

**MODELO PARA ESTIMACIÓN DE CAUDALES MÁXIMOS A PARTIR DE
PARÁMETROS MORFOMÉTRICOS E HIDROLÓGICOS EN CUENCAS
HIDROGRÁFICAS DEL DEPARTAMENTO DEL CAUCA, COLOMBIA**

**MODEL FOR ESTIMATING MAXIMUM RUNOFFS BASED ON
MORPHOMETRIC AND HYDROLOGICAL PARAMETERS IN WATERSHEDS
OF THE CAUCA DEPARTMENT, COLOMBIA**

**TESIS DE MAESTRÍA PRESENTADA PARA OPTAR AL TÍTULO DE
MAGÍSTER EN INGENIERÍA HIDRÁULICA E HIDROLÓGICA**

Oscar Felipe Agredo Campuzano



Director

Ing. Mg. Alberto José Caldas Constáin

**UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERÍA CIVIL
PROGRAMA: MAESTRÍA EN INGENIERÍA HIDRÁULICA E HIDROLÓGICA
POPAYÁN-CAUCA
2024**

Nota de aceptación

El director y jurados han evaluado este documento titulado: “MODELO PARA ESTIMACIÓN DE CAUDALES MÁXIMOS A PARTIR DE PARÁMETROS MORFOMÉTRICOS E HIDROLÓGICOS EN CUENCAS HIDROGRÁFICAS DEL DEPARTAMENTO DEL CAUCA, COLOMBIA”, escuchando la sustentación de este por su autor y la encuentran satisfactoria, por lo cual autorizan al estudiante, Oscar Felipe Agredo Campuzano para optar al título de Magister en Ingeniería Hidráulica e Hidrológica

Ing. Mg. Alberto José Caldas Constaín

Ing. Msc. Carlos Armando Gallardo Barrera

Ing. Mg. Helmer Edgardo Monroy González

*A Margarita y Lucia por cuidarme,
a María y Roque por enseñarme fortaleza
y a los tres seres de mi corazón, Homer J., Candanga H. y Oslo F.*

*To Margarita and Lucia for taking care of me,
to María and Roque for teaching me strength,
and to the three beings of my heart, Homer J., Candanga H., and Oslo F.*

Agradecimientos

Al Semillero de Investigación HydroInnovación (S.H.I.) y al Grupo de Investigación en Hidráulica e Hidrología (G.I.H.H.), donde surgen y se consolidan las ideas respectivamente, y con quienes se desarrollaron los proyectos “Actualización de curvas IDF del Departamento del Cauca en una herramienta de visualización web” y “Estudio y caracterización del C y CN para el Departamento del Cauca: Correlación e implementación” como productos asociados a la presente investigación.

Al Departamento de Hidráulica de la Facultad de Ingeniería Civil, por brindarme la oportunidad de trabajar en lo que amo, permitiéndome aprender cada día mientras enseño y en especial al profesor Carlos Gallardo, profesional que llevan impulsando, mejorando y consolidando la Ingeniería Hidráulica e Hidrológica en la región por más de 30 años, colega, mentor y amigo.

Al Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM, en especial al profesional William González y al Instituto Geográfico Agustín Codazzi – IGAC, por el suministro de información.

Para referirse a la tesis de maestría:

Cita: Agredo-Campuzano, 2024

Referencia: Agredo-Campuzano, O., F. (2024). Modelo para estimación de caudales máximos a partir de parámetros morfométricos e hidrológicos en cuencas hidrográficas del Departamento del Cauca, Colombia. Tesis de maestría, Universidad del Cauca, Facultad de Ingeniería Civil, Maestría en Ingeniería Hidráulica e Hidrológica, Popayán, Cauca, Colombia.

Resumen

Los modelos lluvia escorrentía son fundamentales para el diseño de obras hidráulicas, permitiendo estimar los caudales máximos para diferentes periodos de retorno, aunque desarrollados en zona diferentes del mundo de donde son aplicados. En la actualidad existen modelos complejos que tienen en cuenta el cambio y la variabilidad climática manejando hipótesis que permiten establecer predicciones sobre los eventos extremos, sin embargo, requieren de una gran cantidad de datos, que en muchas ocasiones son difícil de obtener, ya sea por disponibilidad, costos e incluso conocimiento especializado, dejando a un lado modelos más simples que relacionan variables regionales y que pueden generar predicciones igual de robustas, con información mucho más sencilla de conseguir. En el Cauca se tiene una densidad de aproximadamente 800 km^2 por estación de medición de caudal, lo cual se considera muy bajo, teniendo en cuenta la riqueza hídrica con la que dispone el departamento, en donde se encuentra ubicado el Macizo Colombiano, lugar de nacimiento de los principales ríos del país, evidenciándose la necesidad de explorar y determinar métodos para estimar caudales máximos en cuencas no instrumentadas con la información disponible de cuencas vecinas que dispongan de datos, y que sean similares hidrológicamente.

Se analizaron parámetros morfométricos, y variables hidrológicas de 41 cuencas hidrográficas instrumentadas en el Cauca (41 estaciones de medición de caudal y 86 estaciones de medición de precipitación) equivalente aproximadamente al 53% del departamento. Se empleo software SIG para obtener las variables fisiográficas de las cuencas y realizar la caracterización morfométrica de cada una de ellas, empleando modelos digitales de elevación (MED) de libre acceso, y mapas de suelos y coberturas suministrados por entidades gubernamentales nacionales. Se utilizó también el lenguaje de programación R para el análisis de datos, como el cálculo de outliers en los datos hidrológicos, el cual sirvió, aplicando distribuciones de probabilidad, para confeccionar matrices de datos de 38 cuencas (Tres tuvieron que ser retiradas del análisis por inconsistencia de datos de caudal) para periodos de retorno (Tr) de 2.33, 25, 50, 100, 200 y 500 años, teniendo como principales variables hidrológicas, la precipitación y el caudal para cada Tr. Buscando regresiones lineales múltiples (RLM) para los caudales máximos por cada Tr, estas matrices de datos fueron sometidas a un extenso análisis estadístico para dos escenarios, el primero fue para la totalidad de las cuencas evaluando todas las 2047 combinaciones posibles de variables independientes (área de la cuenca, pendiente media de la cuenca, tiempo de concentración, factor de forma, índice de compacidad, índice de alargamiento, índice de asimetría, razón de elongación, número de curva de escorrentía CN, promedio de precipitación máxima en 24 horas multianual y precipitación para el Tr analizado) y seleccionando el mejor modelo por Tr, aquel con menor Criterio de Información de Akaike (AIC), el cual evalúa la calidad de un modelo basado en su ajuste y complejidad, penalizando los modelos con demasiados parámetros para evitar el sobreajuste, y que a su vez cumplieran con los requisitos básicos de RLM (linealidad, independencia, no colinealidad, homoscedasticidad y normalidad). En el

segundo escenario se empleó la distancia de Cook para identificar cuencas con gran influencia en el ajuste del modelo, generando nuevas matrices de datos, esta vez con 31 cuencas, las cuales fueron sometidas al mismo análisis del primer escenario, es decir, mejor combinación de variables independientes posibles por cada Tr, teniendo en cuenta el AIC y que cumplan con los requisitos básicos de RLM. Se obtuvo el error promedio para el mejor modelo por Tr de cada escenario para evaluar su rendimiento mediante validación cruzada LOOCV (Leave-One-Out Cross-Validation), obteniendo métricas como R^2 y KGE (Kling-Gupta Efficiency), los cuales son ampliamente utilizados para evaluar la eficiencia y precisión de modelos hidrológicos, combinados con técnicas de aprendizaje de máquina.

Los resultados del primer escenario arrojaron que únicamente cumplen requisitos básicos de RLM las matrices de datos para 200 y 500 años de Tr, arrojando valores de R^2 y de KGE respectivamente de 0.693 y 0.843, y 0.681 y 0.838, indicando modelos bastante sólidos y efectivos para explicar la variabilidad de los datos y simularlos con buena precisión, mas no perfectos. Los resultados del segundo escenario mejoraron considerablemente al no estar con los outliers, disminuyendo el porcentaje de estudio del departamento a un valor de 28%. En este caso todos los Tr tuvieron correlaciones que cumplían requisitos básicos de RLM, arrojando valores de R^2 y de KGE respectivamente de 0.837 y 0.866, 0.869 y 0.907, 0.866 y 0.909, 0.853 y 0.904, 0.828 y 0.885 y 0.783 y 0.858, indicando que los modelos tuvieron una mejora importante respecto al primer escenario, y que son consistentes y efectivos para simular los datos observados, a excepción del último Tr, el cual es el modelo con el ajuste más bajo, sin embargo, cuenta con una eficiencia aceptable. Para el mejor modelo posible de cada Tr de ambos escenarios se obtuvieron también gráficas que permitieron mejorar el análisis de cada resultado, como por ejemplo valores observados vs. valores calculados, residuales vs. predicciones, histograma de residuales, QQ-Plot de residuales y distribución de errores absolutos, las cuales facilitaron la realización de un análisis más detallado.

Con esta investigación se logró desarrollar un modelo con relaciones matemáticas para una zona específica del Departamento del Cauca dentro de la zona intercordillera que permite estimar caudales máximos en cuencas no instrumentadas, y se invita a evaluar la posibilidad de adoptar el método para un estudio que abarque todo el territorio nacional zonificado. También se apertura un camino hacia otras investigaciones con las matrices de datos de la presente investigación en cuanto a regresiones no lineales, regresiones robustas, así como optimización con Algoritmos Genéticos, Recocido Simulado o Algoritmos de Selección de Subconjuntos, y técnicas de Clúster y Agrupamiento, que permitan mejorar o validar los resultados hallados y que reduzcan la incertidumbre que conlleva la aplicación de modelos lluvia escurrentía bajo información de entrada supuesta, escasa o inexistente en muchos casos.

Palabras clave: Aprendizaje de máquina, Caudales máximos, cuencas hidrográficas, parámetros morfométricos, regresión lineal múltiple, SIG, validación cruzada

Abstract

Rainfall-runoff models are essential for the design of hydraulic structures, allowing for the estimation of peak flows for different return periods, even when they were developed in regions different from where they are applied. Nowadays, there are complex models that consider climate change and variability, employing hypotheses that allow predictions of extreme events. However, these models require a large amount of data, which is often difficult to obtain due to availability, cost, or the need for specialized knowledge. Therefore, simpler models that relate regional variables that can produce equally robust predictions with much easier-to-obtain information are often overlooked. In the Cauca region, there is a density of approximately 800 km^2 per flow measurement station, which is considered low, given the region's abundant water resources, especially as it includes the Macizo Colombiano, the birthplace of the main rivers of the country. This highlights the need to explore and determine methods for estimate peak flows in ungauged basins using available information from hydrologically similar neighboring basins.

Morphometric parameters and hydrological variables were analyzed for 41 gauged watersheds in Cauca (41 flow measurement stations and 86 precipitation measurement stations), covering approximately 53% of the department. GIS software was used to obtain the physiographic variables of the basins and to perform the morphometric characterization of each one, using freely available digital elevation models (DEM) and soil and land cover maps provided by national government agencies. The R programming language was also used for data analysis, such as calculating outliers in the hydrological data. This helped, by applying probability distributions, to construct data matrices for 38 basins (three were excluded due to inconsistent flow data) for return periods (Tr) of 2.33, 25, 50, 100, 200, and 500 years, with precipitation and flow as the main hydrological variables for each Tr. Seeking multiple linear regressions (MLR) for peak flows for each Tr, these data matrices underwent extensive statistical analysis for two scenarios. The first scenario included all the basins, evaluating all 2047 possible combinations of independent variables (basin area, average basin slope, time of concentration, shape factor, compactness index, elongation index, asymmetry index, elongation ratio, runoff curve number CN, average maximum 24-hour precipitation, and precipitation for the analyzed Tr), selecting the best model for each Tr based on the lowest Akaike Information Criterion (AIC), which evaluates the quality of a model based on its fit and complexity, penalizing models with too many parameters to avoid overfitting, while also meeting the basic MLR assumptions (linearity, independence, no multicollinearity, homoscedasticity, and normality). In the second scenario, Cook's distance was used to identify basins with high influence on the model fit, generating new data matrices with 31 basins. These were subjected to the same analysis as the first scenario, i.e., identifying the best combination of possible independent variables for each Tr, considering the AIC and meeting the basic MLR assumptions. The average error for the best model for

each Tr in each scenario was obtained to assess performance through LOOCV (Leave-One-Out Cross-Validation), obtaining metrics such as R^2 and KGE (Kling-Gupta Efficiency), which are widely used to evaluate the efficiency and accuracy of hydrological models, combined with machine learning techniques.

The results of the first scenario showed that only the data matrices for 200- and 500-Tr met the basic MLR assumptions, yielding R^2 and KGE values of 0.693 and 0.843, and 0.681 and 0.838, respectively, indicating solid and effective models for explaining data variability and simulating them with good accuracy, though not perfectly. The results of the second scenario showed a significant improvement after removing the outliers, reducing the study area to 28% of the department. In this case, all Tr had correlations that met the basic MLR assumptions, with R^2 and KGE values of 0.837 and 0.866, 0.869 and 0.907, 0.866 and 0.909, 0.853 and 0.904, 0.828 and 0.885, and 0.783 and 0.858, respectively. These results indicate that the models improved significantly compared to the first scenario and are consistent and effective for simulating observed data, except for the last Tr, which had the lowest fit but still an acceptable efficiency. For the best possible model for each Tr in both scenarios, graphs were also produced to enhance the analysis of each result, such as observed vs. calculated values, residuals vs. predictions, residual histogram, residual QQ-Plot, and absolute error distribution, facilitating a more detailed analysis.

This research successfully developed a mathematical model for a specific region in the Department of Cauca within the inter-Andean zone that allows for the estimation of peak flows in ungauged basins. It is suggested that the method be considered for a study encompassing the entire national territory, divided into zones. This research also opens a path for further studies using the data matrices from this work, exploring nonlinear regressions, robust regressions, and optimization with Genetic Algorithms, Simulated Annealing, or Subset Selection Algorithms, as well as Clustering and Grouping techniques which could improve or validate the results of this research and reduce the uncertainty associated with the application of rainfall-runoff models with assumed, scarce, or nonexistent input data in many cases.

Keywords: Machine learning, maximum runoffs, watersheds, morphometric parameters, multiple linear regression, GIS, cross-validation

Abreviaturas

A: Área de drenaje, (km^2)	P50yr: Precipitación para un periodo de retorno de 50 años, (mm)
AIC: Criterio de Información de Akaike	P100yr: Precipitación para un periodo de retorno de 100 años, (mm)
Amay: Área mayor, (km^2)	P200yr: Precipitación para un periodo de retorno de 200 años, (mm)
Amen: Área menor, (km^2)	P500yr: Precipitación para un periodo de retorno de 500 años, (mm)
C: Coeficiente de escorrentía	Pcuenc: Pendiente de la cuenca, %
CN: Número de curva de escorrentía	Pcauc: Pendiente del cauce principal, %
CLC: Corine Land Cover	PG: PM: Estación Pluviográficas
CO: Estación Climatológica Ordinaria	PM: Estación Pluviométrica
CP: Estación Climatológica Principal	Pmáx24h: Precipitación máxima en 24 horas, (mm)
Cook: Distancia de Cook	Q2.33yr: Caudal para un periodo de retorno de 2.33 años, (m^3/s)
Ct: Índice de torrencialidad	Q25yr: Caudal para un periodo de retorno de 25 años, (m^3/s)
Dd: Densidad de drenaje	Q50yr: Caudal para un periodo de retorno de 50 años, (m^3/s)
DEM: Modelo de elevación digital	Q100yr: Caudal para un periodo de retorno de 100 años, (m^3/s)
De: Diámetro equivalente	Q200yr: Caudal para un periodo de retorno de 200 años, (m^3/s)
Ff: Factor de forma	Q500yr: Caudal para un periodo de retorno de 500 años, (m^3/s)
Ia: Índice de alargamiento	Re: Razón de elongación
Ias: Índice asimétrico	Tc: Tiempo de concentración, <i>horas</i>
Ic: Índice de compacidad	SP: Estación Sinóptica Principal
Lcp: Longitud del Cauce Principal	
LG: Estación Limnigráfica	
LM: Estación Limnimétrica	
LOOCV: Leave-One-Out Cross-Validation	
P: Precipitación, (mm)	
P2.33yr: Precipitación para un periodo de retorno de 2.33 años, (mm)	
P25yr: Precipitación para un periodo de retorno de 25 años, (mm)	

Contenido

1.	Introducción	15
2.	Descripción del problema.....	17
2.1	Planteamiento del problema	17
2.2	Justificación	19
3.	Objetivos	21
4.	Marco teórico.....	22
4.1	Marco conceptual.....	22
4.1.1	Cuenca hidrográfica.....	22
4.1.2	Variables fisiográficas y parámetros morfométricos de cuencas hidrográficas	23
4.1.3	Tiempo de concentración	25
4.1.4	Periodo de retorno (T_r)	28
4.1.5	Análisis de datos dudosos.....	29
4.1.6	Polígonos de Thiessen	30
4.1.7	Distribuciones de probabilidad y pruebas de bondad de ajuste	31
4.1.8	Regresión lineal múltiple (RLM).....	32
4.1.9	Criterio de Información de Akaike (AIC):	33
4.1.10	Distancia de Cook:	33
4.1.11	Validación cruzada (LOOCV)	34
4.1.12	Métricas de rendimiento de modelos	35
4.1.12.1	Coeficiente de Determinación R^2	35
4.1.12.2	Kling-Gupta Efficiency (KGE)	35
4.2	Técnicas y métodos en el área de investigación	37
4.2.1	SIG aplicado a la caracterización de hoyas hidrográficas.....	37
4.2.2	Paquetes computacionales estadísticos en R	38
4.3	Experiencias con análisis de RLM entre morfometría e hidrología.....	38
4.3.1	Colombia	38
4.3.2	Internacionales	40
4.3.2.1	España	40
4.3.2.2	Australia	40
4.3.2.3	Costa Rica	41
4.3.2.4	Honduras	41
4.4	Legislación asociada a estimación de caudales para diferentes T_r	44

5.	Materiales y métodos.....	45
5.1	Descripción del lugar de estudio	45
5.1.1	Descripción asociada a la investigación	47
5.2	Metodología	56
5.2.1	Objetivo específico 1	57
5.2.2	Objetivo específico 2	57
5.2.3	Objetivo específico 3	58
5.2.4	Objetivo específico 4	59
5.2.5	Objetivo específico 5	60
6.	Resultados y discusión.....	61
6.1	Revisión bibliográfica detallada.....	61
6.2	Caracterización morfométrica de cuencas hidrográficas a través de SIG	62
6.3	Estimación de precipitaciones y caudales máximos para diferentes periodos de retorno	69
6.4	Regresión Lineal Múltiple	84
6.4.1	Escenario 1	84
6.4.2	Escenario 2	87
6.5	Rendimiento de los modelos	94
6.5.1	Escenario 1	94
6.5.2	Escenario 2	100
7.	Conclusiones y recomendaciones	111
7.1	Trabajos futuros	114
7.2	Publicaciones y productos derivadas de esta tesis	115
8.	Referencias.....	117
9.	Anexos.....	126

Listado de figuras

Figura 1	– Cuenca hidrográfica Río Páez	22
Figura 2	– Divisoria de aguas.....	22
Figura 3	– Diagrama conceptual tipos de parámetros morfométricos	24
Figura 4	– Polígonos de Thiessen para estaciones 26020180, 26035030, 26015010 y 26025090	31
Figura 5	– Mapa de cobertura de una cuenca hidrográfica mediante SIG	37
Figura 6	– Licencia de Arcmap 10.8.2	37
Figura 7	– Área de Aplicación de las Ecuaciones de Regresión	42
Figura 8	– Localización y variabilidad espacial de las características de la cuenca a lo largo del dominio de estudio: (a) elevación media, (b) temperatura media anual, (c) índice de aridez, (d)	

precipitación media anual, (e) precipitación que cae en forma de nieve, (f) escorrentía media anual y (g) régimen hidrológico.....	43
Figura 9 – Validación cruzada Macrozonas sur y parte de la austral	43
Figura 10 - Localización Departamento del Cauca en Colombia	45
Figura 11 – Estaciones IDEAM con datos de Caudales en el Departamento del Cauca	53
Figura 12 – Estaciones IDEAM con datos de Precipitaciones Máximas en 24 horas en el Departamento del Cauca	54
Figura 13 – Estaciones IDEAM con datos de Caudales y Precipitaciones Máximas en 24 horas en el Departamento del Cauca	54
Figura 14 – Extensión intersecada de las cuencas hidrográficas del presente estudio versus Departamento del Cauca	55
Figura 15 – Diagrama de flujo metodología implementada.....	56
Figura 16 – Generación de la cuenca hidrográfica #40 Totoró como archivo vectorial en ArcMap .	62
Figura 17 – Generación de la cuenca hidrográfica #40 Totoró como archivo vectorial en delineator.py.....	63
Figura 18 – Rectángulo eq. de la cuenca hidrográfica #40 Totoró	64
Figura 19 – Curva hipsométrica de la cuenca hidrográfica #40 Totoró	64
Figura 20 – Cálculo de la pendiente media de la cuenca hidrográfica #40 Totoró mediante el método de Horton en ArcMap.....	65
Figura 21 – Diagrama de cajas y bigotes para Tc cuenca hidrográfica #40 Totoró	66
Figura 22 – Diagrama de barras para Tc cuenca hidrográfica #40 Totoró.....	66
Figura 23 – Mapa de C y CN para las 41 cuencas del estudio	67
Figura 24 – Diagrama de flujo para el C&CN_Calculator	68
Figura 25 – Precipitación máxima en 24 horas vs limites estación #48 Piendamó	74
Figura 26 – Curvas IDF para estación #48 Piendamó	75
Figura 27 – Polígonos de Thiessen y porcentaje de aferencia de estaciones de precipitación sobre la cuenca #40 Totoró en ArcMap	75
Figura 28 – Distribuciones de probabilidad con datos de precipitación de la cuenca #40 Totoró.....	77
Figura 29 – Gráfica de dispersión para visualizar datos de PMáx y QMáx para diferentes Tr y todas las 38 de cuencas.....	83
Figura 30 – AIC para Tr 200 escenario 1	85
Figura 31 – AIC para Tr 500 escenario 1	85
Figura 32 – Matriz de correlación para el mejor modelo Tr 200 escenario 1	86
Figura 33 – Matriz de correlación para el mejor modelo Tr 500 escenario 1	86
Figura 34 – Área de estudio escenario 2 vs área de estudio escenario 1	87
Figura 35 – AIC para todos los Tr escenario 2	90
Figura 36 – Matriz de correlación para el mejor modelo Tr 2.33 escenario 2	91
Figura 37 – Matriz de correlación para el mejor modelo Tr 25 escenario 2	91
Figura 38 – Matriz de correlación para el mejor modelo Tr 50 escenario 2	92
Figura 39 – Matriz de correlación para el mejor modelo Tr 100 escenario 2	92
Figura 40 – Matriz de correlación para el mejor modelo Tr 200 escenario 2	93
Figura 41 – Matriz de correlación para el mejor modelo Tr 500 escenario 2	93
Figura 42 – Predicciones vs valores reales para el escenario 1.....	97
Figura 43 – Residuos vs predicciones para el escenario 1	98
Figura 44 – Histograma de residuales para el escenario 1	99

Figura 45 – QQ-Plot de residuales para el escenario 1	99
Figura 46 – Distribución de errores absolutos para el escenario 1	100
Figura 47 – Predicciones vs valores reales para el escenario 2	107
Figura 48 – Residuos vs predicciones para el escenario 2	108
Figura 49 – Histograma de residuales para el escenario 2	109
Figura 50 – QQ-Plot de residuales para el escenario 2	109
Figura 51 – Distribución de errores absolutos para el escenario 2	110

Listado de tablas

Tabla 1 – Clasificación de las cuencas según tamaño	23
Tabla 2 – Partes de una cuenca hidrográfica	23
Tabla 3 – Parámetros morfométricos de una cuenca hidrográfica	25
Tabla 4 – Periodos de retorno dependiendo de la obra civil	29
Tabla 5 – Valores de K_n de un lado con un nivel de significancia del 10% para la distribución normal	30
Tabla 6 – Legislación relacionada con estimación de caudales máximos	44
Tabla 7 – Listado de estaciones LG/LM empleadas en la investigación	48
Tabla 8 – Listado de estaciones CO/CP/PG/PM/SP empleadas en la investigación	52
Tabla 9 – Variables fisiográficas y parámetros morfométricos básicos para la cuenca #40 Totoró ..	63
Tabla 10 – Tiempos de concentración para la cuenca #40 Totoró	65
Tabla 11 – Valores de C y CN para las 41 cuencas del estudio	67
Tabla 12 – Datos de precipitación máxima mensual y multianual para la estación #48 Piendamó ...	72
Tabla 13 – Verificación de datos dudoso para la estación #48 Piendamó	73
Tabla 14 – Serie ponderada de precipitación para la estación #40 Totoró	77
Tabla 15 – Pruebas de bondad de ajuste organizadas para la cuenca #40 Totoró	78
Tabla 16 – Cálculo de precipitación máxima para diferentes periodos de retorno para la estación #40 Totoró	78
Tabla 17 – Distribuciones de probabilidad para precipitación y caudal	79
Tabla 18 – Matriz global de variables independientes y 38 cuencas hidrográficas del Departamento del Cauca 1/3	80
Tabla 19 – Matriz global de variables independientes y 38 cuencas hidrográficas del Departamento del Cauca 2/3	81
Tabla 20 – Matriz global de variables independientes y 38 cuencas hidrográficas del Departamento del Cauca 3/3	82
Tabla 21 – Variables independientes de modelos de RLM para Tr 200 años escenario 1	84
Tabla 22 – Variables independientes de modelos de RLM para Tr 500 años escenario 1	84
Tabla 23 – Cuencas influyentes con distancia de Cook depuradas de los modelos para escenario 2 87	
Tabla 24 – Variables independientes de modelos de RLM para Tr 2.33 años escenario 2	88
Tabla 25 – Variables independientes de modelos de RLM para Tr 25 años escenario 2	88
Tabla 26 – Variables independientes de modelos de RLM para Tr 50 años escenario 2	88
Tabla 27 – Variables independientes de modelos de RLM para Tr 100 años escenario 2	89
Tabla 28 – Variables independientes de modelos de RLM para Tr 200 años escenario 2	89
Tabla 29 – Variables independientes de modelos de RLM para Tr 500 años escenario 2	89

Tabla 30 – Cálculo del error para Tr 200 años escenario 1.....	95
Tabla 31 – Cálculo del error para Tr 500 años escenario 1.....	96
Tabla 32 – Métricas de rendimiento para modelos del escenario 1	97
Tabla 33 – Cálculo del error para Tr 2.33 años escenario 2.....	101
Tabla 34 – Cálculo del error para Tr 25 años escenario 2.....	102
Tabla 35 – Cálculo del error para Tr 50 años escenario 2.....	103
Tabla 36 – Cálculo del error para Tr 100 años escenario 2.....	104
Tabla 37 – Cálculo del error para Tr 200 años escenario 2.....	105
Tabla 38 – Cálculo del error para Tr 500 años escenario 2.....	106
Tabla 39 – Métricas de rendimiento para modelos del escenario 2	107

1. Introducción

En proyectos de ingeniería hidráulica e hidrológica, como el diseño de presas y puentes, es crucial calcular con precisión los caudales máximos ($Q_{MÁX}$) para diferentes probabilidades de ocurrencia, usando toda la información disponible. Estos caudales, influenciados por eventos extremos de precipitación, son esenciales para dimensionar y mitigar las obras (Maidment, 1993; Monsalve S., 1999; Morassutti F., 2020). La metodología directa implica análisis estadísticos de datos hidrológicos, que deben provenir de estaciones de medición confiables (Bobee, y otros, 1996). En Colombia, la información cartográfica es desactualizada, afectando la precisión de las estimaciones (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2012; IGAC - Instituto Geográfico Agustín Codazzi, 2017), además existen 5178 estaciones, pero la densidad de medición es baja respecto a la riqueza hídrica del país (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2023; Suárez Fiesco & Almeida Rosero, 2017), siendo aún más crítico en el Departamento del Cauca (29.308 Km²), en donde se tienen de manera activa, 86 estaciones de medición de precipitación y 41 estaciones de medición de caudal, lo cual genera una densidad muy limitada, dada la riqueza hídrica del departamento en donde nacen los principales ríos del país (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2023), obligando a realizar cálculos con modelos de lluvia-escorrentía, aun cuando su aplicación en Colombia puede ser inadecuada debido a diferencias en condiciones climáticas y geográficas de donde fueron concebidos (Toro Triana, 2013) y dejando a un lado modelos más simples que pudieran generar resultados robustos.

El diseño y la construcción de infraestructuras hidráulicas requieren una comprensión precisa de los caudales máximos que pueden presentarse en un río o cuenca. La estimación acertada de estos caudales máximos es fundamental para asegurar la funcionalidad y la seguridad de estas (Bedient, Huber, & Vieux, 2019), aun mas en tiempos de cambio climático y aumento de desastres naturales, que aunque siempre han afectado a la humanidad, están impactando más a la población reciente debido a urbanización, migraciones y cambios en el uso del suelo (Viollet, 2005; Matovelle & D., 2020; Rojas Morales, Aguilar Pereira, & Solís Bolaños, 2008). Estos cambios alteran cómo la precipitación interactúa con el suelo, afectando los caudales generados (INVIAS, 2009), la mayoría de las veces para incrementarlos y potenciar la amenaza.

Existen diversos métodos de pronóstico de caudales, como modelos de series de tiempo, redes neuronales artificiales y regresión lineal múltiple (Rojo Hernández & Carvajal Serna, 2010; Wu & Chau, 2013), este último como foco de la presente investigación, la cual buscó desarrollar un modelo de estimación de caudales máximos relacionando variables fisiográficas, parámetros morfométricos y variables hidrológicas como precipitaciones y caudales máximos (Patton, y otros, 2022; Alvarez Sevilla, Brown Manrique, & del Cid Colindres, 2017).

En este proyecto se determinó una correlación entre variables fisiográficas, parámetros morfométricos y variables hidrológicas que permiten establecer ecuaciones para la estimación de caudales máximos en cuencas hidrográficas de una zona entre cordilleras del Departamento del Cauca. Este estudio no solo contribuye a la mejora en el diseño y la construcción de infraestructuras hidráulicas en la región, sino que también abrirá nuevas vías para la investigación en modelos más robustos y adaptados a las condiciones locales (Ouarda, y otros, 2008).

El documento se organiza en los siguientes capítulos:

- 1. Introducción:** En este capítulo se presenta el tema de estudio, se destaca su relevancia en el campo de investigación y se justifica la elección de este.
- 2. Descripción del problema:** Aquí se detalla la situación que origina y fundamenta la necesidad de llevar a cabo la investigación.
- 3. Objetivos:** Este apartado establece el objetivo general del proyecto y detalla objetivos específicos, los cuales definen metas concretas para alcanzar el objetivo general.
- 4. Marco teórico:** Proporciona los fundamentos teóricos que respaldan la investigación, además de revisar estudios previos relevantes y detallar las normativas y regulaciones pertinentes.
- 5. Materiales y métodos:** Describe la metodología que se utilizará en la investigación y proporciona el contexto para su aplicación, además orienta la forma en que se alcanzarán los resultados esperados.
- 6. Resultados y discusión:** Presenta los hallazgos obtenidos durante la investigación.
- 7. Conclusiones:** Resume los descubrimientos principales y las reflexiones derivadas de los resultados y de la revisión bibliográfica. Adicionalmente se presentan trabajos futuros y productos asociados a la investigación.

2. Descripción del problema

2.1 Planteamiento del problema

En la mayoría de proyectos relacionados con Ingeniería Hidráulica e Hidrológica que requieran diseño y construcción de obras de infraestructura que intervengan ríos y sus márgenes, como, por ejemplo, obras de drenaje, bocatomas, presas retenedoras de sedimentos, puentes, protecciones de orillas, etc., es necesario calcular con la mayor precisión posible (Maidment, 1993) y empleando la totalidad de información que se disponga, los caudales máximos ($Q_{MÁX}$) para diferentes probabilidades de ocurrencia, o tiempos de retorno (T_r), dependiendo del nivel de complejidad y socio económico de la obra (Monsalve S., 1999), debido a que es la variable más importante para el dimensionamiento, operación y eventuales medidas de mitigación de las mismas (Morassutti F., 2020). Siendo consecuentes con los procesos del ciclo hidrológico, dichos caudales máximos son generados principalmente por los eventos extremos de precipitación que ocurren en los mismos periodos de retorno (Ponce, 1989).

La metodología directa para estimar dichos caudales máximos consiste en, mediante análisis estadístico (distribuciones de probabilidad y pruebas de bondad de ajuste), realizar la proyección de las variables medidas (Massey, 1951), para ello se debe contar con una estación de medición de variables hidrológicas, es decir al menos caudales y niveles (Estaciones Limnimétricas y Limnigráficas), cerca al sitio de interés, y que dispongan de un buen número de datos y que tengan buena calidad (congruencia, consistencia, periodicidad, etc.) (Bobee, y otros, 1996).

En Colombia se tienen 5178 estaciones en funcionamiento, de las cuales 2639 son operadas por el Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia – IDEAM y las 2539 estaciones restantes se encuentran bajo la operación de otras entidades. Adicionalmente se tiene que dentro de las estaciones en funcionamiento las que miden parámetros hidrológicos corresponden a 1511, de las cuales 646 son operadas por el IDEAM y 865 por otras entidades (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2023). De acuerdo con la información suministrada por esta entidad, la red de monitoreo de caudales a nivel nacional cuenta con una densidad correspondiente de 1767 km^2 por estación, la cual se considera baja, dada la riqueza hídrica nacional (Suárez Fiesco & Almeida Rosero, 2017). En el Departamento del Cauca esta densidad mejora un poco alcanzando los 800 km^2 por estación, sin embargo, teniendo en cuenta la riqueza hídrica con la que dispone el departamento, en donde se encuentra ubicado el Macizo Colombiano, lugar de nacimiento de los principales ríos del país, esta sigue siendo limitada.

Debido a lo anterior, en la mayoría de los trabajos de consultoría o investigación, en donde se requieran conocer caudales o niveles máximos, se deben formular métodos de lluvia-escurrentía (Chow, Maidment, & Mays, 1988), como por ejemplo el método racional para

áreas pequeñas ($A < 2.5km^2$), hidrogramas unitarios ($A > 2.5km^2$), modelos mixtos de hidrogramas unitarios con tránsito de crecientes ($A > 20km^2$) (Aparicio, 1997), o incluso modelos aplicables a áreas específicas con condiciones morfométricas particulares (Viessman & Lewis, 2003). Sin embargo, su aplicación indiscriminada genera controversia al momento de aplicarlos a otras regiones o países diferentes de donde fueron concebidos (Toro Triana, 2013), como sería el caso del territorio colombiano, el cual es el segundo país más diverso del mundo y que cuenta con toda clase de pisos térmicos y coberturas (Boron, Payán, MacMillan, & Tzanopoulos, 2016).

Dichas metodologías tienen en común que requieren los insumos principales de un modelo hidrológico estándar, como la precipitación, la condición de humedad antecedente del suelo, la evapotranspiración y la infiltración, y también características fisiográficas y parámetros morfométricos del área de drenaje, como área de drenaje, longitud del cauce principal, pendientes del cauce principal, entre otros (Témez., 1991), así como una caracterización de las coberturas y tipos de suelos, principalmente, siendo estos últimos estimados con cartografía base suministrada por entidades gubernamentales, como el IDEAM y el IGAC, en productos asociados a los años 2012 (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2012) y 2017 (IGAC - Instituto Geográfico Agustín Codazzi, 2017) respectivamente, siendo información desactualizada y con una escala por lo general muy grande para poder aplicar con confiabilidad en áreas de drenaje de determinados tamaños, evidenciando una influencia importante en las estimaciones de los caudales ya que son variables sensibles al comportamiento antrópico, debido a la susceptibilidad de ser modificada para fines de explotación ganadera, maderera o minera (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2012). También son importantes, ya que permiten predecir y visualizar de manera directa la interacción entre la precipitación y el suelo (Boughton, 1989), y en definitiva relaciona la cantidad de agua que se podría convertir en escorrentía y la cantidad de agua que se infiltra al subsuelo (Freyberg, Reeder, Franzini, & Remson, 1980).

Adicionalmente al ser métodos extranjeros, su aplicación en el ámbito nacional podría no tener en cuenta las características propias del medio (Hawkins, Theurer, & Rezaeianzadeh, 2019), tendiendo a sub estimar o sobre estimar dichos caudales lo que finalmente conlleva al subdimensionamiento o sobredimensionamiento y posteriormente a una potencial falla o sobre costo de las obras hidráulicas que se relacionan con los cauces naturales, que pretender servir con coherencia técnica y económica (INVIAS, 2009) problemas de abastecimiento de agua, control de inundaciones, mejoramiento de infraestructura vial, gestión del riesgo, entre otras varias aplicaciones, por lo tanto influye de manera directa en todo el tema de gestión integral del recurso hídrico.

A partir de la problemática expuesta, la pregunta de investigación que surge es:

¿Existe una correlación significativa entre variables fisiográficas, parámetros morfométricos, y variables hidrológicas como precipitaciones y caudales máximos asociados a diferentes

periodos de retorno, que permita establecer ecuaciones para la estimación de caudales máximos de manera confiable en cuencas hidrográficas no instrumentadas en el Departamento del Cauca, Colombia?

2.2 Justificación

A pesar de que los desastres naturales relacionados con el agua han acompañado al desarrollo de la humanidad desde sus orígenes (Viollet, 2005), estos pareciera que empiezan a afectar en mayor medida a la población en los últimos años (Matovelle & D., 2020), que aumenta en número y en expansión territorial cada vez más, este aumento se ve reflejado en los procesos de urbanización, movimientos migratorios, invasiones de rondas hídricas y modificación de coberturas para explotación antrópica (Rojas Morales, Aguilar Pereira, & Solís Bolaños, 2008). Estas tendencias tienen importantes repercusiones en la interacción de la precipitación con el suelo, ya que al cambiar el estado de cobertura de las hoyas hidrográficas en cuanto a cantidad y calidad se verán influenciados los caudales generados en la misma por un evento determinado de precipitación (INVIAS, 2009).

En la actualidad existen varios métodos de pronóstico de caudales como, por ejemplo, Modelos de Series de Tiempo, Redes Neuronales Artificiales RNA, Polinomios Adaptivos de regresión Multivariada, entre otros (Rojo Hernández & Carvajal Serna, 2010), aplicados en diversas cuencas hidrográficas alrededor del mundo (Wu & Chau, 2013), con diversos resultados, sin embargo, también es importante resaltar el método de Regresión lineal Múltiple RLM, el cual en general es un análisis predictivo que busca captar las relaciones entre variables explicativas y una variable de interés usando diversas técnicas que relacionan hechos históricos (valores observados) y actuales para hacer predicciones sobre acontecimientos futuros (valores calculados) (Rojo Hernández J. D., 2011; Montgomery, Peck, & Vining, 2012).

En el presente proyecto se busca estimar adecuadamente una relación entre las variables fisiográficas y parámetros morfométricos con las variables hidrológicas, lo cual es de vital importancia para entender lo más cercano posible a la realidad la interacción de la precipitación con la cuenca hidrográfica (Patton, y otros, 2022), con el propósito de desarrollar un modelo de estimación de caudales máximos para diferentes periodos de retorno, a partir de variables fisiográficas, parámetros morfométricos e información de precipitación, todos, datos relativamente sencillos de estimar u obtener en el medio, para los dos primeras se aprovechan las metodologías de Sistemas de Información Geográfica (SIG) ya que se requiere organización, almacenamiento, manipulación, análisis y modelización de grandes cantidades de datos procedentes del mundo real que están vinculados a una referencia espacial, facilitando la incorporación de aspectos sociales-culturales, económicos y ambientales que conducen a la toma de decisiones de una manera más eficaz (Burrough, McDonnell, & Lloyd, 1998; Castro & Maidment, 2020), mientras que la última es suministrada por la entidad gubernamental encargada de monitorear las variables

meteorológicas, el IDEAM, para finalmente poder diseñar obras hidráulicas relacionadas con los cauces naturales (Alvarez Sevilla, Brown Manrique, & del Cid Colindres, 2017).

Los resultados obtenidos de este proyecto se espera que puedan ser empleados para diversos fines de consultoría e investigación, determinando la necesidad o no, de modificaciones en la magnitud de obras hidráulicas, desde el punto de vista técnico o económico, en caso de que las magnitudes de los caudales presenten diferencias relativamente importantes, tanto así que resulte conveniente o no, la adopción de un método para todo el territorio nacional. Adicionalmente se espera poder trazar un camino hacia temas de investigación en modelos más robustos, que mejoren o validen los resultados hallados, y que generen impacto al optimizar y reducir de la incertidumbre que conlleva la aplicación de modelos lluvia-escorrentía bajo información de entrada supuesta, escasa o inexistente en muchos casos (Ouarda, y otros, 2008).

3. Objetivos

Objetivo general:

Desarrollar un modelo para estimación de caudales máximos a partir de parámetros morfométricos e hidrológicos en cuencas hidrográficas del Departamento del Cauca, Colombia.

Objetivos específicos:

- Identificar mediante una revisión bibliográfica detallada, las diferentes metodologías para el desarrollo de modelos para estimación de caudales máximos en diferentes zonas o territorios nacionales e internacionales.
- Calcular las variables fisiográficas y parámetros morfométricos mediante SIG de 41 cuencas instrumentadas con estaciones de medición de caudal por el IDEAM en el Departamento del Cauca.
- Estimar precipitaciones y caudales máximos para diferentes periodos de retorno aferentes a las 41 cuencas hidrográficas mediante distribuciones de probabilidad y pruebas de bondad de ajuste.
- Establecer la mejor correlación lineal múltiple entre las variables morfométricas e hidrológicas con los caudales para diferentes periodos de retorno.
- Determinar el rendimiento de los modelos de correlación seleccionados para cada periodo de retorno, mediante métricas de eficiencia R^2 y KGE obtenidas con validación cruzada.

4. Marco teórico

4.1 Marco conceptual

4.1.1 Cuenca hidrográfica

También llamada hoya hidrográfica, es un área de drenaje que su límite son los puntos más altos de la superficie que las contiene formando una línea denominada divisoria de aguas (Goigochea Pinchi, 2016).

