



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Departamento de Obras Civiles

Predicción del caudal aportante en la laguna Cáhuil a través de un modelo hidrológico sobre la Subcuenca Nilahue

Memoria de Título presentada por

Fernanda Andrea Venegas Godoy

como requisito parcial para optar al título de la carrera de

Ingeniería Civil

Profesor Guía
Álvaro Ossandón
Patricio Catalán

NOVIEMBRE, 2024



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): Memoria o trabajo de título Tesis de Postgrado

Título del trabajo: Predicción del caudal aportante en la laguna Cáhuil a través de un modelo hidrogeológico

Nombre del candidato(a): Fernanda Andrea Venegas Godoy

Carrera / Grado: Ingeniería Civil

Campus: Casa Central

Departamento: Obras Civiles

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, **Álvaro Ossandón**, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente

DEJO CONSTANCIA que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

6 meses 12 meses 2 años 3 años 5 años 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 13-10-2025 **Firma:**

Estudiante o Candidato(a):

Fecha: 13-10-2025 **Firma:**

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

PREDICCIÓN DEL CAUDAL APORTANTE EN LA LAGUNA CÁHUIL A TRAVÉS DE UN MODELO HIDROLÓGICO SOBRE LA SUBCUENCA NILAHUE

Fernanda Venegas¹, Patricio Catalán², Álvaro Ossandón¹

¹ Universidad Técnica Federico Santa María

Resumen

Este trabajo de investigación analiza los datos diarios de precipitaciones y caudales, buscando establecer correlaciones entre ellos para construir modelos estadísticos capaces de predecir caudales futuros a partir de la información histórica disponible. La zona de estudio comprende la Subcuenca Nilahue, que forma parte de la cuenca Costera Rapel-Estero Nilahue. Se utilizan predictores de días anteriores de caudales y precipitaciones para construir modelos lineales múltiples (LM) y Modelos Aditivos Generalizados para la Localización, Escala y Forma (GAMLSS). Estos modelos se validan mediante el "Leave one year out crossvalidation", utilizando métricas de precisión como AIC, BIAS, NSE, R y RMSE. Los resultados sugieren que los modelos pueden optimizarse al dividir la base de datos principal en dos grupos: una temporada influenciada por precipitaciones, denominada invierno, y una temporada seca sin influencia de precipitaciones. Además, la temporada de invierno se subdivide para considerar el comportamiento diferenciado de los caudales medios y los picos de crecida, los cuales están acompañados de curvas de ascenso y descenso. Los valores de R obtenidos muestran un buen rendimiento en la predicción de caudales durante las fases de ascenso, crecida y descenso, con valores de 0.84, 0.75 y 0.9, respectivamente, mientras que para los caudales medios el R es de 0.47. El análisis del comportamiento de las variables hidráulicas sugiere que el uso de herramientas avanzadas como el machine learning y bases de datos horarias podría mejorar significativamente los resultados en futuros modelos.

Palabras claves: Correlación de variables, Predicción de caudal diario, Modelo Lineal Múltiple, Modelos GAMLSS.

1. Introducción

Chile, al igual que el resto de los países, está lidiando actualmente con el cambio climático, el cual ha intensificado la variabilidad espacio-temporal de diferentes variables climáticas como la precipitación y la evapotranspiración, afectando así los regímenes de caudal (Meng et al., 2016). Es por ello, que al presentar cambios se pueden percibir irregularidades en los esteros que desembocan en el mar, las cuales no estaban contempladas para épocas pasadas con regímenes distintos. Dentro de esta premisa se encuentra el caso de la localidad de Cáhuil, localizada a 15 kilómetros al sur de Pichilemu, en la provincia Cardenal Caro. Esta pequeña aldea se sitúa en la desembocadura del estero Nilahue, donde coexisten diferentes actores como: ostricultores, salineros y boteros orientados a servicios recreativos, en donde sus actividades dependen directamente del caudal, el nivel de oxígeno disuelto y la salinidad del estero, los cuales varían en función de la apertura o cierre de la barra en la desembocadura del estero (Proyecto Simona Costa, 2023). Esta barra de arena es movida a través de maquinaria, basándose en el criterio de una inspección visual sin respaldo cuantitativo, lo que

estaría afectando la zona de estudio. Tal como fue declarado por GEF Humedales Costeros (2021) se han estado percibiendo cambios en la dinámica de la barra, se presentan cierres prematuros del estuario que dificultan el desarrollo de actividades que dependen del ingreso de agua de mar al estero. Además, plantea que para evaluar el estado de la Laguna de Cáhuil frente a los riesgos de crecidas, se deben considerar dos variables de estado principales: el nivel de la laguna y la predicción de lluvias y/o caudales que puedan elevar el nivel del agua. A partir de lo anterior, se da a conocer la necesidad de determinar una forma de estimar el caudal aportante a la laguna Cáhuil. El objetivo principal de esta memoria es, a través de un modelo hidrológico, predecir el caudal de la zona para contribuir a la resolución de un problema que ha estado aquejando a la comunidad. Esto a partir de un modelo de base estadística. Para alcanzar lo propuesto, se necesitan datos para construir las ecuaciones de predicción y, de esta forma, tener un sustento cuantitativo para la toma de decisiones sobre el movimiento de la barra de arena. Las predicciones probabilísticas de caudales son herramientas valiosas para la estimación predictiva de la incertidumbre, la gestión del riesgo hidrológico y el apoyo a la toma de decisiones en materia de recursos hídricos (Romero-Cuellar et al., 2024). Decisiones que impactan directamente en la vida de las personas que dependen de las condiciones de la laguna.

A modo de resumen, se ha realizado el procesamiento de tipo estadístico a los datos de precipitaciones y caudales. Sin procesamiento estadístico, las incertidumbres en los pronósticos climáticos y la simulación de procesos hidrológicos a menudo se especifican incorrectamente, incluso en sistemas de predicción climática por conjuntos altamente sofisticados (Bennett et al., 2021). Se han creado modelos estadísticos de tipo: lineal múltiple y GAMLSS, los cuales permiten predecir el caudal a partir de ecuaciones que dependen de variables independientes, sus coeficientes, distribuciones y parámetros. Posteriormente, para el mejor modelo que se ajusta se evalúa su validez a través del método de validación cruzada dejando un año fuera del ajuste del modelo y, luego, usando el modelo para hacer una predicción de este año y compararlo con los datos observados. Todo lo anterior, con el fin de conseguir predicciones que puedan aportar en un futuro a la localidad de Cáhuil.

2. Objetivos

El objetivo de la investigación es implementar un modelo hidrológico eficaz de base física y/o estadística para la Subcuenca del Estero Nilahue que permita pronosticar caudales. Es posible desglosar el objetivo general en objetivos específicos, entre ellos se encuentran los siguientes tres:

Objetivos Específicos:

1. Identificar las variables hidráulicas principales de la subcuenca de acuerdo con el registro de publicaciones y casos prácticos.
2. Establecer el modelo hidrológico idóneo para la Subcuenca del Estero Nilahue que permita pronosticar caudal.
3. Evaluar la validez del modelo hidrológico mediante la determinación del mínimo error entre el pronóstico y el registro histórico de caudales diarios.

3. Zona de estudio

El estudio se basa en la subcuenca Estero Nilahue (Figura 1), la cual tiene un área de 1,778 [km²] y forma parte de la cuenca Costera Rapel-Estero Nilahue. En cuanto a su régimen hidrológico, es una cuenca tipo pluvial, en la cual el escurrimiento es sólo importante en los meses de invierno. Sus mayores eventos de precipitación ocurren entre los meses de Mayo y Septiembre de cada año, como se muestra en la Figura 2. El período estival, en que las temperaturas son más altas, ocurre entre los meses de Noviembre y Abril. El estero Nilahue desemboca en el extremo suroriental de la laguna de Cahuil, a través de la cual llega al océano Pacífico, unos 10 km al sur de Pichilemu. Los recursos hídricos de la cuenca del estero Nilahue pueden ser complementados artificialmente por el embalse Convento Viejo ubicado en la cuenca del estero Chimbarongo (Uribe, 2016).

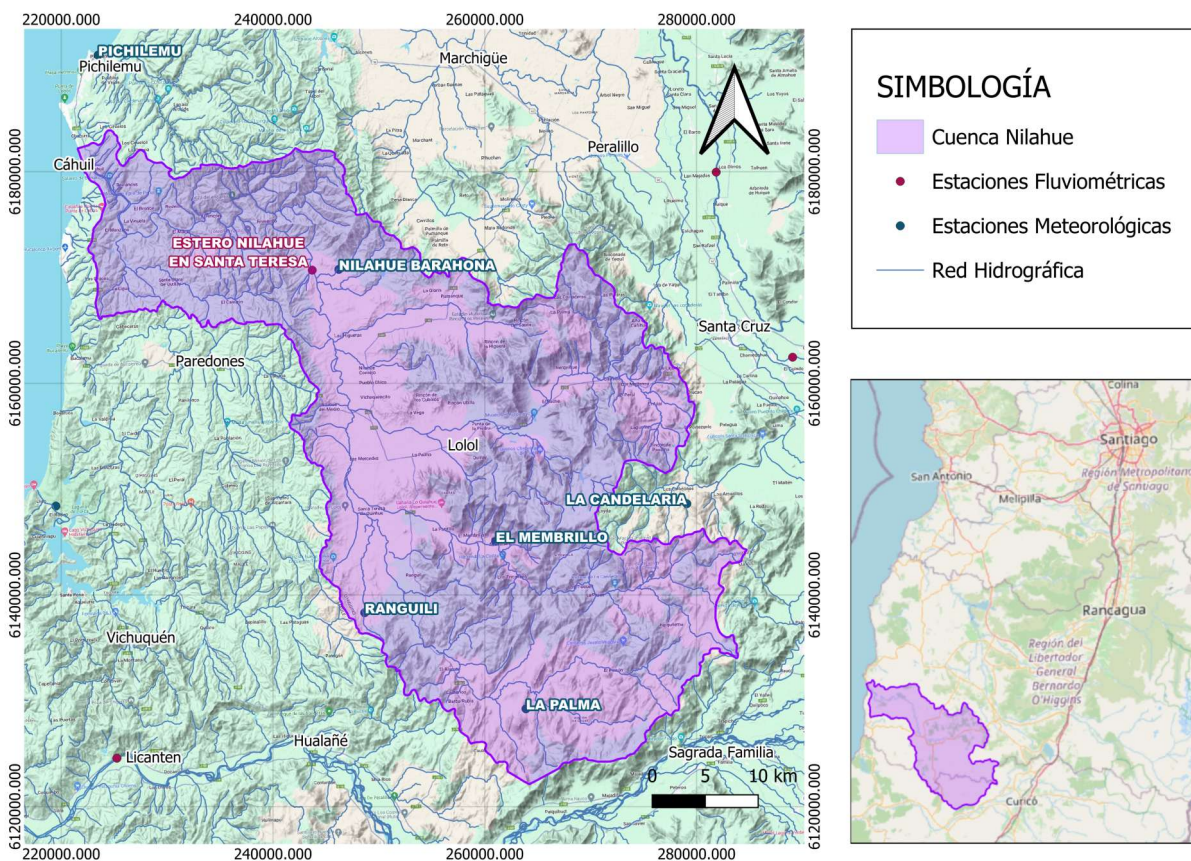


Figura 1: Mapa Subcuenca Estero Nilahue.

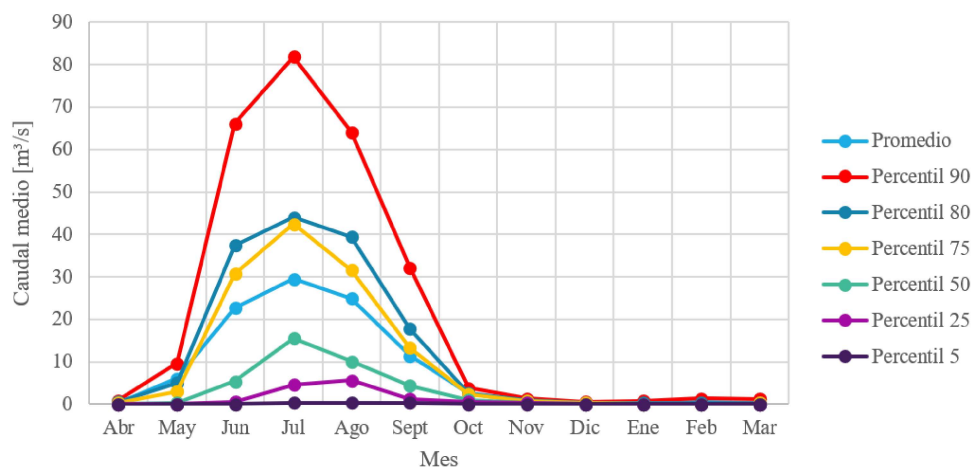


Figura 2: Curva de variación estacional cuenca Estero Nilahue.

Bajo la Figura anterior se pueden definir dos modelos, uno para los meses de "invierno" que considere precipitaciones (desde Mayo hasta Septiembre) y uno para los meses de "temporada seca" que no las considere (desde Octubre hasta Abril).

3.1. Base de Datos

Los datos son obtenidos a partir de la Información Oficial Hidrometeorológica y de Calidad de Aguas en Línea de la Dirección General de Aguas (DGA) del Ministerio de Obras Públicas (MOP). Se obtienen series de tiempo de precipitación diaria para 4 estaciones pluviométricas de la subcuenca Estero Nilahue desde 1985 hasta 2020. Además, se obtienen los caudales medios mensuales y los caudales medios diarios desde este sitio para la estación fluviométrica en un periodo de tiempo desde 1985 hasta 2021. Las estaciones se muestran en la Tabla 1.

Tabla 1: Estaciones fluviométricas y pluviométricas con sus coordenadas (Datum WGS84, Huso 19)

Nombre Estación	Coordenada UTM Norte	Coordenada UTM Este	Tipo
Nilahue en Santa Teresa	6,171,041	243,965	Fluviométrica
Nilahue Barahona	6,170,703	246,328	Pluviométrica
El Membrillo	6,145,369	259,514	Pluviométrica
Ranguili	6,138,358	248,743	Pluviométrica
La Palma	6,129,295	263,897	Pluviométrica

4. Marco teórico

Durante el análisis del tema y la recopilación de la información de la zona, se toman en consideración una serie de métodos y métricas relevantes para comprender el problema, los cuales se exponen a continuación.

