ECWS

MDPI

An Examination of Extreme Flood Events and Resiliency of the Lower St. Johns River, Florida, USA Using Multiple Methods

Samantha Kovalenko [1], Christopher Brown [1,*], Cigdem Akan [2], Alexandra Schonning [2]

[1] School of Engineering, University of North Florida, USA; graduate student; n00945365@unf.edu (S.K.);
[2] Taylor Engineering Research Institute, University of North Florida, USA; cigdem.akan@unf.edu (C.A.); aschonni@unf.edu (A.S.)
* Correspondence: christopher.j.brown@unf.edu; Tel.: +01-904-620-2811

Abstract: The St. Johns River, located in Northeast Florida, USA, is a large watershed characterized by relatively flat topography, porous soils, and increasing urbanization. The city of Jacksonville, Florida is located near the downstream terminus of the river near the Atlantic Ocean. The lower portion of the watershed located downstream of Lake George is subjected to tidal exchange and storm surge from tropical storms and hurricanes as well as extra-tropical winter storms. Extreme flood events in the Lower St. Johns River can be caused by rain-driven runoff, high tide, storm surge or any combination of the three. This study examines the range of potential extreme flood discharges caused by rain-driven runoff within six larger sub-basins located in the Lower St. Johns River. The study uses multiple methods including published flood insurance data, two statistical hydrology methods, and model simulations to estimate an array of flood discharges at varying return frequencies. The study also examines the potential effects on flood discharges from future land use changes and the temporal distribution of rainfall. The rain-driven flood discharge estimates are then fit to a normal distribution to convey the overall risk and uncertainty associated with the flood estimates. The study also proposes a new methodology to estimate rain-driven flood discharges using existing numerical models of each the six sub-basins prepared by the Saint Johns River Water Management District. Overall, the study revealed a wide range of reasonable rainfall-driven flood estimates are possible using the same data sets. The wide range of estimates will help inform future resiliency projects planned in the study area by providing a more realistic set of bounds with which planning can proceed. The estimates derived herein can be combined with the independent or dependent effects of tide and storm surge in order to characterize the total flood resiliency risk of the region.

Keywords: St. Johns River; model; surface water; system model; flood simulation

---

1. Introduction

Extreme flood estimation is a continuously evolving field of research. It is vital in a world where urbanization, sea level rise, and climate change are prevalent. Accurate flood estimation methods are a powerful tool in securing economical and community wellbeing. Research is underway at the University of North Florida in Jacksonville, Florida USA to assess extreme floods in six critical sub-basins of the St. Johns River. These include Black Creek, Julington Creek, Durbin Creek, Big Davis Creek, Ortega River, and Pablo Creek. The focus of this paper is the Pablo Creek sub-basin. Future work will discuss the remaining basins under study. The 10-, 25-, 50-, and 100-year return frequency floods (e.g. the 10%, 4%, 2%, and 1% annual exceedance probability floods) have been assessed using multiple methodologies including: the use of the St. Johns River Water Management District's (SJRWMD) HSPF hydrologic model, statistical computations using of the Log-Pearson Type III and Power Law distribution, and analysis of existing Federal Emergency Management Agency (FEMA) Flood Insurance Study (FIS) estimates. Sensitivity of parameters such as land-use change, precipitation frequency values (median versus 90th percentile), and rainfall temporal distribution (uniform versus Synthetic Type II Modified) were assessed in the resulting extreme flows determined from the HSPF Model. The use of the existing SJRWMD's HSPF model has not previously been used to estimate extreme flood flows. The SJRWMD's HSPF model was originally programmed for their Water Supply Impact Study to estimate water withdrawal limits in the St. Johns River, but it was reprogrammed and modified for the use of extreme flood estimation as part of this research effort. The model reprogramming methodology has the potential of being implemented in any sub-basin along the St. Johns River in Florida where an existing HSPF model has been developed. Figure 1 depicts an outline of the SJRWMD's HSPF model area along with the sub-basins of interest.

