

Time Series Analysis of Average Annual River Flow of the Nile River at Aswan, Egypt from 1870-1945

Megan Pagliaro & Dylan Sosa

December 10, 2017

1 Introduction

The Nile River (Fig. 1) is and has been a critical natural resource for the populations of Egypt and its surrounding areas for thousands of years. With that said, increasing human populations and evolving modern technologies have greatly impacted the river's ecology. Landscape modification, topography, and climate are just a few of the factors that can affect river flow rate. This study uses time series analysis of hydrological data[2] collected at Aswan, Egypt to analyze the likely anthropogenic effect on the Nile River's flow rate from 1870 to 1945. We predict that the average annual river flow rate will be negatively impacted over time mostly likely driven by increases in human population sizes and colonization of the riverbed.



Figure 1: Google Earth image of present-day Nile River at Aswan, Egypt.

2 Methods

2.1 Exploratory Analysis

We began our study by obtaining average annual river flow data from the online data library, Data Market[1]. Average annual river flow, which is synonymous with stream discharge rate, measures the average rate of water flowing through the Nile. The unit of measurement is cubic meters per second (m^3/s). First, we plotted our raw dataset (Fig. 2) to look for any trends indicating non-stationarity. We noticed a non-seasonal, yet overall decreasing pattern. The spikes and troughs are indicative of years that experienced

flooding or droughts, respectively. The autocorrelation function (ACF) also indicated non-stationarity due to the slowly decreasing peaks (App. Fig. 6). Because of the necessity of stationary data in time series analysis, we first ran a Box-Cox test to determine what value of λ would be appropriate to transform our data (App. Fig. 7). This resulted in a λ value near zero, meaning an initial *log* transformation was appropriate.

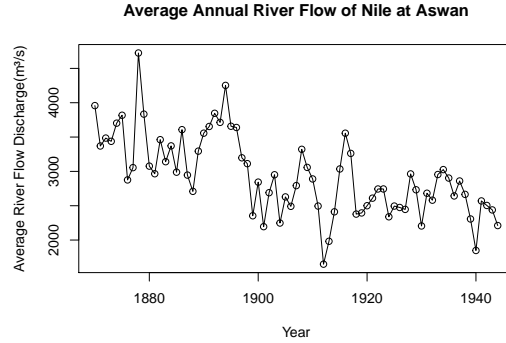


Figure 2: Plot of raw Nile River data.

2.2 Transformation

2.2.1 Log Transformation

The resulting plot of the *log* transformed data was nearly identical to the raw data plot (App. Fig. 8). An ACF was ran and this showed a slowly decreasing trend, indicating non-stationarity (App. Fig. 9). To confirm, an Augmented Dickey-Fuller Test was ran. This resulted in a non-significant p-value of 0.065, meaning the data failed to reject the test's null hypothesis and was non-stationary. The above results suggested the need to further transform the data using a first-order difference.

2.2.2 First-Order Difference

This second transformation involved taking the first-order difference of the *log* transformed data. The corresponding plot (App. Fig. 10) had some odd variance around the mean, but appeared to be more stationary overall and the Dickey-Fuller Test resulted in a significant p-value of < 0.01 . However, when we ran ACF and PACF tests on this data the output showed no significant peak also known as a borderline non-stationary pattern (App. Fig. 11). These results are indicative of the necessity to do further transformation through a second-order difference.

2.2.3 Second-Order Difference

Our final transformation was a second-order difference of the *log* transformed data. The resulting plot (Fig. 3) showed stationarity with fairly even variance around a mean of 0.00 as indicated by the red line. A significant p-value of < 0.01 from the Dickey-Fuller Test meant that this transformation rejected the null hypothesis and the data was stationary. The presence of significant peaks in both the ACF and PACF plots (unlike the first-order difference) indicated that no further transformations were necessary.