4.1. Antecedentes de la Investigación

Para la subcuenca Nilahue, se han realizado en el pasado dos estudios de hidrología impulsados por el Gobierno de Chile. Con ellos, es posible conocer características fundamentales de la cuenca y dar un contexto de la zona de estudio e información preliminar. Según los resultados obtenidos con la construcción de la curva hipsométrica de la cuenca se permite establecer que esta está en equilibrio geomorfológico. Además, se detallan las estaciones meteorológicas utilizadas en el estudio y como estas se correlacionan entre sí, lo que da incapié a seleccionar cuales serán consideradas para esta investigación (Gobierno Regional Libertador Bernardo O´Higgins, 2023).

4.2. Análisis de Componentes Principales

Es una técnica de reducción de datos que crea componentes principales (PC), las cuales son combinaciones lineales de las variables originales. De esta forma, se obtienen variables nuevas no correlacionadas (Hayati et al., 2024). Esto permite trabajar una gran cantidad de datos al reducir su dimensionalidad y preservar su comportamiento. Este análisis se realiza por la descomposición de la matriz de covarianza de los datos originales, la cual provee vectores propios, que son a su vez las llamadas componentes principales. Además, provee los valores propios asociados a estos vectores. La dirección de máxima variación de los datos se representa por los vectores propios, mientras que los valores propios indican la cantidad de varianza que explica cada componente principal.

Una forma alternativa de resolver el problema es mediante la descomposición de valores singulares (SVD). Si $S_{p \times p}$ es la matriz de covarianza de $X_{n \times p}$ (el conjunto de p variables estandarizadas con n observaciones), su descomposición en valores singulares es:

$$SVD(S) = E\Lambda E^T$$

Donde $E_{p \times m}$ y $\Lambda_{m \times m}$ son las matrices de vectores y valores propios.

4.3. Coeficiente de Correlación de Pearson (R)

El coeficiente de correlación de Pearson es una medida estadística que describe la relación lineal entre dos variables continuas. Se denota como "R" y puede variar en valor desde -1 hasta 1. Una correlación positiva perfecta viene dada por un valor de R igual a 1, lo que significa que las dos variables analizadas aumentan juntas en una relación lineal. Cuando R es -1, se indica una correlación negativa perfecta, lo que significa que una variable aumenta mientras que la otra disminuye en una relación lineal. Mientras más alejado se encuentre de estos valores, menor es la correlación lineal entre las dos variables.

4.4. Coeficiente de Correlación de Spearman

El coeficiente de correlación de Spearman es una medida estadística no paramétrica que proporciona la relación entre dos variables. Su principal diferencia con el coeficiente de correlación de Pearson radica en que solo evalúa si el aumento de una variable se asocia con un aumento o disminución de la otra, sin identificar una relación lineal. El coeficiente de Spearman se calcula a partir de los rangos de los datos en lugar de los valores brutos. Se denota como " ρ ", si este valor es 1 indica una correlación positiva perfecta, si es -1 indica una correlación negativa perfecta, y mientras más se acerque a 0 indica una falta de correlación. Es útil cuando los datos no siguen una distribución normal o cuando se presentan valores atípicos.

4.5. Modelo Lineal (LM)

Es una herramienta matemática utilizada para describir la relación entre una variable dependiente y y una o más variables independientes x_1, x_2, \dots, x_n . Un modelo lineal simple puede representarse como:

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$

Donde β_0 es la intersección, β_1 es el coeficiente que mide el cambio en y por cada unidad de cambio en x_1 y $\epsilon \sim \mathcal{N}(0, 1)$ es el error que se distribuye con media 0 y varianza 1. Un modelo lineal múltiple, considera más de una variable independiente, es por ello que el modelo se extiende a:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

Estos coeficientes $(\beta_0, \beta_1, \dots, \beta_n)$ se estiman a través de métodos como el de mínimos cuadrados, en que se minimiza la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos del modelo. Como principal característica, además de la linealidad de la relación de las variables, los errores obtenidos con este tipo de modelos siguen una distribución normal.

4.6. Modelo Lineal Generalizado (GLM)

Este tipo de modelo extiende el concepto de los modelos lineales mencionados anteriormente para permitir que la variable dependiente tenga distribuciones distintas a la normal y para modelar la relación entre la variable dependiente y las variables independientes a través de una función de enlace. La estructura general del modelo se expresa mediante tres componentes: la función de enlace g , la función de predicción lineal η y la distribución de la variable dependiente. La función de predicción lineal es:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

En este caso el error puede seguir una distribución normal, pero no necesariamente. La función de enlace g relaciona la media esperada de la variable dependiente $\mu = E(y)$ con la predicción lineal η :

$$g(\mu) = \eta$$

Dependiendo del tipo de datos y la distribución de la variable dependiente y , se elige una función de enlace adecuada, como por ejemplo las funciones de identidad, logística o logarítmica. Finalmente, la variable dependiente y sigue una distribución de la familia exponencial, como la normal, binomial, Poisson, gamma, entre otras.

4.7. Modelo Aditivo Generalizado para la Localización, Escala y Forma (GAMLSS)

El modelo GAMLSS (Generalized Additive Models for Location, Scale and Shape) es una extensión flexible de los modelos de regresión generalizados que permite modelar no solo la media de la variable dependiente, sino también otros parámetros asociados a la dispersión, la asimetría y la curtosis de su distribución. A diferencia de los GLM, en los cuales solo se modela la media como función de las covariables, los GAMLSS permiten modelar de manera independiente varios parámetros de la distribución de la variable respuesta, utilizando tanto funciones paramétricas como no paramétricas.

La variable respuesta Y se asume que sigue una distribución D , que puede ser una distribución común como la normal, gamma, entre otras. Los parámetros de esta distribución pueden ser modelados de forma independiente mediante funciones de enlace que relacionan cada uno de ellos con

covariables, permitiendo capturar características más complejas en los datos, como heterocedasticidad o variación en la asimetría.

Sea $Y \sim D(\mu, \sigma, \nu, \tau)$, donde D es la distribución elegida, y μ , σ , ν y τ representan la media (parámetro de ubicación), la desviación estándar (parámetro de escala), la asimetría y la curtosis (parámetros de forma) respectivamente. Cada uno de estos parámetros se vincula a las covariables mediante una función de enlace específica:

$$g_k(\theta_k) = \eta_k = h_k(X_k\beta_k) + \sum_{j=1}^m h_{jk}(x_{jk})$$

donde:

- $g_k(\theta_k)$ es la función de enlace para el parámetro θ_k de la distribución, como la media (μ), dispersión (σ), asimetría (ν), o curtosis (τ).
- η_k es el predictor lineal para θ_k .
- $h_k(X_k\beta_k)$ representa el componente lineal, donde X_k son los predictores y β_k son sus coeficientes.
- $\sum_{j=1}^m h_{jk}(x_{jk})$ es una suma de funciones de suavizado no lineales h_{jk} aplicadas a los predictores x_{jk} , lo que permite capturar relaciones complejas y no lineales entre estos predictores y el parámetro θ_k .

Este enfoque flexible permite que cada parámetro de la distribución de Y dependa de diferentes conjuntos de covariables, adaptándose así a las características específicas de los datos. Además, el uso de diferentes funciones de enlace para los distintos parámetros incrementa la flexibilidad del modelo para capturar patrones complejos en la relación entre las covariables y la variable respuesta. La estimación de los coeficientes β_k se realiza mediante el método de máxima verosimilitud. El logaritmo de la función de verosimilitud $l(\theta; y)$ se maximiza respecto a los parámetros θ de la distribución, donde:

$$L(\theta; y) = \prod_{i=1}^n f(y_i; \mu_i, \sigma_i, \nu_i, \tau_i)$$

En esta expresión, $f(y_i; \mu_i, \sigma_i, \nu_i, \tau_i)$ representa la función de densidad de la distribución D , evaluada en los valores observados y_i , y θ es el vector que contiene los parámetros de la distribución para cada observación.

4.8. Criterio de Información de Akaike (AIC)

Es una medida utilizada para evaluar la calidad de los modelos estadísticos para un conjunto de datos dado. Proporciona un medio para balancear el ajuste y la complejidad del modelo. Se basa en la teoría de la información y busca estimar la cantidad de información perdida cuando un modelo dado se utiliza para representar el proceso que generó los datos. La fórmula del modelo (Akaike, 1974) se muestra a continuación:

$$AIC = 2k - 2\ln(L) \tag{1}$$

donde:

- k es el número de parámetros en el modelo.

- L es la función de verosimilitud del modelo evaluada en los parámetros estimados.

Un modelo con un alto valor de $\ln(L)$ se ajusta bien a los datos y un modelo con más parámetros (mayor k) tiende a sobreajustarse a los datos. Un modelo con un menor valor de AIC es preferido porque se considera que tiene un mejor balance entre ajuste y simplicidad.

4.9. Criterio de Información Bayesiano (BIC)

Es una métrica utilizada en la selección de modelos estadísticos, especialmente en modelos de regresión o de series temporales. Su función es comparar varios modelos y elegir el que mejor se ajuste a los datos, realizando una penalización a los modelos más complejos. Se basa en la maximización de la probabilidad del modelo, la penalización se agrega según el número de parámetros. Un valor de BIC más bajo indica un modelo preferible, ya que implica un mejor ajuste con una penalización adecuada por el número de parámetros. A diferencia del AIC, el BIC penaliza más los modelos más complejos, si un modelo tiene menos parámetros pero ajusta de igual forma los datos será elegido sobre el otro.

$$\text{BIC} = -2 \cdot \ln(\hat{L}) + k \cdot \ln(n) \quad (2)$$

Donde:

- \hat{L} es el valor máximo de la función de verosimilitud del modelo.
- k es el número de parámetros estimados en el modelo.
- n es el número total de observaciones.

4.10. Eficiencia de Nash-Sutcliffe (NSE)

Es una métrica utilizada para evaluar la precisión de los modelos hidrológicos, especialmente en la predicción de series temporales. NSE mide la habilidad del modelo para predecir valores con respecto a la media de los valores observados. Su valor varía entre $-\infty$ y 1, donde un valor de 1 indica una predicción perfecta. Los valores negativos indican que el promedio de los datos observados sería un mejor predictor que el modelo.

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (3)$$

Donde:

- O_i son los valores observados.
- P_i son los valores predichos por el modelo.
- \bar{O} es el valor medio de los observados.
- n es el número de observaciones.

4.11. Sesgo (BIAS)

Métrica que mide el sesgo sistemático de un modelo, es decir, la tendencia del modelo a sobrestimar o subestimar sistemáticamente los valores observados. Un valor de BIAS positivo indica una sobrestimación, mientras que un valor negativo indica subestimación. Un BIAS de 0 indica que, en promedio, el modelo no tiene sesgo.

$$BIAS = \frac{\sum_{i=1}^n (P_i - O_i)}{n} \quad (4)$$

Donde:

- O_i son los valores observados.
- P_i son los valores predichos por el modelo.
- n es el número de observaciones.

4.12. Raíz del Error Cuadrático Medio (RMSE)

Es una medida de la diferencia entre los valores observados y los valores predichos por un modelo. Se utiliza comúnmente para evaluar la precisión de un modelo de predicción.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (5)$$

Un RMSE más bajo indica un mejor ajuste del modelo a los datos, ya que en este caso las predicciones son más cercanas a los valores reales.

4.13. Leave One Out Cross Validation (LOOCV)

Es un método de validación cruzada que se utiliza para evaluar la capacidad predictiva de los modelos al aplicarse a nuevas observaciones. A diferencia de otros enfoques que dividen aleatoriamente los datos en grupos, LOOCV permite reducir la variabilidad que surge al hacer esta división, garantizando una evaluación más precisa. En cuanto al procedimiento, en primer lugar se excluye un año de la muestra, que se utiliza como conjunto de validación, y luego se ajusta el modelo utilizando los restantes $N - 1$ años (conjunto de entrenamiento). Después de calibrar el modelo con esos datos, este se utiliza para predecir el valor del año excluido. Este proceso se repite N veces, cada vez dejando fuera un año diferente, para obtener una estimación robusta de la capacidad predictiva del modelo.

5. Resultados y Análisis

5.1. Análisis de Componentes Principales

Para este análisis se consideran 4 estaciones que entregan datos de precipitaciones diarias desde 1985 hasta 2020 para los meses de invierno, donde existen precipitaciones que superan los 0 [mmm]. El objetivo es determinar la homogeneidad de la cuenca. Si es posible predecir la mayor cantidad de información en una sola componente principal, se puede decir que la cuenca es homogénea y que,

a partir del análisis, es posible trabajar con los valores de los vectores propios, utilizando solo un conjunto de datos para la ventana de tiempo y simplificando el trabajo de la información.

Se obtiene el variograma mostrado en la Figura 3 a partir de las 4 series de precipitaciones diarias en el tiempo. Se observa como la primera componente principal es capaz de explicar un 85.6 % de la varianza total, la segunda componente principal explica un 6.7 % mientras que entre la tercera y la cuarta explican solo un 7.7 %.

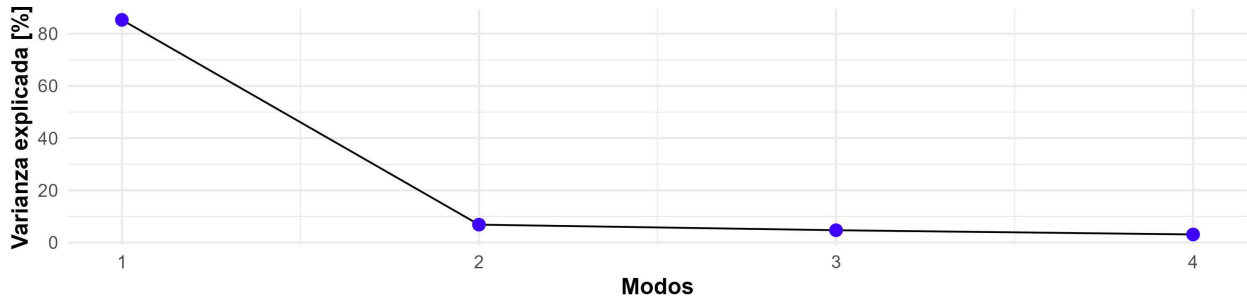


Figura 3: Variograma de precipitaciones diarias para la subcuenca Nilahue.

La contribución de cada serie de precipitaciones diarias a las componentes principales se puede ver en la Tabla 2, donde se notan valores similares en el aporte a la primera componente principal

Tabla 2: Aporte de las variables a cada una de las componentes principales.

