UNDERSTANDING HYDROLOGIC VARIATION THROUGH TIME-SERIES ANALYSIS

Shawn E. Rosenquist, Jason W. Moak, Aaliyah D. Green, and Oscar P. Flite III

AUTHORS:

Shawn E. Rosenquist, Southeastern Natural Sciences Academy, 1858 Lock and Dam Road, Augusta, GA, 30906
Jason W. Moak, Southeastern Natural Sciences Academy, 1858 Lock and Dam Road, Augusta, GA, 30906
Aaliyah D. Green, Southeastern Natural Sciences Academy, 1858 Lock and Dam Road, Augusta, GA, 30906
Oscar P. Flite, III, Southeastern Natural Sciences Academy, 1858 Lock and Dam Road, Augusta, GA, 30906

REFERENCE: *Proceedings of the 2010 South Carolina Water Resources Conference*, held October 13-14, 2010, at the Columbia Metropolitan Convention Center.

Abstract. Building a solid understanding of hydrology is critical to many water resource management endeavors. A broad range of stakeholders draw upon a baseline understanding of hydrology in order to contextualize their own results or to make management decisions that affect the water resource management and the environment. In this study, we present a method that can enhance our understanding of hydrology by utilizing time-series analysis methods to illuminate important aspects of hydrology involving temporal variation. Methods included smoothed first derivative peak finding, spectral density analysis, and wavelet analysis. The case study utilized a combination of these methods to compare specified reaches of two piedmont-to-coastal plain river systems in the southeast US, the Savannah and the Altamaha. These two systems respectively represent a highly flow-regulated river system with many flow-controlled impoundments on the main channel, and a mostly unregulated river system. Results of time-series analysis revealed that the flow-regulated Savannah River can be characterized by a larger number of flow-varying events, most of which are of relatively short-duration compared to flow-varying events in the unregulated Altamaha River. The time-series analysis provided useful additional information on flow variation and differences between that could be obscured or missed by relying solely on more generalized analysis methods.

INTRODUCTION

It is of no surprise that understanding the amount of water moving through a system is of great importance to the wide range of topics involving water. Soil and water chemistry, aquatic biology, geomorphology, and water resource development such as hydropower facilities and impoundments all build on hydrology. As one example, current losses in natural wetlands have spurred the development of mitigation wetlands. Experience in this field has shown that without a solid understanding of the hydrology of the system, the mitigation site is less likely to develop into a functioning wetland (Atkinson et al., 1997).

In lotic systems, temporal variations at various scales are suspected to impact biota and ecological structure (Biggs et al., 2005). Various changes in land and water use can affect hydrology in terms of both average values and temporal variation. Specifically, impoundments can cause either increases or decreases in flow variation depending on the timing of water release from the impoundment. In the case of the Savannah River, high volume releases from J. Strom Thurmond Dam used to generate peak power occur on a daily time-scale and tend to increase variability in flow, while Stevens Creek Dam, 13 miles downstream, is used to generate base power and tends to decrease variability in flow (SCE&G, 2010). Several metrics can be used to describe and compare hydrology including flow duration curves or the application of a mean, standard deviation, or coefficient of variation (CV) to flow data. While these metrics provide useful information about the systems, they provide no information about the timing of the hydrologic variations involved.

Time-series analysis or signal processing in more generic terms has a long history including extensive use in the field of electrical engineering. More recently, time-series analysis methods such as Mann Kendal testing and wavelet analysis have been applied to hydrologic analysis (Dorval et al., 2010; Delgado et al., 2010). The objective of this study is to develop a method utilizing elements of time-series analysis that can provide: 1) parametric characterization of specific flow-varying events including both amplitude and wavelength characteristics; 2) characterization of a particular hydrologic system through distributions in wavelength and amplitude; and 3) illumination of both the scale of variation and the timing of that variation.

METHODS

Time-series methods used in this study include signal smoothing, spectral density analysis, and wavelet analysis.

Each of the three different approaches in the study incorporated some of the above methods while the most successful approach was also used to complete a comparative case study between the two river reaches.

