

Automatic Forecasting via R

Brian O'Horo

This paper illustrates the practical application of automatic forecasting via the R language. For the sake of this paper, forecasting entails predicting the future or simply making statements about events where the outcome has yet to be observed.

Not every phenomenon is subject to forecasting. Generally speaking, this approach applies to systems where we have good historical data and our expectations of the consistency of change within a given system, while potentially unknown, can be modeled [1].

For illustration of forecasting we will be using the dataset “Electricity monthly total net generation January 1973-October 2010” from the US Energy Information Administration as packaged by Rob J. Hyndman in his “fpp” R package.

Two questions we will be looking to answer with this dataset. One, is there a seasonal component to monthly electric net generation? Two, can we forecast future unobserved data points?

Getting the Data

Bringing this dataset into the R environment is as follows:

```
# Data for "Forecasting: principles and practice"
library(fpp, quietly = TRUE, verbose = FALSE)
```

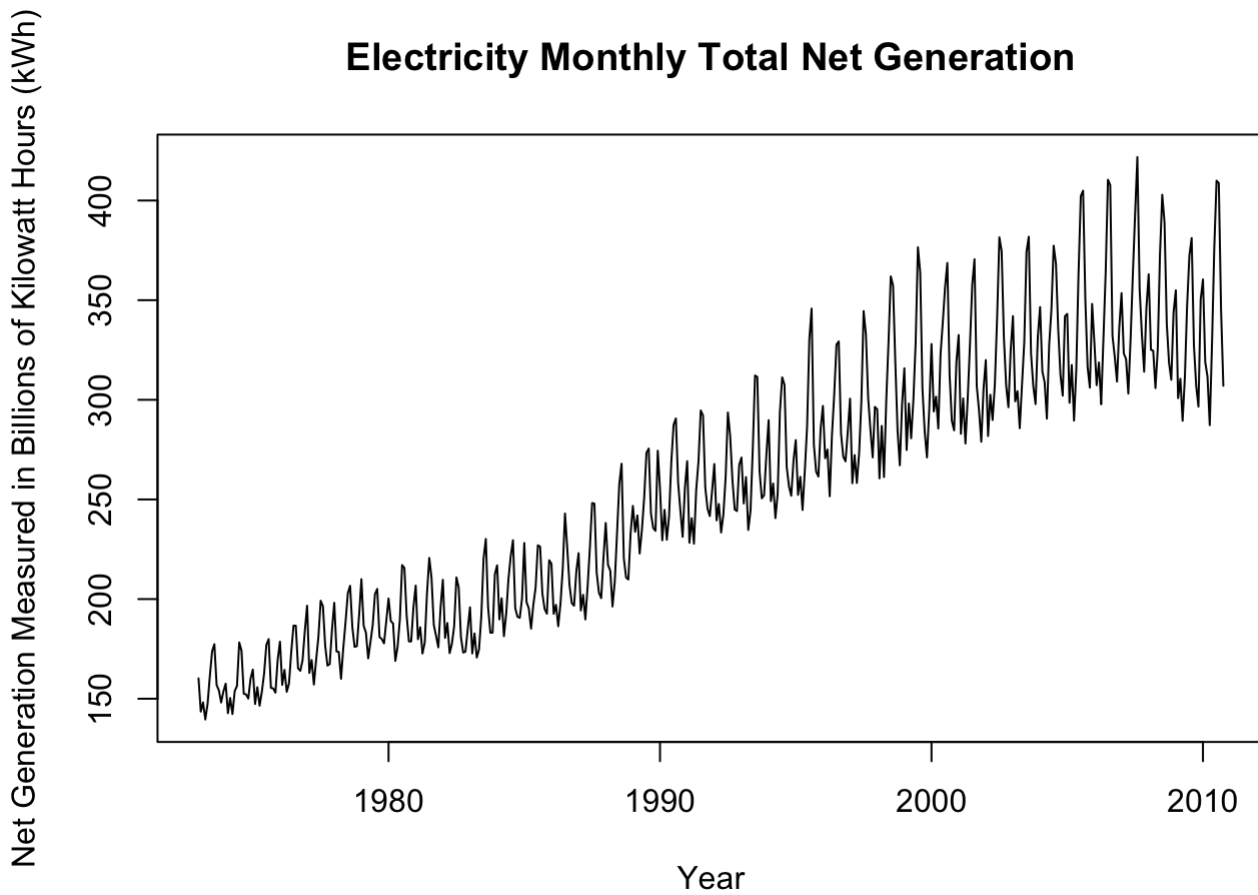
```
## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
##
## Loading required package: timeDate
## This is forecast 5.4
##
## Loading required package: tseries
```

```
# Electricity net generation measured in billions of kilowatt hours (kWh)
data(usmelec)
```