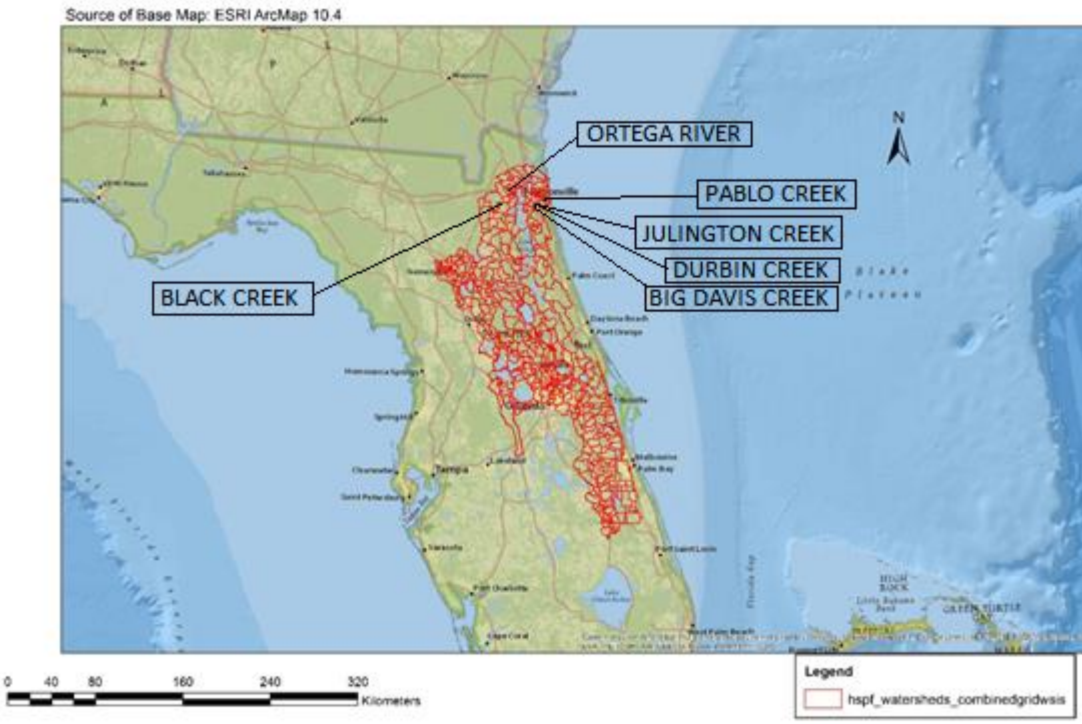

Figure 1. Project study area covering the St. Johns River watershed in Florida, USA.

2. Materials and Methods

In order to obtain the 10-, 25-, 50-, and 100-year flood flow estimates at the sub-basins of interest, the original SJRWMD HSPF models were modified to simulate the scenarios of interest. The general procedure associated with these modifications involved simulating relevant precipitation events, understanding where the simulated precipitation should be incorporated into the model, adding antecedent moisture conditions, and consideration of rainfall temporal distribution and land-use change. Various reaches in each sub-basin are represented in the models. The models were programmed to simulate land-use conditions from 1995 and those projected for 2030. The 1995 land-use condition (base condition) is based on 1994 and 1995 color-infrared aerial photography of the entire SJRWMD (SJRWMD, 2012). The 2030 future condition is the SJRWMD's "planning horizon" and projects land-use changes and urbanization into the future. The 1995 land-use condition was selected as the primary land-use for which each sub-basin was assessed at. The variation in land-use was assessed by performing the simulations at Julington Creek, Dubin Creek, and Big Davis Creek at both the 1995 and 2030 land-use condition. From there, a comparison of results provides insight into the effects that varying the land-use has on the resulting flood flows from those sub-basins. The difference in peak flood flows also gives a general indication regarding urbanization effects within any sub-basin in the St. Johns River watershed.