Spectral Density Analysis

Spectral density analysis provided information that fed in to other analysis methods. Since spectral density analysis could not provide information about specific events, it was of limited utility to the objectives of this study. However, spectral density analysis was able to provide a snapshot of what periodicities were common in the data set and guided the range of wavelengths selected in the three approaches to the range of 1 to 28 days.

Smoothed First Derivative Approach

The first approach for characterizing the data used a smoothed first derivative to localized peaks and valleys in the data. The 15-min data used in the case study were smoothed over a 12- to 72-hour range during different attempts. This smoothing was an attempt to eliminate noise at sub-event time scales that would interfere with the algorithm. In theory, these peaks and valleys could then be matched by position in order to characterize flow-varying events in wavelength and amplitude.

Wavelet Analysis Approach

The second approach for characterizing the data was through direct wavelet analysis. In this method various waveforms (triangular waveform, 0-π sine wave, -π/2-3π/2 sine wave) representing a range in both amplitude (1000cfs increments corresponding to the global range of the data) and wavelength (1 to 7 days by 1 day increments) were convolved through the data set while the algorithm compiled a goodness-of-fit parameter for each waveform. Goodness-of-fit was defined by Equation 1:

$$PercentMatch = \frac{Fit}{(Fit + Misfit)} \qquad \text{Eqn. 1}$$

where Fit was the area under both the waveform and the data set, Misfit was the area between the waveform and the data set, and PercentMatch was the goodness-of-fit parameter that corresponds to the percent of the total area that matches correctly.

This goodness-of-fit data was then mined for the best matches at each location over the data range. After one match was selected, all other possible matches over that interval were eliminated.

Combined Approach

The third approach involved a combination of the above methods. First, data were smoothed over a 2- to 4-hour range to eliminate sub-event-scale noise. Then, since the waveforms observed in the smoothed data were approximately symmetrical, maxima in the data set were chosen as waveform centroids, eliminating the need to convolve each waveform through the dataset. For each local maximum 0-π sine waveforms with set amplitudes corresponding to the particular maximum and covering a range in wavelengths (6 to 672 hours by 6-hour increments) were applied at that temporal location. The algorithm selected the best-fit wavelength for each local maximum based again on Equation 1. In addition to the time of the event, the peak discharge, and the best-fit wavelength, the algorithm also captured the corresponding local minima on both sides of each maximum and the goodness-of-fit for each waveform.

Case Study Reaches

The case study for the above methods included specified reaches of two river systems in the southeast US flowing from the piedmont to the coastal plain, the Savannah and the Altamaha. These two systems respectively represent a highly flow-regulated river system with several flow-controlled impoundments on the main channel, and a mostly unregulated river system. At the same time, their similar physiographic location, watershed size, and geomorphologic nature indicate that the unaffected flow conditions would likely be quite similar. Time and discharge data for January 1, 2006 to December 31, 2006 at 15-minute intervals were sourced from the USGS database for six locations, three on the Altamaha and three on the Savannah. The most upstream "Altamaha" location is actually about 12 miles up the Ocmulgee, one of the two main tributaries that converge to form the Altamaha. A general description of the sites and the flow for 2006 at each site is given below through various parameters (Table 1) and with flow-duration curves (Figure 1).

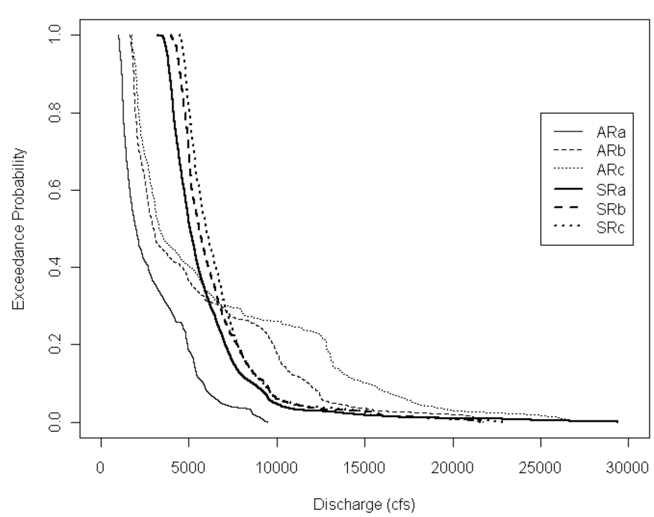

Figure 1. Flow-duration curves for three locations on the Savannah and Altamaha Rivers.

