energy Information Administration ("EIA"), on their website: www.eia.gov, the details are as follow:

- Name of dataset: Monthly OK WTI Crude Oil Spot Price FOB (USD/Barrel)
- Full period: Jan 1986 – Apr 2016 (364 periods, 100%)
- Training set: Jan 1986 – Dec 2009 (288 periods, 79%)
- Testing set: Jan 2010 – Apr 2016 (76 periods, 21%)

Plotting the original data to get the general idea about the trend and variation. We can clearly see there is widely fluctuation of oil price from 2007 to 2010 and downward trend from 2014 until now mainly caused by the confrontation between the USA and Russia.



To get more understanding about the components of the data, we used the STL decomposition function and plot the result as follow. It is clear that the data contains a seasonal component and an increasing trend over time. Based on this plotting, we can apply the ETS framework with all there parameters, i.e. error, trend and seasonal.



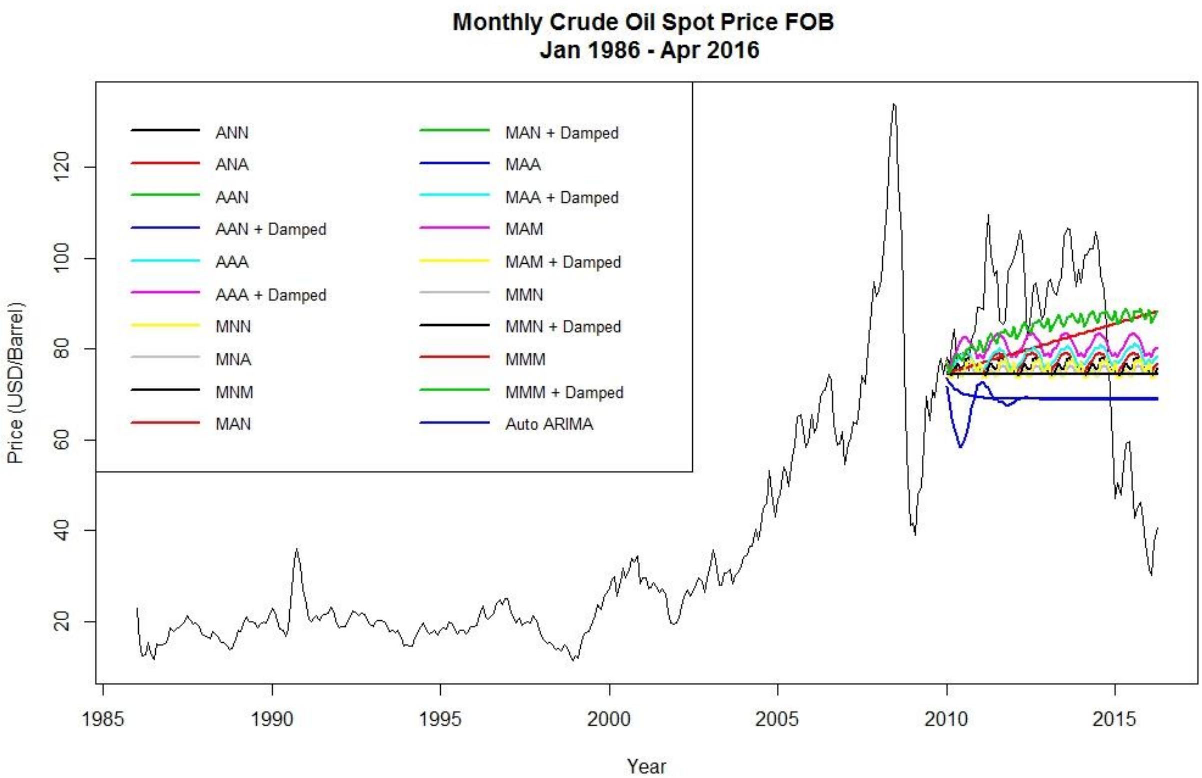
Choosing forecasting models

By applying a set of self-defined functions (the details in Appendix) to scan all 19 ETS possible models and Auto ARIMA model on the training and testing datasets, we have the accuracy measures as follow:

Model name	RMSE	MAE	MAPE	MASE	RMSE	MAE	MAPE	MASE
	Training set				*Testing set*			
ANN	3.861	2.276	6.691	0.244	22.773	20.330	29.638	2.178
ANA	3.789	2.331	7.144	0.250	21.908	19.181	28.955	2.055
AAN	3.498	2.278	7.027	0.244	25.011	23.017	31.107	2.466
AAN + Damped	3.498	2.278	7.027	0.244	25.011	23.017	31.107	2.466
AAA	3.685	2.393	7.495	0.256	21.330	17.767	28.552	1.903
AAA + Damped	3.685	2.393	7.495	0.256	21.330	17.767	28.552	1.903
MNN	3.861	2.276	6.694	0.244	22.773	20.330	29.638	2.178
MNA	3.840	2.269	6.648	0.243	22.576	20.025	29.424	2.145
MNM	3.879	2.273	6.633	0.243	22.270	19.546	29.115	2.094
MAN	3.865	2.277	6.765	0.244	23.824	19.175	31.891	2.054
MAN + Damped	3.859	2.288	6.736	0.245	22.742	20.283	29.613	2.173

MAA	3.810	2.355	7.275	0.252	21.924	18.911	29.204	2.026
MAA + Damped	3.810	2.355	7.275	0.252	21.924	18.911	29.204	2.026
MAM	3.975	2.319	6.768	0.248	22.563	19.832	29.312	2.125
MAM + Damped	3.975	2.319	6.768	0.248	22.563	19.832	29.312	2.125
MMN	3.852	2.276	6.677	0.244	22.756	20.304	29.625	2.175
MMN + Damped	3.852	2.276	6.677	0.244	22.756	20.304	29.625	2.175
MMM	3.973	2.310	6.867	0.247	22.573	16.932	29.652	1.814
MMM + Damped	3.973	2.310	6.867	0.247	22.573	16.932	29.652	1.814
Auto ARIMA	3.241	2.242	7.001	0.240	25.338	23.877	32.211	2.558

Plotting the results of 19 ETS models and 1 Auto ARIMA model:

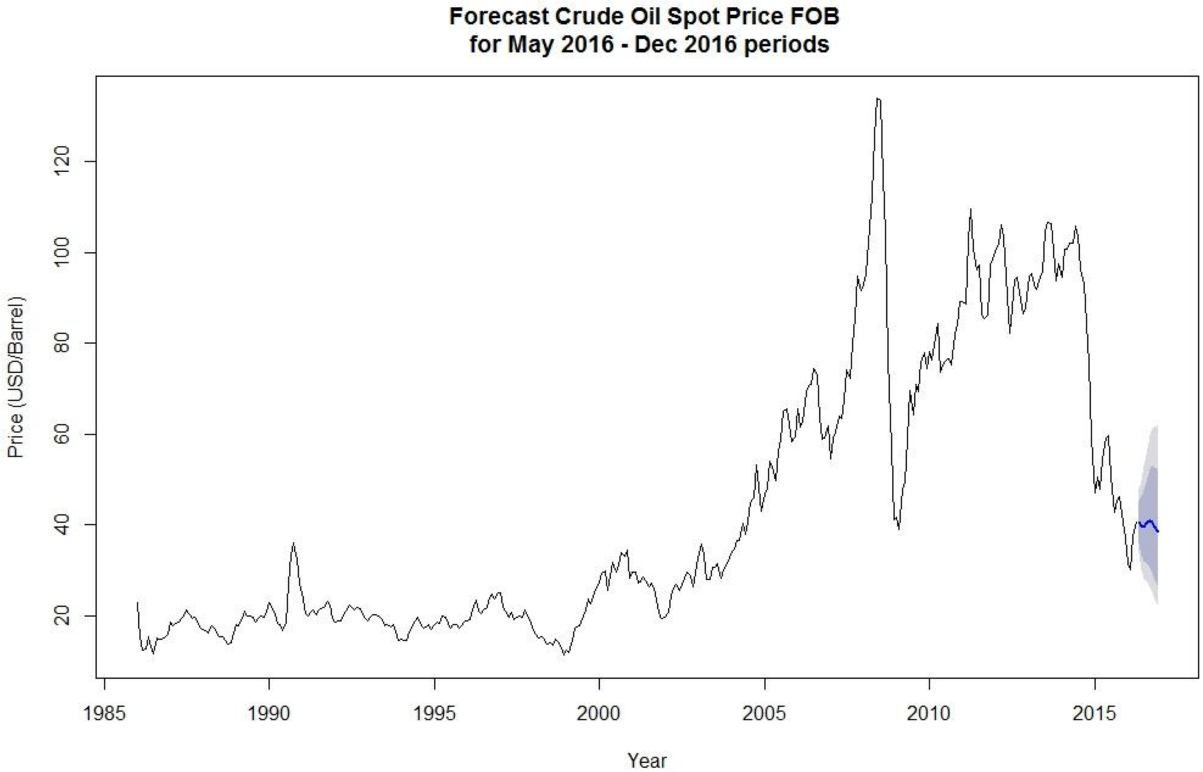


In term of RMSE, ETS(A,A,A) and ETS(A,Ad,A) are best models with the lowest errors. On the other hand, ETS(M,M,M) and ETS(M,Md,M) have far more better performance of MASE than other models. In this case, we select the final model based on MASE measurement with the damped trend, the final forecasting model is: ETS(M,Md,M).

Model summary:

ETS(M,Md,M), alpha = 0.9999, beta = 0.0066, gamma = 1e-04, phi = 0.9486

Using forecasting model: ETS(M,Md,M)



Forecast the Crude Oil FOB price for the next 8 months of 2016:

Month	Point.Forecast	Lo.80	Hi.80	Lo.95	Hi.95
May-16	40.70	36.15	45.34	33.98	47.87
Jun-16	39.74	33.43	46.20	30.33	49.89
Jul-16	39.71	31.97	47.66	28.30	52.88
Aug-16	40.67	31.64	50.45	27.57	56.71
Sep-16	40.96	30.68	52.21	26.40	59.64
Oct-16	40.68	29.76	53.29	25.36	61.26
Nov-16	39.55	28.08	52.71	23.76	61.72
Dec-16	38.43	26.49	52.26	21.93	62.23

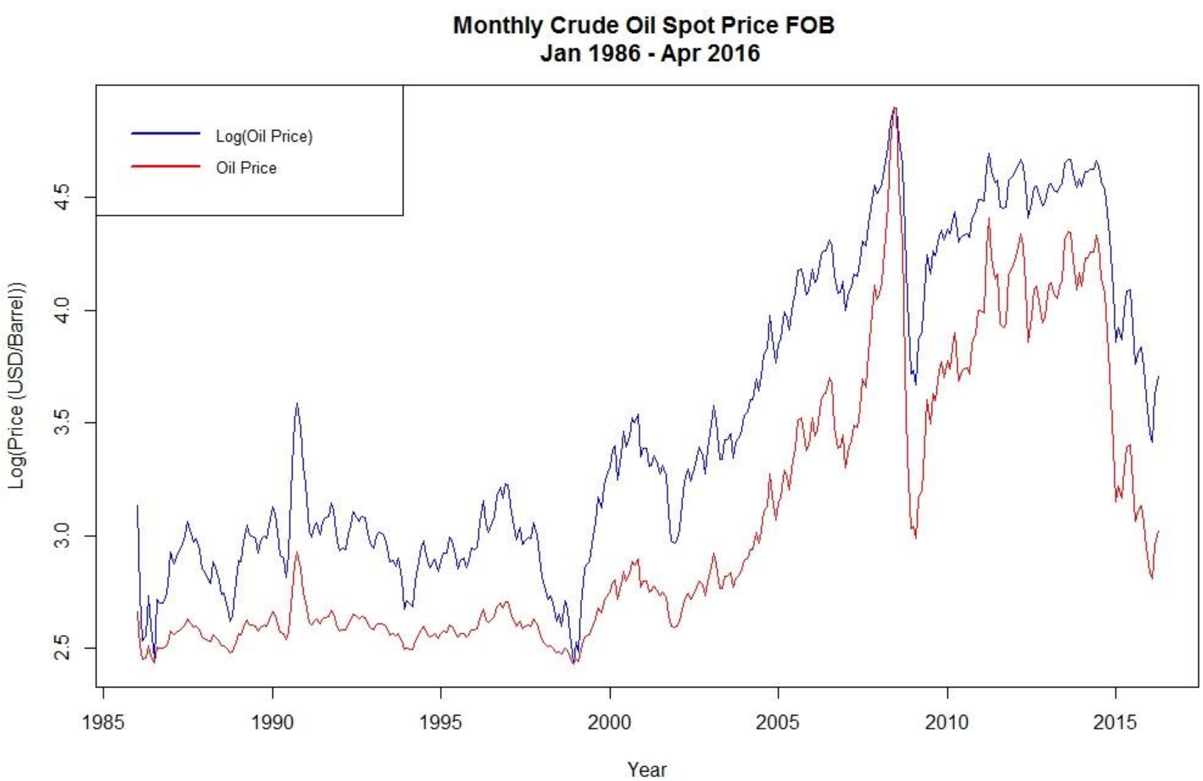
Comparing with the real Crude Oil FOB price of May 2016 (Source: www.eia.gov), we can see that the forecasting results is acceptable but not good enough.