To produce the 10-, 25-, 50-, and 100-year flood flows, the 24-hour duration, 10-, 25-, 50-, and 100-year frequency precipitation events were simulated – causing the flood flows to be rainfall driven. The process of simulating the given frequency precipitation events involved identifying the appropriate precipitation frequency value to simulate on a specific date. Daily and hourly precipitation data over the course of up to about 100 years from various rainfall gages contained in the St. Johns River watershed are incorporated into the model. Each model run process relies on the daily and hourly precipitation data to complete each simulation. The pre-determined precipitation frequency values were added directly to the daily and hourly precipitation data on a specific target date. The target dates were selected by first determining the 50th percentile flow rates in the sub-basin of interest. From there, 10 target dates were selected within a 15% range of accuracy of the median (50th percentile) flow rate. The 50th percentile flood is a standard baseline for various flood frequency analysis procedures (Malamud and Turcotte, 2006). The 10 target dates were selected in varying months of the year to account for the varying rainfall conditions occurring throughout the seasons. Using the Theissen polygon network established by the SJRWMD (2012), the rainfall gages of interest for each sub-basin were identified. The median and 90% percentile 24-hour duration, 10-, 25-, 50-, and 100-year return frequency precipitation values were recorded at each gage of interest using the NOAA Atlas 14 dataset (NOAA, 2005). The 90th precipitation frequency values were simulated at each sub-basin. The median 24-hour precipitation values were also simulated in the Julington Creek, Durbin Creek, and Big Davis Creek sub-basins to understand the effect on varying precipitation magnitudes.

Antecedent moisture conditions (AMC) were also simulated. Three levels of AMC exist: AMC-I for dry, AMC-II for normal, AMC-III for wet conditions (SJRWMD, 1985). According to Schiariti (n.d.), AMC II is considered for modeling purposes because it is essentially the average moisture condition. 2.1 inches of rainfall over the course of five days was simulated before each target date (SJRWMD, 1985) to represent the AMC.

A uniform and synthetic temporal precipitation distribution were also assessed. According to Suphunvorranop (1985), the Type II Modified is representative of Florida specifically, so that is what was implemented in this research effort. The precipitation was modeled following a uniform distribution in all model runs. However, to determine the effects on the flood magnitude predictions resulting from varying precipitation temporal distributions, the Type II Modified distribution was simulated in the Julington Creek, Durbin Creek, and Big Davis Creek sub-basins paired with the 1995 land-use conditions, 90th percentile precipitation, and AMC. After an analysis of the resulting model results, statistical methods were then utilized to develop additional flood estimates.

The statistical Log-Pearson Type III analysis was performed on both real data (observed data from real streamflow gages) and synthetic data (gages that were modeled in HSPF). Using the synthetic streamflow data is advantageous because a longer period of record was at times observed compared to the real gauged streamflow data. Using information complied by Oregon State University (2005) The following equation was used to calculate the LP3 distribution:

$$log\,x = \overline{log\,x} + K\sigma_{log\,x} \quad (1)$$

where x is the flood discharge value of some specified probability
log x represents the discharge values
K is the frequency factor
And σ is the standard deviation of the log x values.

The frequency factor, K, is a function of the skewness coefficient and return period. From there, the maximum flow (Q) for each water year was determined. This information was then ranked from the largest discharge value to the smallest discharge value and each streamflow value was ranked from 1 to n, which is the total number of values included in the dataset. Next, the log of each annual peak streamflow was obtained and defined as log(Q). The average of every Q and the average of every log(Q) was computed. The following computations were conducted for every water year:

$$\log(Q) - average(\log(Q))^2 \quad (2)$$

$$\log(Q) - average(\log(Q))^3 \quad (3)$$

Next, the return period was calculated using the Weibull plotting position presented in Malamud and Turcotte's (2006) research. The Weibull plotting position provides the recurrence interval in years with the following equation:

$$T = \frac{N_{WY} + 1}{N_C} \tag{4}$$

where, Nc is the rank and NWY is the number of water years in the data set.

Next, the final calculation was completed by determining the exceedance probability of each discharge value with the formula:

$$Exceedence\ Probability = \frac{1}{T} \tag{5}$$

The sum of the values computed for Eq. (2) was determined as well as the sum of the values computed for Eq. (3). Then, the variance, standard deviation and skew coefficient were calculated using the equations below:

$$\frac{\sum_i^n((logQ - avg(logQ))^\wedge 2}{n-1} \tag{6}$$

$$\sigma_{\log x} = \sqrt{\frac{\sum(logx - \overline{\log x})^\wedge 2}{n-1}} \tag{7}$$

$$skew\ coeff. = \frac{n * \sum_1^n(\log(Q) - average(\log(Q)))^\wedge 3}{(n-1)(n-2)(\sigma \log(Q))^\wedge 3} \tag{8}$$