Estación	Componente 1	Componente 2	Componente 3	Componente 4
La Palma	0.485	0.816	0.256	0.184
Ranguili	0.515	-0.073	-0.136	-0.843
El membrillo	0.495	-0.547	0.627	0.249
Nilahue Barahona	0.504	-0.172	-0.723	0.440

El análisis de componentes principales muestra que la primera componente explica más del 85 % de la varianza, lo que sugiere que las precipitaciones diarias en las cuatro estaciones meteorológicas son muy similares. Esto permite concluir que la cuenca es homogénea, es decir, que las precipitaciones ocurren de manera uniforme en todas las estaciones.

A partir del resultado se establece un modelo de aplicación de tipo estadístico, de esta forma se pueden realizar regresiones lineales simples, modelos lineales múltiples e incluso modelos lineales generalizados para predecir el caudal aportante a la laguna.

5.2. Correlación entre caudales y la CP de precipitaciones

Utilizando los datos previamente obtenidos mediante el análisis de componentes principales y los registros de caudales medios diarios proporcionados por la estación Nilahue en Santa Teresa, se busca, a través del test de Pearson, evaluar la relación lineal entre las variables. El objetivo es determinar si es factible predecir el caudal a partir de los datos de precipitación, siempre que la correlación entre estas variables sea suficientemente alta.

Adicionalmente, se emplea el test de correlación de Spearman, que permite identificar relaciones que no necesariamente son lineales entre las variables. Dado que se conoce que en una cuenca el caudal responde con cierto desfase a los eventos de precipitación, se analiza la correlación del caudal presente con las precipitaciones de días anteriores.

Asimismo, se examina la relación entre los caudales y las precipitaciones acumuladas durante un periodo de dos días utilizando los mismos test de correlación.

El propósito de aplicar tanto el test de Pearson como el de Spearman es identificar qué variable de precipitación muestra la correlación más alta con los caudales. Los mejores resultados de correlación se obtienen cuando se considera la precipitación con un desfase de dos días respecto al caudal. Esto se ilustra en las Figuras 4 y 5, donde la correlación más significativa se alcanza con el test de Pearson, obteniendo un coeficiente de 0.6. Este valor es significativo con un nivel de confianza del 95 % (P-Value menor a 0.05).

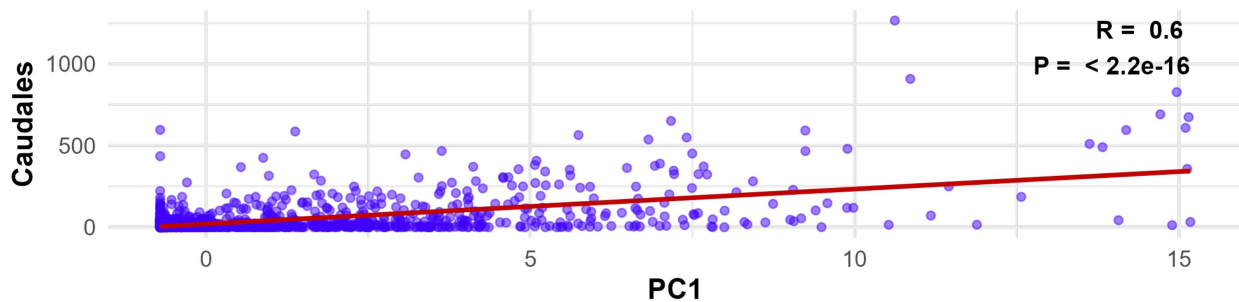


Figura 4: Test de correlación de Pearson para caudales y precipitaciones con desfase de 2 días.

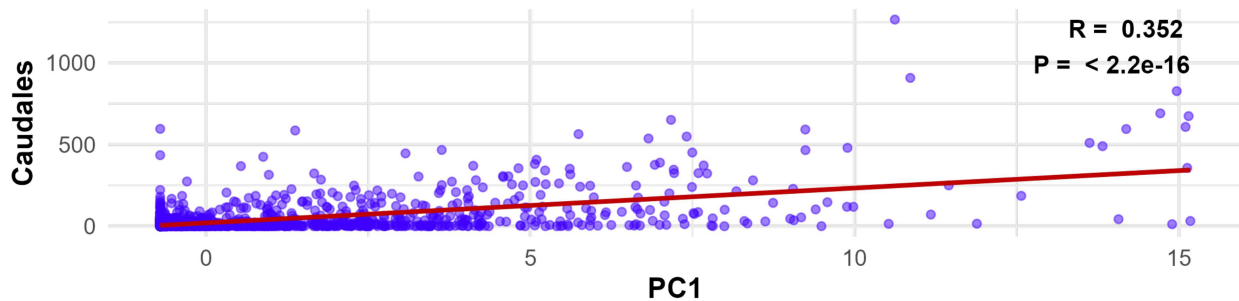


Figura 5: Test de correlación de Spearman para caudales y precipitaciones con desfase de 2 días.

Se evidencia en la mayoría de los datos de tormentas históricas que si se desfasa la precipitación en dos días respecto al caudal se llega a resultados como los mostrados en la Figura 6.

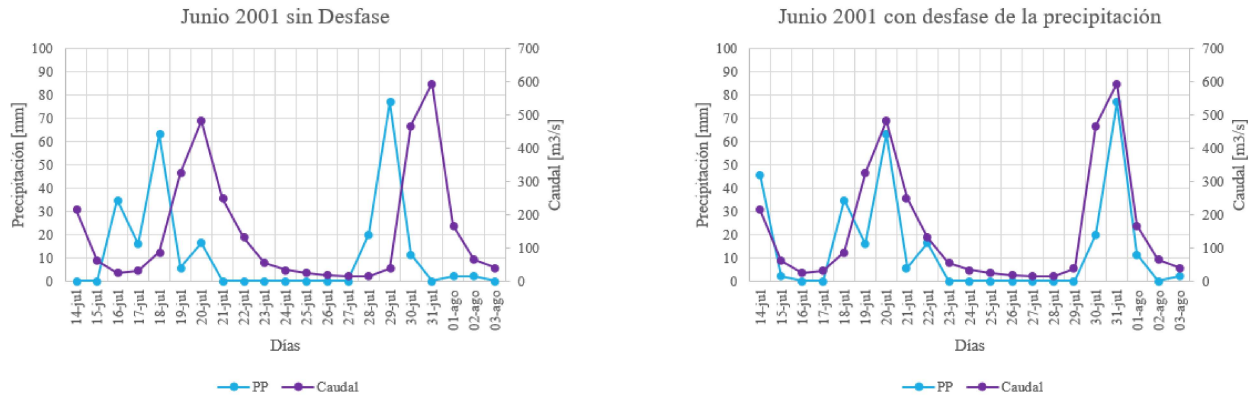


Figura 6: Tormenta histórica junio 2001 con caudales y precipitaciones con desfase de 2 días.

Por otro lado, se obtiene la relación del caudal y la precipitación pero en este caso acumulando la precipitación de dos días. Esta relación se obtiene a partir de los coeficientes mencionados anteriormente. A modo de resumen, se muestran los valores de los coeficientes de Pearson y Spearman realizando desfases de la precipitación en la Tabla 3. Con la mayor correlación dada por el coeficiente de Pearson con un valor de 0.64 para desfase de 1 día y con un valor de 0.60 para desfases de dos días. En el caso de la correlación de Spearman, se observan valores bastante similares para los diferentes tipos de desfase, siendo el más alto un coeficiente de 0.38 para un desfase de dos días.

Tabla 3: Coeficientes de correlación para caudales y precipitaciones acumuladas en dos días con desfases.

Desfase	Coefficiente de Pearson	Coefficiente de Spearman
1 día	0.64*	0.33*
2 días	0.60*	0.38*
3 días	0.37*	0.35*
4 días	0.24*	0.32*
5 días	0.18*	0.30*

* indica valor significativo para un nivel de confiabilidad de 95 %.

Los distintos test de correlación de Pearson y Spearman confirman una buena relación entre el caudal y las precipitaciones desfasadas dos días, con un coeficiente de correlación de Pearson de 0.6. Si se acumulan las precipitaciones en un periodo de dos días y se realiza el mismo análisis, es posible obtener una predicción fiable basada en precipitaciones de hasta tres días antes, manteniendo una correlación lineal.

5.3. Correlación entre caudales y caudales de días anteriores

Se requiere encontrar una relación entre caudales pasados y caudales presentes, de esta forma sería posible predecir el caudal futuro usando la data de mediciones de caudales de los días anteriores. Para ello, se vuelve a utilizar el coeficiente de correlación de Pearson y el Coeficiente de correlación de Spearman, buscando aquel que presente una correlación mayor entre las variables. Se hace una separación entre los caudales de meses en que la precipitación es significativa, denominado

en este estudio como los meses de "invierno", y los meses en que no hay precipitaciones frecuentes denominados meses de "temporada seca".

En este caso, se realiza el mismo análisis anteriormente mencionado utilizando coeficientes de correlación de Pearson y Spearman. Con el objetivo de encontrar una relación entre el caudal a predecir y los datos de caudales de días anteriores. Los resultados se muestran en la Tabla 4 y 5 con el valor de los coeficientes a modo de resumen.

Tabla 4: Coeficientes de correlación entre caudales y caudales con desfases para meses de invierno.

Desfase	Coeficiente de Pearson	Coeficiente de Spearman
1 día	0.73*	0.97*
2 días	0.41*	0.92*
3 días	0.30*	0.89*
4 días	0.23*	0.84*
5 días	0.19*	0.81*

* indica valor significativo para un nivel de confiabilidad de 95 %.

Tabla 5: Coeficientes de correlación entre caudales y caudales con desfases para meses de temporada seca.

Desfase	Coeficiente de Pearson	Coeficiente de Spearman
1 día	0.79*	0.98*
2 días	0.52*	0.96*
3 días	0.4*	0.94*
4 días	0.31*	0.93*
5 días	0.22*	0.92*

* indica valor significativo para un nivel de confiabilidad de 95 %.

Los coeficientes de correlación de Spearman ofrecen mejores resultados, lo que indica una relación no lineal entre estas variables. Los modelos pueden prever con bastante precisión los caudales hasta con cinco días de antelación tanto en meses de invierno como en la temporada seca, con coeficientes entre 0.92 y 0.98, cercanos al valor ideal de 1.

5.4. Influencia embalse Convento Viejo

A lo largo de la investigación se ha identificado una influencia externa que afecta los valores de caudal en la cuenca. En particular, se estudian los cambios que el caudal experimenta durante los meses de verano. El embalse, que forma parte de la cuenca del Estero Chimbarongo, se encuentra en una cuenca con un régimen pluvio-nival. Esto implica que el embalse acumula agua durante el deshielo, y un porcentaje de ese caudal es desviado hacia el Estero Nilahue a través de un túnel que conecta ambas cuencas. Debido a esta situación, se realiza una distinción entre un modelo de invierno, que no considera la influencia del embalse, y un modelo de verano, donde dicha influencia se incluye.

Se calcula el volumen de agua que circula por el Estero Nilahue a partir de los caudales medios mensuales durante los meses correspondientes a la “temporada seca” (Figura 7). Estos valores se comparan con las precipitaciones registradas en el mismo periodo (Figura 8).

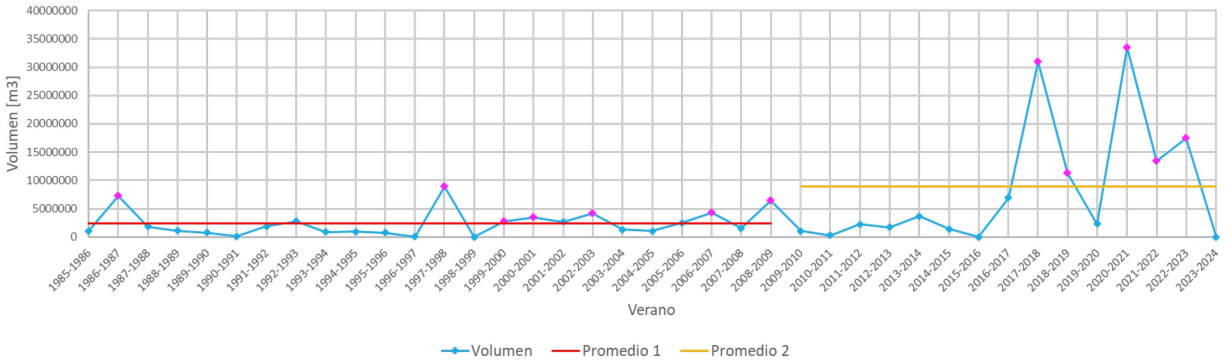


Figura 7: Volumen acumulado en meses de temporada seca.

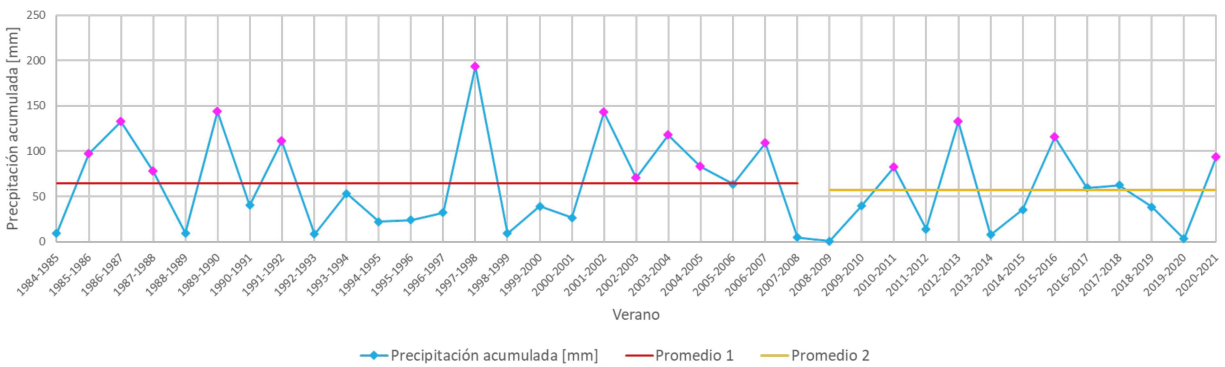


Figura 8: Precipitación acumulada en meses de temporada seca.

A partir de las figuras anteriores, se observa que, antes de la construcción del embalse y su conexión con la cuenca, las precipitaciones presentaban un promedio superior al registrado en los años posteriores. Sin embargo, a pesar de esta disminución en las precipitaciones, se nota un aumento en el volumen de caudal, particularmente a partir del año 2017, cuando este incremento se vuelve más abrupto durante las temporadas secas. Esto evidencia la influencia del embalse y los aportes de caudal provenientes de la cuenca vecina a través del túnel de conexión.

