
Predicting monthly streamflows and their variability from limited historic records in the Tárcoles watershed, Costa Rica.*

Holli Nicole Capps

College of Agricultural and Life Sciences, University of Florida

Faculty mentor: Peter Waylen, Department of Geography

*The author was awarded the prize for the best undergraduate presentation at the Florida Society of Geographers Annual Meeting, 2018, for a presentation based on this research.

Abstract

Lack of instrumentation in many developing nations makes it difficult to gauge available water resources, an essential part of national socioeconomic wellbeing. Costa Rica is a developing country, possessing comparatively good hydrologic data which are limited in time and space. This study establishes a simple model to predicted mean monthly flows and their interannual variability at unmonitored locations, within the densely populated Tárcoles watershed of Costa Rica. Relationships between basin area and percentiles of historic flows are derived as a matrix of monthly flow percentiles and averages that can be transformed into potential flows. The station used to validate the model has limited records and contains apparently abnormal flows. Nonetheless, results indicate considerable success of the model, though it could be further refined and tested with more discharge stations. In turn, the modelling approach could be used to predict monthly flows and their variability within the Tárcoles basin, as well as other countries where information may be lacking.

Keywords: monthly streamflow, prediction, percentiles, Tárcoles, Costa Rica

Introduction

Developing countries often have many questions regarding the vital natural resource of water such as, “How much,” “Where is it,” “What can be used,” and “When is it available?” Answers are needed quickly and efficiently with minimal expenditure of funds and resources. A lack of basic information about water resources makes it difficult for water to be appropriately allocated to specific sectors, including municipalities, agriculture, and industry. Additionally, making decisions with only minimal information available may lead to consequences such as environmental degradation or dependence on foreign sources of food and power. The importance of water cannot be understated to the burgeoning populations of developing nations. With the

increasing urbanization and globalization comes a need for more agricultural land and greater productivity through irrigation. Safe, clean, potable water is essential to the health and sanitation of large urban populations. At a national level, economies can be dependent on water as a cheap renewable energy source (Gleick, 1993; Roberts, 1998). Many countries throughout Latin and Central America, including Costa Rica, rely on hydroelectricity for more than 80% of their national power supply.

Costa Rica's comparatively advanced water resources monitoring provides a suitable test area, while the growing population, increasing urbanization, and linkages to a globalized economy reflect many of the issues facing small developing nations. The purpose of this research is to determine whether likely and rarer (droughts and excess flows) monthly streamflows can be reliably estimated at unmonitored locations of potential interest. Specifically, this study tests whether the percentiles (rare and common occurrences) of monthly streamflows can be predicted at ungauged locations on rivers in the Tárcoles basin, based on limited available historic records.

Study area and data

The upper Tárcoles basin (Figure 1) is one of the most populous areas in Costa Rica, occupying 2,000 km² of the western Central Valley. It is virtually encircled by cordilleras, except to the west (Ballesteros, 2003). Elevations generally lie between 500 and 1500m, with peaks attaining twice that. The broad valley floor supports densities of over 250 people km⁻², including the San José conurbation, where densities exceed 1,000 people km⁻² (Censo de Población, 2011). Monthly streamflow data (m³s⁻¹) are available at eight sites across the upper Tárcoles (Table 1) from Instituto Costarricense de Electricidad (ICE) the national entity responsible for power generation. Basin areas range from 93.3 km² to almost 1750 km² and yield monthly sample sizes from 9 to 36 years.

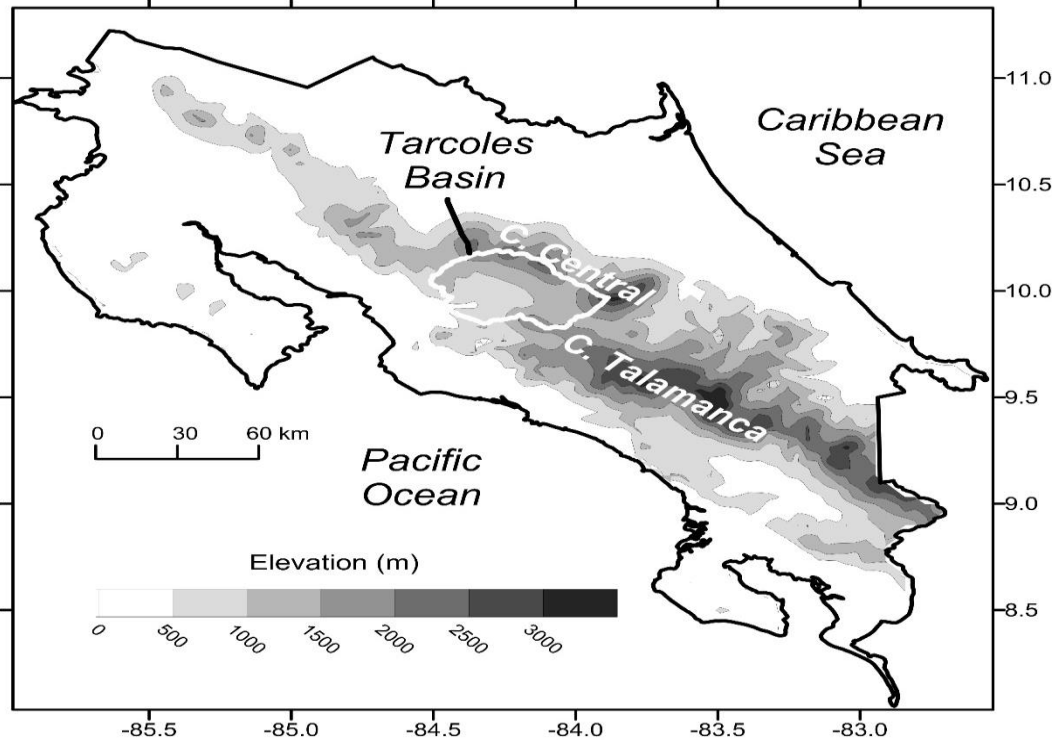


Figure 1. General topography of Costa Rica marking the outline of the study area within the Central Valley. Data source: USGS GTOPO30 <https://lta.cr.usgs.gov/GTOPO30>. Generalized to 1km² resolution and plotted in GoldenSoftware Surfer.

Table 1. Monthly streamflow of the Tárcoles

River	Station	Area (km ²)	Range of number of monthly records
Tiribí	Electriona	302.4	19-20
Alajuela	Alajuela	53.7	9-0
Grande	La Garita	916.0	15-17
Poás	Tacares	201.5	35-36
Tárcoles	Alumbre	1745.4	19-20
Tárcoles	Balsa	1638.0	29-30
Virilla	Linda Vista	93.3	16-18
Virilla	San Miguel	829.0	30-31

Note: Rivers, recording stations and their drainage basin areas of historic observations employed. The right-hand column indicates the range number of years in which records of monthly flow are available. For example, on the Tiribí observations of monthly flows in June are available in 19 years, while October flows are available in 20 years.