Table 1. General Description of the Study Reaches

	Savannah River Locations				
	SRa		SRb		SRc
Latitude	33°22'25"		33°08'59"		32°56'20"
Longitude	81°56'35"		81°45'18"		81°30'10"
Elevation (ft)	99		78		54
Reach Length (miles)		37		29	
Mean Slope		0.011%		0.016%	
2006 Total Discharge (ft^3)	1.85E+11		1.90E+11		2.11E+11
2006 Incremental Flow (ft^3)		5.00E+09		2.11E+11	
Mean Flow (cfs)	5904		6610		6758
Standard Deviation of Flow	2824		2714		2528
Coefficient of Variation of Flow	0.48		0.41		0.37
Minimum Flow(cfs)	3153		3952		4443
Maximum Flow (cfs)	29405		21547		22840
Contributing Watershed ($miles^2$)	7510		8300		8650

	Altamaha River Locations				
	ARa		ARb		ARc
Latitidude	31°55'12"		31°56'20"		31°39'16"
Longitude	82°40'27"		82°21'13"		81°49'41"
Elevation (ft)	98		64		29
Reach Length (miles)		29		53.5	
Mean Slope		0.022%		0.012%	
2006 Total Discharge (ft^3)	8.50E+10		1.64E+11		2.04E+11
2006 Incremental Flow (ft^3)		7.90E+10		2.04E+11	
Mean Flow (cfs)	2851		5668		6515
Standard Deviation of Flow	2033		4538		5833
Coefficient of Variation of Flow	0.71		0.80		0.90
Minimum Flow(cfs)	985		1650		1730
Maximum Flow (cfs)	9460		21700		26800
Contributing Watershed ($miles^2$)	5180		11600		13600

RESULTS

Smoothed First Derivative

This method correctly selected both peaks and valleys from the dataset. However, at least three problems prevented its effective use to accomplish study objectives. First, it was common for the algorithm to select several errant peaks with only minimal smoothing or select two or three points for the same peak. While higher levels of smoothing (60-72 hour averages) could successfully reduce this problem, it created a second problem. With higher levels of smoothing, valid peaks of shorter wavelengths would be ignored, biasing the dataset toward larger waveforms. A third problem arose in matching peaks with the appropriate valleys. Often, shoulder peaks were selected in addition to the main peak, but the intermediate sub-valley would not be selected. This occurred on an unpredictable basis throughout the dataset and created significant difficulty in correctly matching peaks and valleys.

Wavelet Analysis

This method was more successful in describing the data set. For a particular data set, runs were completed where PercentMatch ranged from worst matches near zero to best matches around 0.9. The results were generally useful in providing perspective of the distribution of variation over time in terms of typical wavelengths and their associated peak discharges. However, the cumbersome nature of the algorithm created drawbacks. First, due to the number of iterations required to combine a large number of possible waveforms convolved over the entire data set, the process was computationally demanding. The goodness-of-fit data set for only one year of data created just less than seven million data points, and much of that computation was enumerating very poor matches. The associated time and data storage requirements for computation also severely limited the ability to include either a desirably large range of waveforms or a desirably fine granularity. In addition, the complex algorithm was somewhat prone to problems with different data sets and required a lot of testing to build confidence in the result for a new data set.

