

# USING AUTOREGRESSIVE MODELING FOR FLOW FORECASTING IN THE APALACHICOLA-CHATTAHOOCHEE-FLINT RIVER BASIN (ACF)

Jeffrey Regan

AUTHOR: Georgia Environmental Protection Division - Watershed Protection Branch, 2 Martin Luther King Jr. Dr Atlanta, GA 30334

REFERENCE: *Proceedings of the 2013 Georgia Water Resources Conference*, held April 10–11, 2013 at the University of Georgia.

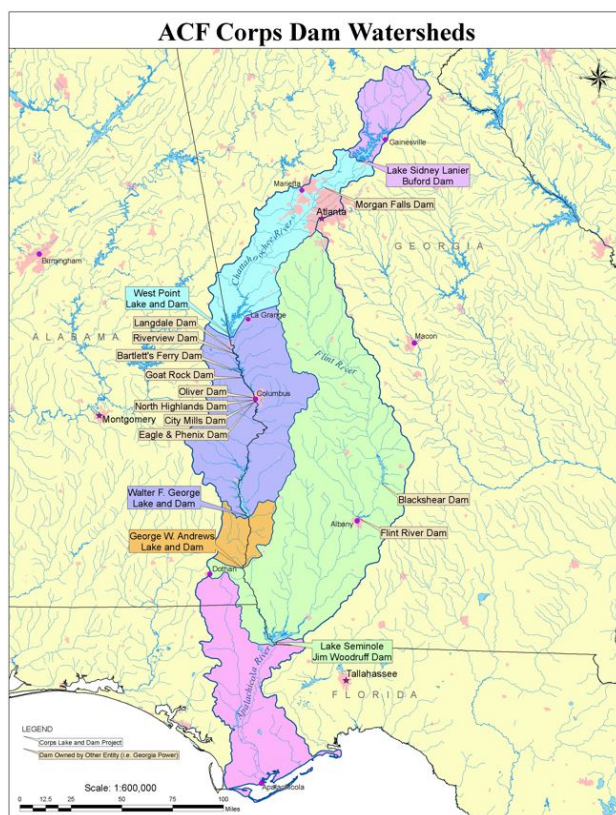
**Abstract** The Apalachicola-Chattahoochee-Flint (ACF) River Basin supports a multipurpose river system, which straddles Georgia, Alabama, and Florida. The Army Corps of Engineer's management of the federal reservoir system has been under litigation between the three states and various stakeholders for more than 20 years. Among the difficulties in properly managing the river system is the uncertainty of future river inflows and the inability to properly forecast and plan for likely drought periods. In addition to the Corps' need for reliable long-term flow forecasts for reservoir operation, the state of Georgia could use such forecast for properly managing off-stream water use when faced with probable droughts.

This paper will develop an autoregressive inflow forecast model and evaluate the benefits of a multiple variable autoregressive model, which includes the ENSO index. The data that will be used is the unimpaired inflow dataset developed by the Corps and the historical ENSO index. The analysis will also measure how each forecasts compares monthly, and determine months in which forecasts are more useful.

## INTRODUCTION

The operations of the Apalachicola-Chattahoochee-Flint (ACF) River Basin are dependent upon the inflows to the basin. This provides much uncertainty regarding expected operations, power generation, lake-levels, and contested state-line flows at the Georgia-Florida border. The Army Corps of Engineers operates the reservoir system and has developed a set of daily-unimpaired inflows from 1939–2008 for each river reach in the basin. These unimpaired flows are labeled unimpaired as they are considered the best estimates of the natural flow that comes into the basin before being altered by reservoir operation, water use, and lake evaporation. They are calculated by taking measured flow gages and adding back the measured changes in storage in the major reservoirs, adding back the estimated evaporation from lake surfaces and adding back the estimated consumptive water use. This unimpaired data can then be placed back into system reservoir models to see the effects of different operation alternatives and water demands. In this sense, forecasted unim-

paired inflows can be input into the river basin models to properly forecast expected operations, power generation, lake-levels, and contested state-line flows. The models could also help formulate alternative operations based on forecasted conditions. Currently, no defined method is in use for a long-term flow forecast. Rather, the states and the Corps merely look at current conditions, the ENSO index and the precipitation forecasts of the Climate Prediction Center (CPC) of the National Weather Service to get the best idea of likely inflows. This idea of likely inflows is not quantified and thus cannot be used in reservoir models for forecasting the river system conditions. The below map shows the reaches of the river system as well as the major reservoirs in the ACF river basin.