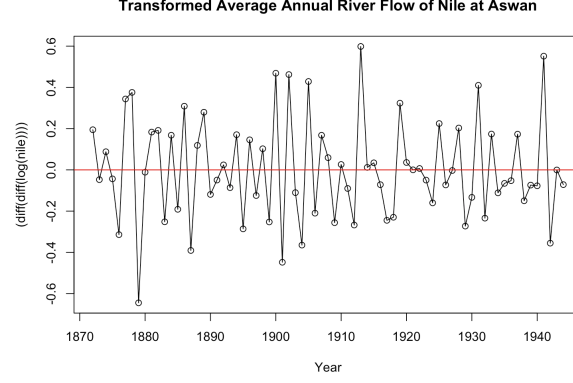


Figure 3: Plot of second-order, log transformed Nile River data.

3 Results

3.1 Model Selection

After we made the data stationary with the second-order difference, we began to determine a proper model that described the data. A single negative peak in the ACF plot (circled in red) and a decaying pattern from below in the PACF plot (marked by a red line) were signatures of a moving average one process (Fig. 4). This reasoning suggested an $ARIMA(0,2,1)$ model which is less common, but seen in special cases such as linear exponential smoothing[3]. Next, we utilized an ARMA subset plot as another tool for model selection (App. Fig. 13). This resulted in multiple rows with the same significant Bayesian Information Criterion (BIC) value which is a model selection standard. Because of this we made our selection according to the principle of maximum parsimony which states that the least complicated explanation is the most likely to be correct. With this in mind, we selected the row that indicated $AR(0)$ and $MA(1)$ processes. Next we used the extended autocorrelation function (EACF) (App. Fig. 12) test which again suggested $AR(0)$ and $MA(1)$ processes. Due to the results described above and the second-order differencing of the data, we can conclude that our data follows an **ARIMA(0,2,1)** model.

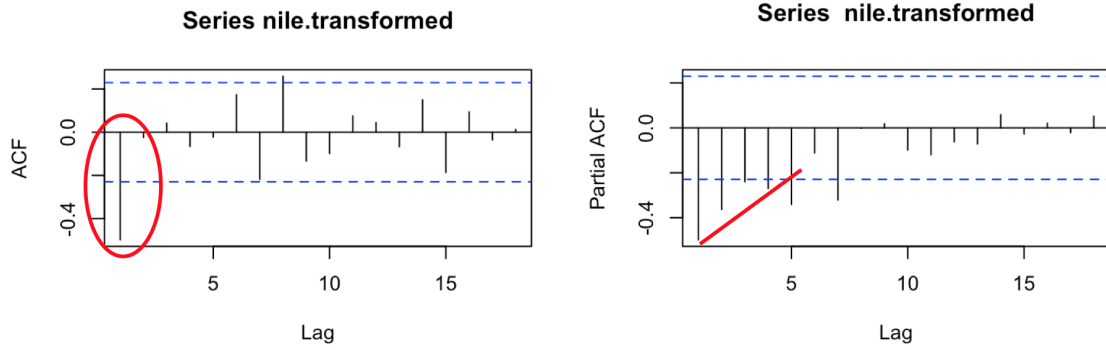


Figure 4: ACF and PACF plots of log transformed, second-order difference Nile River data. These plots show $MA(1)$ signatures: there is a single negative spike in the ACF plot and a decay pattern in the PACF plot.

3.2 Model

After our model selection, we ran an ARIMA function with the method of maximum likelihood estimation to fit our river flow data to the ARIMA(0,2,1) model and to determine the model’s parameters. The output showed $\theta_1 = 1.00$ and provided an Akaike Information Criterion (AIC) value of -53.96 . AIC values estimate the relative quality of a model for a given dataset with smaller values being more significant. This AIC score further supports our model selection. With θ_1 being the only parameter to determine, the model equation is as follows:

$$\nabla^2 Y_t = e_t - e_{t-1}$$

3.3 Forecasting

Due to the valuable ecological context, we decided to also run forecasting on the fitted data. Using a lead value of 60 years we predicted average stream flow from 1945 to 2004. Upper and lower bounds (in blue) were calculated using a 95% confidence interval (Fig. 5). These bounds are unusually wide with a “trumpet shape” which is indicative of second-order differencing[3]. The forecasted values appear in red on the figure. According to plot of the forecasted values, average annual river flow is predicted to continue to decrease over time.