Data Exploration

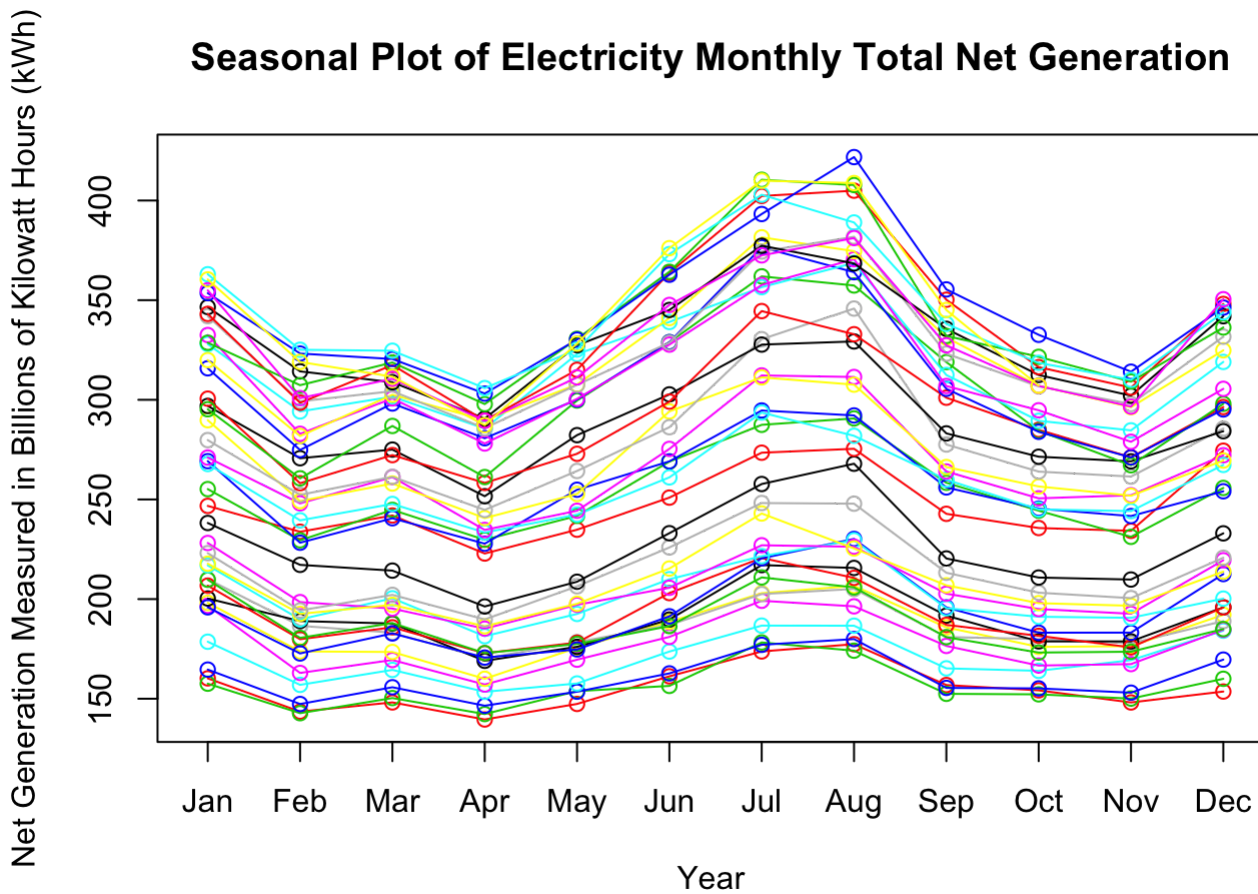
Our first step will be to do some exploration of the data. We start with a simple time series plot in R. We see an upward trend, which has leveled off in recent years, that stair steps. This sets my initial intuition that we had a trend that has perhaps leveled off with a seasonal component that can be modeled.

Electricity Monthly Total Net Generation



Let's dive a little more into the seasonal aspect of the data via the following plot. Viewing the data this way we see spikes of usage near August and a smaller spike in the December/January timeframe.