The seasonal pattern of rainfall (Figure 2) controls monthly streamflows. Lying at approximately 10°N, Costa Rica is orthogonal to the dominant northeast trade winds which bring moisture from the Caribbean to the eastern flank of the country. They produce dry conditions (rainshadow) on the leeward side of the cordilleras (occupied by the Tárcoles) between November and April, with January through April typically being the driest (Poveda, Waylen, & Pulwarty, 2006). The Inter Tropical Convergence Zone (ITCZ) in the eastern equatorial Pacific migrates northwards during boreal summer, reaching approximately 10°N between May and October. This hemispheric-scale feature brings heavy precipitation and a southwesterly flow to the basin. The rainy season is punctuated by a brief decline in precipitation in late July – early August, locally called the *Veranillos de San Juan*. The phenomenon is related to a temporary strengthening of the northeast trades. After October, the ITCZ migrates southwards. During boreal winter, outbreaks of cold North American air cross the Gulf of Mexico and Caribbean. The higher density of the cool air generally constrains the associated rainfall and northerly winds to the east of the cordilleras, although effects may spill into the headwaters of the Tárcoles through mountain passes.

Despite the surrounding mountains, mean monthly precipitation in any month is reasonably homogeneous across the study area (Figure 2) as is streamflow (Figure 3). Within the Tárcoles basin, flows decline during the dry season, and begin to increase in May. Flows rise through June, however, little of the rain reaches the rivers as most replenishes depleted groundwater and soil water stores which sustained flows throughout the dry season. The *veranillos* induces a leveling, or reduction, of flows in July and August. When intense rains return in September and October, soil and groundwater stores have been substantially replenished by four comparatively rainy months and a higher proportion of rainfall is directed to the rivers. Finally, resumption of the dry season in November begins the decline in flows sustained by the depletion of the hydrologic stores. The one exception to this generalization is Virilla (Linda Vista), which drains a small area (93.3 km²) in the east near a topographic gap in the cordillera, which allows rain associated with *nortes* to penetrate. This appears to have a local effect of sustaining peak flows into November.

PREDICTING MONTHLY STREAMFLOWS AND THEIR VARIABILITY FROM LIMITED HISTORIC RECORDS IN THE TÁRCOLES WATERSHED, COSTA RICA

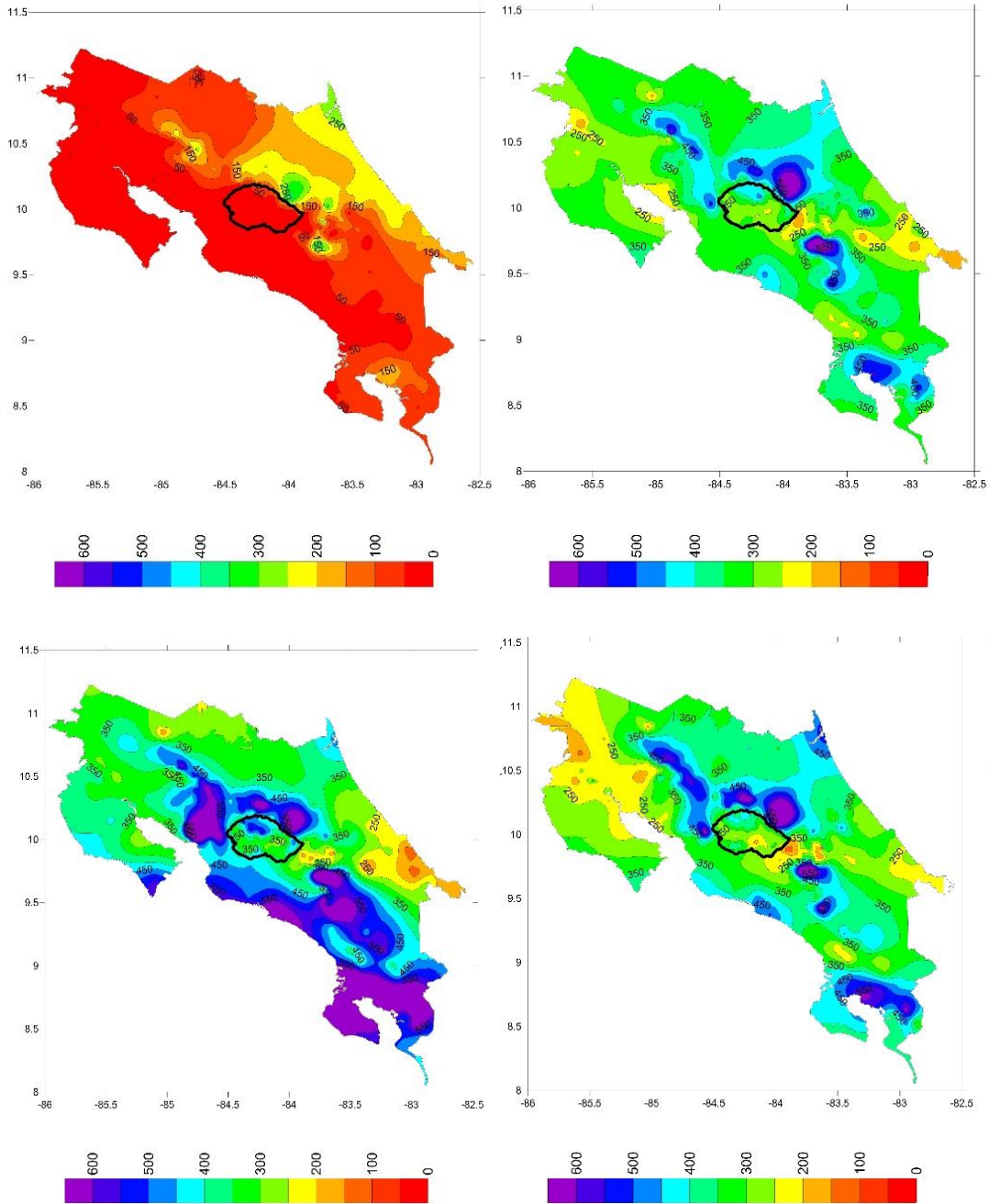


Figure 2. Mean monthly precipitation (mm) over Costa Rica, with the study area highlighted. Clockwise from top left: February (dry season), June (pre-*veranillos* rainy season), October (post-*veranillos* rainy season), and August (*veranillos*), based on all available records. Data source: Instituto Meteorológico Nacional (IMN), San José, Costa Rica.

