Time Series Analysis

Kevin Smith | Big Data Analytics | October 27th, 2014

Overview

This project explores basic time series analysis techniques for two datasets. The first dataset is a complete record of the mean monthly flows of the Ganges from January 1934 to December 2013. The second dataset is a complete record of the mean monthly flows of the Nile (collected at Aswan Dam) from January 1871 to December 1984.

The Ganges

Monthly Flow

Time Series

A plot of the monthly time series reveals seasonality in the data. This structure will need to be removed before we can assess the stationarity of the time series.

Figure 1: 79 years of mean monthly flow data for the Ganges in real space (top) and log space (bottom).

Boxplots

Within-year seasonality is highlighted in this boxplot of monthly mean flows. During the year there seems to be 'wet' and 'dry' seasons of differing lengths. These seasons will be independently analyzed in sections 3 and 4.

Figure 2: A series of boxplots showing the center and spread of the dataset by month. The signal appears highly seasonal.

Sample Statistics

	Mean (CFS)	SD (CFS)	CV
January	2537	1124	0.443
February	2014	843.5	0.4189
March	1663	765.6	0.4603
April	1519	648.7	0.427
May	1758	574.2	0.3266
June	3988	1618	0.4057
July	18760	5860	0.3123
August	37200	8042	0.2162
September	35990	8399	0.2334
October	17370	6489	0.3735
November	6505	2291	0.3522
December	3762	1657	0.4406
Overall	**11090**	**13460**	**1.214**

Initial Stationarity Assessment

The monthly time series is not even first-order stationary becuase the mean exhibits strong seasonality. However, the highly seasonal structure suggests that we may be able to make the time series stationary though differencing.

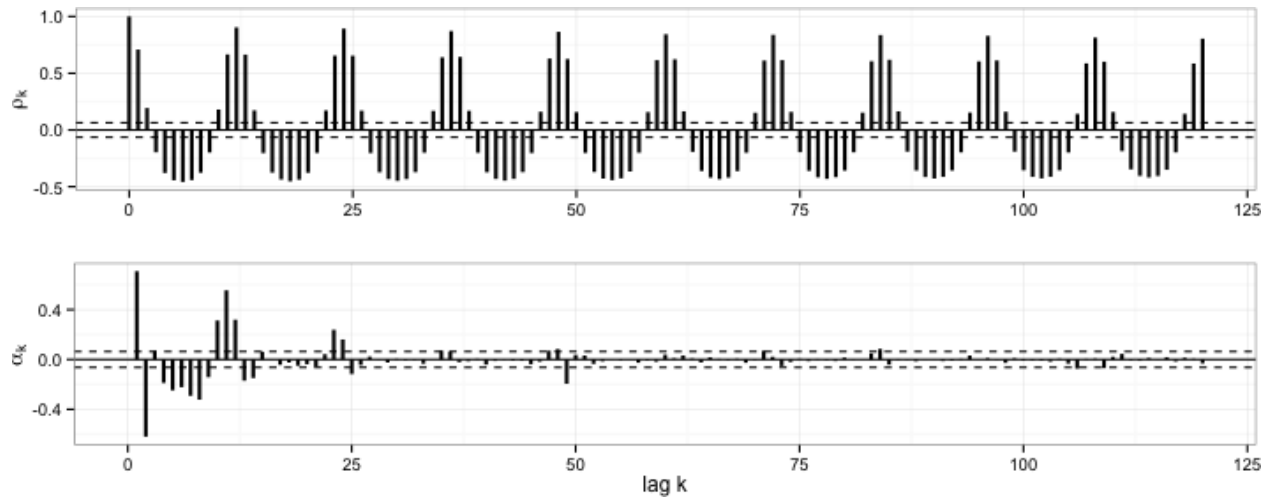Figure 3: Ganges monthly mean flows with $i+12$ observations highlighted.