**Figure 1. Layout of ACF River Basin**

This paper seeks to evaluate the performance of an autoregressive model for the whole ACF system, and the sum inflows to the basin will be used in autoregressive model

development. This paper will determine the performance of an autoregressive model for 1-month, 3-month, and 6-month forecasting as well as test the benefit of adding the ENSO index as an additional variable in the forecasting scheme.

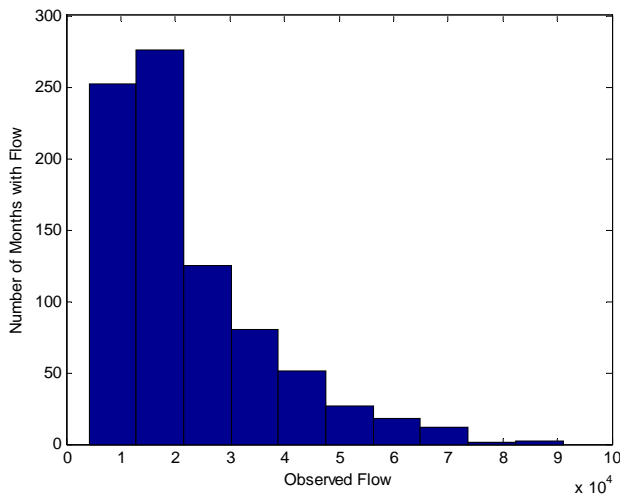
## METHODOLOGY

Using the Corps Unimpaired daily flow data each river reach is summed up and converted to a monthly time-series in cubic feet per second.

In developing an Autoregressive model, the first step is to evaluate the time-series data and determine how it needs to be transformed into a standard normal variable. The below plots show the time-series of the observed data, the mean and standard deviation by month, and the histogram.

**Error! Not a valid link. Figure 2. ACF River Historical monthly unimpaired inflows**

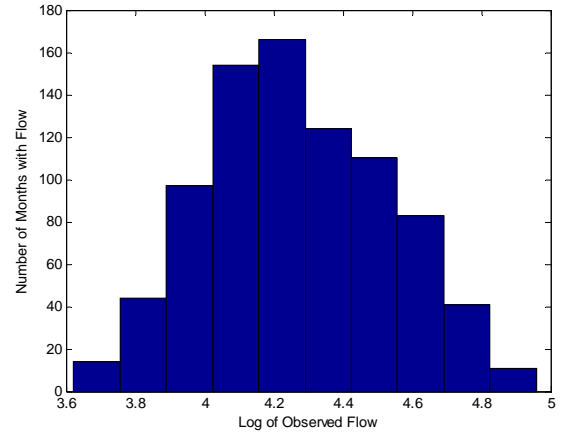
**Error! Not a valid link. Figure 3. ACF River Historical unimpaired inflows monthly average and standard deviation**



**Figure 4. Histogram of ACF River Historical unimpaired inflows**

The observed flow appears to be log distributed, therefore the first step will be to take the log of each observation. This produces the following time series as well as a histogram that is closer to a normal distribution.

**Error! Not a valid link. Figure 5. Log transformed unimpaired inflows**



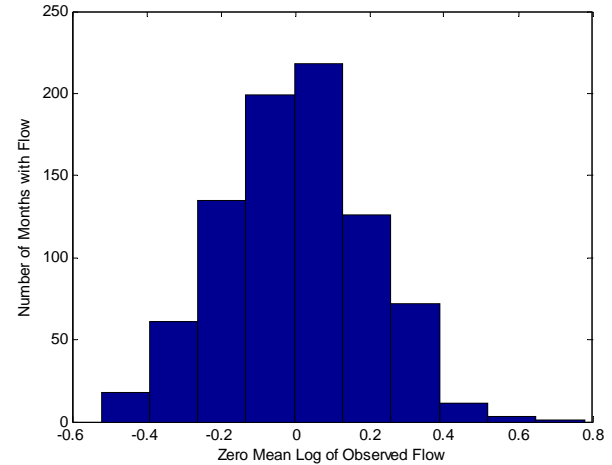
**Figure 6. Histogram of Log transformed unimpaired inflows**

Since the observed data shows a seasonal pattern the monthly mean must be removed. The monthly mean of the log time series is plotted below.