Por lo tanto, el embalse Convento Viejo juega un papel significativo en la predicción de caudal. Si se logra una buena coordinación con su gestión, el modelo podría mejorar aún más durante la temporada seca.

5.5. Modelos con desfase de 1 día para meses de invierno

En esta sección, se realizan diferentes tipos de modelos lineales para obtener el mejor ajuste posible en la predicción de caudal.

5.5.1. Modelo lineal múltiple

El modelo de regresión lineal múltiple se ha ajustado utilizando las siguientes variables:

- $Q(t - 1)$: Caudal desfasado en un día.
- $Q^2(t - 1)$: Caudal desfasado en un día al cuadrado.
- $P(t - 1)$: Precipitación con desfase de 1 día.
- $P_{D-2}(t - 1)$: Precipitación acumulada en 2 días con desfase de 1 día.
- $P_{D-2}^3(t - 1)$: Precipitación acumulada en 2 días con desfase de 1 día al cubo.

Se elige la mejor combinación de las variables utilizando el Criterio de Información de Akaike, donde el menor valor obtenido para AIC se dió con todas las variables mencionadas anteriormente. La ecuación del modelo es la siguiente:

$$\begin{aligned}
 Q(t) = & 1.1339 \\
 & + 7.3946 \cdot P_{D-2}^3(t - 1) \\
 & + 0.559 \cdot P_{D-2}(t - 1) \\
 & + 4.2577 \cdot P(t - 1) \\
 & + 0.81 \cdot Q(t - 1) \\
 & - 0.0005 \cdot Q^2(t - 1)
 \end{aligned} \tag{6}$$

Las variables que son significativas para el modelo cuentan con un valor de p menor a 0.05. Los coeficientes estimados del modelo y sus respectivas significancias se resumen a continuación en la Tabla 6:

Tabla 6: Significancias de las variables.

Variable	P-value
Intercepto	0.0264*
$P_{D-2}^3(t - 1)$	<2e-16*
$P_{D-2}(t - 1)$	0.083
$P(t - 1)$	<2e-16*
$Q(t - 1)$	<2e-16*
$Q^2(t - 1)$	<2e-16*

* indica valor significativo para un nivel de confiabilidad de 95 %.

Entre las estadísticas más importantes del modelo se encuentra el R el cual es igual a 0.87 (Figura 9). Además, el NSE de 0.76 lo que indica que el 76 % de la variabilidad en la variable dependiente se explica por el modelo. Además, el modelo tiene un valor p menor a 2e-16, por lo que el modelo es significativo para una confianza del 95 %.

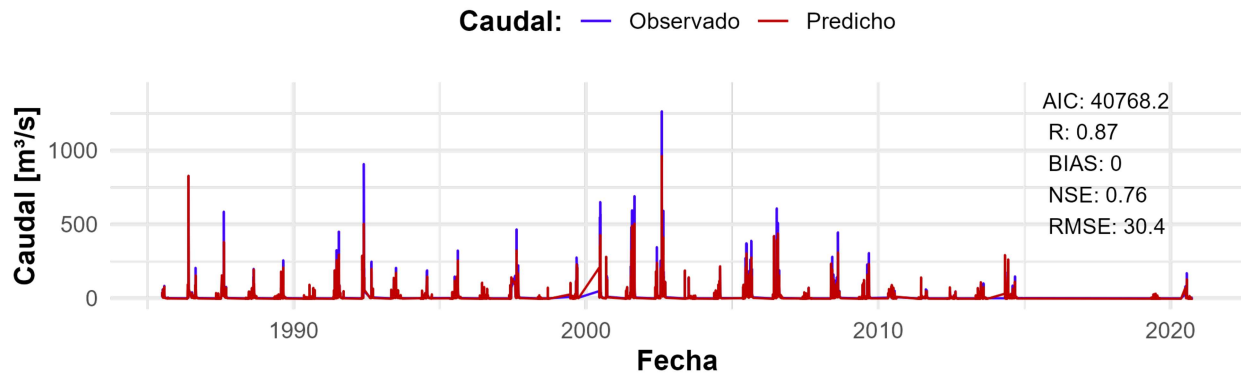
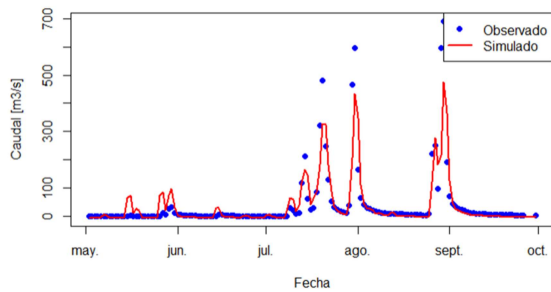
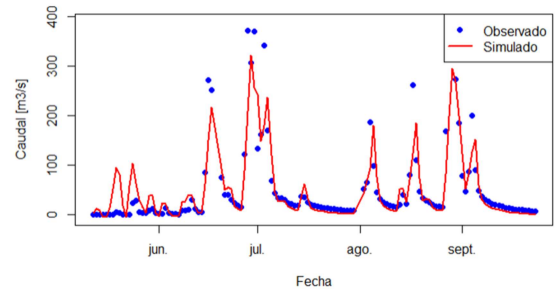


Figura 9: Valores Observados y Simulados.

A continuación, vemos las mismas gráficas pero limitadas a años específicos, los cuales fueron años húmedos (Figura 10a y 10b).



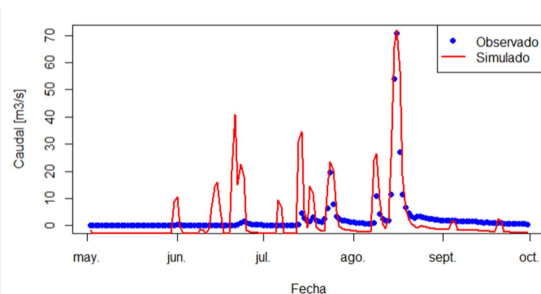
(a) Valores observados y simulados, año 2001.



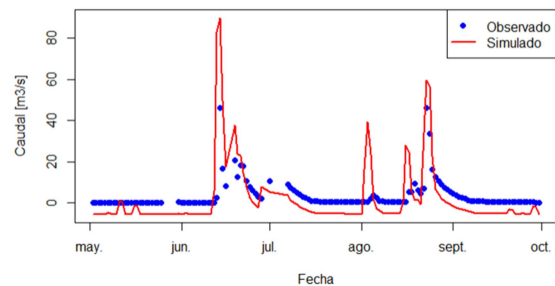
(b) Valores observados y simulados, año 2005.

Figura 10: Valores observados y simulados para años húmedos.

A continuación, se muestran las gráficas para años secos en las Figuras 11a y 11b.



(a) Valores observados y simulados, año 2007.



(b) Valores observados y simulados, año 2012.

Figura 11: Valores observados y simulados para años secos.

Una de las características del modelo lineal es que los errores del modelo se distribuyen de forma

normal. La Figura 12 muestra la gráfica Q-Q realizada para los residuos, en que si se distribuyen normalmente estos deben seguir la línea de tendencia de color rojo.

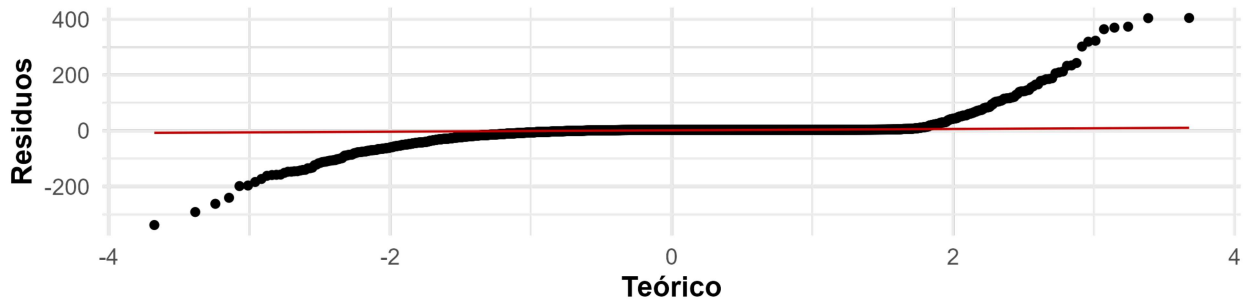


Figura 12: Gráfico Q-Q de los residuos del modelo.

Por otra parte, se puede ver la gráfica de los residuos v/s los valores predichos en la Figura 13.

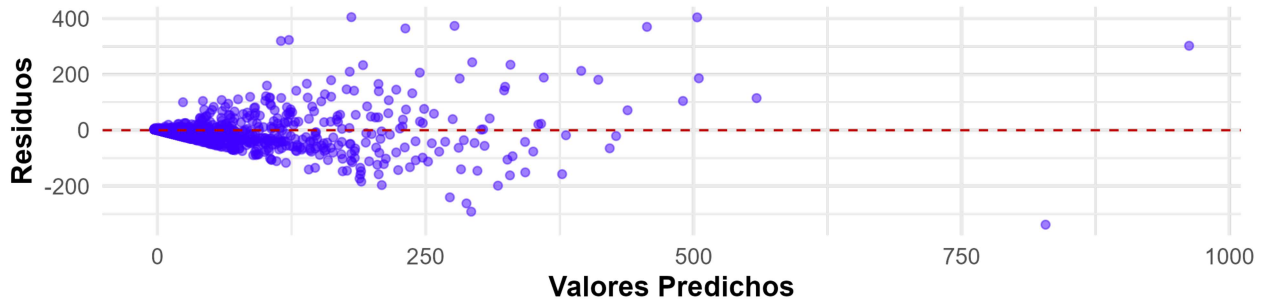


Figura 13: Gráfico de dispersión de los residuos v/s valores predichos.

Al comparar los valores observados y predichos, se nota que, en años húmedos, el modelo predice bien los caudales, pero en caudales altos subestima los valores reales, un aspecto crucial dado que la predicción de crecidas es esencial. En años secos, el modelo sobrestima tanto los caudales altos como los bajos. Esto se corrige separando las predicciones para caudales bajos y altos.

5.5.2. Modelo lineal múltiple con división de data en 2 partes

A partir de los resultados del modelo anterior, se plantea la idea de dividir la serie de datos en dos partes, obteniendo así dos modelos independientes: uno para los casos en que las precipitaciones del día anterior superen un umbral determinado, y otro para cuando dichas precipitaciones sean inferiores a ese umbral. De este modo, se aborda la predicción de un caudal influenciado por la

precipitación y otro caudal que no se ve afectado por ella. El umbral se selecciona mediante un proceso iterativo, probando diferentes valores y evaluando el rendimiento de los modelos en función de las métricas AIC, BIC, R, BIAS y NSE. En este caso se escoje un umbral de 0.5 [mm].

El modelo de predicción de los periodos en que hay precipitaciones se basa en la siguiente fórmula:

$$\begin{aligned}
 Q(t) = & -9.1274 \\
 & + 0.0129 \cdot P_{D-2}^2(t-1) \\
 & + 0.3857 \cdot P_{D-2}(t-1) \\
 & + 1.1624 \cdot Q(t-1) \\
 & - 7e-04 \cdot Q^2(t-1)
 \end{aligned} \tag{7}$$

Podemos ver los valores predichos y observados obtenido a través de la validación cruzada en la Figura 14:

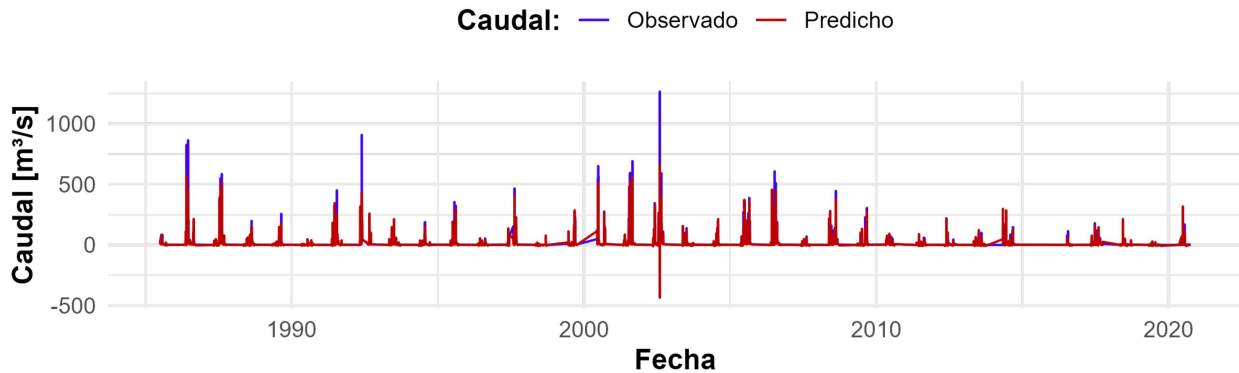


Figura 14: Valores Observados y Predichos de Caudal con influencia de precipitaciones, LM múltiple.

Se observa un valor que esta fuera del rango de las predicciones, el cual es -435.26 [m3/s]. Este es el caudal a predecir del día siguiente en que ocurre el caudal histórico de 1265 [m3/s].

5.5.3. Modelo GAMLSS Log-Normal

En simultáneo al Modelo Lineal se realiza un Modelo GAMLSS con la distribución Log-Normal, donde se obtienen las siguientes predicciones cuando hay presencia de precipitaciones como lo muestra la Figura 15.

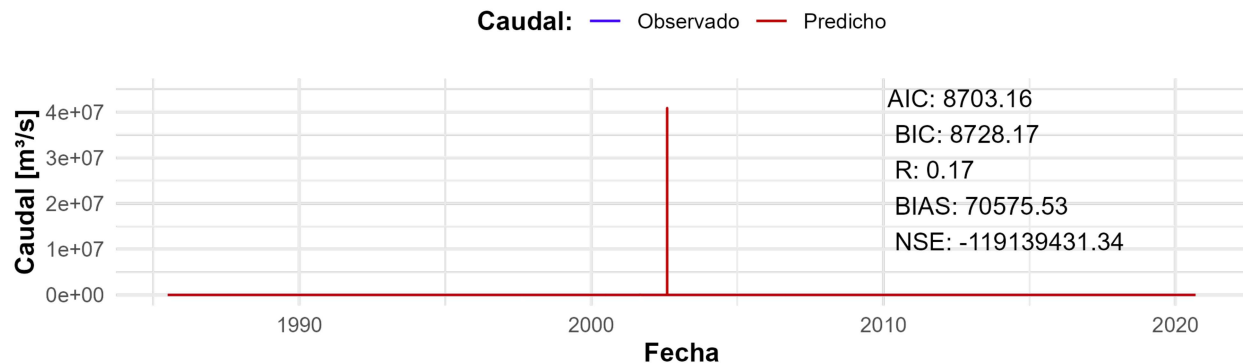


Figura 15: Valores Observados y Predichos de Caudal con influencia de precipitaciones, Modelo GAMLSS.