Combination Method

This approach reduced computational time from the previous approach by over 95% while the assumption of local maxima as centroids for the waveform matches reduced goodness-of-fit by approximately 20%. The resulting array of waveform matches was useful in characterizing the distribution of temporal variability in the data as demonstrated in the case study. Wavelengths selected had higher granularity than was practical with the previous approach while the level of smoothing required was not large enough to wash out smaller wavelengths events. Selection of local minima coordinating with each event was mostly successful, with one drawback. As the algorithm will eliminate a given area of the data as a best-fit waveform is found, the true minimum for an adjacent waveform can sometimes be removed with the first waveform. The result is that the estimation of the amplitude of that waveform must rely on a single minimum, instead of an average of the two. There were a handful of circumstances where both minima were removed from a waveform that otherwise had a good fit, most within SRa data set where the high level of variability resulted in the highest number of total matches while some data sets had no such occurrences.

Case Study

While examination of the characteristics of the two rivers from Table 1 and Figure 1 is helpful for understanding differences, results of the combined time-series approach for this data provides useful additional insights. Distributions were constructed for each location of the wavelength of a typical flow-varying event (Figure 2), the amplitude over the temporally localized baseflow (Figure 3), and the proportional increase in flow over the temporally localized baseflow (Figure 4).

In addition to providing distributions for individual waveform parameters, data gathered through the combined approach enabled matching wavelengths and amplitudes for particular flow-varying events. Then, the relationship between these two parameters could be compared, ultimately providing insights into how that relationship may change between locations (Figure 5). This analysis could be used either to investigate predictive trends between wavelength and amplitude in particular systems or to investigate what amplitudes are typical for a particular wavelength in a particular location and visa versa.

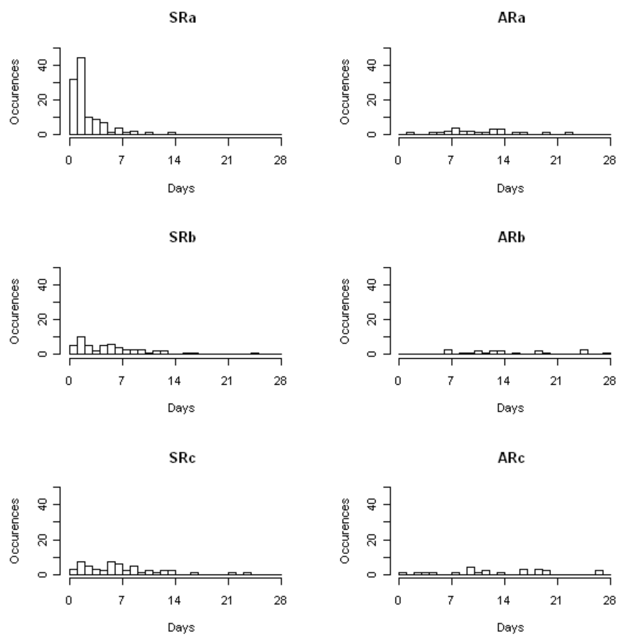

Figure 2. Wavelength distribution for flow-varying events during 2006 in the case study locations.

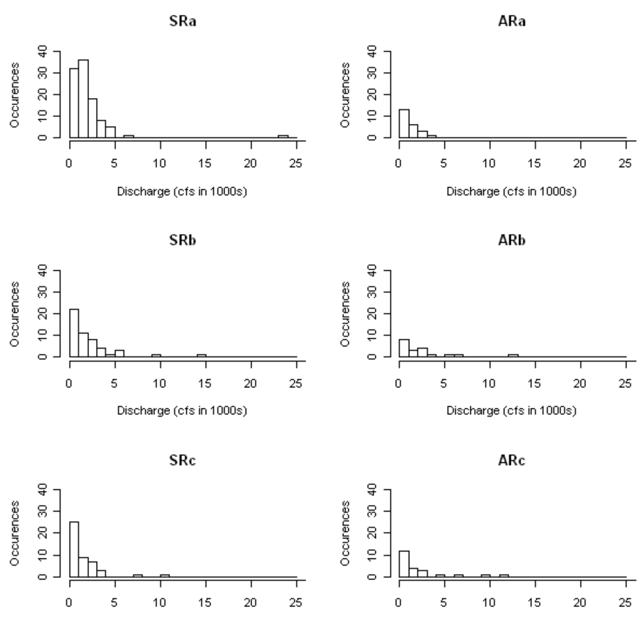

Figure 3. Amplitude distribution for flow-varying events during 2006 in the case study locations.