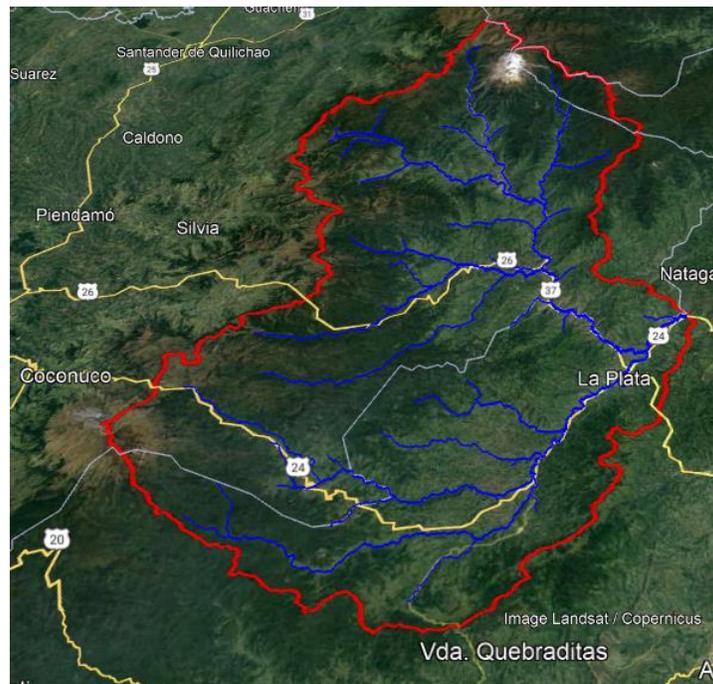


Figura 1 – Cuenca hidrográfica Río Páez
Fuente. El Autor



Figura 2 – Divisoria de aguas
Fuente. El Autor

Las cuencas pueden tener varias clasificaciones, sin embargo, una de las relevantes es la de su tamaño, ya que influye de manera directa las magnitudes de las escorrentías generadas por las precipitaciones (Ferrer, 1993), esta clasificación se presenta a continuación, definiéndose en función del área superficial de la misma.

RANGO DE ÁREAS (km^2)	CLASIFICACIÓN
< 25	Microcuenca
25 a 250	Pequeña
250 a 500	Intermedia-pequeña
500 a 2500	Intermedia-grande
2500 a 5000	Grande
> 5000	Muy grande

Tabla 1 – Clasificación de las cuencas según tamaño

Fuente. (Gualdrón, Villate, Torres, Molina, & Reyes, 2022)

Es importante también definir las partes de la cuenca hidrográfica, las cuales ayudan a definir el comportamiento e interacción del recurso hídrico los demás componentes del entorno.

PARTES DE UNA CUENCA HIDROGRÁFICA	
Cuenca Alta	Corresponde a la zona más elevada en donde se generan los nacimientos de los cuerpos de agua.
Cuenca Media	Es la sección de la cuenca en donde se equilibra el transporte de agua con el transporte de sedimentos.
Cuenca Baja	Corresponde a la zona más baja de la cuenca en donde se depositan los sedimentos arrastrados por la corriente.

Tabla 2 – Partes de una cuenca hidrográfica

Fuente. (Gualdrón, Villate, Torres, Molina, & Reyes, 2022)

4.1.2 Variables fisiográficas y parámetros morfométricos de cuencas hidrográficas

Son los indicadores que caracterizan y determinan el comportamiento fisiográfico de la cuenca (Reyes Trujillo, Barroso, & Carvajal Escobar, 2010). Los estudios que involucran estos parámetros son en su mayoría conjunto de datos lineales de área, relieve y drenaje, las cuales permiten generar deducciones e interpretaciones basadas en teorías hidrológicas, logrando así obtener otras variables del ciclo hidrológico, por lo general calculados mediante SIG (Jensen, 2009).

Los parámetros morfométricos de una cuenca por lo general entran en los siguientes grandes grupos:

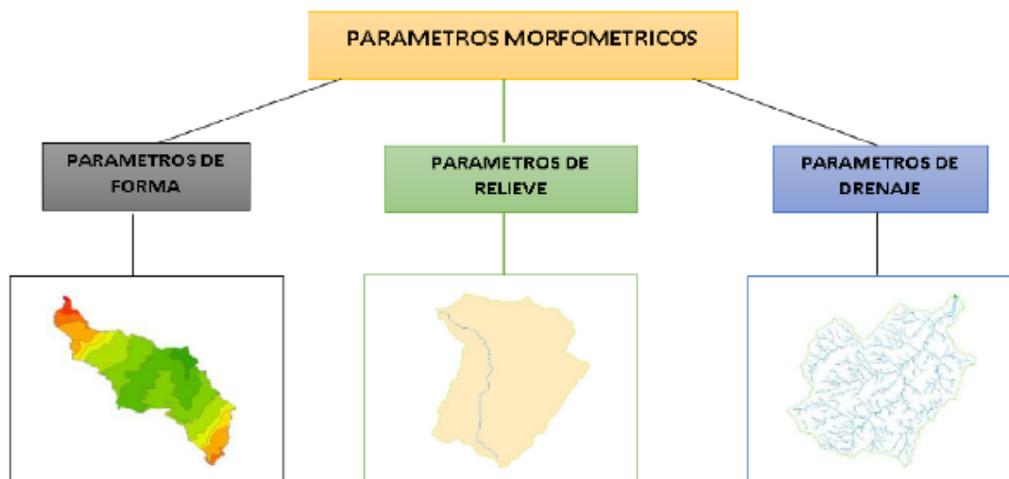


Figura 3 – Diagrama conceptual tipos de parámetros morfométricos

Fuente. (Gualdrón, Villate, Torres, Molina, & Reyes, 2022)

Variable	Parámetros
Escala de la cuenca (10 parámetros)	1. Área (A) (Km^2); 2. Área de la vertiente mayor (Av_{may}) (Km^2); 3. Área de la vertiente menor (Av_{men}) (Km^2); 4. Perímetro (P) (Km); 5. Longitud (L) (Km); 6. Ancho promedio (a) (Km); 7. Ancho máximo ($a_{máx}$) (Km); 8. Diámetro (D) (Km); 9. Área del rectángulo equivalente al de la cuenca (Ar) (Km^2); 10. Perímetro de un círculo igual al área de la cuenca (Pc) (Km)
Gradiente y forma del relieve de la cuenca (18 parámetros)	11. Altitud mínima (h) (msnm); 12. Altitud máxima (H) (msnm); 13. Altitud media (H_m) (msnm); 14. Altitud de la naciente de la corriente principal (H_n) (msnm); 15. Altitud del desagüe de la corriente principal (h_d) (msnm); 16. Relieve máximo (Rm) (m); 17. Radio del relieve (R) (m/ Km); 18. Pendiente media del relieve (P_m) (m/m); 19. Concavidad del perfil longitudinal de la corriente principal (Cp); 20. Pendiente media del perfil longitudinal de la corriente principal (Pcp_m) (m/m); 21. Relieve máximo de la corriente principal ($Rmcp$) (m); 22. Factor topográfico de la corriente principal (Ft); 23. Pendiente media del total de corrientes de la red de drenaje (Ptc_m) (m/m); 24. Integral hipsométrica (IH); 25. Coeficiente de masividad (Cm) (msnm/ Km^2); 26. Coeficiente orográfico (Co); 27. Relieve relativo (Rr) (m/ Km); 28. Número de rugosidad de Melton (NR)
Variable	Parámetros
Forma de la cuenca (14 parámetros)	29. Elongación (E); 30. Perímetro relativo de crenulación (Prc); 31. Coeficiente de compacidad (Kc); 32. Índice de alargamiento (Ia); 33. Índice de homogeneidad (Ih); 34. Índice de simetría (Is); 35. Factor forma (Ff); 36. Alargamiento de caquot (Ac); 37. Radio de elongación (Re); 38. Radio de circularidad (RC); 39. Índice de forma (If); 40. Índice entre el largo y el área de la cuenca (ICo); 41. Coeficiente de forma (Cf); 42. Coeficiente de redondez (Cr)
Extensión de la red de drenaje (15 parámetros)	43. Longitud de la corriente principal (Lcp) (Km); 44. Longitud media del valle de la corriente principal (Lv_m) (Km); 45. Longitud total de las corrientes de la red de drenaje (Ltc) (Km); 46. Sinuosidad total de la corriente principal (S); 47. Sinuosidad topográfica de la corriente principal (St); 48. Sinuosidad hidráulica de la corriente principal (Sh); 49. Centro de gravedad del cauce principal (Cg) (Km); 50. Densidad de drenaje (Dd) (Km/Km^2); 51. Textura del drenaje (Td) (Km^{-1}); 52. Coeficiente de mantenimiento del canal (Cmc) (Km^2/Km); 53. Extensión media de la escorrentía superficial (E_m) (Km); 54. Longitud del flujo superficial (Lfs) (Km); 55. Textura topográfica (Tt); 56. Intensidad de drenaje (Id) (Km); 57. Alejamiento medio (Am)

Orden y magnitud de la red de drenaje (25 parámetros)

58. Orden de la corriente (u); 59. Número de corrientes de un orden dado (N_u); 60. Número total de corrientes de la red de drenaje (N_T); 61. Orden de la cuenca (u); 62. Relación de bifurcación (R_b); 63. Relación de bifurcación media (R_{bm}); 64. Longitud de la corriente de un orden dado (L_u) (Km); 65. Longitud total de las corrientes de un orden dado (L_{tu}) (Km); 66. Longitud media de las corrientes de un orden dado (L_{mu}) (Km); 67. Longitud media acumulada de las corrientes de un orden dado (L_{mau}) (Km); 68. Relación de longitud (R_l); 69. Relación de longitud media (R_{lm}); 70. Desnivel de la corriente de un orden dado (D_u) (m); 71. Pendiente de la corriente de un orden dado (P_u) (m/m); 72. Densidad media de las corrientes de un orden dado (D_{mu}) (m); 73. Pendiente media de las corrientes de un orden dado (P_{mu}) (m/m); 74. Pendiente media acumulada de las corrientes de un orden dado (P_{mau}) (m/m); 75. Relación de pendiente (R_p); 76. Relación de pendiente media (R_{pm}); 77. Frecuencia de las corrientes de un orden dado (F_u) (N_u/Km^2); 78. Frecuencia media de las corrientes totales de la red de drenaje (F_{mc}) (N_T/Km^2); 79. Número de corrientes de orden 1 (N_1); 80. Coeficiente de torrencialidad (Ct) (N_1/Km^2); 81. Coeficiente de almacenamiento (Ca); 82. Magnitud de la cuenca (M) (N_1)

Hidráulico (1 parámetro)

83. Ancho máximo de la sección hidráulica de interés (ash_{mdr}) (m)

Tabla 3 – Parámetros morfométricos de una cuenca hidrográfica

Fuente. (Méndez, Córdova, & Bravo de Guenni, 2015)

4.1.3 Tiempo de concentración

Se considera tal vez el dato más importante en el análisis de los escurrimientos de agua y se define como el tiempo en que tarda en llegar por escorrentía superficial a la sección de salida de la cuenca o hasta un punto de interés, la gota de lluvia caída en el extremo hidráulicamente más alejado de la cuenca. Al cabo de este tiempo toda la cuenca está aportando caudales al río y será el caudal máximo en el punto considerado (Chow, Maidment, & Mays, 1988).

Importante en los modelos lluvia escorrentía porque la duración de la tormenta de diseño se hace igual al tiempo de concentración de la cuenca, puesto que es, para esta duración, cuando la totalidad de la cuenca está aportando al proceso de escorrentía, por ende, es de esperarse que se presenten los caudales máximos (INVIAS, 2009).

Existen diversos autores para la estimación de esta variable:

Kirpich - USBR

$$Tc = 0,06628 \left(\frac{L}{S^{0,5}} \right)^{0,77}$$

Tc: Tiempo de concentración, horas.

L: Longitud del cauce principal, km.

S: Pendiente entre las elevaciones máxima y mínima del cauce principal, m/m.

Témez (Dirección General de Carreteras)

$$Tc = 0,30 \left(\frac{L}{S^{0,25}} \right)^{0,76}$$

Tc: Tiempo de concentración, horas.

L: Longitud del cauce principal, km.

S: Pendiente entre las elevaciones máxima y mínima del cauce principal, m/m.

Williams

$$Tc = 0,683 \left(\frac{L \times A^{0,40}}{D \times S^{0,25}} \right)$$

Tc: Tiempo de concentración, horas.

L: Longitud del cauce principal, km.

A: Área, km².

D: Diámetro de una cuenca circular con área "A", km.

S: Pendiente entre las elevaciones máxima y mínima del cauce principal, %.

Joshstone y Cross

$$Tc = 2,6 \left(\frac{L}{S^{0,5}} \right)^{0,5}$$

Tc: Tiempo de concentración, horas.

L: Longitud del cauce principal, km.

S: Pendiente entre las elevaciones máxima y mínima del cauce principal, m/km.

Giandotti

$$Tc = \frac{4 \times A^{0,5} + 1,50 \times L}{25,3(L \times S)^{0,5}}$$

Tc: Tiempo de concentración, horas.

L: Longitud del cauce principal, km.

A: Área, km².

S: Pendiente entre las elevaciones máxima y mínima del cauce principal, m/m.

SCS – Ranser (Kirpich – Californiana)

$$Tc = 0,947 \left(\frac{L^3}{H} \right)^{0,385}$$

Tc: Tiempo de concentración, horas.

L: Longitud del cauce principal, km.

H: Diferencia de cotas entre puntos extremos de la corriente principal, m.

Ventura – Heras

$$T_c = 0,30 \left(\frac{L}{S^{0,25}} \right)^{0,75}$$

Tc: Tiempo de concentración, horas.

L: Longitud del cauce principal, km.

S: Pendiente entre las elevaciones máxima y mínima del cauce principal, %.

V.T Chow

$$T_c = 0,273 \left(\frac{L}{S^{0,5}} \right)^{0,64}$$

Tc: Tiempo de concentración, horas.

L: Longitud del cauce principal, km.

S: Pendiente entre las elevaciones máxima y mínima del cauce principal, m/m.

Cuerpo de Ingenieros del Ejército de E.E.U.U

$$T_c = 0,28 \left(\frac{L}{S^{0,25}} \right)^{0,76}$$

Tc: Tiempo de concentración, horas.

L: Longitud del cauce principal, km.

S: Pendiente entre las elevaciones máxima y mínima del cauce principal, m/m.

Guaire

$$T_c = 0,355 \left(\frac{A}{S^{0,5}} \right)^{0,595}$$

Tc: Tiempo de concentración, horas.

Área, km².

S: Pendiente entre las elevaciones máxima y mínima del cauce principal, m/km.

Passini

$$T_c = 0,108 \times \frac{(A \times L)^{1/3}}{\sqrt{S}}$$

Tc: Tiempo de concentración, horas.

L: Longitud del cauce principal, km.

A: Área, km².

S: Pendiente entre las elevaciones máxima y mínima del cauce principal, m/m.

Pilgrim y McDermont

$$T_c = 0,76 \times A^{0,38}$$

T_c: Tiempo de concentración, horas.

A: Área, km².

4.1.4 Periodo de retorno (T_r)

El periodo de retorno puede definirse como el tiempo promedio en el cual un evento es igualado o superado una vez cada “T_r” años (Chow, Maidment, & Mays, 1988). En términos matemáticos el periodo de retorno del m-ésimo evento de los “n” registrados, está definido por la siguiente expresión, también denominada probabilidad empírica:

$$T_r = \frac{n + 1}{m}$$

Dónde:

T_r: Periodo de retorno (años).

n: Número total de datos.

m: Número de orden de los datos.

Es usual en estudios hidrológicos trabajar con periodos de retorno en lugar de probabilidades, ya que dicha variable se expresa en unidades de tiempo, siendo semejante a la vida útil de las obras, lo cual presenta una ventaja comparativa (Aparicio, 1997).

Los periodos de retorno para la proyección de obras viales corresponden a 2, 2.33, 5, 10, 20, 25, 50 y 100 años, incluso en algunas ocasiones ese periodo de retorno puede ser de 200 a 500 años para la estimación de socavación en puentes de importante consideración (Guevara Álvarez, Socavación en puentes, 2016; Guevara Álvarez, Socavación en puentes - Evaluación, instrumentación y cálculo, 2021). Lo anterior establecido de acuerdo con las recomendaciones de diseño establecidas dentro del Manual de Drenaje de Carreteras (INVIAS, 2009). Dichos periodos de retorno son asociados al diseño obras hidráulicas viales de acuerdo con lo presentado a continuación:

TIPO DE OBRA	PERIODO DE RETORNO (AÑOS)
Cunetas	5
Zanjas de Coronación ²	10
Estructuras de Caída ²	10
Alcantarillas de 0.90 m de diámetro	10
Alcantarillas mayores a 0.90 m de diámetro	20
Puentes menores (luz menor a 10 m)	25
Puentes de luz mayor o igual a 10 m y menor a 50 m	50
Puentes de luz mayor o igual a 50 m	100
Drenaje subsuperficial	2

Tabla 4 – Periodos de retorno dependiendo de la obra civil
Fuente. (INVIAS, 2009)

4.1.5 Análisis de datos dudosos

El Water Resources Council recomienda realizar el análisis de datos dudosos (outliers), los cuales se definen como puntos de la información que se alejan significativamente de la tendencia de la información restante. Mantener o depurar estos datos puede afectar significativamente los parámetros estadísticos (Hu, 1987), de manera relevante en muestras pequeñas.

Si se tiene una asimetría mayor a +0.4, se consideran primero, las pruebas para detectar datos dudosos altos, si la asimetría es menor a -0.4, primero se consideran pruebas para detectar datos dudosos bajos. Cuando se tiene una asimetría entre -0.4 y +0.4, deben aplicarse pruebas para detectar dudosos altos y bajos (Water Resources Council (Actualmente United States Geological Survey), 1981).

$$y_H = \bar{y} + K_n s_y$$

$$y_L = \bar{y} - K_n s_y$$

En donde,

y_H : Umbral alto en unidades logarítmicas

y_L : Umbral bajo en unidades logarítmicas

\bar{y} : Media en unidad logarítmica

s_y : Desviación estándar en unidades logarítmicas

Valores K_n para la prueba de datos dudosos

Tamaño de muestra n							
n	K_n	n	K_n	n	K_n	n	K_n
10	2.036	24	2.467	38	2.661	60	2.837
11	2.088	25	2.486	39	2.671	65	2.866
12	2.134	26	2.502	40	2.682	70	2.893
13	2.175	27	2.519	41	2.692	75	2.917
14	2.213	28	2.534	42	2.700	80	2.940
15	2.247	29	2.549	43	2.710	85	2.961
16	2.279	30	2.563	44	2.719	90	2.981
17	2.309	31	2.577	45	2.727	95	3.000
18	2.335	32	2.591	46	2.736	100	3.017
19	2.361	33	2.604	47	2.744	110	3.049
20	2.385	34	2.616	48	2.753	120	3.078
21	2.408	35	2.628	49	2.760	130	3.104
22	2.429	36	2.639	50	2.768	140	3.129
23	2.448	37	2.650	55	2.804		

Tabla 5 – Valores de K_n de un lado con un nivel de significancia del 10% para la distribución normal

Fuente. (Water Resources Council (Actualmente United States Geological Survey), 1981)
Finalmente se pueden establecer los límites superiores e inferiores como:

$$\text{Límite superior} = 10^{y_H}$$

Picos de crecientas considerados como datos dudosos altos deben compararse con información histórica y de crecientas en sitios cercanos. La información histórica de crecientas contiene información de eventos inusualmente extremos, fuera del registro sistemático. Si existe información disponible que indica que un dato dudoso alto es el máximo sobre un periodo extendido, el dato dudoso es tratado como información histórica de crecientas y es excluido del análisis. Si no hay información disponible en cuanto a registros históricos útiles para comprar con los datos dudosos altos, entonces estos deben ser retenidos como parte del registro sistemático.

$$\text{Límite inferior} = 10^{y_L}$$

Los picos de crecientas considerados como datos dudosos bajos se eliminan del registro y puede aplicarse un ajuste de probabilidad condicional descrito por el USGS (Beckman & Cook, 1983).

4.1.6 Polígonos de Thiessen

Alfred Thiessen propuso un método para el cálculo del promedio de lluvias en grandes superficies. Dada el área de estudio y la localización de las estaciones meteorológicas, tomó en cuenta el área de influencia de dichas estaciones y ponderó cada valor por dicha área (Thiessen & Alter, 1911). Este proceso dio origen a lo que se conoce como Polígonos de Thiessen, o también Diagrama de Voronoi. Este método ha sido utilizado por más de cien

años en problemas relativos a la lluvia, variables meteorológicas en general, y un sinnúmero de otras aplicaciones en áreas diversas (Guelfi & López-Vazquez, 2018).

Este método no presenta mayores complicaciones en su aplicación (Pizarro T., Ramírez B., & Flores V., 2003). Además de la información pluviométrica, requiere para su ejecución sólo de la ubicación espacial de las estaciones. A cada estación se le asigna una superficie, la cual es obtenida representando las estaciones en un plano y uniéndolas a través de rectas. A estas rectas posteriormente se les trazan sus mediatrices hasta que se intersecan entre sí. Con los límites del área en estudio y los que definen las mediatrices se obtiene la superficie de influencia para cada estación (Linsley, 1988; Fernández, 1995) como se aprecia en la figura a continuación:

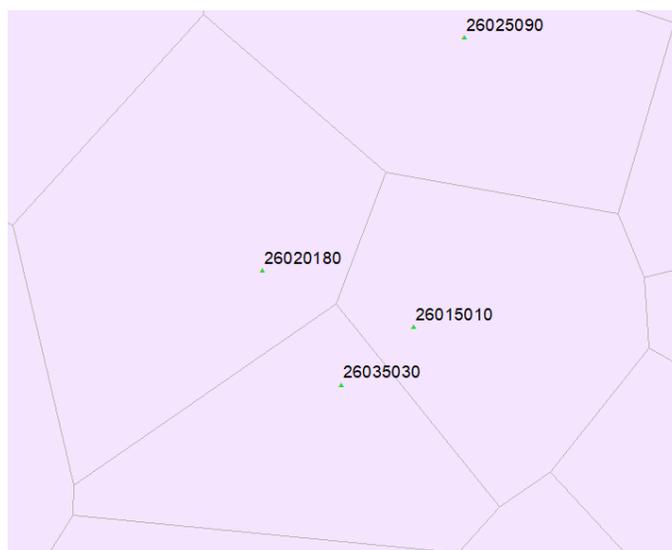


Figura 4 – Polígonos de Thiessen para estaciones 26020180, 26035030, 26015010 y 26025090

Fuente. El Autor

4.1.7 Distribuciones de probabilidad y pruebas de bondad de ajuste

Existen diferentes métodos que permiten determinar la probabilidad de ocurrencia de eventos de interés, sobre la base de una serie de datos para un determinado período, teniendo presente que el riesgo que se comete al ejecutar una obra está en función del número de años de estudio o registro, los cuales deben ser significativos, adecuados y precisos.

El análisis de frecuencias tiene como objetivo, relacionar la magnitud de los eventos extremos con su frecuencia de ocurrencia, mediante el empleo de distribuciones de probabilidad (Bobee, y otros, 1996).

Las distribuciones o funciones de probabilidad más usadas en hidrología corresponden a las distribuciones Normal, Lognormal, Lognormal (3P), Gamma, Gamma (3P), Log-Gamma y Log-Pearson (Aparicio, 1997), adicionalmente se emplea la distribución Gumbel Máx para tratar datos extremos (Aviles, Solera, & Paredes, 2016), dichas funciones se utilizarán para

la interpolación y extrapolación de valores de Caudales Máximos y Precipitaciones Máximas en 24 Horas, asociadas a los periodos de retorno indicados en la tabla 4.

Para la evaluación de la bondad de ajuste entre las funciones de probabilidad y los datos de Caudales Máximos y Precipitaciones Máximas en 24 Horas, existen las pruebas de Chi-Cuadrado, Anderson Darling y Kolmogorov-Smirnov, sin embargo, esta última presenta la ventaja sobre las otras dos, ya que permite comparar los datos con el modelo estadístico sin necesidad de agruparlos (Aparicio, 1997).

La prueba de Kolmogorov-Smirnov consiste en comparar el máximo valor absoluto de la diferencia D entre la función de distribución de probabilidad observada $F_o(xm)$ y la estimada $F(xm)$, con respecto a un valor crítico d , que depende del número de datos y el nivel de significancia seleccionado (Suárez Fiesco & Almeida Rosero, 2017).

$$D = \max |F_o(xm) - F(xm)|$$

Si $D < d$, se acepta la hipótesis nula. La función de distribución de probabilidad observada se calcula mediante la siguiente ecuación: $F_o(xm) = 1 - m/n + 1$

Donde m es el número de orden del dato xm en una lista de mayor a menor y n es el número total de datos.

4.1.8 Regresión lineal múltiple (RLM)

La regresión lineal múltiple es una técnica estadística utilizada para modelar la relación entre una variable dependiente y varias variables independientes. El objetivo es entender cómo las variables independientes (predictoras) afectan a la variable dependiente (respuesta) y hacer predicciones basadas en esta relación (Kutner, Nachtsheim, Neter, & Li, 2005).

El modelo de regresión lineal múltiple se expresa mediante la siguiente ecuación:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$

En donde:

Y : Es la variable dependiente.

β_0 : Intercepto o termino constante.

$\beta_1, \beta_2, \dots, \beta_k$: Son los coeficientes de regresión para las variables independientes.

X_1, X_2, \dots, X_k : Son las variables independientes.

ϵ : Es el término de error, que representa la variabilidad en Y no explicada por las variables independientes.

Empleadas en hidrología para obtener resultados de caudal en términos de otras variables (Campos-Aranda, 2016; Carrasquilla-Batista, y otros, 2016).

Para que los resultados de la regresión lineal múltiple sean válidos, deben cumplirse ciertos supuestos o checklist (Harvey, 1990):

Linealidad: La relación entre la variable dependiente y las independientes es lineal.

Independencia: Los residuos (errores) son independientes entre sí.

Homoscedasticidad: La varianza de los residuos es constante para todos los valores de las variables independientes.

Normalidad: Los residuos se distribuyen normalmente.

No multicolinealidad: Las variables independientes no están altamente correlacionadas entre sí.

4.1.9 Criterio de Información de Akaike (AIC):

Es un criterio utilizado para la selección de modelos estadísticos, proporciona una medida de la calidad relativa de un modelo dado en relación con otros modelos candidatos, teniendo en cuenta tanto el ajuste del modelo a los datos como la complejidad del modelo.

$$AIC = -2 \ln(L) + 2k$$

En donde:

$\ln(L)$: Logaritmo de la función de verosimilitud del modelo.

k : Número de parámetros en el modelo.

El AIC se utiliza para comparar diferentes modelos ajustados a los mismos datos. El modelo con el menor valor de AIC es considerado el mejor en términos de la relación entre ajuste y complejidad, se debe tener en cuenta que no proporciona una medida absoluta de la calidad del modelo, sino una medida relativa entre modelos (Akaike, 1974).

4.1.10 Distancia de Cook:

Es una medida de influencia en el análisis de regresión que cuantifica el efecto que tiene una observación particular en los parámetros estimados del modelo. Esta distancia ayuda a identificar observaciones que tienen un alto impacto en el ajuste del modelo, es decir, observaciones que, si se eliminan, podrían cambiar significativamente los resultados del análisis.

$$D_i = \frac{(\hat{Y}_i - \hat{Y})^2}{p \cdot MSE} \cdot h_{ii}$$

En donde:

\hat{Y}_i : Valor ajustado para la observación i .

\hat{Y} : Valor promedio de las observaciones ajustadas.

p : Número de parámetros del modelo.

MSE : Error cuadrático medio del modelo.

h_{ii} : Valor diagonal de la matriz de influencia (matriz de Hat) para la observación i .

Un valor alto de distancia de Cook indica que una observación tiene una gran influencia en los parámetros del modelo y puede estar afectando significativamente el ajuste del modelo. Valores críticos para la distancia de Cook pueden variar dependiendo del contexto, pero comúnmente se utilizan valores de referencia para identificar observaciones influyentes como:

$$D_i > 4/n$$

Donde:

n : Número de observaciones

La distancia de Cook se utiliza para diagnosticar la robustez del modelo y detectar puntos de datos que pueden estar afectando de manera desproporcionada la estimación de los parámetros (Cook, 1977).

4.1.11 Validación cruzada (LOOCV)

Es una técnica de evaluación de modelos estadísticos que se utiliza para estimar la capacidad de generalización de un modelo. Leave-One-Out Cross-Validation (LOOCV) es un tipo específico de validación cruzada en el cual, para cada iteración, se utiliza una única observación como conjunto de prueba y el resto como conjunto de entrenamiento.

Para un conjunto de datos con n observaciones, el modelo se entrena n veces, cada vez excluyendo una observación diferente como conjunto de prueba. La medida de rendimiento del modelo se calcula para cada iteración y luego se promedia para obtener una estimación general de su capacidad predictiva.

Presenta ventajas como por ejemplo el uso completo de datos, en donde cada observación se utiliza tanto para entrenamiento como para prueba, maximizando el uso del conjunto de datos, también el menor sesgo, dado que cada observación se usa para probar el modelo, se obtiene una evaluación menos sesgada (Hastie, Tibshirani, & Friedman, 2009).

4.1.12 Métricas de rendimiento de modelos

4.1.12.1 Coeficiente de Determinación R^2

Es una medida estadística que representa la proporción de la variabilidad en la variable dependiente que es explicada por las variables independientes en un modelo de regresión.

$$R^2 = 1 - \frac{SSE}{SST}$$

En donde:

SSE: Suma de los cuadrados de los residuos (la variabilidad no explicada por el modelo).

SST: Suma total de los cuadrados (la variabilidad total en los datos).

Un valor cercano a 1 indica que una gran proporción de la variabilidad de la variable dependiente es explicada por el modelo (Kutner, Nachtsheim, Neter, & Li, 2005).

En general se tiene:

Rango: 0 – 1

$R^2 \geq 0.9$: Ajuste muy bueno. El modelo explica la mayor parte de la variabilidad en los datos.

$0.7 \leq R^2 < 0.9$: Ajuste bueno. El modelo explica una parte significativa de la variabilidad.

$0.5 \leq R^2 < 0.7$: Ajuste moderado. El modelo explica algo de la variabilidad, pero no es completamente confiable.

$0.3 \leq R^2 < 0.5$: Ajuste pobre. El modelo explica poca variabilidad en los datos.

$R^2 < 0.3$: Ajuste muy pobre. El modelo apenas explica la variabilidad y puede no ser útil para predicciones (Srivastava, Srivastava, & Ullah, 1995).

4.1.12.2 Kling-Gupta Efficiency (KGE)

Es una métrica de rendimiento utilizada para evaluar la calidad del ajuste en modelos hidrológicos. Combina tres componentes: la correlación entre observaciones y predicciones, el sesgo de la media y la variabilidad.

$$KGE = 1 - \sqrt{(r - 1)^2 + \left(\frac{\sigma_{obs}/\sigma_{sim} - 1}{\beta}\right)^2 + \left(\frac{\mu_{sim} - \mu_{obs}}{\sigma_{obs}}\right)^2}$$

En donde:

r : Coeficiente de correlación entre observaciones y predicciones.

σ_{obs} y σ_{sim} : Son las desviaciones estándar de las observaciones y las simulaciones, respectivamente.

μ_{sim} y μ_{obs} : Son las medias de las observaciones y las simulaciones, respectivamente.

β : Parámetro que ajusta la sensibilidad del KGE a la relación entre desviaciones estándar.

Un valor de KGE cercano a 1 indica un buen rendimiento del modelo en términos de ajuste de datos, correlación y ajuste de la media y la variabilidad (Gupta, Kling, Yilmaz, & Martinez, 2009).

En general, se tiene:

Rango: $-\infty - 1$

$KGE = 1$: Indica un ajuste perfecto del modelo a los datos observados.

$KGE > 0.75$: Indica un buen ajuste.

$0.5 < KGE \leq 0.75$: Indica un ajuste satisfactorio.

$0 \leq KGE \leq 0.50$: Indica un ajuste pobre.

$KGE < 0$: Indica un ajuste muy deficiente, donde las predicciones son peores que simplemente usar la media observada.

4.2 Técnicas y métodos en el área de investigación

4.2.1 SIG aplicado a la caracterización de hoyas hidrográficas

Hoy en día se pueden caracterizar las variables fisiográficas y los parámetros morfométricos de cualquier cuenca en el mundo, teniendo la información adecuada de Modelos de Elevación Digital (DEM por sus siglas en inglés), información de coberturas e información de tipos de suelos, todo lo anterior procesado con software especializado de Sistemas de Información Geográfica (SIG) (Aguirre, Torres, & Ruiz, 2003), (Braz, Garcia, Pinto, Chávez, & Oliveira, 2020) y (Camino, y otros, 2018).

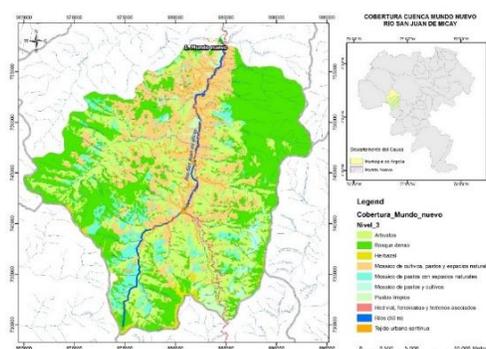


Figura 5 – Mapa de cobertura de una cuenca hidrográfica mediante SIG

Fuente. El Autor

La teledetección y análisis hidrográfico mediante clasificación supervisada es utilizada para establecer las divisorias de agua, tiempos de concentración, coeficientes de escorrentía, entre otros, incluso se han realizado estudios en donde se analiza la evaluación de la evolución temporal de algunos parámetros (Soria & Matar, 2016).

Para el proyecto se utilizará la licencia con la que cuenta actualmente el Departamento de Hidráulica para los programas de posgrado:



Figura 6 – Licencia de Arcmap 10.8.2

Fuente. Departamento de Hidráulica - FIC

4.2.2 Paquetes computacionales estadísticos en R

R es un lenguaje de programación y un entorno de software libre para el análisis estadístico y la visualización de datos, en donde se pueden implementar funciones estadísticas, desde análisis descriptivo hasta modelos complejos. Existen numerosos paquetes disponibles, también llamados librerías a través de CRAN (Comprehensive R Archive Network) que extienden las capacidades de R para realizar análisis específicos, gráficos avanzados, y manipulaciones de datos (R Core Team, 2024), proporcionando a los investigadores herramientas que permiten consultar datos y formular hipótesis de forma rápida (Guisande, Vaamonde, & Barreiro, 2011).

También existen los repositorios en donde programadores de todo el mundo comparten sus códigos para que cualquier pueda acceder a ellos y editarlos y orientarlos a sus datos según sus propias necesidades. Muchos paquetes y librerías de R se alojan en GitHub, donde los desarrolladores pueden colaborar, reportar problemas y sugerir mejoras. Los investigadores y científicos de datos utilizan GitHub para compartir scripts, datos y resultados, promoviendo la reproducibilidad y transparencia en la investigación (GitHub, Inc., 2024).

4.3 Experiencias con análisis de RLM entre morfometría e hidrología

4.3.1 Colombia

En el territorio nacional se han realizado algunos proyectos procesando información hidrológica para la estimación de caudales mediante modelos de regresión (Carvajal, Salazar, Mesa, & Poveda, 2015; Poveda, y otros, 2002; Rojo Hernández & Carvajal Serna, 2010), y quienes “han usado la regresión lineal múltiple para la predicción de caudales medios mensuales, ya sea utilizando una regresión general (que ajuste toda la muestra) o mediante una regresión periódica (que ajusta una regresión mensual). Como antecedente, en el trabajo del 2002 los modelos PREBEO (RLM + ONDITAS) y RLM fueron en su orden los de mayor habilidad de pronóstico puesto que exhiben los menores errores de predicción, para todos los ríos y todas las ventanas de pronóstico.

También se han realizado algunos estudios en donde han explorado la correlación de variables morfométricas e hidrológicas para la estimación de caudales, siendo la primera vez que se tenga registro un trabajo presentado en el XI Seminario Nacional de Hidráulica e Hidrología en 1994, en donde se recopiló y analizó toda la información hidrológica disponible del Departamento del Quindío, para hacer análisis de frecuencias de caudales medios, máximos y mínimos y para precipitaciones totales anuales. Resultaron regionalizaciones de los parámetros de las distribuciones de probabilidad ajustadas, de curvas de duración y algunas características de caudales mínimos (Vélez, Smith, Urrea, Pérez, & Franco, 1994).

Un estudio realizado en 2014 mostró los resultados del análisis hidrológico de las cuencas hidrográficas de los ríos Sinú, Aracataca y Fundación, en donde se evaluaron las características y la variabilidad de los aportes fluviales y las tendencias históricas para identificar patrones de escala regional en la dinámica fluvial. En este estudio se tomaron registros de caudales y transporte de sedimentos del IDEAM para realizar un análisis estadístico de las series de tiempo de transporte de sedimentos y caudales. Se identificó una alta dependencia del transporte de sedimento en función del caudal de las corrientes de las cuencas, mientras que las curvas de doble masa permitieron identificar la influencia de los factores externos en dicho transporte. La dinámica fluvial de las cuencas de Aracataca y Fundación mostró eventos máximos con períodos de retorno de 2 y 10 años, mientras que la dinámica fluvial del río Sinú se mantuvo regular. El estudio demostró correlaciones entre las variables de sedimentos y caudales máximos (López Ramos, Martínez Acosta, & Feria Díaz, 2014).

Vale la pena mencionar el trabajo desarrollado en una tesis de pregrado realizada en 2020, en la cual se demostró que, teniendo conocimiento del tipo de formación geológica, con la ayuda de gráficas que entrecruzan la geología y los parámetros morfométricos, se podrían establecer las propiedades de la cuenca (área, perímetro, pendiente y factor de forma) a partir de parámetros más simples como lo es la longitud, minimizando de esta manera los estudios de campo y procesamiento de cartografía, logrando mayor eficacia en el manejo de cuencas hidrográficas, estimación de caudales máximos para diferentes periodos de retorno, y la gestión del riesgo por inundaciones y avenidas torrenciales. Se evidenció una correlación superior a 0,90 entre parámetros morfométricos básicos de cuencas. Adicionalmente se buscó entrecruzar las demás variables, perímetro, factor de forma e índice de compacidad, sin encontrar con ellos afinidad alguna razonable, lo cual sugiere la necesidad de considerar variables adicionales como la precipitación. Además del uso de herramientas adicionales como otro software especializado en análisis matemático y estadístico (Espinosa Martínez, 2020).

Un trabajo presentado en el XXIV Seminario Nacional de Hidráulica e Hidrología en 2021, regionalizó curvas de duración de caudales para obtener caudales en cuencas sin registro en la zona Valle Medio de Magdalena, donde se emplearon cinco modelos de regresión (potencial, exponencial, logarítmico, cuadrático y cubico), con diversas variables dependientes como área de drenaje, perímetro de la cuenca, valor de precipitación media multianual, longitud del cauce principal y elevación media de la cuenca, obteniendo buenas métricas de rendimiento de los modelos (Donado, Romero-Duque, & Cortes-Ramos, 2021).

Para finalizar esta recopilación nacional, en el año 2019 bajo la modalidad de investigación se culmina un trabajo de grado denominado “Determinación de las curvas regionales para caudales máximos con diferentes periodos de retorno, teniendo en cuenta la variabilidad climática, en la región andina del Departamento del Cauca”, en donde se llegó a la apreciación, al igual que en el presente proyecto de que la gran mayoría de corrientes fluviales

de esta región se caracterizan por no tener estaciones de aforo y que la determinación de los caudales máximos hecha por métodos empíricos, presentan una alta dispersión de resultados por lo que su confiabilidad es discutible. La metodología seguida se desarrolló a partir del procesamiento de los registros de las estaciones de aforo existentes en la zona. Así, se obtuvieron los caudales máximos para diferentes períodos de retorno, pero corregidos teniendo en cuenta la variabilidad climática. Se relacionaron caudales máximos para diferentes periodos de retorno con el área de drenaje, generando envolventes, las curvas encontradas se ajustaron a las ecuaciones de Lowry y Creager para la región Andina caucana (Calvache Rengifo, 2019).

4.3.2 Internacionales

A nivel internacional también se han realizado algunos estudios con resultados favorables en el sentido de formular una metodología que permita zonificar ecuaciones para diferentes periodos de retorno teniendo en cuenta la correlación entre parámetros morfométricos y variables hidrológicas.

4.3.2.1 España

Un modelo de regresión múltiple permitió interpretar las diferencias en los caudales de inundaciones en las cuencas del Tajo y del Guadiana. Las relaciones estadísticas entre caudal y precipitación y adicionalmente los parámetros morfométricos de cada cuenca mostraron una significancia entre el relieve, la longitud de la corriente y las precipitaciones o la permeabilidad según el período de retorno considerado (Potenciano & Garzón, 2005).

4.3.2.2 Australia

Un estudio en el país oceánico confirmó el importante impacto de la topografía en la explicación de la distribución espacial de las estadísticas de diseño para las precipitaciones, los cual evidentemente siguiente la teoría de los métodos lluvia-escorrentía (Chow, Maidment, & Mays, 1988) modificaría el caudal. La efectividad de la elevación sobre variables topográficas más complejas reflejó el modesto relieve topográfico de la mayor parte del continente. No obstante, algunos parámetros inconclusos restantes indicaron la posibilidad de seguir desarrollando la dependencia topográfica en esas zonas. La complejidad de estas dependencias y la baja densidad general de las estaciones meteorológicas en esas zonas limitaron en cierta medida los progresos que pueden lograrse utilizando únicamente datos obtenidos sobre el terreno. El análisis también confirmó que la interpolación de parámetros de distribución de probabilidad es un medio eficiente de interpolación de cuantiles de precipitación sin pérdida de precisión por interpolación directa de cuantiles de precipitación. Los autores propusieron que los hallazgos y métodos propuestos aquí serán útiles para otros investigadores y profesionales que necesitan analizar la distribución espacial de las estadísticas de precipitaciones extremas para diseñar precipitaciones y estudios climáticos (Johnson, Hutchinson, The, Beesley, & Green, 2016).

4.3.2.3 Costa Rica

En un proyecto titulado “Funciones de regresión para caudales extremos en la vertiente Pacífica de Costa Rica”, en donde se obtuvieron las ecuaciones de regresión para diferentes probabilidades de excedencia aplicables a la vertiente Pacífica, halladas mediante la conjugación de la distribución de probabilidad Log-Person III y de regresión lineal y exponencial múltiple. Los parámetros que se relacionaron para la obtención de las ecuaciones fueron el caudal máximo anual, el área de drenaje y la precipitación promedio anual de las subcuencas establecidas, tomando como punto de cierre las estaciones limnigráficas. Los registros de caudal fueron aportados por el Instituto Costarricense de Electricidad (ICE) y los de precipitación por el Instituto Meteorológico Nacional (IMN) y el ICE. El análisis estadístico realizado mediante las series de datos de caudal máximo anual de las estaciones en estudio mostró buenos resultados. En consecuencia, se establece que existe una relación real entre las variables, debido a la confiabilidad de las muestras de datos y de los resultados obtenidos. Las ecuaciones generadas pueden ser utilizadas como un método empírico para el cálculo de caudal para los diferentes períodos de retorno analizados en la vertiente del Pacífico de este país (Rojas Morales, Aguilar Pereira, & Solís Bolaños, 2008).