Donde se vuelve a observar un caudal fuera del rango de valores predichos. Este ocurre para la predicción del mismo día que en el caso anterior: el caudal del día posterior al día donde ocurre el caudal histórico.

Al dividir los datos en dos categorías, se observan valores de magnitud significativa, tanto positivos como negativos, causados por un mal ajuste debido al caudal histórico. En el caso del modelo lineal múltiple, la ecuación incluye un coeficiente negativo para el caudal al cuadrado, lo que provoca la predicción de alta magnitud y negativa. Para el GAMLSS Log-Normal, la variable caudal tiene un valor positivo, lo que lleva al modelo a sobrestimar el caudal futuro si el caudal del día anterior es alto, debido a la naturaleza exponencial del logaritmo. Por ello, es necesario que el modelo distinga cuando hay un pick de caudal y en vez de seguir aumentando el caudal este se disminuya. Lo anterior se logra identificando ascensos y descensos en los caudales.

5.5.4. Modelo GAMLSS Distribución Normal con 3 etapas

Con el fin de corregir el valor que se encuentra fuera del rango de predicciones, se implementa una nueva división basada en las precipitaciones de 1 y 2 días anteriores. Se calcula la pendiente entre estas precipitaciones: si la pendiente es positiva, indica un aumento en la precipitación, lo que sugiere que el caudal también aumentará. Por el contrario, si la pendiente es negativa, significa que la precipitación está disminuyendo, y, en consecuencia, se espera una disminución en el caudal. Este enfoque permite evitar la sobreestimación o subestimación del caudal debido a los caudales de días previos, al reconocer si el caudal está en ascenso o descenso.

Adicionalmente, para prevenir que el modelo compense de manera inapropiada, se decide no utilizar simultáneamente la variable del caudal del día anterior junto con su versión al cuadrado.

En la Figura 16 se muestra el gráfico de valores observados y predichos a través del tiempo junto con el gráfico de dispersión de valores predichos versus observados para el caudal medio. En la Figura Anexo 1 se muestra para el caudal de ascenso y en la Figura Anexo 2 para el de descenso.

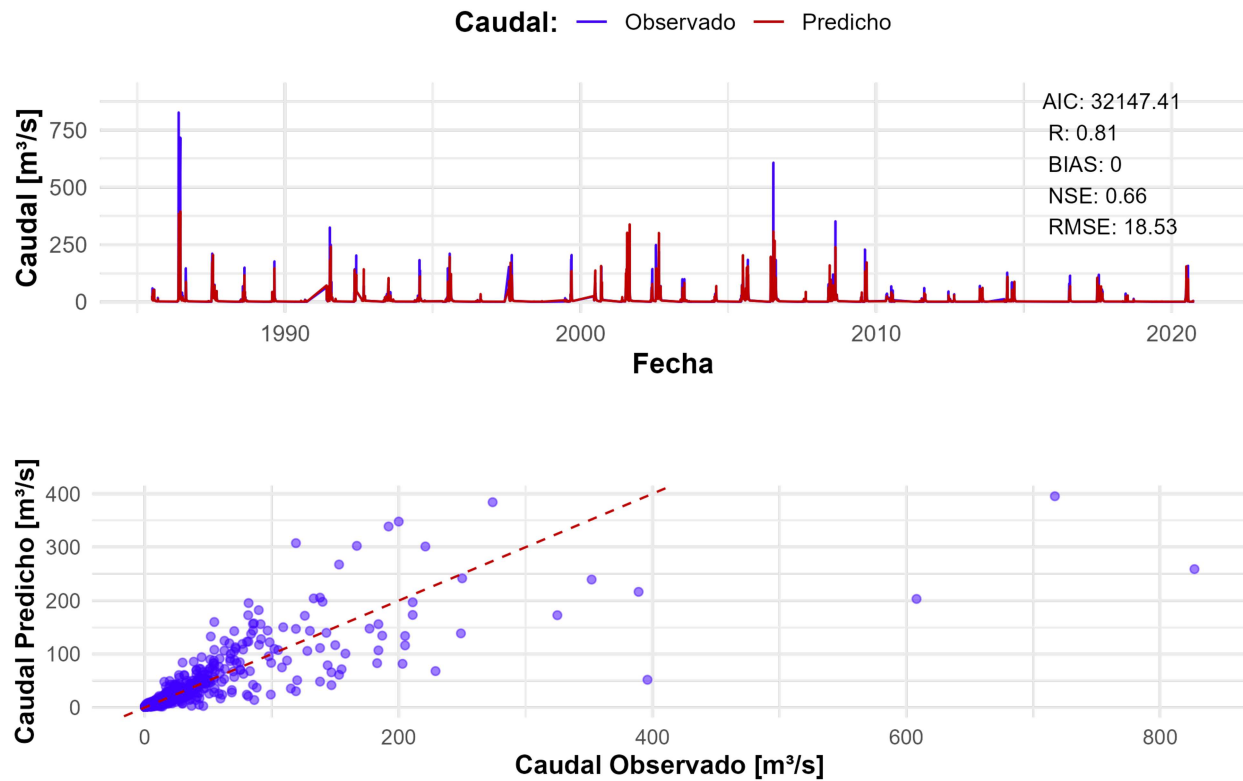


Figura 16: Valores Predichos y Observados de Caudal Medio, Modelo GAMLSS Distribución Normal.

Los gráficos para el caudal de ascenso y descenso se encuentra en Anexo 1 y Anexo 2.

El modelo con tres divisiones proporciona valores de R bastante buenos. Sin embargo, también presenta problemas, ya que dentro de la categoría de caudal medio aún se encuentran valores elevados que pueden afectar en la predicción de caudales altos al no tener esta información en los datos que entrenan los modelos.

5.5.5. Modelo GAMLSS con Distribución Normal con 4 etapas

Dado el resultado de 3 etapas, se puede observar en los gráficos del caudal medio que este toma valores que alcanzan los 700 [m³/s], por ende, se puede comprobar que el modelo está asumiendo caudales que no son bajos como caudal base. Esto influye directamente en la exactitud de la predicción de los caudales que si tienen precipitación aportante.

Si analizamos estos datos de caudal que se almacenan como caudales medios pero tienen un orden de magnitud mayor, podemos ver lo señalado en las Figuras 17 y 18. En la Figura 17 se nota que el aumento del caudal no es posible de anticipar con la precipitación de dos días antes, ya que la precipitación es cercana a 0 dos días antes del caudal de casi 200 [m³/s]. En la Figura 18, se ve un comportamiento similar entre la precipitación y caudal desfasados en dos días, pero al bajar a casi 0 la precipitación se asume que el caudal será bajo, cuando en realidad el suelo se encuentra saturado por días anteriores y el caudal no evidencia una baja abrupta a un caudal cercano a 0.

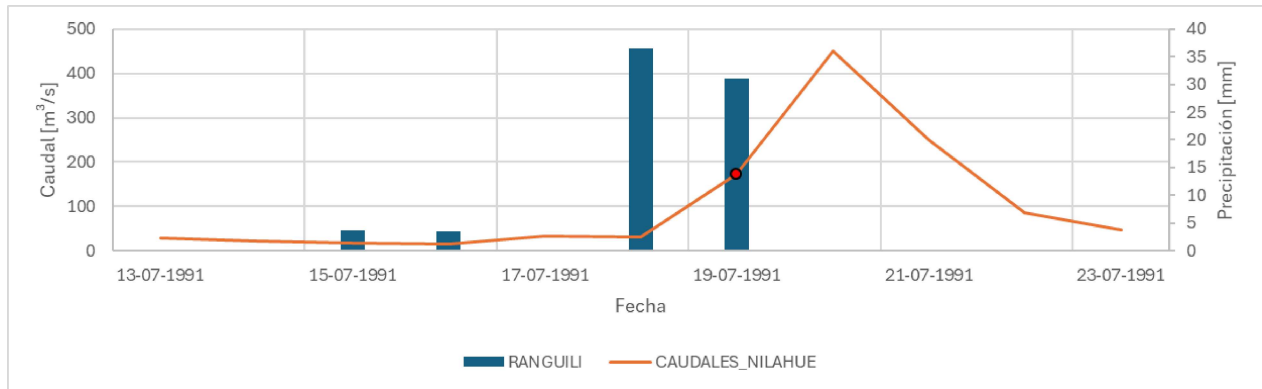


Figura 17: Valores de precipitaciones y caudal año 1991.

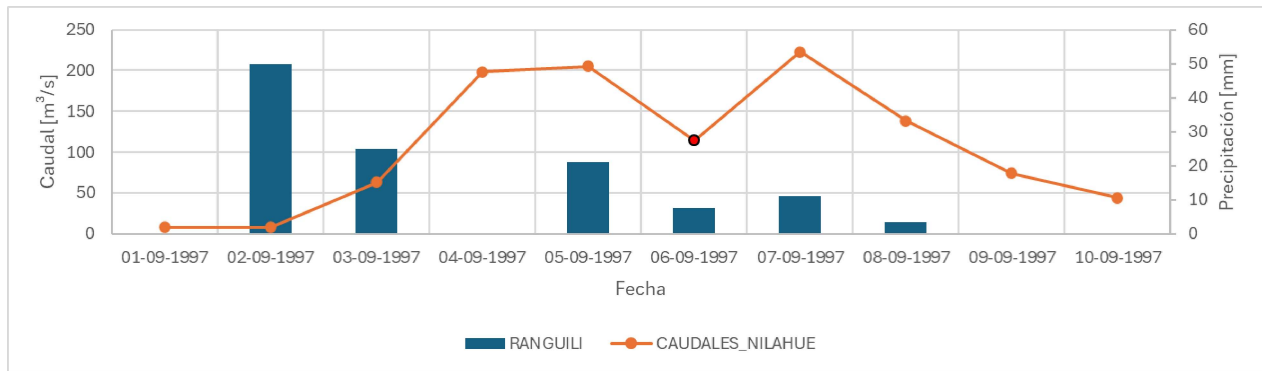


Figura 18: Valores de precipitaciones y caudal año 1997.

Es por ello, que se propone una nueva forma de dividir la data, que considere como primer filtro el valor de las precipitaciones acumuladas de 1, 2 o 3 días antes mayores a cierto umbral para considerar a un caudal como **Caudal de Crecida**. La data restante correspondería a caudales de menor magnitud, como **Caudal Medio** o caudales que se encuentren en la curva antes o después del pick, denominados **Caudal de Ascenso y Descenso**. Para hacer la distinción, se puede utilizar el caudal de días anteriores y definir un umbral para diferenciar caudal base de los caudales de curvas de ascenso y descenso. Finalmente, se mantiene la división entre ascenso o descenso según la pendiente de precipitaciones de días anteriores.

Para proyectar lo anterior de manera gráfica, se obtienen gráficos de dispersión entre el caudal "presente" versus precipitaciones acumuladas en días anteriores (Figura 19) y caudales de días anteriores (Figura 20).

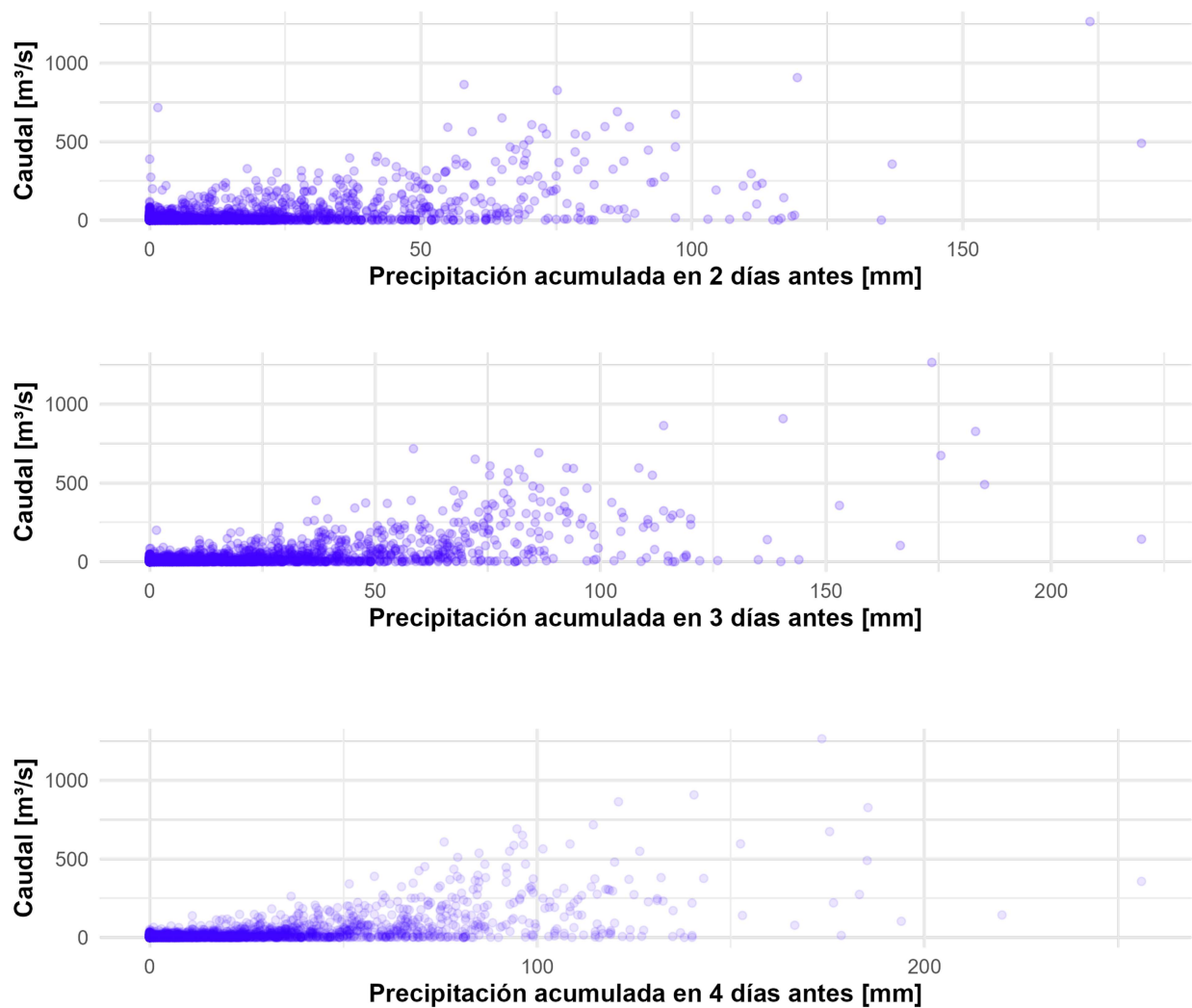


Figura 19: Caudales versus Precipitaciones Acumuladas en días anteriores.

