

MINIMUM INSTREAM FLOW ESTIMATION AT UNGAGED STREAM SITES IN PUERTO RICO



June, 2009

PREPARED FOR:



PREPARED BY:



250 Tanca St.
P.O. Box. 9024157 Old San Juan
Tel. (787) 723-8005 Fax. (787) 721-3196
www.gmaeng.com

Table of Contents

1. INTRODUCTION	1
1.1. Study Description.....	1
1.2. Scope and Purpose of Report.....	1
1.3. Limitations of the Analysis	1
1.4. Authorization.....	1
2. MEAN ANNUAL RAINFALL MAP	2
2.1. Current Rainfall Map.....	2
2.2. Rainfall Data.....	3
2.3. Isoheytal Map.....	7
3. LOW FLOW ESTIMATION AT UNGAGED SITES	8
3.1. Regional Regression Analysis.....	8
3.1.1. Methodology	8
3.1.2. Streamflow Data	9
3.1.3. Regression Analysis	11
3.1.4. Regression Analysis Results.....	13
3.2. Station Index Method.....	15
4. CONCLUSIONS AND RECOMMENDATIONS	16
5. REFERENCES.....	17
APPENDIX A: SAMPLE CALCULATIONS.....	A-1

List of Figures

- Figure 1: Mean annual rainfall, prepared using rainfall data from 1931 to 1960 (Calversbert, 1970).
- Figure 2: Rain gages with more than 10% difference in mean annual rainfall between station data and DNER (2003) rainfall map.
- Figure 3: National Climatic Data Center rainfall gage stations used to generate mean annual rainfall map.
- Figure 4: Rainfall-runoff relationship resulting from the final rainfall contour map.
- Figure 5: Mean annual rainfall surface map for Puerto Rico.
- Figure 6: Mean annual rainfall contours map for Puerto Rico.
- Figure 7: USGS streamflow gage stations location used in regression analysis.
- Figure 8: Comparison of Mean Square Error for the different regression models.
- Figure 9: Relation between observed and predicted Q_{mean} flow.
- Figure 10: Relation between observed and predicted Q_{90} flow.
- Figure 11: Relation between observed and predicted Q_{95} flow.
- Figure 12: Relation between observed and predicted Q_{98} flow.
- Figure 13: Relation between observed and predicted Q_{99} flow.
- Figure 14: Relation between observed and predicted flow for all recurrence intervals compared to line of perfect correlation to gage station data.
- Figure 15: Error between observed and predicted values for the different return intervals.
- Figure A-1: Example 1. Q_{99} of an ungaged site at Río Bauta is determined using USGS Río Bauta Stations (50034000) as reference.
- Figure A-2: Example 2. Q_{99} of an ungaged site with an intake at Río Cialitos is determined using USGS Río Bauta Stations (50034000) as reference.

MINIMUM INSTREAM FLOW ESTIMATION AT UNGAGED STREAM SITES IN PUERTO RICO

1. INTRODUCTION

1.1. Study Description

Minimum streamflow information is commonly used to determine the water available for extraction and to analyze instream environmental parameters. Minimum streamflow estimates are frequently desired at ungaged locations, either on a stream having gages at other locations, or a stream without any gages.

This study presents regional regression equations to estimate minimum streamflow using two parameters: watershed area and mean annual rainfall. These equations are based on a revised rainfall map which incorporates the available rainfall data from 127 raingage stations and also considers geographic parameters.

1.2. Scope and Purpose of Report

This study describes a methodology for minimum streamflow estimation and consists of two major elements.

- A new isoheytal map for Puerto Rico showing contour lines of mean annual rainfall has been prepared based on contour curve fitting within the GIS environment, and adjusted based on geographic criteria such as proximity to the coastline, elevation, and vegetation mapping (Holdrige life zone).
- Regional regression equations for mean discharge and for minimum streamflow exceedances (Q_{90} , Q_{95} , Q_{98} , Q_{99} , and Q_m) have been prepared based on the revised isoheytal map. The regression equations have been developed to simulate current “natural” runoff conditions, using streamflow data from gage stations with minimum influence by upstream reservoirs, diversions and loss to groundwater.

1.3. Limitations of the Analysis

The minimum streamflow equations do not account for the effects produced by the existence of reservoir operations, intakes and diversions, or infiltration into coastal alluvial or karst aquifers. For those cases it is necessary to make a more detailed study in order to obtain more representative results. The equations should not be used in watersheds smaller than one square mile.

1.4. Authorization

Preparation of this report has been authorized by the Department of Natural and Environmental Resources (DNER) by contract # 050-08-001302.

2. MEAN ANNUAL RAINFALL MAP

2.1. Current Rainfall Map

Several isoheytal maps have been prepared for Puerto Rico over the past four decades (Calversbert, 1970; Black and Veatch, 1971; DNER, 2003). The map published by Calversbert (1970) presented in Figure 1, was prepared using rainfall data from 1930 to 1960. The map published by DNER (2003) has significant discrepancies with station rainfall at several locations. Table 1 presents those stations with more than 10% of difference between values obtained from that map and those obtained from the rain gage station data. The location of these stations is presented in Figure 2. The current analysis stemmed from the desire to use the most recently available data to minimize discrepancies between maps and station data.

Table 1: Rain Gages with more than 10% of Difference from DNER (2003) Isoheytal Map.

Rain Gage	Mean Annual Rainfall (in/yr)		% Difference
	Gage Station Data	DNER Map	
San Juan City	59.1	79.0	25.2
Peñuelas 1 NE	54.7	43.7	-25.1
Toro Negro Forest	93.9	76.6	-22.5
Río Blanco Lower	107.8	91.3	-18.1
Carite Dam	74.4	90.2	17.6
Peñuelas Salto Garzas	75.6	90.7	16.7
Lajas Substn	45.1	53.1	14.9
Caguas 1 W	54.1	63.3	14.6
Negro-Corozal	70.8	82.7	14.4
Yauco 1 NW	45.5	39.8	-14.4
Aibonito 1 S	57.7	50.8	-13.7
Jayuya	75.5	67.0	-12.7
Melanía Dam	39.0	44.3	12.0
Potala	30.0	34.1	11.9
Río Piedras Exp Stn	69.4	78.7	11.9
San Lorenzo Espino	119.0	106.8	-11.4
Sabana Grande 2 ENE	62.4	56.1	-11.3
Toro Negro Plt 2	88.9	99.9	11.0
Isabela Substn	61.3	68.5	10.5
Humacao 2 SSE	83.2	92.4	9.9

2.2. Rainfall Data

Rainfall data were obtained from the US Department of Commerce National Climatic Data Center (NCDC). The NCDC database has 143 rain gages with record periods dating from 1900. A screening process was undertaken to eliminate from the analysis stations with less than 15 years of data. Other eliminated stations were those very close to other stations and with a notable difference in mean annual rainfall with other nearby gages having a longer period of record. Table 2 presents the omitted rain gages and the reasons. The rain gages used for the analysis are presented in Table 3 and Figure 3. Some stations with 14 years of record were included in the analysis because the area they are located lack of any surrounding station and the 14 years of data provides a better representation of the area.

Table 2: Rain Gages Omitted in the Analysis.

Rain Gage	Record Begin Date	Record End Date	Record Years	Rainfall (in/yr)	Reason
Bayaney	Jun-1970	Aug-1979	9	38.75	Record < 15 Years
Boca	Jul-1996	Dec-2006	10	37.21	Record < 15 Years
Caguas 2 ENE	Aug-1960	May-1967	7	50.33	Record < 15 Years
Garrochales	Sep-1965	Apr-1970	5	60.41	Record < 15 Years
Guayanilla	Jan-1955	Aug-1961	7	36.89	Record < 15 Years
Guineo RSVR	Jan-1955	Mar-1969	14	97.52	Record < 15 Years
Indiera Baja	Nov-1952	Sep-1962	10	74.62	Record < 15 Years
Jayuya 1 SE	Mar-1960	May-1981	21	60.51	Inconsistent with nearby gage stations ^{A/}
Josefa	Jan-1955	Jan-1969	14	45.32	Record < 15 Years
La Fe	Jan-1956	Mar-1969	13	72.94	Record < 15 Years
Maricao	Jan-1955	Apr-1969	14	103.8	Record < 15 Years
Naguabo 3 E	Apr-1972	May-1983	11	79.93	Record < 15 Years
Naguabo 6 W	Jan-1955	May-1967	12	90.43	Record < 15 Years
Palmarito	Mar-1963	Apr-1975	12	84.55	Record < 15 Years
Ponce Mercedita Ap	Jan-1957	Nov-1968	12	34.19	Record < 15 Years
Potala	Jan-1955	Feb-1969	14	30.02	Record < 15 Years
Rincón 2 NNW	Nov-1957	May-1968	11	57.78	Record < 15 Years
Río Piedras	Jan-1931	Dec-1961	30	76.79	Inconsistent with nearby gage stations ^{B/}
St Just	Jan-1955	Dec-1966	12	80.12	Record < 15 Years
Salttillo 2 Adjuntas	May-1981	Dec-1991	11	91.97	Record < 15 Years
Toa Baja Levitown	Jan-2005	Dec-2006	2	75.25	Record < 15 Years
Vieques Island #2	Mar-1983	Jan-1994	11	49.92	Record < 15 Years
Yauco 1 S	Jan-1955	Jun-1969	11	29.55	Record < 15 Years
Yaurel 3 NNE	Jan-1955	Mar-1969	11	45.6	Record < 15 Years

^{A/} Inconsistent with gage station Jayuya (Period of Record 1909-2002, Rainfall of 75.5 in/yr).

^{B/} Inconsistent with gage station Río Piedras Exp Stn (Period of Record 1959 - 2006, Rainfall of 69.37 in/yr).

Table 3: Rain Gages Used to Generate Isoheytal Map.

Rain Gage	Record Begin Date	Record End Date	Rainfall (in/yr)
Aceituna	Jan-1955	Dec-2006	76.90
Adjuntas 1 NW	Jan-1955	Dec-2006	78.29
Adjuntas Substn	Jan-1970	Dec-2006	73.62
Aguirre	Jan-1931	Oct-1966	42.89
Aguirre	Apr-1955	Dec-2006	40.47
Aibonito 1 S	Jan-1906	Dec-2006	57.72
Arecibo 3 Ese	Jul-1931	Jan-1999	54.39
Arecibo Obsy	Feb-1980	Dec-2006	82.64
Barceloneta 2	Jan-1955	May-1990	53.34
Barceloneta 3 SW	Sep-1990	Dec-2006	60.39
Barranquitas	Jan-1955	Dec-1991	57.38
Benavente-Hormigueros	Aug-1973	Aug-2002	59.95
Borinquen AP	Feb-1974	Dec-2006	53.66
Cabo Rojo	Jan-1955	Aug-1969	56.41
Cacaos-Orocovis	May-1981	Dec-2006	82.76
Caguas	May-1899	Aug-1960	61.25
Caguas 1 W	Mar-1970	Mar-1995	54.08
Calero Camp	Jan-1955	Dec-2006	56.29
Cambalache Exp Forest	Jan-1932	Feb-1966	51.00
Candelaria Toa Baja	Jan-1955	May-1973	78.14
Candelaria Toa Baja	Oct-1973	Aug-1995	74.74

Rain Gage	Record Begin Date	Record End Date	Rainfall (in/yr)
Canóvanas	Jan-1955	Dec-2006	74.37
Caonillas Utuado	Jan-1955	Nov-1987	73.32
Caonillas Villalba	Jan-1955	Sep-1969	55.12
Carite Dam	Jan-1955	Apr-1980	74.37
Carite Plt 1	Jan-1955	Mar-1980	72.83
Cataño	Jan-1955	May-1976	69.17
Cayey 1 E	Jan-1955	Jun-2001	58.32
Central San Francisco	Jan-1955	Jun-1996	31.04
Cerro Gordo Ciales	Oct-1969	Sep-1997	82.26
Cerro Maravilla	Apr-1969	Dec-2006	94.46
Cidra 1 E	Sep-1899	Jun-1994	66.49
Coamo 2 SW	Jan-1955	Dec-2003	36.40
Coloso	Oct-1899	Dec-2006	80.29
Comerío Falls Plt 2	Feb-1959	May-1974	65.74
Corozal Substn	Jan-1931	Dec-2006	75.06
Corral Viejo	Apr-1970	Dec-2006	59.18
Culebra Island	Jan-1920	Jul-1975	33.08
Dorado 2 Wnw	Jan-1931	May-2006	65.21
Dos Bocas	Jan-1937	Dec-2006	76.82
Ensenada 1 W	Jan-1955	Dec-2006	30.76
Fajardo	Jan-1931	Jan-1996	64.81