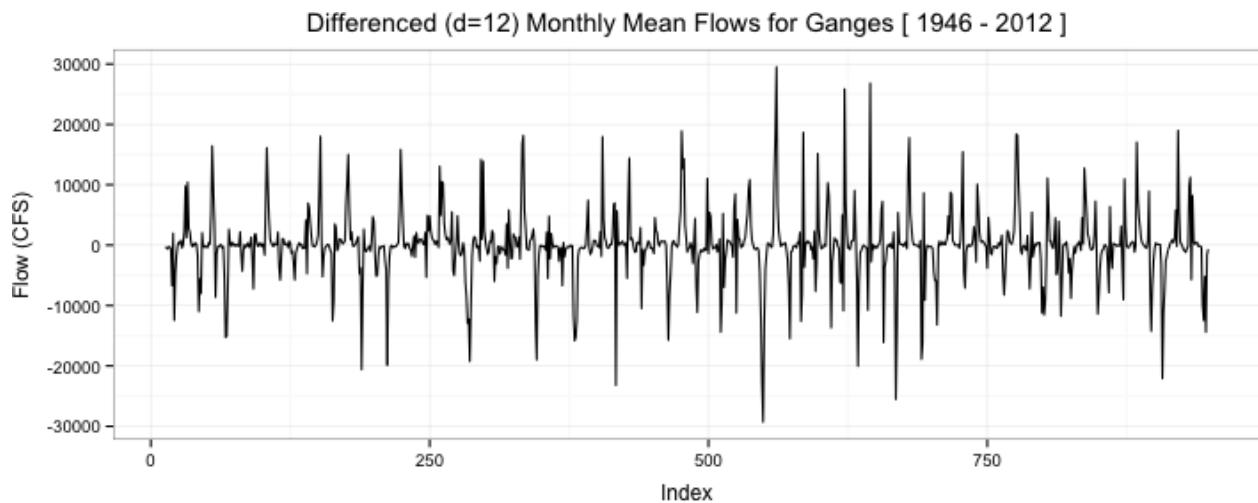Figure 4: Total and partial correlograms.