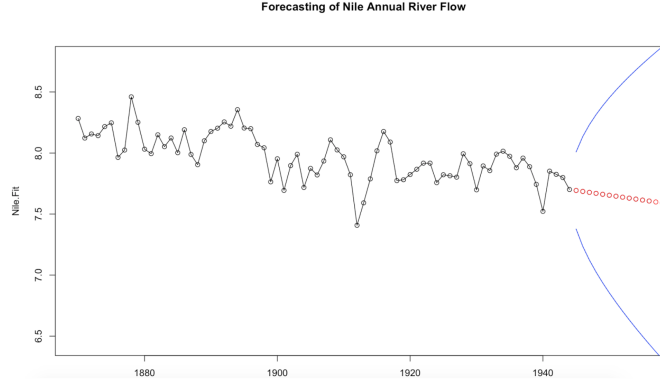


Figure 5: Forecast of Nile River data with lead value of 60. This shows “trumpet”-shaped confidence interval shape which is characteristic of ARIMA models with two orders of non-seasonal differencing.

4 Discussion & Conclusion

In this study we examined and analyzed average annual river flow data collected for the Nile River at Aswan, Egypt. We used various statistical methods to achieve stationarity for the data in order to select an appropriate model that could be used in forecasting. After multiple tests and validations we determined the fully transformed data followed an ARIMA(0,2,1) model with an equation of $\nabla^2 Y_t = e_t - e_{t-1}$. Our forecast for the data was in concordance with our prediction that river flow will be negatively impacted as human populations and landscape modifications continue to affect the ecology of the river. The Nile River at Aswan was once a vital and critical point for hydrological measurements of the river. The results of this study also has real-world applications. Not long after this original data was published[2], a dam was built completely altering the river’s structure and manipulating the river’s flow. As a result natural river flow data can no longer be collected and subsequently used to determine the ecological state of the river. So, trend and forecasting analysis of river flow data collected before the construction of this dam can be valuable tools in understanding the abiotic and biotic effects of the stream’s environment on its flow rate.

5 Appendix

5.1 Supplementary Figures

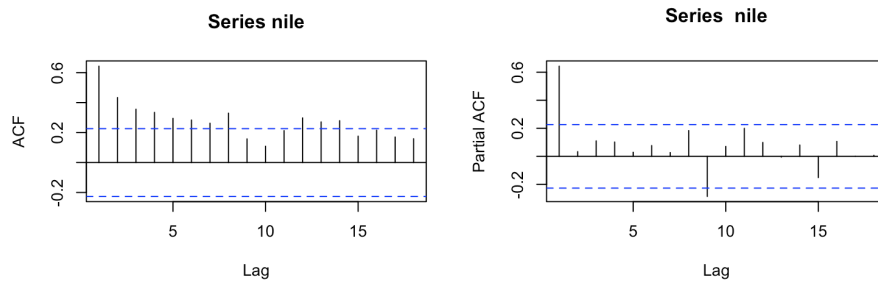


Figure 6: Autocorrelation (ACF) and Partial-autocorrelation (PACF) plots of Nile River data.

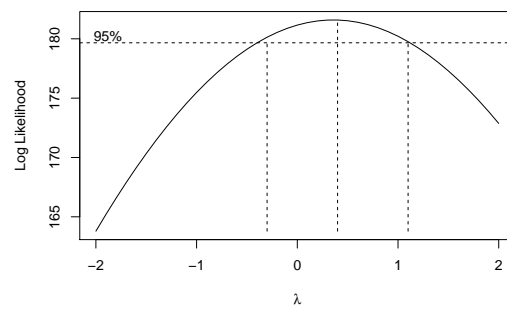


Figure 7: Box-Cox plot showing λ is close enough to 0 that we can use *log* to transform.