- On 06 May 2016: 44.22 USD/Barrel
- On 13 May 2016: 45.44 USD/Barrel
- On 20 May 2016: 47.99 USD/Barrel

Conclusion and further improvements

Oil price as well as gold and stock price which are used as an investment instrument, are very sensitive, wide fluctuated and hard to predict. However, there are some drawing out lesion to improve the results of the forecasting as follow:

- Select a shorter period to avoid unnecessary fluctuation. In this case, we can drop the period from 1986 to 2009 to avoid the variance in oil price from 2006 to 2010, and the previous trend seems not to be correct anymore.
- Apply a transformation to reduce the variance of data. For this case, we can go with logarithm transformation. In the sample below, we can see that the transformation (blue line) has shrunk the variance between the period of 1985-2009 and the period of 2010-2016.



Dataset 2: Temperature of Washington DC 1900-2016

Introduction

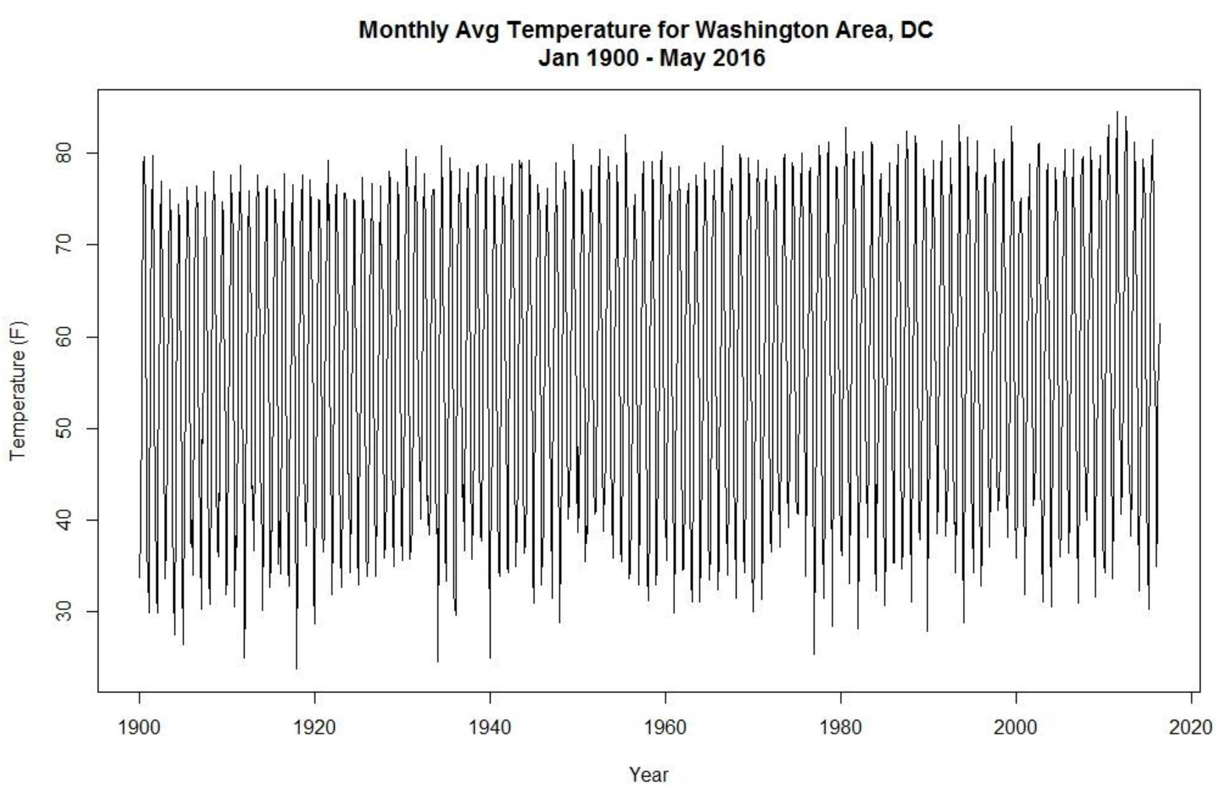
Global warming and climate change are the primary problems that mankind has to face nowadays. During the recent decade, we can feel the weather becomes hotter and hotter especially in tropical countries, e.g. Vietnam; in some European cites, like Lille, people have not seen snow for nearly 5 years. For each one degree increasing in temperature, global warming can cause many problems to every aspects of human life from traveling, agriculture, economics to flood, animal extinction, etc. Therefore, it is important to understand and forecast the changing of the weather.