**Error! Not a valid link. Figure 7. Monthly mean of Log transformed data**

Subtracting the monthly mean yields the following time-series and histogram.

**Error! Not a valid link. Figure 8. Log transformed data with mean removed**



**Figure 9. Histogram of Log transformed data with mean removed**

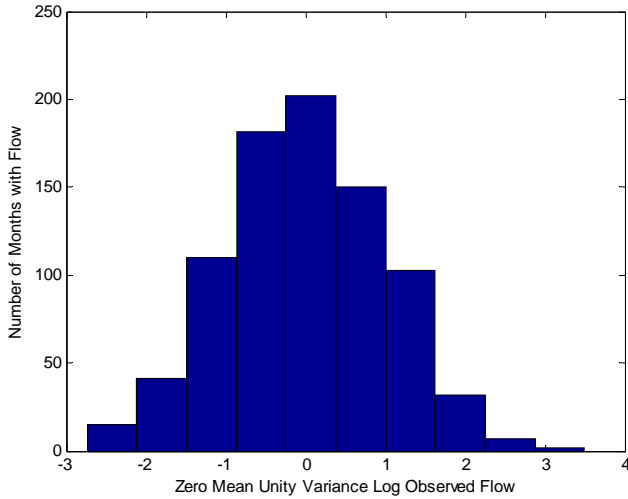
The final step is to normalize the distribution so that it has a variance and standard deviation equal to one. This is done by dividing by the standard deviation of the zero mean Log flow which is plotted below.

**Error! Not a valid link. Figure 10. Standard deviation of Log transformed data with mean removed**

Dividing by the Standard deviation finally results in the following time-series, which has a zero mean and a

standard deviation equal to one. The final timeseries to be used in the forecasting process is plotted below along with a histogram that shows the timeseries to have a standard normal distribution.

**Error! Not a valid link. Figure 11. ACF River Historical monthly flows as a standard normal variable with mean removed and unity variance**



**Figure 12. Histogram of ACF River Historical monthly flow as a standard normal variable with mean removed and unity variance**

The next step is to pick the proper type of autoregressive model based on the autocorrelation of the observed standard normalized data. The below plot shows how the lag correlation compares with an AR(1) model and an ARMA(1,1) model. The lag1 correlation was shown to equal 0.68 while the lag2 correlation equals 0.54.

**Error! Not a valid link.**

**Figure 13. Autocorrelation of standard normal monthly unimpaired flows compared with 2 types of autoregressive models**

The ARMA(1,1) model shows the best approximation of the observed autocorrelation. The observed data also indicates that ARMA(1,1) is a possibility as:

$$\rho_2 \geq \rho_1 (2\rho_1 - 1)$$

Therefore an ARMA(1,1) model will be chosen to complete the autoregressive modeling process. The ARMA(1,1) model takes the form:

$$Z_t = \phi_1 Z_{t-1} + a_t - \theta_1 a_{t-1}$$

$$m_z = 0, \quad \sigma_z^2 = 1.$$

The parameters  $\theta$  and  $\phi$  were found by using the following formulas and solving iteratively.  $\alpha_t$  and  $\alpha_{t-1}$  are standard normal Gaussian white noise.

$$\hat{\rho}_1 = \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 - 2\phi_1 \theta_1 + \theta_1^2},$$

$$\hat{\rho}_2 = \phi_1 \hat{\rho}_1,$$

$$\left[ \frac{(\theta_1 - \phi_1)^2}{1 - \phi_1^2} + 1 \right] \sigma_a^2 = \sigma_z^2 = 1,$$

The resulting parameters  $\theta_1 = 0.0230$  and  $\phi_1 = 0.797$

The best possible forecast will be equivalent to the expected flow and simplifies to:

$$\hat{Z}_{t+k|t} = \rho_k z_t, \quad k = 1, 2, 3,$$

where  $\rho_k = \rho_1 \phi_{k-1}$

This essentially means a 1-month forecast is the previous months flow times the lag1 correlation. In the event of zero correlation the flow forecast will be zero which means average for the zero mean time-series. A n-month forecast is the same but uses the lag-n correlation of the ARMA model.