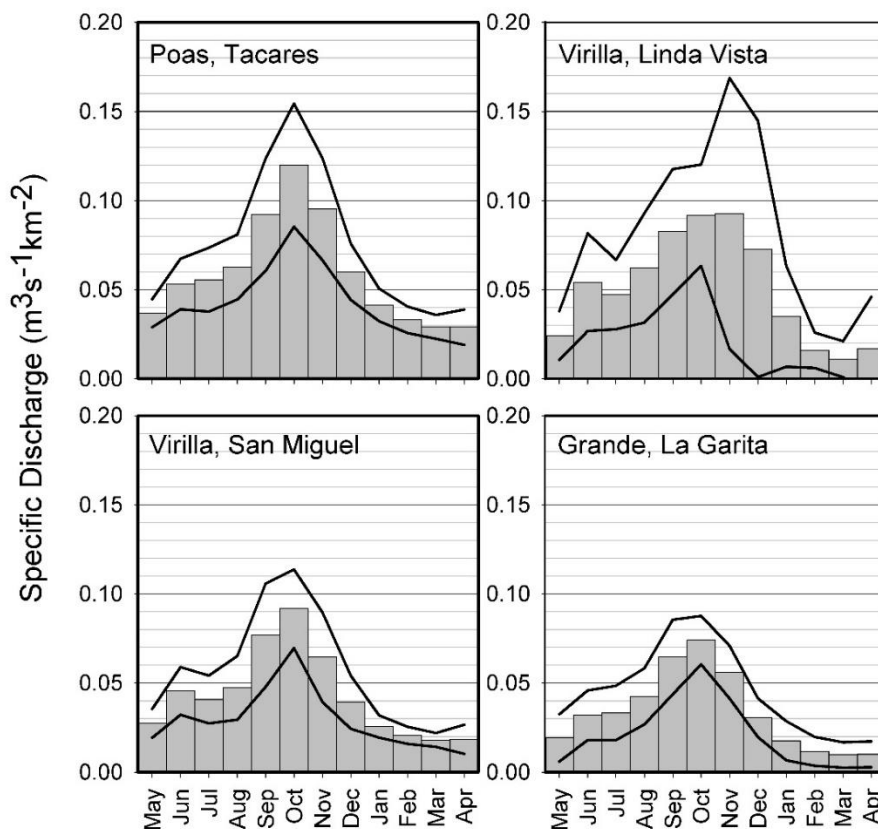


Figure 3. Example plots of mean (bar chart) and mean plus and minus one standard deviation (lines) of observed monthly specific discharge in four basins. Specific discharge (streamflow divided by drainage basin area) is used to facilitate comparisons between larger and smaller basins.

Methods

Data are stored in Microsoft Excel (2013) spreadsheets and analyses are carried out using intrinsic functions within the package.

The relationship between the drainage area upstream of a flow gauging station and the discharge of water, is expressed as the following log/log form (Mimikou, 1984; Leopold, Wolman, & Miller, 2012):

$$Q = a.A^b \tag{1}$$

Where Q is the property of the streamflow of interest, A is the drainage basin area and a and b are coefficients derived from historic data. This approach has been applied to average annual maximum flows (Smith, 1992; Gupta, Mesa, & Dawdy, 1994) and extended to establish the nature of the standard deviation of annual floods (Glaves and Waylen, 1997). However, the latter

is only useful when a particular probability distribution (e.g. normal) can be assumed, which is often not the case for monthly flows (Figure 4). The coefficients are specific to a region within which climate and other factors are reasonably homogeneous. Rasmusson and Perry (2000) incorporated information about precipitation at appropriate time scales (e.g. annual) and other variables, to a multivariate form of equation 1.

$$Q = a.A^b + c.P^d + e.V_1^f \dots \dots \dots + yV_n^z \quad (2)$$

Where P is a measure of precipitation and $V_1 \dots V_n$ are other potentially relevant variables such as elevation. Equation 1 can be expanded to:

$$Q_{\%,i} = a_{\%,i}.A^{b_{\%,i}} \quad (3)$$

Where Q and the coefficients are subscripted with % and *i* to represent the appropriate percentile ($\% = 5^{\text{th}} \dots 95^{\text{th}}$) and month ($i = 1 \dots 12$).

The correlation coefficient, *r*, ($-1 < r < 1$) indicates the goodness of fit of the best fit line to the data, and the coefficient of determination (r^2) conveys the proportion (1-0) of the variance in the dependent variable (flow for a particular percentile) explained by the variance in the independent variable (basin area).

This study considers flow percentiles (5^{th} to 95^{th} in 5 % increments), estimated using the PERCENTILE function, as the properties of streamflow, Q. Figure 4 displays the nineteen monthly percentiles on the Poás River and illustrates both the variability between months (symbols) and between the median and flow extremes encountered from year to year. Monthly plots fall into 3 categories; January-April, flows are low and display little interannual variability; September–November, high levels of flow and interannual variability; remaining months, low variability at the lower percentiles (<60) and greater variability at higher levels. This last group consists of months during which flows transition from low to high flows and vice-versa, reflecting variability is the start and end of the rainy season.

Equation 3 is applied to each of the nineteen flow percentiles extracted from historic data. Regression coefficients (a and b) and coefficients of determination (r^2) are estimate using the REGRESSION function and identified by month and percentile. Estimated slopes and various percentiles provide the basis for estimating flows at unmonitored locations.