Data description and exploration

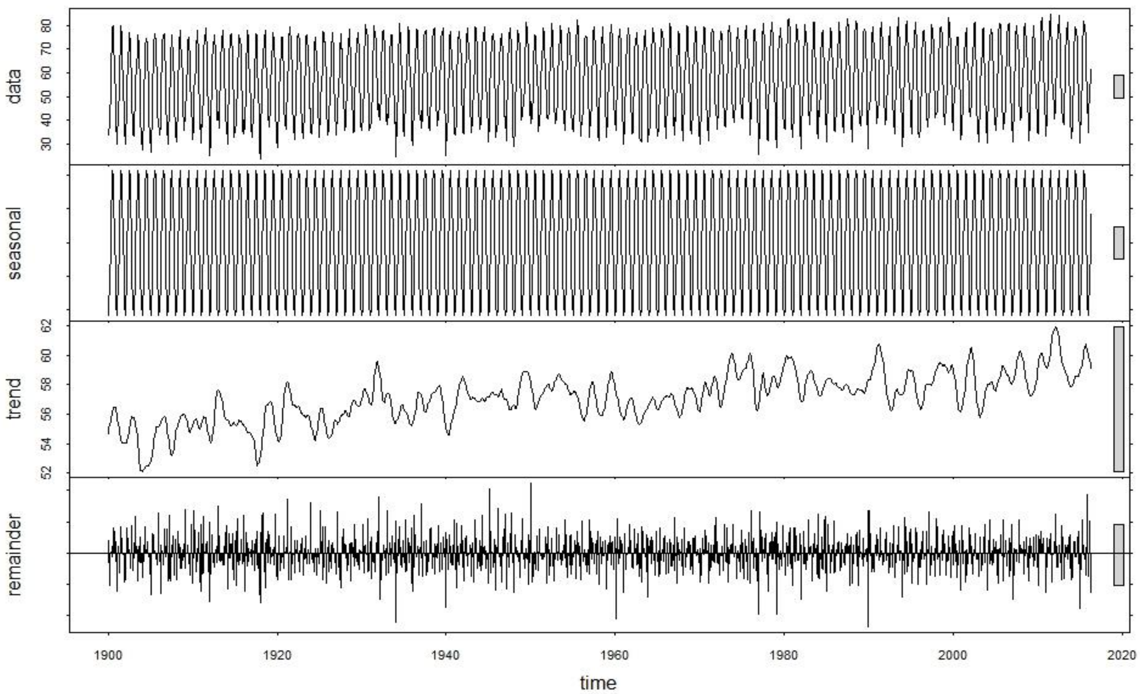
The data was provided by National Weather Service Forecast Office ("NOAA"), on their website: w2.weather.gov, the details are as follow:

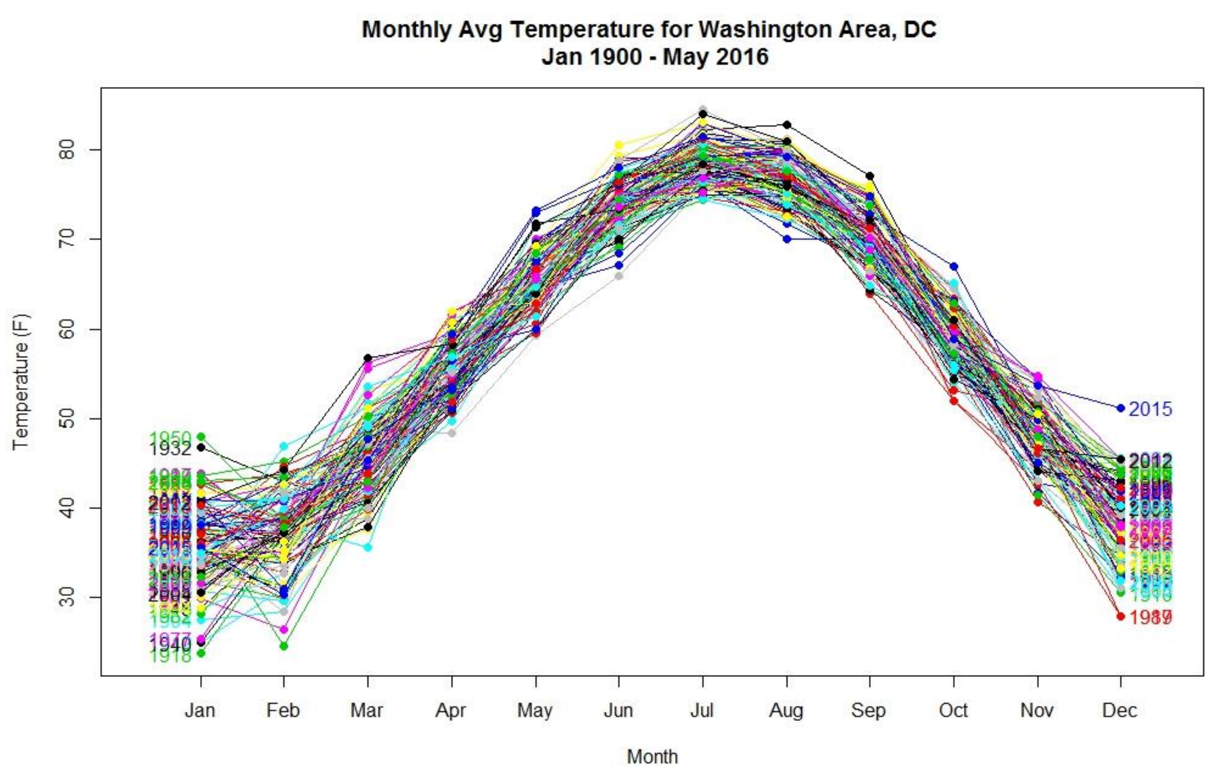
- Name of dataset: Monthly Mean Average Temperature for Washington Area, DC (Fahrenheit)
- Full period: Jan 1900 – May 2016 (1397 periods, 100%)
- Training set: Jan 1900 – Dec 1989 (1080 periods, 77%)
- Testing set: Jan 1990 – May 2016 (317 periods, 23%)

Plotting the original data to get the general idea about the trend and variation. We can see that there are lots of differences in temperature between months, however, they are quite stable through years.



To get more understanding about the components of the data, we used the STL decomposition function and seasonal plotting as follow. It is clear that the data contains a seasonal component and a slightly increasing trend over time. Then we can apply the ETS framework with all there parameters, i.e. error, trend and seasonal.



