TESTING THE DEVELOPMENT OF A SIMPLE EQUATION TO PREDICT CHANGES IN STREAM FLOW FROM CHANGES IN LAND COVER

Kevin Easton
Graduate (M.S.) Professional Paper
Department of Soil, Water, and Ecosystem Sciences

Advisor: Dr. Samuel Smidt
Committee Members: Dr. Mathew Deitch and Dr. Mary Lusk
College of Agricultural and Life Sciences
University of Florida

1. Introduction

The State of Florida has experienced significant and prolific land cover change in recent decades, and it is widely accepted and understood that changes in land cover alters streamflow. Changes in stream flow are largely associated with the conversion of pervious surfaces to impervious surfaces through urbanization, but changes in streamflow are also observed through the conversion of other types of land cover, such as the conversion of forested land to agricultural land. The degree to which streamflow is influenced by conversion of different landcover types at various spatial scales can be quantified using complex hydrological models, but rarely this can be done in a simplistic way useful for applied contexts (e.g., contractors, environmental consultants). The goal of this study was to evaluate a series of basic equations that can be used to estimate changes in streamflow due to land cover change within a watershed. Plainly, I seek to answer questions like: how much will streamflow increase if I convert 5% of the forested land within my watershed to urban? Specifically, this project evaluates how best to estimate streamflow discharge in relation to precipitation, land cover, and other watershed dynamics like groundwater flow. This project builds upon and continues the undergraduate research work of John Flores (2020-2021) who processed and organized the land cover, watershed imperviousness, precipitation, and streamflow data for Florida watersheds from 1996-2020 used here. This study is Part 2 of 3 in a collaborative effort between John (completed Part 1), myself (Part 2), and my adviser, Sam Smidt (Part 3). Collectively, we seek to publish this combined work in 2022-23. Ultimately, the goal of this study is to provide a quick and easy mechanism to assist land planners and land managers with flood mitigation planning, environmental land management, and urban sustainability and resiliency by developing a simple predictive equation developed from openly sourced data.

2. Study Area Background

This study first evaluated watershed scales versus stream gage sites (Figure 1), where gage sites were selected as those with continuous data collected since at least 1996 across a binned range of drainage areas (e.g., 10 from small drainage areas, 10 from medium, etc.; Flores 2021).

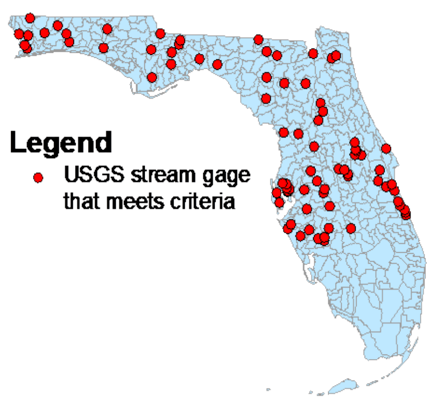

Figure 1. Targeted USGS stream gages within Florida overlaying HUC-10 watershed boundaries for reference (from Flores, 2021).

HUC levels for each gage were then set where the selected gage was located at the outlet of the watershed, so all streamflow based on the hydrological dynamics within a watershed would pass through the gage site (Figure 2). For example, in Figure 2, the stream gage would not be in a location conducive for streamflow analysis at HUC-8 scale since its position is not near any particular outlet of the watershed. However, at a HUC-10 scale, the stream gage is positioned near the discharge outlet of a watershed, which indicates that it would likely capture the total discharge from within the watershed (Flores, 2021). In total, 20 out of the 81 stream gauges were equally selected within HUC-10 and HUC-12 watersheds within the State of Florida.

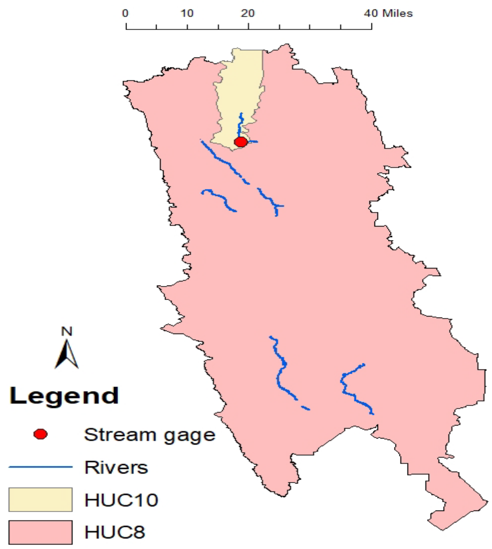

Figure 2. Stream gage positioning at separate HUC scales in ArcMap (from Flores, 2021).

Conceptually, this project assumes streamflow at the watershed outlet (gage site) is a function of precipitation across varying land cover with distinct imperviousness. I then started with the multiple linear regression equation (eq. 1):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \varepsilon_i \qquad \text{(eq. 1)}$$

where the dependent variable, Y_i, is gage streamflow, β is an imperviousness coefficient for land cover, X. A main question then becomes: where does precipitation fit within this equation? Furthermore, how are precipitation and streamflow related so that the hydrological variability across changing land cover areas (e.g., urban, forest, wetland) can be captured (e.g., some watersheds have 5% urban and others have 25%, and these dynamics heavily influence runoff and subsequent streamflow)?