Rain Gage	Record Begin Date	Record End Date	Rainfall (in/yr)
Garzas	Jan-1939	Jan-1981	86.16
Guajataca Dam	Jan-1955	Dec-2006	71.13
Guavate Camp	Dec-1969	Jun-1994	99.56
Guayabal	Jan-1955	Dec-2006	49.85
Guayama 2E	Jan-1911	Dec-2006	52.61
Gurabo	Apr-1946	May-1967	63.75
Gurabo Substn	Mar-1956	Dec-2006	64.19
Hacienda Constanza	Oct-1969	Dec-2006	73.89
Hato Arriba Arecibo	Feb-1974	Aug-1994	55.08
Humacao 2 SSE	Jan-1931	Jan-1996	83.19
Indiera Alta	Oct-1962	Jun-1990	76.42
Isabela Substn	Jan-1901	Dec-2006	61.30
Jájome Alto	Jan-1955	Dec-2006	77.47
Jayuya	Apr-1909	Aug-2002	75.50
Juana Díaz Camp	Jan-1931	Dec-2006	42.21
Juncos 1 SE	Jan-1931	Dec-2006	66.87
Lajas Substn	Jan-1900	Dec-2006	45.14
La Muda Caguas	Sep-1971	Jun-1994	78.92
Lares	Jun-1903	Dec-1991	93.21
Los Caños	Jan-1955	Aug-1973	62.63
Magüeyes Island	Jan-1959	Nov-2006	28.68
Manatí 2 E	Jan-1900	Dec-2006	62.23
Maricao 2 SSW	May-1969	Dec-2006	95.33

Rain Gage	Record Begin Date	Record End Date	Rainfall (in/yr)
Maricao Fish Hatchery	Jan-1955	Dec-2006	98.6
Matrullas Dam	Jan-1955	Apr-1981	86.64
Maunabo	May-1899	Apr-2003	73.95
Mayagüez City	Jan-1957	Dec-2006	75.08
Mayagüez AP	Jan-1900	Dec-2006	76.16
Melania Dam ^{AV}	Jan-1955	Jan-1969	39.03
Mona Island	Jan-1955	Aug-1974	35.91
Mona Island 2	Feb-1980	Dec-2006	39.09
Monte Bello Manatí	Oct-1969	Sep-2001	61.62
Mora Camp	Jan-1955	Dec-2006	58.9
Morovis 1 N	Feb-1956	Dec-2006	71.37
Negro-Corozal	Jan-1976	Dec-2006	70.75
Paraíso	Jan-1956	Dec-2006	98.21
Patillas	Apr-1982	Jun-2003	57.79
Patillas Dam	Jan-1931	Jan-1969	70.23
Peñuelas Salto Garzas	Mar-1971	Dec-2003	75.55
Peñuelas 1 NE	Jan-1955	Feb-1971	54.69
Pico del Este	Oct-1969	Jun-2005	174.38
Ponce 4 E	Apr-1954	Dec-2006	35.12
Ponce City	Jul-1970	Aug-1998	29.45
Puerto Real	Jan-1955	Aug-2001	48.04
Quebradillas	Jan-1955	Sep-2000	55.69
Rincón	Jun-1968	Nov-2006	55.41

Rain Gage	Record Begin Date	Record End Date	Rainfall (in/yr)
Río Blanco Lower	Jan-1955	Dec-2006	107.83
Río Blanco Upper	Jan-1955	Mar-1974	161.78
Río Cañas	Jan-1955	Dec-1969	36.66
Río Grande el Verde	Feb-1956	Dec-1987	96.36
Río Jueyes ^{A/}	Jan-1955	Jan-1969	31.13
Río Piedras Exp Stn	Jan-1959	Dec-2006	69.37
Roosevelt Roads	Jul-1959	Mar-2004	51.61
Sabana Grande 2 ENE	May-1977	Dec-2006	62.39
Sabater ^{A/}	Jan-1955	Jan-1969	37.78
San Cristóbal	Jan-1956	Mar-1972	76.33
San Germán 4 W	Nov-1904	Jul-1973	64.37
San Juan City	Jan-1955	May-1977	59.07
San Juan Intl Ap	Jan-1956	Dec-2006	54.36
San Lorenzo 3S	Mar-1966	Dec-2006	98.57
San Lorenzo Espino ^{A/}	Jan-1945	Jun-1959	118.96

Rain Gage	Record Begin Date	Record End Date	Rainfall (in/yr)
San Lorenzo Farm 2 NW	Jan-1955	Sep-1988	72.61
San Sebastián 2 WNW	Apr-1955	Oct-1997	91.25
Santa Isabel 2 ENE	Jan-1955	Dec-2006	34.50
Santa Rita	Jan-1955	Dec-2006	33.30
Toa Baja 1 SSW	Jan-1955	Aug-1994	68.01
Toro Negro Forest	Aug-1982	Dec-2006	93.85
Toro Negro PLT 2	Jan-1955	Jul-1981	88.94
Trujillo Alto 2 SSW	Feb-1957	Dec-2006	71.96
Utuaedo	Jan-1931	Jul-1998	73.36
Vieques Island	Jan-1955	Sep-1976	42.68
Villalba 1 SE	Jan-1955	Dec-2006	64.19
Yabucoa 1 NNE	Jan-1955	Mar-1995	79.11
Yauco 1 NW	Dec-1981	Dec-2006	45.52

^{A/} Station with 14 years of record included in the analysis.

2.3. Isoheytal Map

The NCDC rainfall gage station data were used to interpolate a surface within the Arc-GIS environment using *spline curves* to create the initial isoheytal countours. *Spline* algorithms can create very smooth surfaces from moderately detailed data and provide exact interpolation within smoothing limits. This method is best suitable for gently varying surfaces, such as rainfall.

The initial contour lines were then adjusted based on coastline proximity, elevation and vegetation mapping. The resulting contours were checked against USGS streamgage data, comparing rainfall and runoff per unit of watershed area to reveal any unusual discrepancies. The rainfall-runoff relationship resulting from the final isoheytal map is presented in Figure 4. The resulting rainfall surface and rainfall contour map are presented in Figure 5 and Figure 6 respectively.

The following physical parameters were used to realign rainfall contours in areas of sparse gage data:

- Contours adjacent to the ocean were adjusted to lie roughly parallel to the coastline, instead of locally curving around coastal rain gage stations;
- Contours along the Cordillera Central were adjusted to run generally parallel to the mountains peaks to better reflects orographic effects; and
- Contours were also checked against Holdridge Life Zone vegetation map, since this mapping system reflects long-term rainfall patterns (Ewel and Whitmore, 1973).

3. LOW FLOW ESTIMATION AT UNGAGED SITES

This section describes two methods to estimate low-flow at ungaged stream sites, regional regression analysis and station index method.

3.1. Regional Regression Analysis

3.1.1. Methodology

Multiple Linear Regression techniques were used to develop a series of equations to estimate average daily streamflow equaled or exceeded on 90%, 95%, 98%, and 99% of the time as well as mean streamflow, also referred as Q_{90} , Q_{95} , Q_{98} , Q_{99} , and Q_m respectively. Equations used average daily streamflow as the dependent variable, and Watershed Area and Mean Annual Rainfall as independent variable. These two independent parameters were selected because they have the highest predictive values and because these data are readily available.

The Multiple Linear Regression analysis relates two or more explanatory variables with a response variable by fitting a linear equation to the observed data (McCuen, 1993). The Multiple Linear Regression Model for the analysis is defined by:

$$Q = a + b \cdot A + c \cdot P \quad (3.1)$$

where,

- Q = Streamflow (cfs)
- A = Basin area (mi²)
- P = Mean annual rainfall (in/yr)
- a, b, c = Regression coefficients

The matrix notation of the model has the following form:

$$\hat{Y} = X \cdot \beta + \epsilon \quad (3.2)$$

where:

\hat{Y} : Estimated discharge vector:

$$\hat{Y} = \begin{bmatrix} \hat{Q}_1 \\ \vdots \\ \hat{Q}_i \end{bmatrix} \quad (3.3)$$

X: Watershed parameters matrix:

$$X = \begin{bmatrix} 1 & A_1 & P_1 \\ 1 & \vdots & \vdots \\ 1 & A_i & P_i \end{bmatrix} \quad (3.4)$$

β : Regression coefficient vector:

$$\beta = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad (3.5)$$

ϵ : Residual (error term):

$$\epsilon = Y_i - \hat{Y}_i \quad (3.6)$$

where,

Y = True streamflow vector

\hat{Y} = Estimated streamflow vector

Logarithmic transformation was performed to linearize the hydrologic data. The resulting transformed equation was used for the analysis:

$$\log Q = a + b \cdot \log A + c \cdot \log R \quad (3.7)$$

Applying antilogarithm to both sides of the equations, the previous equation can be rewritten as:

$$Q = 10^a \cdot A^b \cdot P^c \quad (3.8)$$

Eq. (3.8) is the form used for the regional regression equations.

3.1.2. Streamflow Data

Daily streamflow data were obtained from the 22 USGS gage stations in Puerto Rico with more than ten years of daily data and where low flows were little affected by reservoir operation, water supply intakes and diversions, or infiltration into coastal alluvial or karst aquifers. The drainage area for these stations ranges from 1 to 100 square miles. Figure 7 shows the location of these stations and the watershed area tributary to each.

Minimum streamflow values were determined at each station for exceedance probabilities of 99, 98, 95 and 90% using a daily flow duration analysis (Table 4).

Table 4: USGS Streamflow Data Used in Regression Analysis.

Station Number	Station Name	Drainage Area (mi ²)	Rainfall (in/yr)	Q _m (cfs)	Q ₉₉ (cfs)	Q ₉₈ (cfs)	Q ₉₅ (cfs)	Q ₉₀ (cfs)
50025155	Río Saliente	9.31	78.10	30.75	3.30	3.66	4.50	5.89
50034000	Río Bauta	16.75	80.12	39.33	3.40	3.80	4.70	6.10
50048770	Río Piedras	7.53	75.27	21.37	1.40	1.60	2.50	3.90
50049100	Río Piedras	15.47	73.40	54.68	4.61	6.60	8.80	11.00
50050900	Río Grande de Loíza	6.00	116.29	32.43	4.60	5.10	5.90	6.90
50053025	Río Turabo	7.16	100.61	22.26	4.00	4.40	5.30	6.20
50055225	Río Cagüitas	16.91	56.96	34.10	2.70	3.90	6.10	7.70
50058350	Río Cañas	7.57	69.72	14.96	1.60	1.82	2.60	3.20
50065500	Río Mameyes	6.80	143.46	55.72	10.96	11.00	14.00	16.00
50065700	Río Mameyes	11.87	107.09	73.4	8.70	10.00	13.00	17.00
50075000	Río Icados	1.25	156.48	14.17	2.50	3.30	4.00	4.70
50081000	Río Humacao	6.60	88.88	21.00	0.77	0.82	2.60	5.80
50092000	Río Grande de Patillas	18.38	89.23	59.31	7.50	8.80	11.00	13.00
50100200	Río Lapa	9.98	55.43	9.04	0.00	0.02	0.07	0.15
50100450	Río Majada	16.45	65.33	8.80	0.00	0.01	0.09	0.33
50108000	Río Descalabrado	12.87	43.53	19.95	0.04	0.08	0.15	0.34
50110900	Río Toa Vaca	14.25	63.63	16.75	0.86	1.00	1.30	1.80
50113800	Río Cerrillos	11.87	80.02	29.59	3.60	4.00	4.80	6.20
50136000	Río Rosario	17.64	93.22	52.81	5.50	6.90	9.00	11.00
50141000	Río Blanco	15.19	75.62	37.13	7.00	7.50	8.70	10.00
50144000	Río Grande de Añasco	92.30	87.55	425.53	47.00	54.00	64.00	75.00
50147800	Río Culebrinas	71.60	89.36	294.10	26.00	29.00	35.00	42.00

3.1.3. Regression Analysis

Three multiple linear regression models were analyzed for this study:

1) Ordinary Least Squares (OLS):

Hydrologists have commonly used OLS method to estimate the regression coefficient vector of the linear model for regression analyses. This method obtains parameter estimates that minimize the sum of squared residuals. The OLS regression coefficient vector (β_{OLS}) can be solved using matrix analysis by solving Eq. (3.9):

$$\beta_{OLS} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y \quad (3.9)$$

where:

Y= Observed discharge vector

X= Matrix of the watershed parameters

The OLS method is the faster regression method for this type of analysis and is analyzed very quickly once the data in Table 4 is setup in a spreadsheet. However, it introduces a significant error because the regression is calculated on logged parameter values, which represent a non-linear transformation of the original dataset.

2) Manual Numerical Search for Least Square Error (MNS):

The MNS method consists of a numerical search of the least square error (MNS) using an iterative spreadsheet solver. This method used as initial values the results from the OLS, and by an iterative process the regression coefficients are varied until the minimum square error is found, as compared to original data (without the log transformation). This method required approximately twice the time required by the OLS analysis.

3) Generalized Least Squares (GLS):

Recent studies of minimum streamflow analysis have employed the Generalized Least Squares (GLS) regression technique. This method takes into account varying sampling error and cross correlation among concurrent flows. It was developed by Stedinger and Tasker (1985) and has become a standard for regression analysis of flood frequency data. Stedinger and Tasker (1985) and Stedinger and Tasker (1986) showed that the GLS procedures provide more accurate parameter estimators, relatively unbiased estimate of the model error variance and a better estimation of parameter sampling variance than those estimated with OLS. Moss and Tasker (1991) showed that GLS procedures describe model accuracy in regional regression analysis better than OLS.

