Seasonal Models

Nik Bear Brown

In this lesson we'll learn the how to implement Seasonal Models in R.

# Additional packages needed

To run the code you may need additional packages.

* If necessary install the followings packages.

install.packages("ggplot2");  
install.packages("forecast");  
source(url("http://lib.stat.cmu.edu/general/tsa2/Rcode/itall.R"))

library(ggplot2)  
library(forecast)

## 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 7.3

source(url("http://lib.stat.cmu.edu/general/tsa2/Rcode/itall.R"))

## itall has been installed

# Data

We will be using U.S. Department of the Interior monthly flow data for the Colorado River. see <http://www.usbr.gov/lc/region/g4000/NaturalFlow/documentation.html>

data\_url <- 'http://nikbearbrown.com/YouTube/MachineLearning/M11/colorado\_river.csv'  
data<-read.csv(url(data\_url))  
head(data)

## month year flow  
## 1 1 71 26635.42  
## 2 2 71 21825.23  
## 3 3 71 36031.57  
## 4 4 71 43874.54  
## 5 5 71 90616.44  
## 6 6 71 146125.61

names(data)

## [1] "month" "year" "flow"

# Seasonal Models

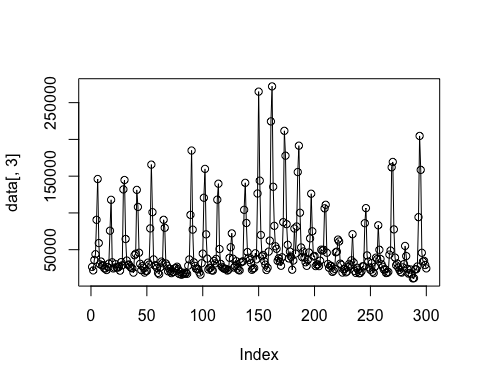
When there are patterns that repeat over known, fixed periods of time (i.e. day, week, month, quarter, year, etc.) within the data set it is considered to be seasonal variation. One has a model for the periodic fluctuations based on knowledge of the domain.

The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. In a seasonal ARIMA model, seasonal parameters predict xt using data values and errors at times with lags that are multiples of S (the span of the seasonality). Before we model for a given data set, one must have an initial guess about the data generation process, that is the span of the seasonality (i.e.day, week, month, quarter, year, etc.)

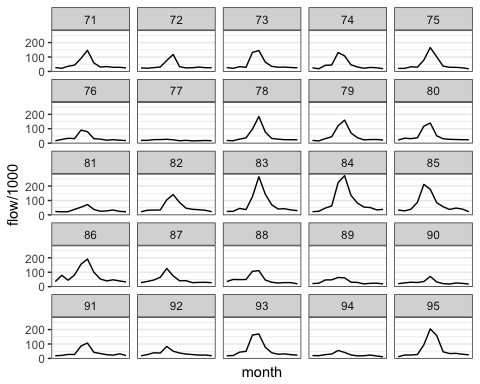
# Seasonal Models in R

Here we are using an ARIMA model to identify seasonality trends by looking for signficant seasonal differences.

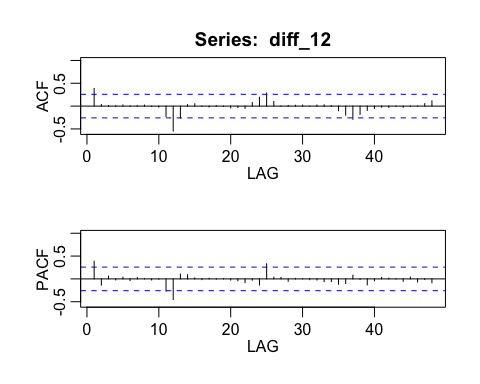
plot(data[,3], type='o')