4.3.2.4 Honduras

En el país centroamericano se encontraron dos proyectos muy interesantes siendo trabajos de grado de pregrado y maestría respectivamente. El primero denominado “Ecuaciones regionales de regresión múltiple para cálculo de frecuencia de caudales máximos en Honduras”, en donde se encuentran unas ecuaciones para el cálculo de caudal máximo que son aplicables a la región centro-occidental de Honduras, el resto del país no cuenta con estaciones hidrométricas con la suficiente información para emplear métodos de regresión. En cuanto al análisis estadístico, los coeficientes de determinación múltiple son bastante altos, por lo tanto, se describe gran parte de la variación de los caudales, pero se sabe que en regresiones múltiples esto no es suficiente para considerarlo como criterio principal. El estadístico F es alto en todas las ecuaciones, lo que significa que las variables escogidas son muy significativas y tienen muy poca probabilidad de que estos se deba al azar. Los errores típicos obtenidos son mesurados y se encuentran en el orden esperado para este tipo de ecuaciones. Estas ecuaciones no aplican para zonas urbanizadas, son estrictamente para cuencas rurales, ya que las características de las cuencas rurales y urbanas son muy diferentes (Velasquez).

Este trabajo se considera bastante interesante ya que presenta las ecuaciones para la estimación de caudales máximos para diferentes periodos de retorno y también una delimitación geográfica de donde aplicarlas, siendo más preciso el estudio al estar zonificado (Castro Heredia, Carvajal Escobar, & Ávila Díaz, 2012).

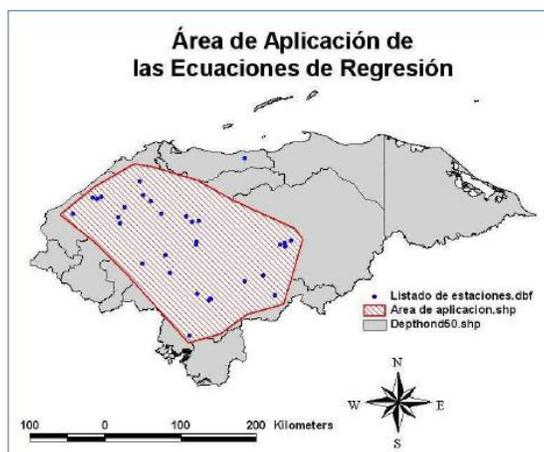


Figura 7 – Área de Aplicación de las Ecuaciones de Regresión

Fuente. (Velasquez)

Para el posgrado, se tiene el proyecto titulado “Análisis regional de crecidas máximas para Honduras”, en donde se presentan los resultados de la aplicación de un conjunto de técnicas tendientes a definir un criterio de regionalización de crecidas para cuencas en Honduras. Las regiones hidrológicas homogéneas se delimitaron inicialmente mediante un criterio de base geográfica e hidrológica, a partir del cual se segmentó al país en cinco regiones. Luego se conformó una compilación de información hidrométrica (caudales líquidos máximos medios diarios para 48 sitios) en base a las estaciones de aforo emplazadas en el país. Seguidamente se determinaron las ecuaciones que relacionan el área con el caudal índice para cada una de las regiones. Es de importancia mencionar que para las cinco ecuaciones obtenidas se obtuvo muy buenos coeficientes de determinación (r^2), por esta razón se puede asegurar el uso de estas para la determinación de caudales medios máximos en un punto dado conociendo el área de drenaje. Finalmente, como parte de la regionalización se determinaron las curvas de distribución de frecuencia; Normal, Gumbel, Pearson Tipo III, Log Normal, Log Gumbel y Log Pearson Tipo III para cada una de las regiones y para diferentes períodos de retorno, dichas curvas se presentan en gráficos adimensionales (Aguilera Sierra, 2010).

4.3.2.5 Chile

En el país austral se estudiaron 22 cuencas montañosas que abarcan diversas condiciones hidroclimáticas a lo largo de la Cordillera de los Andes semiáridos, este trabajo muestra que la consistencia hidrológica, cuantificada aquí a través de sesgos en las firmas de caudal, obtenida con algunas métricas de calibración (por ejemplo KGE que otorga el mismo peso a cada año hidrológico en la serie temporal de calibración) no garantiza pronósticos estacionales ESP satisfactorios y que las métricas que proporcionan pronósticos ESP habilidosos (por ejemplo, VE-Sep, que cuantifica los errores de volumen estacional) no necesariamente producen simulaciones de modelos hidrológicamente consistentes. Entre las opciones exploradas aquí, una función objetivo que combina la eficiencia Kling-Gupta (KGE) y la eficiencia Nash-Sutcliffe (NSE) con flujos en espacio logarítmico ofrece el mejor compromiso entre simulaciones hidrológicamente consistentes y rendimiento del pronóstico

retrospectivo. Finalmente, la elección de la métrica de calibración generalmente afecta la magnitud, más que el signo, de las correlaciones entre los atributos de calidad del pronóstico retrospectivo y los descriptores de la cuenca, siendo el índice de caudal base y la variabilidad del caudal interanual los mejores predictores de la habilidad del pronóstico. En general, este estudio destaca la necesidad de estrategias cuidadosas de estimación de parámetros en la cadena de producción de pronósticos para generar pronósticos habilidosos a partir de simulaciones hidrológicamente consistentes y extraer conclusiones robustas sobre la predictibilidad del caudal.

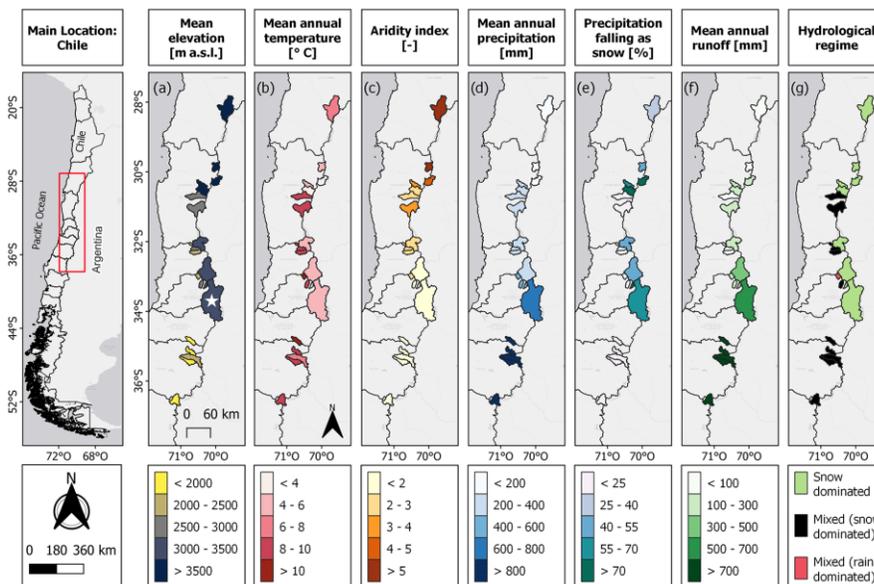


Figura 8 – Localización y variabilidad espacial de las características de la cuenca a lo largo del dominio de estudio: (a) elevación media, (b) temperatura media anual, (c) índice de aridez, (d) precipitación media anual, (e) precipitación que cae en forma de nieve, (f) escorrentía media anual y (g) régimen hidrológico

Fuente. (Araya, Mendoza, Muñoz-Castro, & McPhee, 2023)

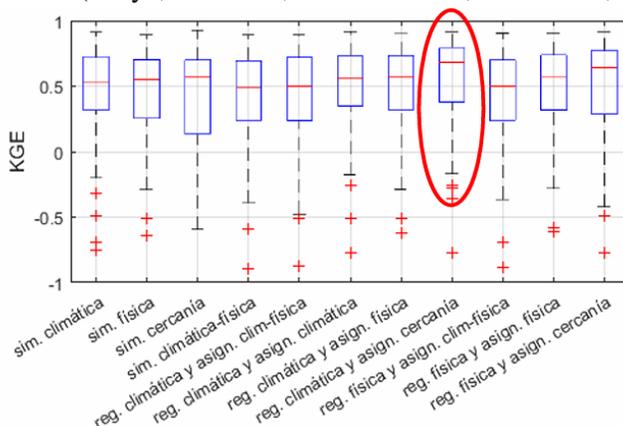


Figura 9 – Validación cruzada Macrozonas sur y parte de la austral

Fuente. (Araya, Mendoza, Muñoz-Castro, & McPhee, 2023)

4.4 Legislación asociada a estimación de caudales para diferentes T_r

Desde que se concibieron por primera vez los modelos hidrológicos que relacionan directamente la precipitación con la escorrentía con Perrault y Mariotte en el siglo XVII hasta la fecha, se siguen investigando y desarrollando modelos y optimizaciones al mismo concepto principalmente (Smith Quintero, y otros, 2001), introduciendo nuevas variables gracias a los modelos más robustos. Estos métodos son tema de discusión entre los investigadores, consultores, y público en general relacionado con el Recurso Hídrico, ya que se ha logrado demostrar la necesidad de calibración de dichos métodos a otras condiciones (Becerra Oviedo, Sánchez Mazorca, Acosta Castellano, & Díaz Arévalo, 2016) y que permitan replicarse para cualquier parte del mundo, por lo que se han venido generando manuales o metodologías que en resumen procuran simplificar métodos de cálculo más enfocados a las necesidades locales de cada país (Vargas M. & Diaz-Granados O., 1998).

A continuación, se resumen las principales legislaciones en el medio nacional e internacional asociada a estimación de caudales para diferentes T_r , desde el país de origen hasta países con condiciones similares a las de Colombia.

País	Año	Norma	Alcance
Colombia	2011	Resolución 0024 del 7 de enero del 2011	Se adopta el Manual de drenaje para carreteras – INVIAS, el cual cubre el diseño y construcción de obras típicas para el drenaje y subdrenaje en las carreteras nacionales
Paraguay	2019	MOC 14/11/2019	Se adopta el Manual de carreteras del Paraguay volumen 4.1 Diseño de drenaje vial y alcantarillas
Estados Unidos	2012	FHWA-HIF-12-003, Federal Highway Administration	Se adopta la metodología para el cálculo y estimación de caudales máximos en cauces naturales
España	2005	Ley 11/2005, de 22 de junio	Se modifica la ley 10/2001 del Plan Hidrológico Nacional
Portugal	2015	Decreto-Lei n.º 218/2015	Normas de qualidade ambiental no domínio da política da água

Tabla 6 – Legislación relacionada con estimación de caudales máximos

Fuente. El Autor

5. Materiales y métodos

Teniendo en cuenta que esta investigación se fundamenta en la estimación de las características morfométricas de cuencas hidrográficas y de la medición y procesamiento de datos de fenómenos naturales, y que adicionalmente se busca expresar una relación entre dichas variables estudiadas de forma deductiva, generalizando y normalizando resultados para una región del Cauca, Colombia, de acuerdo con (Fermini, 2019) la investigación es cuantitativa.

Debido al propósito principal del estudio que busca desarrollar un modelo para estimación de caudales máximos a partir de parámetros morfométricos e hidrológicos en cuencas hidrográficas del Departamento del Cauca, Colombia. El alcance de la investigación es **correlacional**.

Hipótesis: Existe una correlación significativa entre los parámetros morfométricos e hidrológicos en cuencas hidrográficas del Departamento del Cauca, Colombia, lo cual permite estimar mediante una ecuación caudales máximos para diferentes periodos de retorno empleando la regresión lineal múltiple.

5.1 Descripción del lugar de estudio

El Departamento del Cauca, es uno de los treinta y dos departamentos que, junto con Bogotá, Distrito Capital, componen el territorio la República de Colombia. Se localiza en el suroeste del país y sus territorios hacen parte de las regiones Andina y Pacífica. Cuenta con una superficie 29.308 Km², lo que representa el 2.56 % del territorio nacional. Su capital es la ciudad de Popayán y está dividido política y administrativamente en 42 municipios (La Sociedad Geográfica de Colombia – Academia de Ciencias Geográficas, 2015).



Figura 10 - Localización Departamento del Cauca en Colombia

Fuente. El Autor

El Departamento del Cauca está bañado por cinco grandes cuencas: Alto Cauca, Pacífico, Alto Magdalena, Patía y Caquetá (IGAC - Instituto Geográfico Agustín Codazzi, 2023).

La cuenca del Alto Cauca está conformada por el Río Cauca y sus afluentes, los Ríos Palo, Guengué, Negro, Teta, Desbaratado, Quilichao, Mondomo, Ovejas, Pescador, Robles, Piedras, Sucio, Palacé, Cofre, Honda, Cajibío, Piendamó, Tunia, Molino, Timbío y Blanco.

La cuenca del Pacífico: está conformada principalmente por los Ríos Guapi, Timbiquí, Saija y Micay.

La cuenca del Alto Magdalena: tiene como su principal fuente de drenaje al Río Páez al que confluyen los Ríos San Vicente, Moras, Ullucos, Negro y Negro de Narváez, y las quebradas Tóez, Símbola, Salado, Gualcar, Gallo, Macana, Honda y Totumo.

La cuenca del Patía: está conformada por el Río Patía y sus tributarios: los Ríos Guachinoco, Ismita, Bojoleo, El Guaba, Sambingo y Mayo.

La cuenca del Caquetá: está conformada principalmente por el Río Caquetá, al cual confluyen los Ríos Cusiyaco, Cascabelito, Verdeyaco, Mandiyaco, Fragua, Cascabel, Curiaco y Pacayaco.

El Clima del Departamento del Cauca está determinado principalmente por su relieve y posición geográfica, dando como resultado que se presenten todos los pisos térmicos desde el clima cálido superhúmedo que se encuentra en la costa pacífica caucana hasta las nieves perpetuas del Nevado del Huila. En la vertiente occidental se presentan climas cálidos húmedos y superhúmedos. Hacia la cuenca del Patía el clima es de tipo cálido semiárido. Sobre la parte central predomina el clima templado húmedo y semihúmedo. Los climas fríos se registran a lado y lado de la meseta de Popayán.

La distribución de lluvias en el Departamento del Cauca registra una gran variabilidad territorial. Las mayores lluvias ocurren hacia la vertiente del Pacífico en los municipios de López, Timbiquí y Guapi, con volúmenes superiores a los 7000 mm anuales. Los menores valores se registran en sectores de la cuenca del Patía, en municipios como Rosas, El Bordo y La Fonda, entre otros, con cantidades anuales menores a los 1000 mm en promedio. El régimen de lluvias predominante en el departamento es bimodal con una temporada seca muy marcada hacia los meses de junio, julio y agosto y una temporada lluviosa principal en los meses de octubre, noviembre y diciembre. Adicionalmente, se registran dos temporadas secundarias: en los meses de enero y febrero las lluvias disminuyen y en el periodo de marzo a mayo, se incrementan. En la vertiente del Pacífico, no existe una temporada seca definida y las lluvias predominan durante todo el año (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2023).

Respecto a la temperatura, en el Departamento del Cauca, se registra el piso térmico cálido en la totalidad de la vertiente Pacífica y el valle del Patía. En el resto del territorio, predomina el efecto de la altura y la temperatura disminuye en función de la elevación.

5.1.1 Descripción asociada a la investigación

Para seleccionar las estaciones que se utilizaron en el presente estudio, se realizó una búsqueda y clasificación de la información climatológica e hidrológica disponible en la base de datos del IDEAM (ver Anexo 1), sobre las cuales se filtraron las estaciones, activas y vigentes en la actualidad, o inactivas que dejaron de capturar datos hace relativamente poco, y teniendo en cuenta que se requieren aplicar distribuciones de probabilidad, en ambos casos se estableció la empleabilidad de la estación si cuenta con datos que comprenden un período mínimo de 15 años consecutivos, con una antigüedad del dato recopilado que no supere los últimos 20 años o que ocupe un período de 30 años consecutivos con una antigüedad del dato recopilado cuyo año más reciente no tenga antigüedad superior a 15 años (Arumi R., Jara R., & Salgado S., 2000). Por lo general las estaciones seleccionadas para el estudio disponen de series históricas de más de 20 años, aumentando la calidad de los análisis estadísticos y produciendo resultados más robustos y concluyentes (Sánchez San Román, 2022).

Después de verificar la cantidad y calidad de los datos suministrados por la entidad mediante radicados, versus la información climatológica e hidrológica disponible en la base de datos del IDEAM, se establecieron las fechas y rangos de tiempo donde se tienen realmente la información disponible y que es utilizable en el presente proyecto, adicionalmente se evaluó la necesidad de realizar procesos de homogenización, interpolación y extensión de datos en donde sea necesario. Resultado de los anteriores procesos se seleccionaron las siguientes estaciones con medidas de $Q_{Máx}$, ya sea Limnigráficas (LG) o Limnimétricas (LM).

#	Código	Nombre	Tipo	Latitud	Longitud	Inicio	Final	#Años
1	53077010	ANGOSTURA RIO MICAY	LG	2.7991	-77.2561	1959	2013	54
2	44017140	BETANIA- VILLALOBOS	LG	1.4328	-76.4789	1997	2015	18
3	26047020	BOCAT RIO PALO	LG	3.0620	-76.3488	1946	2019	73
4	52017070	BOCATOMA SAJANDI	LG	2.2167	-77.0575	1995	2021	26
5	26017110	BOCATOMA- VINAGRE	LM	2.4164	-76.5088	1977	2011	34
6	26027250	BUENOS AIRES	LG	3.0368	-76.5792	1976	2019	43
7	26027090	CORTIJO EL	LG	2.6234	-76.3586	1961	2022	61
8	44017090	CURIACO	LG	1.7050	-76.6240	1982	2022	40
9	57027010	GORGONA LA PLAYA	LM	2.9639	-78.1742	1996	2020	24
10	52027010	GUACHICONO	LG	2.0464	-77.0031	1992	2023	31
11	53077020	GUAYABAL	LM	2.9936	-77.4914	1990	2022	32
12	52017120	HOYO EL	LG	2.2085	-77.0093	1995	2020	25
13	26017020	JULUMITO	LG	2.4676	-76.6443	1964	2006	42
14	52017030	LA FONDA - AUT	LG	2.1478	-77.0467	1974	2019	45
15	26017070	LOMITAS	LG	2.1893	-76.4818	1970	2023	53
16	26027240	MALVASA	LG	2.4708	-76.3368	1961	2023	62

17	53077060	MONO EL	LM	2.9201	-77.2237	1991	2021	30
18	26017080	PALETARA	LM	2.2019	-76.4921	1971	2022	51
19	44017060	PAPAS	LG	1.8844	-76.6928	1981	2022	41
20	52027020	PLAYA LA	LG	1.9670	-77.0089	1995	2018	23
21	26017060	PUENTE ARAGON - AUT	LG	2.2000	-76.5000	1970	2022	52
22	26027200	PUENTE CARRETERA - AUT	LM	2.5368	-76.5584	1970	2023	53
23	52017110	PUENTE CARRETERA - AUT	LG	2.3300	-76.8500	1995	2017	22
24	26017040	PUENTE CARRETERA	LM	2.4477	-76.5140	1969	2023	54
25	26027100	PUENTE CARRETERA	LG	2.6259	-76.5220	1969	2023	54
26	26037030	PUENTE CARRETERA	LM	2.4800	-76.7300	1974	2023	49
27	52017080	PUENTE CARRETERA	LG	1.9722	-77.2100	1995	2021	26
28	52017100	PUENTE CARRETERA	LG	2.3600	-76.8500	1995	2018	23
29	52017130	PUENTE COLGANTE TIMBIO	LG	2.2181	-77.0044	1995	2023	28
30	26027210	PUENTE FERROCARRIL	LG	3.0380	-76.5118	1970	2021	51
31	52027030	PUENTE FIERRO - GUACHICONO	LG	2.1565	-76.8990	1995	2022	27
32	21057030	PUENTE RICAURTE	LG	2.5229	-75.9617	1964	2019	55
33	26037010	REMOLINO	LG	2.5146	-76.8313	1966	2021	55
34	21057130	RIO CHIQUITO	LG	2.6657	-75.7757	1997	2019	22
35	44017050	RIOGRANDE - CAUCA	LG	1.7050	-76.6436	1983	2022	39
36	21057120	SAN JUAN	LM	2.3519	-76.3131	2000	2023	23
37	53047010	SANGARAL	LG	2.5074	-77.6933	1988	2023	35
38	44017070	SANTA ROSA - AUT	LG	1.7000	-76.5667	1982	2023	41
39	26057040	TIMBA	LG	3.1171	-76.6240	1946	2023	77
40	26027080	TOTORO	LM	2.5125	-76.4038	1961	2021	60
41	21057110	TREBOL EL	LM	2.2500	-76.1285	1994	2023	29

Tabla 7 – Listado de estaciones LG/LM empleadas en la investigación

Fuente. El Autor

En contraste con la base de datos del IDEAM se descartan seis estaciones por los siguientes motivos:

BOCATOMA-FLORIDA [26017030], a pesar de aparecer activa desde 1968, solo fue posible obtener datos desde 1999 hasta el 2009, es decir únicamente 10 años de registro, siendo muy pocos datos para hacer un análisis estadístico.

CAUCA ALTO - AUT [2602700148], los datos no fueron suministrados por la entidad, sin embargo, al ser instalada en el 2017, cuenta con muy pocos datos aun para hacer un análisis estadístico.

MOLINOS - AUT [2601700104], los datos no fueron suministrados por la entidad, sin embargo, al ser instalada en el 2017, cuenta con muy pocos datos aun para hacer un análisis estadístico.

PUENTE COHETANDO - AUT [2105700099], al ser instalada en el 2017, cuenta con muy pocos datos aun para hacer un análisis estadístico. En el mismo punto, funciono la estación homónima LM con código 21057070 desde 1980 hasta la creciente del Río Páez en 1994 (NASA KIWE; Instituto Colombiano de Geología y Minería INGEOMINAS, 2009), estando desactualizados cerca de 30 años.

PUENTE RICAURTE 2 [2105700152] y REMOLINO – AUT [2603700157], ambas instaladas en el 2017, pero registrando datos a partir del 2021, cuentan con muy pocos datos aun para hacer un análisis estadístico. Se aclara que en la lista de estaciones seleccionadas aparecen sus homónimas, que se encontraban hasta hace poco (2019 y 2021) en puntos cercanos, siendo PUENTE RICAURTE [21057030] y REMOLINO [26037010]. Estas últimas tienen más de 50 años de registros, por lo tanto, si se tuvieron en cuenta en la investigación.

Para la variable $PM_{\max 24h}$ se seleccionaron las siguientes estaciones, ya sea Climática Ordinarias (CO), Climáticas Principales (CP), Pluviográficas (PG), Pluviométricas (PM) y Sinópticas Principales (SP).

#	Código	Nombre	Tipo	Latitud	Longitud	Año inicial	Año final	#Años
1	26035030	AEROPUERTO G L VALENCIA	SP	2.4529	-76.6088	1941	2023	82
2	53045040	AEROPUERTO GUAPI - AUT	CP	2.5744	-77.8948	1981	2023	42
3	26040290	AGUADA LA	PM	2.8805	-76.4075	1973	2023	50
4	21050250	ALTAMIRA	PM	2.5010	-76.0952	1975	2022	47
5	26020200	AMPARO EL	PM	2.9612	-76.4892	1970	2023	53
6	52010020	BALBOA	PM	2.0328	-77.2217	1954	2023	69
7	21050070	BELALCAZAR	PM	2.6570	2.6570	1959	2022	63
8	53040010	BOCAS DE NAPI	PM	2.4857	-77.7730	1981	2023	42
9	53060010	BOCAS DE PATIA	PM	2.7888	-77.4662	1981	2023	42
10	52025010	BOLIVAR	CO	1.8299	-77.0040	1970	2023	53

11	26020100	BUENOS AIRES	PM	3.0202	-76.6391	1966	2023	57
12	26030030	CABANA INDERENA	PM	2.6407	-76.9090	1984	2019	35
13	26060200	CAJONES	PM	3.2264	-76.1408	1975	2020	45
14	26020120	CALOTO	PM	3.0301	-76.4067	1966	2023	57
15	26020160	CATALINA LA	PM	2.9452	-76.6518	1962	2023	61
16	26010020	COCONUCO	PM	2.3434	-76.4955	1947	2023	76
17	53080020	CONCHA LA	PM	3.1674	-77.1357	1981	2023	42
18	26040250	CORINTO	PM	3.1791	-76.2477	1972	2015	43
19	44010140	CRISTALINA LA	PM	1.5765	-76.5356	1997	2023	26
20	26030070	DINDE	PM	2.6974	-76.7547	1975	2022	47
21	21050310	ESCUELA RIOSUCIO 2	PM	2.4468	-76.1689	1984	2022	38
22	52025080	ESTRECHO PATIA - AUT	CO	1.9600	-77.1200	1971	2023	52
23	52010190	ESTRELLAS LAS	PM	2.1700	-76.6100	1994	2023	29
24	26025070	GABRIEL LOPEZ	CO	2.5100	-76.2900	1971	2023	52
25	53070090	GALLERA LA	PM	2.7451	-76.9053	1987	2022	35
26	26030060	GAMBOA	PM	2.6256	-76.8536	1974	2022	48
27	57025020	GORGONA GUAPI	CP	2.9629	-78.1744	1986	2021	35
28	52025060	GRANJA EXPERIMENTAL UNIVERSIDAD DE NARIÑO	CP	1.9100	-77.1900	1992	2023	31
29	52020010	GUACHICONO	PM	2.0464	-77.0031	1971	2023	52
30	52020190	GUACHICONO	PG	2.0300	-76.6734	1994	2023	29
31	53070060	HONDURAS	PM	2.6396	-77.1549	1960	2023	63
32	26045010	INGENIO BENGALA	CO	3.2628	-76.4145	1965	2022	57
33	21055070	INZA - AUT	CP	2.5482	-76.0639	2005	2021	16
34	26020470	JAPIO	PM	3.0484	-76.4414	2004	2023	19
35	52025090	LA SIERRA - AUT	CP	2.1938	-76.7503	1971	2023	52
36	26025110	LOMITAS	CO	3.0754	-76.5563	2003	2023	20
37	53070020	MECAJE	PM	2.4833	-77.1000	1976	2019	43
38	52010160	MESA LA	PM	2.2616	-77.1196	1992	2020	28
39	52025020	MILAGROS LOS	CO	1.7533	-76.8917	1972	2023	51
40	26065020	MIRANDA	CO	3.2457	-76.2231	1964	2023	59
41	26020250	MONDOMO	PM	2.9037	-76.5470	1971	2023	52
42	26020220	MORALES	PM	2.7694	-76.6242	1971	2023	52
43	53070050	NOANAMITO	PM	3.0771	-77.5361	1981	2023	42

44	26020390	OVEJAS ABAJO ALERT	PM	2.8538	-76.5886	1979	2022	43
45	52015010	PAISPAMBA	CO	2.2497	-76.6119	1994	2023	29
46	52010050	PARRAGA	PG	2.2323	-76.7879	1967	2023	56
47	52010180	PATIA	PM	2.0686	-77.0529	1994	2023	29
48	26020030	PIENDAMO	PM	2.6427	-76.5361	1947	2023	76
49	26020460	POLINDARA	PM	2.4876	-76.4047	1984	2023	39
50	26010100	PUENTE ARAGÓN	PM	2.1916	-76.4823	2007	2022	15
51	53070030	PUERTO LOPEZ	PM	2.8453	-77.2490	1960	2023	63
52	26010030	PURACE	PM	2.3809	-76.4548	1959	2023	64
53	52020180	QUEBRADILLAS	PG	1.9510	-76.7329	1995	2023	28
54	52010100	QUILCASE	PG	2.3322	-76.8487	1974	2021	47
55	26030090	REMOLINO	PM	2.5150	-76.8297	1981	2023	42
56	26040310	RIO PALO	PM	3.0706	-76.3778	1964	2023	59
57	52020050	RODEO EL	PM	1.8802	-76.9916	1967	2023	56
58	26030080	ROMELIA LA	PM	2.6228	-76.8951	1976	2019	43
59	26020230	ROSARIO EL	PM	2.6066	-76.7561	1971	2023	52
60	52010040	ROSAS	PM	2.2624	-76.7378	1958	2023	65
61	53060020	SAIJA	PM	2.8718	-77.6217	1966	2023	57
62	52010080	SAJANDI	PM	2.1871	-77.0497	1958	2023	65
63	52010110	SALADITO	PM	2.3346	-76.6201	1990	2023	33
64	26030150	SALVAJINA LA	PM	2.9433	-76.7069	2008	2023	15
65	21050230	SAN ANDRES	PM	2.5818	-76.0468	1975	2023	48
66	21050320	SAN JUAN	PM	2.3434	-76.3084	2002	2023	21
67	21050220	SAN LUIS	PM	2.5659	-75.9279	1975	2023	48
68	21055030	SANTA LETICIA	CO	2.2400	-76.1600	1971	2023	52
69	44010100	SANTA ROSA	PG	1.6948	-76.5712	1983	2023	40
70	21050240	SANTA TERESA	PM	2.4606	-76.0542	1975	2023	48
71	52020070	SANTIAGO	PM	1.8142	-76.7829	1970	2023	53
72	26020180	SATE	PM	2.4963	-76.6384	1970	2023	53
73	26050270	SILENCIOSO EL	PM	3.1156	-76.7194	1972	2023	51
74	26020020	SILVIA PTA ELECTRI	PM	2.6247	-76.3491	1947	2022	75
75	26050060	SUAREZ	PM	2.9546	-76.6961	1952	2023	71
76	26030050	TAMBO	PM	2.4560	-76.8111	1953	2023	70
77	26020320	TERMALES PILIMBALA	PM	2.3631	-76.4041	1970	2023	53
78	53050020	TIMBIQUI	PM	2.7736	-77.6636	1967	2023	56
79	26040260	TORIBIO ALERTAS	PM	2.9524	-76.2622	1974	2023	49
80	26020130	TOTORO	PM	2.5418	-76.4188	1956	2023	67

81	26025100	TUNIA	CO	2.6745	-76.5263	1989	2023	34
82	44015030	VALENCIA	CP	1.9004	-76.6688	1971	2023	52
83	52020020	VEGA LA	PM	2.0050	-76.7817	1971	2023	52
84	26025090	VENTA DE CAJIBIO	CO	2.5848	-76.5618	1973	2023	50
85	26060030	VILLA RICA	PM	3.1789	-76.4630	1969	2023	54
86	44010130	VILLALOBOS	PM	1.4283	-76.4789	1997	2022	25

Tabla 8 – Listado de estaciones CO/CP/PG/PM/SP empleadas en la investigación

Fuente. El Autor

En contraste con la base de datos del IDEAM se descartan 12 estaciones por los siguientes motivos:

ARRAYANALES - AUT [26015040], los datos no fueron suministrados por la entidad, sin embargo, al ser instalada en el 2010, cuenta con muy pocos datos aun para hacer un análisis estadístico.

BALBOA - AUT [52015050], a pesar de aparecer activa desde 2005, solo fue posible obtener datos del 2013, es decir únicamente 1 año de registro, siendo muy pocos datos para hacer un análisis estadístico.

CORINTO - AUT [2604500039], los datos no fueron suministrados por la entidad, sin embargo, al ser instalada en el 2017, cuenta con muy pocos datos aun para hacer un análisis estadístico.

EL DIVISO - AUT [53075020], a pesar de aparecer activa desde 2005, los datos no fueron suministrados por la entidad.

EL TABLAZO - AUT [26015010], a pesar de aparecer activa desde 2005, los datos no fueron suministrados por la entidad.

GABRIEL LOPEZ [26020190], a pesar de aparecer activa desde 1970, solo fue posible obtener datos desde 1970 hasta 1975, es decir únicamente 5 años de registro, siendo muy pocos datos para hacer un análisis estadístico, además de desactualizada.

MONDOMO [26020360], a pesar de aparecer activa desde 1985, los datos no fueron suministrados por la entidad.

PALETARA - AUT [26015050], los datos no fueron suministrados por la entidad, sin embargo, al ser instalada en el 2011, cuenta con muy pocos datos aun para hacer un análisis estadístico.

PNN PURACE - AUT [26015030], a pesar de aparecer activa desde 2007, solo fue posible obtener datos del 2009, es decir únicamente 1 año de registro, siendo muy pocos datos para hacer un análisis estadístico.

REMOLINO ALERTAS - EL VOLCAN - AUT [2603500127], los datos no fueron suministrados por la entidad, sin embargo, al ser instalada en el 2017, cuenta con muy pocos datos aun para hacer un análisis estadístico.

SANTIAGO - AUT [5202500132], los datos no fueron suministrados por la entidad, sin embargo, al ser instalada en el 2017, cuenta con muy pocos datos aun para hacer un análisis estadístico.

TORIBIO ALERTAS TORIBIO - AUT [2604500134], los datos no fueron suministrados por la entidad, sin embargo, al ser instalada en el 2017, cuenta con muy pocos datos aun para hacer un análisis estadístico.

De acuerdo con la tabla 7, en el Departamento del Cauca existen 41 estaciones con datos hidrológicos (Color rojo) entre LG y LM (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2023), y a cada una le corresponderá su respectiva hoya cuenca hidrográfica, tal como se muestra a continuación:

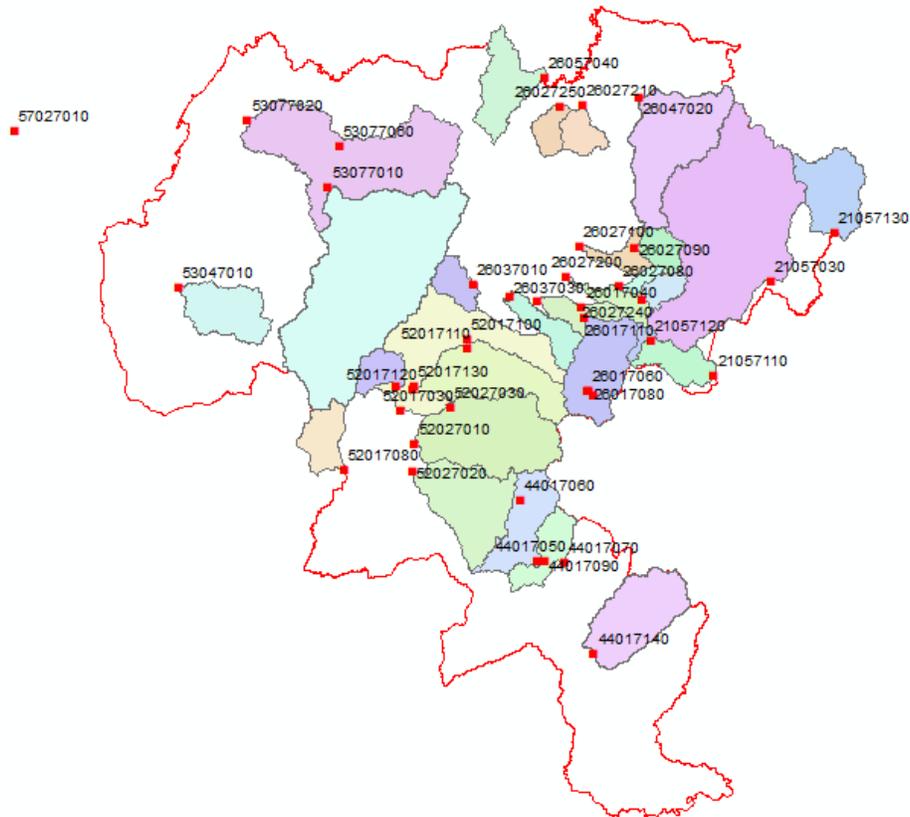


Figura 11 – Estaciones IDEAM con datos de Caudales en el Departamento del Cauca

Fuente. El Autor

De acuerdo con la tabla 8, en el Departamento del Cauca existen 86 estaciones con datos de precipitación máxima en 24 horas (Color verde), entre CO, CP, PG, PM y SP (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2023), tal como se muestra a continuación:

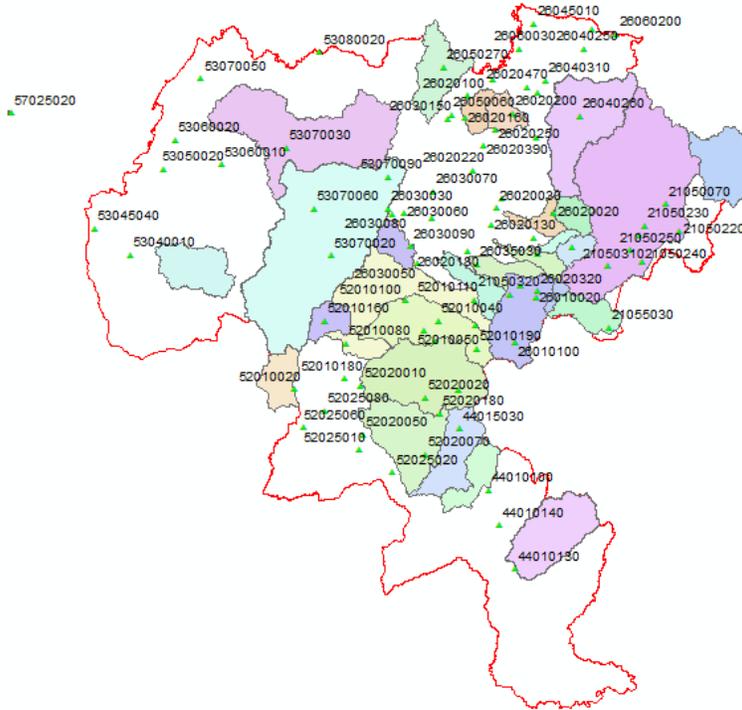


Figura 12 – Estaciones IDEAM con datos de Precipitaciones Máximas en 24 horas en el Departamento del Cauca

Fuente. El Autor

Con las dos anteriores figuras como base, se presenta a continuación el mapa del Departamento del Cauca, con las estaciones LG y LM (■) y también con las estaciones CO, CP, PG, PM y SP (▲):

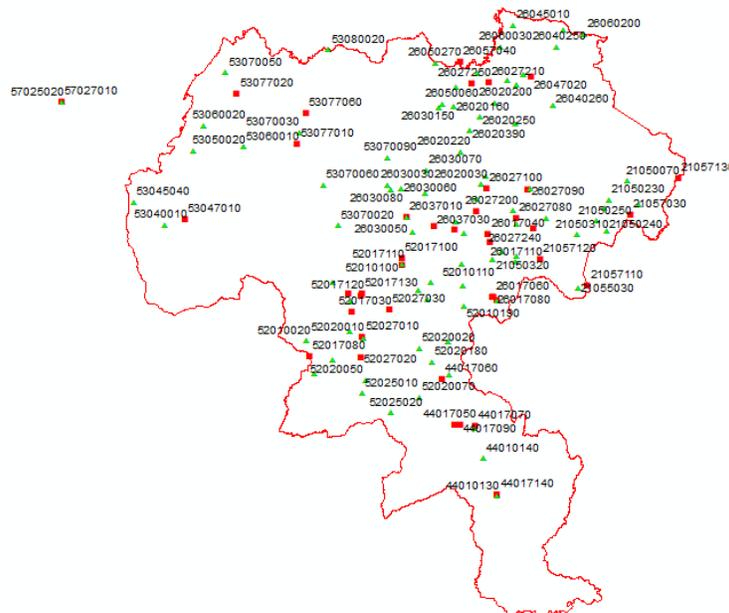


Figura 13 – Estaciones IDEAM con datos de Caudales y Precipitaciones Máximas en 24 horas en el Departamento del Cauca

Fuente. El Autor

Teniendo en cuenta que la superficie del departamento es aproximadamente 29.308 Km² (La Sociedad Geográfica de Colombia – Academia de Ciencias Geográficas, 2015) y que la extensión intersecada de las cuencas hidrográficas del presente estudio es de aproximadamente 15.378 Km². La presente investigación estudia aproximadamente el 52.5% del Cauca, tal como indica la siguiente figura:

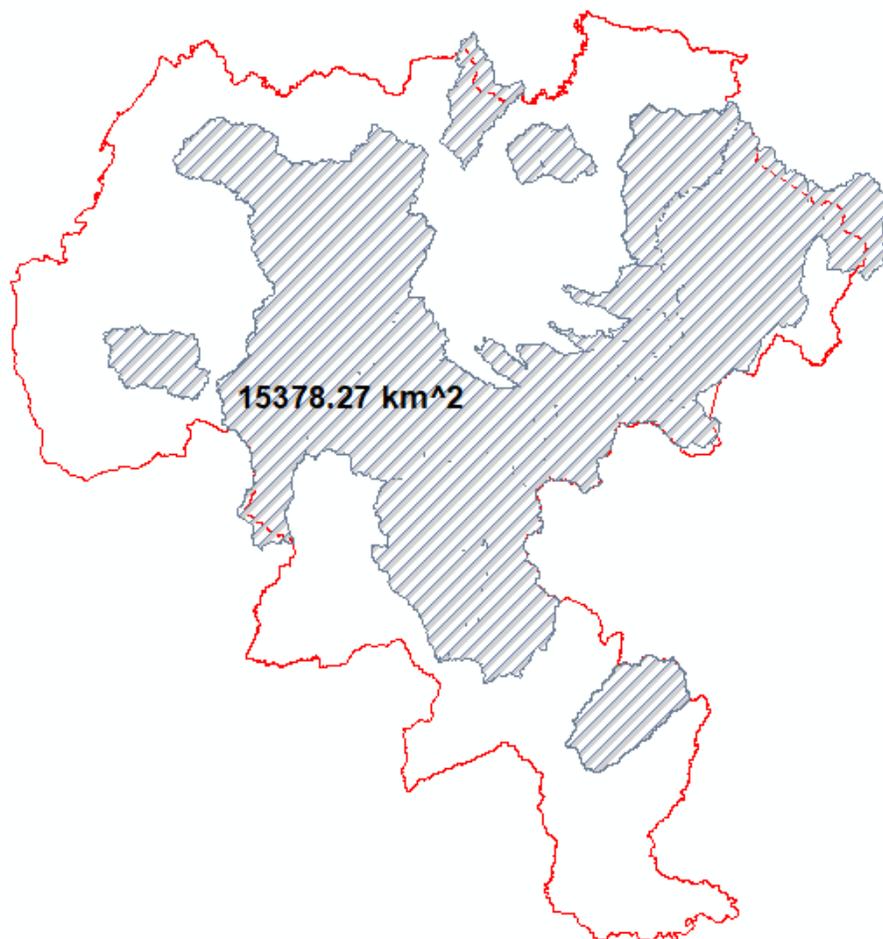


Figura 14 – Extensión intersecada de las cuencas hidrográficas del presente estudio versus Departamento del Cauca

Fuente. El Autor

5.2 Metodología

La metodología utilizada en el desarrollo del presente proyecto corresponde al siguiente diagrama de flujo:

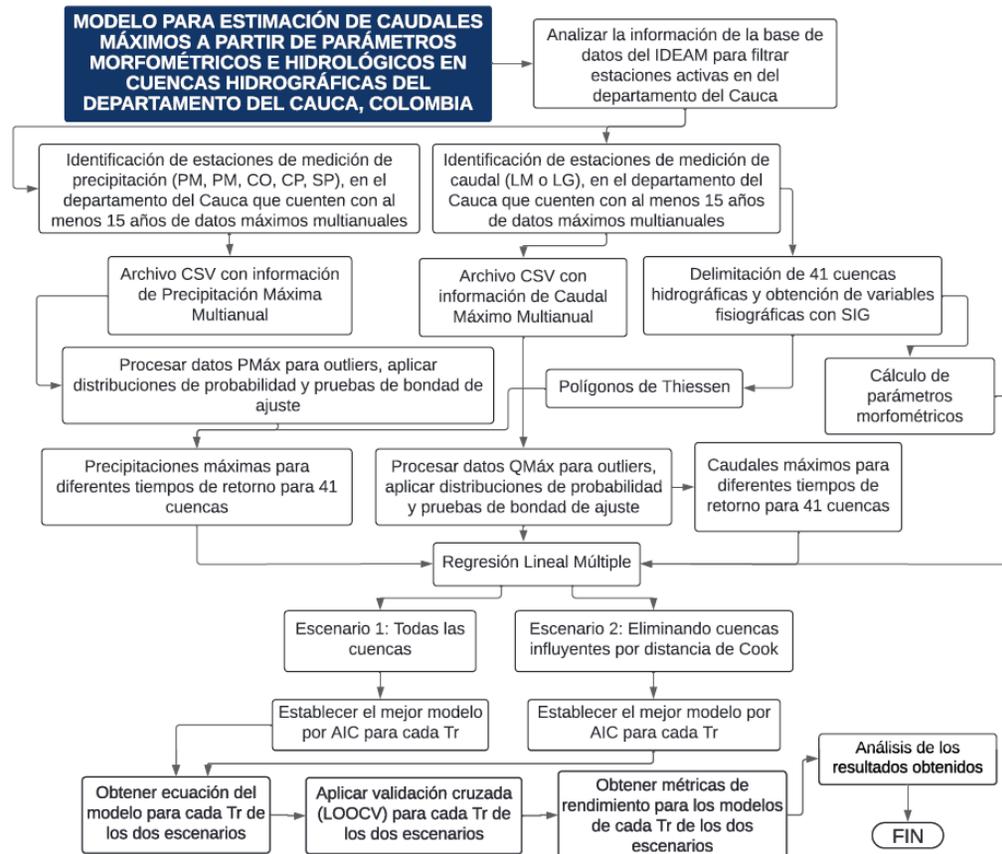


Figura 15 – Diagrama de flujo metodología implementada

Fuente. El Autor

A continuación, se tratan, el Sistema de recolección de información y el Sistema de análisis de información, respectivamente, para cada uno de los objetivos específicos.