En la Figura 19 se observa que a medida que se aumenta la cantidad de días antes en que se acumula la precipitación, se tienden a concentrar los valores bajos de caudal bajo un umbral de precipitación, como lo son los 50 [mm] para el caso de la precipitación acumulada en 4 días antes.

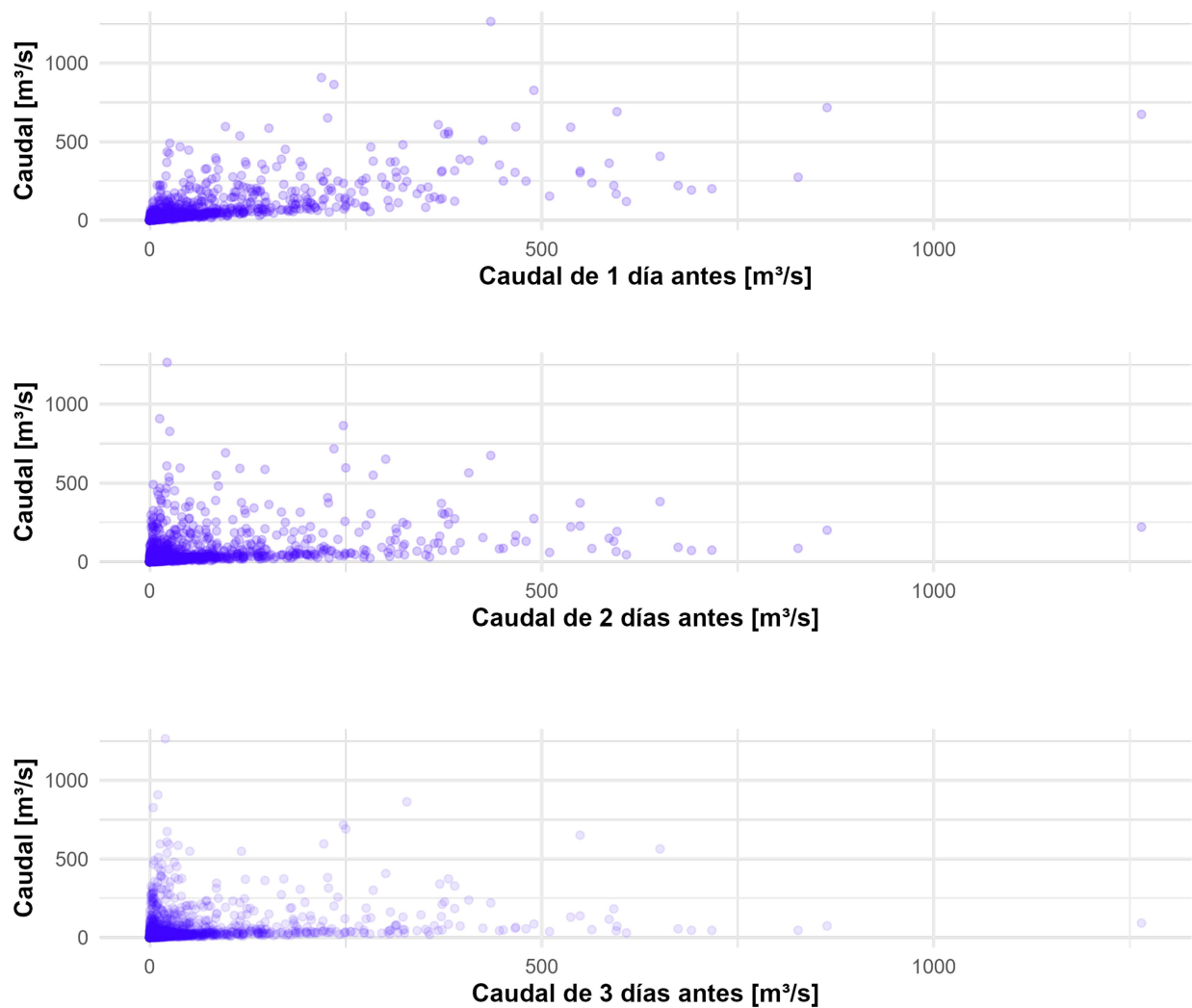


Figura 20: Caudales versus Caudales de días anteriores.

En la Figura 20 se observa que a medida que se aumenta la cantidad de días antes del caudal, los caudales del día presente tienden a dispersarse, mientras que si vemos el caudal del día anterior los valores bajos de caudal presente tienden a concentrarse cerca de los caudales del día anterior cercanos a 0.

Se fija como umbral de las precipitaciones acumuladas de 4 días antes el valor de 50 [mm]. Y para el umbral de caudal del día anterior se utiliza el caudal base, el cual se calcula como la mediana de los datos que toma el valor de 2.35 [m³/s]. La división se muestra gráficamente en la Figura 21.

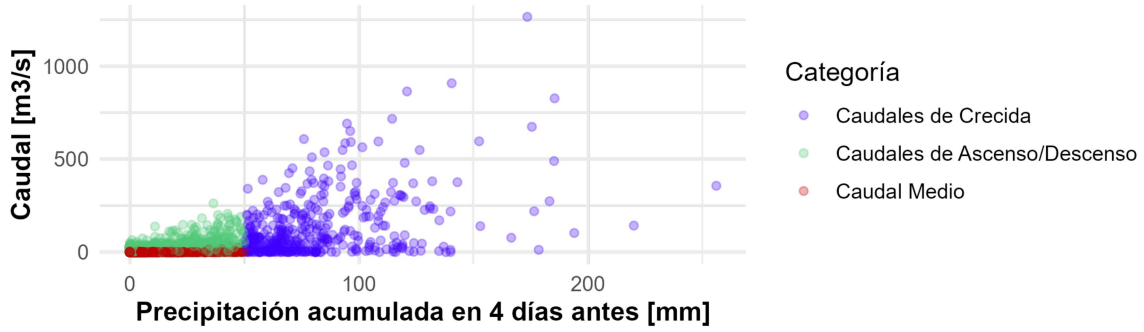


Figura 21: División de la data según caudales y precipitaciones acumuladas de días anteriores.

Usando la división anterior, se ajustan los modelos utilizando las siguientes variables:

- $Q(t - 1)$: Caudal desfasado en un día.
- $P(t - 1)$: Precipitación con desfase de 1 día.
- $P_{D-2}(t - 1)$: Precipitación acumulada en 2 días con desfase de 1 día.
- $P_{D-2}^2(t - 1)$: Precipitación acumulada en 2 días con desfase de 1 día al cuadrado.

Se elige la mejor combinación de las variables utilizando el Criterio de Información de Akaike. Las variables que son significativas para los modelos cuentan con un valor de p menor a 0.05.

El esquema de la Figura 22 muestra los cuatro modelos construidos y sus respectivas variables independientes.

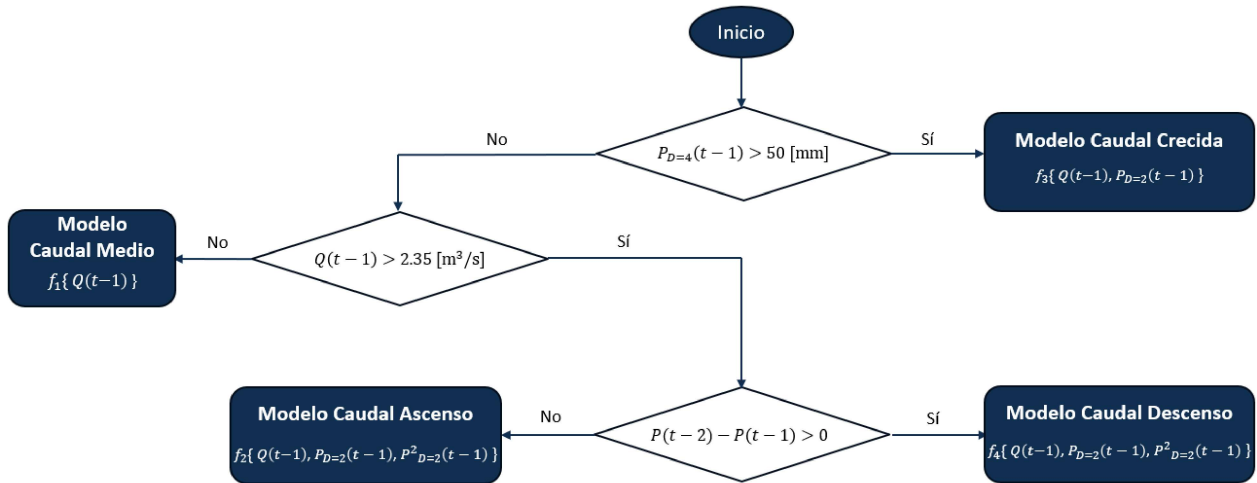


Figura 22: Esquema de modelos GAMLSS.

Para el **Modelo Caudal Medio** se muestran los coeficientes estimados y sus respectivas significancias, las cuales se resumen en las Tablas 7 y 8.

- **Función de enlace para μ :** identity

Tabla 7: Coeficientes de μ

Variable	Coeficiente	P-Value
Intercepto	0.07615	< 0.115
$Q(t - 1)$	1.14223	< 2e-16*

* indica valor significativo para un nivel de confiabilidad de 95 %.

- **Función de enlace para σ :** log

Tabla 8: Coeficientes de σ

Variable	Coeficiente	P-Value)
Intercepto	0.57912	< 2e-16 *

* indica valor significativo para un nivel de confiabilidad de 95 %.

Para el **Modelo Caudal de Ascenso** se muestran los coeficientes estimados y sus respectivas significancias, las cuales se resumen en las Tablas 9 y 10.

- **Función de enlace para μ :** identity

Tabla 9: Coeficientes de μ

Variable	Coeficiente	P-Value)
Intercepto	-0.11218	0.807
$P_{D-2}^2(t - 1)$	0.02167	2.82e-11 *
$P_{D-2}(t - 1)$	0.51393	3.59e-06 *
$Q(t - 1)$	0.79502	< 2e-16*

* indica valor significativo para un nivel de confiabilidad de 95 %.

- **Función de enlace para σ :** log

Tabla 10: Coeficientes de σ

Variable	Coeficiente	P-Value
Intercepto	2.5455	< 2e-16*

* indica valor significativo para un nivel de confiabilidad de 95 %.

Para el **Modelo Caudal Crecida** se muestran los coeficientes estimados y sus respectivas significancias, las cuales se resumen en las Tablas 11 y 12.

- **Función de enlace para μ :** identity

Tabla 11: Coeficientes de μ

Coeficientes de Mu	Coeficiente	P-Value
Intercepto	-42.8391	1.56e-05 *
$P_{D-2}(t-1)$	2.3596	< 2e-16 *
$Q(t-1)$	0.6532	< 2e-16 *

* indica valor significativo para un nivel de confiabilidad de 95 %.

- **Función de enlace para σ :** log

Tabla 12: Coeficientes de σ

Coeficientes de Sigma	Coeficiente	P-Value
Intercepto	4.69776	<2e-16*

* indica valor significativo para un nivel de confiabilidad de 95 %.

Para el **Modelo Caudal de Descenso** se muestran los coeficientes estimados y sus respectivas significancias, las cuales se resumen en las Tablas 13 y 14.

- **Función de enlace para μ :** identity

Tabla 13: Coeficientes de μ

Coeficientes de Mu	Coeficientes	P-Value
Intercepto	1.38532	0.07605
$P_{D-2}(t-1)$	0.73900	< 2e-16 *
$P(t-1)$	0.84229	0.00206 *
$Q(t-1)$	0.49894	< 2e-16 *

* indica valor significativo para un nivel de confiabilidad de 95 %.

- **Función de enlace para σ :** log

Tabla 14: Coeficientes de σ

Coeficientes de Sigma	Coeficiente	P-Value
Intercepto	2.51975	<2e-16*

* indica valor significativo para un nivel de confiabilidad de 95 %.

Las gráficas con los valores observados y los simulados de los 4 modelos se encuentran en el Anexo Anexo 3. Se unen los gráficos que muestran los valores predichos y observados en uno solo, en la Figura 23.

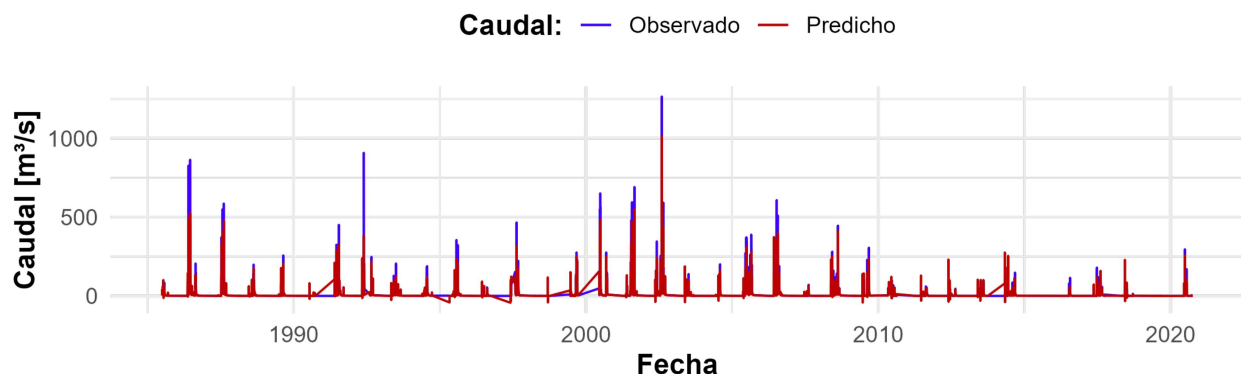


Figura 23: Valores Observados y Predichos GAMLSS Normal.

Además, para los cuatro modelos se aplica el Leave One Out Cross Validation (LOOCV), obteniendo los resultados mostrados en el Anexo Anexo 4.