3. Equation Testing

3.1. Rational Equation Development

The Rational Equation is used by hydrologists to predict peak stream discharge for a drainage area during the peak-period of a precipitation event. The Rational Equation proves that stream discharge (Q) is proportional to land cover type (i.e., runoff coefficient), rainfall intensity, and watershed area. The Rational Equation is written as $Q = ciA$, where Q is peak discharge (in cfs), c is the runoff coefficient, i is rainfall intensity (in inches/hour) and A is the drainage area (in acres). I then expanded this equation, modeled after the multivariate equation, to isolate individual land cover classifications (eq. 2):

$$\frac{S}{A*P} = m_1\left(A_{lc_1} * M_{lc_1}\right) + m_2\left(A_{lc_2} * M_{lc_2}\right) \ldots + b \qquad \text{(eq. 2)}$$

where, S is peak streamflow measured in volume per time at each stream gage, P is precipitation across a watershed measured in depth per time, A is the total watershed area, m_x is the predictive coefficient of streamflow influence, A_x represents the percent site area of a land cover class, M represents an average impervious surface coefficient of the land cover class, lc_x. The right side of the equation quantifies a "sum of parts" function ($\sum x$), which accounted for different land cover types and coverage area (totaling 100%) within the selected watershed. In comparison to the Rational Equation, the S variable in *Eq. 2* represents the Q variable in the Rational Equation, and the A variable in *Eq. 2* represents the A variable in the Rational Equation, and the P variable in *Eq. 2* represents the i variable in the Rational Equation. The *Eq. 2* variables, A and P, have been divided to the left side of the equation to yield a dependent variable of $\frac{S}{A*P}$. This could also be accomplished using the Rational Equation by dividing i and A to the left side of the equation and leaving c (runoff coefficient) on the right side of the equation, yielding equation 3:

$$\frac{Q}{i*A} = \sum c \qquad \text{(eq. 3)}$$

This allows for calculation of the runoff coefficient c using stream discharge (Q), rainfall intensity (i), and drainage area (A) data, like in *Eq. 2*; however, the Rational Equation only solves for peak discharge using peak rainfall data, whereas *Eq. 2* is intended to measure changes in streamflow from changes in land cover. The water-index (left side of the equation) in *Eq. 2,* ($\frac{S}{A*P}$), represents stream discharge normalized by the product of precipitation and the watershed area. Discharge was normalized to the watershed area to account for the various sizes of watersheds and the difference in resultant streamflow.

3.2. Equation Assessment

Eq. 2. Assessment 1

The calculations of the dependent (water-index; left side of the equation) and independent variables (right side of the equation) for *Eq. 2* were assessed over a random 3-day period in August of 2001 and 2006. Precipitation was measured as cumulative rainfall (m/3-days) over the 3-day interval and streamflow (m^3/3-days) was extrapolated from peak discharge (m^3/s) during the same 3-day interval using dimensional analysis and converting seconds into 3-days. These values were input into the linear regression model to predict the constant coefficients (m_x) and slope-intercept (b) values, using R-squared as a gut-check measure of effectiveness where half of the data was used to build the model and predicted against the other half (Table 1).

Table 1. R-Squared Values for Linear Regression of Equation 2 – Assessment 1

Period	R-Squared Value
3-Day	0.679

The R-squared value for *Eq. 2 - Assessment 1* yielded a decent statistical fitness between the independent and dependent variables assumed in *Eq. 2* given highly variable precipitation and the random selection of days.

Eq. 2. Assessment 2

Based on the favorable 3-day results, I then decided to evaluate different time periods that may better capture the hydrology of the analyzed watersheds. I then re-calculated the equation using data over a 1-day, 5-day and 7-day periods to determine trends in fitness over longer time intervals. Precipitation was measured as cumulative rainfall (m/day) over the 1-day, 5-day, and 7-day period and streamflow (m^3/day) was extrapolated from peak discharge (m^3/s) during the same 1-day, 5-day, and 7-day periods using dimensional analysis and converting seconds into 1-day, 5-day, and 7-day periods. These values were input into the linear regression model to predict the constant coefficients (m_x) and slope-intercept (b) values, and R-squared values were reported (Table 2).

Table 2. R-Squared Values for Linear Regression of Equation 2 – Assessment 2

Period	R-Squared Value
1-Day	0.694
5-Day	0.784
7-Day	0.791

The R-squared value for *Eq. 2 - Assessment 2* yielded stronger statistical fitness than *Eq. 2 – Assessment 1*, suggesting that water flow in response to a precipitation event can take longer than 3 days to make it through the watershed and to the gage site outlet.

Eq. 2. Assessment 3

Given the improved correlation, I then continued to expand the time intervals to find the peak R-squared, which is likely to best represent the hydrological timescales of the watershed dynamics. Data were then processed in daily intervals until the R-squared appeared to plateau (Table 3).