An appropriate frequency factor table (Haan, 1977) was used along with the calculated skew coefficient to find the k-values. The following equation was used to calculate the 10-, 25-, 50-, and 100-year discharges:

$$\log(Q(T) = avg(\log(Q)) + [K(T, C_S)] * \sigma\log(Q) \tag{9}$$

The Power Law (PL) is the second selected statistical method for flood flow estimation in this research. Like the Log-Pearson Type III method, real data and synthetic data was assessed. Referencing the work of Malamud and Turcotte (2006), the maximum streamflow value (Q) for every given year of water data was sorted from largest to smallest. The data was assigned a ranking value, N_C, which was used to determine the Weibull plotting position return recurrence interval, T. N_C is ranked as 1, 2, 3, …, N_{WY} and T is defined as:

$$T = \frac{N_{WY}}{N_C} \tag{10}$$

The log function was applied to all peak streamflow values and all T values. Then, a scatterplot of log(T) versus log(Q) was created. A linear regression trendline was determined following the generalized power law equation:

$$logQ[T] = \alpha \log(T) + \log(C) \tag{11}$$

The trendline of the scatterplot provided the initial estimate for the α and C regression coefficients. The α coefficient was identified as the slope of the trendline equation. The C coefficient was identified as the y-intercept of the trendline equation. Once these coefficients were determined, the discharge value of the 10-, 25-, 50-, and 100-year flood was calculated.

Additionally, the Generalized Reduced Gradient (GRG) Nonlinear regression methodology was used with the Solver (Microstoft Excel, 2016) plug-in to verify the graphical derivation of the α and C coefficient. The least squared method was implemented by first assuming an initial guess where the α and C coefficient are greater than 0.01. Then, the modeled Q values were calculated using the estimated α and C coefficient using the general PL equation:

$$Q[T] = CT^\alpha \tag{12}$$

From there, the sum of squared differences was obtained using:

$$sum\ of\ the\ squared\ differences = \left(sum(Q) - sum(Qmodeled)\right)^2$$

Then, the Solver (Microsoft Excel, 2016) plug-in was used to minimize the sum of the squared differences while iterating for the most ideal values of α and C.

3. Results

A comparison of results was conducted at each reach in each sub-basin. Table 1 depicts a sample of the results obtained in the Pablo Creek sub-basin. Reach 8 is the modeled location of the Pablo Creek sub-basin outlet into the Jacksonville, Florida intracoastal waterway. The HSPF model did not include a simulated gage, however real gage data was available. The adjusted FEMA FIS flow estimates for Reach 8 were obtained by comparing the modeled basin area to the estimated basin area from the FIS. The FIS reported a drainage area of approximately 119 square kilometers (FEMA, 2014) and the modeled area was approximately 98 square kilometers. The HSPF model area is approximately 82% of the drainge area reported in the FEMA FIS. Therefore, the adjusted FEMA FIS estimates represent 82% of the reported FEMA FIS flow values. The flood estimates for the real gage were not adjusted because the modeled gage does not have a drainage area association in HSPF. Therefore, it was assumed that the modeled area that the synthetic gage encompasses is the same as the drainage area that the real gage encompasses in the FEMA FIS. The HSPF model estimates and the statistical estimates were overall low compared to the FEMA FIS estimates in Pablo Creek. However, the strong tidal component of this sub-basin greatly influences the results. It is hypothesized that the HSPF model does not take into consideration the tidal influence to the extent of the FEMA FIS.

Table 1. Comparison of estimates at Pablo Creek (in m^3/s).

Location	Return Frequency	1995 Land-use, 90th Percentile Precipitation, and AMC	Log-Pearson Type III	Power Law – Linear Regression	Power Law – Nonlinear Regression	Adjusted FEMA FIS Estimate
Reach 8	10-year	59	56	88	56	110
	25-year	76	75	109	94	167
	50-year	86	91	128	138	213
	100-year	99	108	150	203	249
Pablo Creek Real Gage	10-year	NA	28	33	28	108
	25-year	NA	37	68	47	139
	50-year	NA	43	115	70	171
	100-year	NA	49	196	102	200

0verall, it was found that simulating 90th percentile precipitation frequency values compared to the median precipitation frequency values resulted in statistically significantly higher flood flow estimates, the difference between following a uniform versus synthetic Modified Type II rainfall distribution does not produce drastic differences in flood flow estimates, and the simulation of the 2030 land-use condition over the 1995-land use consistently produced higher flood flows across the watershed.