A discussion and implementation of the GLS is presented in Stedinger and Tasker (1989) and Griffis and Stedinger (2007). The GLS estimate of β_{GLS} is given by Stedinger and Tasker (1985) as:

$$\beta_{GLS} = (X^T \Lambda^{-1} X)^{-1} X^T \Lambda^{-1} Y \quad (3.10)$$

where Λ is the covariance of the model. In the GLS model Λ is estimated by

$$\Lambda = \sigma_\delta^2 I_N + \hat{\Sigma} \quad (3.11)$$

where σ_δ^2 is an estimate of the model error variance and $\hat{\Sigma}$ is an $(i \times i)$ matrix of sampling covariance with elements:

Diagonal elements ($i = j$):

$$\Sigma_{ij} = \text{Var}[Y_i] \quad (3.12)$$

Chowdury and Stedinger (1991) provide the following first order approximation of the sampling error ($\text{Var}[Y_i]$),

$$\begin{aligned} \text{Var}[Y_i] = & \left[1 + k_i \gamma_i + k_i^2 \left(\frac{1}{2} + \frac{3}{8} \gamma_i^2 \right) + k_i \gamma_i W_i \frac{\partial k_i}{\partial \gamma_i} \left(3 \gamma_i + \frac{3}{4} \gamma_i^3 \right) \right. \\ & \left. + W_i^2 \left(\frac{\partial k_i}{\partial \gamma_i} \right)^2 \left(6 + 9 \gamma_i^2 + \frac{15}{8} \gamma_i^4 \right) \right] \frac{\hat{\sigma}_i^2}{n_i} + (1 - W_i)^2 \hat{\sigma}_i^2 \text{MSE}_{\gamma_i} \left(\frac{\partial k_i}{\partial \gamma_i} \right)^2 \end{aligned} \quad (3.13)$$

Off-diagonal elements ($i \neq j$):

$$\begin{aligned} \Sigma_{ij} = & \hat{\rho}_{ij} \frac{m_{ij} \hat{\sigma}_i \hat{\sigma}_j}{n_i n_j} \left[1 + \frac{k_i \gamma_i}{2} + \frac{k_j \gamma_j}{2} + \frac{k_i k_j}{2} (\hat{\rho}_{ij} + 0.75 \gamma_i \gamma_j) \right. \\ & + k_j \gamma_i W_i \frac{\partial k_i}{\partial \gamma_i} (3 \hat{\rho}_{ij} + 0.75 \gamma_i \gamma_j) + k_i \gamma_j W_j \frac{\partial k_j}{\partial \gamma_j} (3 \hat{\rho}_{ij} + 0.75 \gamma_i \gamma_j) \\ & \left. + W_i W_j \hat{\sigma}_i \hat{\sigma}_j \frac{\partial k_i}{\partial \gamma_i} \frac{\partial k_j}{\partial \gamma_j} \text{Cov}[\gamma_i, \gamma_j] \right] \end{aligned} \quad (3.14)$$

where:

$\hat{\sigma}_i$ = estimate of the standard deviation of flows at site i

K_i = T -year frequency factor for the distribution used

n_i = record length at site i

γ_i = station skew at site i

m_{ij} = concurrent record length of sites i and j

$\hat{\rho}_{ij}$ = estimate of the lag zero correlation of flows between sites i and j

$\text{Cov}[\gamma_i, \gamma_j]$ = covariance matrix estimator,

$$\text{Cov}[\gamma_i, \gamma_j] = \hat{\rho}_{\gamma_i \gamma_j} \sqrt{(\text{Var}[\gamma_i] \text{Var}[\gamma_j])} \quad (3.15)$$

Methodologies for estimating station skew are presented in IACWD (1982). The cross correlation ($\hat{\rho}_{\gamma_i \gamma_j}$) is estimated using the approximation developed by Martins and Stedinger (2002),

$$\hat{\rho}_{\gamma_i \gamma_j} = \text{Sign}(\hat{\rho}_{ij}) \left(m_{ij} / \sqrt{(m_{ij} + n_i)(m_{ij} + n_j)} \right)^\delta \quad (3.16)$$

and Griffis (2003) provides an approximation for ($\text{Var}[\gamma_i]$ and $\text{Var}[\gamma_j]$),

$$\text{Var}[\gamma_i] = \left[\frac{6}{n_i} + a(n_i) \right] \left[1 + \left(\frac{9}{6} + b(n_i) \right) \gamma_i^2 + \left(\frac{15}{48} + c(n_i) \right) \gamma_i^4 \right] \quad (3.17)$$

The model error variance σ_δ^2 and the vector of regression coefficients β_{GLS} are estimated jointly by iteratively searching for a nonnegative solution to the equation (Stedinger and Tasker, 1985),

$$(\hat{Y} - X * \beta)^T (\sigma_\delta^2 * I_n + \hat{\Sigma})^{-1} (\hat{Y} - X * \beta) = N - (k + 1) \quad (3.18)$$

where σ_δ^2 is the estimator of the unknown model error variance and β_{GLS} is given by Eq.(3.10). This method takes approximately eighty times the effort required to the OLS analysis because of the need to research and calculate many intermediate parameter values.

3.1.4. Regression Analysis Results

The mean square error was used as an indicator of the goodness of fit of the regression model. The mean square error is defined as average squared difference between the observed and estimated values as defined by:

$$MSE = \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (3.19)$$

Table 5 shows the mean square error produced by the different models. These results are presented graphically in Figure 7.

Table 5: Mean Square Error Comparison.

Equation	OLS	MNS	GLS
Q ₉₉	118.31	8.44	4.31
Q ₉₈	55.35	10.05	5.07
Q ₉₅	44.96	11.60	6.72
Q ₉₀	37.32	16.49	10.09
Q _{mean}	423.52	144.14	122.06

Results of Table 5 show that the best fit to the observed data is provided by the GLS method, and the worst fit is provided by the OLS method. Results of MNS method are close to the GLS results. Based on these results, the GLS model was selected as the model which provides the best approximation to the observed data. Table 6 presents the regression parameters obtained from the GLS analysis.

Table 6: Regression Equation Parameters for Estimating Runoff.

Parameter	Regression Coefficients				
	Q _{mean}	Q ₉₀	Q ₉₅	Q ₉₈	Q ₉₉
10 ^a	9.869E-04	3.127E-06	8.985E-07	2.237E-07	1.237E-07
b	1.075	0.986	1.016	1.057	1.059
c	1.787	2.756	2.966	3.200	3.307

The following are the equations obtained from this analysis:

$$Q_m = 9.869 * 10^{-04} * A^{1.075} * P^{1.787} \quad (3.20)$$

$$Q_{90} = 3.127 * 10^{-06} * A^{0.986} * P^{2.756} \quad (3.21)$$

$$Q_{95} = 8.958 * 10^{-07} * A^{1.016} * P^{2.966} \quad (3.22)$$

$$Q_{98} = 2.237 * 10^{-07} * A^{1.057} * P^{3.200} \quad (3.23)$$

$$Q_{99} = 1.237 * 10^{-07} * A^{1.059} * P^{3.307} \quad (3.24)$$

Figure 9 to Figure 13 compare actual streamflows to the values estimated by the regression equations. The equations produce a good fit to the observed data when the estimated flow exceeds 2 cfs (Figure 14). When the estimated values were less than 2 cfs

the equations are not reliable, as can be seen in Figure 14. Figure 15 shows how the percent error between the observed and estimated values increases as the estimated flow decreases. Stations with estimated low flow less than 2 cfs and large error correspond to small south coast watershed.

3.2. Station Index Method

The station index method determines streamflow at an ungaged site using as reference a streamflow data from a nearby gage station. The most common procedure is the Drainage-Area ratio method. This method computes the streamflow at the ungaged location based on the ratio of drainage areas between the gaged and ungaged location. This method is suitable when the mean annual rainfall is similar for both watersheds. This method is defined by the following equation:

$$Q_u = \frac{A_u}{A_g} Q_g \quad (3.25)$$

where:

Q_u = minimum streamflow at the ungaged site

Q_g = minimum streamflow at the gaged site

A_u = drainage area at the ungaged site

A_g = drainage area at the gaged site

However in Puerto Rico rainfall variations are frequently large over short distances due to orographic effects. Because of this a more appropriate method to use in Puerto Rico is to translate the data from the gaged site to the ungaged site by the ratio of mean flows computed by the regression equation (Eq. 3.20). The following equation defines this method:

$$Q_u = \frac{Q_{mean_u}}{Q_{mean_g}} Q_g = \frac{A_u^{1.075} * P_u^{1.787}}{A_g^{1.075} * P_g^{1.787}} \quad (3.26)$$

where:

Q_u = minimum streamflow at the ungaged site

Q_g = minimum streamflow at the gaged site

Q_{mean_u} = mean streamflow computed at ungaged site (Eq. 3.20)

Q_{mean_g} = mean streamflow computed at gaged site (Eq. 3.20)

P_u = mean annual rainfall (in/yr) at ungaged site

P_g = mean annual rainfall (in/yr) at gaged site

4. CONCLUSIONS AND RECOMMENDATIONS

This report presents a new mean annual rainfall map for Puerto Rico (Figure 5 and Figure 6).

- Regional regression equations (Eq. 3.20 to 3.24) for mean discharge and for minimum streamflow (Q_m , Q_{90} , Q_{95} , Q_{98} , and Q_{99}) are presented in section 3.1.4 of this report.
- The regression equations are not reliable for estimated discharges less than 2.0 cfs (Figure 14 and Figure 15). The regional regression equations are not suitable for watersheds with less than 1 square mile of drainage area. In these small watersheds factors such as vegetative cover, soils and local groundwater interactions and the possibility of unknown private withdrawals, have a much larger role in determining low flow.

5. REFERENCES

- Calvesbert, R.J., 1970, *Climate of Puerto Rico and U.S. Virgin Islands*, revised: U.S. Environmental Science Services Administration, *Climatology of the United States*.
- Chowdury, J.U. and Stedinger, J.R., 1991. Confidence interval for design floods with estimated skew coefficient. *Journal of Hydraulic Engineering*, 117 (7), 811-831.
- Ewel, J.J. and Whitemore, J.L., 1973. *The Ecological Life Zones of Puerto Rico and the U.S. Virgin Islands*. Puerto Rico.
- Griffis, V.W., 2003. *Evaluation of Log-Pearson type 3 Flood Frequency Analysis Methods Addressing Regional Skew and Low Outliers*. M.S. Thesis, Cornell University.
- Griffis, V. W. and Stedinger, J. R., 2007. The use of GLS regression in regional hydrologic analyses. *Journal of Hydrology*, 344(1-2), 82-95.
- Interagency Committee on Water Data (IACWD), 1982. *Guidelines for Determining Flood Flow Frequency: Bulletin 17-B (revised and corrected)*. Hydrology Subcommittee, Washington, DC, March 1982, pp. 28.
- Martins, E.S. and Stedinger, J.R., 2002. Cross correlations among estimators of shape. *Water Resources Research*, 38 (11), 1252. doi:10.1029/2002WR00158.
- McCuen, R.H., 1993. *Microcomputer Applications in Statistical Hydrology*. New Jersey: Prentice Hall.
- Moss, M.E. and Tasker, G.D., 1991. An intercomparison of hydrological network-design technologies. *Hydrology Scientific Journal*, 36 (3), 209.
- Stedinger, J.R. and Tasker, G.D., 1985. Regional hydrologic analysis 1. Ordinary, weighted, and generalized least squares compared. *Water Resources Research*, 21 (9), 1421-1432.
- Stedinger, J.R. and Tasker, G.D., 1986. Regional hydrologic analysis 2. Model-error estimators, estimation of sigma and log-Pearson type 3 distributions. *Water Resources Research*, 22 (10), 1487-1499.
- Tasker, G.D. and Stedinger, J.R., 1989. An operational GLS model for hydrologic regression. *Journal of Hydrology*, 111, 361-375.