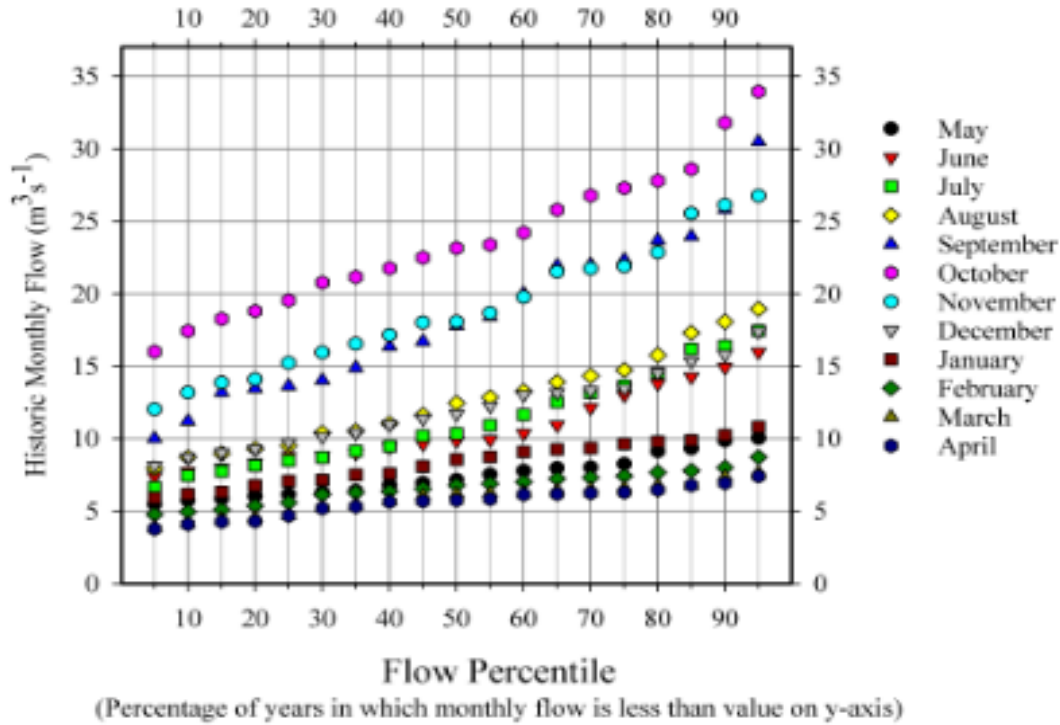


Figure 4. Estimates of the percentiles (5th-95th) of monthly streamflows on the Poás River at Tacares, based on historic data (1952-1987).

Given the small sample size available for validation, no comprehensive statistical testing is possible, however the Root Mean Square Error (RMSE) can provide some comparative measure of goodness of fit.

$$RMSE = \sqrt{\sum_i^k (o_i - e_i)^2} \quad (4)$$

Where i indicates the individual percentile (1,2, ..) and k the total number of percentiles being evaluated (19). The observed flow corresponding to a particular percentile is o and the predicted flow is e .

Student's t-distribution is used to test the null hypotheses that the correlation coefficients and the regression coefficients are not significantly from zero. All significance tests are carried out at the 0.05 level.

Results and validation

Figures 5 and 6 plot monthly mean flows, basin area, and the best fit lines for both the recommended log/log and simple linear plots. All correlation coefficients are significantly different from zero, however the log/log model does not improve fit and the more easily interpreted linear model is adopted:

$$Q (m^3s^{-1}) = m \cdot A(km^2) + b \quad (6)$$

The intercepts, b, are very small and never significantly different from zero. Physically, it is impossible for these perennial streams to have zero or negative flows (implied by negative values of b), or for rivers with “zero” drainage basin area to generate any flow (positive values of b). Thus, on statistical and physical grounds, computations are repeated forcing the best fit line through the origin (b=0). Reduced to this linear form with no intercept, the slope (m) can be viewed as the increment to flow percentile generated by an additional 1 km² of drainage area within the upper Tárcoles and has units of specific discharge (m³s⁻¹km⁻²).

The 228 estimates of slopes and correlations (19 percentiles x 12 months) are displayed as surfaces of specific discharge and coefficients (Figures 7 and 8) with the desired percentile on the x-axis and month on the y-axis. Specific discharge at any percentile reflects the dry season (low productivity), pre-*veranillos* rains (rising productivity), *veranillos* (intermediate productivity) and post-*veranillos* rains (very high productivity) witnessed in monthly flows (Figure 3). Meanwhile, the contrast between estimates of the lowest (left side of graph) and highest values (right) mirrors the slopes of monthly percentile estimates (figure 4). Even where there are seemingly dramatic declines in correlations during the dry season, they remain significant and declines are associated with the paucity of rain rather than errors in the model.

The short (1972-1974) record of historic monthly discharges on the Itiquis river at Itiquis (ICE station 842417, basin area 32 km²) provides an independent validation of the methodology.

It is the only other available record not used in the preceding estimation of coefficients. In each month of record at the validation station, there are only three observed flows (k = 3). A variety of approximations, or “plotting formulae,” exist to assign percentile (probability) to each observation (Cunnane, 1978). The Weibul plotting position is employed:

$$\rho = \frac{m}{(k+1)} x 100 \quad (5)$$

Where ρ is the percentile associated with an observed flow, m is the rank of that flow in the historic record (1= smallest, k = largest), and k is the number of observations.