Differencing

Figure 5: Differenced (d=12) Monthly Mean Flow Series

```
q1 = ggacf(tsd$value) + xlab(" ") + theme(plot.margin=unit(c(2,2,2,2),"mm"))
q2 = ggpacf(tsd$value) + theme(plot.margin=unit(c(2,2,2,2),"mm"))

grid.arrange(q1, q2, nrow = 2)
```
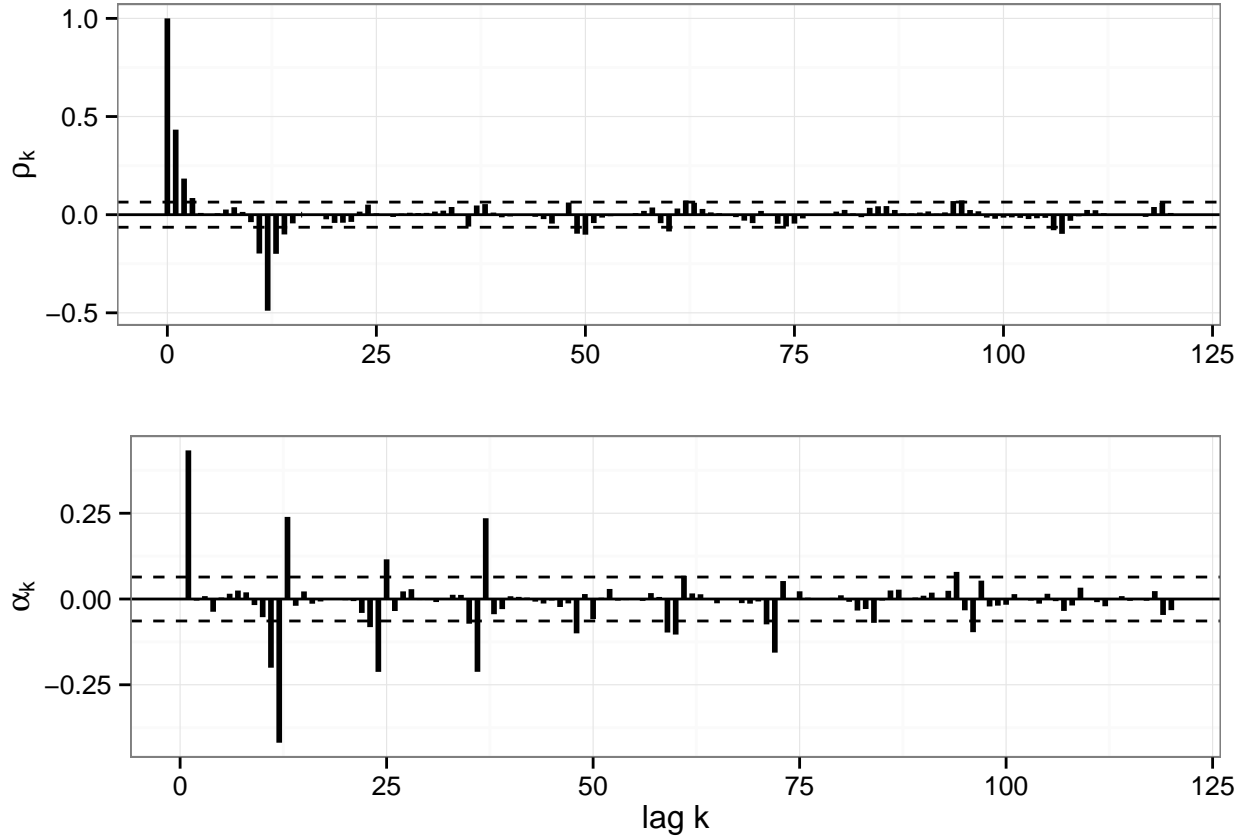
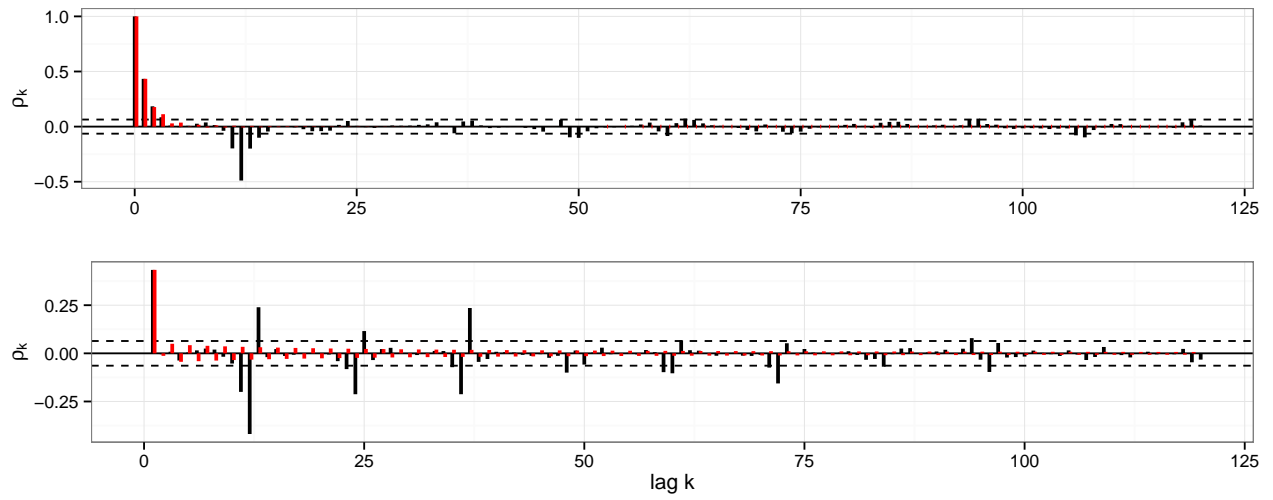Figure 6: Total and partial correlograms for the differenced (d=12) series.

Figure 1: Differenced (d=12) Monthly Mean Flow Series

3