



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Departamento de Obras Civiles

“EVALUACION DEL MODELO ESTADÍSTICO SARIMAX PARA EL PRONÓSTICO DE CAUDALES DE CRECIDA EN LA CUENCA DEL RIO MAIPO”

Memoria de Título presentada por

MARCELO IVÁN GUTIÉRREZ VILCHES

como requisito parcial para optar al título de la carrera de

Ingeniería Civil

PROFESOR GUÍA: MIGUEL LAGOS ZUÑIGA

CO-REFERENTE: ÁLVARO OSSANDÓN

Julio de 2025



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: Evaluación de un modelo estadístico SARIMAX para el pronóstico de caudales de crecida en la Cuenca Alta del Río Maipo.

Nombre del candidato(a): Marcelo Iván Gutiérrez Vilches

Carrera / Grado: Ingeniería Civil

Campus: Santiago San Joaquín ; **Departamento:** Obras Civiles

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Miguel Lagos Zúñiga, 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

☒ 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:

☐ 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: 21-07-2025

; Firma:

Estudiante o Candidato(a):

Fecha: 21-07-2025

; 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.

Resumen

Los estudios hidrológicos continuos en la Cuenca del Río Maipo son fundamentales no solo por ser una de las principales fuentes de agua potable para Santiago de Chile y su capacidad para generar energía, sino también debido a los graves daños ocasionados por las inundaciones en esta cuenca. Siendo inundación más reciente, la ocurrida durante junio del 2023 la cual tuvo severas consecuencias para las personas e infraestructura (Martínez et al., 2023).

Este trabajo propone la utilización de una modelación estadística basada en modelos estacionales autorregresivos integrados de medias móviles con variables exógenas (SARIMAX), con el objetivo principal de representar y pronosticar series temporales de crecidas en el caudal horario en la estación Río Maipo en El Manzano, dentro de la Cuenca del Río Maipo.

Se generaron modelos utilizando distintas combinaciones de predictores (Temperatura, Precipitación y Caudal Aguas Arriba) para representar el caudal horario en la cuenca. El estudio abarcó aproximadamente 15 años de datos de caudal, enfocándose en las crecidas desde el año 2010 en adelante. Adicionalmente, se analizó el comportamiento de los modelos para tiempos de adelanto de 1 a 6 horas, obteniendo resultados favorables en los pronósticos a corto plazo (con tiempos de adelanto de 1 a 3 horas), los resultados muestran un decrecimiento en la precisión a medida que aumentan los tiempos de adelanto, especialmente en las predicciones a 5 y 6 horas. A pesar de que la precisión disminuye con el tiempo, los resultados aún se mantienen dentro de un rango aceptable en distintas métricas de error. En particular, se obtienen para cada uno de los tiempos de adelanto, pronósticos con un valor del coeficiente de determinación R^2 mayor a un 0,85, lo que permite su uso en aplicaciones prácticas, como los sistemas de alerta temprana. Los resultados obtenidos en este estudio demuestran que los modelos SARIMAX son efectivos para el pronóstico de crecidas en la Cuenca del Río Maipo a corto plazo. A pesar de la disminución de precisión a medida que aumentan los tiempos de adelanto, los pronósticos siguen siendo útiles para aplicaciones prácticas, como los sistemas de alerta temprana.

En cuanto a las perspectivas futuras, es crucial continuar con la mejora y validación de estos modelos utilizando datos adicionales y con el ajuste de nuevos predictores, con el fin de aumentar la precisión de los pronósticos a plazos más largos. Además, el desarrollo de una herramienta de monitoreo en tiempo real podría ser clave para anticipar y mitigar los impactos de futuras crecidas, optimizando así la gestión de recursos y la protección de la infraestructura en la región.

Palabras claves

Cuenca Río Maipo; Modelos estadísticos; SARIMAX; Pronóstico de caudal; Caudal horario; Tiempos de adelanto; Lead Times.