After plotting the results of each method assessed in each sub-basin's critical locations, the normal distribution was applied to the results by computing the mean and standard deviation of the data set to produce bell-shaped curves (Smantary and Sahoo, 2020). The normal distribution bell curves portrayed a strong presence of a normal distribution in the flood frequency estimates.

The Ortega River and Pablo Creek sub-basin HSPF models ran well. The 1995 land-use, 90th percentile precipitation frequency values, uniform distribution, and antecedent moisture conditions were considered in the model runs. However, it would be beneficial to assess the 2030 land-use condition paired with the 90th percentile precipitation frequency values and Type II Modified rainfall distribution for all sub-basins where it was not assessed. This combination of parameters yields the highest flood flows.

4. Conclusions

In a case study conducted by Ninov et al. (2008), the results of their HSPF modeling for flood assessment yielded flood flows that were 130% higher than the historical flood flows, while the modeled annual, seasonal, and low flows were approximately 25% to 33% less than the observed respectively. This is an interesting perspective to considering the research of Ninov et al. (2008) produced flood flows that were too high while the results of the HSPF flood modeling presented in this research appears to be too low compared to FEMA FIS estimates.

The Log-Pearson Type III (LP3) statistical computations were successful. Varying results were obtained in each sub-basin. The LP3 based flood estimates were greater than the FEMA FIS estimates in some sub-basins and less in others. The Power Law (PL) statistical computations were mostly successful. The PL derived flood estimates were typically much higher than the LP3 results and the HSPF modeled results. The PL derived flood estimates were even higher than the adjusted FEMA FIS estimates at times. The PL distributions produced more reasonable estimates for the 10- and 25-year flood estimates and appeared to massively overestimate the 50- and 100-year flood flows. The PL was selected because of the praise it received in various research studies for being a simple and effective method. However, there was certainly a caveat that the PL performs best with a larger data set (Kidson and Richards, 2005). The difference in computing the PL distribution regression coefficients using the Linear Regression versus Nonlinear Regression method produced varying results. It was evident that the data sets were better suited for one method over the other in certain cases.

The use of the FEMA FIS estimates proved to be an asset as the estimates were derived by qualified professionals. There is a degree of validity in comparing the methodologies assessed in this research to the FEMA FIS estimates. However, it has also been established that the FEMA FIS estimates were all obtained using varying methods (FEMA, 2011, 2013, 2013), which could explain why the other methods in this researched produced estimates either greater or less than those presented in the FIS. Research conducted by Okoli et al. (2019), compared statistical and hydrological methods for the estimation of design floods based on 10,000 years of synthetically generated weather and discharge data. Although their hydrologic modeling did not reflect any real applications and was intended as a baseline for discussion for comparison of results, their ultimate findings suggest that more than one flood estimate should be obtained and the maximum value (within reason) should be selected to minimize the likelihood of underestimating the design flood (Okoli, 2019).

In conclusion, this research has developed a new methodology for producing flood estimates. The modification of the St. Johns River Water Management District's HSPF model to estimate flood estimates is a brand-new methodology. Several of the HSPF models need to be expanded for sub-basins where the extreme flood flows exceed the model flow capacity. However, reasonable flood estimates can still be obtained from this new methodology in every sub-basin belonging to the St. Johns River. Current existing flood flow estimates are typically established as a one-value estimate per return frequency as seen in the FEMA Flood Insurance Studies. Additionally, the selected methodologies from which their (FEMA FIS) estimates were obtained are not always consistent. It is suggested that future extreme flood estimation procedures include the assessment of multiple methodologies to minimize the risk of underestimating design floods. This research is unique in producing a set of estimates for the 10-, 25-, 50-, and 100-year floods for the Black Creek, Julington Creek, Durbin Creek, Big Davis Creek, Ortega River, and Pablo Creek sub-basins based on hydrologic modeling, statistical analysis, and comparison to existing flood estimates.

