

STATISTICAL LOW FLOW ESTIMATION USING GIS ANALYSIS IN
HUMID MONTANE REGIONS IN PUERTO RICO

ANDRES R. GARCIA-MARTINÓ, FREDERICK N. SCATENA, GLENN S. WARNER,
AND DANIEL L. CIVCO

Made in United States of America
Reprinted from JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
Vol. 32, No. 6, December 1996
© 1996 by the American Water Resources Association

STATISTICAL LOW FLOW ESTIMATION USING GIS ANALYSIS IN HUMID MONTANE REGIONS IN PUERTO RICO¹

Andres R. Garcia-Martinó, Frederick N. Scatena, Glenn S. Warner, and Daniel L. Civco²

ABSTRACT: Statistical analysis of watershed parameters derived using a Geographical Information system (GIS) was done to develop equations for estimating the 7d-10yr, 30d-10yr, and 7d-2yr low flow for watersheds in humid montane regions of Puerto Rico. Digital elevation models and land use, geology, soils, and stream network coverages were used to evaluate 21 geomorphic, 10 stream channel, 9 relief, 7 geology, 4 climate, and 2 soil parameters for each watershed. To assess which parameters should be used for further investigation, a correlation analysis was used to determine the independence and collinearity among these parameters and their relationship with low flows. Multiple regression analyses using the selected parameters were then performed to develop the statistical models of low flows. The final models were selected in the basis of the Mallows Cp statistic, the adjusted R², the Press statistic, the degree of collinearity, and an analysis of the residuals. In the final models, drainage density, the ratio of length of tributaries to the length of the main channel, the percent of drainage area with northeast aspect, and the average weighted slope of the drainage were the most significant parameters. The final models had adjusted standard errors of 58.7 percent, 59.2 percent, and 48.6 percent for the 7d-10yr, 30d-10yr, and 7d-2yr low flows respectively. For comparison, the best model based on watershed parameters that can be easily measured without a GIS had an adjusted standard error of 82.8 percent.

(KEY TERMS: modeling/statistics; low stream flows; surface water hydrology; forest hydrology; geographic information systems; watershed characterization.)

INTRODUCTION

The ability to estimate streamflow in ungaged watersheds is a basic need in hydrological sciences (Riggs, 1973). Until the recent advent of Geographic Information Systems (GIS), this analysis has been limited to watershed attributes measured from topo-

graphic maps (Riggs, 1968). However, with the advent of GIS, the number and complexity of watershed attributes that can be used to estimate low flows has dramatically increased. The objective of this research was to combine GIS technology with multiple regression analysis to develop statistical models that estimate low stream flow in ungaged watersheds in the Caribbean National Forest of Puerto Rico. To accomplish this objective, a methodology for selecting GIS-derived parameters to be used in the multiple regression analysis had to be developed.

Many studies have predicted low stream flows using multiple regression analyses of watershed characteristics (Riggs, 1973; Holmstrom, 1980; Cervione *et al.*, 1982; Hammett, 1985; Hayes, 1991; Vogel and Kroll, 1992; Dingman and Lawlor, 1995). However, none of these studies have used a GIS to develop the parameters used in the analysis. The most commonly used low flow parameter is the 7-day 10-year (7d-10yr) flow, or the lowest average flow in 7 consecutive days in a 10-year interval (Task Report, 1980). Most equations use drainage area to estimate flows, and basin geology and morphology, climate, evapotranspiration, and land use are generally considered critical parameters in determining the magnitude of low flows (Singh and Stall, 1974; Task Committee, 1980). However, the relative importance and statistical reliability of these factors varies considerably between studies and regions.

In 47 states where physiographic characteristics were used to estimate low flow events, the standard error of the estimate (SE) was over 100 percent (Riggs, 1973). In a Connecticut study, the standard

¹Paper No. 95145 of the *Water Resources Bulletin*. Discussions are open until June 1, 1997.

²Respectively, Graduate Student and Research Hydrologist, U.S. Department of Agriculture, Forest Service, International Institute of Tropical Forestry, Call Box 25000, Rio Piedras, Puerto Rico 00928-2500; and Assistant Professor and Extension Water Resources Specialist and Director, Laboratory for Earth Resources Information Systems, Natural Resources Management and Engineering, University of Connecticut, U-87, 1376 Storrs Road, Storrs, Connecticut 06269-4087.

error was reduced to 68 percent when geologic parameters were included (Cervione *et al.*, 1982). In 12 West Virginian watersheds, watershed perimeter, main channel length, a basin form factor, a temperature index, and a rainfall index were the most significant of 18 watershed parameters tested (Chang and Boyer, 1977). When combined, these five factors were able to estimate low flows with a standard error of 35 percent. The percentage of the basin underlain by stratified drift, mean watershed elevation, and mean annual rainfall were the most significant parameters for 53 watersheds in New York (Barnes, 1986). The percent of the basin underlain by sand and gravel was a significant parameter for the prediction of low flows in New Hampshire watersheds (Dingman and Lawlor, 1995). In general, relatively high magnitude low flows have been associated with the presence of sand and gravel (Schneider, 1965; Barnes, 1986; Bouwer, 1978). However, it has also been noted that the influence of geology on the magnitude of low flows is dependent on the presence of channels that intersect and receive base flow contributions from specific geologic units (Farvolden, 1963; Daniels and Hammer, 1992).

In eastern Puerto Rico, the Log-Pearson Type III distribution was used to estimate the magnitude of low flows for 12 continuous and 81 partial stations (Santiago-Rivera, 1992). The 7d-10yr flows for 54 ungaged streams in Puerto Rico were also estimated using relationships between drainage area and discharge (Colón-Dieppa and Quiñones, 1985). A multivariate analysis using drainage area, mean annual rainfall, and mean elevation of 29 watersheds scattered throughout the island was done to predict the flow that is exceeded 90 percent of the time (Morris, 1984). Only the first two of these parameters were significant and together explained 85 percent of the variation in low flows.

DATA AND METHODS

The Caribbean National Forest, also known as the Luquillo Experimental Forest, is the only tropical rain forest in the U.S. National Forest System. This is a well described and studied forest (Brown *et al.*, 1983; Lugo and Lowe, 1995) that has four forest types (Wadsworth, 1951), the highest rainfall in Puerto Rico, and strong relationships between elevation and rainfall (Garcia-Martino, 1996; Garcia-Martino *et al.*, 1996). The procedures used to develop the statistical models of low flow for this area can be divided into six basic steps (Figure 1): development of GIS data base, watershed selection, flow frequency analysis, watersheds characterization, parameter selection, and multiple regression analysis.