Abstract

Continuous hydrological studies in the Maipo River Basin are essential not only because it is one of the main sources of water supply for Santiago, Chile, and a key source of energy generation, but also due to the significant damage caused by flooding in the basin. One of the most recent floods was the one that occurred in June 2023, which had severe consequences for both people and infrastructure (Martínez et al., 2023). This study proposes the use of a statistical model based on Seasonal Autoregressive Integrated Moving Average (SARIMAX) models with exogenous variables, aiming to represent and forecast hourly flow time series at the Río Maipo station in El Manzano, within the Maipo River Basin. Models were generated using different combinations of predictors (Temperature, Precipitation, and Upstream Flow) to represent the hourly flow in the basin. The study covered approximately 15 years of flow data, focusing on floods from 2010 onward. Additionally, the behavior of the models was analyzed for lead times from 1 to 6 hours, with favorable results in short-term forecasts (with lead times of 1 to 3 hours). The results show a decrease in accuracy as lead times increase, especially in predictions at 5 and 6 hours. Although accuracy decreases over time, the results still remain within an acceptable range across various error metrics. In particular, for each lead time, forecasts with a coefficient of determination (R^2) greater than 0.85 were obtained, making them suitable for practical applications such as early warning systems. The results obtained in this study demonstrate that SARIMAX

models are effective for flood forecasting in the Maipo River Basin in the short term. Despite the decrease in accuracy as lead times increase, the forecasts remain useful for practical applications, such as early warning systems. As for future perspectives, it is crucial to continue improving and validating these models using additional data and adjusting new predictors in order to increase the accuracy of forecasts for longer time horizons. Furthermore, the development of a real-time monitoring tool could be key to anticipating and mitigating the impacts of future floods, thus optimizing resource management and protecting infrastructure in the region.

Keywords

Maipo River Basin; Statistical Models; SARIMAX; Flow Forecasting; Hourly Flow; Lead Times.

Tabla de Contenido

1	Introducción.....	6
2	Objetivos	8
2.1	Objetivo General.....	8
2.2	Objetivos específicos	8
3	Marco Teórico	9
3.1	Sistema de Alerta Temprana (SAT).....	9
3.2	Problemas en el pronóstico	11
3.3	Pronóstico de caudales de crecida.....	11
3.4	Tiempos de adelanto (Lead Times).....	15
4	Zona de estudio	17
5	Metodología.....	23
5.1	Base de datos.....	23
5.2	Evaluación de la Memoria Hidrológica del Río Maipo mediante ACF y PACF	29
5.3	Combinación de predictores	29
5.4	Validación Modelo Optimo	32
5.5	Tiempos de adelanto	32
5.6	Validación Manual Tiempos de adelanto.....	32
6	Análisis y Resultados	33
6.1	Función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF)	33
6.2	Combinación Precipitación EP + Temperatura EP	35
6.3	Combinaciones de predictores	39
6.3.1	Resultados Coeficiente de determinación (R ²):	40
6.3.2	Resultados Raíz del Error Cuadrático Medio (RMSE):.....	41
6.3.3	Resultados Error Absoluto Porcentual Medio (MAPE):	43
6.4	Validación Modelo.....	45
6.5	Tiempos de Adelanto	47
6.5.1	Validación Lead Times 1:.....	50
6.5.2	Validación Lead Times 3:.....	52
6.5.3	Validación Lead Times 5:.....	54
7	Discusión.....	57
8	Conclusiones y recomendaciones.....	59
9	Referencias.....	60
10	Anexos.....	64

1 Introducción

Una inundación corresponde a un rápido ascenso del nivel del agua, generando caudales inusuales que cubren o llenan superficies de terreno que normalmente son secas (SENAPRED, n.d.). Este fenómeno es provocado por el desborde de un río a causa de lluvias, tormentas tropicales, huracanes, y algunas veces por las acciones del ser humano (Oficina de las Naciones Unidas para la Reducción del Riesgo de Desastres (UNDRR), 2020).