5.2.1 Objetivo específico 1

Identificar mediante una revisión bibliográfica detallada, las diferentes metodologías para el desarrollo de modelos para estimación de caudales máximos en diferentes zonas o territorios nacionales e internacionales.

Revisión bibliográfica en las principales bases de datos académicas, como Scopus, Science Direct, Google Scholar e IAHR Journals de métodos para generación de ecuaciones de caudales máximos teniendo en cuenta variables fisiográficas, parámetros morfométricos y variables hidrológicas (Universidad del Cauca, 2024; IAHR - The International Association for Hydro-Environment Engineering and Research, 2024), siendo relevantes palabras como: Caudales máximos, cuencas hidrográficas, parámetros morfométricos, regresión lineal múltiple, SIG, validación cruzada.

Lo anterior permitirá sistematizar las metodologías existentes y evaluar las similitudes en cuanto tipo y cantidad de datos para evaluar la posibilidad de, a partir los métodos existentes formular la metodología aplicable al Departamento del Cauca, siendo detectadas variables independientes iniciales para la exploración de datos relacionadas con variables fisiográficas, parámetros morfométricos y variables hidrológicas.

5.2.2 Objetivo específico 2

Calcular las variables fisiográficas y parámetros morfométricos mediante SIG de 41 cuencas instrumentadas con estaciones de medición de caudal por el IDEAM en el Departamento del Cauca.

Modelo digital de elevación Alaskan Satellite Facility (University of Alaska Fairbanks, 2023), satélite ALOS PALSAR beam mode FBS, Hi-Res Terrain Corrected. Tamaños del píxel 12.5x12.5m, año 2023. Imagen satelital ESA Copernicus (European Space Agency (ESA), 2023) y/o Earth Explorer (U.S. Geological Survey (USGS), 2023) satélite Sentinel-2, plataforma S2A_ y tipo de producto S2MSI2A con segundo nivel de procesamiento, con correcciones atmosféricas y geométricas. Año 2023. Información descargada de MERIT-Basins en donde se encuentra información base en formato vector para la hidrografía global derivada del 90m "MERIT-Hydro Basin Scale rasters" producto utilizando la herramienta TauDEM con las siguientes características: Definiciones corregidas/refinadas de cuenca/región y ~2,94 millones de tramos fluviales (captaciones unitarias) (Lin, y otros, 2019; Yamazaki, y otros, 2019; Heberger, 2023). Información existente de análisis de coberturas y categorías de suelo y geología, solicitados a las entidades gubernamentales. (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2012; IGAC - Instituto Geográfico Agustín Codazzi, 2017), año 2023.

Con los anteriores insumos fue posible realizar el procesamiento de MDT para generación de curvas de nivel en el software Arcmap 10.8.2, para realizar la caracterización morfométrica de las cuencas hidrográficas a estudiar. (Delimitación y características físicas

de las cuencas hidrográficas), también se realizó el procesamiento de imágenes satelitales e información de suelos software Arcmap 10.8.2, para realizar la caracterización de coberturas, y suelos de las cuencas hidrográficas a estudiar, que junto con el mapa de pendientes originado de los diferentes DEM, permitía caracterizar completamente los coeficientes de escorrentía y los números de curva de escorrentía de cada cuenca hidrográfica.

Finalmente fue posible la estimación de las variables fisiográficas y parámetros morfométricos (Méndez, Córdova, & Bravo de Guenni, 2015) de las 41 cuencas del estudio mediante una hoja de cálculo creada por el autor.

El anterior objetivo está estrechamente relacionado con dos resultados esperados del proyecto, el primero consiste en generar trabajos de investigación que puedan ser desarrollados por estudiantes de pregrado, y que sean parte del Semillero HidroInnovación, en este objetivo se hace énfasis en el proyecto denominado “Estudio y caracterización del C y CN para el Departamento del Cauca: Correlación e implementación”, el cual tendrá como finalidad, generar cartografía para el coeficiente de escorrentía y el número de curva de escorrentía para todo el Departamento del Cauca. Se ofertará la información en el portal de Arcgis Online con el nombre: "Mapa de coeficiente de escorrentía y número de curva de escorrentía del Departamento del Cauca 2023", para que cualquier usuario (después de un breve registro) pueda consultar en línea la información de manera clara y obtener el cálculo de dichos coeficientes al realizar la intersección con cualquier cuenca en el Departamento del Cauca.

El segundo el cual consiste en escribir un manuscrito titulado “SIG para aplicaciones hidrológicas - Guía de ArcMap: Procedimientos para el cálculo de variables fisiográficas y parámetros morfométricos de cuencas hidrográficas” el cual será coautoría del Semillero de Investigación HidroInnovación, y será a revisión en la Convocatoria del Área de Desarrollo Editorial para la edición de libros de texto 2025-2026, en junio de 2025, en donde se resume de manera técnica todos los procedimientos paso a paso necesarios para una correcta caracterización de cualquier cuenca hidrográfica.

5.2.3 Objetivo específico 3

Estimar precipitaciones y caudales máximos para diferentes periodos de retorno aferentes a las 41 cuencas hidrográficas mediante distribuciones de probabilidad y pruebas de bondad de ajuste.

Datos máximos diarios de caudal de las estaciones limnimétricas y/o limnigráficas administradas por la entidad (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2023), desde el año inicial hasta el último día de registro al momento de solicitar la información. Datos máximos diarios de precipitación de las estaciones pluviométricas y/o pluviográficas aferentes a las cuencas hidrográficas, administradas por la

entidad (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2023), desde el año inicial hasta el último día de registro al momento de solicitar la información.

Con los anteriores insumos fue posible procesar los datos de caudales y precipitación máximos en 24h obteniendo los valores máximos para cada año de la serie de datos, después de realizar un análisis de outlier, en este paso se depuran tres estaciones de caudal que no tenían datos como reportaba la entidad, resultando ahora en 38 cuencas para el estudio. Hoja de cálculo desarrollada por el autor. También se pudo determinar la distribución espacial de la precipitación máxima en 24h en las cuencas hidrográficas del estudio (Polígonos de Thiessen), desarrollado en Arcmap 10.8.2. Para finalizar se realizó el análisis estadístico, análisis de valores atípicos, análisis de frecuencias, distribución de probabilidad y pruebas de bondad de ajuste de las series de datos de caudales y precipitación máximos en 24h para obtener variables proyectadas en diferentes periodos de retorno. Hoja de cálculo desarrollada por el autor y R (R Core Team, 2024).

El anterior objetivo está estrechamente relacionado con un resultado esperado del proyecto, el cual consiste en generar trabajos de investigación que puedan ser desarrollados por estudiantes de pregrado, y que sean parte del Semillero HydroInnovación, en este objetivo se hace énfasis en el proyecto denominado “Actualización de curvas IDF del Departamento del Cauca en una herramienta de visualización web”, el cual tendrá como finalidad, realizar una consulta de la información que permita establecer el inventario actual de las estaciones meteorológicas activas del Departamento del Cauca y que sean potencialmente utilizables en proyectos de Ingeniería Hidráulica e Hidrológica, todas a su vez administradas por el IDEAM. Se establecerá ubicación geográfica mediante el uso de cartografía georreferenciada y SIG , y se consignarán los datos más relevantes como tipo de estación, año de inicio de registro de datos, año de finalización de registro de datos, PMáx24h anual, PMáx24h horas promedio multianual, y con este último se generaran las curvas IDF con la metodologías aceptadas según la normatividad vigente (Vargas M. & Diaz-Granados O., 1998) para periodos de retorno de 2, 3, 5, 10, 20, 25, 50 y 100 años y duraciones de 10, 15, 30, 60. 120, 240 y 360 minutos. Se ofertará la información en el portal de Arcgis Online con el nombre: "Curvas IDF Departamento del Cauca", para que cualquier usuario (después de un breve registro) pueda consultar en línea la información de manera clara y obtener de manera práctica la curva IDF de la estación de interés en el Departamento del Cauca.

5.2.4 Objetivo específico 4

Establecer la mejor correlación lineal múltiple entre las variables morfométricas e hidrológicas con los caudales para diferentes periodos de retorno.

Producto de los anteriores objetivos se cuenta con seis matrices de datos, una para cada periodo de retorno a evaluar (2.33, 25, 50, 100, 200 y 500 años), que cuenta con 38 cuencas en las filas, 11 variables independientes en las columnas (Área de la cuenca, pendiente media de la cuenca, tiempo de concentración, factor de forma, índice de compacidad, índice de

alargamiento, índice de asimetría, razón de elongación, número de curva de escorrentía CN (II)*, promedio de precipitación máxima en 24 horas multianual y precipitación para el Tr analizado) y una variable dependiente en las columnas, que es el caudal para el Tr analizado.

Empleando el software R, se evalúan dos escenarios, el primero teniendo en cuenta todas las cuencas disponibles para el estudio, y determinando mediante el Criterio de Información de Akaike el mejor modelo que a su vez cumpla con los criterios de linealidad, independencia, homoscedasticidad, normalidad y no multicolinealidad. Finalmente se obtiene la ecuación del modelo para cada Tr. El segundo escenario busca depurar cuencas que se consideren que está afectando el rendimiento del modelo, esto a través de la distancia de Cook, y se repite el procedimiento buscando un modelo más robusto.

*: Se propone con la condición de humedad antecedente II, siendo un modelo concebido para condiciones de incertidumbre, direccionado al análisis básico de coberturas.

5.2.5 Objetivo específico 5

Determinar el rendimiento de los modelos de regresión lineal múltiple seleccionados para cada periodo de retorno, mediante métricas de eficiencia R^2 y KGE obtenidas con validación cruzada.

Del objetivo anterior se genera una nueva tabla de datos, esta vez únicamente con las variables que se tuvieron en cuenta para definir el mejor modelo para cada Tr de los dos escenarios. Empleando el software R, se realiza la evaluación de los modelos con la metodología de validación cruzada LOOCV (Leave-One-Out Cross-Validation). Se evaluó el rendimiento del modelo empleando dos diferentes criterios (Haan, 1977; Rodda & Little, 2015), Coeficiente de determinación (R^2) y KGE:

$$R^2 = 1 - \frac{SSE}{SST}$$

(Legates & McCabe, 1999)

$$KGE = 1 - \sqrt{(r - 1)^2 + \left(\frac{\sigma_{obs}/\sigma_{sim} - 1}{\beta}\right)^2 + \left(\frac{\mu_{sim} - \mu_{obs}}{\sigma_{obs}}\right)^2}$$

(Gupta, Kling, Yilmaz, & Martinez, 2009)

Para desarrollar estos objetivos se realizaron varios códigos de programación en R (ver Anexo 10).

6. Resultados y discusión

A continuación, se presentan los resultados asociados al desarrollo del presente trabajo de investigación (de acuerdo con los objetivos trazados), los cuales se generaron a partir de los datos suministrados por entidades y procesados por el autor (generando nuevos conjuntos de datos), los cuales podrán ser consultados en los Anexos.

6.1 Revisión bibliográfica detallada

Se realizó la búsqueda de las palabras, caudales máximos, cuencas hidrográficas, parámetros morfométricos, regresión lineal múltiple, SIG, validación cruzada, asociados a artículos y publicaciones asociadas a métodos para generación de ecuaciones de caudales máximos teniendo en cuenta variables fisiográficas, parámetros morfométricos y variables hidrológicas, en bases de datos académicas especializadas como Scopus, Science Direct, Google Scholar e IAHR Journals (Universidad del Cauca, 2024; IAHR - The International Association for Hydro-Environment Engineering and Research, 2024)

Lo anterior permitió sistematizar las metodologías existentes y evaluar las similitudes en cuanto tipo y cantidad de datos para evaluar la posibilidad de, a partir los métodos existentes formular la metodología aplicable al Departamento del Cauca, siendo detectadas variables independientes iniciales para la exploración de datos como por ejemplo Área de la cuenca, Pendiente de la cuenca, Tiempo de concentración, Factor de forma, Índice de compacidad, Índice de alargamiento, Índice de asimetría, Razón de elongación, Número de curva de escorrentía, Precipitación máxima en 24 horas, y Precipitación para Tr años de periodo de retorno, contra la variable dependiente Caudal para Tr años de periodo de retorno.

De manera relevante se encontraron trabajos importantes en Colombia, como, por ejemplo, Regionalización de curvas de duración y caudales mínimos en el Quindío, RLM caudales medios mensuales, Correlación entre sedimentos y caudales máximos ríos Sinú, Aracataca y Fundación y Caudales en cuencas sin registro en la zona Valle Medio de Magdalena (Vélez, Smith, Urrea, Pérez, & Franco, 1994; Carvajal, Salazar, Mesa, & Poveda, 2015; López Ramos, Martínez Acosta, & Feria Díaz, 2014; Donado, Romero-Duque, & Cortes-Ramos, 2021). En el extranjero, se encontraron de manera relevante respectivamente, en España, Australia, Costa Rica, Honduras y Chile, Caudales de inundaciones en las cuencas del Tajo y del Guadiana, Caudales teniendo en cuenta el impacto de la topografía, Regresión para caudales extremos en la vertiente Pacífica, Regresión múltiple para cálculo de frecuencia de caudales máximos y Análisis regional de crecidas máximas y Pronóstico de caudales en cuencas montañosas (Potenciano & Garzón, 2005; Johnson, Hutchinson, The, Beesley, & Green, 2016; Rojas Morales, Aguilar Pereira, & Solís Bolaños, 2008; Velasquez; Castro Heredia, Carvajal Escobar, & Ávila Díaz, 2012; Aguilera Sierra, 2010; Araya, Mendoza, Muñoz-Castro, & McPhee, 2023).

6.2 Caracterización morfométrica de cuencas hidrográficas a través de SIG

Se realizó el proceso exhaustivo para la obtención del archivo vectorial correspondiente al área de drenaje conforme a los puntos de cierre de las cuencas hidrográficas con coordenadas lat./long. según tabla 7, empleando las cajas de herramientas del software ArcMap.

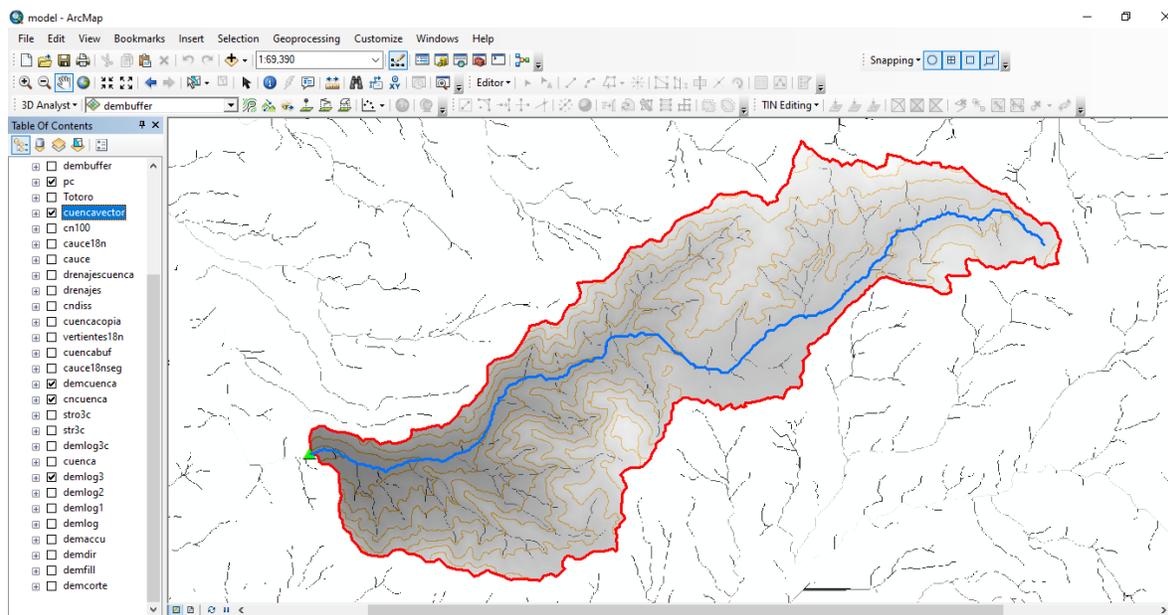


Figura 16 – Generación de la cuenca hidrográfica #40 Totoró como archivo vectorial en ArcMap
Fuente. El Autor

Los protocolos detallados para la obtención de variables fisiográficas y parámetros morfométricos quedan consignados en el manuscrito del libro “SIG para aplicaciones hidrológicas - Guía de ArcMap: Procedimientos para el cálculo de variables fisiográficas y parámetros morfométricos de cuencas hidrográficas”, como producto de la presente investigación.

Siendo consecuentes con la optimización y depuración de procesos, para obtener el archivo vectorial del resto de cuencas se emplearon códigos de libre uso que se encuentran en el repositorio GitHub (Heberger, 2023), los cuales emplean información global de libre acceso, y realizan de forma automatizada el proceso (Lin, y otros, 2019; Yamazaki, y otros, 2019).

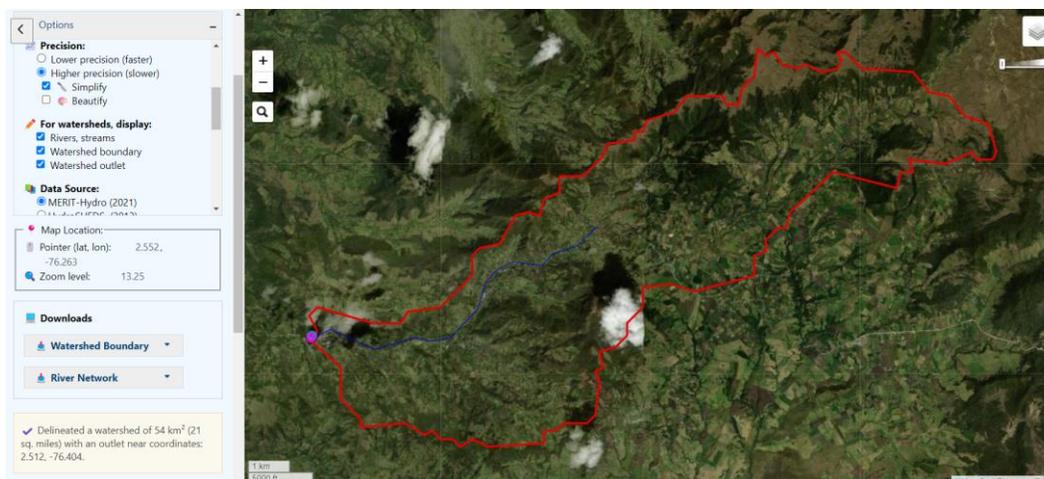


Figura 17 – Generación de la cuenca hidrográfica #40 Totoró como archivo vectorial en delineator.py

Fuente. El Autor

Las dos figuras anteriores permiten establecer la poca diferencia que tiene la generación de los archivos vectoriales por ambos métodos, siendo un proceso más ágil trabajar con delineator.py para el resto de las cuencas de la investigación.

Contando con la delimitación de estas, se siguió el protocolo para obtener las principales variables fisiográficas y realizar la caracterización morfométrica de cada una de ellas.

Variable	U.M	Magnitud
Área	km ²	53.51
Perímetro	km	47.47
Long. Cauce ppal.	km	20.28
Long. Cauces sec.	km	68.7
Long. Axial	km	16.38
Ancho máx	km	4.67
D. equivalente	km	8.25
A _{may}	km ²	28.47
A _{men}	km ²	25.04
W _{prom}	km	3.27

Parámetros de forma		
Parámetro	Nombre	Valor
Ff	Factor de forma	0.20
Ic	Índice de compacidad	1.82
la	Índice de alargamiento	3.51
Ias	Índice asimétrico	1.14
Re	Razón de elongación	0.50

Tabla 9 – Variables fisiográficas y parámetros morfométricos básicos para la cuenca #40 Totoró

Fuente. El Autor

También fue posible, siguiendo los protocolos más detallados obtener parámetros morfométricos específicos, como, por ejemplo, el rectángulo equivalente, curva hipsométrica y pendiente media (Horton, 1945):



Figura 18 – Rectángulo eq. de la cuenca hidrográfica #40 Totoró
Fuente. El Autor

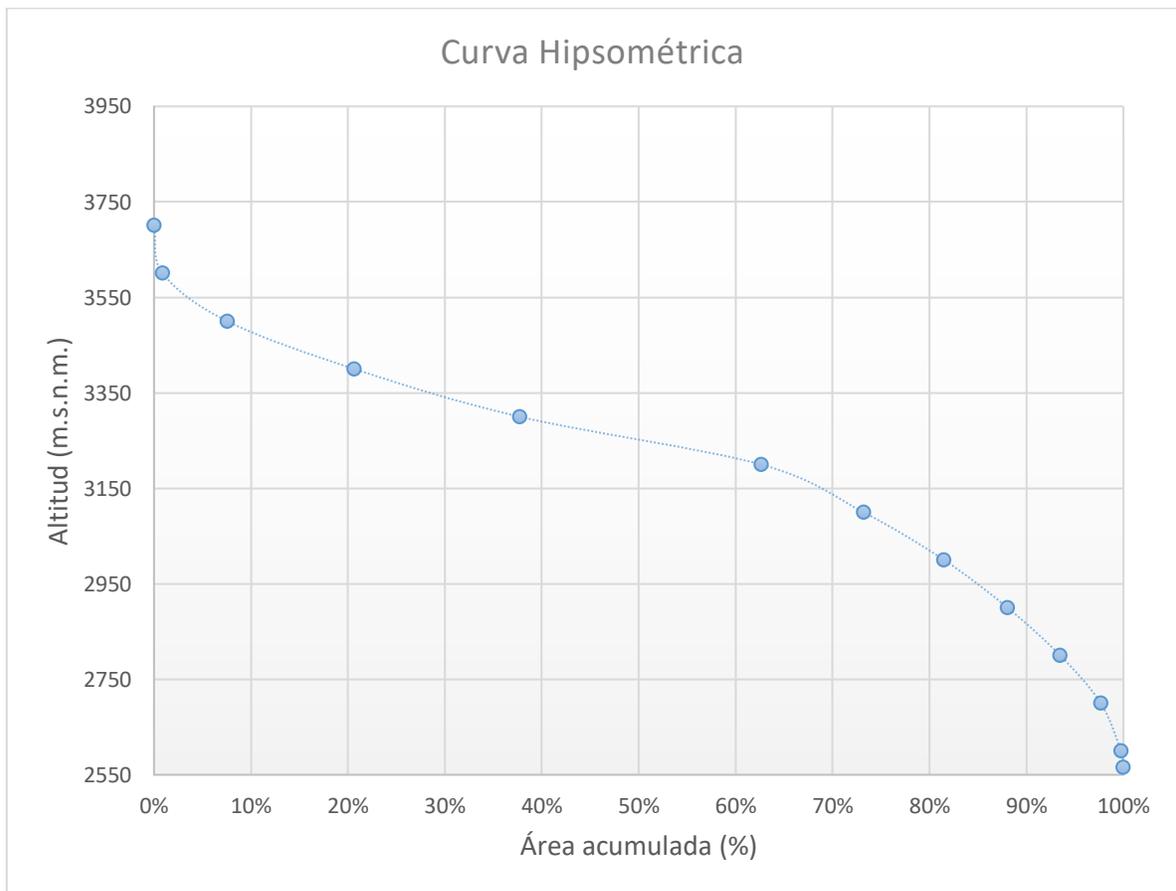


Figura 19 – Curva hipsométrica de la cuenca hidrográfica #40 Totoró
Fuente. El Autor

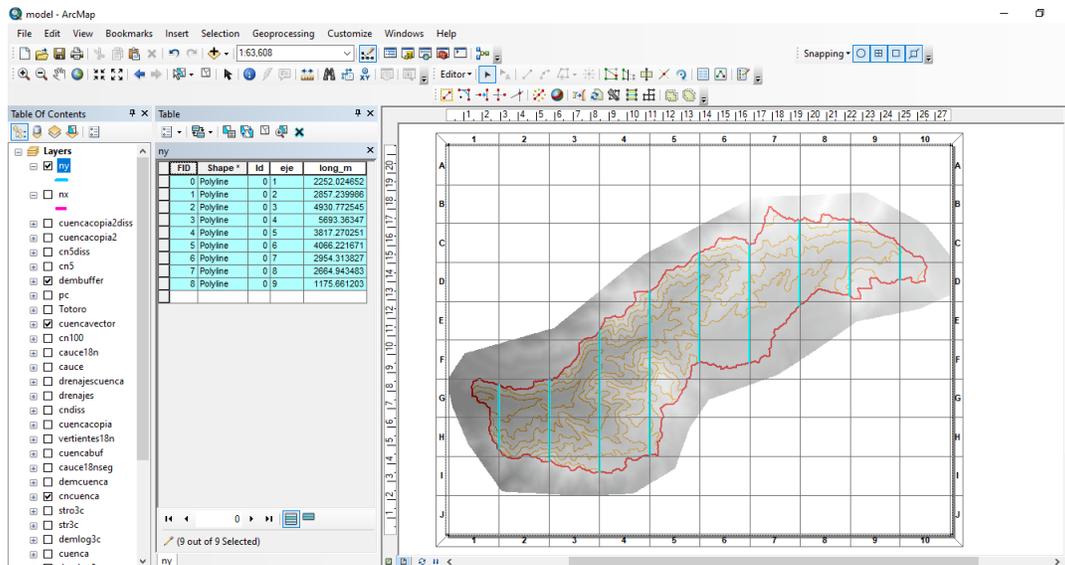


Figura 20 – Cálculo de la pendiente media de la cuenca hidrográfica #40 Totoró mediante el método de Horton en ArcMap

Fuente. El Autor

Se generaron datos mucho más específicos de cada cuenca, brindando una medida simplificada de estas, y que puede ser útil para comparaciones rápidas entre cuencas diferentes, o proporcionando una visión profunda sobre el estado de madurez y los procesos geomorfológicos de la cuenca. En especial la pendiente media que es una métrica esencial para entender la energía del flujo en la cuenca y su potencial para erosión y transporte de sedimentos (Montgomery & Dietrich, 1992).

Se estimaron también los tiempos de concentración conforme al numeral 4.1.3:

Ecuación	Tc (h)
Kirpich - USBR	2.26
Témez - DGC	5.37
Williams	5.72
Johnstone y Cross	4.57
Giandotti	2.52
SCS - Ranser	2.26
Ventura - Heras	2.18
Ven T. Chow	5.12
Cuerpo de Ingenieros de USA	5.01
Guaire	1.92
Passini	5.34
Pilgrim y McDermond	3.45

Tabla 10 – Tiempos de concentración para la cuenca #40 Totoró

Fuente. El Autor

Se obtuvieron los diagramas de cajas y bigotes, para visualizar la distribución del conjunto de datos, y detectar valores atípicos y asimetría.

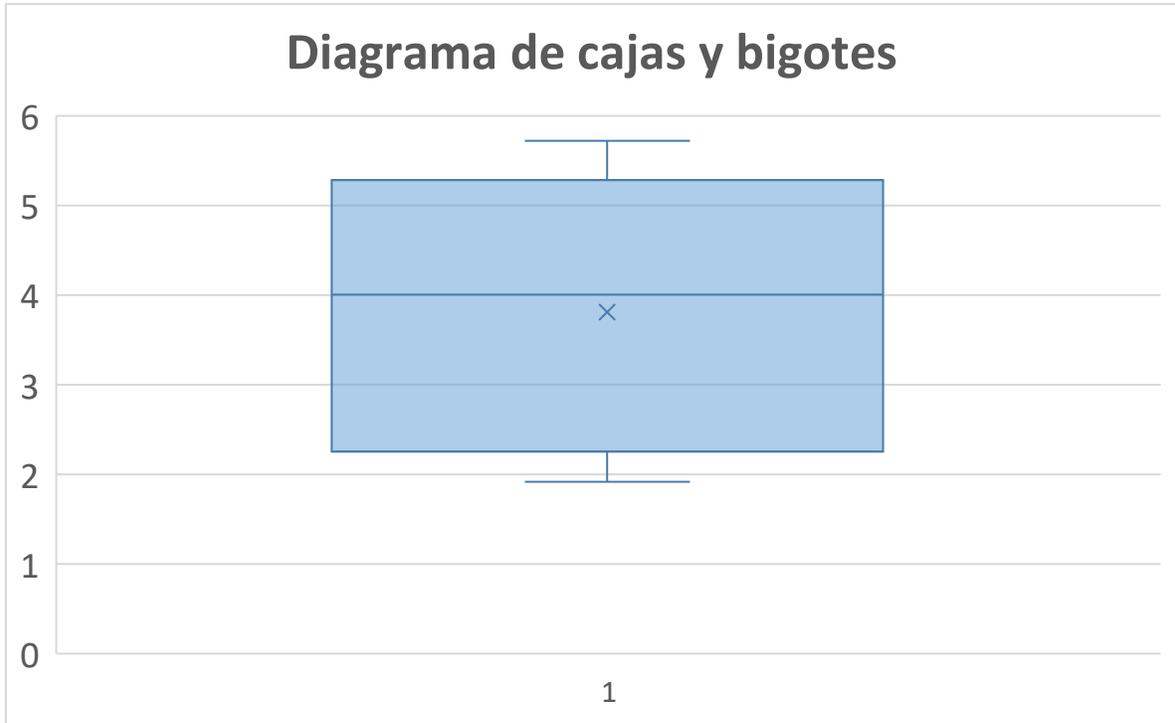


Figura 21 – Diagrama de cajas y bigotes para Tc cuenca hidrográfica #40 Totoró

Fuente. El Autor

Y finalmente se obtuvo un diagrama de barras para visualizar frecuencias haciendo más sencilla la comparación de las diferentes formulaciones.

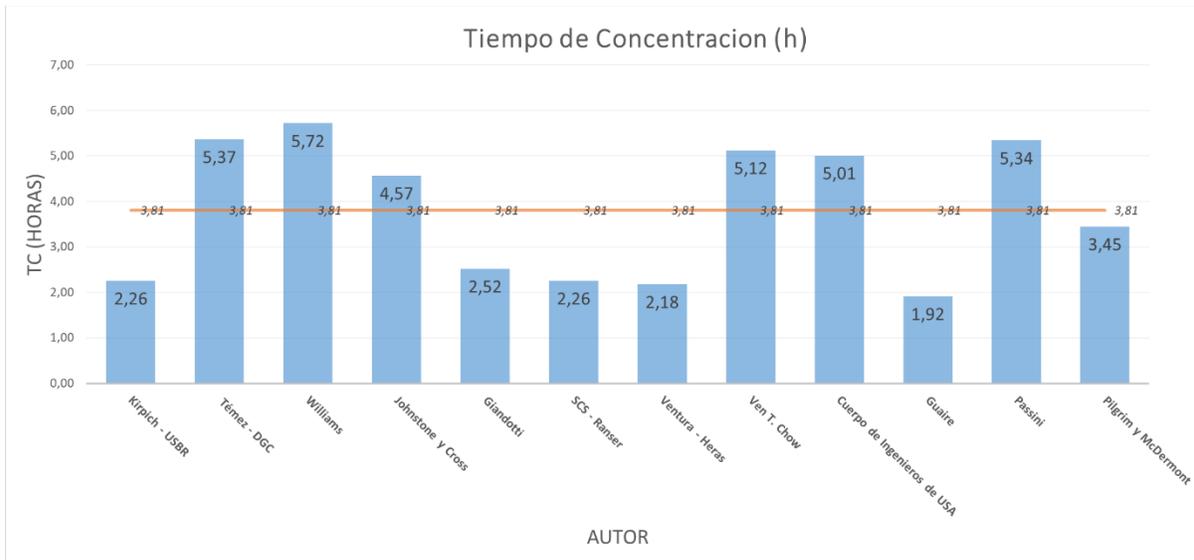


Figura 22 – Diagrama de barras para Tc cuenca hidrográfica #40 Totoró

Fuente. El Autor

Para complementar la caracterización se obtuvieron los coeficientes de escorrentía C y CN para las 41 cuencas hidrográficas, principalmente mediante el geoprocesamiento de imágenes satelitales (U.S. Geological Survey (USGS), 2023) y archivos vectoriales de entidades gubernamentales (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales,

2012; IGAC - Instituto Geográfico Agustín Codazzi, 2017). Estos coeficientes se resumen en la figura 23 y tabla 11.

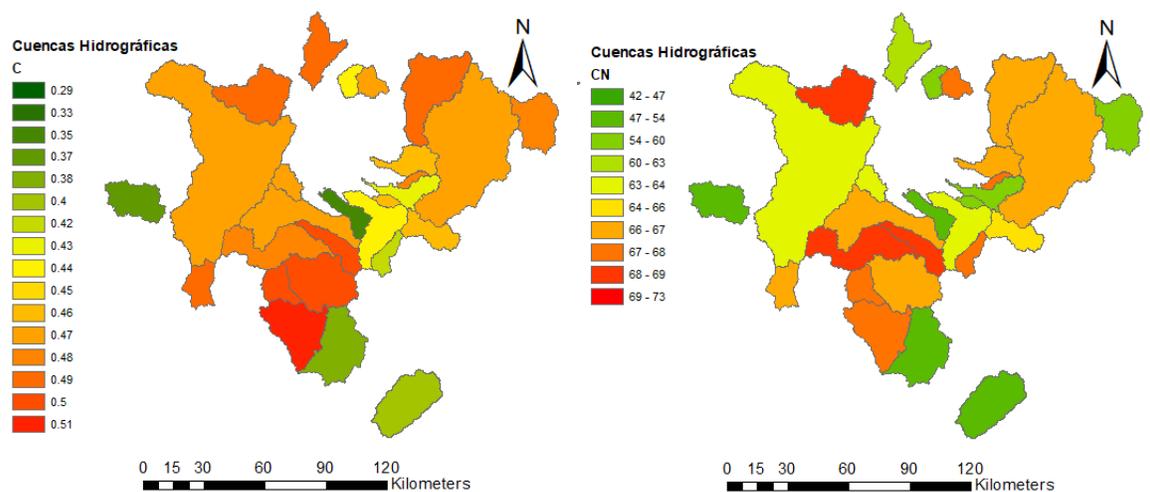


Figura 23 – Mapa de C y CN para las 41 cuencas del estudio

Fuente. El Autor

#	Código	C	CN-II	#	Código	C	CN-II	#	Código	C	CN-II
1	53077010	0.50	66.75	15	26017070	0.42	68.13	29	52017130	0.47	66.90
2	44017140	0.40	52.00	16	26027240	0.38	57.64	30	26027210	0.47	67.94
3	26047020	0.49	66.92	17	53077060	0.49	69.24	31	52027030	0.50	66.91
4	52017070	0.51	73.07	18	26017080	0.43	68.56	32	21057030	0.47	66.43
5	26017110	0.43	63.63	19	44017060	0.29	45.10	33	26037010	0.47	64.21
6	26027250	0.44	60.60	20	52027020	0.51	67.73	34	21057130	0.48	60.32
7	26027090	0.45	66.87	21	26017060	0.42	67.85	35	44017050	0.33	42.86
8	44017090	0.33	47.22	22	26027200	0.43	60.17	36	21057120	0.42	71.10
9	57027010	0.50	62.50	23	52017110	0.50	69.91	37	53047010	0.37	49.74
10	52027010	0.50	68.57	24	26017040	0.46	59.68	38	44017070	0.38	54.22
11	53077020	0.47	64.15	25	26027100	0.46	66.60	39	26057040	0.49	63.08
12	52017120	0.50	71.10	26	26037030	0.35	53.00	40	26027080	0.48	67.86
13	26017020	0.44	63.62	27	52017080	0.49	67.22	41	21057110	0.46	66.19
14	52017030	0.48	69.61	28	52017100	0.44	65.49				

Tabla 11 – Valores de C y CN para las 41 cuencas del estudio

Fuente. El Autor

Contando con el archivo vectorial donde se consignaron mediante disolución las capas de trabajo (Cobertura, Suelos y Pendiente) en donde quedaron los polígonos clasificados para Área, C y CN, se realizó un código que permitiera el cálculo de C y CN para cualquier polígono que quedara dentro del área de trabajo, el cual se representa en la siguiente figura mediante un diagrama de flujo, generando un script, en donde la única entrada es el polígono al que se le quiere calcular el C y CN, denominado *C&CN_Calculator*.

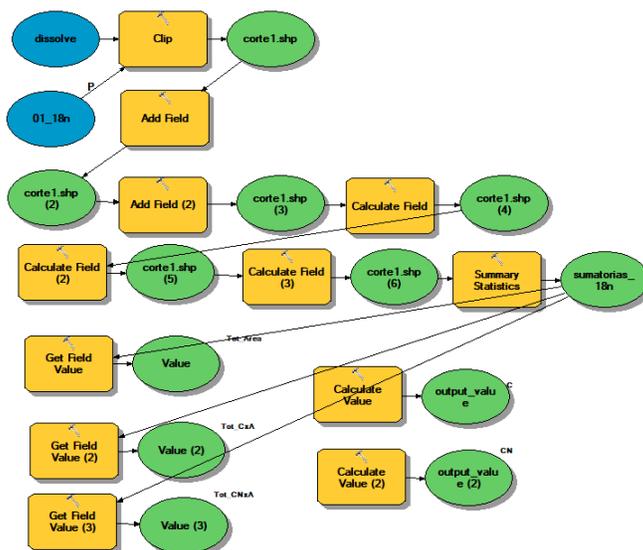


Figura 24 – Diagrama de flujo para el C&CN_Calculator

Fuente. El Autor

En cuanto a la aplicación, si bien la implementación de C y CN es semicuantitativa, y depende principalmente del criterio de los autores, su practicidad la hace un producto atractivo para trabajos a grandes escalas, teniendo como uno de sus mayores beneficios el tiempo de estimación de los coeficientes.

El código se puede comprobar en el Anexo 10.2, siendo un producto asociado al proyecto “Estudio y caracterización del C y CN para el Departamento del Cauca: Correlación e implementación” (Agredo-Campuzano, Jiménez, Bolaños-Rodríguez, & Paredes-Guevara, 2024).

En cuanto a la superficie total analizada, el 50.48% del área estudiada es tipo C, con un potencial de escorrentía moderadamente alto. También se logró establecer un 33.67% son bosques y le sigue en extensión el mosaico de cultivos, pastos y bosques, que cubre el 13.16%. Esta combinación refleja la coexistencia de actividades humanas, como la agricultura y la ganadería, con áreas naturales, lo que eventualmente podría verse reflejado en la degradación de las cuencas, siendo sinónimo de crecientes relacionadas con movimientos en masa y avenidas torrenciales, por lo que se podría emplear el producto de la presente investigación para determinar, a gran escala en modelos de iniciación, la probabilidad de ocurrencia de eventos de remoción en masa, en función de los coeficientes

de escorrentía y tipos de suelo, siendo esta una de las variables más complejas de cuantificar por los elevados costos que tendría debido a los trabajos de campo y ensayos especializados requeridos (Rickenmann, 2020).

6.3 Estimación de precipitaciones y caudales máximos para diferentes periodos de retorno

Se procesó la información cruda suministrada por la entidad gubernamental (IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales, 2023), que contiene valores de precipitación máxima en 24 horas diarias, y valores de caudal máximo diario (Ver Anexo 3 y 4 respectivamente), mediante tablas dinámicas para obtener los valores máximos multianuales. Se analizó estación por estación para verificar la consistencia de los datos y realizar el proceso de completar datos de años faltantes, principalmente con el método de la media o promedio, el cual usa la media de los valores disponibles para estimar los valores faltantes. Este método es sencillo, pero puede no capturar variaciones significativas (Chow, Maidment, & Mays, 1988). En otros casos se completó la información con años de estaciones cercanas, en todo caso las series de tiempo presentaban una buena consistencia de datos, siendo menos del 5% los datos que tuvieron que ser completados.