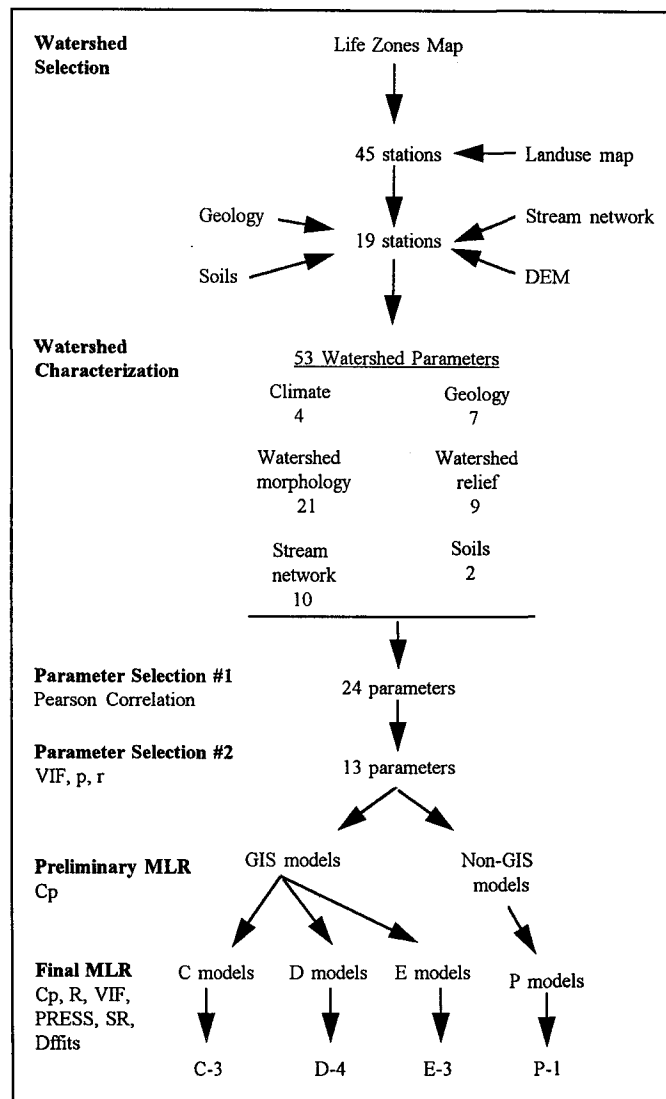


Figure 1. Flow Chart Showing the Procedures Used for the Watershed Selection, Characterization, and Statistical Analysis.

GIS Data Base

A digital elevation model (DEM) and vector-based GIS layers of land use, soils, geology, hydrography, and adjacent areas were used in this analysis. All GIS coverages were obtained as 7.5' x 7.5' quadrangles and geo-referenced to the Universal Transverse Mercator (UTM) coordinate system. Geologic and soil coverages were at a 1:20,000 scale and were from USDA Soil Surveys and U.S. Geological Survey maps that were published in the 1970s.

Land use layers were produced in 1977 by the Puerto Rico Department of Transportation using first

degree units based on Anderson *et al.* (1976). Because the original layers were subdivided into many specific categories (e.g., crop type, housing density), they were aggregated into three categories for this analysis: forest, agriculture, and urban. The drainage network for the study streams was obtained from the USGS as Digital Line Graphs. This network is virtually the same as the stream network shown in the 1:20,000 USGS topographic maps. The DEMs are all in UTM coordinates, and two different formats were used. DEMs with a 30 m by 30 m resolution (USGS, 1987) were obtained for the eastern region of the island, specifically for El Yunque, Fajardo, Humacao, Naguabo, Juncos, and Fajardo quadrangles. For the rest of the island, a DEM produced by the Defense Mapping Agency was used. These DEMs had a resolution of 3 x 3 arc seconds, which in Puerto Rico is equivalent to an approximate 90 m x 90 m resolution. The DEMs were only used in their raster format. The vector format coverages were converted to raster format when needed.

Watershed Selection

Because the primary goal of this study was to develop equations for estimating low flows in the CNF, only gaged streams that were either within or adjacent to the forest or drained forested catchments with similar rainfall and temperature were considered in the analysis. The Holdridge Life Zones Map of Puerto Rico (Ewel and Whitmore, 1973) was used to identify watersheds that had similar annual rainfall and temperature to that of the CNF (e.g., Moist, Wet, and Rain forest life zones). For the entire island, 45 continuous or partial gaging stations were identified. When the watershed boundaries of these 45 watersheds were overlain with the landuse maps, 19 watersheds were selected for further analysis (Figure 2 and Table 1). All of these watersheds had either continuous or partial stream discharge records, were forested, and drained some combination of Moist, Wet, and Rain forest life zones.

Flow Frequency Analysis

The HEC-FFA frequency analysis program (HEC, 1992) and publications by Riggs (1967, 1972, and 1973) and the Water Resources Council (1977) were used as guides in conducting the frequency analysis of continuous and partial stream discharge stations (Table 1). Frequency analysis for the continuous stations was done using the Weibull plotting position and the log-Pearson Type probability distribution with a

skew coefficient of 0.55. This distribution fit the data well and is considered more representative of low flow conditions than the Pearson Type V distributions (Matalas, 1963). The skew coefficient is a weighted average of the skew coefficients of distribution of continuous stations with at least 15 years of record (Water Resources Council, 1977).

The data for most of the partial stations consist of six or seven low flow measurements that were taken over a four-year period. For some of these stations, the low flow values were obtained from Santiago-Rivera (1992). For other partial stations and three continuous stations that did not have long records, low flows were correlated with concurrent discharges from nearby watersheds with similar hydrologic characteristics, as recommended by Riggs (1972). The objective of this correlation was to estimate low flow events with a specific recurrence interval, not to create or improve a frequency distribution. All of the correlations between stations were significant ($\alpha = 0.05$, Garcia-Martinó, 1996).

Watershed Characterization

Fifty-three parameters were used to describe each watershed (Table 2). Most of these parameters have been used elsewhere (Horton, 1945; Strahler, 1954; Wisler and Brater, 1963; Chang and Boyer, 1977; Dunne and Leopold, 1978; and Eash, 1994) and can be divided into six basic groups: climatic, stream network morphology, watershed morphology, watershed relief, geology, and soils. Additional details describing the watershed characterization are given elsewhere (Civco *et al.*, 1995, Garcia-Martinó, 1996).

Although the watersheds were selected to have similar climates and life zones, mean annual rainfall and three measures of rainless days per year were used to describe the climate of each watershed (Table 2). When characterizing channel and watershed morphology, special emphasis was placed on morphologic features that maybe related to water input (e.g., aspect) or water storage (e.g., slope, drainage density, etc.). Both geologic and soil parameters were selected on the basis of their potential to influence storm runoff and ground water dynamics. An initial inspection of soil parameters derived from the soil survey reports (Boccheciamp, 1977) indicated that many variables (water holding capacity, liquid limit, texture, drainage and hydrologic class, and profile and bedrock depth) did not vary significantly between the study watersheds. Soil parameters that did vary between watersheds and that were used in the analysis were the plasticity index and permeability (Table 2).