FIGURES

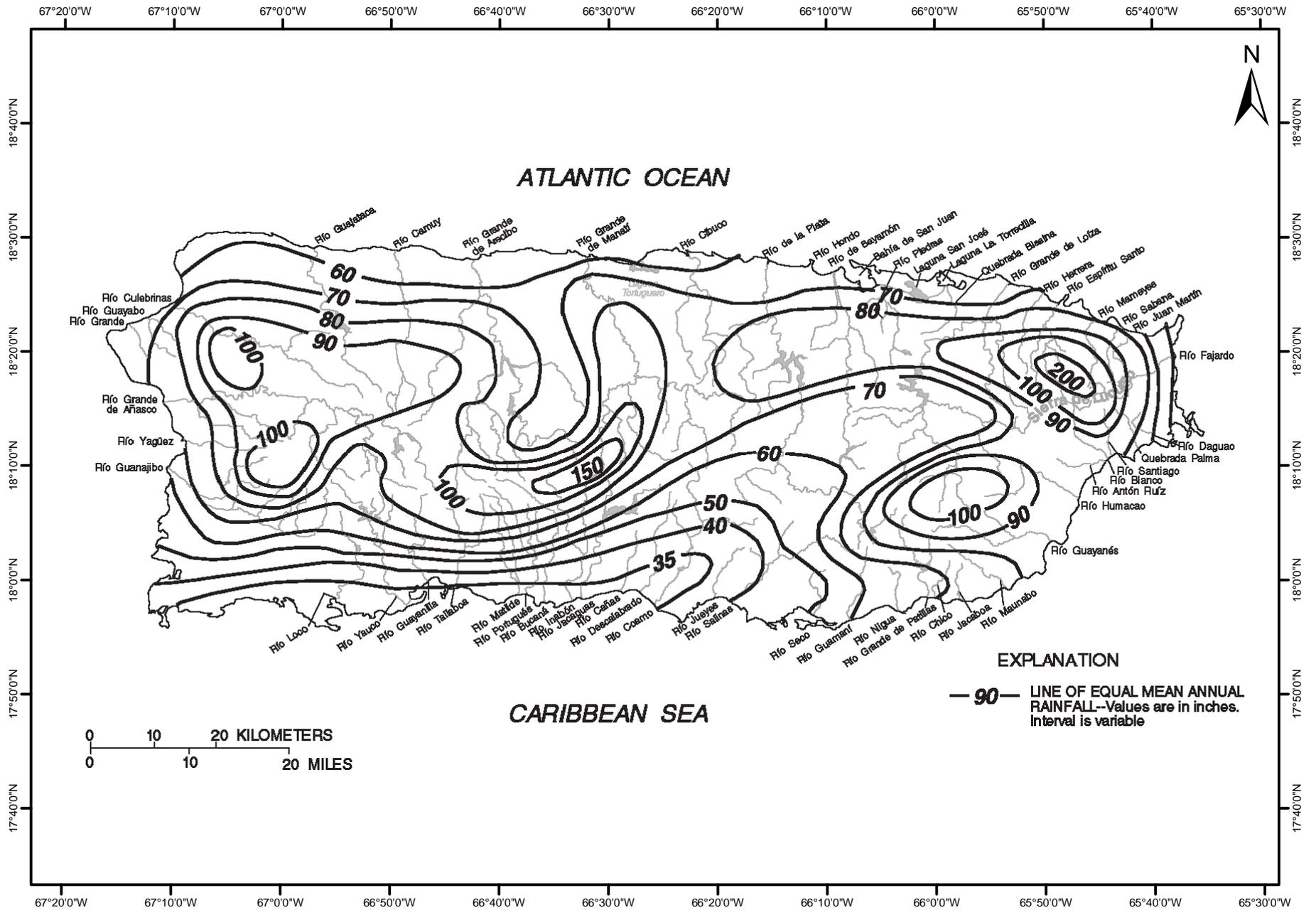


Figure 1: Mean annual rainfall, prepared using rainfall data from 1931 to 1960 (Calversbert, 1970).

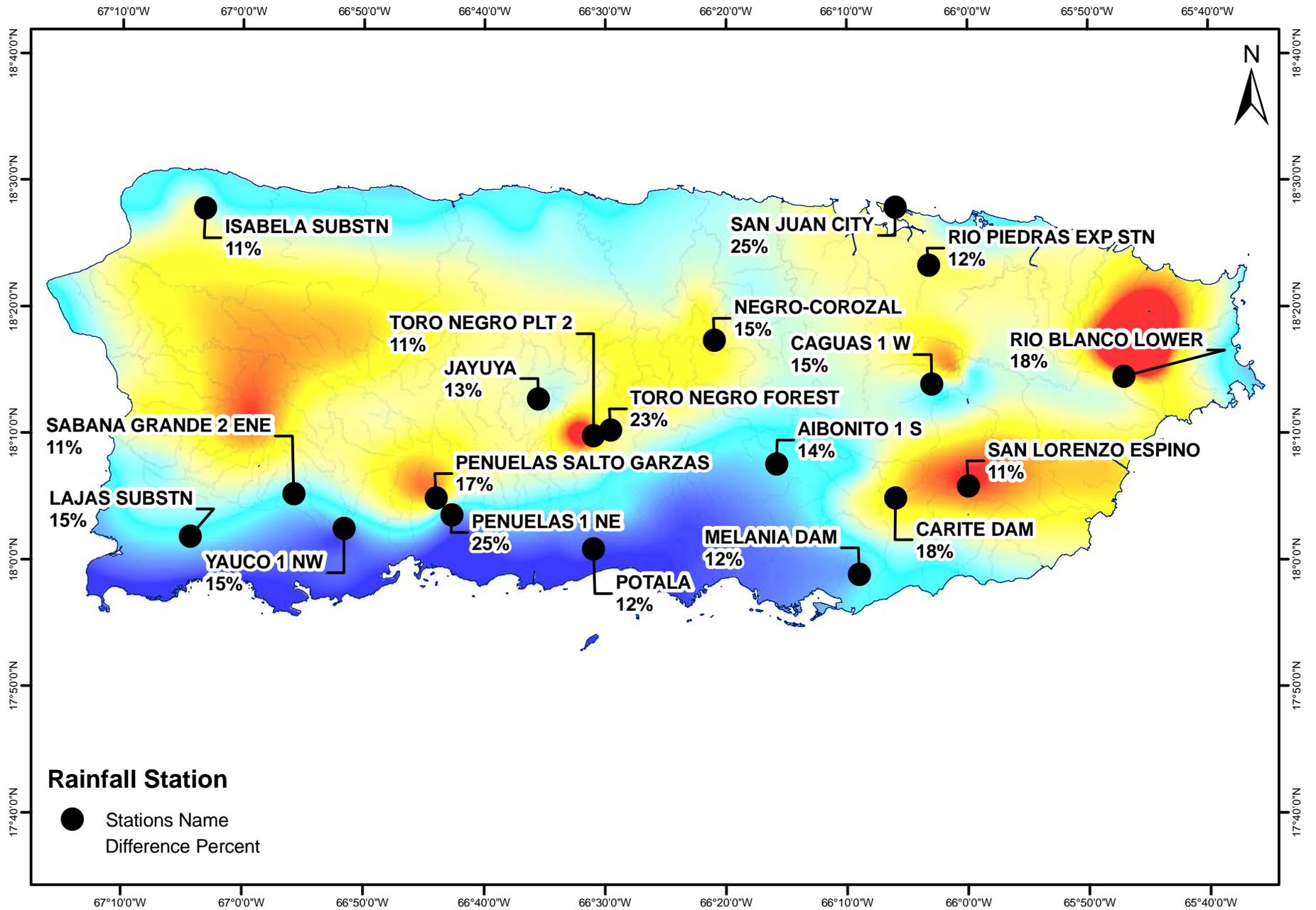


Figure 2: Rain gages with more than 10% difference in mean annual rainfall between station data and DNER (2003) rainfall map.

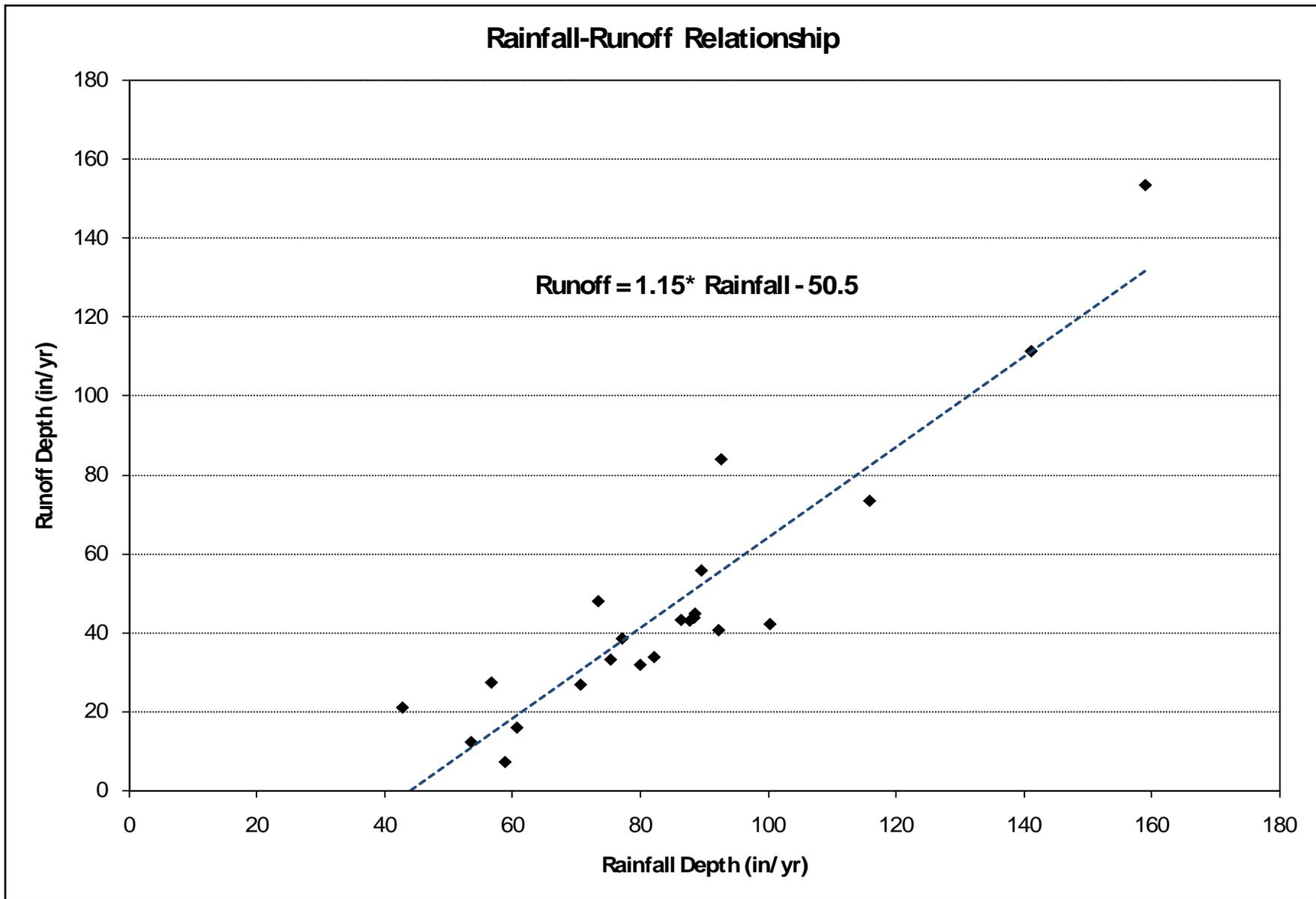
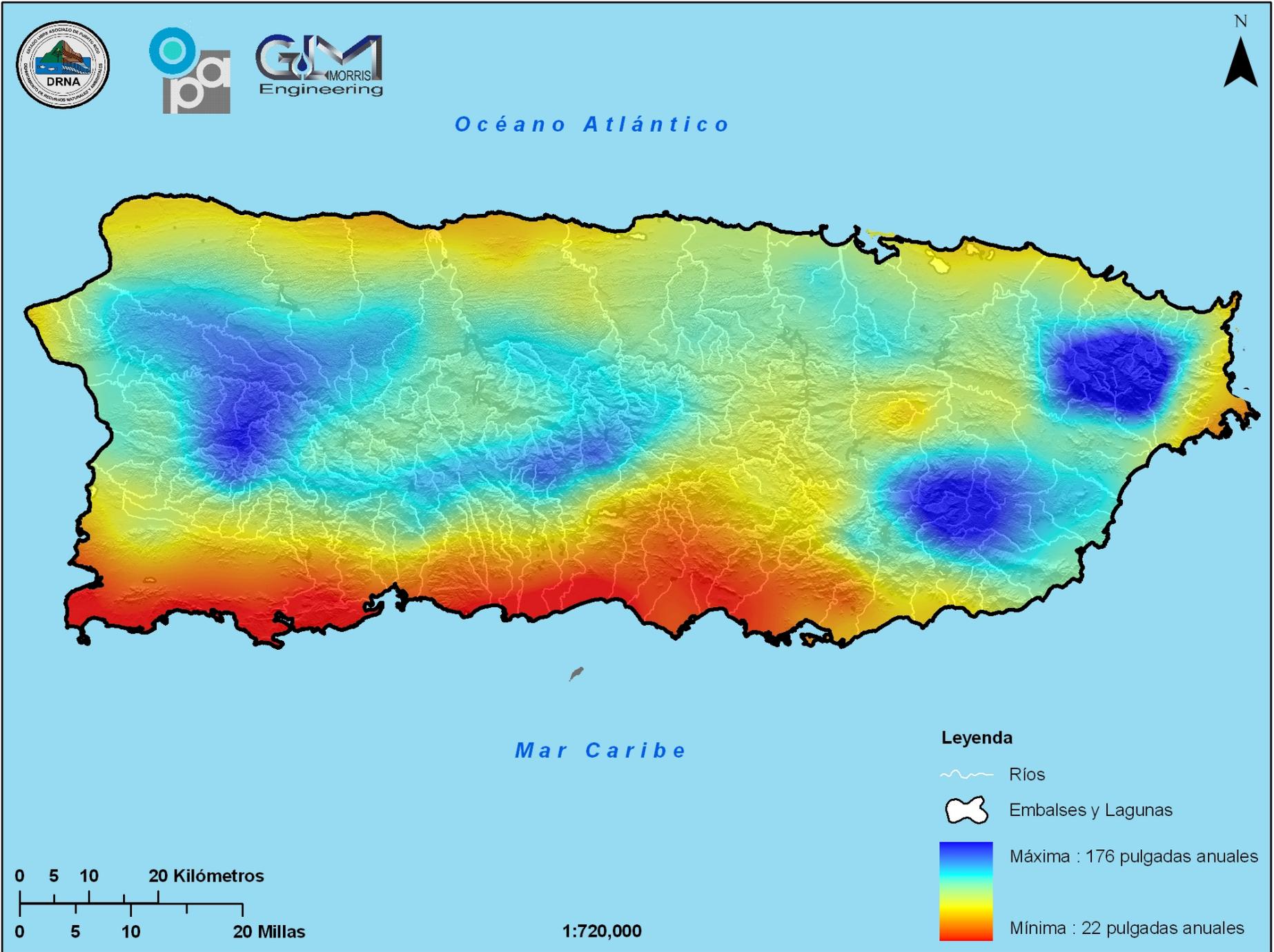
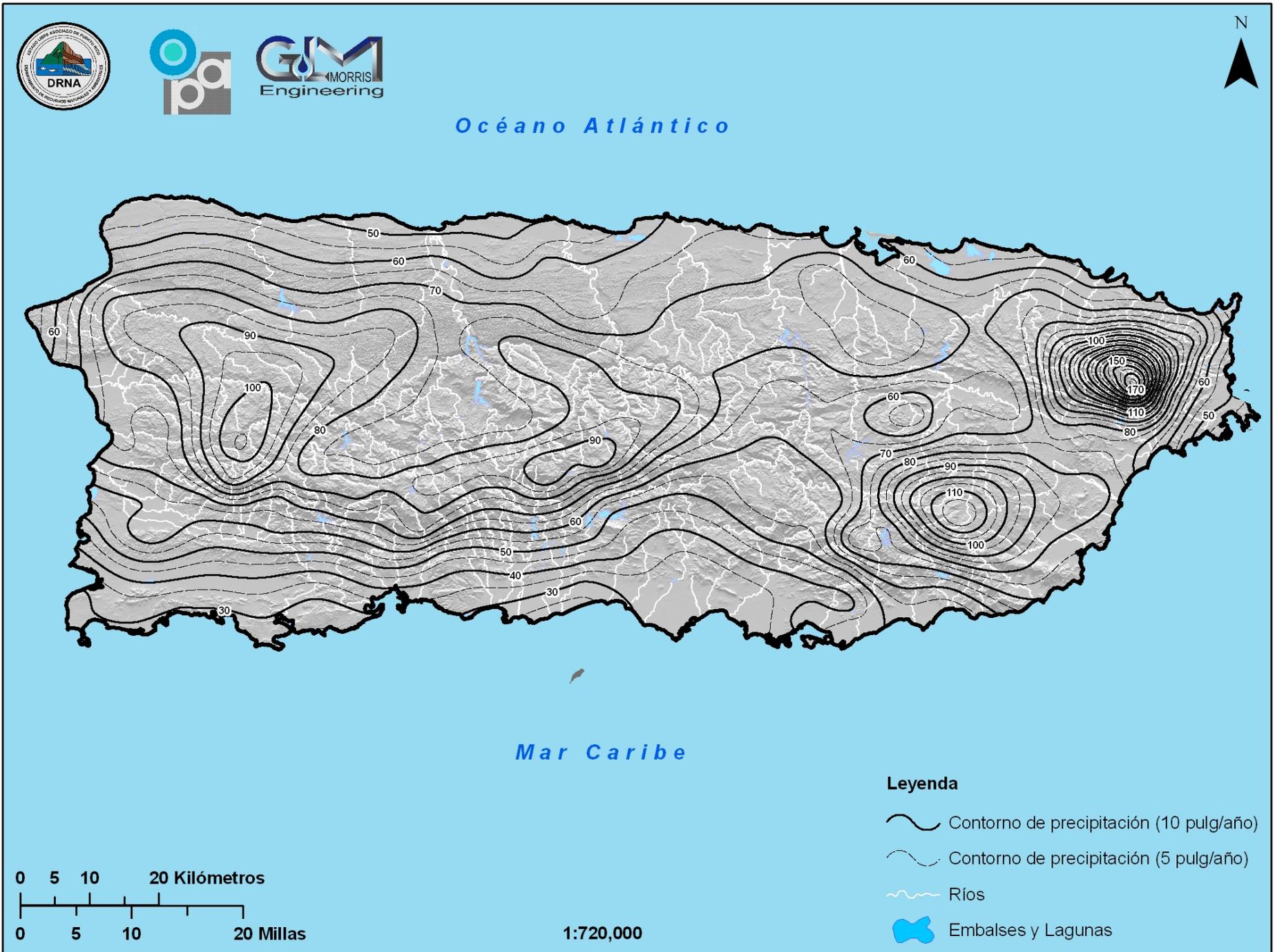