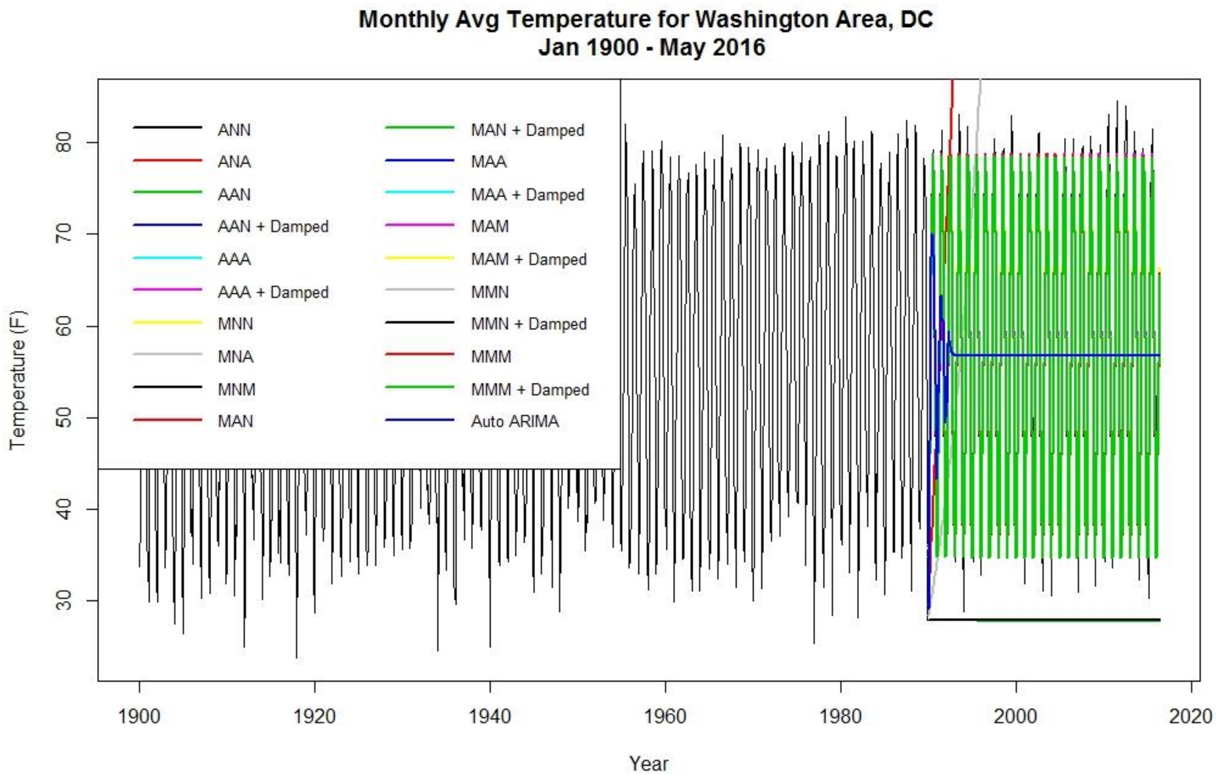


Choosing forecasting models

By applying a set of self-defined functions (the details in Appendix) to scan all 19 ETS possible models and Auto ARIMA model on the training and testing datasets, we have the accuracy measures as follow:

Model name	RMSE	MAE	MAPE	MASE	RMSE	MAE	MAPE	MASE
	Training set				*Testing set*			
ANN	8.900	7.603	14.744	2.302	34.253	30.581	48.462	9.260
ANA	3.081	2.359	4.952	0.714	3.339	2.653	5.115	0.803
AAN	7.158	5.730	11.310	1.735	107.203	105.904	187.551	32.069
AAN + Damped	7.158	5.730	11.310	1.735	107.203	105.904	187.551	32.069
AAA	3.078	2.356	4.941	0.713	3.305	2.626	5.077	0.795
AAA + Damped	3.078	2.356	4.941	0.713	3.305	2.626	5.077	0.795
MNN	8.900	7.604	14.746	2.303	34.253	30.581	48.462	9.260
MNA	3.095	2.372	4.978	0.718	3.403	2.728	5.215	0.826
MNM	3.136	2.414	5.044	0.731	3.425	2.732	5.232	0.827
MAN	9.099	7.690	15.227	2.329	284.397	240.447	451.054	72.811
MAN + Damped	8.924	7.622	14.838	2.308	34.357	30.697	48.677	9.296
MAA	3.090	2.363	4.958	0.715	3.434	2.744	5.237	0.831
MAA + Damped	3.090	2.363	4.958	0.715	3.434	2.744	5.237	0.831
MAM	3.144	2.433	5.068	0.737	3.476	2.792	5.310	0.845
MAM + Damped	3.144	2.433	5.068	0.737	3.476	2.792	5.310	0.845
MMN	9.020	7.683	15.014	2.327	1,265.515	766.155	1,419.359	232.004
MMN + Damped	8.895	7.601	14.750	2.302	34.260	30.589	48.477	9.263
MMM	3.164	2.444	5.108	0.740	3.471	2.766	5.288	0.837
MMM + Damped	3.160	2.436	5.082	0.738	3.438	2.730	5.243	0.827
Auto ARIMA	5.022	3.980	8.105	1.205	15.179	13.113	24.396	3.971

Plotting the results of 19 ETS models and 1 Auto ARIMA model:

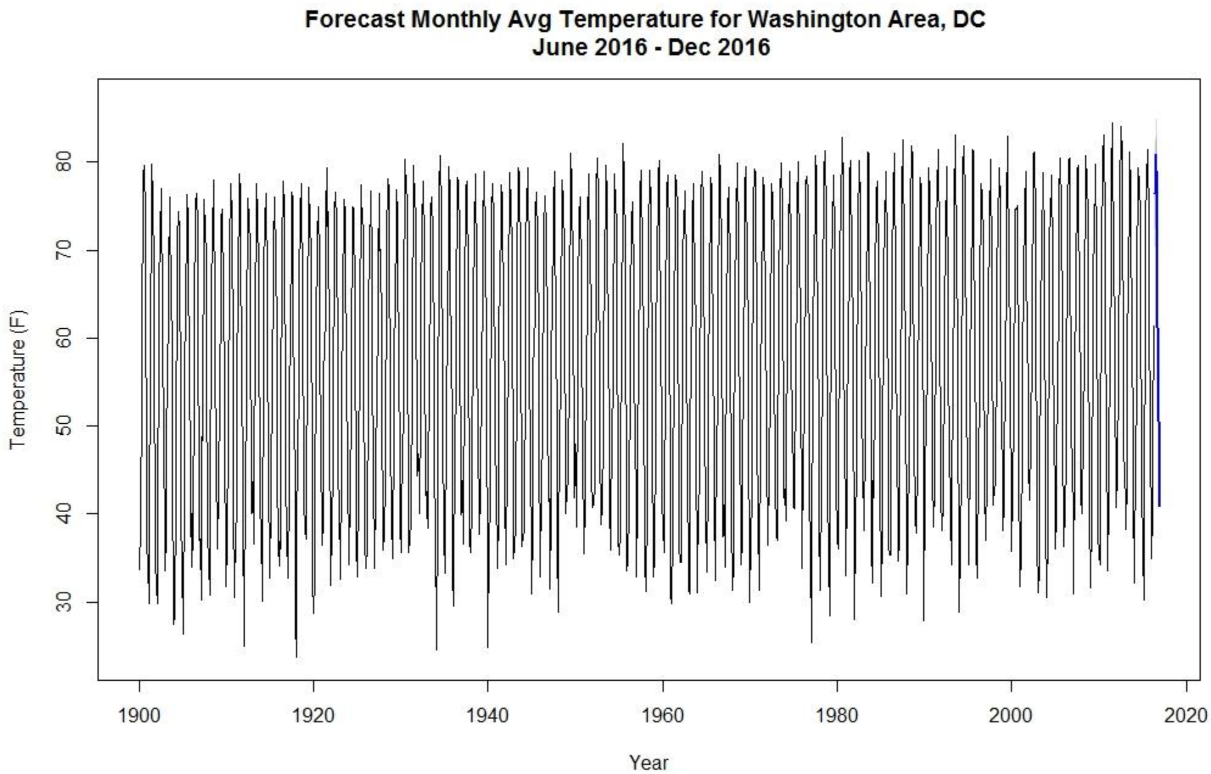


In overall, most of models work very well in this forecasting, but there are some few models performed very badly, e.g. ETS(A,A,N), ETS(A,Ad,N), ETS(M,A,N), ETS(M,Ad,N) and ETS(M,N,N). In term of RMSE and MASE measurements, it is crystal clear that ETS(A,A,A) and ETS(A,Ad,A) are best models with a very good accuracy result. We select ETS(A,Ad,A) as the final model to forecast the weather of the next few months.

Model summary:

ETS(A,Ad,A), alpha = 0.0273, beta = 1e-04, gamma = 1e-04, phi = 0.9621

Using forecasting model: ETS(A,Ad,A)



Forecast Monthly Average Temperature for Washington Area, DC for the next 7 months of 2016:

Month	Point.Forecast	Lo.80	Hi.80	Lo.95	Hi.95
Jun-16	76.43	72.49	80.38	70.40	82.47
Jul-16	80.87	76.92	84.82	74.83	86.90
Aug-16	79.04	75.10	82.99	73.01	85.08
Sep-16	72.70	68.75	76.65	66.66	78.75
Oct-16	61.18	57.23	65.13	55.13	67.22
Nov-16	50.57	46.62	54.52	44.53	56.62
Dec-16	40.84	36.88	44.79	34.79	46.89

Conclusion and further improvements

Not like oil, gold or stock price, weather temperature time series data in Washington DC is very stable and easy to forecast with high accuracy. There is only one improvement for this analysis is to select a shorter period that will make the plotting easier to be observed (currently is 1397 periods).

Appendix

The set of self-defined forecasting functions:

Name/Syntax	Function
ETS.all (data.ts, error, trend, season)	Scan all possible ETS methods, return a list of trained ETS models
model.all (data.ts, error, trend, season)	Combine ETS.all(...) and Auto ARIMA, return a list of trained models
forecast.all (model.list, h)	Apply the whole list of models to forecast data, return a list of forecasting objects
accuracy.fit (model.list)	Evaluate the accuracy of models on training datasets, return a data.frame of RMSE, MAE, MAPE, MASE values for each forecast model
accuracy.fc (data.ts, forecast.list)	Evaluate the accurary of models on testing datasets, return a data.frame of RMSE, MAE, MAPE, MASE values for each forecast model
acc.rank (acc.table)	Rank the accurary tables to provide a guide to select the best model