Las inundaciones son peligros naturales que causan impactos ambientales y socioeconómicos negativos a nivel local y regional. Además, las vidas humanas, las instalaciones y la infraestructura se ven frecuentemente amenazadas por este evento natural (Kalogeropoulos et al., 2023; Ortega et al., 2014; Stillwell, 1992). En el mundo, los peligros de lluvias intensas son muy recurrentes y crecientes, provocando severos daños a la infraestructura e incluso provocando muertes cada año (Cabral et al., 2023; Roldán et al., 2022)

En los últimos años, los desastres relacionados con el clima se están intensificando y son cada vez más frecuentes, especialmente las inundaciones y las tormentas (United Nations Office for Disaster Risk Reduction (UNDRR), 2021). Durante 2021 se registraron 201 grandes inundaciones a nivel mundial, un 23% más que el promedio anual (163 eventos), los cuales provocaron 6.171 muertes, un 18% más que la media anual de 5.233 (United Nations Office for Disaster Risk Reduction (UNDRR), 2021). Lo anterior, subraya la importancia de contar con herramientas de alerta temprana y predicción que proporcionen información con suficiente antelación para reducir los riesgos asociados (ALNAP, 2008).

La predicción y el manejo de estos eventos extremos son fundamentales para mitigar sus impactos, especialmente en áreas donde los asentamientos humanos se encuentran en zonas vulnerables (Ministerio del Medio Ambiente, 2021). En Chile, estos eventos han dejado consecuencias negativas, las cuales van desde daños en infraestructura hasta personas damnificadas y pérdidas de vidas humanas (BBC News Mundo, 2023; EL PAÍS Chile, 2024). Un ejemplo de aquello ocurrió en las inundaciones de junio del 2023, las cuales generaron 27.588 damnificados, 140 albergados, 1.724 personas aisladas y 3 fallecidos a nivel nacional. Puntualmente, en la región Metropolitana, se registraron 1.048 damnificados, 2 personas albergadas y 1 fallecido. Por otro lado, el daño estructural fue de alto impacto con cerca de 5.900 viviendas afectadas dentro del territorio chileno y 130 en la región Metropolitana. (Martínez et al., 2023)

La Cuenca del Río Maipo constituye la fuente principal de agua potable de la Región Metropolitana, aportando cerca del 70% de este consumo. Además, proporciona alrededor del 90% de agua para el regadío. Por otra parte, el Río Maipo alimenta centrales hidroeléctricas presentes en la cuenca (R. García, 2008). A lo largo de los años, han ocurrido crecidas de caudal en el Río Maipo que han ocasionado inundaciones. El registro más antiguo sobre daños ocasionados por el Río Maipo en esta área corresponde al 23 de agosto de 1953. Ese día, el Maipo destruyó totalmente las defensas fluviales próximas a la localidad de Valdivia de Paine (La Nación, 1953). Un evento más reciente es el que ocurrió el 23 de junio de 2023, para el cual se analizó el caudal registrado en las estaciones El Río Volcán en Queltehues y para la estación Río Colorado antes Junta Río Maipo, el cual tuvo un aumento de 106 a 514,8 $[m^3/s]$ en 24 horas lo que generó inundaciones en distintos puntos del Río Maipo. (Martínez et al., 2023).

Durante las inundaciones de junio de 2023 en el Río Maipo, el momento crítico de la emergencia ocurrió el viernes 23 de junio a las 12:30 a.m., cuando la estación de monitoreo de la Dirección General de Aguas (DGA) se vio afectada por la crecida y dejó de funcionar. Ante esta situación, se implementó un monitoreo manual llevado a cabo por los agricultores de la zona. Basándose en su experiencia, ellos estimaron de forma aproximada un margen de 24 horas entre el desborde del río en la parte alta y el eventual desborde río abajo. (Martínez et al., 2023)

La gestión y la normativa vigente en torno al monitoreo y pronóstico de inundaciones aún no se han consolidado, presentando diversas debilidades que deben ser abordadas. Predomina un enfoque reactivo, con medidas implementadas principalmente durante la fase de emergencia. Esto resalta la necesidad de avanzar hacia un modelo de alerta temprana que, a partir de datos proporcionados por sistemas de monitoreo y un

conocimiento profundo del territorio, permita predecir con mayor precisión tanto la ocurrencia como la magnitud y las áreas afectadas por las inundaciones (Oberli et al., 2021).