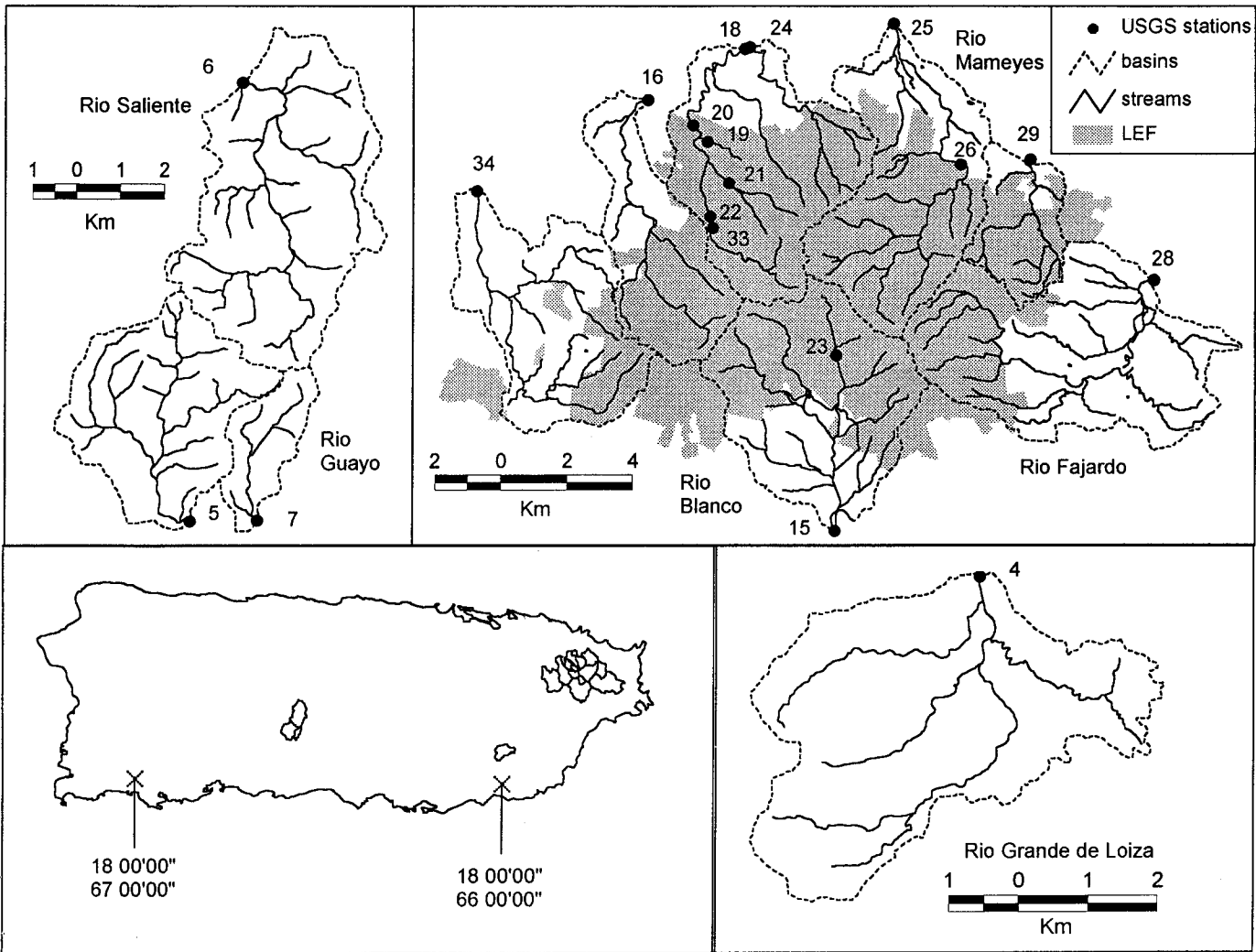


Figure 2. Nineteen Watersheds Selected for Statistical Analysis, of Which 15 are Located in the LEF, Three in the Central Mountain Region, and One in the Southeast Region of Puerto Rico.

TABLE 1. Low Stream Flows in Cubic Feet Per Second (cfs) in Rain, Wet, and Moist Forest Life Zones of Puerto Rico, where d = day, yr = year. [See text and Garcia-Martínó (1996) for details.]

Station	USGS ID	Area (km ²)	7d-10yr (cfs)	7d-2yr (cfs)	30d-10yr (cfs)
5	1124	15.4	0.82	1.49	1.14
6	25155	23.9	2.82	4.18	3.47
7	1127	4.27	0.21	0.46	0.33
15	760	31.8	7.23	14.00	12.00
20	6354	13.7	3.34	4.90	5.30
22	633	5.64	2.53	3.09	3.16
24	6385	9.25	1.51	2.37	2.53
25	657	30.9	8.30	9.90	11.20
33	6325	5.24	3.29	3.76	3.69
4	5090	15.7	4.14	5.63	4.96
16	642	19.1	3.39	5.37	4.90
18	638	22.4	5.43	7.95	8.31
19	635	0.19	0.01	0.03	0.03
21	6344	2.65	0.03	0.21	0.30
23	750	3.33	2.86	3.82	3.67
26	655	17.5	9.50	12.30	13.10
28	710	38.5	3.92	7.59	8.19
29	670	10.0	1.58	2.05	2.00
34	618	26.6	2.92	4.38	3.32

TABLE 2. Watershed Parameters Used to Correlate with Low Stream Flows from 19 Forested Watersheds in Rain, Wet, and Moist Forest Life Zones of Puerto Rico.

Climate	
MMYR	= Mean annual rainfall (mm/yr).
10YRCDNR	= The number of consecutive days with no rain with a recurrence interval of 10 years.
M#DNR	= Average number of days per year with no rain MMCDNR = maximum consecutive days with no rain.
Geology	
UNCON & QA	= Area and percent of the watershed classified as unconsolidated material
LSEU	= Length of perennial streams that are in direct contact with areas classified as unconsolidated material.
INTRU	= Percent of the total area of the watershed classified as intrusive igneous rock with low water bearing capacities.
STRAT	= Percent of the total area of the watershed classified as alluvial deposits and sedimentary formations (sandstone).
Soils	
PLASTI	= Weighted average plasticity index.
PERM	= A weighted average permeability (in/h).
Stream Network	
PERC	= Total length of perennial channels (m).
CHAN	= Total length of the main channel (m).
TRIB	= Total length of the tributaries to the main channel (m).
DD	= Total length of perennial channels per unit area (m/m ²).
SD	= Number of perennial channels per unit area (#/m ²).
TRIPER	= A ratio of the total length of tributaries to the total length of the perennial channels (m/m).
CHATRI	= A ratio of total length of tributaries to the total length of the main channel (m/m).
CHANSLP	= Slope of the main channel between two points located at 15 percent and 85 percent of the total length of the channel, starting from the outlet and divided by the horizontal distance between the two points (°).
CHSLP	= (CHANSLP/CHAN) * 100.
SINUO	= Total length of the main channel divided by the length of the watershed (m/m).
Watershed Morphology	
AREA	= Watershed area (ha).
SLP	= Weighted average slope (°).
PERIM	= Perimeter following the watershed boundary (m).
COMP	= Ratio of the perimeter of the watershed to the circumference of a circle whose area is equal to that of the watershed.
FORM	= Area divided by the square of the watershed length (m ² /m ²).
ELONG	= Diameter of a circle with the same area as the area of the watershed divided by the watershed length.
AREA4 & A4 AREA7 & A7 AREA10 & A10	= Area in the watershed and percent of the area in the watershed with slopes less than 4°, 7°, and 10°, respectively.
MSLP100 ASLP100 MSLP200 ASLP200	= Maximum and average slope inside a 100 meter buffer surrounding the perennial streams (°).
ASPECT	= Weighted average aspect.
AREA90	= Total area of the watersheds with aspect between 0 and 90°.
PA90	= Percent of the total area of the watershed with aspects between 0 and 90°.
ASP100 & ASP100	= Weighted average aspect inside a buffer of 100 and 200 meters, respectively, around the perennial channels.