1971	68	44	78	76	28	10	8	21	14	25	39	50	78.0	1.89
1972	59	42	56	36	36	12	2	38	39	44	117	36	117.0	2.07
1973	14	36	42	32	36	13	20	30	22	75	63	50	75.0	1.88
1974	42	53	78	45	33	12	26	19	76	44	54	39	78.0	1.89
1975	38	50	69	26	39	36	46	33	19	39	56	66	69.0	1.84
1976	40	25	35	41	16	4	1	23	33	58	11	62	62.0	1.79
1977	26.8	30	37	104	35	36	9	18	41	56	28	72	104.0	2.02
1978	41	20	65	52	21	24	43	7	42	34	60	66	66.0	1.82
1979	65	26	80	36	33	10	33	39	32	48	38	50	80.0	1.90
1980	44	88	58	64	43	28	6	32	50	30	25	50	88.0	1.94
1981	27	55	107	59	43	58.3	18.6	20.7	24.8	33.2	44.2	57.3	107.0	2.03
1982	44.1	42.8	53.3	31.4	16.3	11	6	4	30	65	84	43	84.0	1.92
1983	91	69	60	65	46	15	29	12.6	9.5	43.5	27	47.8	91.0	1.96
1984	52.7	48	56	76.3	56.2	35	41	30	62	56	37	72	76.3	1.88
1985	49	38	43	43	30	24	3	24	20	63	32	30	63.0	1.80
1986	35	53	54	25	34	6	1	10	66	45	53	46	66.0	1.82
1987	44	145	25	79	30	9	23	10	54	53	44	30	145.0	2.16
1988	33	36	53	71	47	27	17	23	50	45	74	47	74.0	1.87
1989	74	28	72	31	42	27	48	11	22	35	30	41	74.0	1.87
1990	40	48	30	41	24	15	23	27	12	72	85	66	85.0	1.93
1991	60	40	42	16	24	19	28	22	43	39	39	69	69.0	1.84
1992	30	48	44	43	11	28	15	9	16	11	69	43	69.0	1.84
1993	46	73	35	55	48	4	12	3	31	30	48	50	73.0	1.86
1994	59.6	42.3	38	65.1	27.5	30.8	12	9	18	6.8	64	39	65.1	1.81
1995	24	36	51	41.8	27	32	40	27	45.7	50.7	29	59.5	59.5	1.77
1996	49	48	45	43	60	37	39	10	23	50	24	55	60.0	1.78
1997	53	15	71	30	49	30	0	0	36	59	72	48	72.0	1.86
1998	23	41	48	65		26	49	45	31	37	88	27	88.0	1.94
1999	41	69	62	52	31	38	8	9	31	62	97	53	97.0	1.99

2000	18	26	21	25	23	47	14	15	30	30	30	46	47.0	1.67
2001	24	34	36	17	50	30	11	0	37	40	68	90	90.0	1.95
2002	65	40	35	41	15	25	30	30	10	40	30	50	65.0	1.81
2003	60	20	30	60	30	40	10	14	18	60	45	40	60.0	1.78
2004	46	50	41	35	20	0	17			51	71	58	71.0	1.85
2005	70	74	65	24	14	17	2	3	20	36	42	68	74.0	1.87
2006	42	27	39	41	28	22	35	8	13	71	37	36	71.0	1.85
2007	32	79	62	69	49	8	18	44	4	44	87	70	87.0	1.94
2008	45	71	55	78	63	35	55	41	29	85	67	94	94.0	1.97
2009	44	33	98	61	66	22	3	22	64	108	69	81	108.0	2.03
2010	46	71	30	76	84	49	122	29	31	207	32	50	89.0	1.95
2011	51	94	53	34	59	13	34	17	12	62	88	51	94.0	1.97
2012	60	26	62	112	33	29	2	2	1	49	52	94	112.0	2.05
2013	52	58	43	82	58	71	6	17	122	42	61	40	122.0	2.09
2014	67	38	72	52	48	6	0	10	27	69	43	47	72.0	1.86
2015	58	30	46	91	24	8	13	1	29	68	58	24	91.0	1.96
2016	31	84	37	65	18	33	49	4	48	70	32	55	84.0	1.92
2017	62	45	71	32	49	52	4	23	3	67	61	84	84.0	1.92
2018	27	73	92	38	21	20	24	13	32	62	58	32	92.0	1.96
2019	60	55	56	65	37	11	23	3	40	23	83	38	83.0	1.92
2020	44	47	46	28	41	27	34	23	23	74	38	96	96.0	1.98
2021	65	112	54	41	44	25	7	45	35	37	78	40	112.0	2.05
2022	102	60	56	73	32	53	32	16	52	76	60	54	102.0	2.01

Tabla 12 – Datos de precipitación máxima mensual y multianual para la estación #48 Piendamó

Fuente. El Autor

Posteriormente se realizó el análisis de datos dudosos (Water Resources Council (Actualmente United States Geological Survey), 1981) , en la misma hoja de cálculo.

PARÁMETROS ESTADÍSTICOS	P24hr	Log(P24hr)
Número de datos (N)	76.00	76.000
Sumatoria	5802.90	142.066
Valor Máximo	145.00	2.161
Valor Mínimo	38.00	1.580
Media:	76.35	1.869
Varianza:	383.22	0.012
Desviación Estándar:	19.58	0.109
Coficiente Variación:	0.26	0.058
Coficiente de Sesgo:	0.84	0.043
Se Considera:	Detectar Datos Dudosos Altos y Bajos	
n=	76.00	
Kn=	2.9116	
Kn:	Valor recomendado, varía según el valor de n (significancia:10%)	

Umbral de datos dudosos altos (xH: unidad. Logarítmicas)

$$x_H = \bar{x} + k_n \cdot S \quad xH= \quad 2.19$$

Precipitación máxima aceptaba

$$PH= \quad 10^{xH} \quad PH= \quad 153.67 \quad \text{mm}$$

NO EXISTEN DATOS DUDOSOS ALTO DE LA MUESTRA

Umbral de datos dudosos bajos (xL: unidad. Logarítmicas)

$$x_L = \bar{x} - k_n \cdot S \quad xL= \quad 1.55$$

Precipitación mínima aceptaba

$$PH= \quad 10^{xL} \quad PL= \quad 35.64 \quad \text{mm}$$

NO EXISTEN DATOS DUDOSOS MINIMO DE LA MUESTRA

Tabla 13 – Verificación de datos dudoso para la estación #48 Piendamó

Fuente. El Autor

Se graficaron las series contra los datos límites máximos y mínimos, obteniendo una mejor interpretación visual de que tan centralizado estaba el valor anual contra los valores límites.

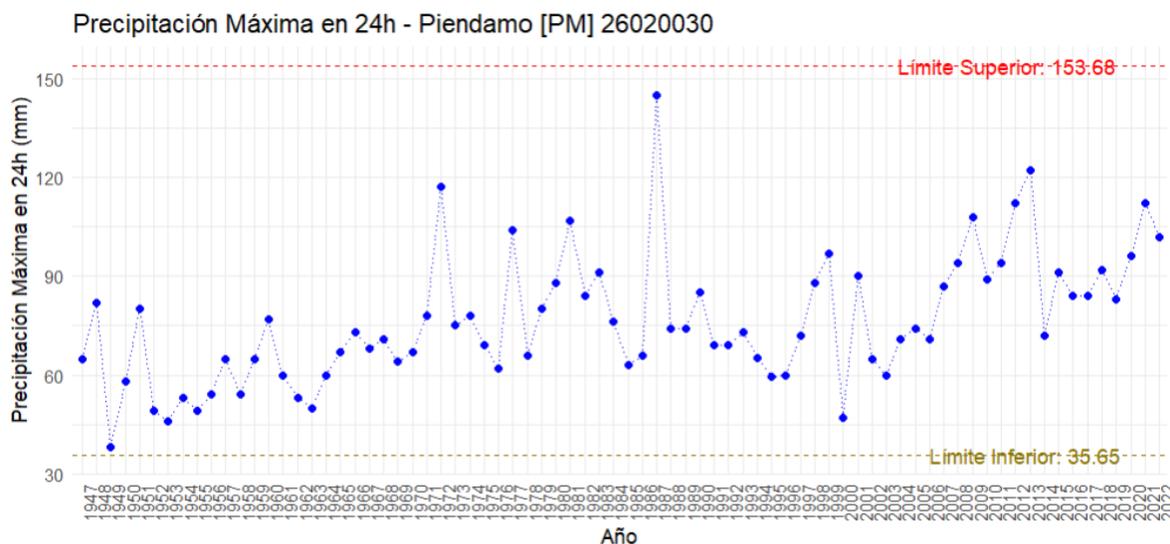


Figura 25 – Precipitación máxima en 24 horas vs limites estación #48 Piendamó

Fuente. El Autor

Este proceso es muy importante para la presente investigación, ya que permitió la identificación de datos atípicos, mejorando la precisión de los modelos, ya que los outliers pueden distorsionar los resultados de los análisis estadísticos, influyendo en la media, la varianza y otros estadísticos importantes. Al identificarlos y tratarlos adecuadamente, se mejora la precisión y la robustez de los modelos que derivan de los datos tratados. No todos los outliers se corrigieron, ya que el método sirve para la identificación de fenómenos extremos, puesto que, en algunos casos, los outliers pueden representar fenómenos reales pero raros, como eventos hidrológicos extremos. Identificarlos es fundamental para estudios de riesgos y diseño de infraestructura que debe soportar tales eventos (Kundzewicz & Robson, 2004).

En los Anexos se puede verificar que se realizó el proceso para todas las estaciones de precipitación (ver Anexo 5) y caudal (ver Anexo 6), y en este punto del análisis de datos, se pasó de tener 41 cuencas para correlacionar, a 38, puesto que las estaciones, #5 26017110 BOCATOMA-VINAGRE y #34 21057130 RIO CHIQUITO, no disponían de datos de caudal (aun cuando la entidad reportaba que sí), adicionalmente se optó por dejar por fuera de la investigación la cuenca #9 57027010 GORGONA LA PLAYA, ya que posiblemente por la magnitud del área aferente podría convertirse en una cuenca influyente del modelo (1 km^2), además es la única cuenca con influencia oceánica directa, conformando un régimen totalmente diferente a las cuencas continentales (Dai & Trenberth, 2004).

Como producto asociado al proyecto “Actualización de curvas IDF del Departamento del Cauca en una herramienta de visualización web” se procesaron los datos para la generación

de Curvas IDF (Vargas M. & Diaz-Granados O., 1998) de las 86 estaciones con datos de precipitación.

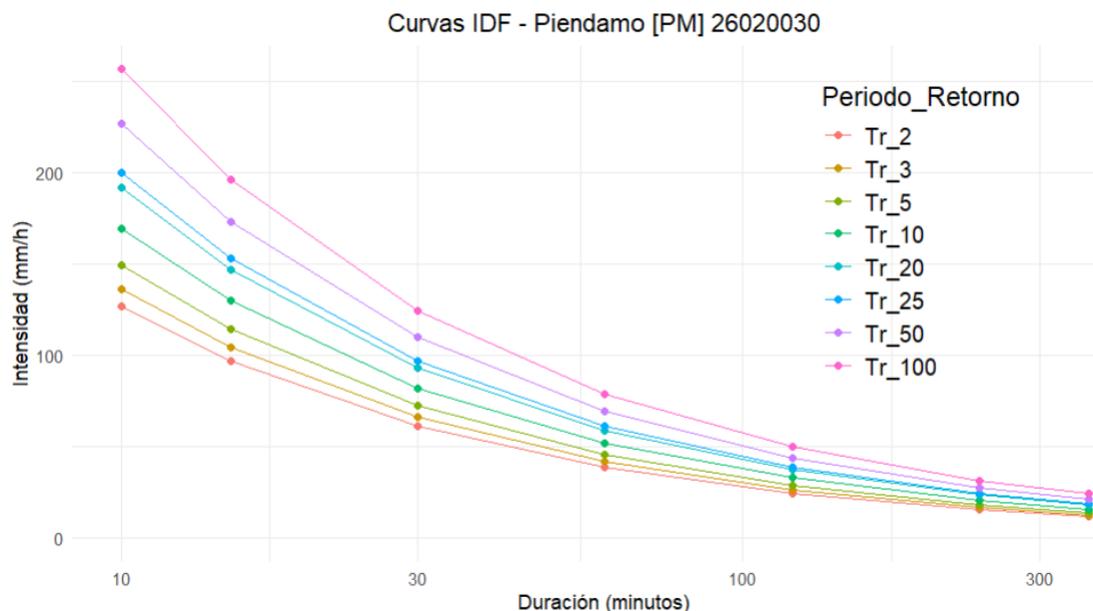


Figura 26 – Curvas IDF para estación #48 Piendamó

Fuente. El Autor

Antes de obtener los valores de precipitación para los periodos de retorno de 2.33, 25, 50, 100, 200 y 500 años de las cuencas se determinó el porcentaje de aferencia de cada estación a cada cuenca hidrográfica mediante la distribución espacial con Polígonos de Thiessen (Thiessen & Alter, 1911).

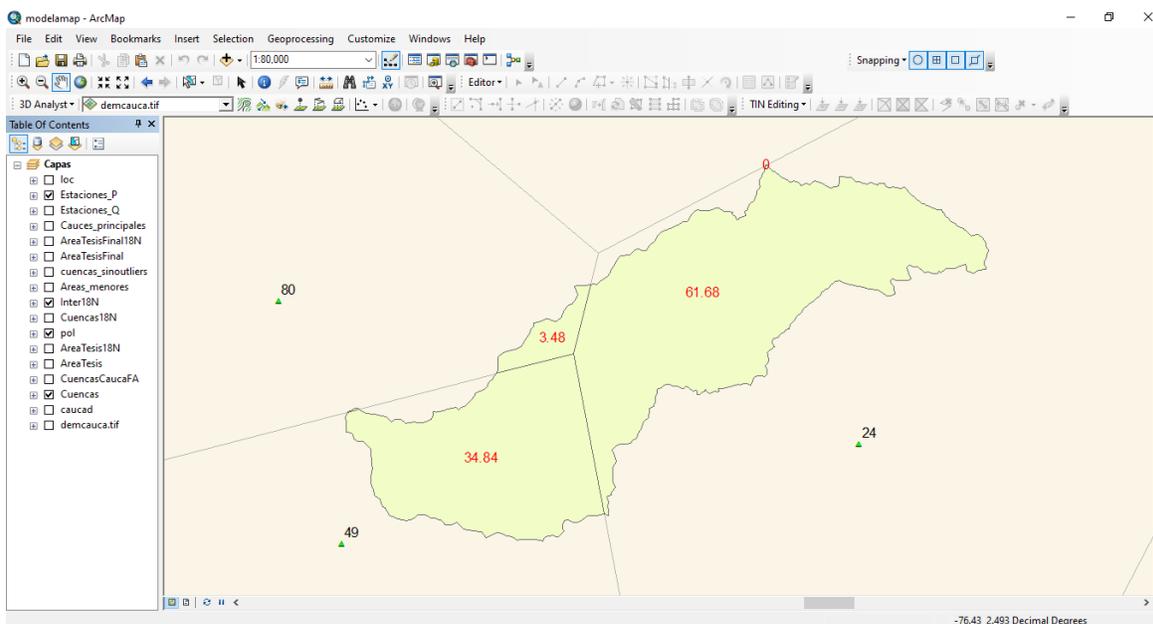


Figura 27 – Polígonos de Thiessen y porcentaje de aferencia de estaciones de precipitación sobre la cuenca #40 Totoró en ArcMap

Fuente. El Autor

Una vez se conoció el porcentaje de aferencia de las cuencas para las estaciones con medidas de precipitación, se procede a generar una serie de datos ponderada.

Número	24	49	80	
Código	26025070	26020460	26020130	
Nombre	GABRIEL LOPEZ	POLINDARA	TOTORO	
Tipo	CO	PM	PM	
%	<i>61.68%</i>	<i>34.84%</i>	<i>3.48%</i>	Cuenca 40
1985	25	51	64	35.42
1986	27.8	47	55	35.44
1987	43	66	78	52.23
1988	23.6	100	54	51.28
1989	25	90	54	48.66
1990	26	80	54	45.79
1991	22.8	54	49	34.58
1992	42	80	70	56.21
1993	40	90	77	58.71
1994	29.9	60	80	42.13
1995	30	50	76	38.57
1996	28.5	82	75	48.76
1997	27.4	64	65	41.46
1998	28.2	72	55	44.39
1999	26.8	101	57	53.70
2000	40.6	46	70	43.50
2001	23.1	58	50	36.20
2002	38.8	65	51	48.35
2003	21.5	71	47	39.63
2004	27.6	65	80	42.45
2005	22.4	62	60	37.51
2006	25.8	58	45	37.69
2007	27.8	68.43	59	43.04
2008	25.3	44	71	33.41
2009	23	63	63	38.33
2010	43.6	54	65	47.97
2011	28.2	63	58	41.36
2012	22	67.2	60	39.07
2013	40.3	67	57	50.18
2014	30.2	65	63	43.47
2015	28.3	60	50	40.10

2016	29.6	53	58	38.74
2017	32.6	80	89	51.08
2018	30.4	56	65	40.52
2019	45.5	61	56	51.27
2020	49.9	45	68	48.82
2021	43.7	60	45	49.42
2022	47.2	60	85	52.97
				44.27

Tabla 14 – Serie ponderada de precipitación para la estación #40 Totoró

Fuente. El Autor

La conformación de ponderaciones de precipitación para cada cuenca hidrográfica se puede verificar en el Anexo 7.

Estas nuevas series fueron procesadas con distribuciones de probabilidad, escogiendo la que mejor bondad de ajuste resultara con el método de Kolmogorov – Smirnov. Se prefiere este método ya que es una prueba no paramétrica, lo que significa que no asume una distribución específica para los datos. Esto lo hace útil para comparar cualquier distribución empírica con una distribución teórica o para comparar dos distribuciones empíricas sin hacer suposiciones adicionales, también es robusto en cuanto a tamaño de muestra, ya que se puede ajustar para trabajar con muestras más pequeñas de manera razonable en comparación con otras pruebas que pueden requerir tamaños de muestra más grandes para obtener resultados precisos (Massey, 1951).

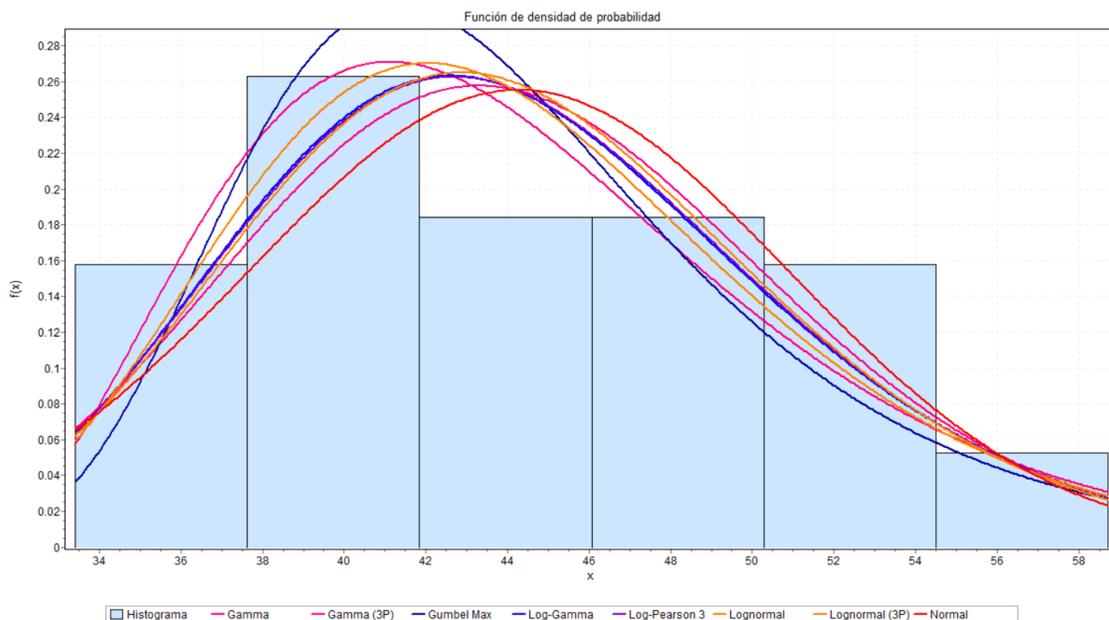


Figura 28 – Distribuciones de probabilidad con datos de precipitación de la cuenca #40 Totoró

Fuente. El Autor

#	Distribución	Kolmogorov Smirnov	
		Estadística	Rango
8	Normal	0.10737	1
1	Gamma	0.11898	2
5	Log-Pearson 3	0.12648	3
4	Log-Gamma	0.12737	4
6	Lognormal	0.12737	5
7	Lognormal (3P)	0.13515	6
2	Gamma (3P)	0.13777	7
3	Gumbel Max	0.15553	8

Tabla 15 – Pruebas de bondad de ajuste organizadas para la cuenca #40 Totoró

Fuente. El Autor

La distribución reportada como la mejor (Normal) se utilizó para estimar valores para diferentes periodos de retorno.

Tr	% excedencia	% no excedencia	P Máx 24h (Normal)
2.33	0.429	0.571	45.453
25	0.040	0.96	55.806
50	0.020	0.98	57.802
100	0.010	0.99	59.598
200	0.005	0.995	61.241
500	0.002	0.998	63.233

Tabla 16 – Cálculo de precipitación máxima para diferentes periodos de retorno para la estación #40 Totoró

Fuente. El Autor

De una forma similar (sin ponderar ya que los datos de caudal son propios de cada cuenca) se obtuvieron los valores de caudal máximo para periodos de retorno de 2.33, 25, 50, 100, 200 y 500 años. En los Anexos 8 y 9 se puede encontrar el reporte de las distribuciones de probabilidad y pruebas de bondad de ajuste para los datos de precipitación ponderada máxima en 24 horas y caudal máximo para las cuencas hidrográficas.

Sin embargo, se presenta en la tabla 17, un resumen que consigna, la distribución que mejor se ajusto a cada cuenca, dependiendo si se analizan los datos de precipitación o caudal.

#	Código	Nombre	P	Q
1	53077010	ANGOSTURA RIO MICAY	Lognormal	Log-Pearson 3
2	44017140	BETANIA-VILLALOBOS	Gamma	Gamma
3	26047020	BOCAT RIO PALO	Normal	Log-Gamma
4	52017070	BOCATOMA SAJANDI	Normal	Lognormal (3P)

6	26027250	BUENOS AIRES	Gamma (3P)	Gumbel Máx
7	26027090	CORTIJO EL	Gumbel Máx	Lognormal
8	44017090	CURIACO	Log-Gamma	Log-Pearson 3
10	52027010	GUACHICONO	Gumbel Máx	Log-Gamma
11	53077020	GUAYABAL	Gumbel Máx	Gamma
12	52017120	HOYO EL	Gumbel Máx	Log-Pearson 3
13	26017020	JULUMITO	Log-Pearson 3	Gumbel Máx
14	52017030	LA FONDA - AUT	Normal	Gamma (3P)
15	26017070	LOMITAS	Lognormal (3P)	Normal
16	26027240	MALVASA	Gamma (3P)	Gamma
17	53077060	MONO EL	Lognormal (3P)	Gumbel Máx
18	26017080	PALETARA	Lognormal (3P)	Lognormal
19	44017060	PAPAS	Log-Pearson 3	Lognormal (3P)
20	52027020	PLAYA LA	Log-Pearson 3	Log-Pearson 3
21	26017060	PUENTE ARAGON - AUT	Log-Pearson 3	Lognormal
22	26027200	PUENTE CARRETERA - AUT	Gamma (3P)	Log-Gamma
23	52017110	PUENTE CARRETERA - AUT	Lognormal (3P)	Normal
24	26017040	PUENTE CARRETERA	Normal	Gamma (3P)
25	26027100	PUENTE CARRETERA	Gumbel Máx	Log-Gamma
26	26037030	PUENTE CARRETERA	Normal	Gamma (3P)
27	52017080	PUENTE CARRETERA	Normal	Log-Pearson 3
28	52017100	PUENTE CARRETERA	Normal	Lognormal (3P)
29	52017130	PUENTE COLGANTE TIMBIO	Gamma	Lognormal (3P)
30	26027210	PUENTE FERROCARRIL	Gamma (3P)	Log-Gamma
31	52027030	PUENTE FIERRO - GUACHICONO	Gamma (3P)	Lognormal (3P)
32	21057030	PUENTE RICAURTE	Gamma (3P)	Normal
33	26037010	REMOLINO	Gumbel Máx	Lognormal (3P)
35	44017050	RIOGRANDE - CAUCA	Gumbel Máx	Gumbel Máx
36	21057120	SAN JUAN	Gumbel Máx	Log-Gamma
37	53047010	SANGARAL	Lognormal (3P)	Gumbel Máx
38	44017070	SANTA ROSA - AUT	Log-Pearson 3	Gumbel Máx
39	26057040	TIMBA	Gamma	Log-Pearson 3
40	26027080	TOTORO	Normal	Lognormal (3P)
41	21057110	TREBOL EL	Normal	Normal

Tabla 17 – Distribuciones de probabilidad para precipitación y caudal

Fuente. El Autor

Se conformó la matriz de datos, que relaciona en las filas las 38 cuencas hidrográficas, y en las columnas las variables morfométricas e hidrológicas:

#	Código	Nombre	Área (km ²)	Perímetro (km)	Long. Cauce ppal. (km)	Long. Axial (km)	Ancho máx (km)	D. equivalente (km)	A _{may} (km ²)	A _{men} (km ²)	W _{prom} (km)	Cota sup. Cauce ppal (msnm)	Cota inf. Cauce ppal (msnm)	Pendiente cauce ppal. (m/m)
1	53077010	ANGOSTURA RIO MICAY	2380.97	259.76	107.86	72.14	49.74	55.06	1625.22	755.75	33.00	3042.57	40.98	2.7829
2	44017140	BETANIA-VILLALOBOS	643.61	154.00	54.41	38.05	23.19	28.63	414.18	229.43	16.91	2071.79	846.47	2.2520
3	26047020	BOCAT RIO PALO	841.91	217.62	50.51	42.58	30.57	32.74	651.99	189.92	19.77	2762.63	1051.59	3.3875
4	52017070	BOCATOMA SAJANDI	166.99	63.76	18.23	14.19	19.62	14.58	86.34	80.65	11.77	2619.00	800.64	9.9745
6	26027250	BUENOS AIRES	127.84	57.00	28.20	17.20	10.17	12.76	83.50	44.34	7.43	1876.00	996.56	3.1186
7	26027090	CORTJO EL	151.08	59.85	18.06	16.72	12.96	13.87	94.63	56.45	9.04	3361.76	2541.00	4.5446
8	44017090	CURIACO	427.99	133.33	40.78	29.81	23.55	23.34	259.62	168.37	14.36	3549.00	1942.79	3.9387
10	52027010	GUACHICONO	960.57	164.23	75.87	45.96	29.92	34.97	702.38	258.19	20.90	3644.00	716.40	3.8587
11	53077020	GUAYABAL	3819.58	414.61	177.65	95.67	49.90	69.74	2691.93	1127.65	39.92	1845.79	9.17	1.0338
12	52017120	HOYO EL	625.17	160.37	72.07	48.69	19.07	28.21	449.71	175.46	12.84	3741.00	629.44	4.3174
13	26017020	JULUMITO	723.60	156.15	71.28	44.58	24.63	30.35	506.02	217.58	16.23	3246.00	1672.70	2.2072
14	52017030	LA FONDA - AUT	1700.51	247.59	86.82	51.27	42.71	46.53	1156.15	544.36	33.17	3724.00	598.32	3.6002
15	26017070	LOMITAS	71.62	46.28	10.66	9.69	7.66	9.55	44.51	27.11	7.39	3274.00	2972.00	2.8330
16	26027240	MALVASA	116.74	54.25	18.35	15.26	10.03	12.19	68.40	48.34	7.65	3533.00	2939.78	3.2328
17	53077060	MONO EL	632.20	139.45	57.54	39.72	30.67	28.37	375.98	256.22	15.92	2553.00	47.27	4.3548
18	26017080	PALETARA	63.87	45.56	17.77	16.49	5.55	9.02	45.98	17.89	3.87	4377.00	2973.00	7.9010
19	44017060	PAPAS	86.49	50.48	13.14	12.35	15.05	10.49	56.48	30.01	7.00	3473.00	2882.00	4.4977
20	52027020	PLAYA LA	666.22	141.28	54.31	31.85	20.98	29.12	359.90	306.32	20.92	3318.00	711.67	4.7990
21	26017060	PUENTE ARAGON - AUT	151.78	70.82	18.38	16.86	6.70	13.90	133.87	17.91	9.00	4386.40	2968.28	7.7156
22	26027200	PUENTE CARRETERA - AUT	250.17	123.84	52.28	39.28	11.64	17.85	161.76	88.41	6.37	3533.17	1732.19	3.4449
23	52017110	PUENTE CARRETERA - AUT	271.07	109.61	47.43	38.05	10.00	18.58	142.46	128.61	7.12	3740.00	878.00	6.0342
24	26017040	PUENTE CARRETERA	57.25	37.59	15.36	13.92	6.37	8.54	46.88	10.37	4.11	3576.00	1999.00	10.2669
25	26027100	PUENTE CARRETERA	285.21	105.99	44.10	34.07	14.21	19.06	154.32	130.89	8.37	3360.72	1775.20	3.5953
26	26037030	PUENTE CARRETERA	210.83	96.25	36.74	29.59	10.96	16.38	177.05	33.78	7.13	3062.00	1581.47	4.0297
27	52017080	PUENTE CARRETERA	247.29	80.49	31.53	23.42	13.65	17.74	182.53	64.76	10.56	2970.00	592.58	7.5402
28	52017100	PUENTE CARRETERA	276.79	106.66	41.03	33.91	10.84	18.77	227.23	49.56	8.16	3144.66	922.00	5.4172
29	52017130	PUENTE COLGANTE TIMBIO	767.14	201.09	70.64	49.60	23.24	31.25	690.52	76.62	15.47	3144.66	631.15	3.5582
30	26027210	PUENTE FERROCARRIL	155.80	62.44	18.75	16.21	12.33	14.08	78.57	77.23	9.61	1422.00	997.00	2.2667
31	52027030	PUENTE FIERRO - GUACHICONO	732.02	139.80	54.70	38.12	26.39	30.53	506.51	225.51	19.20	3454.00	927.21	4.6194
32	21057030	PUENTE RICAURTE	2287.97	286.43	73.85	57.82	48.42	53.97	1564.60	723.37	39.57	3895.00	1168.75	3.6916
33	26037010	REMOLINO	161.16	66.27	16.89	13.22	12.06	14.32	132.88	28.28	12.19	2566.00	1613.00	5.6424
35	44017050	RIOGRANDE - CAUCA	131.87	68.50	26.74	22.04	12.76	12.96	73.44	58.43	5.98	3471.00	2278.37	4.4601
36	21057120	SAN JUAN	46.03	43.22	14.11	10.07	6.55	7.66	30.24	15.79	4.57	4354.80	3256.00	7.7874
37	53047010	SANGARAL	414.69	105.57	46.21	32.64	18.19	22.98	274.93	139.76	12.70	3187.00	7.00	6.8816
38	44017070	SANTA ROSA - AUT	682.72	149.67	51.08	23.68	35.21	29.48	358.19	324.53	28.83	3549.00	1648.70	3.7202
39	26057040	TIMBA	395.39	121.83	33.86	26.92	15.62	22.44	320.16	75.23	14.69	2393.00	999.00	4.1170
40	26027080	TOTORO	53.51	47.47	20.28	16.38	4.67	8.25	28.47	25.04	3.27	3441.00	2565.00	4.3195
41	21057110	TREBOL EL	285.36	103.15	42.02	28.11	13.51	19.06	176.75	108.61	10.15	4385.55	1747.00	6.2793

Tabla 18 – Matriz global de variables independientes y 38 cuencas hidrográficas del Departamento del Cauca 1/3

Fuente. El Autor

#	Código	Nombre	Pendiente media cuenca (%) *R	Tc (horas) *prom	Ff	Ic	la	las	Re	C	CN-II	P_máx 24h prom
1	53077010	ANGOSTURA RIO MICAY	52.10	17.06	0.46	1.49	1.45	2.15	0.76	0.50	66.75	96.36
2	44017140	BETANIA-VILLALOBOS	35.01	10.53	0.44	1.70	1.64	1.81	0.75	0.40	52.00	81.83
3	26047020	BOCAT RIO PALO	43.35	9.34	0.46	2.10	1.39	3.43	0.77	0.49	66.92	66.64
4	52017070	BOCATOMA SAJANDI	37.05	3.28	0.83	1.38	0.72	1.07	1.03	0.51	73.07	90.99
6	26027250	BUENOS AIRES	14.15	5.50	0.43	1.41	1.69	1.88	0.74	0.44	60.60	85.62
7	26027090	CORTIJO EL	35.25	3.99	0.54	1.36	1.29	1.68	0.83	0.45	66.87	50.27
8	44017090	CURIACO	19.61	7.26	0.48	1.80	1.27	1.54	0.78	0.33	47.22	48.08
10	52027010	GUACHICONO	29.39	11.30	0.45	1.48	1.54	2.72	0.76	0.50	68.57	69.46
11	53077020	GUAYABAL	33.09	32.37	0.42	1.88	1.92	2.39	0.73	0.47	64.15	114.81
12	52017120	HOYO EL	22.99	10.09	0.26	1.80	2.55	2.56	0.58	0.50	71.10	81.9
13	26017020	JULUMITO	18.33	12.43	0.36	1.63	1.81	2.33	0.68	0.44	63.62	62.63
14	52017030	LA FONDA - AUT	24.03	13.43	0.65	1.68	1.20	2.12	0.91	0.48	69.61	84.07
15	26017070	LOMITAS	14.29	3.19	0.76	1.53	1.27	1.64	0.99	0.42	68.13	52.84
16	26027240	MALVASA	22.52	4.28	0.50	1.41	1.52	1.41	0.80	0.38	57.64	31.03
17	53077060	MONO EL	37.44	8.91	0.40	1.55	1.30	1.47	0.71	0.49	69.24	116.56
18	26017080	PALETARA	16.57	3.05	0.23	1.60	2.97	2.57	0.55	0.43	68.56	55.62
19	44017060	PAPAS	19.47	3.17	0.57	1.52	0.82	1.88	0.85	0.29	45.10	36.81
20	52027020	PLAYA LA	32.73	8.46	0.66	1.53	1.52	1.17	0.91	0.51	67.73	59.77
21	26017060	PUENTE ARAGON - AUT	13.94	3.48	0.53	1.61	2.52	7.47	0.82	0.42	67.85	54.01
22	26027200	PUENTE CARRETERA - AUT	27.50	8.13	0.16	2.19	3.37	1.83	0.45	0.43	60.17	47.37
23	52017110	PUENTE CARRETERA - AUT	27.85	6.61	0.19	1.86	3.81	1.11	0.49	0.50	69.91	74.9
24	26017040	PUENTE CARRETERA	26.90	2.60	0.30	1.39	2.19	4.52	0.61	0.46	59.68	70.3
25	26027100	PUENTE CARRETERA	35.14	7.39	0.25	1.76	2.40	1.18	0.56	0.46	66.60	54.72
26	26037030	PUENTE CARRETERA	14.14	6.24	0.24	1.86	2.70	5.24	0.55	0.35	53.00	71
27	52017080	PUENTE CARRETERA	47.88	4.92	0.45	1.43	1.72	2.82	0.76	0.49	67.22	88.8
28	52017100	PUENTE CARRETERA	17.58	6.30	0.24	1.80	3.13	4.58	0.55	0.44	65.49	75.89
29	52017130	PUENTE COLGANTE TIMBIO	24.16	10.74	0.31	2.03	2.13	9.01	0.63	0.47	66.90	82.43
30	26027210	PUENTE FERROCARRIL	14.22	5.01	0.59	1.40	1.31	1.02	0.87	0.47	67.94	89.96
31	52027030	PUENTE FIERRO - GUACHICONO	30.69	8.70	0.50	1.45	1.44	2.25	0.80	0.50	66.91	76.71
32	21057030	PUENTE RICAURTE	32.36	12.93	0.68	1.68	1.19	2.16	0.93	0.47	66.43	57.84
33	26037010	REMOLINO	15.37	3.67	0.92	1.46	1.10	4.70	1.08	0.47	64.21	84.6
35	44017050	RIOGRANDE - CAUCA	18.87	4.83	0.27	1.67	1.73	1.26	0.59	0.33	42.86	58.27
36	21057120	SAN JUAN	15.92	2.61	0.45	1.78	1.54	1.92	0.76	0.42	71.10	55.41
37	53047010	SANGARAL	22.15	6.60	0.39	1.45	1.79	1.97	0.70	0.37	49.74	148.73
38	44017070	SANTA ROSA - AUT	25.88	8.85	1.22	1.60	0.67	1.10	1.25	0.38	54.22	49.7
39	26057040	TIMBA	31.94	6.45	0.55	1.72	1.72	4.26	0.83	0.49	63.08	86.3
40	26027080	TOTORO	31.29	3.81	0.20	1.82	3.51	1.14	0.50	0.48	67.86	44.27
41	21057110	TREBOL EL	23.69	6.14	0.36	1.71	2.08	1.63	0.68	0.46	66.19	56.39

Tabla 19 – Matriz global de variables independientes y 38 cuencas hidrográficas del Departamento del Cauca 2/3

Fuente. El Autor

#	Código	Nombre	P_2.33yr	P_25yr	P_50yr	P_100yr	P_200yr	P_500yr	Q_2.33yr	Q_25yr	Q_50yr	Q_100yr	Q_200yr	Q_500yr
1	53077010	ANGOSTURA RIO MICAY	97.672	115.890	119.780	123.380	126.780	131.020	1160	1910.7	2126.6	2348.8	2579.1	2898.3
2	44017140	BETANIA-VILLALOBOS	83.594	109.380	114.880	119.980	124.770	130.740	347.56	574.02	626.36	675.93	723.41	783.67
3	26047020	BOCAT RIO PALO	68.376	83.619	86.558	89.201	91.621	94.553	142.83	317.97	374.44	434.87	499.76	593.15
4	52017070	BOCATOMA SAJANDI	94.678	127.11	133.36	138.99	144.13	150.37	91.713	216.44	266.53	324.19	390.22	491.73
6	26027250	BUENOS AIRES	85.827	104.48	109.01	113.35	117.55	122.94	23.326	33.906	36.748	39.568	42.378	46.085
7	26027090	CORTIJO EL	50.286	74.757	81.329	87.852	94.351	102.93	45.207	80.506	89.984	99.456	109	121.79
8	44017090	CURIACO	48.755	60.517	63.148	65.629	67.999	71.005	162.36	254.46	273.78	291.38	307.61	327.32
10	52027010	GUACHICONO	69.476	92.841	99.116	105.34	111.55	119.74	370.96	671.8	756.59	842.99	931.62	1053
11	53077020	GUAYABAL	114.83	137.62	143.74	149.81	155.86	163.85	1941.9	2423.1	2524	2617.1	2704.3	2812.4
12	52017120	HOYO EL	81.908	97.515	101.71	105.87	110.01	115.48	308.27	525.17	567.17	603.92	636.41	673.99
13	26017020	JULUMITO	62.835	82.614	87.784	92.92	98.07	104.95	148.46	239.87	264.41	288.78	313.05	345.08
14	52017030	LA FONDA - AUT	85.676	99.788	102.51	104.96	107.2	109.91	639.32	1058.1	1157.2	1251.6	1342.5	1458.5
15	26017070	LOMITAS	45.54	128.88	167.68	214.94	271.87	364.31	63.678	88.104	92.813	97.049	100.93	105.62
16	26027240	MALVASA	29.944	50.084	55.798	61.477	67.131	74.575	25.265	40.945	44.546	47.952	51.209	55.337
17	53077060	MONO EL	115.6	152.29	163.32	174.78	186.72	203.35	1550.3	1938.5	2042.7	2146.2	2249.3	2385.3
18	26017080	PALETARA	53.24	97.026	112.14	128.62	146.57	172.74	23.492	35.344	38.24	41.047	43.796	47.376
19	44017060	PAPAS	38.255	47.078	48.459	49.601	50.562	51.619	17.266	25.658	28.184	30.809	33.547	37.359
20	52027020	PLAYA LA	60.944	72.036	74.184	76.112	77.872	79.996	167.36	291.54	323.16	354.12	384.71	424.83
21	26017060	PUENTE ARAGON - AUT	50.419	108.25	130.66	156.61	186.72	234.1	71.502	117.73	129.61	141.32	152.96	168.35
22	26027200	PUENTE CARRETERA - AUT	47.926	60.015	62.73	65.283	67.713	70.777	37.492	75.447	87.158	99.502	112.57	131.12
23	52017110	PUENTE CARRETERA - AUT	74.376	95.65	101.94	108.44	115.19	124.52	80.6	112.24	118.34	123.83	128.85	134.94
24	26017040	PUENTE CARRETERA	72.833	95.121	99.419	103.28	106.82	111.11	36.381	94.965	110.84	126.46	141.89	162.07
25	26027100	PUENTE CARRETERA	54.742	78.948	85.448	91.9	98.329	106.81	68.517	118.3	132.11	146.11	160.41	179.9
26	26037030	PUENTE CARRETERA	72.709	87.704	90.595	93.195	95.575	98.46	45.868	123.790	145.600	167.230	188.710	216.940
27	52017080	PUENTE CARRETERA	91.848	118.61	123.77	128.41	132.66	137.81	76.797	187.8	230.35	279.35	335.85	424.05
28	52017100	PUENTE CARRETERA	77.349	90.166	92.637	94.86	96.894	99.359	87.46	246.28	324.1	420.77	539.39	735.77
29	52017130	PUENTE COLGANTE TIMBIO	83.719	98.851	101.96	104.81	107.47	110.74	259.62	446.33	498.28	550.82	604.32	676.9
30	26027210	PUENTE FERROCARRIL	89.641	114.96	121.45	127.76	133.93	141.92	90.68	240.28	294.21	354.45	421.90	523.22
31	52027030	PUENTE FIERRO - GUACHICONO	75.985	104.21	111.67	118.98	126.16	135.53	215.56	505.99	613.41	733.43	867.13	1066.8
32	21057030	PUENTE RICAURTE	57.007	74.88	79.857	84.786	89.677	96.099	608.98	983.93	1056.2	1121.3	1180.8	1252.9
33	26037010	REMOLINO	84.632	123.68	134.17	144.58	154.95	168.63	34.541	52.275	56.413	60.365	64.185	69.086
35	44017050	RIOGRANDE - CAUCA	58.284	79.41	85.084	90.715	96.326	103.73	85.693	138.37	152.52	166.56	180.55	199.00
36	21057120	SAN JUAN	55.433	82.778	90.121	97.41	104.67	114.25	13.04	17.773	18.918	20.025	21.109	22.519
37	53047010	SANGARAL	145.12	205.47	227.05	250.86	277.1	315.82	783.02	1004.8	1064.3	1123.4	1182.3	1260
38	44017070	SANTA ROSA - AUT	50.009	65.219	69.053	72.815	76.545	81.471	330.28	485.58	527.29	568.69	609.93	664.35
39	26057040	TIMBA	88.151	115.13	120.88	126.21	131.22	137.46	139.06	283.59	334.69	391.66	455.37	551.5
40	26027080	TOTORO	45.453	55.806	57.802	59.598	61.241	63.233	8.4665	21.963	26.972	32.575	38.824	48.166
41	21057110	TREBOL EL	58.026	72.428	75.205	77.703	79.989	82.76	102.53	133.6	139.59	144.98	149.91	155.88

Tabla 20 – Matriz global de variables independientes y 38 cuencas hidrográficas del Departamento del Cauca 3/3

Fuente. El Autor

En la exploración de los variables que se presumen como unos de los más importantes para la confección del modelo se percibe cierta relación, ya que se ven reflejados los picos en ambas gráficas (ver figura 29).