Las inundaciones catastróficas en Valencia, España en octubre de 2024, provocadas por una Depresión Aislada en Niveles Altos (DANA), pusieron de manifiesto la urgencia de contar con sistemas de alerta temprana más eficaces. Las lluvias torrenciales causaron desbordamientos de ríos, resultando en una tragedia con 224 muertes, 3 personas desaparecidas y pérdidas materiales superiores a los 10 mil millones de euros. En este contexto, un sistema de alerta eficiente hubiera permitido a las autoridades evacuar a tiempo las zonas de mayor riesgo, minimizando las consecuencias humanas y materiales (E. García, 2025; Olcina-Cantos et al., 2025).

Las inundaciones catastróficas ocurridas en Río Grande do Sul, Brasil en 2024 evidencian la urgente necesidad de implementar sistemas de alerta temprana más eficaces para mitigar los impactos de los fenómenos climáticos extremos. A pesar de las predicciones meteorológicas previas, la falta de una respuesta coordinada y la insuficiencia en la infraestructura de prevención exacerbó la tragedia, afectando a más de 2,3 millones de personas, 173 muertes y más de 206 mil propiedades rurales fueron perjudicadas, con pérdidas en producción e infraestructura (OCHA, 2024; Palummo Lantes, 2025).

Un claro ejemplo de éxito de un modelo de alerta temprana en inundaciones es el sistema implementado en Bangladesh. Este país, uno de los más vulnerables a inundaciones, ha logrado reducir significativamente los daños mediante la implementación de un sistema de alerta temprana que se basa en la monitorización de los niveles de agua y la difusión de alertas a las comunidades en riesgo. Según el estudio de Hasan e Islam (Hasan, 2024), las autoridades pueden emitir alertas con antelación, lo que permite a los residentes tomar medidas preventivas, como evacuar zonas de alto riesgo. Este enfoque ha sido tan exitoso que ha sido replicado en varios países de la región, mejorando la preparación ante desastres y reduciendo las pérdidas humanas y materiales.

El presente trabajo busca seleccionar un modelo estocástico SARIMA con el propósito de representar el caudal horario de la cuenca Río Maipo en la estación “Río Maipo en el Manzano”. Además, se evalúa el modelo para distintos tiempos de adelanto, lo cual puede contribuir en elaborar un sistema de alerta temprana en la Cuenca. Con respecto al desarrollo de este trabajo, se presenta en 3 acápites principales. En primer lugar, se desarrolla el marco teórico que abarca los principales conceptos hidrológicos asociados al tema de memoria. Luego, se detalla la metodología para determinar el modelo SARIMA óptimo en base a distintas métricas de error. Posteriormente, se analizan los resultados para concluir cual es el modelo que mejor representa el comportamiento de la cuenca.

2 Objetivos

2.1 Objetivo General

Entrenar un modelo estadístico del tipo SARIMAX para distintos tiempos de adelanto y variables predictoras, para testear su aplicabilidad en el pronóstico de crecidas en la cuenca del Rio Maipo en el Manzano.

2.2 Objetivos específicos

1. Determinar cuáles son los predictores hidroclimáticos que minimizan el error en el pronóstico de caudal, utilizando un conjunto de crecidas históricas para probar la precisión de los modelos con distintas combinaciones de variables (precipitación, temperatura, caudal).
2. Determinar los errores asociados en el pronóstico de caudal para distintos tiempos de adelanto, considerando los predictores hidroclimáticos con menor error, utilizando series de crecidas históricas y analizando los resultados para distintos periodos de tiempo de adelanto.
3. Realizar recomendaciones sobre la estructura del modelo y los predictores que garanticen errores aceptables, maximizando el tiempo de adelanto posible, basándose en los resultados obtenidos al probar distintos predictores y crecidas históricas.