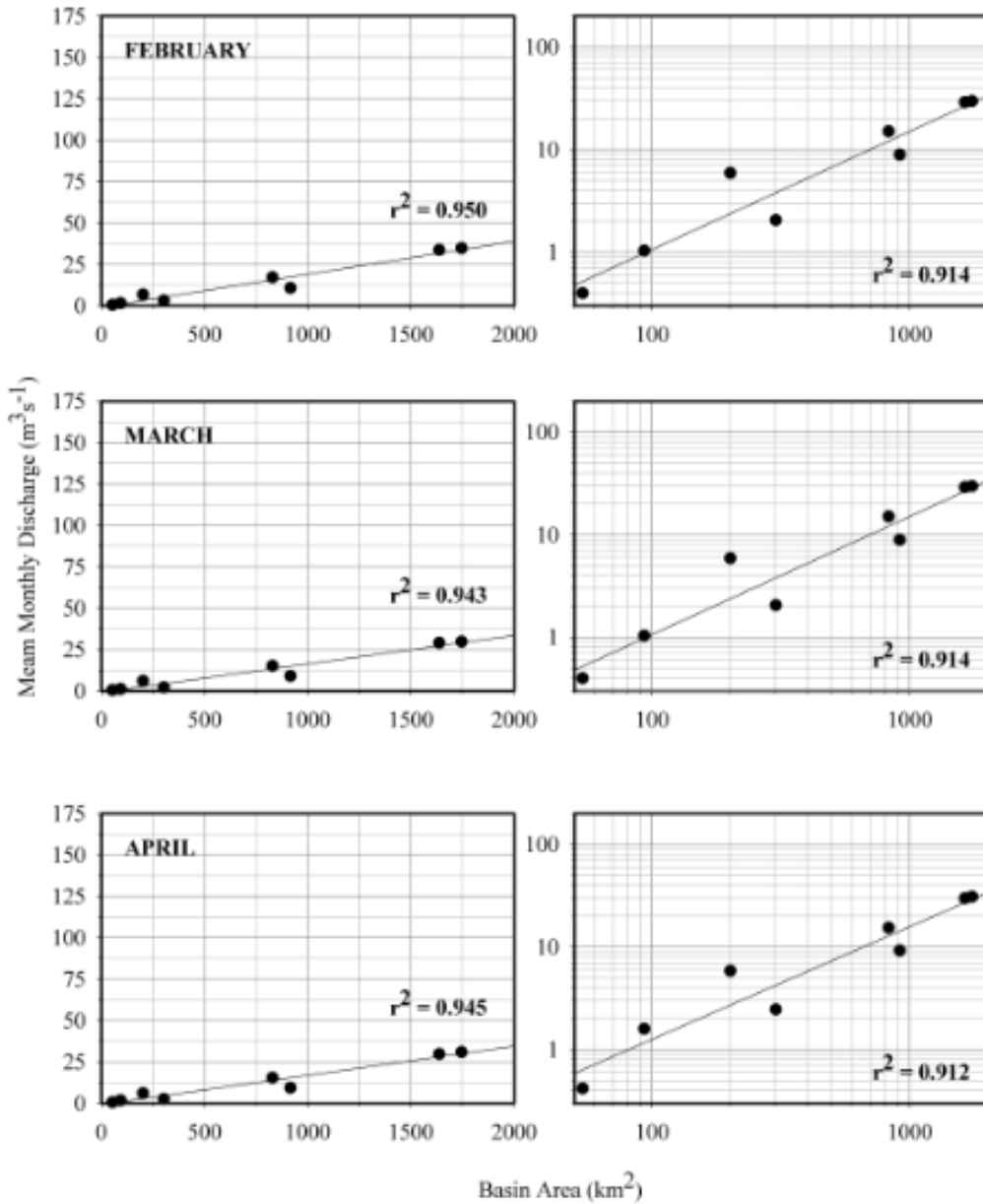


Figure 5. Plots and best fit lines of the relationship between basin area and mean monthly discharge for February through April using both simple linear (left) and log/log (right) models.

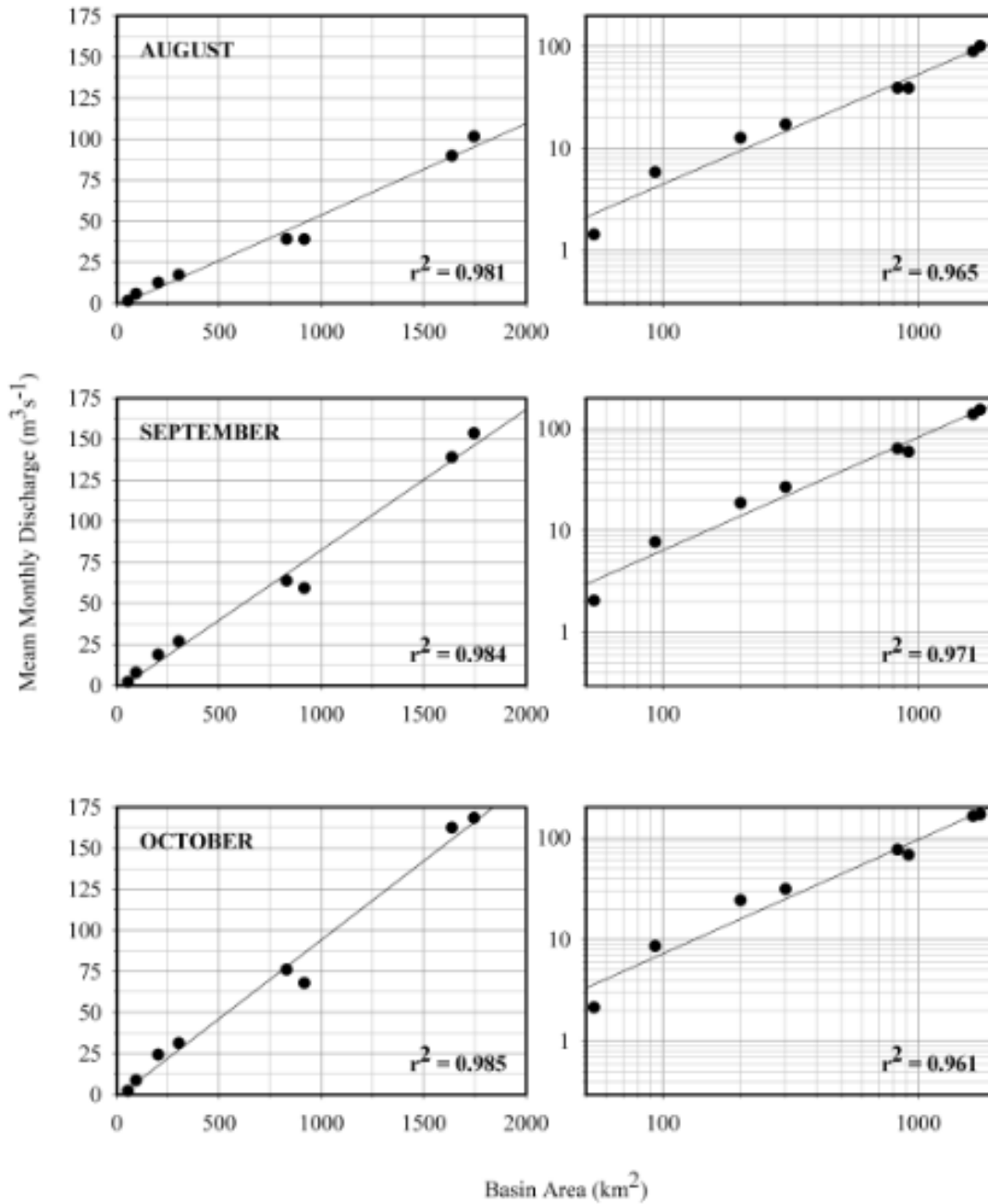


Figure 6. Plots and best fit lines of the relationship between basin area and mean monthly discharge for August through October using both simple linear (left) and log/log (right) models.

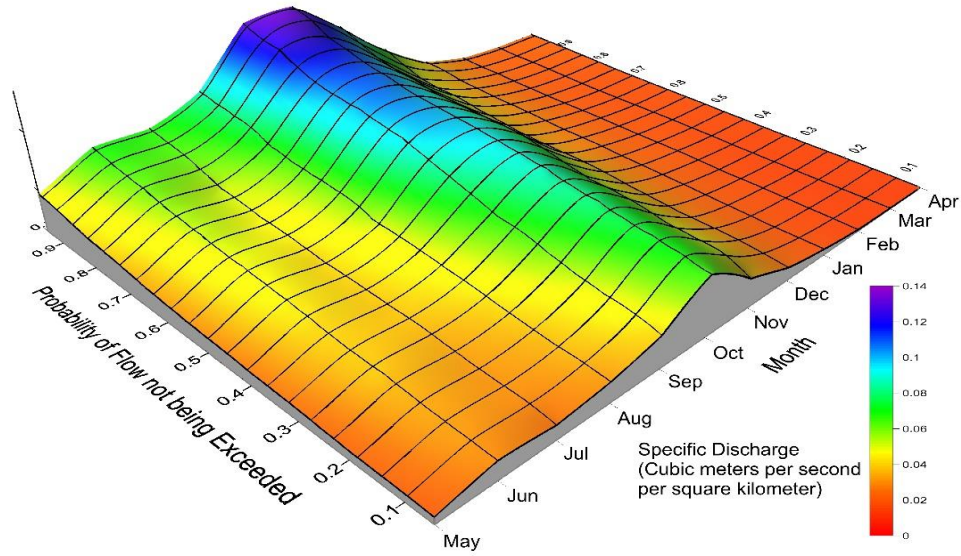


Figure 7. Surface of the estimated slope (coefficient m) of the regression lines (specific discharges) between basin area and monthly flows percentiles throughout the Tárcoles basin.