# it is difficult to identify seasonality trends here  
# So,aggregate the data by month to better understand this trend  
ggplot(data, aes(x=month, y=flow/1000))+  
 stat\_summary(geom = 'line', fun.y='mean')+ # take the mean of each month  
 scale\_x\_discrete(breaks=seq(1,12,1), labels=seq(1,12,1))+  
 theme\_bw()+ # add a little style  
 facet\_wrap(~year) # visualize year by year



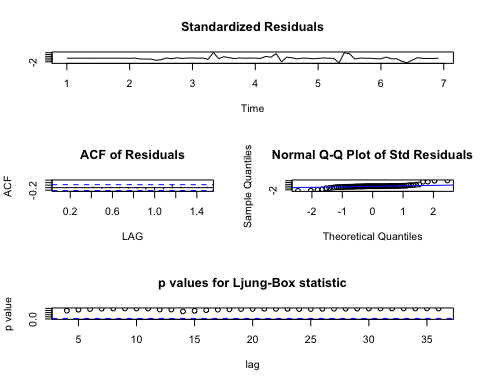
#Since we hypothesize that there is seasonality,  
#we can take the seasonal difference (create a variable that gives the 12TH differences), then look at the ACF and PACF.  
mydata<-ts(data[1:72,][,3])  
diff\_12 <- diff(mydata, 12)  
acf2(diff\_12, 48)



## ACF PACF  
## [1,] 0.39 0.39  
## [2,] 0.03 -0.14  
## [3,] 0.01 0.06  
## [4,] 0.01 -0.02  
## [5,] 0.02 0.04  
## [6,] -0.01 -0.04  
## [7,] 0.01 0.03  
## [8,] 0.02 0.01  
## [9,] -0.01 -0.02  
## [10,] -0.02 -0.01  
## [11,] -0.23 -0.27  
## [12,] -0.55 -0.45  
## [13,] -0.26 0.11  
## [14,] 0.03 0.09  
## [15,] 0.05 0.02  
## [16,] 0.00 -0.02  
## [17,] -0.02 0.01  
## [18,] 0.01 0.00  
## [19,] -0.01 0.00  
## [20,] -0.04 -0.03  
## [21,] -0.03 -0.04  
## [22,] -0.05 -0.08  
## [23,] 0.08 -0.03  
## [24,] 0.19 -0.14  
## [25,] 0.29 0.33  
## [26,] 0.10 0.04  
## [27,] 0.00 0.03  
## [28,] 0.01 -0.05  
## [29,] 0.02 0.01  
## [30,] 0.02 0.00  
## [31,] 0.00 -0.02  
## [32,] 0.02 -0.02  
## [33,] 0.03 -0.06  
## [34,] 0.01 -0.06  
## [35,] -0.10 -0.12  
## [36,] -0.20 -0.10  
## [37,] -0.29 0.08  
## [38,] -0.18 -0.01  
## [39,] -0.09 -0.13  
## [40,] -0.05 -0.04  
## [41,] -0.02 0.03  
## [42,] -0.03 0.01  
## [43,] -0.01 0.00  
## [44,] -0.03 -0.05  
## [45,] 0.00 0.04  
## [46,] 0.00 -0.07  
## [47,] 0.05 -0.01  
## [48,] 0.11 -0.09

#we see that for both the ACF and PACF we have significant autocorrelation at seasonal (12, 24, 36) lags. The ACF has a cluster around 12,   
#and not much else besides a tapering pattern throughout. Further, the PACF also has spikes on two multiples of S, AR(2)  
# Try, ARIMA (1,0, 0) x (2, 1, 0)12  
mydata<-ts(mydata, freq=12)  
mod1<-sarima(mydata, 1,0,0,2,1,0,12)