Table 3. R-Squared Values for Linear Regression of Equation 2 – Assessment 3

Period	R-Squared Value
1-Day	0.694
2-Day	0.764
3-Day	0.679
4-Day	0.628
5-Day	0.784
6-Day	0.710
7-Day	0.791
8-Day	0.722
9-Day	0.723
10-Day	0.723

Collectively, the 5 to 7-day range generated the strongest correlations (R-squared of near 0.8).

3.3. Other Equation Attempts

I also tested a simple quotient relationship between streamflow and precipitation (eq. 4):

$$\frac{S}{P} = m_1\left(A_{lc_1} * M_{lc_1}\right) + m_2\left(A_{lc_2} * M_{lc_2}\right) \ldots + b \qquad \text{(eq. 4)}$$

where, S is peak streamflow measured in volume per time at each stream gage, and P is precipitation across a watershed measured in depth per time. The right side of the equation combined multiple variables quantifying a "sum of parts" function ($\sum x$), which accounted for different land cover types and coverage area (totaling 100%) within the selected watershed.

Here, m_x is the predictive coefficient of streamflow influence, A represents the percent site area of a land cover class, M represents an average impervious surface coefficient, lc_x represents each unique land cover type, and b represents a constant term. The multivariate linear regression model was then used to calculate the predictive coefficients (m_x) and the slope-intercept (b). Conceptually, eq. 4 suggests there is a predictive relationship between peak streamflow and precipitation, and this relationship is dependent on the land cover classifications within the watershed.

Eq. 4. Assessment 1

Here, precipitation was measured as cumulative rainfall (m/3-days) over the 3-day interval and streamflow (m^3/3-days) was extrapolated from peak discharge (m^3/s) during the same 3-day interval using dimensional analysis and converting seconds into 3-days, and R-squared again measured the goodness of fit (Table 4). However, this equation attempt had poor predictive performance and did not well-represent the watershed hydrology.

Table 4. R-Squared Values for Linear Regression of Equation 4 - Assessment 1

Period	R-Squared Value
3-Day	0.04

Eq. 4. Assessment 2

Using the same equation, precipitation was again measured as cumulative rainfall (m/3-days) over a 3-day interval; however, streamflow (m^3/3-days) was measured as the change in stream flow (ΔS) between day 1 and day 3. These values were input into the linear regression model to predict the constant coefficients (m_x) and slope-intercept (b) values, which demonstrated a much stronger correlation (Table 5).

Table 5. R-Squared Values for Linear Regression of Equation 4 - Assessment 2

Period	R-Squared Value
3-Day	0.699

Eq. 4. Assessment 3

Lastly, I then expanded the temporal range of this approach to evaluate best goodness of fit (Table 6). Here, I again found reasonably high R-squared values, but they were still less than the rational equation approach. Like the Rational Equation approach, the R-squared values peaked in the 5-day to 7-day period, providing further support of a correlation in streamflow response at the 5-day to 7-day range.

Table 6. R-Squared Values for Linear Regression of Equation 4 – Assessment 3

Period	R-Squared Value
1-Day	0.601
3-Day	0.699
5-Day	0.765
7-Day	0.759

4. Limitations

For applied purposes, the results from this study are largely specific to the state of Florida. For example, the residence time of water within the watersheds seemed to be somewhere in the 5-to-7-day range. However, Florida is a flat landscape with sandy soils. Residence time would likely change based on topography and soil type, which means the initial steps of data collection, data processing, and equation evaluation would need to be completed for each region outside of Florida. Likewise, precipitation patterns are widely variable throughout the country which can also drive differences from these results.

Additionally, this study focuses on peak streamflow as a direct alignment with the end-goal of reducing flooding risk (e.g., increase streamflow = higher flood likelihood). Peak streamflow is a threshold value for management decisions but may not generate the highest correlations when assessing the links between streamflow, precipitation, and land cover. For example, using total streamflow on the left side of the equation may generate stronger results.

Despite these limitations, it is believed that the results of this study provided a strong correlation between changes in landcover and resultant changes in streamflow, which will provide a framework in developing a simple predictive equation that utilizes open-source data to assist land planners and land managers with flood mitigation planning, environmental land management, and urban sustainability and resiliency.

5. Conclusion

This ongoing study seeks to develop a simple, “back-of-the-envelope” formula for estimating changes in streamflow relative to changes in land cover during land development projects. Here, I work to identify the most effective arrangement of streamflow, precipitation, and land cover variables that generate the strongest correlation and subsequent predictive power for applied settings. Based on these results, I conclude:

1. Blending a multivariate framework into the Rational Equation provides a considerably strong method for estimate streamflow changes relative to changes in land cover.
2. Capturing a range of residence times within the watershed is critical for identifying the strongest variable approach and equation framework.
3. When trialing equation designs, there appears to be a “yes/no” result instead of a gradual deviation away from being accurate (i.e., an equation is effective or it is not effective; there were no mediocre equations).

References

Flores, J. (2021). *Land Cover Chnage Influence on Florida Streamflow.* Gainesville: Soil and Water Sciences Department (UF).