TABLE 2. Watershed Parameters Used to Correlate with Low Stream Flows from 19 Forested Watersheds in Rain, Wet, and Moist Forest Life Zones of Puerto Rico (cont'd.).

Watershed Relief	
The HYP parameters were obtained from a hypsometric graph.	
HYP1	= Percent change between the .25 and .75 points in the Y axis divided by the watershed AREA.
HYP2	= Percent change between the .25 and .75 points in the Y axis.
HYP3	= Percent change between the .25 and the .50 points divided by the percent change between the .50 and .75 points.
RELIEF	= Maximum elevation minus minimum elevation divided by the length of the watershed.
ELRANGE	= Maximum minus minimum elevation in the watershed.
RERA	= Relief * elrange.
RELARELI	= Maximum elevation minus minimum elevation divided by the length of the watershed.
SLORAT	= CHANSLP / SLP.
EL	= Weighted average elevation.

Parameter Selection

First Selection. The first selection from the original 53 parameters was based on their correlation with the dependent variables and their correlation with other independent variables. The second parameter selection excluded independent parameters on the basis of logic, redundancy, and collinearity. Normality was initially checked for all the watershed parameters using the Shapiro-Wilk test (SAS, 1985). The decision to transform a variable was based on the normality test and corresponding correlations with the dependent variable.

During this first selection, the elimination or inclusion of each parameter was based on its Pearson correlation with the flow frequencies and other watershed parameters. Only unique parameters that had relatively high correlations with the stream flow and lower correlations with other independent variables were selected. Of the original 53 parameters, 24 were selected for further analysis. These were QA, DD, EL, MMYR, AREA, SLP, ASLP100, ASPECT, ASP200, PLASTICI, PERM, PA90, CHATRI, CHSLP, ELONG, SINUO, SLORAT, SAND, INTRU, SRAT, HYP1, LA10, LLSEU, and RELARELI (Table 2).

Second Selection. In order to select the final parameters for the regression analysis, six groups of parameters were formed. Each of these groups had a unique combination of 12 of the 24 parameters that were previously selected. Each of these groups was treated as a model in which the 12 parameters were used as the independent variables. The purpose of this grouping was to evaluate inter-relationships between variables and how combinations of values affect the prediction of the dependent variable. Each group was analyzed using four statistics; the VIF

collinearity statistic, the significance of the coefficients of each parameter, the Pearson correlation with the dependent variable, and the Pearson correlation with other independent variables.

Collinearity, or the relationship of one independent variable to one or more independent variables, was measured using the VIF statistic (Helsel and Hirsch, 1992).

$$VIF = 1 / (1 - R_j^2) \quad (1)$$

where R_j^2 is the R^2 of the j parameter with the other parameters in the model. A VIF value of 10 or lower is typically considered satisfactory (Helsel and Hirsch, 1992). As the number of watershed parameters in the model increases, the probability of having high collinearity and redundancy also increases. High collinearity in a model can result in coefficients with unexpected sign and unrealistically large or unstable regression coefficients (Cohen and Cohen, 1983).

The significance of the coefficients of each of the 12 parameters in each group was also analyzed to identify parameters that were constantly significant through all or most of the groups. Parameters that were significant at the 0.15 level were considered potentially significant while parameters at the 0.05 level were considered as highly significant. The purpose of the comparison was to identify which parameters cover the highest degree of variation of the dependent variable and at the same time had the highest independence from other independent variables. The parameters that were finally selected by comparing all these statistics were QA, DD, AREA, SLP, PLASTICI, PERM, PA90, CHATRI, ELONG, SINUO, LLSEU, RELARELI, and CHSLP (Table 2).

Multiple Regression Analysis

Using the 13 parameters that remained after the second selection process, a multiple regression analysis was performed to predict the 7d-10yr, 30-d10yr, or 7d-2yr low flow events. Two types of models were developed: GIS models that were based on the 13 watershed parameters and non-GIS models that only included parameters that could be easily measured without the aid of a GIS. The parameters used to develop the non-GIS models were CHAN, CHANSLP, EL, MMYR, AREA, SINUO, FORM, ELRANGE, DD, aNd PER (Table 2).

The selection of the final models was based on the Mallow's Cp statistic, the PRESS statistic, the maximum VIF, the adjusted R² (R_a²), and the distribution of residuals. The Mallow Cp statistic is a measure of the additional variance in y that is covered after a new parameter is added to the model, but at the same time minimizing the number of coefficients and standard error:

$$Cp = p + (n-p) * (s_p^2 - \sigma^2) / \sigma^2 \tag{2}$$

where p is the number of coefficients in the model including the intercept, n is the sample size, s_p² is the mean squared error, and σ² is the minimum mean squared error from all 2^k possible models (k = number of variables) (Helsel and Hirsch, 1992). The lower the Cp, the better the model. While the Cp statistic should not be used as a final decision tool, it does indicate which models should be considered for further analysis.

The distribution of residuals was analyzed using the H, standardized residuals, and Dffit statistics. These statistics measure the leverage in the x direction, leverage in the y direction, and influence respectively. In general, it is recommended that there should not be more than five standardized residuals greater than two for each 100 observations (Helsel and Hirsch, 1992). The critical values for the H and Dffits statistics used were

$$H_{crit} = 3p / n \tag{3}$$

$$|Dffit_{crit}| = 2 * (p/n)^{1/2} \tag{4}$$

where p is the number of coefficients including the intercept and n is the sample size.

To determine the most complete appropriate model, an exclusion process was used. The first model selection was based on the Cp statistic using the PROC RSQUARE procedure in SAS. The objective of this step was to determine if any of the 19 stations constantly produced outliers. Because Station 21 constantly produced high leverage and influence, it was

removed from the subsequent analysis. The models with the lowest Cp values were then selected for further analysis, and the significance and collinearity of each of the parameters in each of the selected models was tested. Models with coefficients not significant at the 0.15 level or with a VIF ≥ 10 were excluded. Table 3 shows the models selected based on previous criteria. The last step in model selection was to compare the PRESS statistic, maximum VIF, R_a², Cp, H, Dffit, and standardized residuals of the various models.

To compare the robustness of the final models, an adjusted standard error of the estimate (SE_a) was used (Ezekiel and Fox, 1959):

$$SE_a = [N / (N - M) * SS_e]^{0.5} \tag{5}$$

where N is the sample size, M is the number of coefficients in the model including the intercept, and SS_e is the sum of squares of the error. In this analysis, the error in the model is the product of chance variation and variation from unaccounted differences in watershed characteristics. Because the difference between the observed and predicted low flows can vary with discharge, the relative error of prediction (EP) was also calculated where

$$EP = (Predicted - Observed) / Observed * 100 \tag{6}$$

RESULTS AND DISCUSSION

Relationship Between Dependent Variables

Significant relationships (α = 0.01) existed among the 7d-10yr flow and the other low flows for the 19 study watersheds, as the following indicates:

$$7d-2yr = 1.4018X + 0.2106 \quad R^2 = 0.91 \tag{7}$$

$$14d-10yr = 1.1252X - 5.9378 \quad R^2 = 0.99 \tag{8}$$

$$14d-2yr = 1.6338X + 0.1886 \quad R^2 = 0.91 \tag{9}$$

$$30d-10yr = 1.4468X - 0.0392 \quad R^2 = 0.95 \tag{10}$$

$$30d-2yr = 2.2817X - 0.0136 \quad R^2 = 0.85 \tag{11}$$

where X is the 7d-10yr in cubic feet per second (cfs). For relations that predict a two-year recurrence interval flow (Equations 7, 9, and 11), the confidence of the estimates decreases as 7d-10yr increases. However, for events with a 10-year recurrence, the confidence in the predictions was constant through the range of flows.