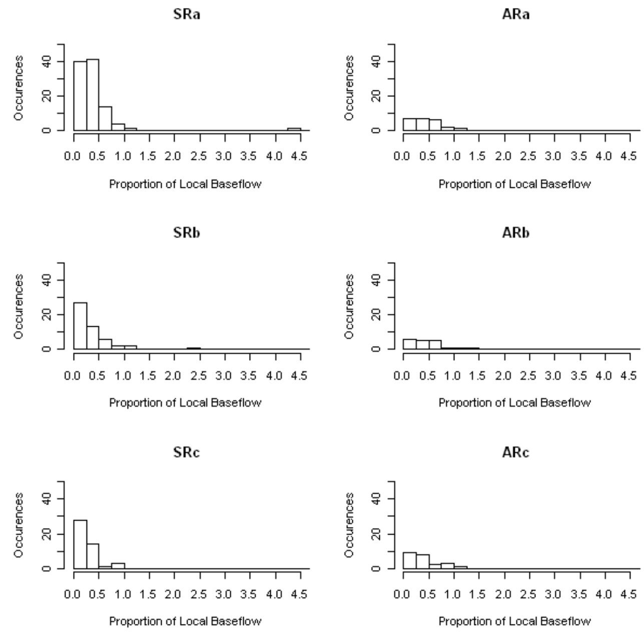

Figure 4. Distribution of proportion of flow-varying event amplitude to the temporally localized baseflow during 2006 in the case study locations.

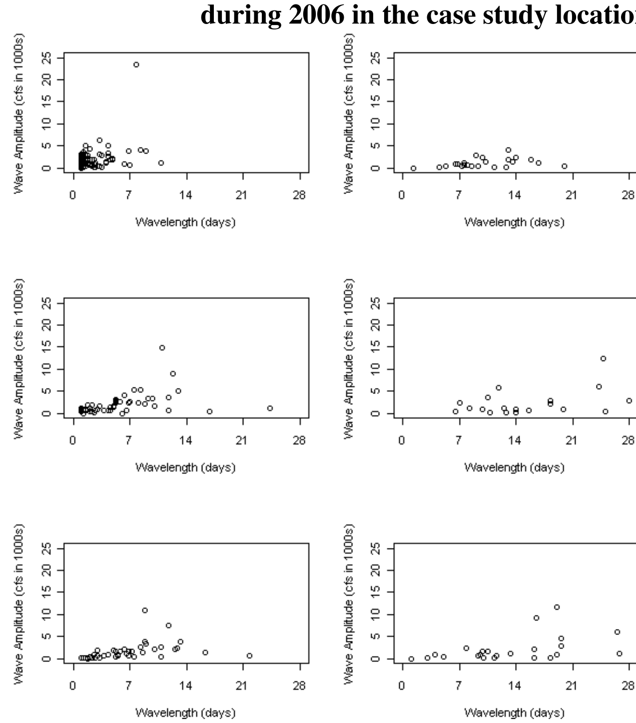

Figure 5. Relationships between waveform amplitude and wavelength for the various case study locations.

DISCUSSION

Some insights about case-study reaches can be drawn from information in Table 1 and Figure 1 without the time-series analysis. For instance, consideration of Figure 1 reveals that the baseflow conditions of the Savannah are likely increased from flow regulation. In addition, Figure 1 and the standard deviations from Table 1 seem to indicate that there is as much or more "flow variation" in the Altamaha than in the Savannah. Furthermore, the coefficient of variation in Table 1 may seem to indicate that flow-varying events on the Altamaha have a larger effect on baseflow than those on the Savannah. While these last two conclusions are likely true in part, the generalized statistics alone can leave an incomplete picture while time-series analysis reveals several additional insights that enhance our understanding.