3 Marco Teórico

En este capítulo se presentan los fundamentos teóricos necesarios para comprender el desarrollo de este documento y el cumplimiento de los Objetivos planteados. Se abordan conceptos clave como la definición y función de los Sistema de Alerta Temprana (SAT), los principios del pronóstico de caudal, particularmente utilizando el modelo estadístico SARIMAX, y la importancia de los tiempos de adelanto (lead times) en la predicción de eventos extremos. Además, se describe el proceso de selección de eventos de crecida históricos utilizados como base para el modelado. Estos elementos proporcionan el sustento conceptual y metodológico para la construcción del modelo predictivo que se propone en esta investigación.

3.1 Sistema de Alerta Temprana (SAT)

Chile, debido a su situación geográfica particular, está expuesto a una variedad de amenazas naturales recurrentes, como terremotos, tsunamis, erupciones volcánicas, sequías, crecidas e inundaciones (Bustos, 2011; Ministerio del Medio Ambiente, 2021). Se tiene registro de 254 eventos de tipo hidrometeorológico ocurridos en Chile entre los años 1902 y 2019, siendo las precipitaciones el factor desencadenante. En particular, se observa un aumento en la cantidad de eventos documentados en años recientes, como 2015 y 2017, lo cual se atribuye tanto a mejores sistemas de monitoreo como al efecto del cambio climático en la intensificación de las precipitaciones extremas (Ministerio del Medio Ambiente, 2021; Oberli et al., 2021b; Ríos Insua et al., 2013).

Un sistema de alerta temprana es una herramienta utilizada para generar advertencias anticipadas sobre eventos climáticos extremos (Pascal, 2023). El Centro de Alerta Temprana (CAT) de ONEMI gestiona el tráfico de datos relacionados con el Sistema Nacional de Protección Civil y monitorea las variables de interés para las distintas amenazas que pueden derivar en desastres en el país. A partir de este monitoreo, es posible emitir alertas o estados de vigilancia, definidos en cuatro niveles: Alerta Verde, Alerta Temprana Preventiva, Alerta Amarilla y Alerta Roja. Estos estados facilitan la administración y movilización de recursos esenciales para reducir el riesgo de pérdida de vidas y minimizar los daños a la población (Oberli et al., 2021b).

Por lo tanto, un Sistema de Alerta Temprana (SAT) para crecidas de caudales es fundamental en la gestión de emergencias naturales, ya que permite una respuesta anticipada y efectiva para proteger a la población y mitigar los riesgos asociados. Uno de los propósitos de este trabajo, es poder proporcionar un modelo que ayude a generar un SAT para el Río Maipo.

A continuación, se presenta un diagrama de flujo en la Figura 3-1 que resume el procedimiento general propuesto para el funcionamiento de un CECOSEM (Centro de Coordinación del Sistema de Emergencia) dentro de un Sistema de Alerta Temprana (SAT). Este esquema permite visualizar las etapas clave en la generación y gestión de alertas, desde el análisis de datos históricos y en tiempo real, hasta la verificación de umbrales y la posterior decisión sobre el nivel de alerta a difundir. La figura ilustra de manera clara cómo se activa la respuesta institucional según el nivel de amenaza identificado, destacando el rol central del monitoreo permanente y la toma de decisiones informada para la protección de la población.