TABLE 3. Statistical Comparison of Models Selected to Predict Low Flows in 19 Forested Watersheds in Rain, Wet, and Moist Forest Life Zones of Puerto Rico (see Table 2 for definition of variables).

Model	Parameters	Cp ¹	R _a ²	Max. VIF ³	PRESS ⁴	Above Critical ⁵		
						H ⁶	Dffits ⁷	SR ⁸
7day-10yr								
C-1	DD, PA90, CHATRI	0.07	0.95	1.3	0.57	1	2	2
C-1B	PLASTICI, CHATRI, CHSLP	3.36	0.93	2.6	0.10	1	2 ⁹	0
C-2	RELARELI, CHATRI, DD, PA90	1.10	0.96	3.9	0.50	0	2 ⁹	2
C-3	DD, SLP, PA0, CHATRI	0.14	0.96	1.6	0.46	1	2	1
30day-10yr								
D-1	DD, PA90, CHATRI	3.90	0.94	1.3	0.63	1	2 ⁹	0
D-2	PLASTICI, PA90, CHATRI, RELARELI	0.41	0.96	2.5	0.49	0	3 ⁹	2
D-3	DD, PA90, CHATRI, RELARELI	1.97	0.95	3.9	0.59	0	3 ⁹	2
D-4	DD, SLP, PA90, CHATRI	2.33	0.95	1.6	0.45	0	2	1
D-5	PLASTICI, PA90, CHATRI, RELARELI, CHSLP	0.45	0.97	5.8	0.44	0	3 ⁹	1
7day-2yr								
E-1	DD, PA90, CHATRI	3.60	0.96	1.30	0.39	1	2	1
E-2	DD, PA90, CHATRI, RELARELI	1.70	0.97	3.90	0.35	0	3	2
E-3	DD, SLP, PA90, CHATRI	2.10	0.97	1.60	0.29	0	1	2
E-4	DD, PLASTICI, PA90, CHATRI, RELARELI	0.01	0.98	6.70	0.27	0	4	1
E-5	DD, PLASTICI, PA90, CHATRI, SINUO	0.50	0.97	4.60	0.26	0	2	0
E-6	DD, PLASTICI, PA90, LLSEU	0.80	0.98	6.80	0.22	0	2 ⁹	1
7day-10yr								
P-1	CHANS LP, FORM, ELRANGE, DD	4.22	0.92	1.9	0.78	1	2	1

¹CP = Measure of variance.

²R_a = Adjusted coefficient of regression.

³VIF = Diagnostic for measuring multi-collinearity.

⁴PRESS = Measure of outliers in the Y direction.

⁵Observations out of a total of 18 with values above the corresponding critical value.

⁶H = Measure of outliers in the X direction.

⁷Dffits = Measure of influence.

⁸SR = Measure of outliers in the Y direction.

⁹One observation is at least two times above the corresponding critical value.

Correlation Between Dependent and Independent Variables

As has been shown in other many studies (e.g., Morris, 1984; Holmstrom, 1980; Thomas and Benson, 1970), the drainage area of the watershed (AREA) was one of the best predictors of low flow. However, in this study AREA was highly correlated to many other independent variables, the strongest of which were the length of perennial channels, length of tributaries, the length of the main channel (Figure 3), and the ratio of the length of the tributaries to the length of the main channel (Table 4). Soil permeability, channel sinuosity, and the percent area with an unconsolidated substrate (QA) were also positively correlated

with AREA. Relative relief (RELARELI) and channel slope ratio (CHSLP) were negatively correlated with area.

Various measures of the morphology of the watersheds were significantly related to the magnitude of low flows. Moreover, channel slope, relative relief, and basin form had negative correlations with log transformed low flows (Table 4). However, correlations of low flows with the percent of the total watershed area classified as alluvial deposits or sedimentary formations (STRAT) and the percent of watershed area classified as intrusive igneous rock (INTRU) were not significant. In addition, the climatic parameters tested did not have significant correlations with any of the low flow variables. Apparently, there was not

enough geologic or climatic variation between the study watersheds for these factors to be important.

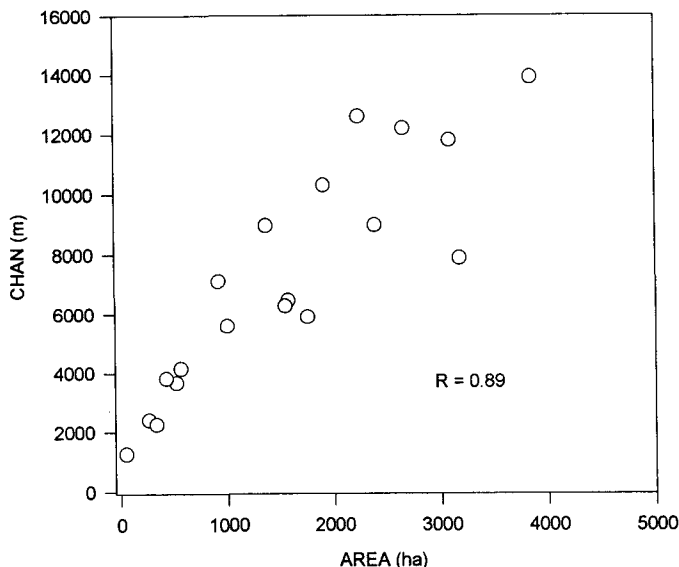


Figure 3. Correlation Between the Watershed Area (AREA) and the Main Channel Length (CHAN) for the 19 Watersheds.

The influence of the prevailing winds and climatic inputs was significantly related to various watershed and flow parameters. Moreover, the percent of the total watershed area that had a northeast aspect (PA90), the aspect facing the prevailing trade winds, was negatively correlated with drainage area, main channel slope, and soil plasticity (Table 2). In turn, higher soil plasticity was correlated with greater drainage densities and lower low flows. Apparently, larger watersheds have more flat or low-lying areas where stream channels have greater sinuosity, unconsolidated deposits are more abundant, and soils have greater permeability. These correlations also indicate that longer and steeper watersheds have less storage capacity than smaller, steeper basins and therefore have lower low flows. Finally, the strength of the correlations with the aspect relative to the tradewinds may also reflect long-term interactions between the prevailing trade winds, light, and moisture inputs and the erosion and morphologic development of the Luquillo landscape.