While standard deviation from the mean in both systems may be similar or even higher in the Altamaha, **Figure 2 reveals that flow-varying events on the upstream portion of the Savannah, closest to the impoundments, are much more frequent and of shorter duration than those seen on the Altamaha.** These events fall in the 24- to 48-hour range. Interestingly, variations at this particular time-scale were indicated as most important to biological processes such as colonization, biotic interactions, and reproduction (Biggs et al., 2005). This effect seems to disappear as the Savannah flows downstream and begins to approach the more evenly distributed flows observed in all locations on the Altamaha.

Regarding the amplitude of flow variations, **Figure 2 and Figure 3 reveal that the Savannah has an increased number of small amplitude (<3000 cfs) waveforms that increase flow less than 75% above the temporally localized baseflow.** This finding is not visible in the standard deviation, which provides no information of when the variation occurs. The prevalence of smaller events may further explain some of the difference in CV, but much of the CV difference may also be explained by the increase in baseflow mentioned earlier.

Lastly, **Figure 5 reveals a clustering of flow-varying events upstream in the Savannah at small wavelengths but over a wide range of amplitudes (0-5000 cfs).** This phenomenon then dissipates moving downstream into a relationship that looks more similar to the Altamaha where amplitude seems to vary more at higher wavelengths than at lower wavelengths. However, the highest amplitudes in the Savannah, even downstream, seem to occur with lower wavelengths than in the Altamaha.

Future work with this method should include incremental flow analysis, as the wave characterization demonstrated should allow for wave matching between locations. This may help determine wave travel time and changes in wave structure between locations, which could be used to investigate differences in geomorphology or land use between locations or to further compare the affects of flow regulation. In development of the method, future work should include varying the centroid of the waveform within the data set by a small percentage of the wavelength during best-fit selection in order to improve match accuracy.

CONCLUSIONS

The application of time-series analysis methods to river hydrology yielded valuable insights in addition to more traditional metrics for flow characterization. In the case study presented, a large number of shorter-duration but similar amplitude flow-varying events originating in the upstream region of the Savannah closest to the flow-controlled impoundments differentiated its flow variation from the Altamaha in ways that were masked in the generalized statistics. While the global range of variation in the two systems may be similar or slightly higher for the Altamaha, the frequency of variation in the unregulated river was much lower with less total flow-varying events. This temporal masking in the generalized statistics could lead to incomplete or incorrect conclusions if generalized statistics are relied upon exclusively for decision-making or as a baseline for studies in water chemistry or biology.

ACKNOWLEDGMENTS

This research was funded through several foundations and many individual contributors to Southeastern Natural Sciences Academy. Data was provided through the USGS river-monitoring program.

LITERATURE CITED

Atkinson, R.B., W.L. Daniels and J. Cairns, Jr., 1997. Hydric soil development in depressional wetlands: A case study from surface mined landscapes. In*:* S.K. Majumdar, W.W. Miller and F.J. Brenner, Ecology of Wetlands and Associated Habitats. Pennsylvania Academy of Science, Philadelphia, pp. 170-182.

Biggs, B. J. F., V.I. Nikora and T. H. Snelder, 2005. Linking scales of flow variability to lotic ecosystem structure and function. *River Research and Applications* 21:283-298.

Delegado, J.M., H.Apel and B. Merz, 2010. Flood trends and variability in the Mekong River. *Hydrology and Earth System Sciences* 14:407-418.

Dorval F.A., B. Chocat, E.Emmanuel and G. Lipeme Kouyi, 2010. Sewer system flow components signal processing. *Water Science and Technology* 62.1:106-114.

SCE&G. Stevens Creek Hydroelectric Project. Downloaded: September 9, 2010. http://www.sceg.com/NR/rdonlyres/25DD5351-2826-478B-9691-BE26A3F1CEB3/0/StevensCreekReport.pdf

United States Geological Survey (USGS). 2010. USGS Real-Time Water Data for Georgia. http://waterdata.usgs.gov/ga/nwis/rt