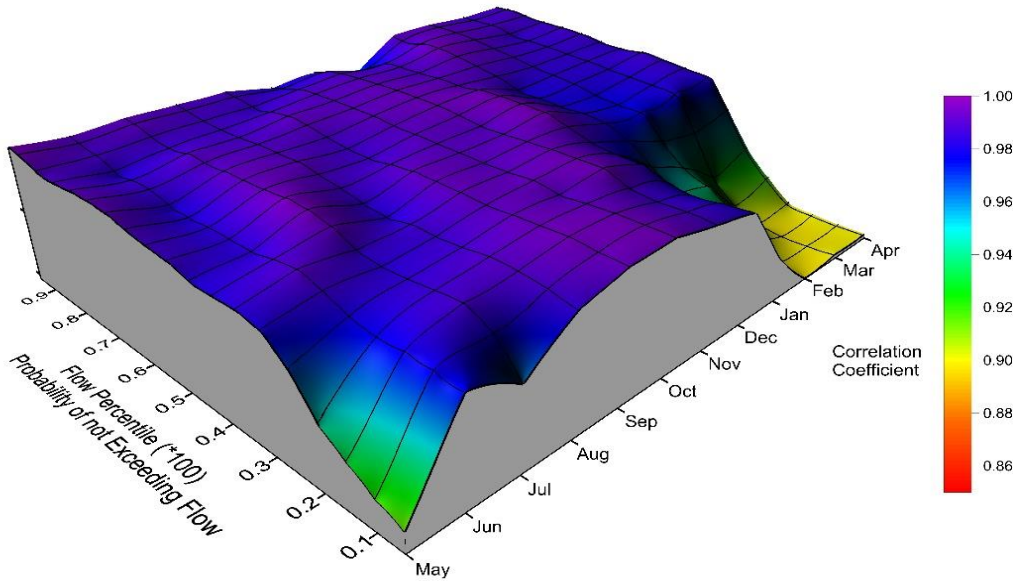


Figure 8. Surface of the coefficients of determination (r^2) of the regression lines between basin area and monthly flows percentiles throughout the Tárcoles basin.

Coefficients (figure 7) at each percentile are multiplied by 32 km² (basin area above the Itiquis station) to derive estimates of monthly flow quantiles. Figure 9 plots estimated monthly quantiles (flows corresponding to specified percentiles) for a selection of common and rarer percentiles upon which are superimposed the observed monthly flows from 1972 to 1974. Because the water year starts in May, dry season months (Jan-Apr) marked as 1972 are actually 1973 calendar months.

The limited validation data set means that the three ranked observations are assigned percentiles of 0.25, 0.5 and 0.75 (smallest to largest flow) consistent with the Weibul formula. Their coordinates are plotted along with the predicted percentiles on figure 10. Equivalent flows are estimated from each profile, and the observed and predicted flows are then used to calculate monthly RMSE (figure 11), which has units of discharge. A comparative measure of error is determined by deriving the ratio of the error ($|\text{observed} - \text{estimated}|$) to the observed monthly mean flow. March has the highest error (0.35), while most months report values at or under 0.25. Given the limited sample size, the model performs well during the wet season, including the *Veranillos* and more poorly in the dry season.

Lack of available data makes statistical validation fraught. Some observed data appear to fall well beyond the expected ranges in so small a sample size (Figures 9 and 10). The 121 year precipitation record at San José sheds light on whether observed flows are as unusual as these estimates suggest or whether the method is flawed. Equation 5 is applied to the San José data and monthly precipitation total observed during water years 1972 to 1974 are assigned empirical percentiles (probabilities). Unusually high (low) monthly flows generally correspond to, or follow, unusually high (low) precipitation totals. In Table 2, monthly observed streamflow, expressed as forecast percentiles on the Itiquis (Figure 9), and rainfall at San José, based on long-term records, are divided into three groups (terciles). The lowest tercile (orange) comprises the lowest third of all observations and the upper tercile (blue), the highest. Inspection reveals partial agreement between months of apparently rare flows on the Itiquis and unusual rainfalls at San José, suggesting that the flows in the three years of observations are as unexpected as the methodology implies.

Discussion

In order to operationalize and test the proposed model, two assumptions have been made. First, the statistical properties of streamflow at these stations remains unchanged over the period of record (stationarity). The small number of available records and their varying periods of records (Table 1) is a weakness. Ideally, identical recent periods of record (perhaps 30 years) would be preferable, however this information is not available for the Tárcoles basin and, indeed, for many other parts of the developing world. Second, the relationships holds within the), temporal (monthly) and areal (basin areas) extent of the study. Despite the small basin (32km²) used in validation, the procedure is best used on basins within the size limits of the historic observations – 53 to 1746 km². The specific coefficients employed in the model will change with geographic as might the shape of the relationship (linear to log/log), however the overall approach is still valid.

This approach requires reasonable spatial homogeneity in a) long-run monthly rainfalls, and b) hydrologic processes converting precipitation to streamflow. Empirical evidence suggests that both are reasonable assumptions, however that validity may change should additional data be available (for example small basins high in the cordillera, or ones that are almost entirely urbanized). The approach does permit modifications to include measures of precipitation and land use in a multiple regression approach (Rasmusson and Perry, 2000).