Since our observed data has been transformed into a standard normal variable, our model output needs to be transformed back into regular form. The steps laid out in the transformation process will be reversed such that first the results must be multiplied by the standard deviation, then the means added back, and then finally it must be reverse log transformed for the entire period 1939-2008. The forecasting process was repeated for a 3-month and 6-month forecast and the results of each model will be shown in the Results Section of this paper.

A multiple autoregressive model was considered using the ENSO index that is often deemed a good indicator of weather patterns in the Southeast US. A La Nina is often associated with dry and warm weather in the Southeast and used as predictor for possible drought. The ENSO

time-series was transformed into a standard normal variable was done with the unimpaired inflows. The lag correlation of the ENSO and the unimpaired inflow data is plotted in Figure 14. An ARMA(1,1) could not be fitted since the lag two correlation is greater than the lag one. The AR(1) model also does not properly model the long lag correlation. A more complicated AR(13) was developed to better fit the observed correlation better. According to Bras and Rodriguez-Iturbe care must be used when working with sample correlations for high lags. This model shows matches the observed correlation until lag 13 and may be considered a better estimate than the AR(1) model.

The correlation does show as expected a certain dryness associated with la Nina conditions but the correlation is very small. The ENSO may have a greater correlation with precipitation but this does not translate to correlation with stream flow due to the dynamics of the stream flow system.

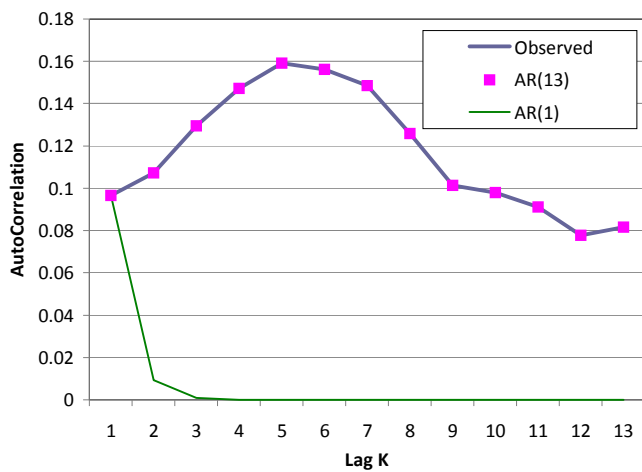


Figure 14. Lag correlation between inflows and ENSO index

## RESULTS

After completing the ARMA(1,1) model for the entire data period of 1939-2008 a forecast for each month is compared with the observed data in Figures 15 and 16. These results reflect 1-month forecasts without the addition of the ENSO index data.

**Error! Not a valid link. Figure 15. 1-Month ARMA(1,1) and observed unimpaired flows**

**Error! Not a valid link. Figure 16. 1-Month ARMA(1,1) and observed unimpaired flows**

One way to measure the goodness of the forecast is by determining the correlation between observed and simulated. This is likely to be similar to the lag-1 correlation. The below plot in Figure 17 shows the monthly correlation of the observed and simulated time-series.

**Error! Not a valid link. Figure 17. 1-Month AR-MA(1,1) and observed correlation**

The standard error of the observed and the forecasted (Figure 18) can be compared to determine how the forecasted data has become closer to the mean than the observed and therefore has a lower standard deviation.

**Error! Not a valid link. Figure 18. 1-Month ARMA(1,1) and observed standard deviation**

Next, the flow forecasting results using just the AR(13) ENSO index model were analyzed. The plots for these model runs are not shown; however, the results show that certain times of the year have a greater ENSO signal. The winter seems to have the greatest correlation while the spring appears to have zero correlation. Also due to the long lag in the AR(13) model, the correlation does not greatly decrease as the forecast length increases.