Figure 4: Rainfall-runoff relationship resulting from the final mean annual rainfall map.



Preparado por GLMorris Engineering para DRNA, 2009

Figure 5: Mean annual rainfall surface map for Puerto Rico.



Preparado por GLMorris Engineering para DRNA, 2009

Figure 6: Mean annual rainfall contours map for Puerto Rico.

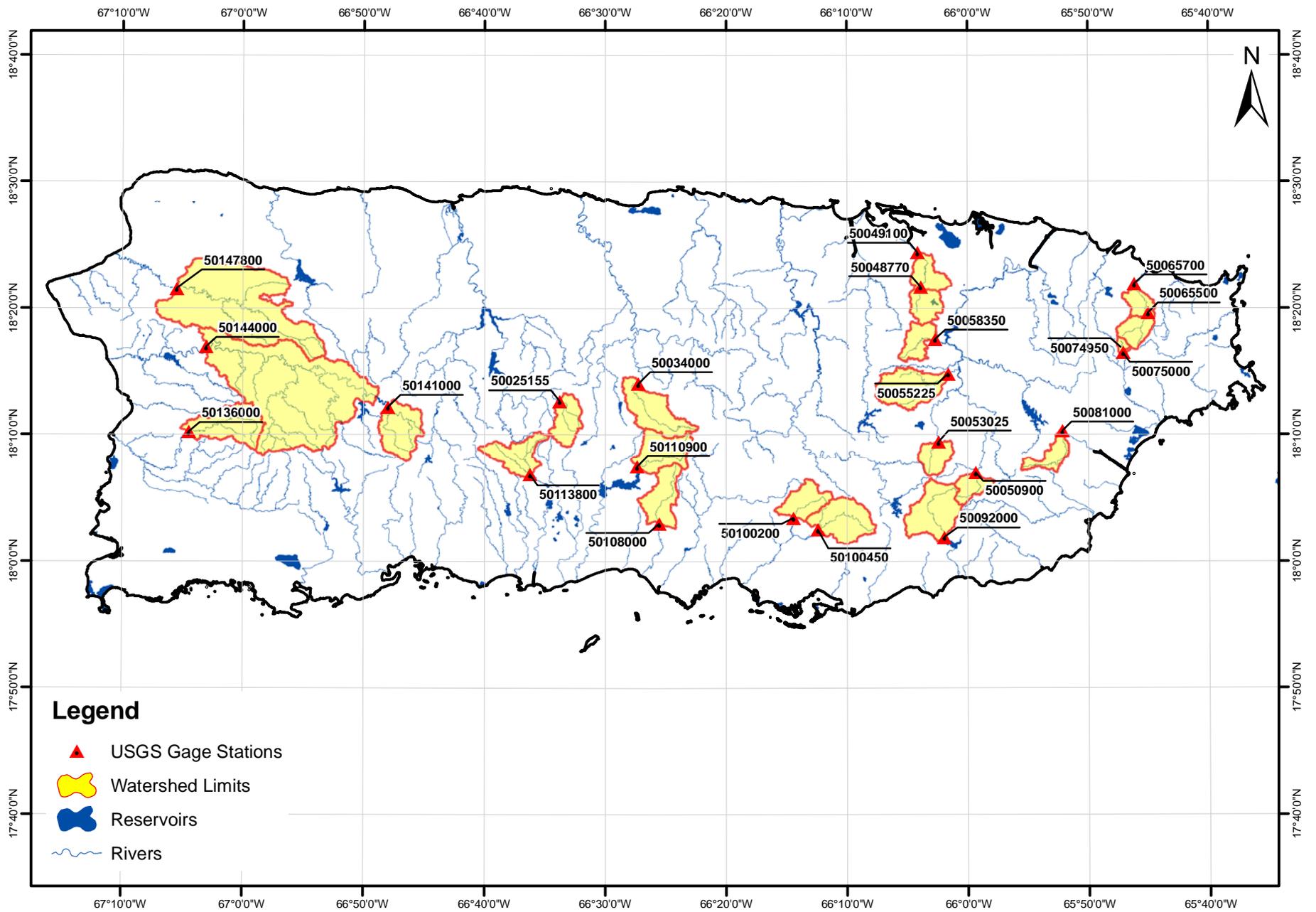


Figure 7: USGS streamflow gage stations location used in regression analysis.

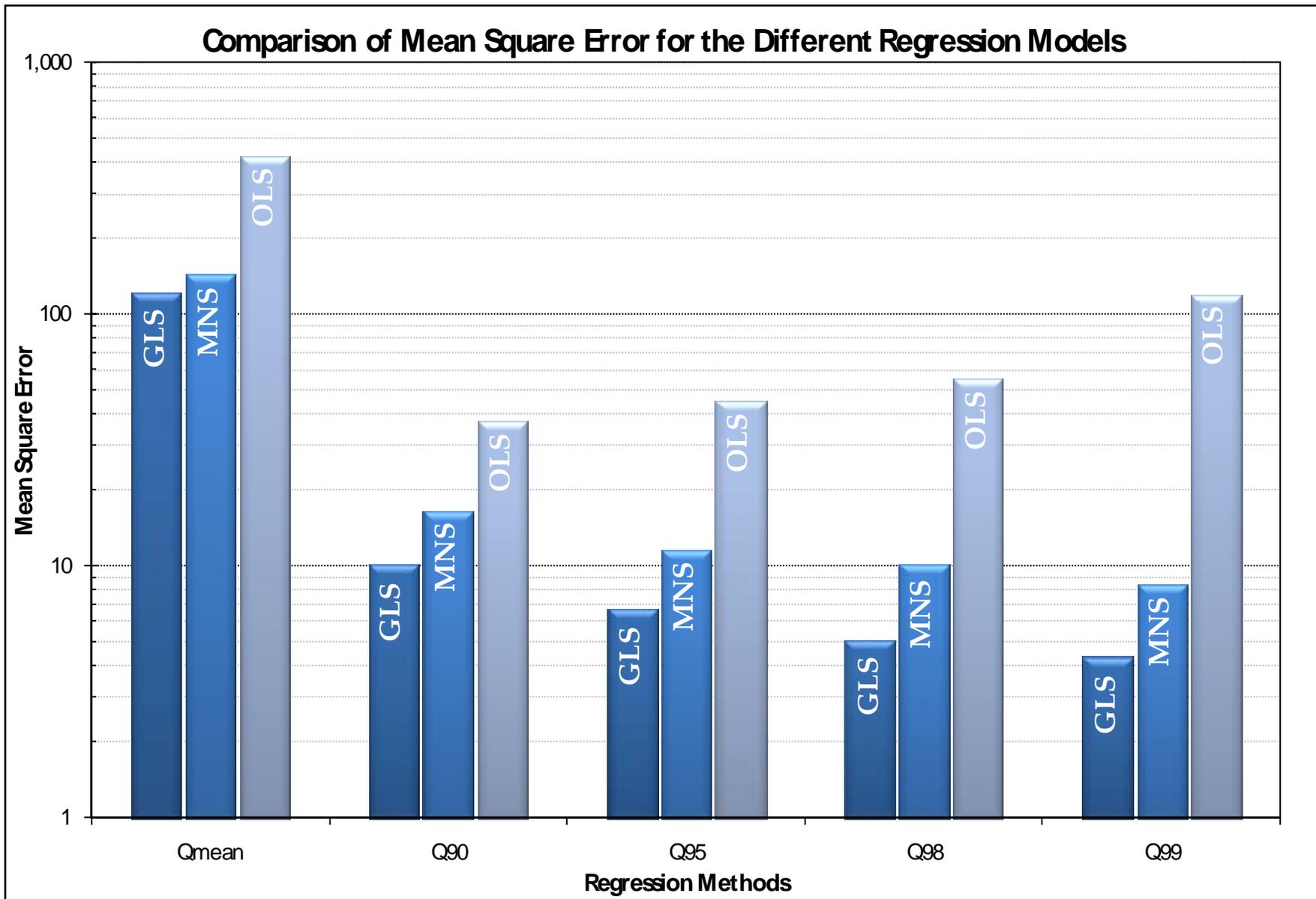


Figure 8: Comparison of Mean Square Error for the different regression models.