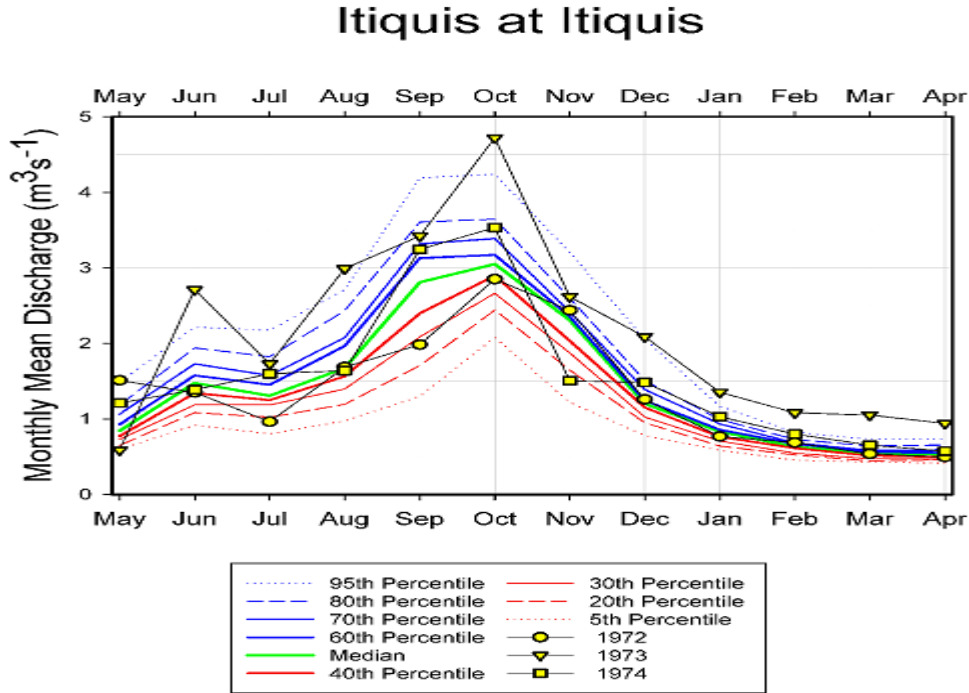


Figure 9. Predicted median and selected common and rare percentiles (lines) compared to the observed data, water years 1972-74 (lines and symbols).

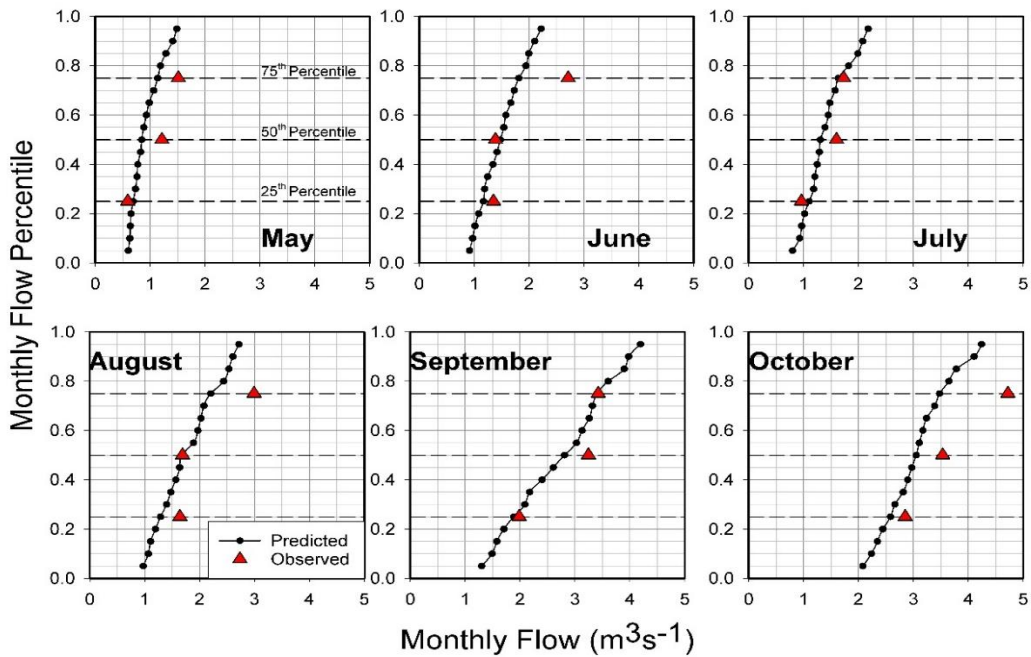


Figure 10. Predicted values of monthly flow percentiles on the Itiquis (black) and estimated percentiles from limited historic records (red), for six months of the rainy season.

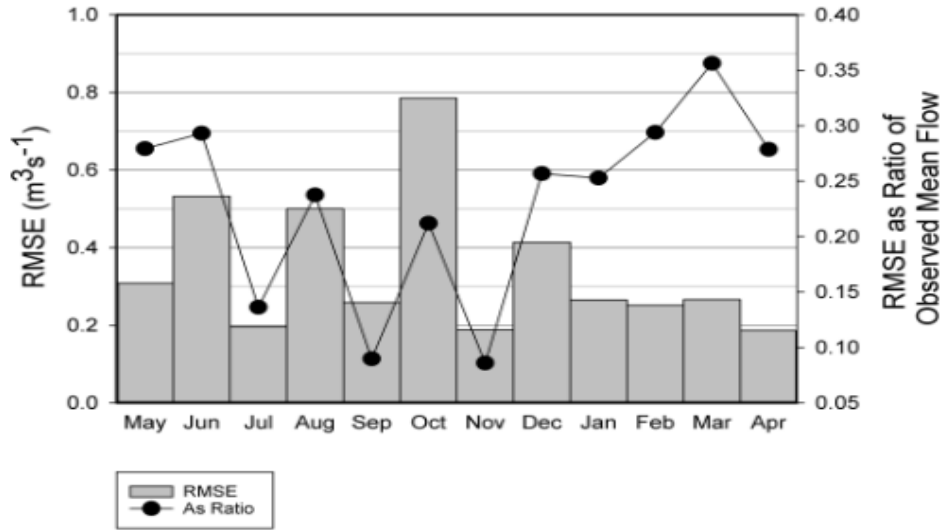


Figure 11. Monthly values of the root mean square error computed between observed and predicted flow percentiles on the Itiquis and as a ratio of the observed mean flow.

Table 2. Identification of monthly flows on the Itiquis and precipitation at San José as laying in the predicted (Itiquis) and observed (San José) terciles of historic observations. Wettest third in blue, middle third white and driest third orange. Years are identified as warm or cold phases of ENSO and months in terms of the regional hydrologic regime (B = Pre-Veranillos rains; V= Veranillos; A = Post-Veranillos rains; R = hydrograph recession).