Los resultados y métricas más representativas se muestran a modo de resumen en la siguiente Tabla 15.

Tabla 15: Métricas de modelos y sus validaciones cruzadas.

Caudal	Medio		Ascenso		Crecida		Descenso	
Métrica	Modelo	LOOCV	Modelo	LOOCV	Modelo	LOOCV	Modelo	LOOCV
AIC	9977	9977	11221	11221	5556	5256	3602	3602
R	0.47	0.46	0.84	0.83	0.75	0.74	0.9	0.9
NSE	0.22	0.21	0.71	0.69	0.56	0.55	0.81	0.8
RMSE	1.78	1.79	12.75	13.5	109.7	111.16	12.43	12.85

Para el modelo completo se tienen las siguientes métricas de la Tabla 16.

Tabla 16: Métricas del modelo completo.

Modelo	R	NSE	RMSE
Modelo GALMSS 4 etapas	0.85	0.72	33.7

Otros modelos GAMLSS fueron aplicados con distribuciones como la log-normal, gamma y otras, obteniendo resultados no deseados. Como se muestra en el Anexo Anexo 5, donde no es posible replicar la magnitud de los caudales debido a la excesiva presencia de caudales bajos.

Se puede concluir que se obtiene un modelo óptimo para la temporada de invierno que incorpora cuatro tipos de caudal. Este modelo permite predecir caudales medios menores a 60 [m³/s], solucionando el problema que se tenía dividiendo la data original en tres. Aunque en caudales pequeños su ajuste es inferior al modelo de tres fases, se observa una mejora en la predicción del caudal en descenso, mientras que la predicción del caudal ascendente se mantiene, y se obtiene una predicción aceptable de las crecidas, con un coeficiente R de 0.75.

5.6. Modelos con desfase de 1 día para meses de temporada seca

Se crea ahora el modelo para los meses de temporada seca. Dado que en esta temporada las precipitaciones son casi nulas, se realiza el modelo utilizando caudal del día anterior.

El modelo de regresión lineal múltiple se ha ajustado utilizando las siguientes variables:

- $Q(t - 1)$: Caudal desfasado en un día.
- $Q^2(t - 1)$: Caudal desfasado en un día al cuadrado.

Se elige la mejor combinación de las variables utilizando el Criterio de Información de Akaike, donde el menor valor obtenido para AIC se dio con la variable independiente caudal del día anterior.

La ecuación del modelo es la siguiente:

$$Q(t) = 0.15718 + 0.79229 \cdot Q(t - 1) \quad (8)$$

Las variables que son significativas para el modelo cuentan con un valor de p menor a 0.05. Los coeficientes estimados del modelo y sus respectivas significancias se resumen a continuación en la Tabla 17:

Tabla 17: Significancias de las variables.

Variable	P-value
Intercepto	0.000163 *
$Q(t - 1)$	< 2e-16 *

* indica valor significativo para un nivel de confiabilidad de 95 %.

A continuación, se muestra una gráfica con los valores observados y los simulados del modelo en la Figura 24.

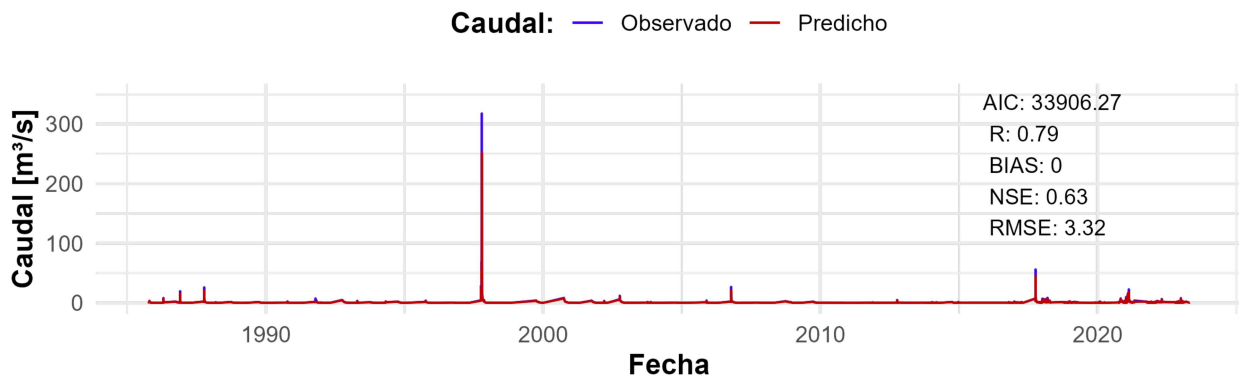


Figura 24: Valores Observados y Simulados Temporada Seca.

Además, se aplica el Leave One Out Cross Validation (LOOCV), obteniendo los siguientes resultados mostrados en las Figuras 25:

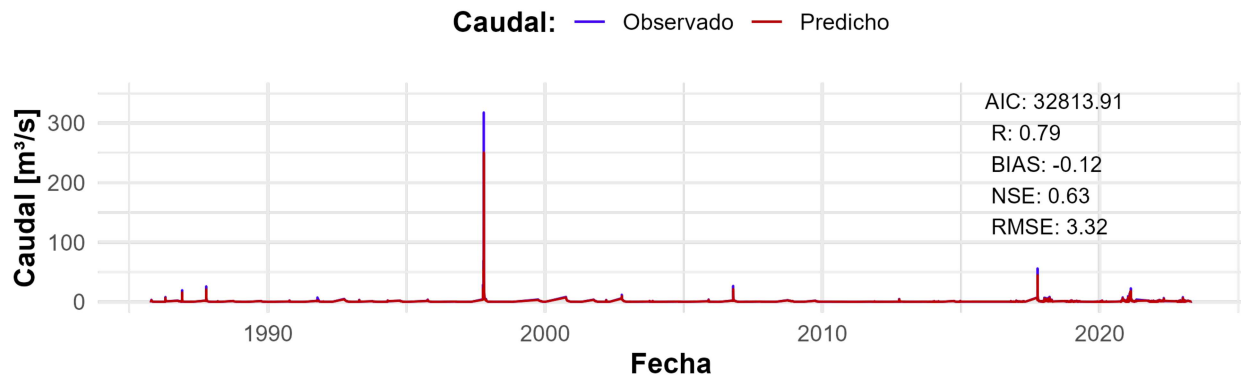


Figura 25: Validación cruzada, valores Observados y Predichos Temporada Seca.

En cuanto a los modelos GAMLSS, no fue posible calcular resultados para la validación cruzada con ninguna distribución debido a la particularidad de los datos.

Se ve claramente, en la Figura 24, como existe un caudal que se encuentra fuera del rango del resto de la data. Lo anterior, se ve en detalle en la Figura 26, donde se aprecia que este caudal alto se debe a las precipitaciones ocurridas y no a un factor externo. Precipitaciones que además no son comunes en estas fechas.

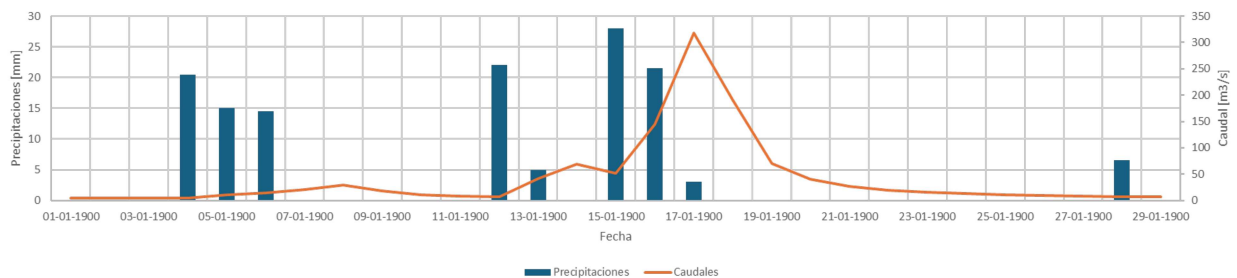


Figura 26: Tormenta Octubre 1997, Temporada Seca.

Se obtiene un modelo lineal múltiple con buenos resultados en sus métricas, con un R de 0.79. Sin embargo, al existir muchos caudales bajos no es posible la implementación de modelos GAMLSS.

6. Conclusiones y recomendaciones

En esta investigación, analizamos cómo es posible predecir el caudal presente a través de los datos de caudal y precipitaciones de días anteriores, enfocándonos en usar modelos de tipo estadísticos como: el modelo lineal y los modelos aditivos generalizados para la localización, escala y forma, evaluándolos según métricas y validándolos a través del “Leave one out cross validation”.

Los resultados obtenidos demuestran que se tiene una buena predicción de caudales altos, así como los ascensos y descensos de caudal. Esto es especialmente importante, ya que los fenómenos extremos, como inundaciones, son las principales preocupaciones en el contexto del cambio climático. Se destaca especialmente la detección de la influencia del embalse convento viejo en los caudales de verano, ya que esta influencia externa fue impulsada por sospechas de los locatarios y ahora ha sido comprobada de manera empírica.

Este trabajo destaca por su enfoque en la predicción de caudal a través de datos de días anteriores, ya que permite una aplicación simple a futuro, a través de las ecuaciones de los modelos, e inmediata, al tener disponibles los valores de precipitación y caudal de alguna de las estaciones.

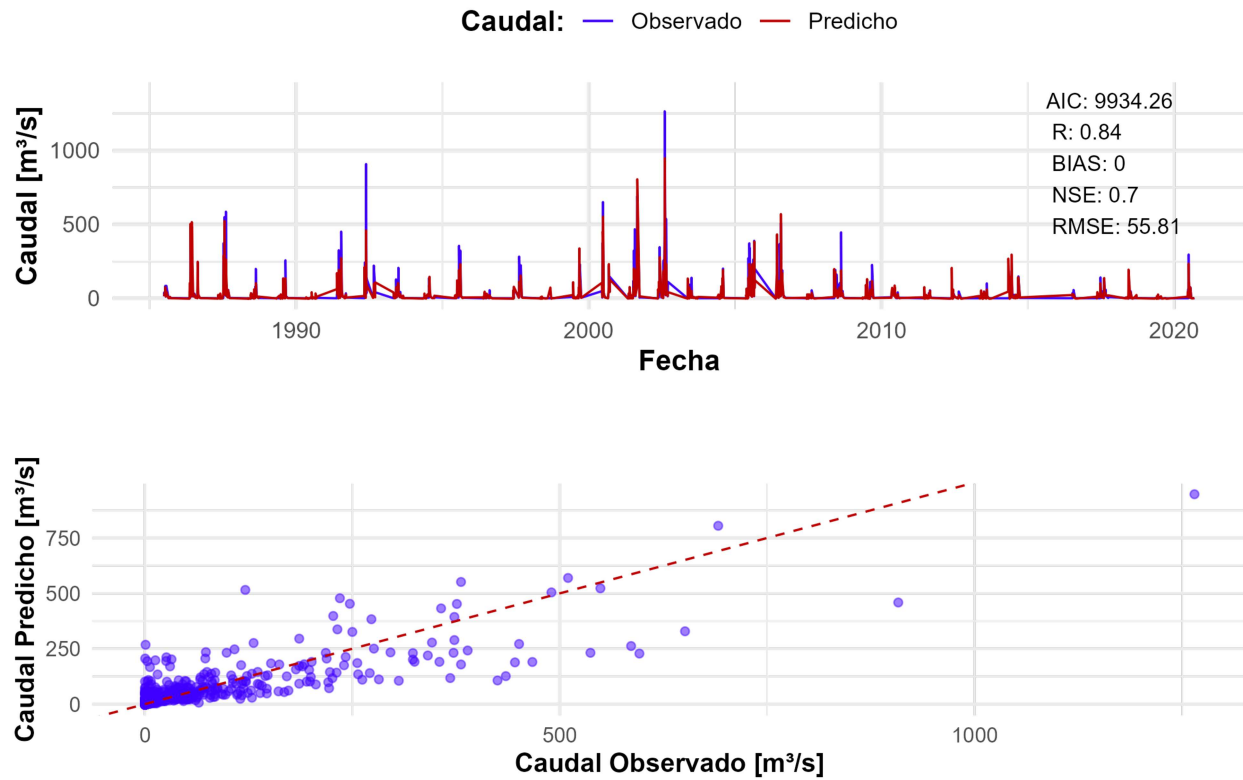
Todavía existe potencial para desarrollar modelos que logren un ajuste más preciso para todo tipo de caudal, desde niveles bajos hasta picos máximos. El análisis de precipitaciones y caudales revela un fenómeno inherentemente no lineal, con una dinámica que ocurre en una escala temporal horaria. Esto se observa, por ejemplo, cuando en algunas tormentas el aumento del caudal puede preverse con 2 días de anticipación, mientras que, en otras, el incremento ocurre 3 días antes. Esta variabilidad se debe a que las precipitaciones aumentan en momentos intermedios entre 2 y 3 días, lo cual, al trabajar con datos diarios, genera una pérdida de precisión temporal. Por ello, se recomienda que en futuros estudios se utilicen datos con resolución horaria para capturar mejor la dinámica del sistema. En cuanto a la diferenciación entre caudales medios, de crecida y en fases de ascenso o descenso, el enfoque actual se basa en el uso de umbrales estáticos, los cuales pueden ser limitados al depender de reglas rígidas que se cumplen o no. Aquí es donde el uso de técnicas de machine learning podría ofrecer una ventaja significativa. Ya que se tendría el potencial de detectar patrones más complejos en los datos, permitiendo identificar el comportamiento del caudal de manera más precisa y adaptativa.

Este trabajo es solo un paso para la predicción de caudales de la subcuenca Nilahue, y se espera ver avances exitosos siguiendo esta dirección.