Seasonal Plot of Electricity Monthly Total Net Generation



Forecasting

Now let's fit a forecast model. To gauge our accuracy we will create a training and test set. In our case, we will use 1973 to 2009 as our training data and attempt to forecast 2010 data. We will hold out 2010 data so we can see how well the forecast performs.

```
uselecTrain <- window(usmelec, 1973, 2009)
uselecTest  <- window(usmelec, 2009, 2010)
```

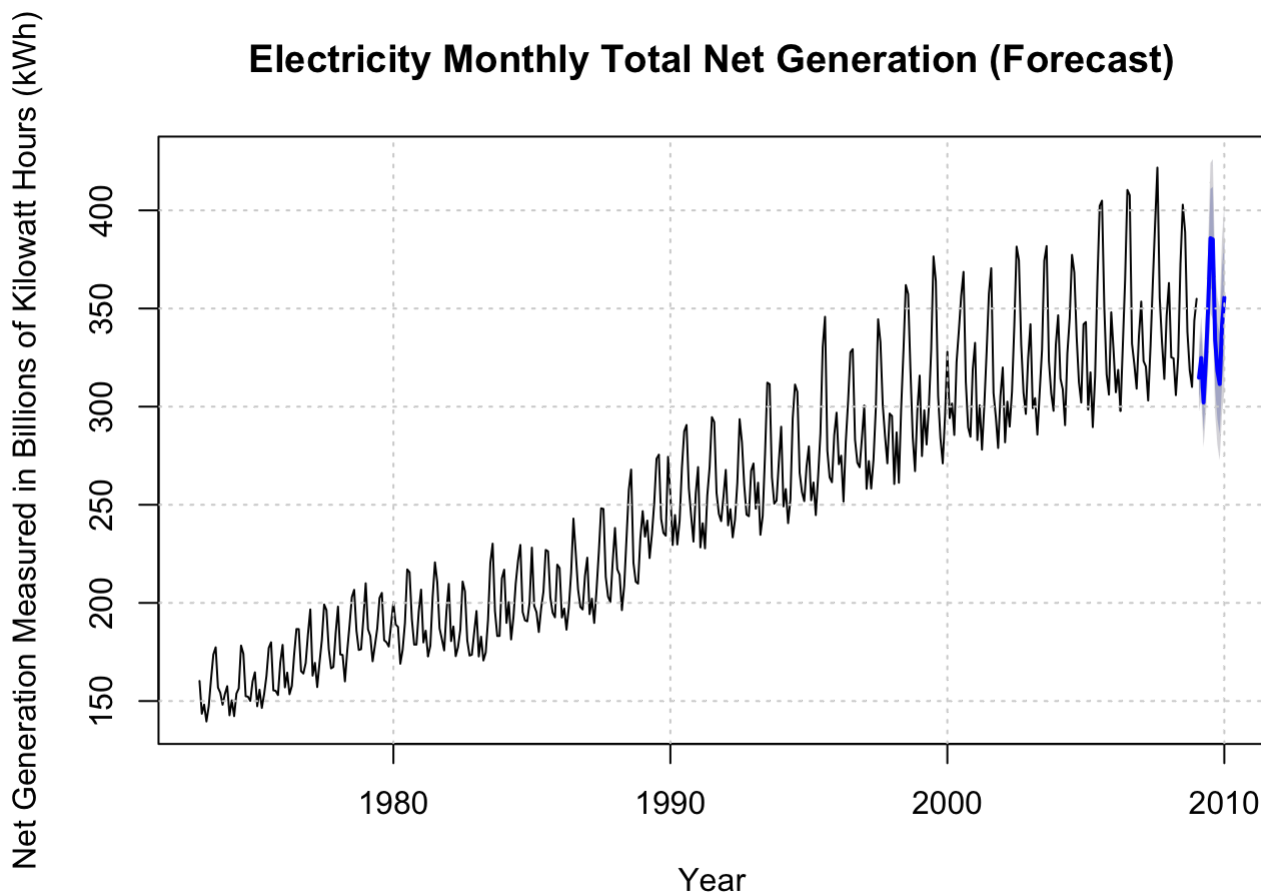
We now will invoke the fully automated forecast method that will fit the best automated forecast to the data out of 30 possible forecast models. We will also tell the forecast model to forecast or predict the next 12 periods (2010).

```
# fully automated forecasting
fcast <- forecast(uselecTrain, h=12)
sprintf("The method selected for the data is: %s", fcast$method)
```