Comparison of Final Models

All of the GIS 7d-10yr models (C models) had similar residuals (Table 3). However models C-1B and C-2 had Dffit values that were two times above the critical value. Only models C-1 and C-3 show residuals

with no exaggerated leverage or influence. For the 30d-10yr models (D models) the VIF, Cp, adjusted R², and PRESS statistics (Table 3) were most satisfactory for models D-2, D-4, and D-5. However, because only model D-4 had fewer than three Dffits above the critical, it was selected as the most appropriate model to estimate the 30d-10yr low flow. Model D-4 also has the same parameters as model C-1, making it practical to use. Based on the Cp, PRESS, and Adjusted R² statistics for the 7d-2yr flows, (E models) E-3, E-4, E-5, and E-6 had the most acceptable results (Table 3). According to the residual analysis, models E-3, E-5, and E-6 were all highly acceptable. Nevertheless, model E3 was selected as the most appropriate because it had the same parameters as C-3 and D-4.

When all the parameters tested in the non-GIS model were compared in one model, MMYR, EL, CHAN, and AREA had VIFs above 20. When only CHANSLP, EL, SINUO, FORM, ELRANGE, DD and PER were considered, each of their VIFs was below 5. Using these parameters and the lowest CP value, three non-GIS models were selected of which model P-1 was selected. This model had the lowest Cp, the highest R_a², and fewest outliers.

The final GIS and non-GIS models that were selected are as follows:

Model C-3

$$\begin{aligned} \text{Log7d-10yr} = & 1.78 - 1331.61(\text{DD}) - .0316(\text{SLP}) \\ & + 0.0146(\text{PA90}) + 0.3126(\text{CHATRI}) \\ R_a^2 = & 0.96 \end{aligned} \tag{12}$$

Model D-4

$$\begin{aligned} \text{Log30d-10yr} = & 1.60 - 1098.26(\text{DD}) - 0.0301(\text{SLP}) \\ & + 0.0160(\text{PA90}) + 0.2957(\text{CHATRI}) \\ R_a^2 = & 0.95 \end{aligned} \tag{13}$$

Model E-3

$$\begin{aligned} \text{Log7d-2yr} = & 1.70 - 1157.49(\text{DD}) - 0.0248(\text{SLP}) \\ & + 0.0121(\text{PA90}) + 0.3017(\text{CHATRI}) \\ R_a^2 = & 0.97 \end{aligned} \tag{14}$$

Model P-1

$$\begin{aligned} \text{LD710} = & 0.1827 - 0.0824(\text{CHANSLP}) + 2.24(\text{FORM}) \\ & + 0.0012(\text{ELRANGE}) - 882.51(\text{DD}) \\ R_a^2 = & 0.92 \end{aligned} \tag{15}$$

TABLE 4. Pearson's r Correlation Matrix and Associated Probabilities in Parenthesis of Watershed Parameters for 19 Forested Watersheds in Rain, Wet, and Moist Forest Life Zones of Puerto Rico (see Table 2 for definition of variables).

Parameters	log(7d10yr)	qa	dd	area	slp	plastici	perm	pa90	chatri	elong	sinuo	liseu	relareli
qa	0.31 (0.1894)												
dd	-0.70 (0.0008)	0.16 (0.5313)											
area	0.62 (0.0047)	0.66 (0.0031)	-0.11 (0.6391)										
slp	-0.0001 (.9974)	0.04 (0.8885)	0.20 (0.4292)	0.19 (0.43)									
plastici	-0.71 (0.0007)	0.17 (0.4913)	0.83 (0.0001)	-0.19 (0.4389)	0.18 (0.4696)								
perm	0.72 (0.0005)	0.43 (0.0755)	-0.43 (0.0636)	0.61 (0.0053)	-0.24 (0.3500)	-0.47 (0.0423)							
pa90	0.50 (0.0304)	0.19 (0.4519)	-0.44 (0.0581)	0.26 (0.2881)	0.33 (0.1863)	-0.52 (0.0211)	-0.07 (0.7701)						
chatri	0.46 (0.0464)	0.58 (0.0109)	0.13 (0.5867)	0.77 (0.0001)	0.36 (0.1476)	0.16 (0.5187)	0.53 (0.0184)	-0.003 (0.0006)					
elong	0.48 (0.0381)	0.31 (0.2036)	-0.25 (0.3055)	0.34 (0.1558)	-0.14 (0.5900)	0.03 (0.9160)	0.43 (0.0675)	-0.10 (0.6716)	0.62 (0.0045)				
sinuo	0.51 (0.0255)	0.50 (0.0332)	-0.26 (0.2886)	0.60 (0.0070)	-0.14 (0.5700)	0.24 (0.3141)	0.41 (0.0852)	0.30 (0.2132)	0.25 (0.3015)	0.23 (0.3402)			
liseu	0.46 (0.0572)	0.64 (0.0039)	0.02 (0.9211)	0.76 (0.0001)	0.11 (0.6691)	-0.002 (0.9944)	0.52 (0.0223)	0.26 (0.2839)	0.72 (0.0005)	0.32 (0.1798)	0.50 (0.0275)		
relareli	-0.90 (0.0001)	-0.27 (0.2776)	0.72 (0.0005)	-0.59 (0.0082)	0.23 (0.3650)	0.64 (0.0032)	-0.65 (0.0027)	-0.39 (0.1017)	-0.35 (0.1480)	-0.57 (0.0113)	-0.66 (0.0020)	-0.43 (0.0694)	
chslp	-0.94 (0.0001)	-0.30 (0.2217)	0.76 (0.0001)	-0.56 (0.0131)	0.02 (0.9293)	0.66 (0.0021)	-0.72 (0.0005)	-0.46 (0.0464)	-0.38 (0.1123)	-0.49 (0.0318)	-0.58 (0.0090)	-0.51 (0.0251)	0.91 (0.0001)

The SE_a for these models (Table 5) is relatively good considering that the SE of most low flow equations is over 100 percent (Riggs, 1973). In models C-3, D-4, and E-3, the contributions of each variable to the explanation of the variation in Y is very similar (Table 5). Model C-3 was the only model that showed a direct relationship between EP and the magnitude of the low flow (Figure 4). In models C-3 and P-1, the confidence of the prediction decreases as the magnitude of the low flow increases (Figure 5).

CONCLUSIONS

Of the 53 parameters that were derived using the GIS, drainage density, average watershed slope, aspect relative to trade winds, and the ratio of the length of the tributaries to the length of the main channel were identified as the most important factors in estimating low flows from humid sub-tropical forested steplands in Puerto Rico. Drainage density, which explained over 53 percent of the variation in low flows in the GIS models, appears to reflect much of the subsurface and ground water dynamics in the area.

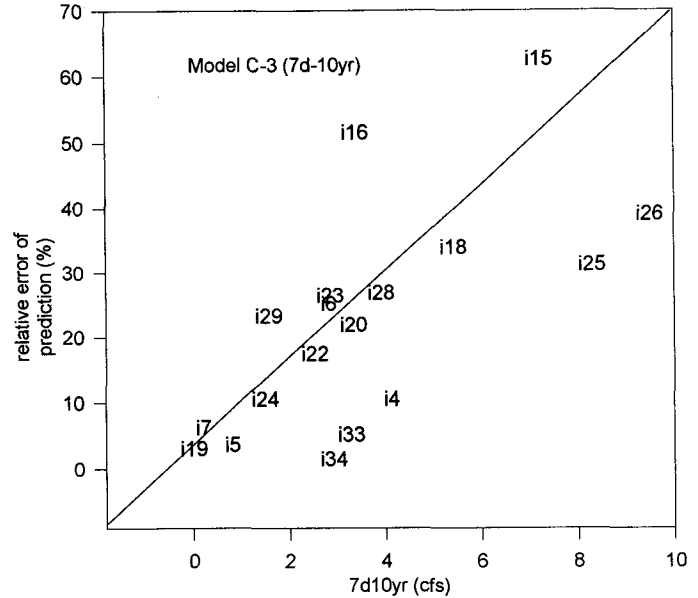


Figure 4. Relationship Between the Relative Error of Prediction and the Observed 7d-10yr Low Flow for the 19 Watersheds.