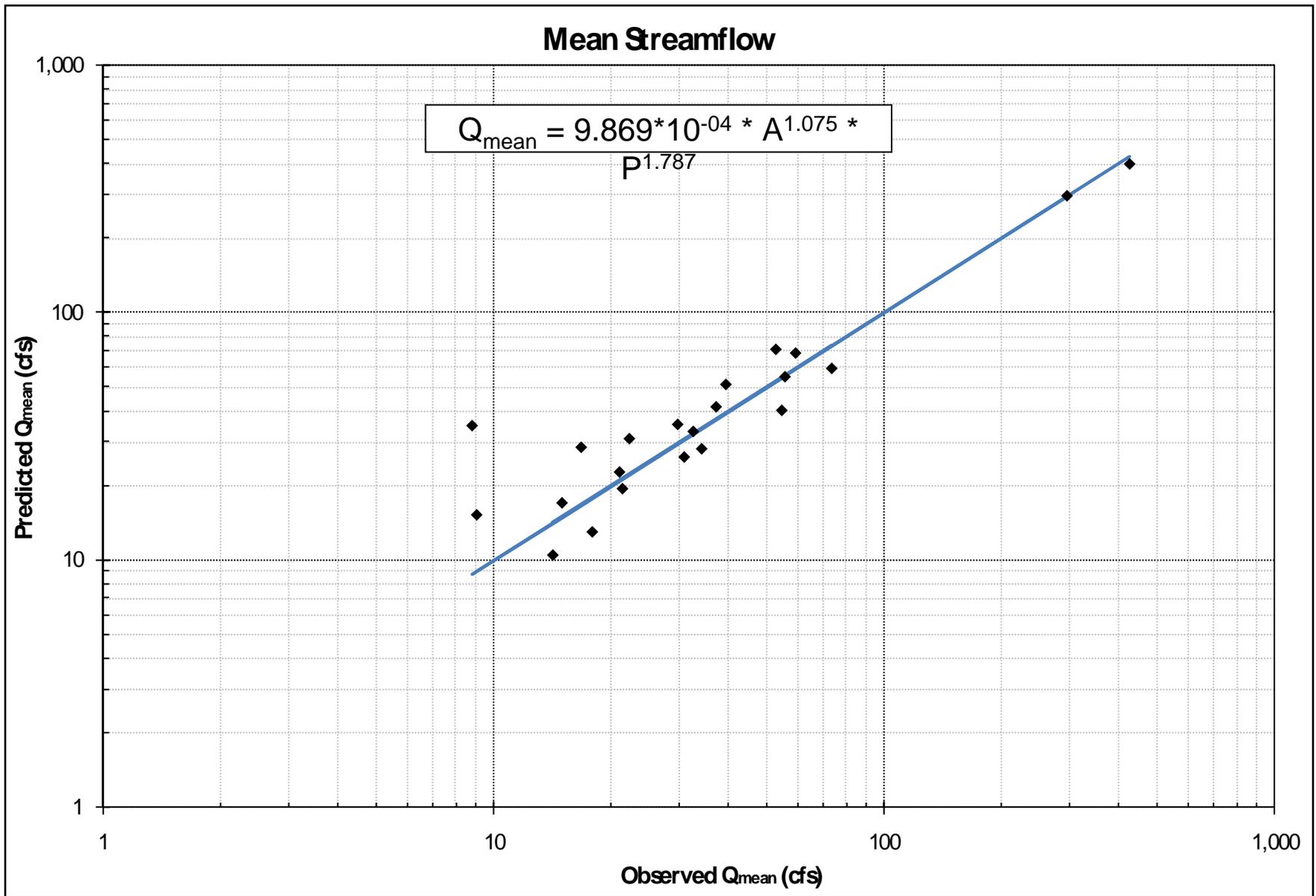


Figure 9: Relation between observed and predicted Q_{mean} flow.

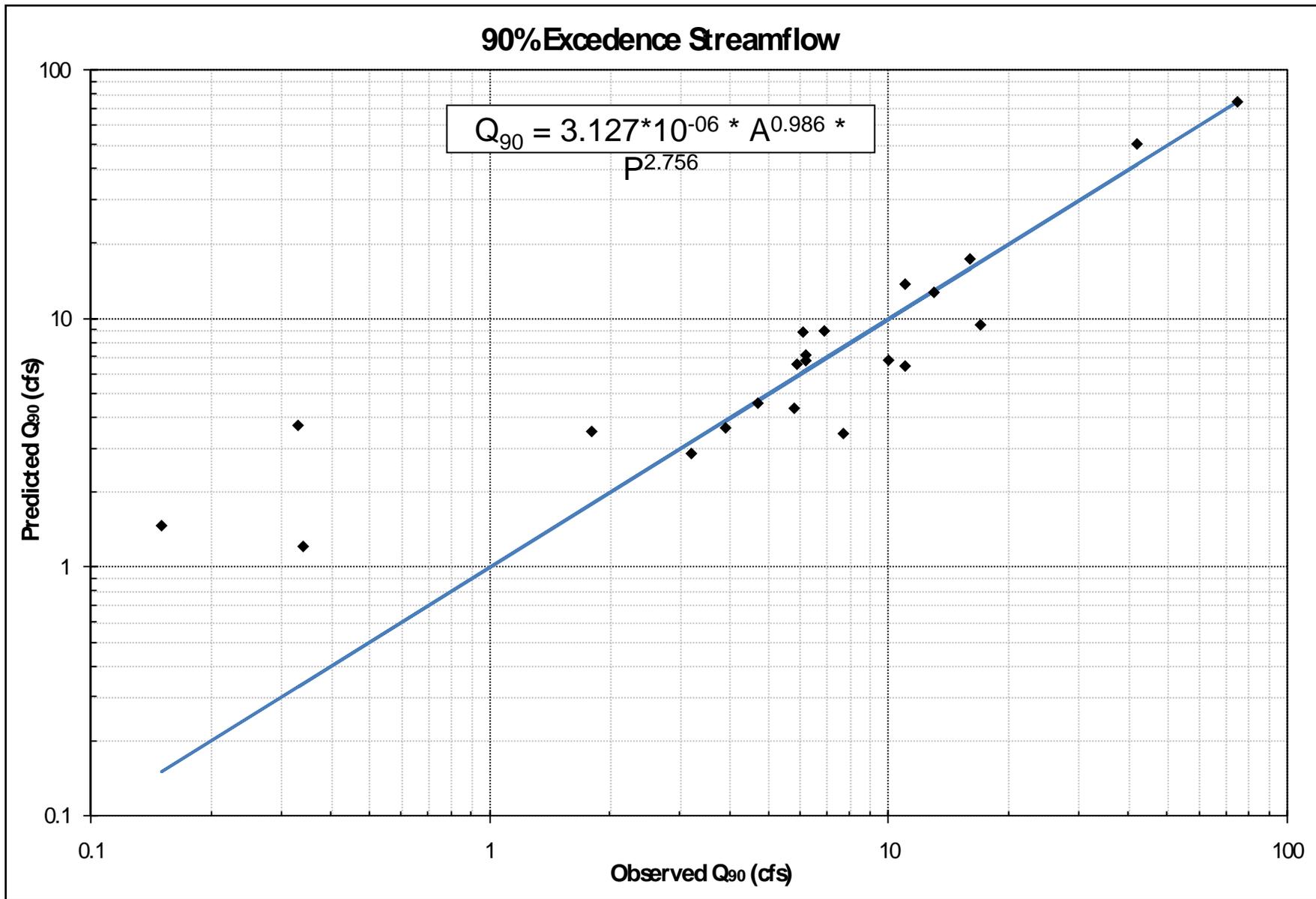


Figure 10: Relation between observed and predicted Q₉₀ flow.

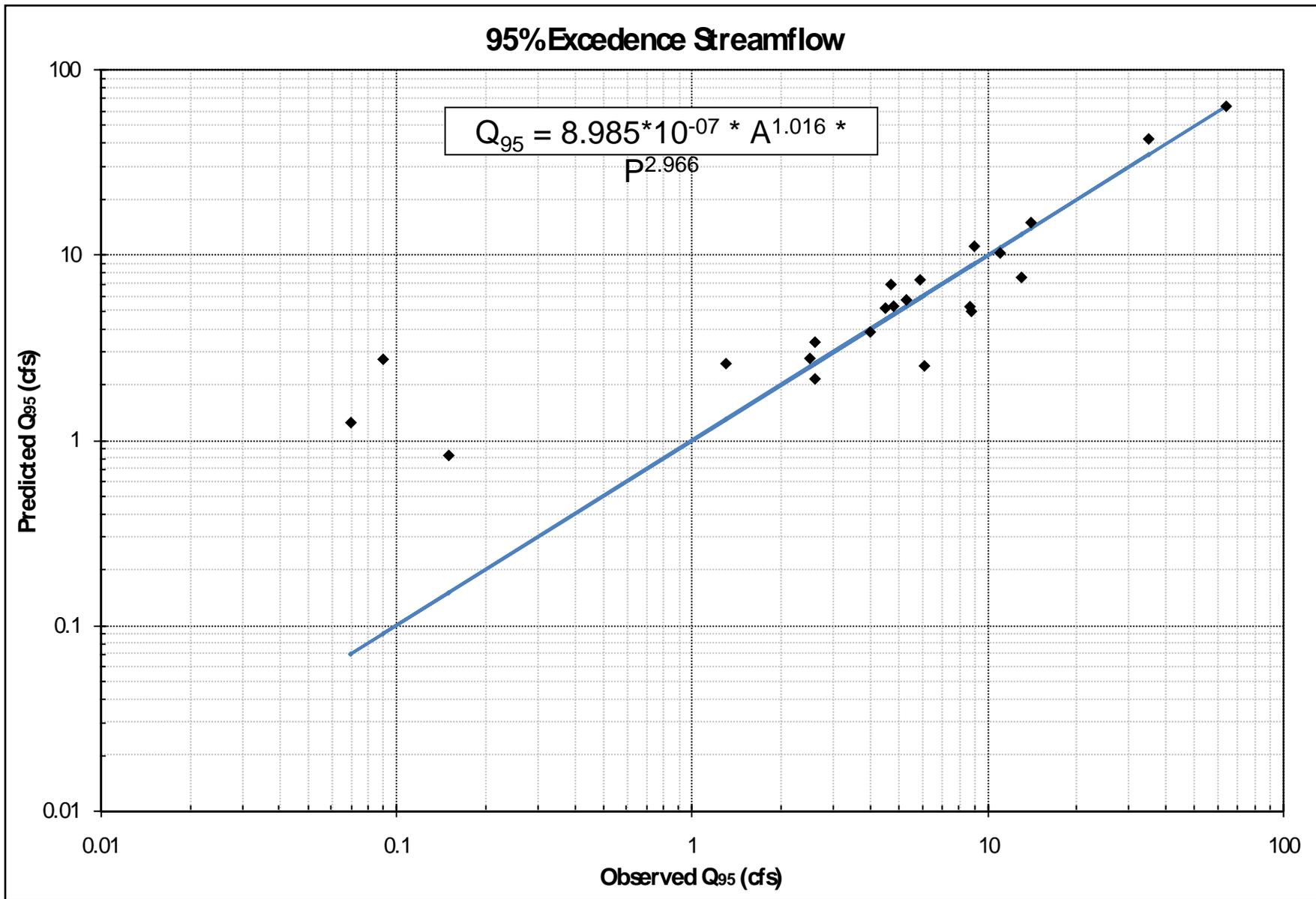


Figure 11: Relation between observed and predicted Q₉₅ flow.