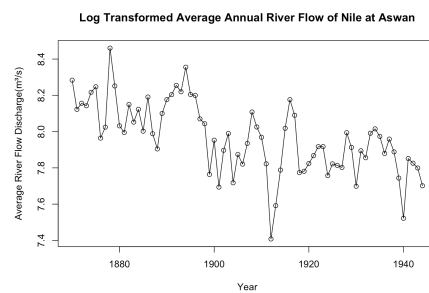


Figure 8: Plot of log transformed Nile River data.

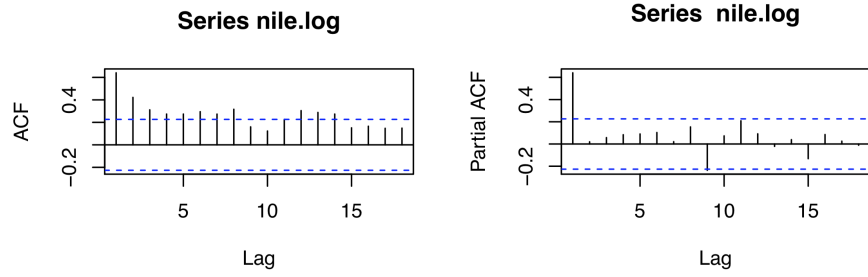


Figure 9: ACF and PACF plots of log transformed Nile River data.

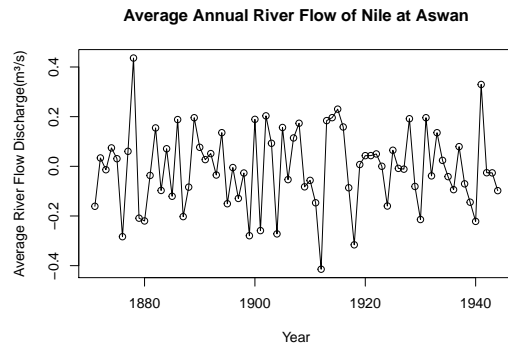


Figure 10: Plot of first-order differenced log transformed Nile River data.

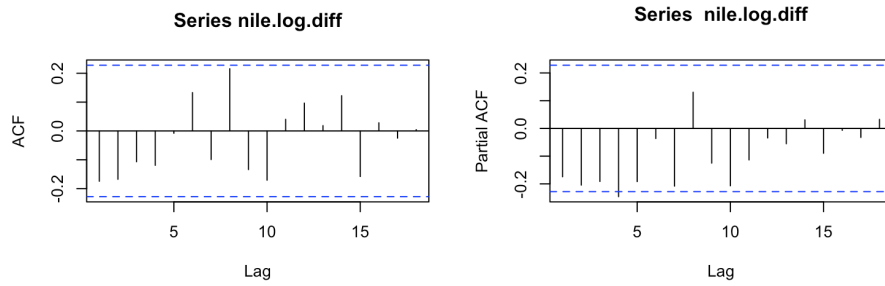


Figure 11: ACF and PACF plots of first-order difference, log transformed Nile River data.

AR/MA	0	1	2	3	4	5	6
0	x	o	o	o	o	o	o
1	x	x	o	o	o	o	o
2	x	x	o	o	o	o	o
3	x	o	x	o	o	o	o
4	x	o	x	o	o	o	o
5	x	x	o	o	o	o	o
6	x	x	o	o	o	o	o

Figure 12: Output of the sample extended artificial correlation which indicates an IMA(2,1)

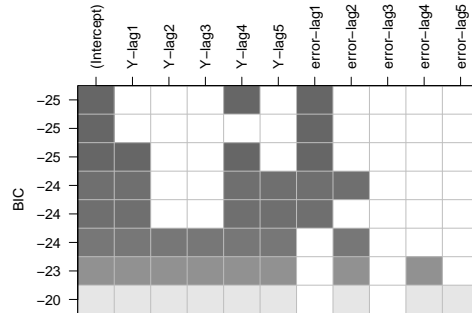


Figure 13: ARMA subset plot which indicates IMA(2,1)