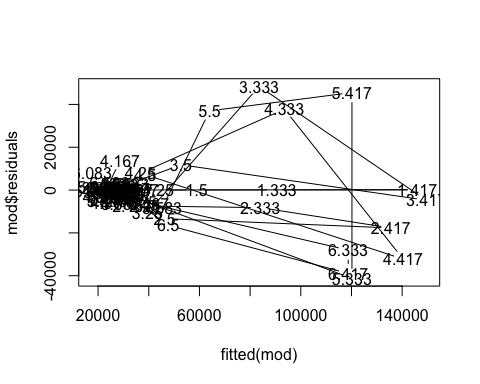
## initial value 10.165417   
## iter 2 value 9.844064  
## iter 3 value 9.816870  
## iter 4 value 9.779773  
## iter 5 value 9.769839  
## iter 6 value 9.767267  
## iter 7 value 9.766853  
## iter 8 value 9.766852  
## iter 9 value 9.766852  
## iter 9 value 9.766852  
## iter 9 value 9.766852  
## final value 9.766852   
## converged  
## initial value 9.722784   
## iter 2 value 9.719603  
## iter 3 value 9.711661  
## iter 4 value 9.709954  
## iter 5 value 9.709280  
## iter 6 value 9.709266  
## iter 7 value 9.709265  
## iter 7 value 9.709265  
## iter 7 value 9.709265  
## final value 9.709265   
## converged



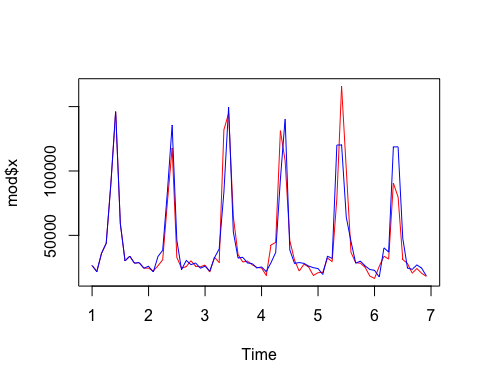
# install.packages("forecast")  
# Fit the Model  
mod<-Arima(mydata,order=c(1, 0, 0),  
 seasonal=list(order=c(2, 1, 0), period=12))  
mod

## Series: mydata   
## ARIMA(1,0,0)(2,1,0)[12]   
##   
## Coefficients:  
## ar1 sar1 sar2  
## 0.2806 -0.7924 -0.1932  
## s.e. 0.1336 0.1323 0.1676  
##   
## sigma^2 estimated as 250742540: log likelihood=-667.78  
## AIC=1343.57 AICc=1344.29 BIC=1351.94

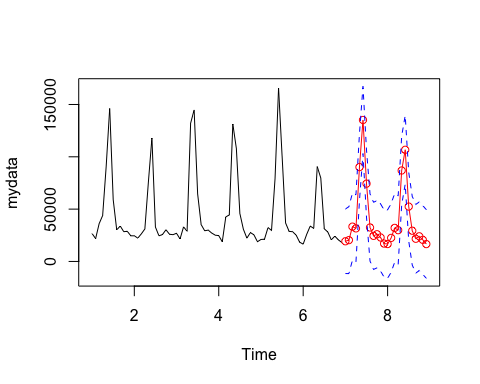
plot(fitted(mod), mod$residuals)



plot(mod$x, col='red')  
lines(fitted(mod), col='blue')



# Now, we have a reasonable prediction, we can forecast the model, say 24 months into the future.  
sarima.for(mydata, 24, 1,0,0,2,1,0,12)



## $pred  
## Jan Feb Mar Apr May Jun Jul  
## 7 19280.41 20375.61 33258.81 31599.55 90175.30 135176.30 74346.53  
## 8 16643.94 22493.80 32055.86 29896.44 86925.76 106695.30 52426.14  
## Aug Sep Oct Nov Dec  
## 7 32455.48 24461.92 26000.24 22917.53 17066.97  
## 8 29367.63 21679.08 24035.64 20627.54 16596.45  
##   
## $se  
## Jan Feb Mar Apr May Jun Jul Aug  
## 7 15418.48 16002.65 16046.86 16050.27 16050.54 16050.56 16050.56 16050.56  
## 8 16374.13 16398.84 16400.75 16400.90 16400.91 16400.91 16400.91 16400.91  
## Sep Oct Nov Dec  
## 7 16050.56 16050.56 16050.56 16050.56  
## 8 16400.91 16400.91 16400.91 16400.91

predict(mod, n.ahead=24)