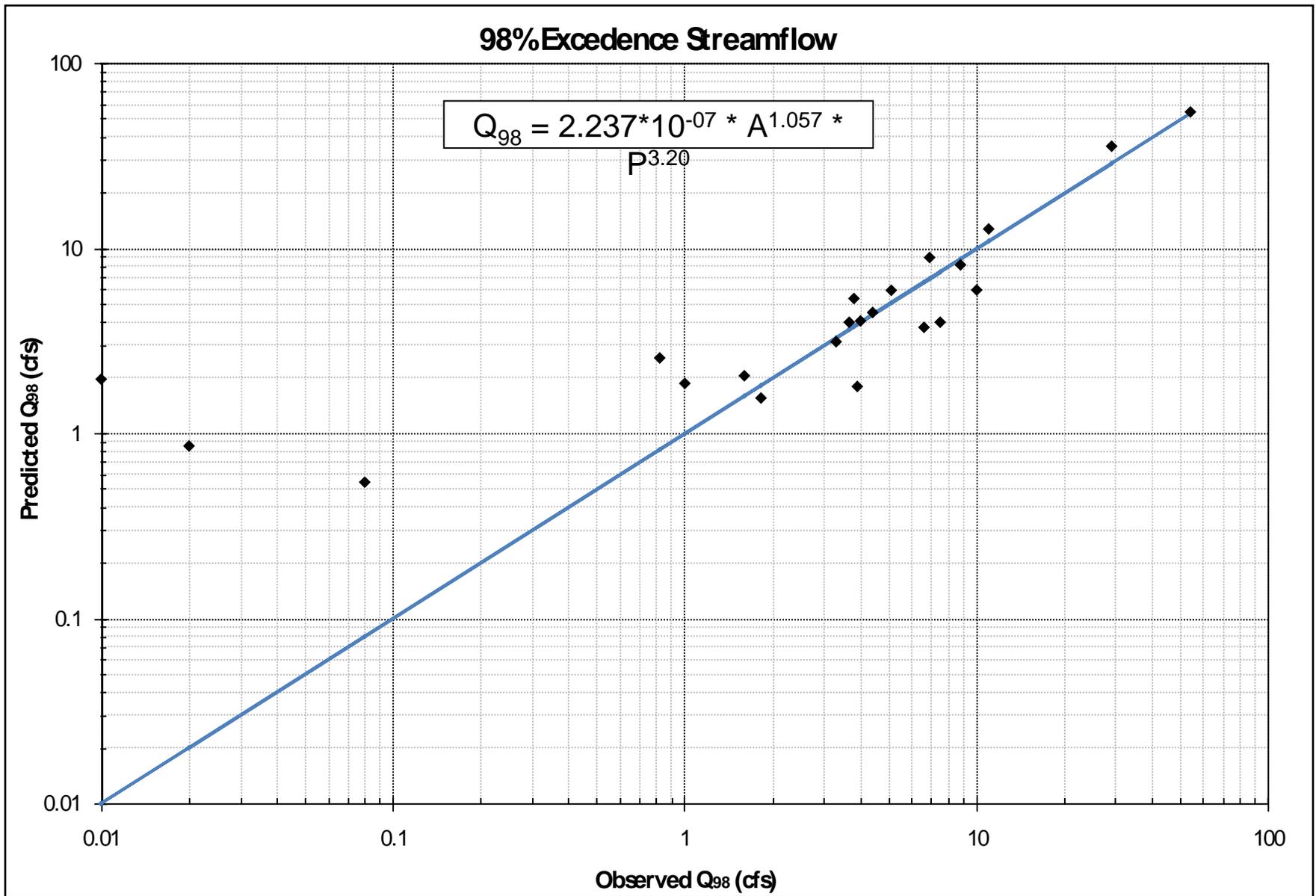


Figure 12: Relation between observed and predicted Q_{98} flow.

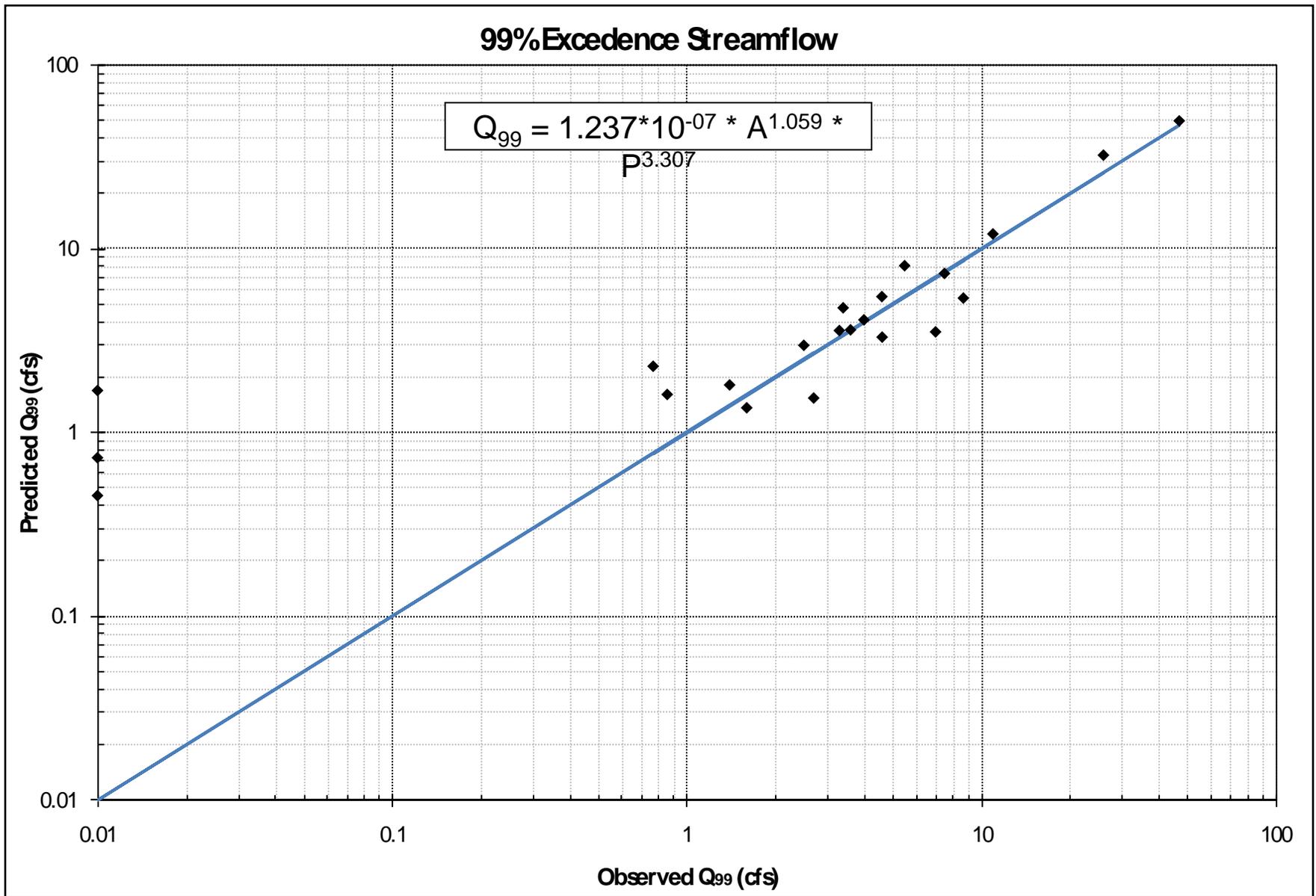


Figure 13: Relation between observed and predicted Q₉₉ flow.

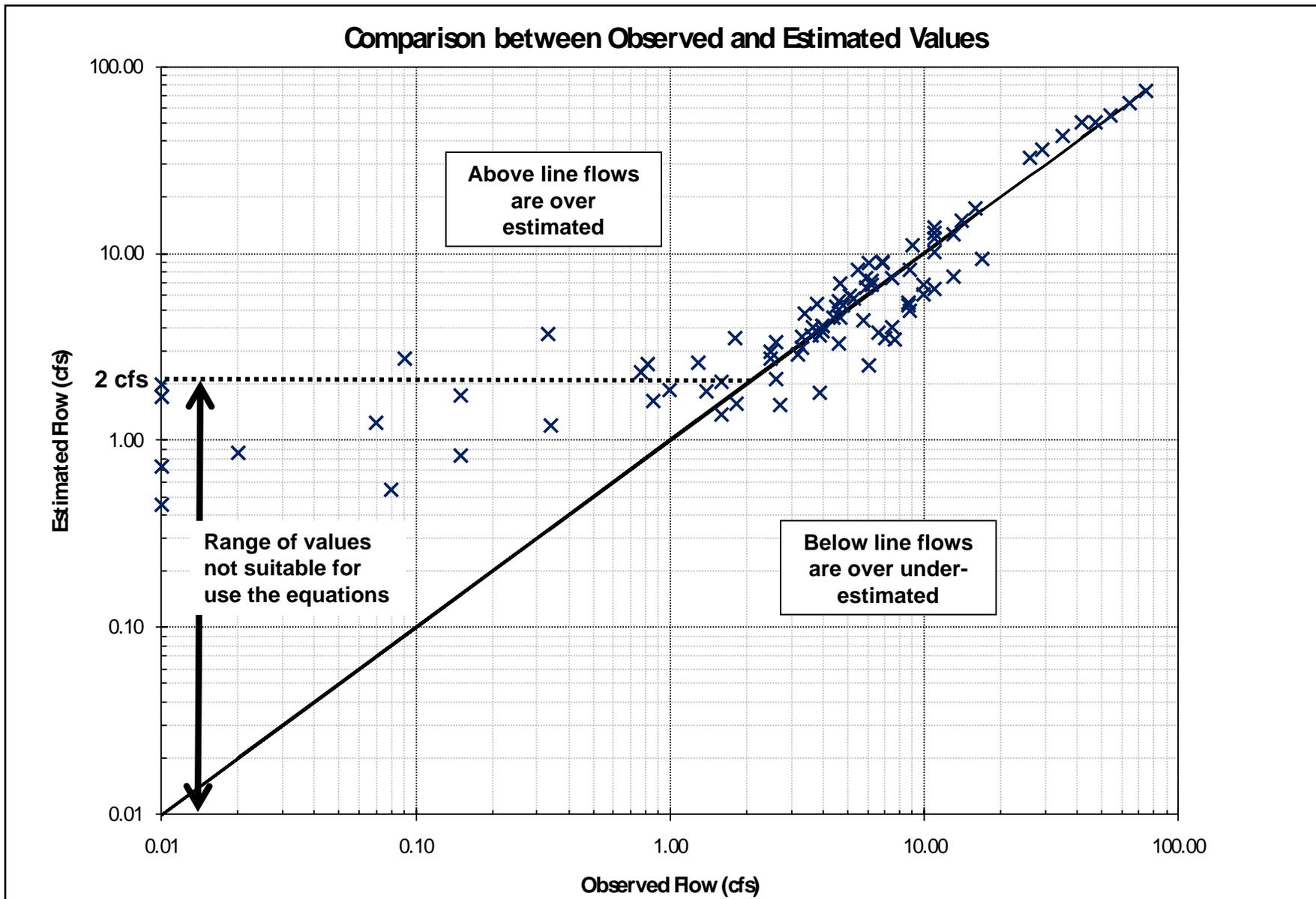


Figure 14: Relation between observed and predicted flow for all recurrence intervals compared to line of perfect correlation to gage station data.

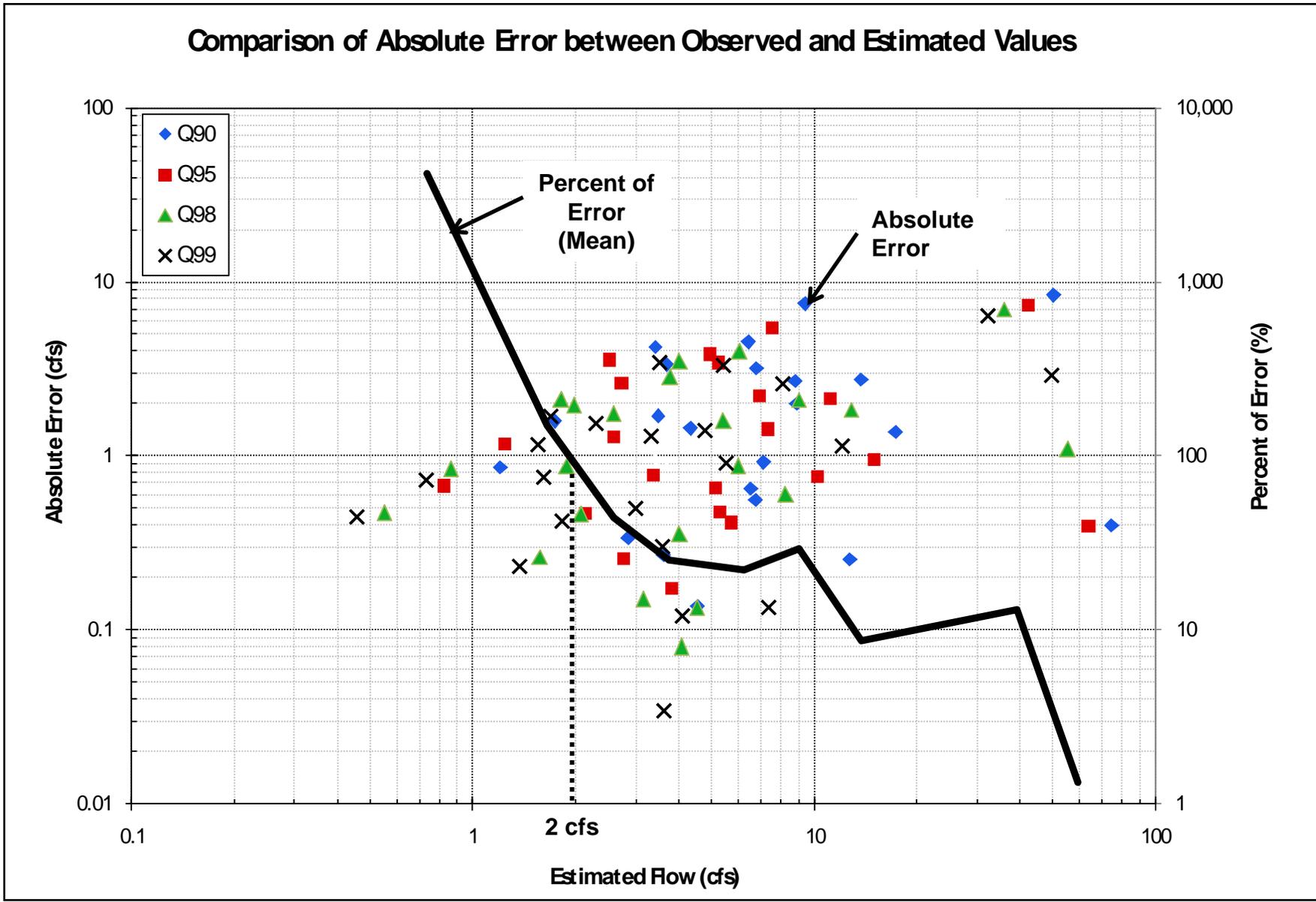


Figure 15: Error between observed and predicted values for the different return intervals.