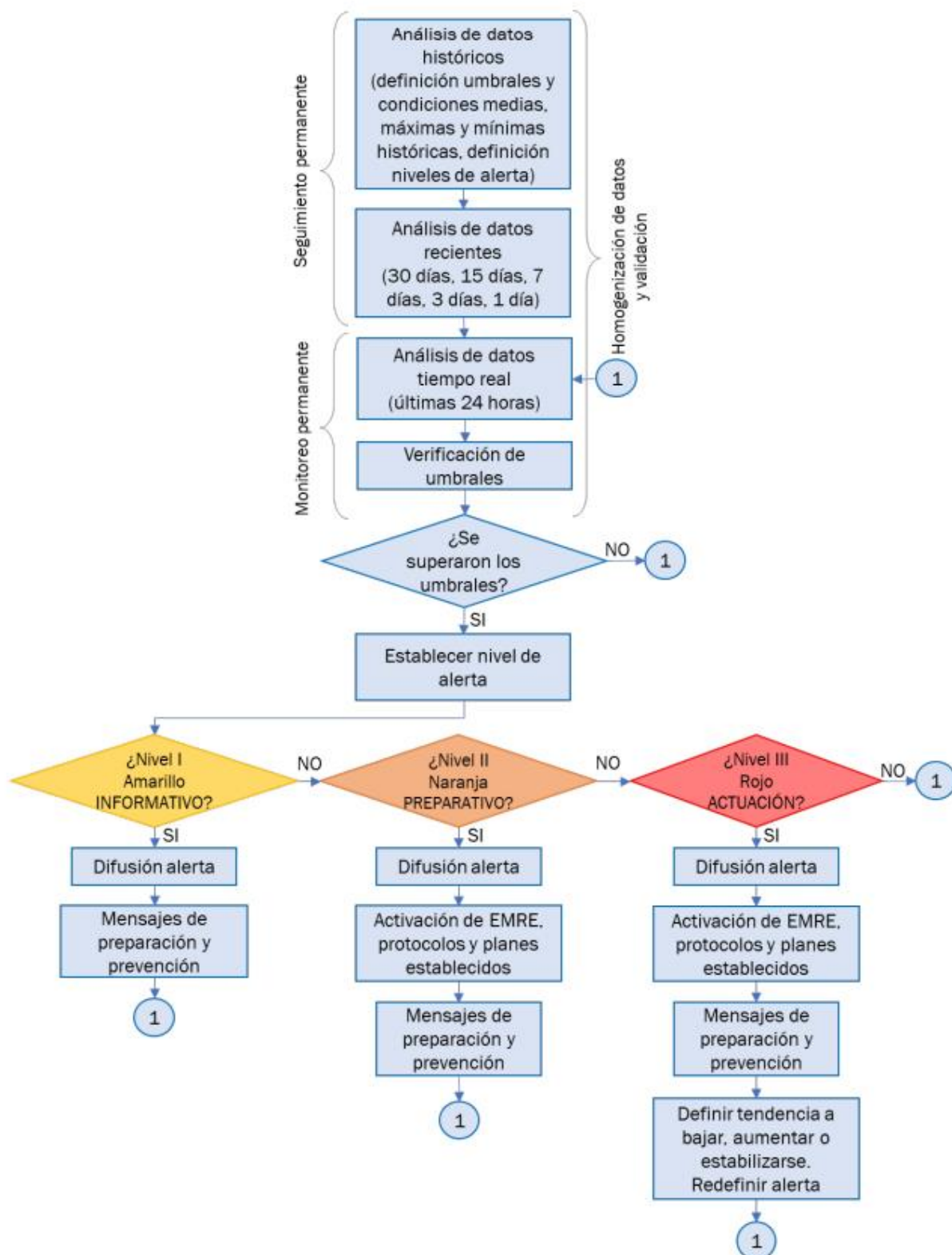


Figura 3-1: Ejemplo de procedimiento general de funcionamiento de un CECOSEM. Fuente: (Unidad Nacional para la Gestión del Riesgo de Desastres (UNGRD), 2021)

3.2 Problemas en el pronóstico

El pronóstico es una tarea fundamental en una variedad de campos, desde la meteorología hasta la economía, y se enfrenta a retos únicos según la escala temporal en cuestión. Los modelos de pronóstico deben adaptarse a las características particulares de los datos de cada escala, ya que los patrones y tendencias pueden variar significativamente a medida que se cambian las unidades de tiempo analizadas.