## $pred  
## Jan Feb Mar Apr May Jun Jul  
## 7 20255.41 21600.16 34566.44 32926.83 91479.84 136741.26 75871.08  
## 8 18181.32 24104.77 33671.14 31514.60 88541.37 108213.01 53968.18  
## Aug Sep Oct Nov Dec  
## 7 33820.25 25825.51 27353.24 24271.39 18409.99  
## 8 30981.40 23293.94 25653.80 22244.57 18220.17  
##   
## $se  
## Jan Feb Mar Apr May Jun Jul Aug  
## 7 15834.85 16446.23 16493.39 16497.10 16497.39 16497.41 16497.41 16497.41  
## 8 16821.60 16846.86 16848.84 16849.00 16849.01 16849.01 16849.01 16849.01  
## Sep Oct Nov Dec  
## 7 16497.41 16497.41 16497.41 16497.41  
## 8 16849.01 16849.01 16849.01 16849.01

# Resources

* [Seasonal ARIMA models | STAT 510](https://onlinecourses.science.psu.edu/stat510/node/67)
* [Identifying Seasonal Models and R Code | STAT 510](https://onlinecourses.science.psu.edu/stat510/node/68)
* [Seasonal ARIMA models | OTexts](https://www.otexts.org/fpp/8/9)
* [General seasonal ARIMA models](http://people.duke.edu/~rnau/seasarim.htm)
* [Modeling Seasonal Time Series - Hu-berlin.de](http://fedc.wiwi.hu-berlin.de/xplore/tutorials/xegbohtmlnode44.html)

# References

The data, R code and lessons are based upon:

1. Time Series Analysis :

Data Source: <http://www.geophysics.geol.uoa.gr/catalog/catgr_20002008.epi>

Code References :

Book : Mastering Predictive Analytic with R  
Author: Rui Miguel Forte  
<https://www.safaribooksonline.com/library/view/mastering-predictive-analytics/9781783982806/>

Chapter 9: Time series Analysis

<http://www.statoek.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer03/ts_r_intro.pdf>

<http://www.stat.pitt.edu/stoffer/tsa3/R_toot.htm>

<http://www.statoek.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer03/ts_r_intro.pdf>

1. Trend Analysis

Code References :

Book : Mastering Predictive Analytic with R  
Author: Rui Miguel Forte  
<https://www.safaribooksonline.com/library/view/mastering-predictive-analytics/9781783982806/>

<http://www.r-bloggers.com/seasonal-trend-decomposition-in-r/>

1. Seasonal Models

Code references :

Book: Time Series Analysis and Its Applications  
Author: Robert H. Shumway . David S. Stoffer  
Link: <http://www.springer.com/us/book/9781441978646#otherversion=9781461427599>

<http://a-little-book-of-r-for-time-series.readthedocs.org/en/latest/src/timeseries.html>

<https://onlinecourses.science.psu.edu/stat510/?q=node/47>

<https://rpubs.com/ryankelly/tsa5>

<https://onlinecourses.science.psu.edu/stat510/node/68>

Data Reference : <https://github.com/RMDK/TimeSeriesAnalysis/blob/master/colorado_river.csv>

1. Spectral Analysis

Code References:  
Book:  
Modern Applied Statistics with S Fourth edition  
Author: W. N. Venables and B. D. Ripley  
Link: Modern Applied Statistics with S Fourth edition

<http://www.maths.adelaide.edu.au/patty.solomon/TS2004/tsprac3_2004.pdf>