Year	Month	Terciles Flow Itiquis	Terciles Rain San Jose	COAPS ENSO Phase	Hydroclimatology Recession
1972	May	Blue	White	Warm	B
	Jun	White	White	Warm	B
	Jul	White	Orange	Warm	V
	Aug	White	White	Warm	B
	Sep	White	Orange	Warm	A
	Oct	White	White	Warm	A
	Nov	White	Orange	Warm	R
	Dec	White	Blue	Warm	R
	Jan	White	Orange	Warm	R
	Feb	White	White	Warm	R
	Mar	White	Blue	Warm	R
	Apr	White	White	Warm	R
1973	May	Orange	Blue	Cold	B
	Jun	White	White	Cold	B
	Jul	Blue	White	Cold	V
	Aug	Blue	White	Cold	B
	Sep	Blue	White	Cold	A
	Oct	Blue	White	Cold	A
	Nov	Blue	Orange	Cold	R
	Dec	Blue	White	Cold	R
	Jan	Blue	White	Cold	R
	Feb	Blue	White	Cold	R
	Mar	Blue	Orange	Cold	R
	Apr	Blue	White	Cold	R
1974	May	Blue	White	Cold	B
	Jun	Blue	White	Cold	B
	Jul	Blue	White	Cold	V
	Aug	Blue	White	Cold	V
	Sep	Blue	White	Cold	A
	Oct	Blue	White	Cold	A
	Nov	Blue	Orange	Cold	R
	Dec	Blue	White	Cold	R
	Jan	Blue	White	Cold	R
	Feb	Blue	White	Cold	R
	Mar	Blue	Orange	Cold	R
	Apr	Blue	White	Cold	R

Another reason for the high variability of observed flows and precipitation within the validation period may be the phases of the El Niño – Southern Oscillation (ENSO). On the Pacific slope of Costa Rica, warm phases of ENSO are generally accompanied by drought and cold phases by excess precipitation (Poveda et al., 2006). Table 2 identifies phases of ENSO prevailing during the validation period based on the *a priori* classification of the Center for Ocean-Atmosphere Prediction Studies (<https://coaps.fsu.edu/jma>). Water year 1972 corresponds to warm phase (the fourth strongest “El Niño” 1950-2017) and 1973 and 1974 to the strongest cold phase (La Niña), contributing greatly to the unusually high variability within of these observations.

Specific inconsistencies between flow and rainfall terciles require explanation in terms of how flows are maintained during extended dry periods. Water stored as soil water and groundwater during the rainy season drains during the dry season, analogous to water draining out a spigot at the base of a rain barrel – flow recession (R). The rate of outflow (dry season streamflow) is a function of the quantity of water stored in the barrel (soil and groundwater). Flows follow an exponential decline, or hydrograph recession, (see Figures 3 and 9) during December-April. Dry season precipitation contributions (Figure 2) to streamflow are minimal, thus monthly totals that might be classified as unusually high or low have minimal impact of contemporary streamflow, which more correctly reflects the quantity of water stored at the end of the preceding rainy season. This explains the lack of correspondence between terciles of monthly precipitation and streamflow in the dry season months (R). A corollary occurs at the beginning of the rainy season (B), particularly May, when the study area exhibits varying degrees of dryness, depending upon the depth of the preceding recession. The drier the conditions (lower soil and groundwater storage), the less impact the rains of May have on contemporary flows. No correspondence exists between tercile classes of May of 1972 and 1973, as streamflow is reflecting prior rainy seasons, not contemporary monthly precipitation. However, May 1974 returns values of both variables in the upper tercile, as 1973 was also a cold phase year with excessive rains and the dry season hydrograph remained higher than usual. The high rains of May 1974 (cold phase) fell on a comparatively wet landscape permitting monthly flow also to be high. All three validation years return November (transition between the rainy and dry seasons) rainfalls in the lower tercile, regardless of the phase of ENSO. November rains can be comparatively large (Figure 3), but any

interannual variability is of lesser consequence than the accumulated storage at the end of the October.

Conclusions

Despite the comparative paucity of available data and its extreme temporal variability at both the intra- (seasonal) and inter-annual (year to year) scales, the model derived to predict mean monthly flows and their fluctuations (percentiles) performs well and yields results that are physically meaningful in terms of local (wet/dry seasons) and global (ENSO) hydro-climatological phenomena. Although the lack of validation data undermines thorough statistical testing, the research represents a good starting point for predicting monthly flows at unmonitored locations of interest. The model and approach have the capability to be used in other regions/countries wherever an assumption of regional homogeneity in hydro-climatology is reasonable.

Water is an important resource for all nations, especially fast-growing, developing ones which desperately need knowledge of its spatial and temporal availability to satisfy socio-economic needs of their populaces. However, lack of finance, instrumentation, and technology often prevent even a rudimentary knowledge of the resource or development of an understanding of hydro-climatology necessary to answer the basic questions of how much water they have and its likely temporal and spatial variability. This research outlines a regional approach that maximizes the utility of limited data, successfully applied in the most densely populated region of Costa Rica. The approach might reasonably be employed in other similar locations where basic information about available water resources is needed.

References

- Ballestero, M. (2003). Tárcoles River Basin Costa Rica. 1-120. Retrieved July 11, 2017, from http://siteresources.worldbank.org/INTSAREGTOPWATRES/Resources/CostaRica_Tarcoles_Background_FINAL.pdf.
- Censo de Población (2011). Retrieved on October 16, 2018, from <http://www.inec.go.cr/censos/censos-2011>.
- Cunnane, C. (1978). Unbiased plotting positions—a review. *Journal of Hydrology*, 37(3-4), 205-222.

- Glaves, R.M. & Waylen, P.R. (1997) Regional flood frequency analysis in Southern Ontario using L-Moments. *Canadian Geographer*, 41, 178-198.
- Gleick, P. (ed) (1993) *Water in Crisis: A Guide to the World's Freshwater Resources*. Oxford University Press, Oxford, UK 473pp
- Gupta, V. K., Mesa, O.J. & Dawdy, D.R. (1994). Multiscaling theory of flood peaks: Regional quantile analysis. *Water Resources Research* 30.12, 3405-3421.
- Leopold, L. B., Wolman, M. G. & Miller, J. P. (2012). *Fluvial Processes in Geomorphology*. Courier Corporation. North Chelmsford, MA. 522pp.
- Mimikou, M. (1984). Regional relationships between basin size and runoff characteristics. *Hydrological Sciences Journal*, 29(1), 63-73.
- Poveda, G.J., Waylen, P.R. & Pulwarty, R. (2006). Annual and interannual variability of the present climate in northern South America and southern Mesoamerica. *Paleogeography, Paleoclimatology, Paleoecology*, 234(1), 3-27.
- Rasmusson, P.P. & Perry, C.A. (2000). Estimation of peak streamflows on unregulated rural streams in Kansas. United States Geological Survey Water Resources Investigation Report 00-4079, Lawrence, Ks. 33pp.
- Roberto Jovel, J. & Ahlgren, L. E. (1972). The water resources of the Río Grande de Tárcoles basin, Costa Rica. *Hydrological Sciences Bulletin*, 17(4), 405-418.
- Roberts, L. (Ed.). (1998). *World Resources: A Guide to the Global Environment*. Oxford University Press, Oxford, UK. 369pp.
- Waylen, P.R., Caviedes, C.N. & Quesada, M.E. (1996). Interannual Variability of Monthly Precipitation in Costa Rica. *Journal of Climate*, 9, 2606-2613.