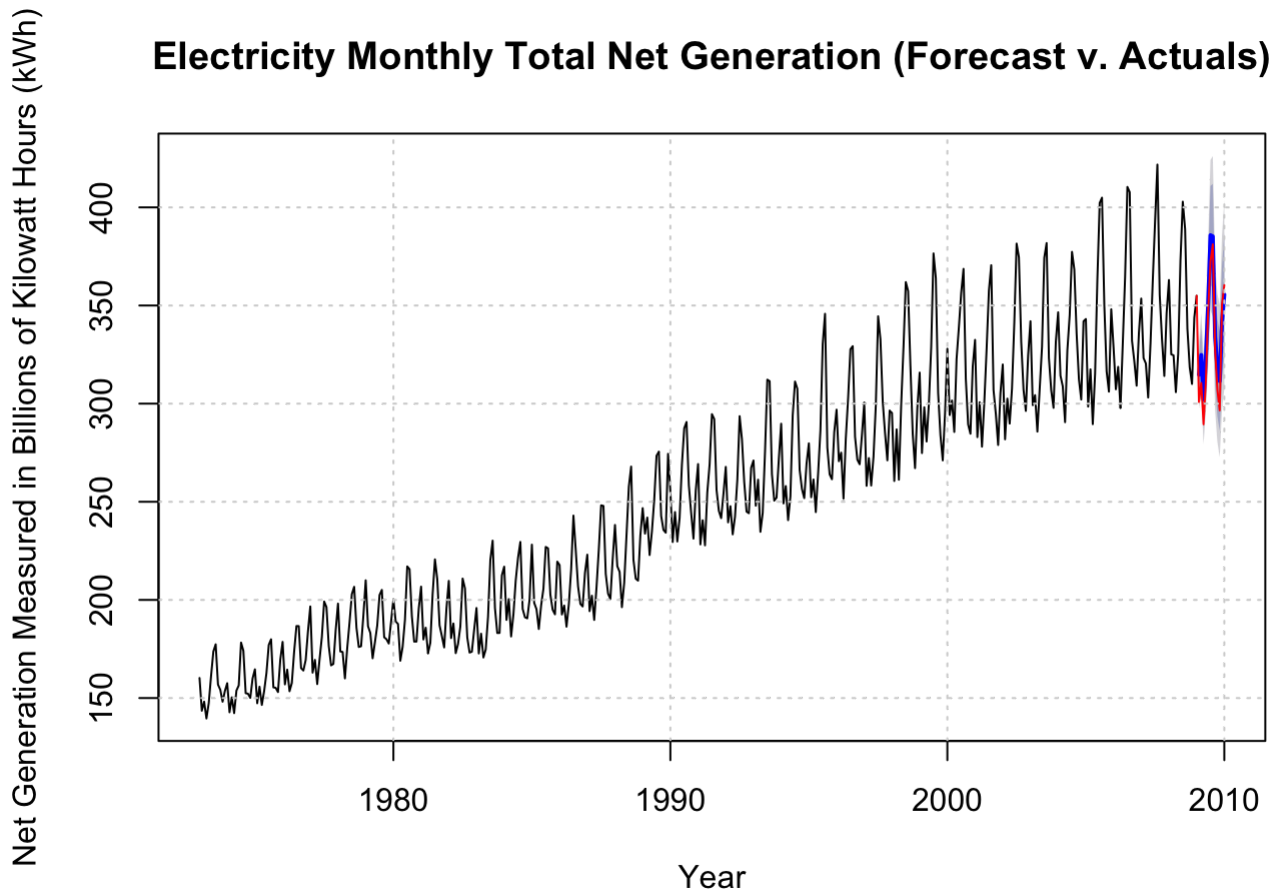
```
## [1] "The method selected for the data is: ETS(M,N,M)"
```

We see from the model printed above, the forecast method has selected an Exponential Smoothing State Space model with error type of *multiplicative*, trend type of *none*, and seasonal type of *multiplicative*. This matches our original intuition of a trend (none) that has leveled off with a seasonal component (multiplicative).

Now let's see a plot of the forecast. The blue line is the point forecast (mean of forecast) with the gray areas indicating the confidence intervals. We can see that the forecast appears to have done a reasonable job in the forecast. Specifically, it appears to have extrapolated the visual pattern for the observed data.



Now let's overlay the real 2010 data. To do this we will simply add a red line over the blue/gray forecast. As we can see, they are a very good match. The automated forecast was a good fit to the actual observed values.



Conclusion

This paper has generally shown the practical application of automated forecasting within R. Specifically, for our example we confirmed there was a seasonal component to monthly electric net generation and we could reasonably forecast future unobserved data points.

References

1. Hyndman and Athanasopoulos (2014), Forecasting: principles and practice, OTexts.
2. Forecast Package (<http://cran.r-project.org/web/packages/forecast/index.html> (<http://cran.r-project.org/web/packages/forecast/index.html>))
3. My github at <https://github.com/bohoro/UnivariateForecastingWithTimeseries> (<https://github.com/bohoro/UnivariateForecastingWithTimeseries>)