5.2 R Code

```

1 # Time Series Final Project
2 # Fall 2017
3 # Due: 12/05/17
4 # Megan Pagliaro and Dylan Sosa
5 # Dr. Haijun Gong
6
7 # Dataset information:
8 # Average Annual River flow, Nile at Aswan, 1870-1945
9 # From DataMarket. Source: Time Series Data Library (citing: Hipel and McLeod (1994))
10 library("TSA")
11 setwd("/Users/Dylan/Documents/SLU/5.1/Time Series/Project/")
12 nile = ts(read.table(file = "nile.txt"), start = 1870, deltat = 1)
13 plot(nile, type='o', main='Average Annual River Flow of Nile at Aswan', ylab='Average River Flow
    Discharge(m^3/s)', xlab='Year')
14 par(mfrow=c(2,2))
15 acf(nile)
16 pacf(nile) #one peak, but will transform
17
18 # -----transformation-----
19 BoxCox.ar(nile) #close to 0, use log
20 nile.log = (log(nile))
21 plot(nile.log, type='o', main='Log Transformed Average Annual River Flow of Nile at Aswan', ylab='Average
    River Flow Discharge(m^3/s)', xlab='Year') #no significant change
22 par(mfrow=c(2,2))
23 acf(nile.log)
24 pacf(nile.log)
25
26 # first order difference transformation [I(1)]
27 nile.log.diff = (diff(log(nile)))
28 plot(nile.log.diff, type='o', main='Average Annual River Flow of Nile at Aswan', ylab='Average River Flow
    Discharge(m^3/s)', xlab='Year')
29 par(mfrow=c(2,2))
30 acf(nile.log.diff) #No significant peaks, weak MA and need to do second order difference
31 pacf(nile.log.diff) #No significant peaks, weak AR and need to do second order difference

```

```

32
33 # -----second order difference [I(2)]-----
34 Nile.log.diff.diff = (diff(diff(log(Nile))))
35 Nile.transformed <- Nile.log.diff.diff
36 abline(h=0, col = 'red')
37 plot(Nile.transformed, type='o', main='Transformed Average Annual River Flow of Nile at Aswan', xlab='
    Year', ylab='(diff(diff(log(Nile))))')
38 mean(Nile.transformed) # very close to zero
39 par(mfrow=c(2,2))
40
41 acf(Nile.transformed) # MA1 # Can we assume MA(1)
42 pacf(Nile.transformed) # AR(1) or AR(2)
43 adf.test(Nile.transformed)
44
45 # -----eacf use -----
46 eacf(Nile.transformed, ar.max = 5, ma.max = 5) #Suggesting an AR(0)/MA(1). BUT because MA showing weak
    movement at 7, row down may be more significant, meaning MA(1)/weak AR(1),
47 # eacf shows AR(0), MA(1)
48
49 # -----Finding Coefficients-----
50 Nile.subset = arimasubsets(y=Nile.transformed, nar=5, nma=5, ar.method='ols')
51 plot(Nile.subset)
52
53 # -----Fitting model-----
54 Nile.fit = arima(Nile.log, order=c(0,2,1), method='ML')
55 Nile.fit # These values for model.
56 # theta = 1
57 plot(Nile.fit)
58
59 # -----forecasting-----
60 Nile.predict = predict(Nile.fit, n.ahead = 60)
61 Nile.predict
62
63 upper = Nile.predict$pred + qnorm(0.975) * Nile.predict$se
64 lower = Nile.predict$pred - qnorm(0.975) * Nile.predict$se
65
66 points(Nile.predict$pred, col = 'red') #represent prediction mean value
67
68 lines(y=upper, x = 1945:2004, col = 'green')
69 lines(y=lower, x = 1945:2004, col = 'green')

```

References

- [1] Average annual river flow, Nile at Aswan, 1870 – 1945. <https://datamarket.com/data/set/22p1/average-annual-river-flow-nile-at-aswan-1870-1945#!ds=22p1&display=line>.
- [2] K.W. Hipel and A.I. McLeod. *Time Series Modeling of Water Resources and Environmental Systems*. Elsevier, 1994.
- [3] Robert Nau. "notes on nonseasonal arima models". https://people.duke.edu/~rnau/Notes_on_nonseasonal_ARIMA_models--Robert_Nau.pdf, 2014.