Finally the two models were combined to see the benefits of combining the two models. All results can perhaps be best summarized in the following three graphs that compare the correlation performance for the three different models for the 1, 3 and 6-month forecast (Figures 19-21). These plots show that the ARMA(1,1) model consistently out performs the AR(13) and multiple variable model. Though the AR(13) model shows statistically significant correlation, it only takes away from the correlation when included. One can also see that as the forecast length increases the ability of ARMA(1,1) to out perform the other models diminishes.

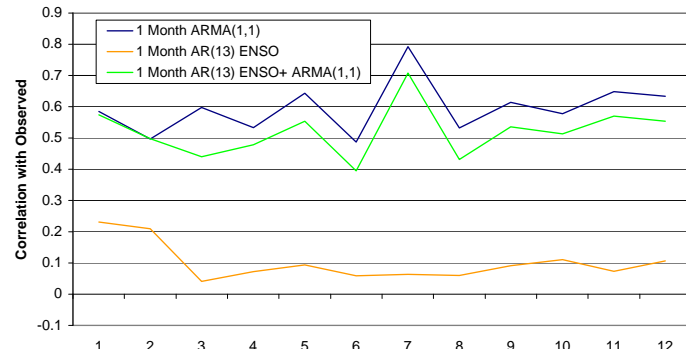
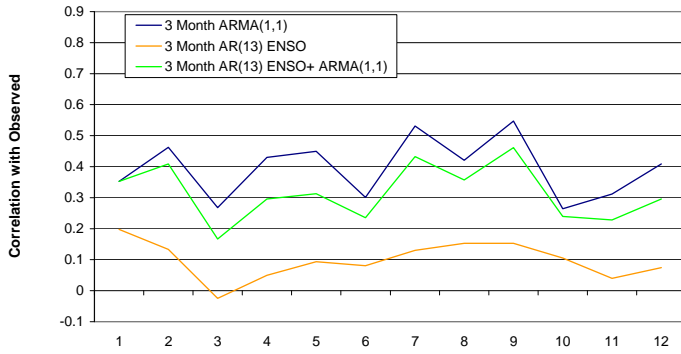
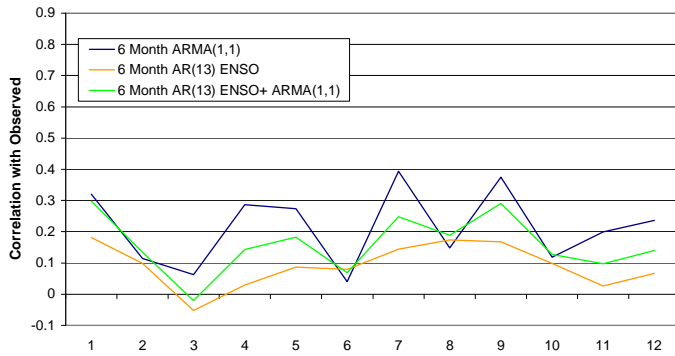


Figure 19. 1-Month Forecast correlation of three developed models.



**Figure 20. 3-Month Forecast correlation of three developed models.**



**Figure 21. 6-Month Forecast correlation of three developed models.**

## CONCLUSION

The ACF River system shows to have an autocorrelation lasting almost up to 6 months lag throughout all periods of the year. This correlation can be used for forecasting with acceptable degree of uncertainty. While forecasts beyond 1 month carry a higher uncertainty, they still are the best indicators for future conditions. The ENSO index shows slight correlation for a long lag period with ACF flows. However, even after possibly over fitting the model to match this lag correlation, the ENSO index inclusion only hurt performance of the ARMA model. Future studies could include testing whether the ENSO AR(13) model has been over fit as well as determining a proper way to include the ENSO correlation to improve the performance of the ARMA(1,1) model.

## REFERENCES

Bras, R. L., and I. Rodriguez-Iturbe, 1985: Random Function and Hydrology