**A
P
P
E
N
D
I
X
E
S**

APPENDIX A: SAMPLE CALCULATIONS

Example 1: Determination of Q_{99} at Río Bauta using the Station Index Method.

The station index method can be used if there is a gage station near the interest point watershed with enough data to perform an exceedance analysis (>10 years). The data from the gaged site is translated to the ungaged site by the ratio of mean flows computed by the regression equation (Eq. 3.20).

In this example, the Q_{99} of a point at Río Bauta is desired. The interest point is located downstream the USGS gage station Río Bauta Station (50034000) as presented in Figure A-1.

Information of the interest point:

Watershed area (A) = 28.29 mi²

Mean annual rainfall (P) = 80.54 in

Information of the gage site:

Watershed area (A) = 16.74 mi²

Mean annual rainfall (P) = 80.12 in

Q_{99g} obtained from the USGS Streamflow Data = 3.4 cfs

Calculate Q_{mean} for both sites using Equation (3.20):

$$Q_m = 9.869 * 10^{-04} * A^{1.075} * P^{1.787}$$

Ungaged site:

$$Q_m = 9.869 * 10^{-04} * 28.29^{1.075} * 80.54^{1.787}$$

$$Q_m = 91.38 \text{ cfs}$$

Gaged site:

$$Q_m = 9.869 * 10^{-04} * 16.74^{1.075} * 80.12^{1.787}$$

$$Q_m = 51.52 \text{ cfs}$$

Determine Q_{99} for the ungaged site with the Equation (3.26):

$$Q_{99u} = \frac{Q_{mean_u}}{Q_{mean_g}} Q_{99g}$$

$$Q_{99u} = \left(\frac{91.38}{51.52} \right) 3.4$$

$$Q_{99u} = 6.03 \text{ cfs}$$

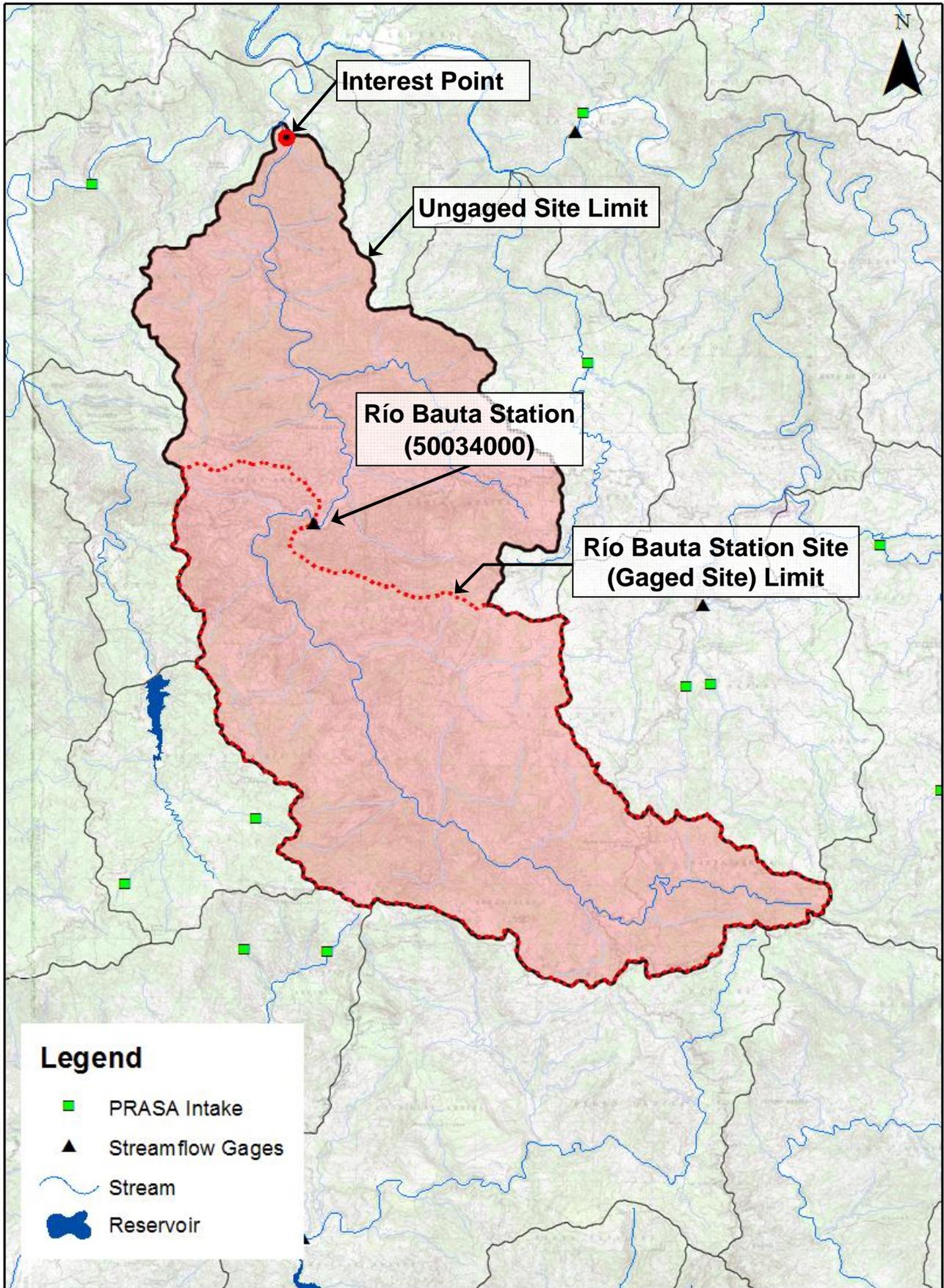


Figure A-1: Example 1. Q_{99} of an ungaged site at Río Bauta is determined using USGS Río Bauta Stations (50034000) as reference.

Example 2: Determination of Q_{99} at Río Cialitos using the Station Index Method.

As in Example 1, the interest point is near a gage station thus the Station Index method can be applied (Figure A-2). The interest point is located in Río Cialitos (Figure A-2) and the Q_{99} this point will be estimated using the streamflow data of the Río Bauta Station (50034000).

Information of the interest point:

Watershed area (A) = 15.48 mi²

Mean annual rainfall (P) = 80.16 in

Information of the gage site:

Watershed area (A) = 16.74 mi²

Mean annual rainfall (P) = 80.12 in

Q_{99} obtained from the USGS Streamflow Data = 3.4 cfs

Calculate Q_{mean} for both sites with Equation (3.20):

$$Q_m = 9.869 * 10^{-04} * A^{1.075} * P^{1.787}$$

Ungaged site:

$$Q_m = 9.869 * 10^{-04} * 15.48^{1.075} * 80.16^{1.787}$$

$$Q_m = 47.42 \text{ cfs}$$

Gaged site:

$$Q_m = 9.869 * 10^{-04} * 16.74^{1.075} * 80.12^{1.787}$$

$$Q_m = 51.52 \text{ cfs}$$

Determine Q_{99} for the ungaged site with the Equation (3.26):

$$Q_{99u} = \frac{Q_{\text{mean}_u}}{Q_{\text{mean}_g}} Q_{99g}$$

$$Q_{99u} = \left(\frac{47.42}{51.52} \right) 3.4$$

$$Q_{99u} = 3.13 \text{ cfs}$$

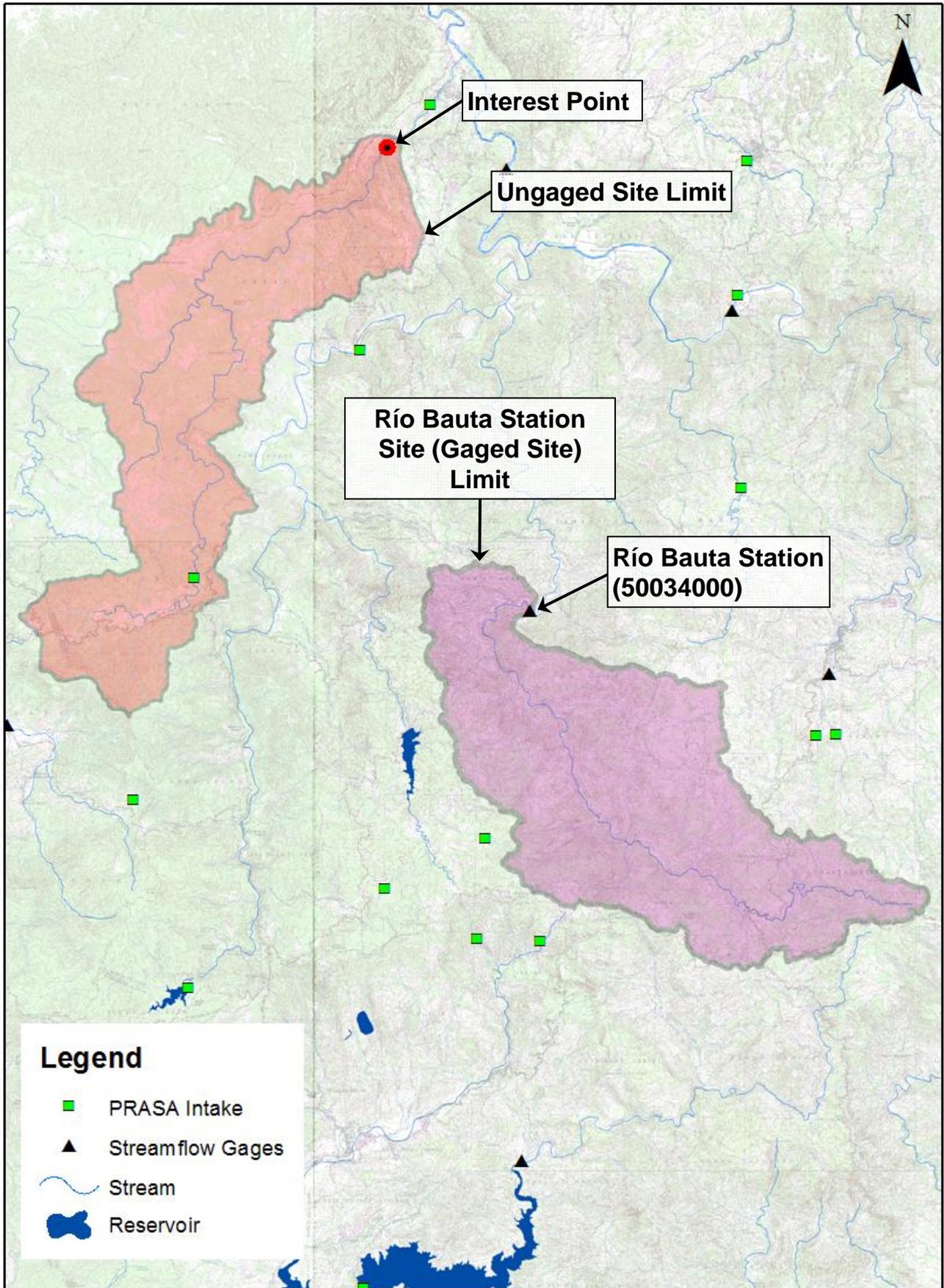


Figure A-2: Example 2. Q_{99} of an ungaged site with an intake at Río Cialitos is determined using USGS Río Bauta Stations (50034000) as reference.

Example 3: Determination of Q_{99} at Río Bauta using the Regression Equation.

If there is no streamflow station near the interest point, the minimum flows can be estimated using Equation (3.24). For this example, data from Example 1 will be used for comparison purposes.

Example 1 information of the site:

Watershed area (A) = 28.29 mi²

Mean annual rainfall (P) = 80.54 in

Determine Q_{99} using Equation (3.24):

$$Q_{99} = 1.237 * 10^{-07} * A^{1.059} * P^{3.307}$$

$$Q_{99} = 1.237 * 10^{-07} * 28.29^{1.059} * 80.54^{3.307}$$

$$\mathbf{Q_{99} = 8.57 cfs}$$