TABLE 5. Statistical Comparison of Models Selected to Predict the 7day-10yr Low Flows in 19 Forested Watersheds in Rain, Wet, and Moist Forest Life Zones of Puerto Rico.

Parameters	Model C-3 (7day-10yr)			Model D-4 (30day-10yr)			Model E-3 (7day-2yr)		
	R ²	SE*	SE _a **	R ²	SE	SE _a	R ²	SE	SE _a
DD	0.59	45.68	193.81	0.53	44.67	189.51	0.54	43.19	183.25
DD + CHATRI	0.91	21.84	92.68	0.88	23.41	99.32	0.92	18.11	76.85
DD + CHATRI + PA90	0.96	15.71	66.65	0.95	15.59	66.15	0.96	12.83	54.43
DD + CHATRI + PA90 + SLP	0.97	13.84	58.72	0.96	13.94	59.15	0.97	11.47	48.64

Parameters	Model P-1 (7day-10yr)		
	R ²	SE	SE _a
DD	0.59	45.68	193.81
DD + ELRANGE	0.91	21.84	92.68
DD + ELRANGE + FORM	0.96	15.71	66.65
DD + ELRANGE + FORM + CHANSLP	0.97	13.84	58.72

*Standard error.

**Adjusted standard error.

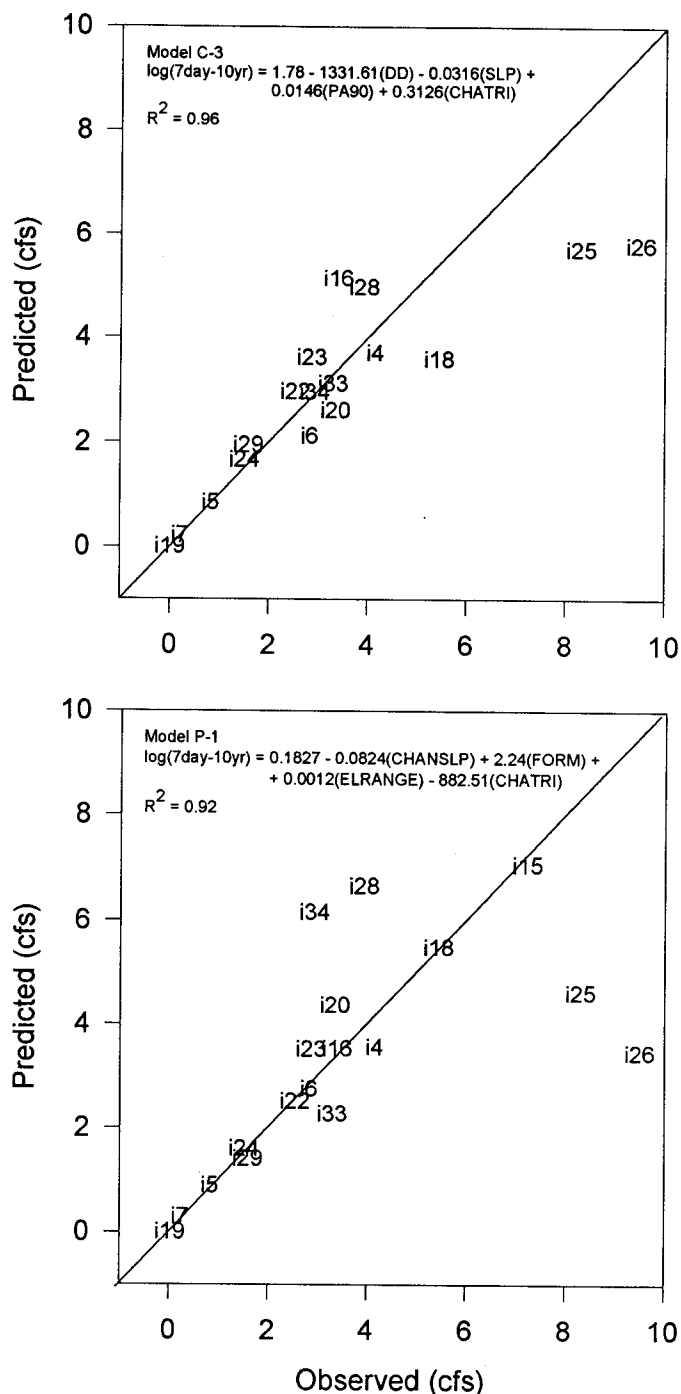


Figure 5. Predicted Versus Observed Values for Models C-3 and P-1.

The combination of the geographic information systems and the complete statistical analysis that was used here resulted in a greater accuracy of low flow prediction than would have been obtained by only using average, non-GIS derived parameters. Moreover, using the GIS improved the Se_a of the equations from 58.7 to 82.8 percent. However, one potential

weakness of the approach used here was that parameters were eliminated on the basis of correlations and collinearity before they were used in a multiple regression analysis. As the use of GIS in hydrologic analysis increases, the number and complexity of parameters used to characterize watershed will also increase. Additional research should be conducted for determining techniques and procedures for selecting relevant parameters and appropriate models.

ACKNOWLEDGMENTS

This research was funded by the USDA Forest Service International Institute of Tropical Forestry as part of the Luquillo Long-Term Ecological Research Project (BSR-8811764) and the University of Connecticut. Published as Scientific Contribution No. 1649 of the Storrs Agricultural Experiment Station. Special thanks to Eda Meléndez from the Terrestrial Ecology Division of the University of Puerto Rico for providing rainfall data and to John Parks from the U.S. Geological Survey and Pedro Ríos from the U.S. Forest Service in Puerto Rico for providing GIS data layers.

LITERATURE CITED

Anderson, J. A., E. E. Hardy, J. T. Roach, and R. E. Witmer, 1976. A Land Use and Land Cover Classification System for Use with Remote Sensing Data. U.S. Geological Survey Professional Paper 954, 28 pp.

Barnes, C. R., 1986. Method for Estimating Low-Flow Statistics for Ungaged Streams in the Lower Hudson River Basin, New York. U.S. Geological Survey Water-Resources Investigation Report 85-4070, 22 pp.

Boccheciamp, R. A., 1977. Soil Survey of the Humacao Area of Eastern Puerto Rico. USDA Soil Conservation Service. 103 pp.

Bouwer, H., 1978. Groundwater Hydrology. McGraw, New York, New York, 480 pp.

Brown, S., A. E. Lugo, S. Silander, and L. Liegel, 1983. Research History and Opportunities in the Luquillo Experimental Forest. Gen. Tech. Rep. SO-44, U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station, New Orleans, Louisiana, 128 pp.