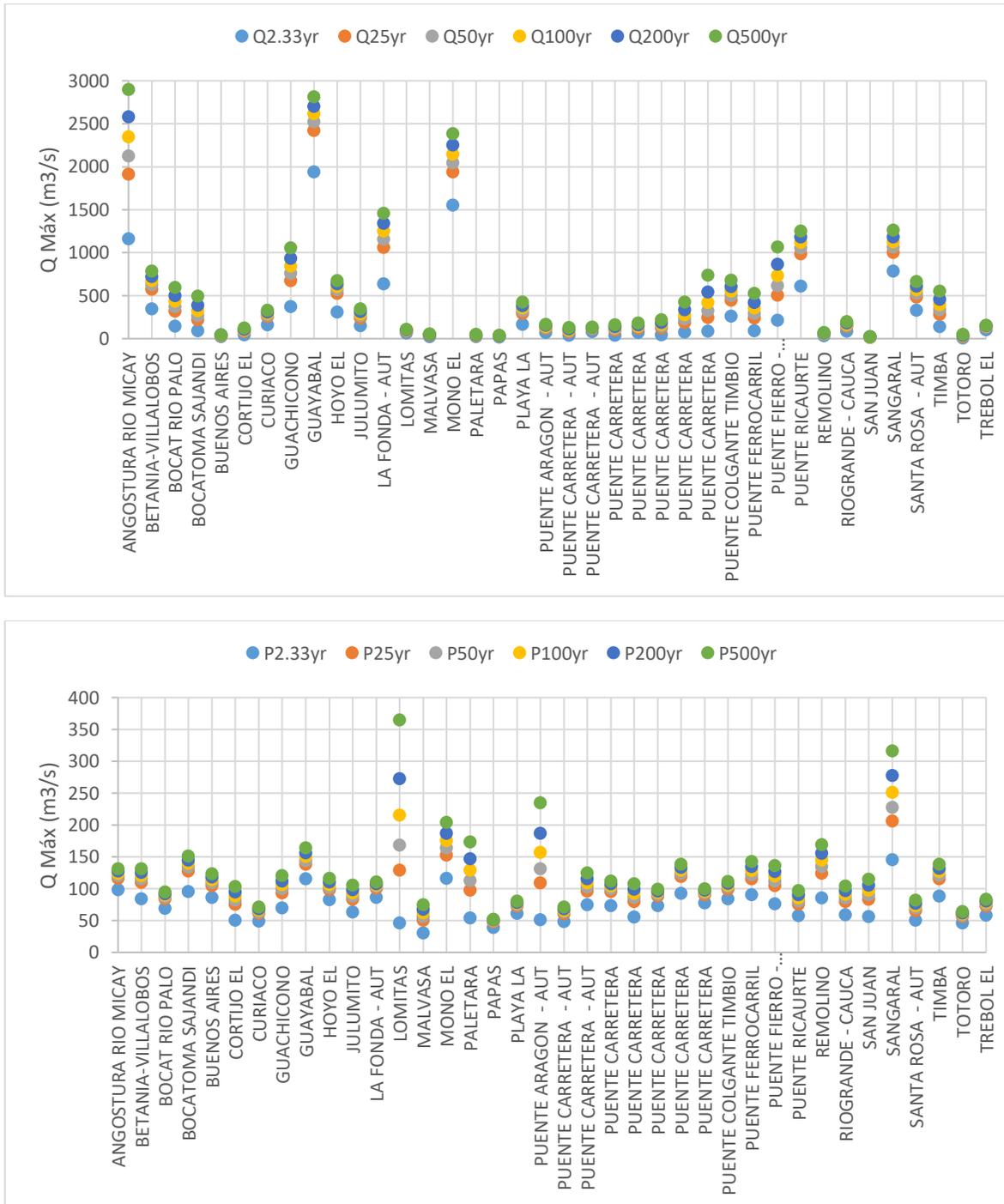


Figura 29 – Gráfica de dispersión para visualizar datos de PMáx y QMáx para diferentes Tr y todas las 38 de cuencas

Fuente. El Autor

6.4 Regresión Lineal Múltiple

Mediante códigos elaborados en R que están destinados a realizar un análisis exhaustivo de modelos de regresión lineal múltiple (RLM) para diferentes periodos de retorno (Tr), se verificó el cumplimiento de una serie de condiciones (checklist de criterios de RLM) para cada combinación de variables independientes ($2^{11} - 1 = 2047$), seleccionando los mejores modelos según el criterio AIC. Se tiene dos escenarios, el primero con todas las cuencas, y el segundo depurando las cuencas influyentes en el modelo.

6.4.1 Escenario 1

Se obtuvieron muy pocos modelos válidos, únicamente para los Tr de 200 y 500 años, para el primer Tr se resumen las variables independientes de cada modelo en la tabla 20, mientras que para el Tr de 500 años se resumen en la tabla 22.

Modelo 1	Pcuen	Tc	Ic	la	Ree	Pmax24	P200		
Modelo 2	Pcuen	Tc	Ic	la	Ias	Ree	Pmax24	P200	
Modelo 3	Pcuen	Tc	Ic	la	Ree	CN	Pmax24	P200	
Modelo 4	Pcuen	Tc	Ic	la	Ias	Ree	CN	Pmax24	P200

Tabla 21 – Variables independientes de modelos de RLM para Tr 200 años escenario 1

Fuente. El Autor

Modelo 1	Pcuen	Tc	Ic	Ias	Pmax24	P500			
Modelo 2	Pcuen	Tc	Ic	CN	Pmax24	P500			
Modelo 3	Pcuen	Tc	Ic	la	Ree	Pmax24	P500		
Modelo 4	Pcuen	Tc	Ic	Ias	CN	Pmax	P500		
Modelo 5	Pcuen	Tc	Ic	la	Ias	Ree	Pmax24	P500	
Modelo 6	Pcuen	Tc	Ic	la	Ias	Ree	CN	Pmax24	P500

Tabla 22 – Variables independientes de modelos de RLM para Tr 500 años escenario 1

Fuente. El Autor

En las tablas existe una discrepancia, en cuanto a que ningún modelo considero como una variable relevante el Área de drenaje, lo cual podría estar indicando que si se incluye se recae en la colinealidad, ya que el Área podría estar altamente correlacionada con el tiempo de concentración.

Para determinar el mejor modelo para un Tr de 200 años se analizó la figura 30, en donde se puede evidenciar que el modelo con un menor AIC es el modelo 1, el cual tiene en cuenta la pendiente de la cuenca, el tiempo de concentración, el índice de compacidad, índice de alargamiento, razón de elongación, la precipitación máxima en 24 horas y la precipitación para un periodo de retorno para 200 años. El modelo correlaciona las variables así:

$$Q_{200} = -30.53 + 13.19P_{cuen} + 83.5T_c - 292.6I_c - 149.88l_a - 441.08R_{ee} + 6.5Pmáx_{24} + 1.81P_{200}$$

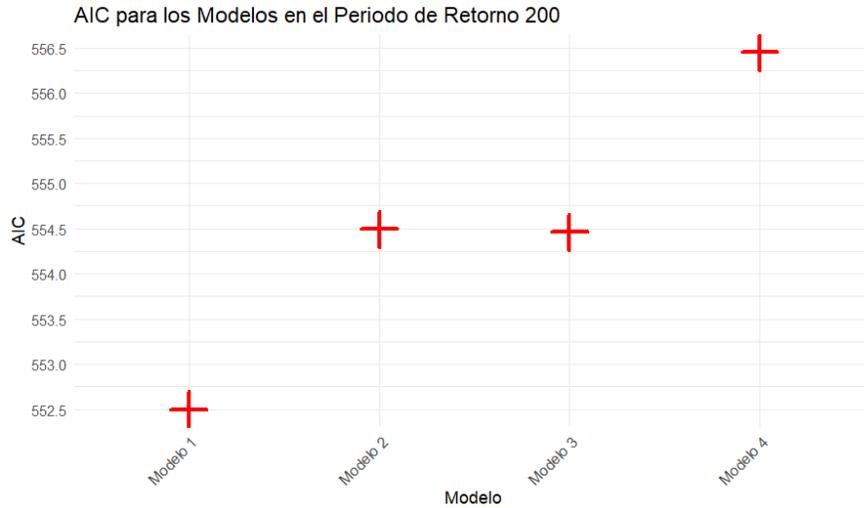


Figura 30 – AIC para Tr 200 escenario 1

Fuente. El Autor

Para determinar el mejor modelo para un Tr de 500 años se analizó la figura 31, en donde se puede evidenciar que el modelo con un menor AIC es el modelo 2, el cual tiene en cuenta la pendiente de la cuenca, el tiempo de concentración, el índice de compacidad, el número de curva de escorrentía, la precipitación máxima en 24 horas y la precipitación para un periodo de retorno para 500 años. El modelo correlaciona las variables así:

$$Q_{500} = -383.58 + 16.65P_{cuen} + 88.38T_c - 481.89I_c - 1.08CN + 8.44Pmáx_{24} + 1.03P_{500}$$

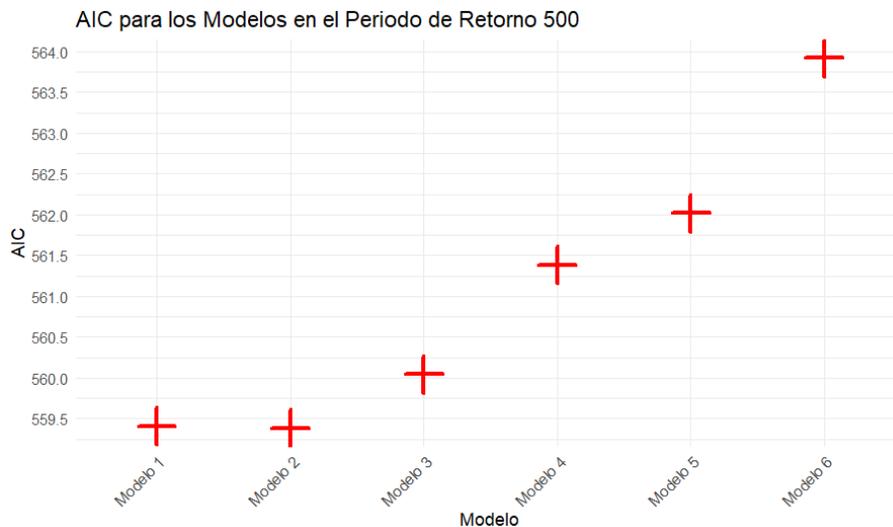


Figura 31 – AIC para Tr 500 escenario 1

Fuente. El Autor

En la figura 32 se presenta una matriz de correlación para el mejor modelo del Tr 200, y sobre la cual se puede inferir que Q200 está fuertemente relacionado con Tc y Pmax24, lo que sugiere que estos factores son importantes para predecir Q200, también Ic tiene una correlación significativa con la, lo que podría indicar una relación importante a considerar, y por último Ree presenta una correlación negativa fuerte con la, sugiriendo una relación inversa entre estas dos variables.

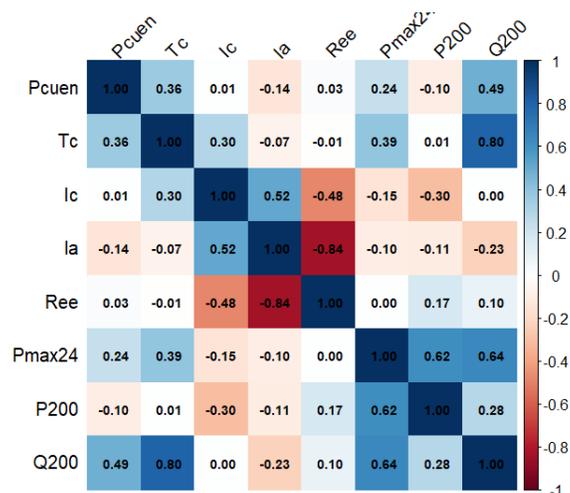


Figura 32 – Matriz de correlación para el mejor modelo Tr 200 escenario 1

Fuente. El Autor

En la figura 33 se presenta una matriz de correlación para el mejor modelo del Tr 500, y sobre la cual se puede inferir que Q500 está fuertemente correlacionado con Tc y Pmax24, lo que sugiere que estos factores son importantes para predecir Q500, también Pmax24 también tiene una correlación moderada con P500, lo que podría ser relevante para el análisis de Q500 y por último Ic muestra poca correlación con otras variables, sugiriendo que su relación con estas variables es débil.

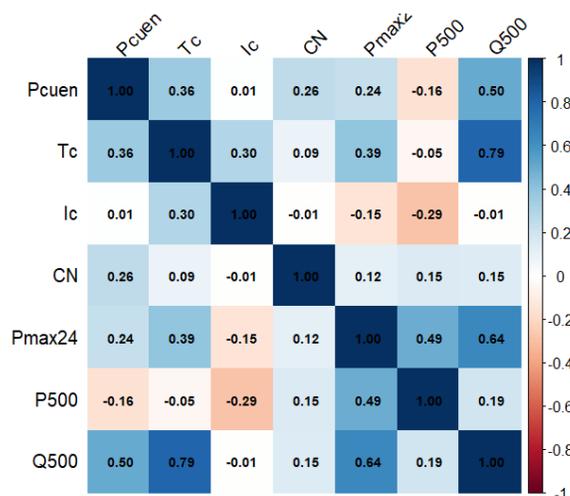


Figura 33 – Matriz de correlación para el mejor modelo Tr 500 escenario 1

Fuente. El Autor

6.4.2 Escenario 2

Realizando el análisis por distancia de Cook, se obtuvieron las cuencas influyentes resumidas en la tabla 23, las cuales fueron depuradas por generar el nuevo análisis de RLM.

#	Cód.	Nombre
1	53077010	ANGOSTURA RIO MICAY
11	53077020	GUAYABAL
17	53077060	MONO EL
21	26017060	PUENTE ARAGON - AUT
32	21057030	PUENTE RICAURTE
37	53047010	SANGARAL
38	44017070	SANTA ROSA - AUT

Tabla 23 – Cuencas influyentes con distancia de Cook depuradas de los modelos para escenario 2

Fuente. El Autor

Se generó el cálculo del área estudiada depurando las cuencas influyentes, obteniendo un área de 8.136,5 Km², correspondiendo a aproximadamente al 28% de la extensión del Departamento del Cauca y al 53% del área inicialmente explorada en el escenario 1.

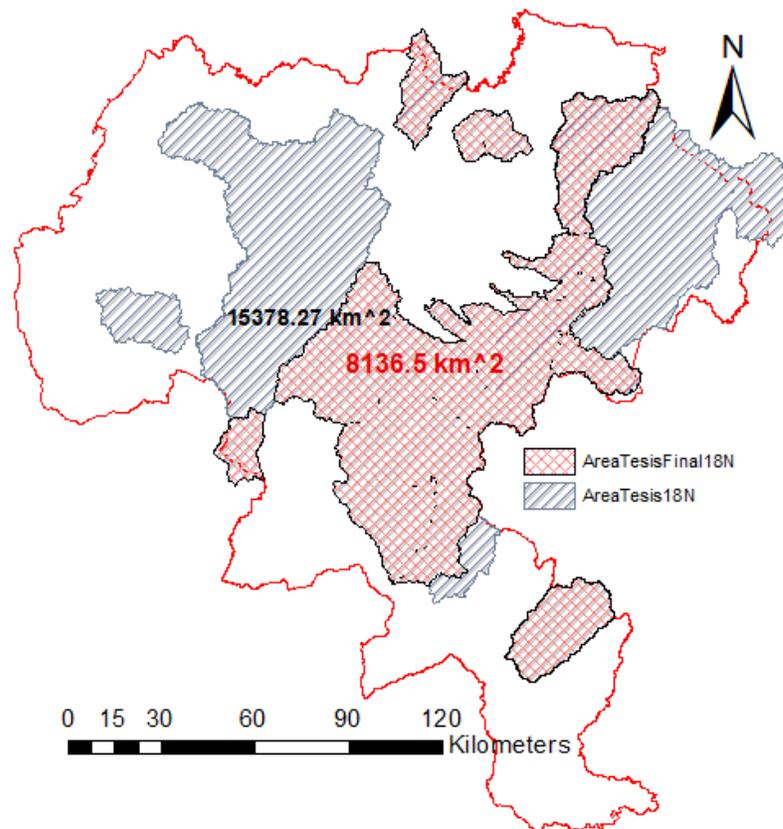


Figura 34 – Área de estudio escenario 2 vs área de estudio escenario 1

Fuente. El Autor

Se obtuvieron 230 modelos válidos para el Tr de 2.33 años, 311 para Tr 25, 346 para Tr 50, 398 para Tr 100, 450 para Tr 200 y 545 para Tr 500, por lo que para realizar la evaluación del mejor AIC se seleccionan los primeros ocho modelos de cada Tr.

Se resumen las variables independientes de cada modelo en las tablas 24 a 29.

Modelo 1	A	Ias	Pmax24	
Modelo 2	A	Pcuen	Ias	Pmax24
Modelo 3	A	Ias	CN	Pmax24
Modelo 4	A	Ic	Pmax24	
Modelo 5	A	Pcuen	Ias	P2.33
Modelo 6	A	Ias	P2.33	
Modelo 7	A	CN	Pmax24	
Modelo 8	A	Ic	P2.33	

Tabla 24 – Variables independientes de modelos de RLM para Tr 2.33 años escenario 2

Fuente. El Autor

Modelo 1	A	Ic	Pmax24		
Modelo 2	A	P25			
Modelo 3	A	Pmax24			
Modelo 4	A	CN	P25		
Modelo 5	A	Ff	Ic	P25	
Modelo 6	A	Ff	Ic	CN	P25
Modelo 7	A	Ic	Ree	CN	P25
Modelo 8	A	Ias	Pmax24		

Tabla 25 – Variables independientes de modelos de RLM para Tr 25 años escenario 2

Fuente. El Autor

Modelo 1	A	Ic	Pmax24	
Modelo 2	A	Pmax24		
Modelo 3	A	Ias	Pmax24	
Modelo 4	A	Ff	Ic	Pmax24
Modelo 5	A	Ic	Ree	Pmax24
Modelo 6	A	Ic	CN	Pmax24
Modelo 7	A	CN	Pmax24	
Modelo 8	A	Pmax24	P50	

Tabla 26 – Variables independientes de modelos de RLM para Tr 50 años escenario 2

Fuente. El Autor

Modelo 1	A	Ic	Pmax24	
Modelo 2	A	Pmax24		
Modelo 3	A	Ff	Ic	Pmax24
Modelo 4	A	Ic	Ree	Pmax24
Modelo 5	A	Ias	Pmax24	
Modelo 6	A	Ic	CN	Pmax24
Modelo 7	A	CN	Pmax24	
Modelo 8	A	A	Pmax	P100

Tabla 27 – Variables independientes de modelos de RLM para Tr 100 años escenario 2

Fuente. El Autor

Modelo 1	A	Pmax24		
Modelo 2	A	Ic	Pmax24	
Modelo 3	A	Ff	Ic	Pmax24
Modelo 4	A	Ic	Ree	Pmax24
Modelo 5	A	Ias	Pmax24	
Modelo 6	A	la	Pmax24	
Modelo 7	A	CN	Pmax24	
Modelo 8	A	Pmax24	P200	

Tabla 28 – Variables independientes de modelos de RLM para Tr 200 años escenario 2

Fuente. El Autor

Modelo 1	A	Pmax24		
Modelo 2	A	Ic	Pmax24	
Modelo 3	A	Pcuen	Pmax24	
Modelo 4	A	Ias	Pmax24	
Modelo 5	A	la	Pmax24	
Modelo 6	A	Ff	Ic	Pmax24
Modelo 7	A	CN	Pmax24	
Modelo 8	A	Pmax24	P500	

Tabla 29 – Variables independientes de modelos de RLM para Tr 500 años escenario 2

Fuente. El Autor

Se resalta que en este escenario aparece de manera significativa en todos los Tr una de las variables que se presumían desde un inicio como las más importantes para la estimación de caudales máximos, el Área de drenaje, la cual es bien sabida, tiene una relación directa y positiva en los cálculos (Chow, Maidment, & Mays, 1988). Es posible que la aparición de esta variable haya sido generada por influencia de ciertos puntos extremos que hayan estado ocultando su efecto real. Se podría ratificar la colinealidad con la variable Tc, al haber eliminado las cuencas influyentes del modelo.

También la inclusión de la variable Área de drenaje después de ajustar los datos puede mejorar el ajuste del modelo y proporcionar una mejor representación de la relación entre las variables. Esto sugiere que la variable es realmente relevante y su exclusión inicial puede haber sido un efecto de la influencia de las cuencas influyentes.

Para determinar el mejor modelo para los diferentes Tr se analizó la figura 35, en donde se puede evidenciar que el modelo con un menor AIC para 2.33 años es el modelo 1, el cual tiene en cuenta el área, el índice asimétrico y la precipitación máxima en 24 horas. Para 25 años es el modelo 1 (al igual que para 50 y 100 años), el cual tiene en cuenta el área, el índice de compacidad y la precipitación máxima en 24 horas. Para 200 y 500 años fueron los modelos 1, el cual tiene en cuenta el área y la precipitación máxima en 24 horas.

El modelo correlaciona las variables así para cada Tr:

$$Q_{2.33} = -55.2597 + 0.3364A - 7.1408I_{as} + 1.1059Pmáx_{24}$$

$$Q_{25} = 22.3445 + 0.5733A - 93.5653I_c + 2.2402Pmáx_{24}$$

$$Q_{50} = 14.571 + 0.6277A - 107.3706I_c + 2.8583Pmáx_{24}$$

$$Q_{100} = 3.0872 + 0.679A - 121.5349I_c + 3.5812Pmáx_{24}$$

$$Q_{200} = -255.0102 + 0.7036A + 4.8163Pmáx_{24}$$

$$Q_{500} = -319.2288 + 0.7601A + 6.2004Pmáx_{24}$$

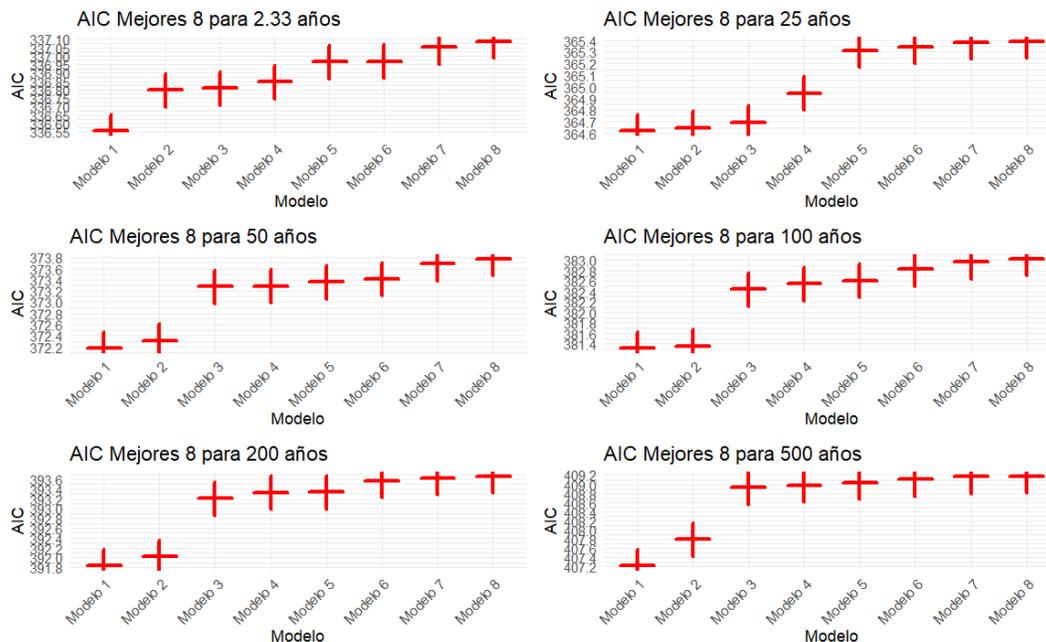


Figura 35 – AIC para todos los Tr escenario 2

Fuente. El Autor

En la figura 36 se presenta una matriz de correlación para el mejor modelo del Tr 2.33, y sobre la cual se puede inferir que A y Q2.33 tienen una correlación muy alta de 0.93, lo que indica una relación fuerte y positiva entre estas dos variables, también A e Ias tienen una correlación muy baja de 0.17, indicando una relación débil entre estas dos variables, adicionalmente Pmax24 y Q2.33 muestran una correlación positiva moderada de 0.40, lo que sugiere que cuando Pmax24 aumenta, Q2.33 también tiende a aumentar, pero la relación no es tan fuerte, y finalmente Ias y Q2.33 tienen una correlación muy baja de 0.11, indicando una relación débil entre estas dos variables.

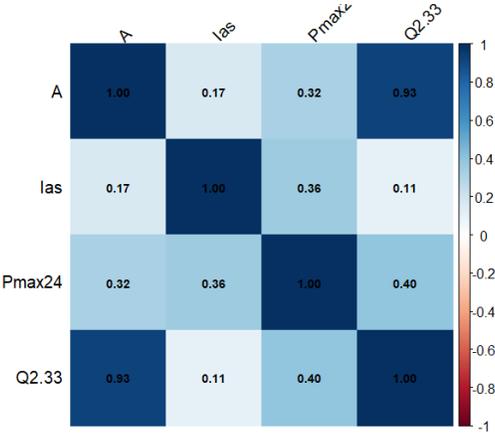


Figura 36 – Matriz de correlación para el mejor modelo Tr 2.33 escenario 2
Fuente. El Autor

En la figura 37 se presenta una matriz de correlación para el mejor modelo del Tr 25, y sobre la cual se puede inferir que A y Q25 tienen una correlación muy alta de 0.93, indicando una relación fuerte y positiva entre estas dos variables, también A y Pmax24 tienen una correlación positiva moderada de 0.32, sugiriendo que existe una relación positiva entre estas variables, aunque no tan fuerte como con Q25, adicionalmente Pmax24 y Q25 tienen una correlación positiva moderada de 0.46, lo que sugiere una relación positiva entre estas variables, aunque la relación no es tan fuerte como con A, y finalmente Ic y Q25 tienen una correlación muy baja de 0.09, lo que sugiere que no hay una relación significativa entre estas dos variables.

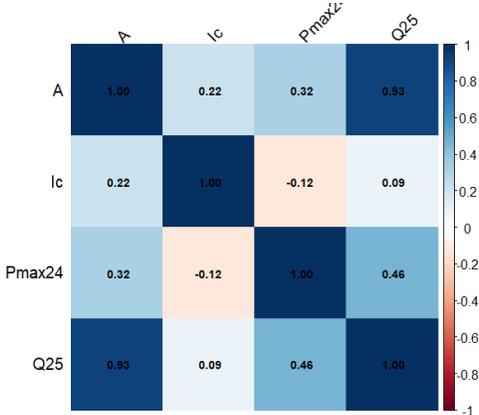


Figura 37 – Matriz de correlación para el mejor modelo Tr 25 escenario 2
Fuente. El Autor

En la figura 38 se presenta una matriz de correlación para el mejor modelo del Tr 50, y sobre la cual se puede inferir que la correlación de 0.92 indica una relación muy fuerte y positiva entre A y Q50, esto sugiere que las variaciones en A están estrechamente asociadas con las variaciones en Q50, lo que puede ser indicativo de una relación significativa en el modelo. La correlación de 0.48 entre Pmax24 y Q50 sugiere una relación positiva moderada, esto indica que, en general, un aumento en Pmax24 tiende a estar asociado con un aumento en Q50, aunque la relación no es tan fuerte como la observada entre A y Q50.

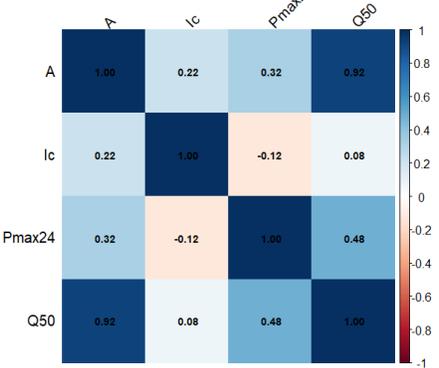


Figura 38 – Matriz de correlación para el mejor modelo Tr 50 escenario 2

Fuente. El Autor

En la figura 39 se presenta una matriz de correlación para el mejor modelo del Tr 100, y sobre la cual se puede inferir que la correlación de 0.91 entre A y Q100 indica una relación muy fuerte y positiva, esto sugiere que A tiene una asociación significativa con Q100, implicando que las variaciones en A están estrechamente vinculadas con las variaciones en Q100. Además, la correlación de 0.50 entre Pmax24 y Q100 indica una relación positiva moderada. Aunque no es tan fuerte como la relación entre A y Q100, sugiere que un aumento en Pmax24 tiende a estar asociado con un aumento en Q100.

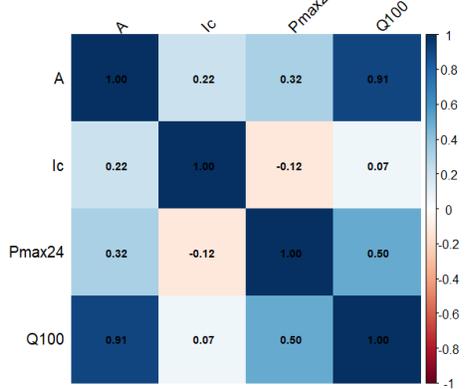


Figura 39 – Matriz de correlación para el mejor modelo Tr 100 escenario 2

Fuente. El Autor

En la figura 40 se presenta una matriz de correlación para el mejor modelo del Tr 200, y sobre la cual se puede inferir que la correlación de 0.89 entre A y Q200 es muy alta, lo que indica una relación fuerte y positiva, esto sugiere que A tiene una gran influencia en Q200, lo cual es consistente con la importancia de A en el modelo. También, la correlación de 0.52 entre Pmax24 y Q200 muestra una relación positiva moderada. Aunque no tan fuerte como

la relación entre A y Q200, esta correlación sugiere que Pmax24 también tiene un efecto significativo sobre Q200. Para finalizar, la correlación de 0.32 entre A y Pmax24 indica una relación positiva moderada. Aunque no tan alta como las correlaciones con Q200, sugiere que A y Pmax24 están algo relacionados, lo que puede influir en el comportamiento conjunto de las variables en el modelo.

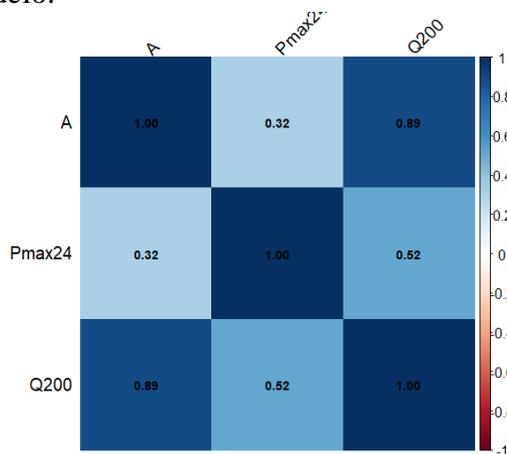


Figura 40 – Matriz de correlación para el mejor modelo Tr 200 escenario 2

Fuente. El Autor

En la figura 41 se presenta una matriz de correlación para el mejor modelo del Tr 500, y sobre la cual se puede inferir que la correlación de 0.86 entre A y Q500 indica una relación muy fuerte y positiva, esto muestra que A tiene una influencia significativa en Q500, sugiriendo que A es un predictor importante para Q500, comprobando las suposiciones iniciales que indicaban el Área de drenaje como variable importante para la predicción de caudales máximos. Segundo, la correlación de 0.54 entre Pmax24 y Q500 indica una relación positiva moderada, esto sugiere que Pmax24 también tiene una influencia notable sobre Q500, aunque no tan fuerte como A. Por último, la correlación de 0.32 entre A y Pmax24 muestra una relación positiva moderada. Aunque esta correlación es más baja en comparación con las correlaciones con Q500, indica que hay una conexión entre A y Pmax24, lo que podría ser relevante para el análisis del modelo.

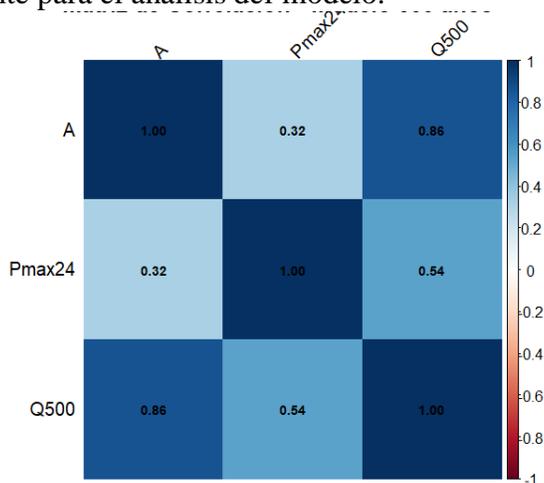


Figura 41 – Matriz de correlación para el mejor modelo Tr 500 escenario 2

Fuente. El Autor

6.5 Rendimiento de los modelos

Para los mejores modelos definidos por el AIC, se realizó validación cruzada (LOOCV), obteniendo tablas de error y gráficas, como por ejemplo valores observados vs. valores calculados, residuales vs. predicciones, histograma de residuales, QQ-Plot de residuales y distribución de errores absolutos, las cuales facilitaron la realización de un análisis más detallado, para finalmente obtener las métricas de eficiencia por cada Tr para los dos escenarios.

6.5.1 Escenario 1

Se muestra en la tabla 30 cómo se evalúa el rendimiento del modelo para 200 años al entrenarlo en todos los datos menos uno, y luego evaluarlo en ese único dato excluido.

Cuenc	(Intercept)	Pcuen	Tc	Ic	la	Ree	Pmax24	P200	Error
1	-249.64	7.62	76.23	-103.95	-154.27	-339.36	7.12	1.60	879.13
2	-49.80	13.82	83.49	-272.39	-157.82	-462.54	6.51	1.86	-285.25
3	-338.55	15.63	79.53	-59.06	-195.81	-483.89	6.43	2.00	-424.72
4	-142.10	14.23	80.55	-284.78	-147.36	-351.44	7.10	1.70	-288.02
6	192.19	11.01	83.44	-377.61	-151.72	-477.90	7.42	1.40	-419.86
7	98.23	14.04	83.40	-342.91	-154.69	-474.69	6.06	1.84	-234.18
8	-61.05	13.51	83.61	-315.15	-136.82	-410.63	6.57	1.83	119.68
10	-16.42	13.19	83.69	-297.79	-150.38	-445.85	6.47	1.81	-24.24
11	-72.97	11.21	104.67	-358.72	-127.13	-422.93	5.92	2.19	-778.37
12	-29.77	13.09	83.57	-291.60	-149.09	-444.27	6.55	1.79	-49.84
13	367.29	10.99	89.35	-377.77	-175.02	-616.01	5.76	1.81	-554.31
14	29.13	14.04	82.26	-308.94	-149.49	-503.68	6.06	2.01	302.39
15	-51.73	13.33	84.41	-261.32	-161.94	-459.18	4.96	2.75	-258.87
16	-125.39	13.26	82.95	-260.33	-148.16	-421.63	6.81	1.82	118.01
17	-237.61	10.07	89.38	-381.57	-34.33	17.90	4.83	1.14	1161.66
18	-91.82	13.35	83.18	-266.46	-156.84	-416.43	6.82	1.67	155.63
19	-180.93	13.67	83.84	-286.97	-126.43	-373.63	6.69	1.91	196.65
20	-68.84	13.52	83.79	-315.73	-135.15	-363.74	6.53	1.66	-195.95
21	100.96	13.46	82.77	-279.97	-186.68	-571.41	7.27	1.37	319.55
22	-17.46	13.15	83.56	-300.32	-150.62	-441.24	6.52	1.80	25.37
23	-138.05	13.48	83.14	-295.52	-120.92	-372.83	6.68	1.76	-167.95
24	4.92	13.21	83.37	-306.81	-149.59	-453.87	6.51	1.79	-51.93
25	75.23	14.48	83.66	-291.35	-161.20	-556.36	5.76	2.03	-346.89
26	-30.94	13.52	83.69	-300.37	-150.22	-435.29	6.37	1.86	80.26
27	-69.03	17.89	78.93	-361.94	-132.24	-419.91	6.96	1.80	-632.81
28	83.68	14.50	84.39	-284.72	-194.80	-544.80	5.51	2.24	456.68
29	-82.32	12.96	83.05	-243.88	-158.01	-457.28	6.71	1.78	-141.70
30	-33.06	13.25	83.53	-291.15	-149.93	-442.32	6.47	1.82	12.11
31	-43.71	13.16	83.40	-286.68	-149.50	-437.64	6.51	1.81	35.35

32	12.26	12.87	81.75	-297.21	-151.92	-498.03	6.85	1.81	280.80
33	-181.54	12.43	82.20	-283.18	-131.89	-290.30	7.09	1.66	-231.49
35	-44.36	13.23	83.52	-293.02	-147.63	-430.41	6.51	1.80	16.05
36	-40.24	13.68	84.80	-338.90	-132.44	-404.19	6.49	1.79	190.13
37	-43.72	12.91	82.42	-287.03	-155.13	-471.84	7.05	1.95	-167.02
38	315.62	13.32	83.31	-324.23	-189.03	-784.65	6.28	2.15	359.86
39	-131.61	13.51	82.33	-265.00	-145.08	-393.68	6.79	1.76	-160.47
40	49.28	12.21	84.36	-273.89	-188.14	-512.39	6.64	1.93	289.56
41	-22.34	13.15	83.46	-292.35	-150.33	-444.83	6.49	1.79	-46.75

Tabla 30 – Cálculo del error para Tr 200 años escenario 1

Fuente. El Autor

Se observan errores altamente variables, ya que hay una amplia gama de errores, desde valores negativos (que indican un buen ajuste si se considera el signo de la métrica) hasta errores positivos muy altos. La alta variabilidad en los errores puede sugerir que el modelo tiene un desempeño inconsistente en diferentes cuencas. Esto podría indicar la presencia de cuencas con características atípicas o problemas específicos que afectan la capacidad del modelo para generalizar (cuencas influyentes), como por ejemplo se tienen algunas cuencas como la cuenca 17 y la cuenca 38 (nótese que estas cuencas se encuentran también en la tabla 23 como cuencas influyentes y depuradas para el escenario 2) tienen errores extremadamente altos (1161.66 y 359.86, respectivamente), que son mucho mayores que la mayoría de los otros errores. Estos errores tan altos sugieren que el modelo podría estar teniendo dificultades para predecir correctamente en estas cuencas específicas, por lo que se hace necesario la depuración por la distancia de Cook. Otra observación muy importante es que la mayoría de los errores tienen signo negativo, lo que indica que el modelo ha hecho predicciones que, en promedio, están subestimando los valores reales. Los errores negativos generalizados podrían sugerir que el modelo tiene una tendencia sistemática a subestimar los valores en lugar de sobreestimarlos. Esto podría ser indicativo de un sesgo en el modelo que necesita ser ajustado.

Se muestra en la tabla 31 cómo se evalúa el rendimiento del modelo para 500 años al entrenarlo en todos los datos menos uno, y luego evaluarlo en ese único dato excluido.

Cuenc	(Intercept)	Pcuen	Tc	Ic	CN	Pmax24	P500	Error
1	-503.67	9.93	81.11	-330.77	0.00	8.89	0.86	1025.44
2	-306.55	18.04	88.48	-463.22	-3.45	8.45	1.12	-345.15
3	-519.03	17.94	87.14	-411.68	-1.13	8.40	1.08	-200.78
4	-416.54	17.34	86.37	-508.20	-0.10	8.81	0.96	-254.41
6	-214.79	14.11	88.15	-561.99	-0.74	9.37	0.69	-472.32
7	-297.12	17.58	88.13	-537.03	-0.76	8.01	1.02	-277.18
8	-475.36	16.74	88.13	-497.89	0.47	8.61	1.04	223.20
10	-389.48	16.67	88.17	-476.34	-1.16	8.47	1.03	31.42
11	-416.02	13.95	115.74	-513.82	-1.72	7.96	1.39	-1046.18
12	-417.08	16.34	88.42	-474.48	-0.64	8.56	0.99	-94.49

13	-255.60	14.40	93.38	-542.65	-0.28	8.04	0.91	-522.77
14	-349.04	17.96	86.19	-464.33	-2.52	8.14	1.19	358.56
15	-395.89	16.77	88.89	-477.06	-1.06	7.87	1.33	-120.79
16	-455.17	16.67	87.98	-457.68	-0.89	8.69	1.04	106.15
17	-26.24	13.34	93.98	-514.68	-2.44	5.94	0.83	1182.56
18	-377.31	16.85	88.49	-481.43	-1.37	8.54	0.99	91.55
19	-640.59	16.66	88.39	-443.29	1.25	8.80	1.12	308.96
20	-342.76	16.91	89.00	-518.15	-0.44	8.26	0.93	-282.42
21	-364.89	17.02	88.15	-487.46	-1.51	8.90	0.80	207.66
22	-391.41	16.68	88.32	-475.70	-1.11	8.43	1.03	-16.74
23	-554.22	16.71	86.57	-417.66	0.01	8.74	0.99	-349.30
24	-340.67	16.72	87.84	-498.60	-1.24	8.51	0.98	-115.92
25	-441.82	17.92	87.92	-454.29	-0.86	7.99	1.11	-357.60
26	-388.23	16.90	88.61	-493.20	-0.81	8.34	1.05	76.34
27	-409.97	21.95	82.63	-525.40	-1.38	8.98	1.04	-718.29
28	-266.93	18.73	90.23	-534.56	-2.47	7.53	1.33	522.02
29	-415.37	16.54	88.27	-466.85	-0.95	8.52	1.01	-57.36
30	-395.09	17.17	88.50	-467.51	-1.45	8.25	1.09	87.74
31	-411.72	16.57	87.99	-456.98	-1.32	8.47	1.04	164.10
32	-405.38	16.40	86.28	-464.86	-1.40	8.89	1.01	273.58
33	-352.41	15.60	87.57	-504.78	-0.64	8.82	0.97	-270.39
35	-454.33	16.61	88.58	-485.36	0.06	8.47	1.02	121.35
36	-277.57	17.72	89.93	-515.85	-2.69	8.35	1.10	306.05
37	-345.91	16.80	86.22	-463.16	-3.61	9.58	1.27	-307.79
38	-472.80	16.55	87.52	-465.42	-0.34	8.71	1.03	182.40
39	-412.99	16.97	87.48	-465.69	-1.20	8.63	1.02	-152.24
40	-359.16	16.43	88.94	-493.52	-1.30	8.50	1.05	96.25
41	-387.06	16.55	88.30	-480.76	-0.92	8.44	1.00	-66.98

Tabla 31 – Cálculo del error para Tr 500 años escenario 1

Fuente. El Autor

Al igual que en la tabla 30 se presenta una gama amplia de errores, que van desde valores negativos muy bajos hasta positivos elevados. Esto indica una variabilidad significativa en el desempeño del modelo para diferentes cuencas, recalcando la necesidad de realizar el análisis de cuencas outliers (escenario 2).