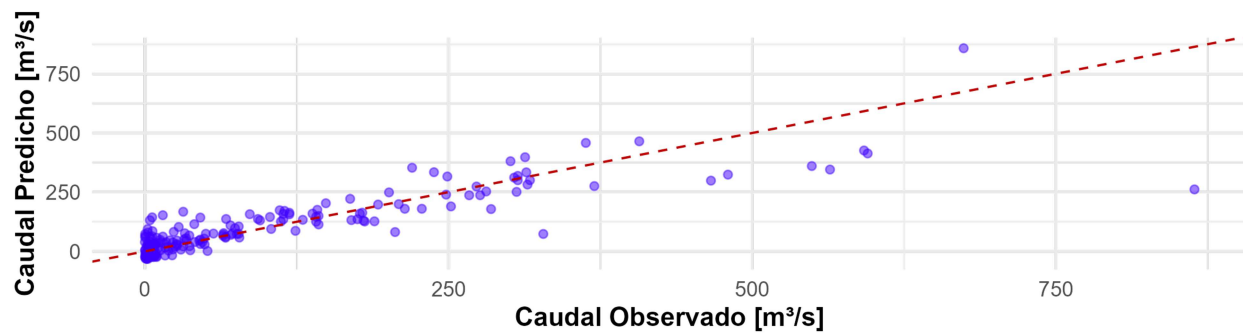
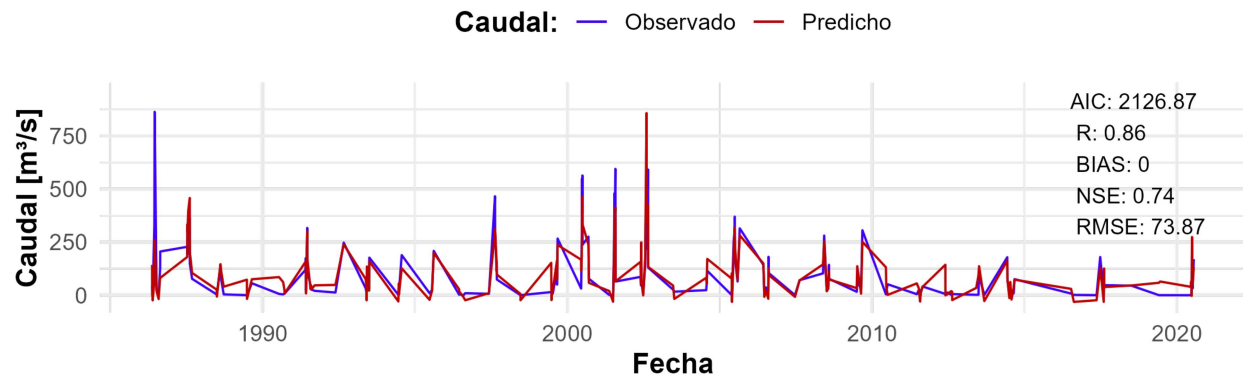
Referencias

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723. <https://doi.org/10.1109/TAC.1974.1100705>
- Bennett, J. C., Wang, Q., Robertson, D. E., Bridgart, R., Lerat, J., Li, M., & Michael, K. (2021). An error model for long-range ensemble forecasts of ephemeral rivers. *Advances in Water Resources*, 151, 103891. <https://www.sciencedirect.com/science/article/pii/S0309170821000464>
- GEF Humedales Costeros. (2021). Propuesta integral de protocolo de apertura mecanizada de la barra terminal para el humedal Cáhuil.
- Gobierno Regional Libertador Bernardo O´Higgins. (2023). Estudio Inundabilidad en el humedal de Cáhuil, Comuna de Pichilemu, Región de O´Higgins. *ANEXO J, ESTUDIO DE HIDROLOGÍA*.
- Hayati, R., Munawar, A. A., Lukitaningsih, E., Earlia, N., Karma, T., & Idroes, R. (2024). Combination of PCA with LDA and SVM classifiers: A model for determining the geographical origin of coconut in the coastal plantation, Aceh Province, Indonesia. *Case Studies in Chemical and Environmental Engineering*, 9, 100552. <https://www.sciencedirect.com/science/article/pii/S2666016423002578>
- Meng, F., Su, F., Yang, D., Tong, K., & Hao, Z. (2016). Impacts of recent climate change on the hydrology in the source region of the Yellow River basin. *Journal of Hydrology: Regional Studies*, 6, 66-81. <https://doi.org/https://doi.org/10.1016/j.ejrh.2016.03.003>
- Proyecto Simona Costa. (2023). Proyecto FONDEF IDeA I+D 2022 (ID22I10087). <https://simonacosta.cl/simona-costa/#tap-3>
- Romero-Cuellar, J., Arabzadeh, R., Craig, J. R., Tolson, B. A., & Mai, J. (2024). A multi-model evaluation of probabilistic streamflow predictions via residual error modelling. *Journal of Hydrology*, 635, 131152. <https://www.sciencedirect.com/science/article/pii/S002216942400547X>
- Uribe, H. (2016). Hidrología del secano de la región de O´Higgins. *Rengo: Boletín INIA - Instituto de Investigaciones Agropecuarias*, 320. <https://hdl.handle.net/20.500.14001/6550>

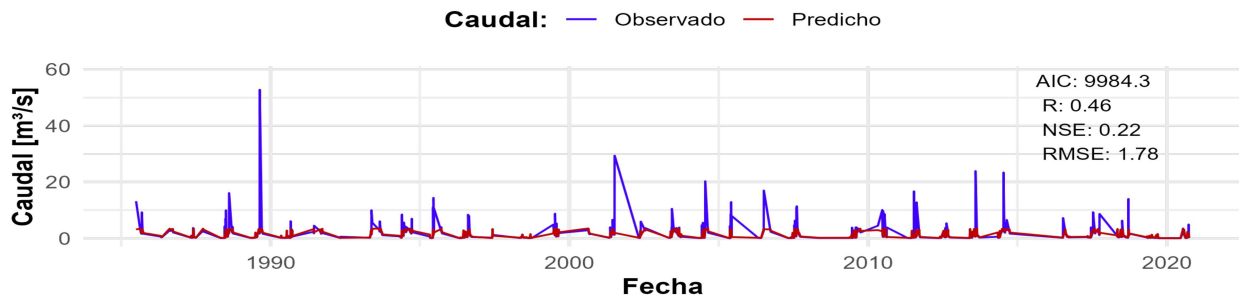
7. Anexos



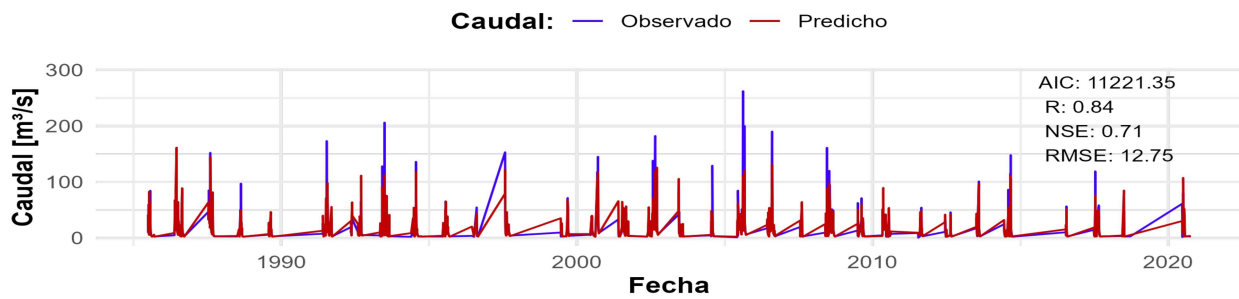
Anexo 1: Valores Predichos y Observados de Caudal de Ascenso, Modelo GAMLSS Distribución Normal.



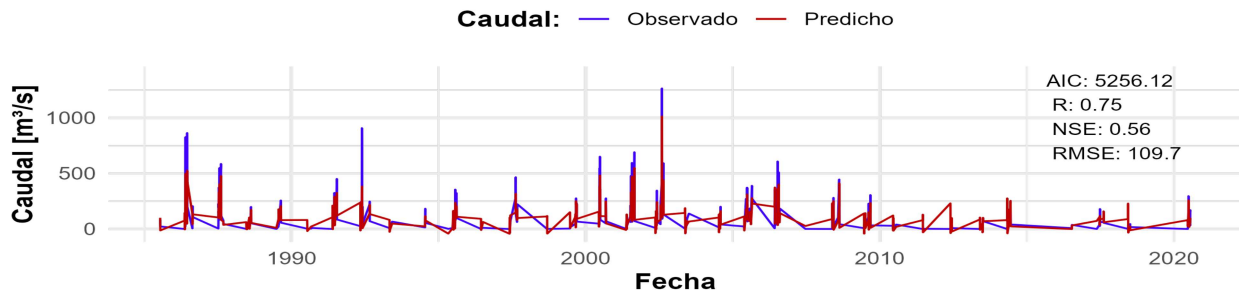
Anexo 2: Valores Predichos y Observados de Caudal de Descenso, Modelo GAMLSS Distribución Normal.



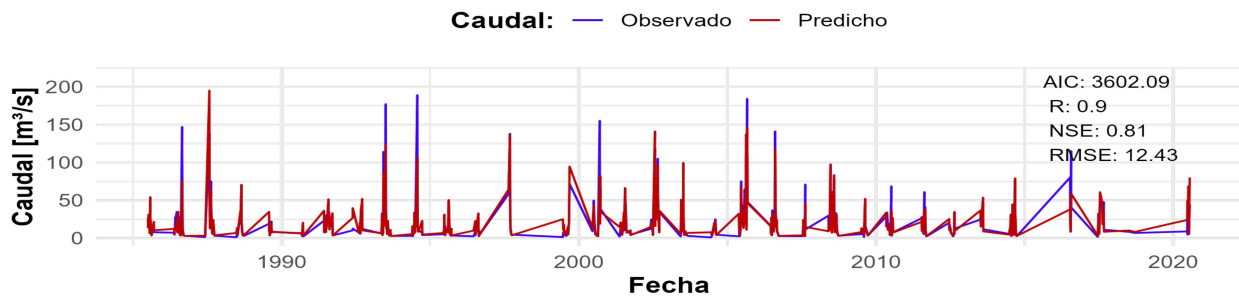
(a) Valores Observados y Predichos de Caudal Medio, GAMLSS Normal.



(b) Valores Observados y Predichos de Caudal de Ascenso, GAMLSS Normal.

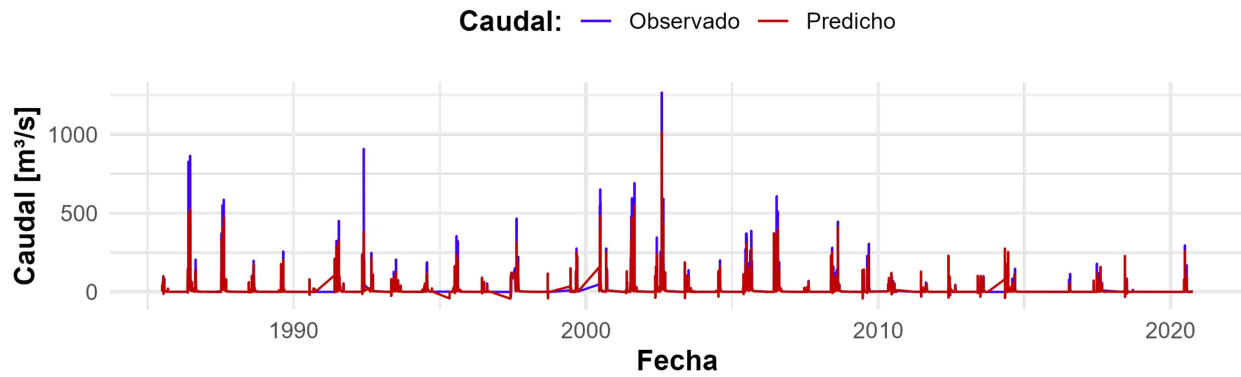


(c) Valores Observados y Predichos de Caudal Crecida, GAMLSS Normal.

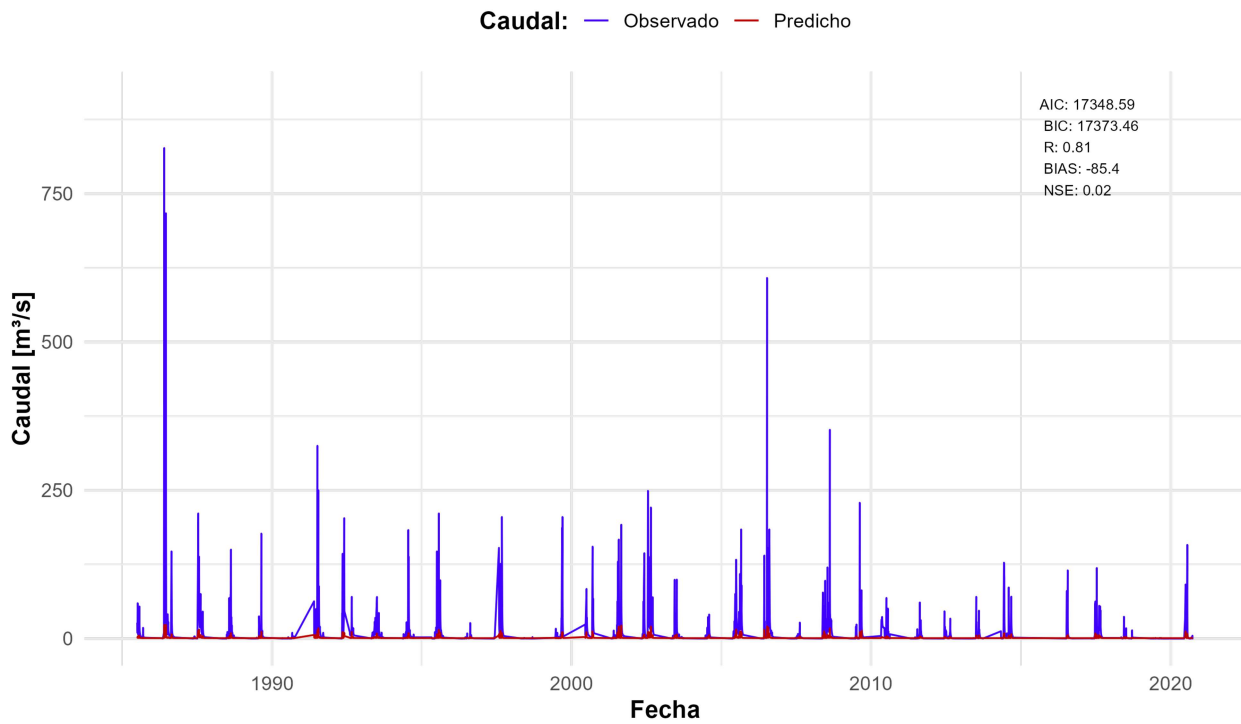


(d) Valores Observados y Predichos de Caudal Receso, GAMLSS Normal.

Anexo 3: Valores Predichos y Observados de Caudal, Modelo GAMLSS Distribución Normal con división en 4 etapas.



Anexo 4: Valores Predichos y Observados de Caudal, Validación Cruzada del Modelo GAMLSS Distribución Normal con división en 4 etapas.



Anexo 5: Valores Observados y Predichos de Caudal Medio, GAMLSS con Distribución Gamma.