Cervione, M. A., R. L. Melvin, and K. A. Cyr, 1982. A Method for Estimating the 7-day, 10-year Low Flow of Streams in Connecticut. Connecticut Water Resources Bulletin No. 34, 17 pp.

Chang, M. and D. G. Boyer, 1977. Estimates of Low Flows Using Watershed and Climatic Parameters. Water Resources Research 13(6):997-1001.

Civco, D. L., A. R. García, and G. S. Warner, 1995. Key Steps to Effective Watershed Characterization. GIS World 8(11):5 pp.

Cohen, J. and P. Cohen, 1983. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (2nd Edition). Lawrence Erlbaum Assoc., Hillsdale, New Jersey. 545 pp.

Colon-Dieppa E., and V. Quiñones-Aponte, 1985. Estimates of 7-day 10-year Low Flow at Ungaged Streams in Puerto Rico. U.S. Geological Survey Water Resources Investigation Report 84-4089, 2 plates.

Daniels, R. B. and R. D. Hammer, 1992. Soil Geomorphology. John Wiley & Sons, Inc., New York, New York, 236 pp.

Dingman, S. L. and S. C. Lawlor, 1995. Estimating Low Flow Quantiles from Drainage-Basin Characteristics in New Hampshire and Vermont. Water Resources Bulletin 31(2):243-256.

- Dunne, T. and L. B. Leopold, 1978. *Water in Environmental Planning*. W.H.Freeman Co., New York, New York. 818 pp.
- Eash, D. A., 1994. A Geographic Information System Procedure to Quantify Drainage-Basin Characteristics. *Water Resources Bulletin* 30(1):1-17.
- Ewel, J. J. and J. L. Whitmore, 1973. *The Ecological Life Zones of Puerto Rico and the US Virgin Islands*. USDA Forest Service Research Paper ITF-18, Institute of Tropical Forestry, Rio Piedras, Puerto Rico. 71 pp.
- Ezekiel, M. and K. A. Fox, 1959. *Methods of Correlation and Regression Analysis*. John and Wiley, New York, New York, 548 pp.
- Farvolden, R. N., 1963. Geologic Controls on Ground-Water Storage and Base Flow. *Journal of Hydrology* 1:219-249.
- Garcia-Martinó, A. R., 1996. Use of GIS for Low Flow Prediction in Humid Montane Regions in Eastern Puerto Rico. Unpublished M.S. Thesis, University of Connecticut, Storrs, Connecticut.
- Garcia-Martinó, A. R., 1996. Rainfall, Runoff and Elevation Relationships in the Luquillo Mountains of Puerto Rico. *Caribbean Journal of Science* (in press).
- Hammett, K. M., 1985. Low-Flow Frequency Analysis for Streams in West-Central Florida. U.S. Geological Survey Water-Resources Investigations Report 84-4299, 48 pp.
- Hayes, D. C., 1991. Low-Flow Characteristics of Streams in Virginia. U.S. Geological Survey Water Supply Paper 2374, 69 pp.
- HEC, 1992. HEC-FFA: Flood Frequency Analysis User's Guide. U.S. Corps of Engineers CPD-13.
- Helsel D. R. and R. M. Hirsch, 1992. *Statistical Methods in Water Resources*. Elsevier Science Publishing Co., New York, New York, 522 pp.
- Holmstrom, B. K., 1980. Low-Flow Characteristics of Streams in the St. Croix River Basin, Wisconsin. U.S. Geological Survey Water-Resources Investigations Open-File Report 80-696, 23 pp.
- Horton, R. E., 1945. Erosional Development of Streams and Their Drainage Basins; Hydrophysical Approach to Quantitative Morphology. *Bulletin of the Geological Society of America* 46:275-370.
- Lugo A. E. and C. Lowe, 1995. *Tropical Forests: Management and Ecology*. Springer-Verlag New York, Inc., New York, New York. 461 pp.
- Matalas, N. C., 1963. Probability Distribution of Low Flows. U.S. Geological Survey Professional Paper 434-A, 27 pp.
- Morris, G., 1984. Estimation of Minimum Streamflows at Ungaged Sites in Puerto Rico. *In: Workshop on Assessment of Puerto Rican Stream Resources with Regard to Management Needs*. U.S. Fish and Wildlife Service and USDA Forest Service Puerto Rico, Palmer, Puerto Rico, 7 pp.
- Riggs, H. C., 1967. Some Statistical Tools in Hydrology. U.S. Geological Survey Techniques of Water-Resources Investigations Book 4, Chapter A1, 39 pp.
- Riggs, H. C., 1968. Frequency Curves. U.S. Geological Survey Techniques of Water-Resources Investigations Book 4, Chapter A2, 15 pp.
- Riggs, H. C., 1972. Low Flow Investigations. U.S. Geological Survey Techniques of Water-Resources Investigations Book 4, Chapter B1, 18 pp.
- Riggs, H. C., 1973. Regional Analyses of Streamflow Characteristics. U.S. Geological Survey Techniques of Water-Resources Investigations Book 4, Chapter B3, 15 pp.
- Santiago-Rivera, L., 1992. Low-Flow Characteristics at Selected Sites on Streams in Eastern Puerto Rico. U.S. Geological Survey Water-Resources Investigations Report 92-4063, 46 pp.
- SAS, 1985. *SAS Users Guide: Basics*. Version 5 Edition. Cary, North Carolina, 1290 pp.
- Schneider, W. J., 1965. Areal Variability of Low Flows in a Basin of Diverse Geologic Units. *Water Resources Research* 1(4):509-515.
- Singh, K.P. and J. B. Stall, 1974. Hydrology of 7-Day 10-Year Low Flows. Proc. Paper 11013, *Journal of the Hydraulics Division, ASCE* 100(HY12):1753-1771.
- Strahler, A. N., 1954. Statistical Analysis in Geomorphic Research. *The Journal of Geology* 62(1):1-25
- Task Committee, 1980. Task Committee on Low-Flow Evaluation, Methods and Needs of the Committee on Surface Water Hydrology of the Hydraulics Division. *In: Characteristics of Low Flows*. Proc. Paper 15400, *Journal of the Hydraulics Division, ASCE* 106(HY5):717-731.
- Thomas, D.M., and M.A. Benson, 1970. Generalization of Streamflow Characteristics from Drainage-Basin Characteristics. U.S. Geological Survey Water Supply Paper 1975, 55 pp.
- U.S. Geological Survey, 1987. *Digital Elevation Models Data Users Guide*. National Mapping Program Tech. Intruc. Data Users Guide 5, 38 pp.
- Vogel, R. M. and C. N. Kroll, 1992. Regional Geohydrologic-Geomorphic Relationships for the Estimation of Low-Flow Statistics. *Water Resources Research* 28(9):2451-2458.
- Wadsworth, F. H. 1951. *Forest Management in the Luquillo Mountains*. I. The setting. *Carib. For.* 12:93-114.
- Water Resources Council, 1977. *Guidelines for Determining Flood Flow Frequency*. Bulletin No. 17A, Hydrology Committee, 25 pp.
- Wisler, C. O. and E. F. Brater, 1963. *Hydrology*. John Wiley & Sons, Inc., New York, New York, 408 pp.