Algunas cuencas, como la cuenca 11 y la cuenca 27, tienen errores negativos significativos (-1046.18 y -718.29, respectivamente), o el caso contrario, las cuencas con errores positivos significativos, como la cuenca 8 y la cuenca 28, tienen errores de 223.20 y 522.02, respectivamente, indicando que el modelo ha subestimado o sobreestimado sustancialmente los valores reales en estas cuencas. Los errores elevados podrían señalar un sesgo sistemático del modelo en ciertas cuencas. Esto puede deberse a una falta de representatividad de los datos o a características particulares de estas cuencas que no están bien capturadas por el modelo.

Teniendo en cuenta la magnitud de los caudales para el periodo de retorno de 500 años, existen errores relativamente pequeños, indicando que el modelo puede estar funcionando relativamente bien para esas cuencas. Estos casos pueden servir como ejemplos de buenas prácticas o características que el modelo maneja bien (ver cuenca 10, 22 y 29).

La diferencia entre el error más alto (1182.56) y el error más bajo (-1046.18) es significativa, nuevamente indicando la necesidad de evaluar si ciertas cuencas deben ser excluidas del análisis.

En la figura 44 se aprecia la relación que existe entre valores de predicción y valores observados para los diferentes periodos de retorno, permitiendo comprobar que los puntos se encuentran alrededor de la línea, a excepción de unos cuantos puntos para valores grandes, esto indica que los modelos están haciendo predicciones precisas, en la mayoría de las cuencas y que tiene dificultades para capturar la relación en esos rangos específicos. Puede ser útil ajustar el modelo para mejorar el ajuste en esos rangos (escenario 2). También se puede comprobar que no hay presencia de sesgo sistemático y que se puede afirmar que el modelo no está sobrestimando ni subestimando mayormente. Los valores de la gráfica permiten calcular las métricas de rendimiento y eficiencia del modelo, ver tabla 32.

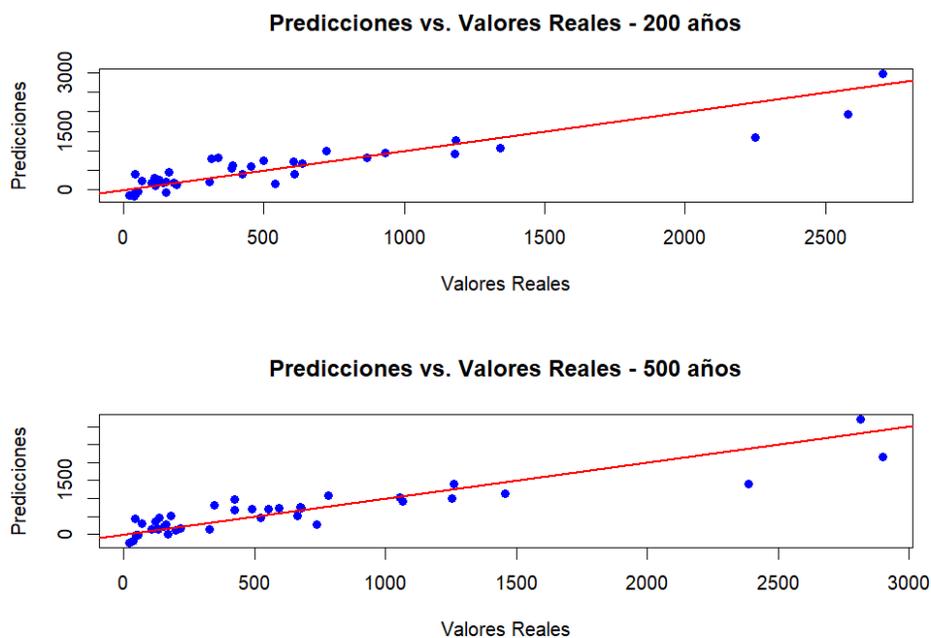


Figura 42 – Predicciones vs valores reales para el escenario 1

Fuente. El Autor

Tr	R^2	KGE
200	0.69368	0.8430011
500	0.6815405	0.8385571

Tabla 32 – Métricas de rendimiento para modelos del escenario 1

Fuente. El Autor

Ambos valores de R^2 son relativamente altos, lo que sugiere que el modelo tiene un buen poder explicativo en ambos periodos de retorno, sin embargo, no es perfecto y se puede mejorar. El R^2 para el periodo de retorno de 200 años es ligeramente superior al de 500 años, indicando que el modelo tiene un ajuste marginalmente mejor para el periodo de retorno de 200 años en comparación con el de 500 años. Sin embargo, la diferencia es pequeña, por lo que no hay una gran variabilidad en la capacidad explicativa entre estos dos periodos.

Los valores de KGE son muy altos en ambos periodos, indicando que el modelo tiene una excelente capacidad para capturar la variabilidad y la tendencia de los datos observados. El KGE para el periodo de retorno de 200 años es ligeramente superior al de 500 años, similar a lo observado con la anterior métrica. Esto refuerza la idea de que el modelo ajusta marginalmente mejor en el periodo de retorno de 200 años en comparación con el de 500 años.

Se generaron gráficas adicionales para mejorar el análisis sobre los mejores modelos generados para los diferentes periodos de retorno, para inferir resultados favorables para la RLM.

- Residuos vs Predicciones: Los residuos están distribuidos aleatoriamente alrededor del 0, sin un patrón claro, se podría concluir que el modelo lineal captura adecuadamente la relación entre la variable. Se resalta el orden de magnitud de la gráfica, producido por los errores de magnitud considerable.

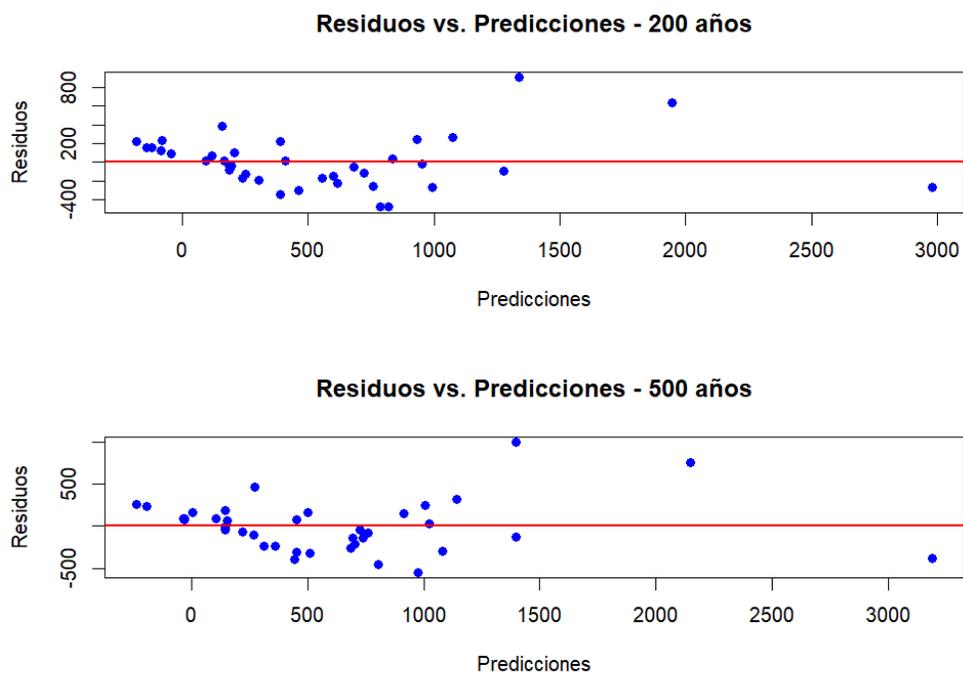


Figura 43 – Residuos vs predicciones para el escenario 1

Fuente. El Autor

- Histograma de Residuales: No se tiene una distribución normal, por lo que se infiere que los errores no están distribuidos simétricamente, es decir el modelo presenta sesgos, los cuales son generados por barras atípicas, generadas por cuencas influyendo el modelo, esto podría indicar que el modelo tiene errores sistemáticos y que no captura bien la variabilidad de los datos (o algunos).

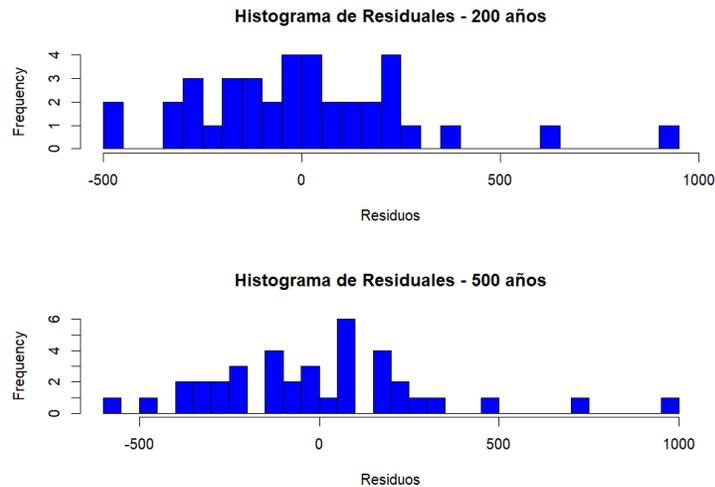


Figura 44 – Histograma de residuales para el escenario 1

Fuente. El Autor

- QQ-Plot de Residuales: Se puede visualizar que los puntos siguen la línea diagonal, pudiendo afirmar que los residuos se distribuyen normalmente, lo cual es una suposición importante en regresión lineal (Harvey, 1990), aun así, tiene puntos que se pueden mejorar.

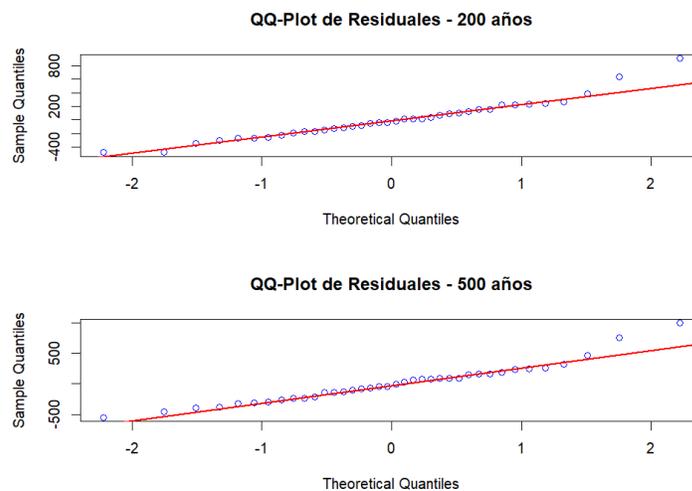


Figura 45 – QQ-Plot de residuales para el escenario 1

Fuente. El Autor

- **Distribución de Errores Absolutos:** Se resalta que la mayoría de los errores absolutos son pequeños, el modelo predice con precisión, aunque nuevamente se notan barras atípicas que podrían estar afectando el modelo.

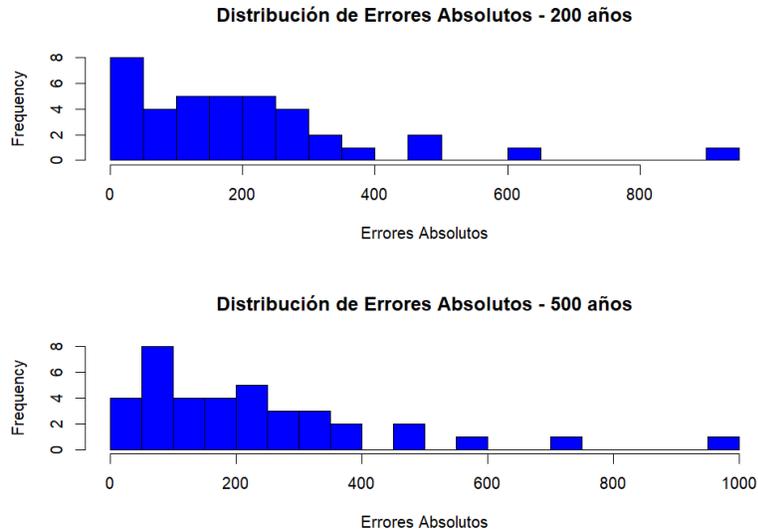


Figura 46 – Distribución de errores absolutos para el escenario 1
Fuente. El Autor

6.5.2 Escenario 2

Se muestra en la tabla 33 cómo se evalúa el rendimiento del modelo para 2.33 años.

Cuenc	(Intercept)	A	Ias	Pmax24	Error
2	-45.84	0.33	-5.09	0.86	119.18
3	-48.46	0.36	-5.57	0.91	-149.52
4	-55.71	0.34	-7.22	1.12	-2.76
6	-61.76	0.33	-7.99	1.29	-52.71
7	-56.30	0.34	-7.12	1.12	6.46
8	-61.68	0.33	-6.98	1.19	34.40
10	-56.68	0.33	-7.11	1.15	52.02
12	-48.19	0.33	-6.52	0.96	86.85
13	-47.63	0.35	-7.33	0.99	-99.81
14	-49.14	0.31	-5.58	1.08	88.36
15	-62.43	0.34	-6.92	1.16	51.70
16	-61.85	0.34	-7.19	1.20	20.90
18	-56.99	0.34	-7.24	1.12	15.11
19	-60.31	0.34	-7.26	1.17	18.88
20	-48.77	0.34	-8.16	1.04	-65.02
22	-49.02	0.34	-7.14	1.03	-33.32
23	-56.32	0.34	-7.87	1.17	-33.03

24	-55.05	0.34	-7.93	1.10	30.51
25	-51.89	0.34	-7.45	1.08	-25.84
26	-55.38	0.34	-6.70	1.10	-12.61
27	-60.12	0.33	-7.32	1.22	-32.97
28	-55.34	0.34	-7.10	1.11	-1.75
29	-49.25	0.33	-12.14	1.19	64.57
30	-55.01	0.34	-7.09	1.10	1.63
31	-56.99	0.34	-7.59	1.15	-47.74
33	-58.63	0.33	-6.53	1.16	-28.29
35	-59.53	0.34	-6.62	1.12	43.82
36	-55.91	0.34	-7.13	1.11	5.58
39	-55.78	0.34	-7.08	1.11	-4.15
40	-56.39	0.34	-7.10	1.12	5.45
41	-56.61	0.34	-7.06	1.12	11.61

Tabla 33 – Cálculo del error para Tr 2.33 años escenario 2

Fuente. El Autor

El análisis es similar al de la tabla 30, resaltando que las magnitudes del error de algunas cuencas son mucho más bajas (4, 7, 30, 36, 39 y 40), adicionalmente algo que se debe resaltar es el coeficiente de la variable A, el cual es relativamente constante, lo que sugiere que esta variable tiene un efecto similar en todas las cuencas. Un error positivo elevado, como en la cuenca 2, sugiere que el modelo sobreestimó el valor real. Un error negativo elevado, como en la cuenca 3, indica que el modelo subestimó el valor real.

Se muestra en la tabla 34 cómo se evalúa el rendimiento del modelo para 25 años.

Cuenc	(Intercept)	A	Ic	Pmax24	Error
2	45.74	0.57	-98.96	1.98	169.81
3	-53.75	0.59	-46.26	2.21	-173.19
4	20.83	0.58	-90.04	2.15	28.38
6	31.74	0.56	-108.69	2.61	-139.58
7	31.08	0.57	-97.06	2.20	-15.61
8	18.73	0.57	-97.03	2.37	52.59
10	-15.49	0.55	-73.82	2.38	99.37
12	60.44	0.57	-108.49	1.99	140.12
13	65.06	0.60	-108.92	1.94	-202.05
14	14.43	0.55	-86.77	2.28	58.38
15	7.09	0.58	-89.16	2.32	53.71
16	9.05	0.57	-89.36	2.33	18.38
18	22.15	0.57	-93.53	2.24	1.21
19	14.39	0.57	-91.52	2.30	16.09
20	62.92	0.59	-110.07	2.02	-114.48
22	28.13	0.57	-98.06	2.25	11.88
23	-1.83	0.57	-81.26	2.36	-65.05

24	18.77	0.57	-91.53	2.23	13.79
25	22.76	0.57	-91.83	2.21	-27.22
26	20.60	0.57	-92.60	2.25	-5.28
27	23.83	0.57	-98.31	2.37	-47.02
28	42.15	0.58	-103.05	2.12	68.32
29	15.39	0.57	-90.24	2.27	-12.48
30	19.78	0.58	-85.96	2.02	68.60
31	13.57	0.57	-87.98	2.24	30.90
33	25.10	0.56	-103.57	2.57	-129.40
35	18.90	0.58	-95.39	2.28	69.76
36	23.51	0.57	-95.06	2.25	12.88
39	22.88	0.57	-93.73	2.23	1.93
40	22.34	0.58	-98.81	2.33	45.13
41	23.50	0.57	-92.95	2.22	-19.61

Tabla 34 – Cálculo del error para Tr 25 años escenario 2

Fuente. El Autor

Al igual que en la tabla 33, el coeficiente de la variable A es bastante constante, lo que indica una influencia uniforme de esta variable en todas las cuencas. La cuenca 2 tiene un error positivo alto (169.81), lo que sugiere una sobreestimación por parte del modelo. La cuenca 3 tiene un error negativo significativo (-173.19), lo que sugiere que el modelo subestimó considerablemente el valor real. Cuencas como la 18 y la 39 tienen errores cercanos a cero, lo que sugiere que el modelo predijo bastante bien en estos casos. Las variaciones en los coeficientes de Ic y Pmax24 podrían estar reflejando diferencias en las características físicas o climáticas de las cuencas que no están siendo capturadas completamente por el modelo.

Se muestra en la tabla 35 cómo se evalúa el rendimiento del modelo para 50 años.

Cuenc	(Intercept)	A	Ic	Pmax24	Error
2	37.67	0.62	-112.69	2.60	167.62
3	-58.15	0.64	-62.16	2.83	-165.49
4	12.31	0.63	-102.12	2.72	42.22
6	26.27	0.61	-126.20	3.32	-173.79
7	25.39	0.63	-111.70	2.81	-19.34
8	11.03	0.63	-110.76	2.98	51.48
10	-31.49	0.60	-83.34	3.03	120.97
12	49.69	0.63	-121.13	2.63	129.16
13	62.66	0.65	-124.65	2.52	-227.48
14	10.49	0.62	-103.87	2.88	30.09
15	0.22	0.63	-103.23	2.93	50.53
16	-3.82	0.63	-101.55	2.98	25.42
18	15.27	0.63	-107.48	2.85	-4.29
19	4.36	0.63	-104.75	2.94	20.65
20	60.12	0.64	-125.90	2.61	-128.52
22	24.92	0.63	-115.42	2.88	21.25

23	-18.41	0.62	-90.59	3.02	-88.74
24	12.06	0.63	-105.94	2.85	9.68
25	15.05	0.63	-105.40	2.82	-30.93
26	12.77	0.63	-106.37	2.87	-5.45
27	16.00	0.62	-111.92	2.99	-45.01
28	49.54	0.64	-124.11	2.64	120.64
29	4.27	0.63	-102.45	2.90	-18.46
30	11.26	0.64	-97.55	2.58	88.61
31	-9.49	0.62	-92.07	2.85	84.72
33	18.02	0.61	-119.89	3.27	-161.94
35	11.03	0.63	-109.24	2.90	71.64
36	15.42	0.63	-108.46	2.86	9.38
39	17.45	0.63	-108.26	2.83	10.39
40	14.57	0.63	-113.62	2.97	53.74
41	16.53	0.63	-106.32	2.82	-33.27

Tabla 35 – Cálculo del error para Tr 50 años escenario 2

Fuente. El Autor

El coeficiente de la variable A es bastante consistente, con valores cercanos a 0.63 para la mayoría de las cuencas. Esto indica una influencia relativamente constante de esta variable en el modelo. Pmax24 también muestra variabilidad en sus coeficientes, aunque en un rango estrecho. Esto sugiere que su impacto es más uniforme entre las cuencas, pero sigue habiendo diferencias significativas. La cuenca 2 tiene un error positivo alto (167.62), lo que sugiere que el modelo sobreestimó el valor real. La cuenca 13 muestra un error negativo significativo (-227.48), lo que indica que el modelo subestimó fuertemente el valor real. Algunas cuencas, como la 18 y la 26, tienen errores cercanos a cero, lo que sugiere una predicción más precisa en estos casos.

Se muestra en la tabla 36 cómo se evalúa el rendimiento del modelo para 100 años.

Cuenc	(Intercept)	A	Ic	Pmax24	Error
2	25.14	0.67	-126.62	3.34	160.07
3	-64.03	0.69	-79.81	3.56	-152.74
4	-0.09	0.68	-114.15	3.39	59.35
6	17.43	0.66	-144.61	4.14	-213.05
7	16.38	0.68	-126.85	3.52	-23.76
8	-0.30	0.68	-124.78	3.70	49.21
10	-51.94	0.65	-92.82	3.78	144.51
12	33.01	0.68	-133.25	3.38	110.04
13	56.59	0.71	-140.76	3.21	-253.09
14	3.87	0.68	-122.20	3.58	-5.74
15	-9.89	0.68	-117.79	3.65	45.69
16	-21.49	0.68	-113.75	3.75	33.98
18	4.92	0.68	-121.83	3.57	-11.32

19	-9.83	0.68	-118.22	3.68	26.13
20	53.84	0.70	-142.18	3.31	-143.19
22	18.63	0.68	-133.62	3.62	31.92
23	-40.03	0.67	-99.59	3.79	-116.00
24	2.57	0.68	-121.24	3.58	1.99
25	3.63	0.68	-119.31	3.54	-34.97
26	0.29	0.68	-119.99	3.59	-8.46
27	4.37	0.68	-125.63	3.70	-40.55
28	58.27	0.69	-147.95	3.24	190.38
29	-11.36	0.68	-114.62	3.63	-25.91
30	-1.04	0.69	-109.29	3.23	110.48
31	-39.80	0.66	-94.26	3.57	150.99
33	7.33	0.66	-136.95	4.09	-199.32
35	-0.46	0.68	-123.42	3.62	71.95
36	3.48	0.68	-122.05	3.58	4.40
39	9.03	0.68	-123.38	3.52	21.47
40	3.09	0.68	-128.85	3.71	62.90
41	5.94	0.68	-120.01	3.53	-48.28

Tabla 36 – Cálculo del error para Tr 100 años escenario 2

Fuente. El Autor

Los valores del intercepto muestran menos variabilidad extrema en esta tabla en comparación con las anteriores. Esto podría indicar una mejor estabilidad en el modelo, aunque aún existe una amplia gama de valores. El coeficiente de la variable A es consistentemente cercano a 0.68, lo que indica una influencia relativamente uniforme en la mayoría de las cuencas. Los coeficientes de Pmax24 se han incrementado ligeramente en comparación con las tablas anteriores, lo que podría reflejar un ajuste en su impacto en el modelo. La cuenca 28, por ejemplo, muestra un error positivo alto (190.38), lo que indica una fuerte sobreestimación. La cuenca 13 presenta un error negativo significativo (-253.09), lo que indica subestimación. Varias cuencas, como la 26 y la 36, tienen errores cercanos a cero, lo que indica una mejora en la precisión para estas cuencas.

Se muestra en la tabla 37 cómo se evalúa el rendimiento del modelo para 200 años.

Cuenc	(Intercept)	A	Pmax24	Error
2	-243.58	0.70	4.61	140.69
3	-246.82	0.73	4.65	-174.86
4	-237.52	0.71	4.44	104.19
6	-284.04	0.68	5.50	-231.34
7	-257.71	0.70	4.85	16.74
8	-261.08	0.70	4.90	32.64
10	-260.46	0.67	4.98	200.77
12	-249.99	0.70	4.73	61.01
13	-235.05	0.73	4.49	-262.14
14	-255.54	0.71	4.81	-7.42

15	-262.58	0.71	4.89	54.76
16	-283.91	0.70	5.20	91.51
18	-253.28	0.70	4.80	-14.95
19	-270.76	0.70	5.02	58.70
20	-242.71	0.72	4.63	-125.79
22	-247.59	0.70	4.72	-39.67
23	-261.17	0.70	5.04	-175.90
24	-254.94	0.71	4.80	19.32
25	-248.40	0.70	4.74	-51.37
26	-255.26	0.70	4.86	-48.77
27	-256.81	0.70	4.85	-12.18
28	-244.46	0.71	4.48	246.26
29	-262.01	0.71	4.92	-84.43
30	-230.21	0.72	4.28	155.12
31	-245.46	0.68	4.67	254.50
33	-281.13	0.68	5.44	-224.46
35	-261.28	0.71	4.86	65.44
36	-252.12	0.70	4.79	-24.76
39	-252.78	0.70	4.77	17.97
40	-264.90	0.71	4.93	47.76
41	-246.87	0.70	4.73	-70.67

Tabla 37 – Cálculo del error para Tr 200 años escenario 2

Fuente. El Autor

En esta quinta tabla, se observa un ajuste del modelo con una reducción de variables, enfocándose en A y Pmax24, lo que podría ofrecer una perspectiva más clara sobre cómo estas variables están afectando el modelo y la precisión de las predicciones. El intercepto es consistentemente negativo para todas las cuencas, lo que indica una tendencia general en el modelo cuando las demás variables están en sus valores promedio. La magnitud del intercepto varía entre -230.21 y -284.04, lo que muestra cierta variabilidad, aunque es menor en comparación con tablas anteriores. El coeficiente de la variable A se mantiene relativamente constante, con valores alrededor de 0.70, lo que indica una influencia positiva consistente en el caudal. Pmax24 también tiene un coeficiente positivo constante alrededor de 4.70, sugiriendo que un aumento en esta variable se asocia consistentemente con un aumento en el caudal. La cuenca 31 tiene un error de 254.50, lo que indica una sobreestimación significativa. La cuenca 13 presenta un error negativo de -262.14, indicando subestimación. Varias cuencas como la 27 y la 36 tienen errores mucho más cercanos a cero, lo que indica una mejora en la precisión.

La reducción de variables a solo A y Pmax24 parece haber simplificado el modelo, lo que podría contribuir a una mejor estabilidad en los coeficientes observados. Aunque los errores aún son significativos en algunas cuencas, hay indicios de que la precisión ha mejorado con este modelo simplificado, especialmente en comparación con otros Tr donde se consideraban más variables.

Se muestra en la tabla 38 cómo se evalúa el rendimiento del modelo para 500 años.

Cuenc	(Intercept)	A	Pmax24	Error
2	-309.98	0.75	6.04	113.81
3	-311.96	0.78	6.05	-155.18
4	-295.80	0.77	5.70	139.53
6	-356.45	0.73	7.08	-296.58
7	-321.73	0.76	6.23	15.56
8	-323.93	0.76	6.27	25.24
10	-325.77	0.72	6.40	240.92
12	-318.33	0.76	6.18	10.93
13	-296.69	0.79	5.83	-295.93
14	-324.11	0.78	6.18	-68.17
15	-325.57	0.76	6.26	45.89
16	-355.42	0.76	6.68	114.58
18	-315.93	0.76	6.18	-28.60
19	-338.79	0.76	6.45	72.90
20	-305.25	0.77	5.98	-143.00
22	-312.43	0.76	6.12	-36.33
23	-327.17	0.75	6.49	-226.97
24	-319.22	0.76	6.20	2.03
25	-311.52	0.76	6.11	-59.93
26	-319.57	0.76	6.26	-67.33
27	-318.44	0.76	6.18	5.32
28	-302.37	0.78	5.66	393.44
29	-328.09	0.77	6.34	-106.89
30	-288.47	0.78	5.53	192.43
31	-305.00	0.73	5.98	379.17
33	-352.74	0.73	6.99	-287.99
35	-324.95	0.76	6.24	59.73
36	-314.63	0.76	6.17	-39.38
39	-314.50	0.76	6.10	38.18
40	-331.25	0.76	6.34	58.07
41	-308.20	0.76	6.08	-95.78

Tabla 38 – Cálculo del error para Tr 500 años escenario 2

Fuente. El Autor

El modelo parece seguir la misma línea de los Tr anteriores, pero con ajustes en los coeficientes, lo que sugiere un refinamiento adicional del modelo. El coeficiente de la variable A muestra un ligero incremento en comparación con las tablas anteriores, con valores entre 0.73 y 0.79. Esto sugiere que el área (A) sigue teniendo una influencia positiva significativa en el caudal, y su efecto parece haberse fortalecido en este modelo. Pmax24 presenta un coeficiente más elevado, rondando valores entre 5.53 y 7.08. Este aumento en los coeficientes podría indicar que la precipitación máxima en 24 horas tiene un impacto aún

mayor en el caudal en este ajuste, lo que podría estar mejorando la precisión del modelo en ciertas cuencas. La cuenca 28 tiene un error positivo de 393.44, lo que indica una sobreestimación notable. La cuenca 13 presenta un error negativo de -295.93, lo que muestra una subestimación significativa. Algunas cuencas, como la 24 y la 27, tienen errores cercanos a cero, lo que sugiere que el modelo está ajustando bastante bien en estos casos.

El modelo parece haber mejorado en términos de los coeficientes de A y Pmax24, con valores más altos que en las versiones anteriores, lo que podría estar proporcionando un ajuste más robusto.

En la figura 47 se aprecia la relación que existe entre valores de predicción y valores observados para los diferentes periodos de retorno, permitiendo comprobar que los puntos se encuentran alrededor de la línea, esto indica que los modelos están haciendo predicciones precisas. También se puede comprobar que no hay presencia de sesgo sistemático y que se puede afirmar que el modelo no está sobrestimando ni subestimando mayormente. Los valores de la gráfica permiten calcular las métricas de rendimiento y eficiencia del modelo, ver tabla 39. En general los modelos del escenario 2 han mejorado sobre el escenario 1.

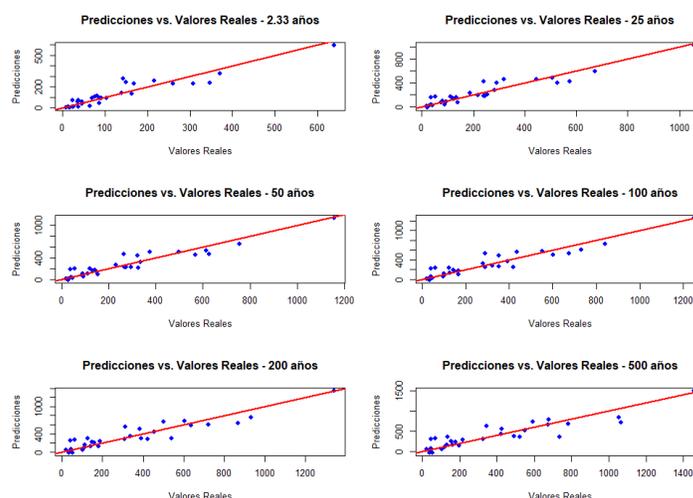


Figura 47 – Predicciones vs valores reales para el escenario 2

Fuente. El Autor

Tr	R^2	KGE
2.33	0.8376363	0.8666531
25	0.8690938	0.9075004
50	0.8660834	0.9095089
100	0.8532578	0.904916
200	0.8282861	0.8852935
500	0.7831108	0.8580413

Tabla 39 – Métricas de rendimiento para modelos del escenario 2

Fuente. El Autor

Los valores de R^2 varían entre 0.783 y 0.869, lo que indica que el modelo explica entre un 78.3% y un 86.9% de la variabilidad del caudal máximo para los diferentes periodos de retorno. El ajuste del modelo parece ser mejor para los periodos de retorno de 25 y 50 años, donde R^2 es más alto (~ 0.87), mientras que el ajuste es menor para el periodo de retorno de 500 años ($R^2 \sim 0.78$). Los modelos de 200 y 500 años aumentan su rendimiento 13% y 10% respectivamente frente al escenario 1.

Los valores de KGE varían entre 0.858 y 0.91, lo que sugiere que el modelo tiene un buen desempeño general para todos los periodos de retorno. Al igual que con R^2 , el mejor desempeño se observa para los periodos de retorno de 25 y 50 años, con KGE ~ 0.91 , y el desempeño es ligeramente menor para el periodo de 500 años (KGE ~ 0.86).

El modelo parece ajustarse mejor a los periodos de retorno intermedios (25 y 50 años), con métricas R^2 y KGE más altas. Para los periodos de retorno más extremos (2.33 y 500 años), aunque el desempeño sigue siendo bueno, es ligeramente inferior en comparación con los periodos intermedios, lo que podría indicar que el modelo es menos preciso en estas condiciones.

El análisis adicional que se tiene con las gráficas complementarias es:

- Residuos vs Predicciones: Los residuos están distribuidos aleatoriamente alrededor del 0, sin un patrón claro, se podría concluir que el modelo lineal captura adecuadamente la relación entre la variable. A diferencia del escenario 1, el orden de magnitud de los residuos es mucho menor.

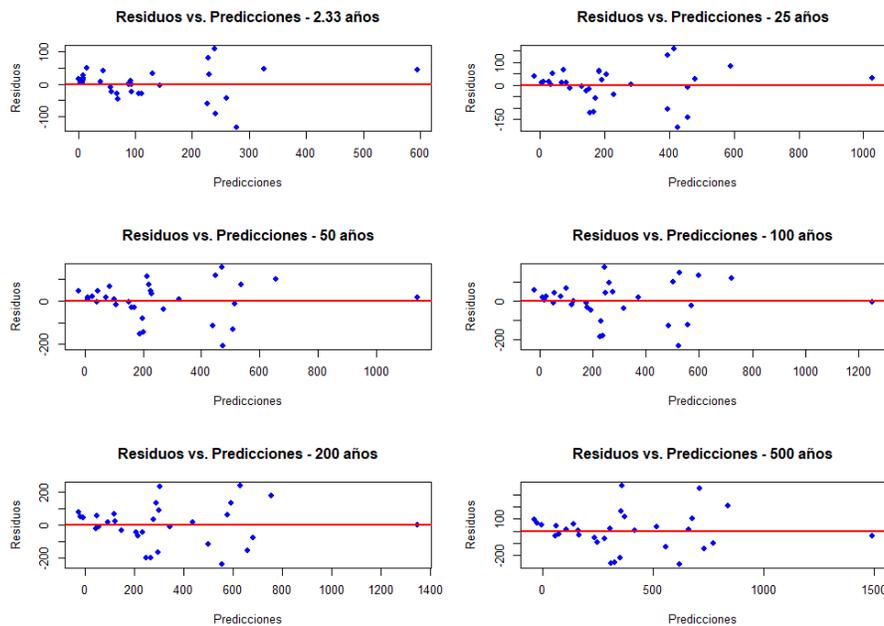


Figura 48 – Residuos vs predicciones para el escenario 2

Fuente. El Autor

- **Histograma de Residuales:** Los residuos siguen una distribución normal centrada en 0, lo que sugiere que los errores están distribuidos simétricamente y que el modelo no tiene sesgos.

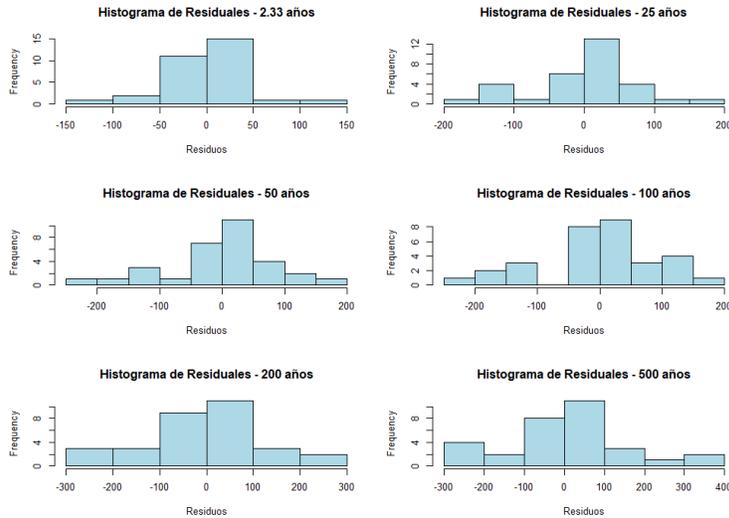


Figura 49 – Histograma de residuales para el escenario 2

Fuente. El Autor

- **QQ-Plot de Residuales:** Se puede visualizar que los puntos siguen la línea diagonal, pudiendo afirmar que los residuos se distribuyen normalmente, lo cual es una suposición importante en regresión lineal (Harvey, 1990). El modelo de 2.33 años esta notablemente mejor ajustado que los demás.

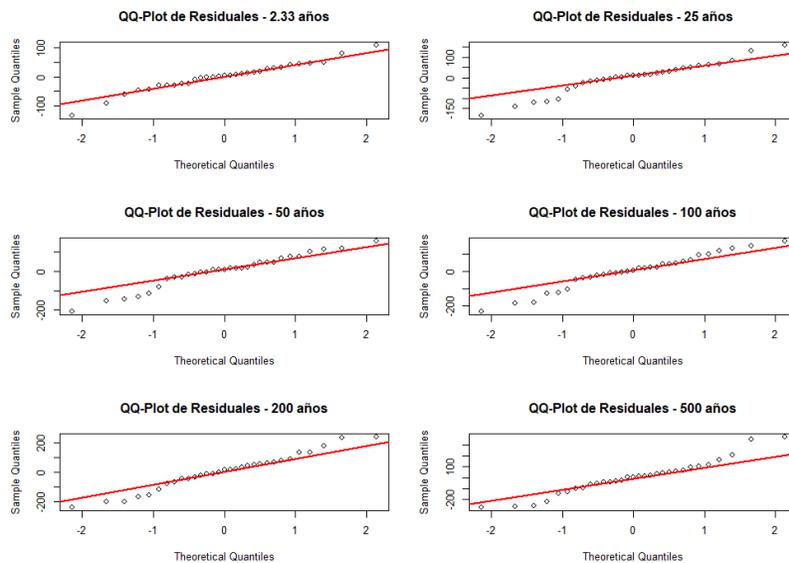


Figura 50 – QQ-Plot de residuales para el escenario 2

Fuente. El Autor

- **Distribución de Errores Absolutos:** Se resalta que la mayoría de los errores absolutos son pequeños, el modelo predice con precisión, se resalta que el orden de magnitud es mucho menor (hasta 4 veces menos) que en el escenario 1.

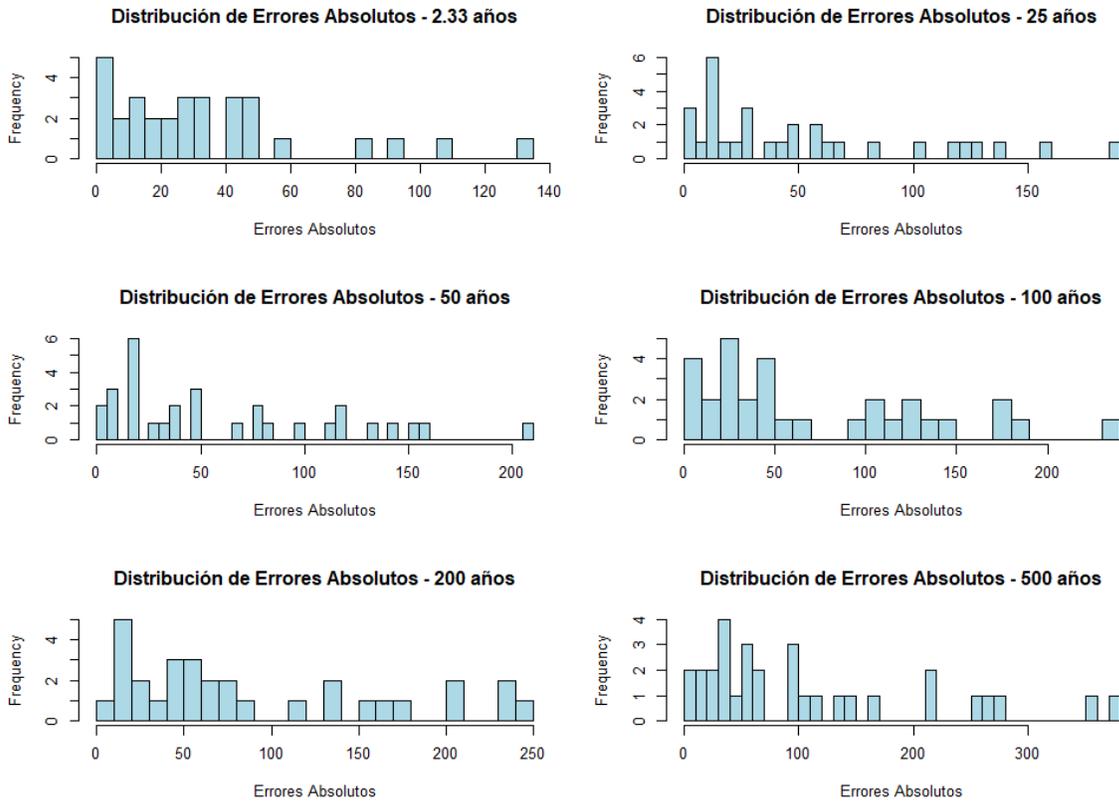


Figura 51 – Distribución de errores absolutos para el escenario 2

Fuente. El Autor

Teniendo en cuenta las experiencias de otras investigaciones que tienen relación con la temática del presente estudio alrededor del mundo (Matovelle & D., 2020; Méndez, Córdova, & Bravo de Guenni, 2015), se logró comprobar una correlación significativa entre las variables fisiográficas y parámetros morfométricas de las 38 cuencas estudiadas y las variables hidrológicas, precipitación máxima en 24 horas y caudales máximos, para diferentes periodos de retorno. Con este análisis se logró desarrollar un modelo mediante relaciones matemáticas para una zona determina entre cordilleras que permiten estimar caudales máximos en cuencas no instrumentadas.