Author Contributions: Graduate student Samantha Kovalenko led the research efforts for this article as part of an ongoing research program led by Drs. Chris Brown and Cigdem Akan. Drs. Chris Brown, Cigdem Akan, and Alexandra Schonning are the research committee of Samantha Kovalenko. Graduate students Kovalenko assisted with model development and integration and it was overseen by Dr. Brown. This article was written by Samantha Kovalenko with review by Drs. Brown, Akan, and Schonning. Dr. Brown provided project administration and final article editing.

Funding: This research was funded by funds provided by Dr. Brown, the UNF Environmental Center and Battelle Labs.

Acknowledgments: The authors would like to acknowledge help and support from the St. Johns River Water Management District who provided historical model datasets and map coverages used for this study.

Conflicts of Interest: The authors declare no conflict of interest. No one other than the authors had a role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Federal Emergency Management Agency (2013). *Flood Insurance Study: Duval County, Florida.* (12109CV000C). Retrieved from https://fris.nc.gov/fris_hardfiles/FL/FIS/12109CV000C.pdf
2. Federal Emergency Management Agency (2013). *Flood Insurance Study: Duval County, Florida.* (12031CV001A). Retrieved from https://fris.nc.gov/fris/
3. Federal Emergency Management Agency (2014). *Flood Insurance Study: Clay County, Florida.* (12019CV001A). Retrieved from https://fris.nc.gov/fris/
4. Haan, C. T. (1977). *Statistical methods in hydrology*. Ames: Iowa State Univ. Pr.
5. Kidson, R. & Richards, K. (2005, July 1). Flood frequency analysis: assumptions and alternatives. *Progress in Physical Geography,* 29(392). Retrieved from https://journals.sagepub.com/doi/abs/10.1191/0309133305pp454ra
6. Malamud, B. & Turcotte D. (2006). The applicability of power-law frequency statistics to floods. *Journal of Hydrology.* 322(2006). Retrieved from https://www.researchgate.net/publication/222655854_The_applicability_of_power-law_frequency_statistics_to_floods
7. Microsoft Corporation (2016). Microsoft Excel (Solver plug-in). Retrieved from https://office.microsoft.com/excel
8. National Oceanic and Atmospheric Administration Atlas 14 (2005). *Precipitation Frequency Values*. Retrieved from https://hdsc.nws.noaa.gov/hdsc/pfds/pfds_map_cont.html
9. Ninov P., Ribarova I. Kalinkov P., & Dimova G. (2008). Application of the HSPF Model for Flood Simulation with Analysis of the Results in Terms of Monitoring Uncertainties/Case Study of the Lesnovska River, Bulgaria. International Congress on Environmental Modelling Software. Retrieved from https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=2858&context=iemssconference
10. Okoli K., Mazzoleni M., Breinl K., & Baldassarre G. (2019). A systematic comparison of statistical and hydrological methods for design flood estimation. Hydrology Research.. 50(6): 1665-16678. Retrieved from https://iwaponline.com/hr/article/50/6/1665/69899/A-systematic-comparison-of-statistical-and
11. Oregon State University (2005). Analysis Techniques: Flood Frequency Example with Daily Data (Log-Pearson Type III Distribution). Retrieved from https://streamflow.engr.oregonstate.edu/analysis/floodfreq/meandaily_example.htm
12. Samantaray S., Sahoo A. (2020). Estimation of flood frequency using statistical method: Mahanadi River basin, India. H2Open Journal. 3(1): 189-207. Retrieved from https://doi.org/10.2166/h2oj.2020.004
13. Schiariti, P., P.E., CPESC (n.d.). Basic Hydrology – Runoff Curve Numbers. Retrieved from http://njscdea.ncdea.org/
14. St. Johns River Water Management District (1985). *A Guide to SCS Runoff Procedures.* (SJ85-5). Retrieved from https://www.sjrwmd.com/documents/technical-reports /
15. St. Johns River Water Management District (2012). *St. Johns River Water Supply Impact Study* (SJ20120-1). Retrieved from https://www.sjrwmd.com/documents/water-supply/#wsis-final-report
16. Suphunvorranop, T. (1985). *A Guide to SCS Runoff Procedures*. United States Department of Water Resources Technical Publication No. 85-5.