7. Conclusiones y recomendaciones

- Se obtuvieron variables fisiográficas y parámetros morfométricos de 41 cuencas hidrográficas del Departamento del Cauca, cubriendo aproximadamente el 53% de su superficie. El análisis se realizó utilizando dos modelos digitales de elevación (DEM) diferentes, obteniendo resultados muy similares entre ambos, lo que motiva la utilización de productos accesibles y de menor costo computacional. Por un lado, el MERIT DEM, con una resolución espacial de aproximadamente 90 metros (3 arc-seconds), es un modelo de elevación de acceso libre que ha mostrado un rendimiento confiable en la obtención de parámetros morfométricos. Por otro lado, se empleó el DEM generado a partir de los datos del satélite ALOS PALSAR (Advanced Land Observing Satellite Phased Array type L-band Synthetic Aperture Radar), disponible a través del proyecto ALOS PALSAR RTC (Radiometrically Terrain Corrected) de la Alaskan Satellite Facility (ASF), que ofrece una resolución espacial de 12.5 metros y permite la captura de datos incluso en condiciones de nubosidad o poca luz, gracias al uso de radar de apertura sintética. Aunque el DEM de ALOS PALSAR proporciona un mayor nivel de detalle debido a su alta resolución, los resultados obtenidos con el MERIT DEM fueron suficientemente precisos, lo que lo convierte en una opción atractiva para estudios morfométricos y fisiográficos en grandes áreas, dado su equilibrio entre accesibilidad, costo computacional y precisión, además la precisión vertical del MERIT DEM se ha mejorado significativamente mediante la eliminación de errores sistemáticos, lo que le da precisión con esa resolución (90 m).
- En el contexto de la caracterización del coeficiente de escurrimiento (C) y el número de curva de escurrimiento (CN), se logró una cobertura completa (100%) para municipios clave del Departamento del Cauca, como Sucre, Rosas, Almaguer, La Sierra, La Vega, Argelia, San Sebastián, Toribío, Sotará, Jambaló, Inzá, Puracé y El Tambo. Otros municipios, como Timbío, Páez, Totoró, Silvia, Buenos Aires y El Patía, fueron caracterizados parcialmente, con una cobertura entre el 60% y 80%. Sin embargo, algunos municipios como Popayán, Balboa, López de Micay, Santander de Quilichao, Santa Rosa, Caloto, Suárez y Guapi, presentaron una caracterización pobre. Finalmente, municipios como Florencia, Guachené, Mercaderes, Miranda, Padilla, Puerto Tejada, Villa Rica, Piamonte, Caldon, Corinto, Cajibío, Morales, Piendamó y Timbiquí no fueron caracterizados. Se generó una base de datos de esta información que es fundamental para la gestión del recurso hídrico en estos municipios, especialmente en relación con eventos extremos. Por ejemplo, en estudios de evaluación de amenazas por inundación y avenidas torrenciales, el conocimiento de los coeficientes de escurrimiento de las cuencas es necesario, ya que constituye un insumo principal en diversas metodologías técnicas empleadas en la planificación y mitigación de riesgos.

- Se procesó la información de 86 estaciones con medición de precipitación máxima en 24 horas y 39 estaciones con medición de caudal máximo diario (incluida la estación Gorgona La Playa). Se realizó un análisis exhaustivo de los datos, evaluando la presencia de valores atípicos o información faltante, y se encontró que, en general, las series de tiempo son robustas y presentan una buena consistencia en los datos. En algunos casos específicos, se identificaron dificultades en la toma de datos o periodos prolongados de inactividad en las estaciones, lo cual se tuvo en cuenta para mitigar estas condiciones en la presente investigación. También se detectaron algunos datos duplicados, particularmente en estaciones ubicadas en la región del Pacífico. Dado que algunas de estas estaciones están alejadas de las comunidades, y que en ciertos casos los operadores pueden transmitir la información según su propio criterio, se recomienda migrar a estaciones con telemetría que puedan reducir el error humano en la recolección de datos.
- Se desarrollaron modelos de Regresión Lineal Múltiple (RLM) para estimar caudales máximos en el Departamento del Cauca, considerando diferentes periodos de retorno y utilizando parámetros morfométricos y variables hidrológicas, principalmente la precipitación. Se abordaron dos escenarios. En el primer escenario, aplicado a 38 cuencas, se lograron establecer relaciones para los periodos de retorno más grandes, es decir, 200 y 500 años, con métricas de rendimiento aceptables, aunque se identificó la necesidad de ajustar el modelo. Posteriormente, al identificar las cuencas influyentes mediante la distancia de Cook, se conformó un segundo escenario con 31 cuencas. En este caso, se obtuvieron relaciones para todos los periodos de retorno (2.33, 25, 50, 100, 200 y 500 años), alcanzando métricas de rendimiento que reflejan un buen ajuste.

Escenario 1, 15378.27 km² (53% del Cauca):

$$Q_{200} = -30.53 + 13.19P_{cuen} + 83.5T_c - 292.6I_c - 149.88I_a - 441.08R_{ee} + 6.5Pmáx_{24} + 1.81P_{200}$$

$$Q_{500} = -383.58 + 16.65P_{cuen} + 88.38T_c - 481.89I_c - 1.08CN + 8.44Pmáx_{24} + 1.03P_{500}$$

Escenario 2, 8.136,5 Km² (28% del Cauca):

$$Q_{2.33} = -55.2597 + 0.3364A - 7.1408I_{as} + 1.1059Pmáx_{24}$$

$$Q_{25} = 22.3445 + 0.5733A - 93.5653I_c + 2.2402Pmáx_{24}$$

$$Q_{50} = 14.571 + 0.6277A - 107.3706I_c + 2.8583Pmáx_{24}$$

$$Q_{100} = 3.0872 + 0.679A - 121.5349I_c + 3.5812Pmáx_{24}$$

$$Q_{200} = -255.0102 + 0.7036A + 4.8163Pmáx_{24}$$

$$Q_{500} = -319.2288 + 0.7601A + 6.2004Pmáx_{24}$$

- Se analizaron los errores obtenidos mediante validación cruzada, lo cual sirve para proporcionar una medida robusta de cómo los modelos generalizan a datos no vistos y ayudar a identificar problemas como el sobreajuste, la necesidad de ajuste de parámetros y la comparación entre diferentes modelos. Comparando el error promedio de los modelos, se tiene para el escenario 1, para los Tr de 200 y 500 años, errores de -12.15, -17.47, mientras que los errores para los Tr del escenario 2 son, para 2.33 años, 2.18, para 25 años, 0.33, para 50 años, 0.18, para 100 años -0.03, para 200 años, 0.6, y para 500 años, -0.01, valores que reflejan un mejor ajuste en el escenario 2. En todo caso se tienen errores significativos en algunas cuencas, podría ser útil realizar un análisis más detallado para identificar patrones o características en esas cuencas que podrían estar causando estos errores, como variabilidad en la topografía, características del suelo, o eventos extremos no capturados adecuadamente.
- Se evaluaron los mejores modelos para los diferentes periodos de retorno. Para el escenario 1, los modelos muestran un buen desempeño en ambos periodos de retorno, con valores altos tanto de R^2 como de KGE (0.68-0.69 y 0.83-0.84 respectivamente), esto indica que el modelo es robusto y capaz de capturar la relación entre las variables en diferentes escenarios. Las diferencias en los valores de R^2 y KGE entre los dos periodos de retorno son pequeñas, lo que sugiere que el modelo no tiene una variabilidad significativa en su desempeño a través de los diferentes periodos de retorno analizados. Dado que el modelo tiene un buen ajuste en ambos periodos, se puede considerar que está bien calibrado. Sin embargo, si se busca mejorar el modelo aún más, podría ser útil investigar las características específicas de los datos para los periodos de retorno más largos y ajustar el modelo en consecuencia. Para el escenario 2, se tienen valores incluso más altos R^2 como de KGE que el escenario 1 (0.78-0.87 y 0.85-0.91 respectivamente). Los modelos se ajustan mejor a los periodos de retorno intermedios (25 y 50 años), con métricas R^2 y KGE más altas y para los periodos de retorno más extremos (2.33 y 500 años), aunque el desempeño sigue siendo bueno, es ligeramente inferior en comparación con los periodos intermedios, lo que podría indicar que el modelo es menos preciso en estas condiciones. En términos generales el escenario 2 presenta un mejor ajuste (alrededor del 10%) que el escenario 1.
- A lo largo de este proyecto, se ha logrado implementar de manera efectiva el uso de programación en R y Python para desarrollar un análisis robusto y detallado que sustenta la tesis. A través del uso de estos lenguajes de programación, se han automatizado varios procesos clave, lo que permitió una evaluación exhaustiva de modelos de regresión lineal múltiple para diferentes periodos de retorno, garantizando el cumplimiento de criterios estadísticos rigurosos. Se desarrollaron funciones para cargar y limpiar los datos de diversas hojas de cálculo, lo que facilitó el manejo de grandes volúmenes de información. Se implementaron procedimientos

para ajustar modelos de regresión, evaluar la influencia de las observaciones mediante la distancia de Cook, y eliminar cuencas que podían distorsionar los resultados. Se aseguraron los modelos ajustados cumplieran con un conjunto de criterios esenciales, tales como la multicolinealidad, homocedasticidad, normalidad de los residuos e independencia de los errores. Se evaluaron todas las posibles combinaciones de variables con 38 y 31 cuencas, seleccionando aquellas que mejor cumplen con los criterios establecidos, optimizando el modelo según el criterio de información de Akaike (AIC) y se utilizaron herramientas de visualización para identificar los mejores modelos, facilitando la interpretación y comparación de estos. Este enfoque programático no solo ha permitido realizar un análisis riguroso y detallado, sino que también ha proporcionado un marco reproducible y escalable para futuros estudios en el ámbito de la hidrología. La automatización y la precisión alcanzadas mediante R y Python fueron fundamentales para el éxito de esta investigación.

- Los análisis y resultados presentados en este documento confirman la hipótesis de investigación, demostrando una correlación significativa entre variables fisiográficas, parámetros morfométricos, y variables hidrológicas como precipitaciones y caudales máximos asociados a diferentes periodos de retorno. Además, se han establecido ecuaciones confiables para la estimación de caudales máximos en cuencas hidrográficas no instrumentadas del Departamento del Cauca, Colombia, respaldadas por métricas de R^2 y KGE.

7.1 Trabajos futuros

Se considera de vital importancia la matriz de datos donde se reúnen las variables fisiográficas, parámetros morfométricos y variables hidrológicas para las 38 cuencas del presente estudio, por lo que se recomienda aprovecharla continuando estudios de la hidrología del Departamento, mediante regresiones no lineales o regresión robusta, las cuales tienen un mejor ajuste ante datos influyentes.

También se sugiere realizar modelos de optimización (algoritmos genéticos, recocido simulado o algoritmos de selección de subconjuntos) que generen combinaciones para las 38 cuencas ($2^{38} - 1 = 274.877'906.943$), ya que evaluar todas las posibles combinaciones es muy costoso computacionalmente y requeriría demasiados recursos computacionales (Cheng, Lien, Wu, & Su, 2017; Hsu, Gupta, & Sorooshian, 1995; Moriasi, y otros, 2007).

Se recomienda también la aplicación de técnicas de clúster y regionalización para agrupar las cuencas en subregiones homogéneas, lo que permitiría desarrollar modelos de Regresión Lineal Múltiple (RLM) específicos para cada grupo. Esta estrategia podría mejorar la precisión de los modelos al considerar las características particulares de cada subregión, reduciendo la heterogeneidad y mejorando el ajuste local de los modelos. La regionalización de los datos es esencial para capturar las variabilidades espaciales en las respuestas hidrológicas, lo que podría conducir a una mejor gestión de los recursos hídricos y a una mayor precisión en la predicción de caudales máximos en el Departamento del Cauca.

Se sugiere explorar la posibilidad calibrar los coeficientes de escurrentía C y CN para el Departamento del Cauca, teniendo datos de eventos de precipitación y valores de caudales máximos (Ajmal, Waseem, Jehanzaib, & Kim, 2023) se podría obtener valores calculados mediante modelización hidrológica. Esta calibración adicional permitirá ajustar los coeficientes y mejorará su capacidad predictiva, lo que contribuirá significativamente a la comprensión y gestión de los recursos hídricos en el Departamento del Cauca.

Considerando la disponibilidad de datos de caudales en cuencas más pequeñas que están contenidas dentro de cuencas más grandes, se propone desarrollar un proyecto enfocado en la optimización de técnicas para la transposición de caudales máximos. El objetivo es mejorar la precisión en la estimación de estos valores en diversas cuencas hidrográficas.

7.2 Publicaciones y productos derivadas de esta tesis

Se presentarán los resultados de la investigación como ponencia oral en el XXXI Congreso Latinoamericano de Hidráulica de la IAHR (The International Association for Hydro-Environment Engineering and Research), entre los días 1- 4 de octubre del 2024. El artículo quedará consignado en las memorias del evento, y se tendrá la posibilidad de ser seleccionado para ser parte de una edición especial de las revistas: DYNA, Ribagua o Ingenierías Universidad de Medellín, según el proceso de evaluación.

Se redactó un manuscrito en donde se resume de manera técnica todos los procedimientos paso a paso necesarios para una correcta caracterización de cualquier cuenca hidrográfica, llamado “SIG para aplicaciones hidrológicas - Guía de ArcMap: Procedimientos para el cálculo de variables fisiográficas y parámetros morfométricos de cuencas hidrográficas”, y que se espera poder someter a revisión en la Convocatoria del Área de Desarrollo Editorial para la edición de libros de texto 2025-2026 de la Universidad del Cauca, en junio de 2025 (se requiere contar con el certificado de investigación culminada para poder aplicar).

En el transcurso de la presente investigación se lograron realizar los siguientes proyectos:

- “Actualización de curvas IDF del Departamento del Cauca en una herramienta de visualización web”, el cual tuvo como finalidad, construir las 86 curvas IDF de las estaciones meteorológicas activas del Departamento del Cauca administradas por el IDEAM, y que sean potencialmente utilizables en proyectos de Ingeniería Hidráulica e Hidrológica. Se ofertará la información en el portal de Arcgis Online con el nombre: "Curvas IDF Departamento del Cauca", para que cualquier pueda consultar en línea la información de manera clara y obtener de manera práctica la curva IDF de la estación de interés en el Departamento del Cauca (<https://dicam.maps.arcgis.com/apps/instant/atlas/index.html?appid=799bdb386fe34d24bcc399069fb707aa>).

Se presentará el proyecto como ponencia mediante póster en el XXXI Congreso Latinoamericano de Hidráulica de la IAHR entre los días 1- 4 de octubre del 2024.

- “Estudio y caracterización del C y CN para el Departamento del Cauca: Correlación e implementación”, el cual tuvo como finalidad, desarrollar un modelo matemático de correlación entre coeficientes de escorrentía y números de curva de escorrentía, y unificar criterios de implementación que se representaron mediante cartografía para su cálculo en el Departamento del Cauca.

El proyecto fue aceptado como ponencia oral en el XXV Seminario Nacional de Hidráulica e Hidrología de la Sociedad Colombiana de Ingenieros en la ciudad de Bogotá, entre los días 5 - 7 de junio de 2024. El artículo quedó consignado en las memorias del evento (ISSN: 2619-3787 (En línea)) y se podrá descargar en el siguiente enlace: <https://sci.org.co/memorias-xxv-seminario-nacional-de-hidraulica-e-hidrologia/>, entre las páginas 22 y 38 (Agredo-Campuzano, Jiménez, Bolaños-Rodríguez, & Paredes-Guevara, 2024).

El proyecto también fue presentado como ponencia oral en Democratia-Aqua-Technica Conference de la SRH Heidelberg el día 26 de septiembre del 2024.

Se agradece a los estudiantes de pregrado pertenecientes al Semillero de Investigación HidroInnovación (S.H.I.) y al Grupo de Investigación en Hidráulica e Hidrología (G.I.H.H.) de la Universidad del Cauca que apoyaron en el desarrollo de los tres proyectos derivados de esta tesis.

8. Referencias

- Agredo-Campuzano, O. F., Jiménez, D., Bolaños-Rodríguez, A., & Paredes-Guevara, M. J. (2024). Estudio y caracterización del C y CN para el Departamento del Cauca: Correlación e implementación. (*ponencia*). XXV Seminario Nacional de Hidráulica e Hidrología, Bogotá. Recuperado el Agosto de 2024, de <https://sci.org.co/memorias-xxv-seminario-nacional-de-hidraulica-e-hidrologia/>
- Aguilera Sierra, E. (2010). Análisis regional de crecidas máximas para Honduras. (*Trabajo de grado*). Universidad San Carlos de Guatemala, Facultad de Ingeniería, Escuela Regional de Ingeniería Sanitaria y Recursos Hidráulicos a nivel de postgrado ERIS, Ciudad de Guatemala. Recuperado el 29 de Agosto de 2023, de http://biblioteca.usac.edu.gt/tesis/08/08_0396_MT.pdf
- Aguirre, M., Torres, H., & Ruiz, R. (2003). *Ministerio del Ambiente del Perú, Manual de procedimientos para la delimitación y codificación de cuencas hidrográficas del Perú*. Obtenido de <https://siar.minam.gob.pe/puno/sites/default/files/archivos/public/docs/556.pdf>
- Ajmal, M., Waseem, M., Jehanzaib, M., & Kim, T.-W. (2023). Development and testing of updated curve number models for efficient runoff estimation in steep-slope watersheds. *Journal of Hydrology*, 617(B). doi:<https://doi.org/10.1016/j.jhydrol.2022.129049>
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723. doi:10.1109/TAC.1974.1100705
- Alvarez Sevilla, I., Brown Manrique, O., & del Cid Colindres, E. P. (2017). Estimación de variables hidrológicas mediante modelos de regresión en la cuenca del río Suchiate, Guatemala. *Revista Ciencias Técnicas Agropecuarias*, 26(3), 22-29. Obtenido de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2071-00542017000300003&lng=es&tlng=es
- Aparicio, F. (1997). *Fundamentos de Hidrología de Superficie*. Limusa, 303 pp.
- Araya, D., Mendoza, P. A., Muñoz-Castro, E., & McPhee, J. (2023). Towards robust seasonal streamflow forecasts in mountainous catchments: impact of calibration metric selection in hydrological modeling. *Hydrology and Earth System Sciences*, 27(24), 4385–4408. doi:10.5194/hess-27-4385-2023
- Arumi R., J. L., Jara R., J., & Salgado S., L. (2000). *Análisis hidrológico*. Recuperado el 14 de Septiembre de 2023, de Instituto de Investigaciones Agropecuarias, INIA: <https://hdl.handle.net/20.500.14001/39933>
- Aviles, A., Solera, A., & Paredes, J. (2016). Análisis del rendimiento de sistemas hídricos en desarrollo mediante el acople de modelos estocásticos hidrológicos y optimización de redes de flujo. *Ingenius*, 15, 48-57. doi:<https://doi.org/10.17163/ings.n15.2016.05>
- Becerra Oviedo, J. A., Sánchez Mazorca, L. F., Acosta Castellano, P. M., & Díaz Arévalo, J. L. (2016). Regionalización de curvas IDF para el uso de modelos hidrometeorológicos en la Sabana

- Occidental del departamento de Cundinamarca. *Ingeniería Y Región*(14 (2)), 143–150. doi:<https://doi.org/10.25054/22161325.701>
- Beckman, R. J., & Cook, R. D. (1983). Outlier s. *Technometrics*, 25(2), 119-149. doi:<https://doi.org/10.1080/00401706.1983.10487840>
- Bedient, P. B., Huber, W. C., & Vieux, B. E. (2019). *Hydrology and Floodplain Analysis* (6th ed.). New York: Pearson Education. Obtenido de <https://www.pearson.com/en-us/subject-catalog/p/hydrology-and-floodplain-analysis/P200000003180/9780137561391?tab=title-overview>
- Bobee, B., Mathier, L., Perron, H., Trudel, P., Rasmussen, P. F., Cavadias, G., . . . Birikundavyi, S. (1996). Presentation and review of some methods for regional flood frequency-analysis. *Journal of Hydrology*, 186(1-4), 63-84. doi:<https://doi.org/10.1016/s0022-1694%2896%2903042-9>
- Boron, V., Payán, E., MacMillan, D., & Tzanopoulos, J. (2016). Achieving sustainable development in rural areas in Colombia: Future scenarios for biodiversity conservation under land use change. *Land Use Policy*, 59, 27–37. doi:<https://doi.org/10.1016/J.LANDUS>
- Boughton, W. C. (1989). A review of the USDA SCS curve number method. *Soil Research*, 27, 511-523. doi:<https://doi.org/10.1071/SR9890511>
- Braz, A. M., Garcia, P. H., Pinto, A. L., Chávez, E. S., & Oliveira, I. J. (2020). Manejo integrado de cuencas hidrográficas: posibilidades y avances en los análisis de uso y cobertura de la tierra. *Cuadernos de Geografía: Revista Colombiana de Geografía*, 29(1), 69–85. doi:<https://doi.org/10.15446/rcdg.v29n1.76232>
- Burrough, P. A., McDonnell, R. A., & Lloyd, C. D. (1998). *Principles of Geographical Information Systems* (3rd ed.). Oxford University Press.
- Calvache Rengifo, P. J. (2019). Determinación de las curvas regionales para caudales máximos con diferentes periodos de retorno, teniendo en cuenta la variabilidad climática, en la región andina del Departamento del Cauca. (*Trabajo de grado*). Universidad del Cauca. Facultad de Ingeniería Civil. Programa de Ingeniería Ambiental, Popayán, Colombia.
- Camino, M., Bó, M., Cionchi, J., López de Armentia, A., Del Río, J., & De Marco, S. (2018). Estudio morfométrico de las cuencas de drenaje de la vertiente sur del sudeste de la provincia de Buenos Aires (Argentina). *Revista Universitaria de Geografía*, 27(1), 73-97. Obtenido de <http://bibliotecadigital.uns.edu.ar/pdf/reuge/v27n1/v27n1a05.pdf>
- Campos-Aranda, D. F. (2016). Una aplicación hidrológica de la regresión lineal múltiple ponderada. *Tecnología y Ciencias del Agua*, VII(4), 161-173. Recuperado el 28 de Agosto de 2023, de <https://www.redalyc.org/articulo.oa?id=353549828011>
- Carrasquilla-Batista, A., Chacón-Rodríguez, A., Núñez-Montero, K., Gómez-Espinoza, O., Valverde, J., & M., G.-B. (2016). Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal. *Tecnología en Marcha. Encuentro de Investigación y Extensión*, 33-45. doi:10.18845/tm.v29i8.2983

- Carvajal, L. F., Salazar, J. E., Mesa, Ó. J., & Poveda, G. (2015). Predicción Hidrológica en Colombia mediante Análisis Espectral Singular y Máxima Entropía. *Tecnología Y Ciencias Del Agua*, XIII(1), 7-16. Recuperado el 28 de Agosto de 2023, de <http://revistatyca.org.mx/ojs/index.php/tyca/article/view/798>
- Castro Heredia, L. M., Carvajal Escobar, Y., & Ávila Díaz, Á. J. (2012). ANÁLISIS CLÚSTER COMO TÉCNICA DE ANÁLISIS EXPLORATORIO DE REGISTROS MÚLTIPLES EN DATOS METEOROLÓGICOS. *Ingeniería de Recursos Naturales y del Ambiente*, 11, 11-20. Obtenido de <https://www.redalyc.org/articulo.oa?id=231125817001>
- Castro, C. V., & Maidment, D. R. (2020). GIS preprocessing for rapid initialization of HEC-HMS hydrological basin models using web-based data services. *Environmental Modelling & Software*, 130, 104732. doi:<https://doi.org/10.1016/J.ENVSOFT.2020.104732>.
- Cheng, K.-S., Lien, Y.-T., Wu, Y.-C., & Su, Y.-F. (2017). On the criteria of model performance evaluation for real-time flood forecasting. *Stochastic Environmental Research and Risk Assessment*, 31(5), 1123–1146. doi:<https://doi.org/10.1007/s00477-016-1322-7>
- Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). *Applied Hydrology*. New York, U.S.A.: McGraw-Hill International Editions, Civil Engineering Series.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15-18. doi:10.1080/00401706.1977.10489493
- Dai, A., & Trenberth, K. E. (2004). The diurnal cycle and its depiction in the Community Climate System Model. *Journal of Climate*, 17(7), 930-951.
- Donado, L. D., Romero-Duque, S. A., & Cortes-Ramos, D. A. (2021). Regionalización de Caudales a partir de la Caracterización Morfoclimática de Cuencas. Caso de Estudio: Valle Medio de Magdalena - Colombia. (ponencia). XXIV Seminario Nacional de Hidráulica e Hidrología, Barranquilla. Recuperado el Marzo de 2024, de <https://sci.org.co/memorias-xxiv-seminario-nacional-de-hidraulica-e-hidrologia/>
- Espinosa Martínez, C. D. (2020). *Correlación entre geología y parámetros morfométricos de cuencas sedimentarias en Colombia*. Obtenido de https://ciencia.lasalle.edu.co/ing_civil/883
- European Space Agency (ESA). (2023). *Copernicus Open Access Hub*. Obtenido de <https://scihub.copernicus.eu/>
- Fermini, L. A. (2019). *Cómo elaborar una tesis de grado*. Bogotá: Alfaomega.
- Fernández, F. (1995). *Manual de Climatología Aplicada. Clima, medio ambiente y planificación*. Madrid: Síntesis S. A.
- Ferrer, F. (1993). *Recomendaciones para el Cálculos Hidrometeorológicos de Avenidas*. Obtenido de CEDEX, Ministerio de Obras Públicas: <https://hispagua.cedex.es/node/92786>
- Freyberg, D. L., Reeder, J. W., Franzini, J. B., & Remson, I. (1980). Application of the Green-Ampt Model to infiltration under time-dependent surface water depths. *Water Resources Research*, 16(3), 517–528. doi:<https://doi.org/10.1029/WR016i003p00517>

- GitHub, Inc. (2024). *GitHub Documentation*. Obtenido de <https://docs.github.com>
- Goigochea Pinchi, J. (2016). Delimitación y codificación de unidades hidrográficas en el ámbito del Alto Huallaga. (*Trabajo de grado*). Universidad Nacional Agraria de la Selva, Facultad de Recursos Naturales Renovables, Programa de Ingeniero en Recursos Naturales Renovables con mención en Forestales, Tingo María. Recuperado el 28 de Agosto de 2023, de <https://repositorio.unas.edu.pe/handle/20.500.14292/549?show=full>
- Gualdrón, D., Villate, J., Torres, D., Molina, J., & Reyes, C. (2022). *Determinación de los parámetros morfométricos en cuencas empleando Sistemas de Información Geográfica* (1 ed.). Tunja: Editorial UPTC 10.19053/9789586606769.
- Guelfi, M., & López-Vazquez, C. (2018). Comparación del método de Thiessen con alternativas más simples mediante simulación de Monte Carlo. *Revista Cartográfica*, 96, 125–138. doi:<https://doi.org/10.35424/rcarto.i96.191>
- Guevara Álvarez, M. (2016). *Socavación en puentes*. (U. d. Cauca, Ed.) Bogotá: Lemoine Editores.
- Guevara Álvarez, M. (2021). *Socavación en puentes - Evaluación, instrumentación y cálculo*. Bogotá: Universidad del Cauca.
- Guisande, C., Vaamonde, A., & Barreiro, A. (2011). *Tratamiento de Datos con R, STATISTICA y SPSS*. España: Díaz de Santos.
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2), 80-91. doi:10.1016/j.jhydrol.2009.08.003
- Haan, C. T. (1977). *Statistical Methods in Hydrology* (1 ed.). Iowa: Iowa State University Press.
- Harvey, A. C. (1990). *The Econometric Analysis of Time Series* (2nd ed.). MIT Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- Hawkins, R., Theurer, F., & Rezaeianzadeh, M. (2019). Understanding the basis of the curve number method for watershed models and TMDLs. *Journal of Hydrologic Engineering*, 24(7) . doi:[https://doi.org/10.1061/\(ASCE\)HE.1943-5584.000175](https://doi.org/10.1061/(ASCE)HE.1943-5584.000175)
- Heberger, M. (2023). *delineator.py: Fast, accurate watershed delineation using hybrid vector- and raster-based methods and data from MERIT-Hydro*, 1.3. (Zenodo) doi:10.5281/zenodo.10143149
- Horton, R. E. (1945). Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin*, 56(3), 275-370. doi:10.1130/0016-7606(1945)56[275:EDOSAT]2.0.CO;2
- Hsu, K.-l., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network. *Water resources*, 31(10), 2517-2530. doi:<https://doi.org/10.1029/95WR01955>

- Hu, S. (1987). Problems with outlier test methods in flood frequency analysis. *Journal of Hydrology*, 96(1-4), Pages 375-383. doi:[https://doi.org/10.1016/0022-1694\(87\)90167-3](https://doi.org/10.1016/0022-1694(87)90167-3)
- IAHR - The International Association for Hydro-Environment Engineering and Research. (1 de Septiembre de 2024). *IAHR Online Journals Platform*. Obtenido de <https://iahr.tandfonline.com/>
- IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales. (2012). *Mapas Cobertura Tierra*. Obtenido de Cobertura de la tierra periodo 2010 - 2012: <http://www.siac.gov.co/catalogo-de-mapas>
- IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales. (2023). *Consulta y Descarga de Datos Hidrometeorológicos*. Obtenido de <http://dhime.ideam.gov.co/atencionciudadano/>
- IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales. (2023). *Estudio Nacional del Agua 2022*. Obtenido de https://www.andi.com.co/Uploads/ENA%202022_compressed.pdf
- IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales. (28 de Agosto de 2023). *Tiempo y Clima. Atlas de Colombia*. Obtenido de <http://www.ideam.gov.co/web/tiempo-y-clima/atlas-de-colombia>
- IGAC - Instituto Geográfico Agustín Codazzi. (2017). *Mapas de Suelos del Territorio Colombiano a escala 1:100.000. Departamento: Cauca*. Obtenido de <https://geoportal.igac.gov.co/contenido/datos-abiertos-agrologia>
- IGAC - Instituto Geográfico Agustín Codazzi. (28 de Agosto de 2023). *Diccionario Geográfico Gobernación del Departamento del Cauca*. Obtenido de <https://diccionario.igac.gov.co/>
- INVIAS, I. N. (2009). *Manual de drenaje para carreteras*. Obtenido de <https://www.invias.gov.co/index.php/archivo-y-documentos/documentos-tecnicos/especificaciones-tecnicas/984-manual-de-drenaje-para-carreteras>
- Jensen, J. R. (2009). *Remote Sensing of the Environment: An Earth Resource Perspective* (2nd ed.). Pearson Education.
- Johnson, F., Hutchinson, M., The, C., Beesley, C., & Green, J. (2016). Topographic relationships for design rainfalls over Australia. *Journal of Hydrology*, 533, 439-451. doi:<https://doi.org/10.1016/j.jhydrol.2015.12.035>
- Kundzewicz, Z. W., & Robson, A. J. (2004). Change detection in hydrological records—a review of the methodology. *Hydrological Sciences Journal*, 49(1), 7-19. doi:10.1623/hysj.49.1.7.53993
- Kutner, M. H., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models* (5th ed.). McGraw-Hill. (5th ed.). New York: McGraw-Hill.
- La Sociedad Geográfica de Colombia – Academia de Ciencias Geográficas. (1 de Agosto de 2015). *Mapa modificado a partir del Mapa División Política del Cauca*. Obtenido de https://www.sogeocol.edu.co/web/index_original.htm

- Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "goodnessof-. *Water Resources Research*, 35(1), 233–241. doi:<https://doi.org/10.1029/1998WR900018>
- Lin, P., Pan, M., H. E., B., Yang, Y., Yamazaki, D., Frasson, R., . . . Wood, E. F. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resources Research*, 55, 6499–6516. doi:<https://doi.org/10.1029/2019WR025287>
- Linsley, R. M. (1988). *Hidrología para Ingenieros* (Segunda ed.). D. F. México: Mc Graw-Hill.
- López Ramos, A. A., Martínez Acosta, L. E., & Feria Díaz, J. J. (2014). Parámetros hidrológicos y tensores en la dinámica fluvial: estudio en cuencas del caribe colombiano. *Ingeniería y Desarrollo*, 32(2), 218-241. Recuperado el 28 de Agosto de 2023, de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0122-34612014000200004&lng=en&tlng=es
- Maidment, D. (1993). *Handbook of Hydrology*. McGraw Hill.
- Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46, 253: 68-68. doi:<https://doi.org/10.2307/2280095>
- Matovelle, C., & D., H. (2020). Análisis comparativo de las características morfométricas de sistemas hidrográficos de la vertiente del Pacífico, Ecuador. *Investigación y Ciencia*, 28(80), 22-31. Obtenido de <https://www.redalyc.org/journal/674/67464474003/html/>
- Méndez, W., Córdova, J., & Bravo de Guenni, L. (2015). Modelos predictivos de caudales máximos instantáneos para cuencas de drenaje de ambientes montañosos, sustentados en parámetros morfométricos. *Revista Técnica de la Facultad de Ingeniería Universidad del Zulia*, 38(3), 229-238. Recuperado el 28 de Agosto de 2023, de http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0254-07702015000300006&lng=es&tlng=es
- Monsalve S., G. (1999). *Hidrología en la Ingeniería*. Bogotá D.C., Colombia: Editorial Escuela Colombiana de Ingeniería, Segunda Edición.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis* (5th ed.). Wiley.
- Montgomery, D. R., & Dietrich, W. E. (1992). Channel initiation and the problem of landscape scale. *Science*, 255(5046), 826-830. doi:10.1126/science.255.5046.826
- Morassutti F., G. (2020). *Diseño de estructuras de corrección de torrentes y retención de sedimentos* (Primera ed.). Bogotá, Colombia: Ediciones de la U.
- Moriasi, D. N., Arnold, J. G., Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model Evaluation Guidelines for. *Transactions of the ASABE*, 50(3), 885–900. doi:<https://doi.org/10.13031/2013.23153>
- NASA KIWE; Instituto Colombiano de Geología y Minería INGEOMINAS. (2009). *Informe de los escenarios de Amenaza por flujos de lodo "Avalanchas" en la Cuenca del Río Páez*. Popayán. Recuperado el 14 de Septiembre de 2023, de

<https://miig.sgc.gov.co/Paginas/Resultados.aspx?k=2105245991101000#k=Informe%20de%20los%20escenarios%20de%20Amenaza%20por%20flujos%20de%20lodo%2022Avalanchas%22%20en%20la%20Cuenca%20del%20R%3%ADo%20P%3%A1ez>

- Ouarda, T., Bâ, K., Delgado, C. D., Cârsteanu, A., Chokmani, K., Gingras, H., . . . Bobée, B. (2008). Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study. *Journal of Hydrology*, Volume 348, Issues 1–2, 40-58. doi:<https://doi.org/10.1016/j.jhydrol.2007.09.031>
- Patton, D., Smith, D., Muche, M. E., Wolfe, K., Parmar, R., & Johnston, J. M. (2022). Catchment scale runoff time-series generation and validation using statistical models for the Continental United States. *Environmental Modelling & Software*, 149, 105321. doi:<https://doi.org/10.1016/J.ENVSOF.2022.105321>.
- Pizarro T., R., Ramírez B., C., & Flores V., J. P. (2003). Análisis comparativo de cinco métodos para la estimación de precipitaciones areales anuales en períodos extremos. *Bosque (Valdivia)*, 24(3), 31-38. doi:<https://dx.doi.org/10.4067/S0717-92002003000300003>
- Ponce, V. M. (1989). *Engineering Hydrology, Principles and Practices*. Englewood Cliffs, New Jersey, USA: Prentice Hall.
- Potenciano, A., & Garzón, G. (2005). Parámetros geomorfológicos y regresión múltiple en el comportamiento hidrológico de caudales en las cuencas altas de los ríos Tajo y Guadiana. *Geogaceta*, 38, 259-262. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=1333493>
- Poveda, G., Mesa, O., Carvajal, L., Hoyos, C., Mejía, J., Cuartas, L., & Pulgarín, A. (2002). Predicción de caudales medios mensuales en ríos colombianos usando métodos no lineales. *Meteorología Colombiana*, VI, 101-110. doi:https://www.scielo.org.mx/scielo.php?script=sci_nlinks&ref=8948338&pid=S2007-2422201000040000300009&lng=es
- R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Obtenido de <https://www.R-project.org>
- Reyes Trujillo, A., Barroso, F., & Carvajal Escobar, Y. (2010). *Guía básica para la caracterización morfométrica de cuencas hidrográficas* (Primera ed.). Cali: Universidad del Valle. Obtenido de <https://programaeditorial.univalle.edu.co/gpd-guia-basica-para-la-caracterizacion-morfometrica-de-cuencas-hidrograficas-9789586708555-63324b79e35bd.html>
- Rickenmann, D. (2020). *Methods for the Quantitative Assessment of Channel Processes in Torrents (Steep Streams)* (1st ed.). Leiden, The Netherlands: CRC Press / Balkema.
- Rodda, H. J., & Little, M. A. (2015). *Understanding mathematical and statistical techniques in hydrology : an examples-based approach* (1 ed.). Chichester, West Sussex, UK: Wiley-Blackwell. doi:10.1002/9781119077985
- Rojas Morales, N., Aguilar Pereira, J. F., & Solís Bolaños, H. (2008). Funciones de regresión para caudales extremos en la vertiente Pacífica de Costa Rica. *Ingeniería. Revista de la*

Universidad de Costa Rica, 18((1-2)), 31-44. Obtenido de <https://www.redalyc.org/articulo.oa?id=44170521002>

Rojo Hernández, J. D. (2011). Desarrollo de un sistema experto para la predicción de caudales medios mensuales en Colombia. (*Trabajo de grado*). Universidad Nacional de Colombia. Facultad de Minas – Escuela de Geociencias y Medio Ambiente. Maestría en Ingeniería – Recursos Hidráulicos, Medellín, Colombia. Recuperado el 28 de Agosto de 2023, de <https://repositorio.unal.edu.co/handle/unal/8796>

Rojo Hernández, J. D., & Carvajal Serna, L. F. (2010). Predicción no lineal de caudales utilizando variables macroclimáticas y análisis espectral singular. *Tecnología Y Ciencias Del Agua*, 1(4), 59-73. Obtenido de <http://www.revistatyca.org.mx/ojs/index.php/tyca/article/view/71>

Sánchez San Román, F. J. (2022). *Hidrología Superficial y Subterránea* (Segunda ed.). Salamanca, España: Kindle Direct Publishing.

Smith Quintero, R. A., Vélez Upegui, J. I., Velásquez Henao, J. D., Ceballos López, A., Correa Velásquez, P. L., Góez Arango, C., . . . Zapata Gómez, E. C. (2001). Modelos de predicción de caudales mensuales para el sector Eléctrico Colombiano. *Avances en recursos Hidráulicos*, 11, 91-102. Obtenido de <https://revistas.unal.edu.co/index.php/arh/article/view/93064/77832>

Soria, M., & Matar, M. (2016). *Nociones sobre teledetección*. Obtenido de Universidad Nacional de San Juan. Argentina.: <http://www.unsj.edu.ar/unsjVirtual/cartografiaaplicadaminas/wp-content/uploads/2016/10/Apuntes-de-c%C3%A1tedra-para-Cartograf%C3%ADa-Aplicada.pdf>

Srivastava, A. K., Srivastava, V. K., & Ullah, A. (1995). The coefficient of determination and its adjusted version in linear regression models. *Econometric Reviews*, 14(2), 229–240. doi:<https://doi.org/10.1080/07474939508800317>

Suárez Fiesco, R., & Almeida Rosero, O. F. (2017). Análisis de correlación entre parámetros morfométricos, precipitaciones máximas y caudales máximos asociados a diferentes periodos de retorno en cuencas rurales. (*Trabajo de Grado*). Universidad Católica de Colombia. Facultad de Ingeniería. Programa de Ingeniería Civil. Especialización en Recursos Hídricos, Bogotá, Colombia. Recuperado el 28 de Agosto de 2023, de <http://hdl.handle.net/10983/14625>

Témez, J. R. (1991). Generalización y mejora del método racional. *Ingeniería Civil, Versión de la Dirección General de Carreteras de España*, 82: 51–56.

Thiessen, A., & Alter, J. (1911). Precipitation averages for large areas. *Monthly Weather Review*(39), 1082-1089. doi:[https://doi.org/10.1175/1520-0493\(1911\)39<1082b:PAFLA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1911)39<1082b:PAFLA>2.0.CO;2)

Toro Triana, L. (2013). *Análisis de sensibilidad de parámetros hidrológicos de la cuenca del río Bermellón para el cálculo de caudales de escorrentía*. Obtenido de Universidad de los Andes: <http://hdl.handle.net/1992/61276>

U.S. Geological Survey (USGS). (2023). *Earth Explorer*. Obtenido de <https://earthexplorer.usgs.gov/>

- Universidad del Cauca. (3 de Agosto de 2024). *Acceso a bases de datos*. Obtenido de <https://www.unicauca.edu.co/biblioteca/contenido/servicios/acceso-bases-de-datos>
- University of Alaska Fairbanks. (2023). *Alaska Satellite Facility*. Obtenido de <https://search.asf.alaska.edu/#/>
- Vargas M., R., & Diaz-Granados O., M. (1998). Curvas Sintéticas Regionalizadas de Intensidad-Duración-Frecuencia para Colombia. (*Trabajo de grado*). Universidad de los Andes. Maestría en Ingeniería Civil, Bogotá.
- Velasquez, F. (s.f.). Ecuaciones regionales de regresión múltiple para cálculo de frecuencia de caudales máximos en Honduras. (*Trabajo de grado*). Universidad Nacional Autónoma de Honduras, Tegucigalpa.
- Vélez, M. V., Smith, R. A., Urrea, D. P., Pérez, C. A., & Franco, C. J. (1994). Estudio Hidrológico del Departamento del Quindío. (*ponencia*). XI Seminario Nacional de Hidráulica e Hidrología, Popayán. Recuperado el Marzo de 2024
- Viessman, W., & Lewis, G. L. (2003). *Introduction to Hydrology*. Pearson Education Inc., 5ª ed., 612 pp.
- Viollet, P. L. (2005). *Water Engineering in Ancient Civilizations: 5,000 Years of History*. Paris: IAHR Monographs. Leiden. CRC Press.
- Water Resources Council (Actualmente United States Geological Survey). (1981). *Guidelines for determining flood flow frequency, bulletin 17B*. Reston, Virginia. Obtenido de chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://water.usgs.gov/osw/bulletin17b/dl_flow.pdf
- Wu, C., & Chau, K. (2013). Prediction of rainfall time series using modular soft computing methods. *Engineering applications of artificial intelligence*(26(3)), 997-1007. doi:<http://dx.doi.org/10.1016/j.engappai.2012.05.023>
- Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. (2019). MERIT Hydro: A high-resolution global hydrography map based on latest topography datasets. *Water Resources Research*, 55, 5053–5073. doi:<https://doi.org/10.1029/2019WR024873>

9. Anexos

1. Base de datos de estaciones hidrometeorológicas IDEAM
2. Parámetros morfométricos y tiempos de concentración de 41 cuencas hidrográficas
3. Datos de precipitación máxima en 24 horas suministrados por el IDEAM
4. Datos de caudal suministrados por el IDEAM
5. Análisis de outlier de datos de precipitación
6. Análisis de outlier de datos de caudal
7. Conformación de ponderaciones de precipitación para cada cuenca hidrográfica
8. Distribuciones de probabilidad y pruebas de bondad de ajuste ponderaciones de precipitación
9. Distribuciones de probabilidad y pruebas de bondad de ajuste estaciones de caudal
10. Códigos de programación en Python o R utilizados en la investigación