- Pronóstico Estacionario: El pronóstico estacionario de caudales se basa en la suposición de que las series temporales de caudales siguen un comportamiento estable a lo largo del tiempo, sin grandes variaciones estacionales o tendencias. Sin embargo, en la gestión de recursos hídricos, los caudales presentan fluctuaciones estacionales debido a factores como las lluvias o los deshielos, lo que exige modelos capaces de capturar estas variaciones. Aunque los modelos tradicionales como ARIMA (Auto-Regressive Integrated Moving Average) son eficaces para predecir caudales en series con componentes lineales y estacionales, enfrentan limitaciones cuando los datos muestran patrones no lineales o comportamientos complejos, como en eventos extremos de crecidas o sequías. Para abordar estos problemas, se emplean enfoques como SARIMA (Seasonal ARIMA) para modelar la estacionalidad de los caudales, o modelos híbridos que combinan ARIMA con técnicas de aprendizaje automático, como LSTM (Long Short-Term Memory), que permiten manejar tanto las tendencias lineales como las relaciones no lineales en los datos de caudal, mejorando la precisión de los pronósticos en contextos más complejos (Khedhiri, 2022; Kumar Dubey et al., 2021).
- Pronóstico Sub-Estacional: A escalas más cortas, como en el caso de pronósticos sub-estacionales, los modelos deben adaptarse a comportamientos más finos y detallados, como variaciones de corto plazo en fenómenos climáticos. Estos pronósticos requieren una mayor flexibilidad para capturar cambios más rápidos y dinámicos.
- Pronóstico de Crecidas: la predicción de eventos extremos, como inundaciones o aumentos repentinos en el caudal de un río, presenta retos significativos debido a la variabilidad impredecible de los eventos climáticos y la interacción de múltiples factores. Estos fenómenos a menudo no siguen patrones estrictamente lineales, lo que requiere el uso de técnicas más complejas que puedan modelar tanto las tendencias subyacentes como los eventos repentinos o no lineales.

3.3 Pronóstico de caudales de crecida

Pronosticar significa prever o anticipar un evento futuro basándose en datos, observaciones o análisis. Los pronósticos basados en datos utilizan variables ya registradas como precipitación, caudal, temperatura, tiempos de adelanto, entre otros (Bustos, 2011; Ministerio del Medio Ambiente, 2021; Ríos Insua et al., 2013).

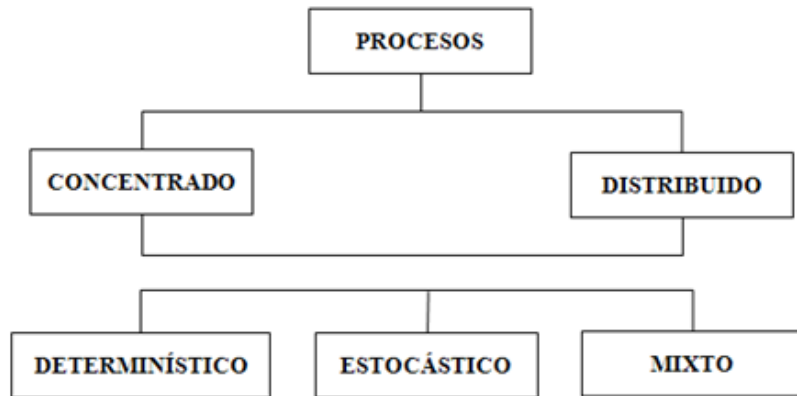
Para llevar a cabo un pronóstico, se pueden utilizar modelos del tipo dinámicos (modelos hidrológicos), basado en datos (estadísticos) y/o híbridos (Chavasse et al., 1997; Court, 2008; Lee et al., 2016; Meis et al., 2017; Pascal, 2023; Suescún, 2019; Sun et al., 2020). A continuación, se describen brevemente las características de cada uno de estos enfoques.

- Dinámicos:

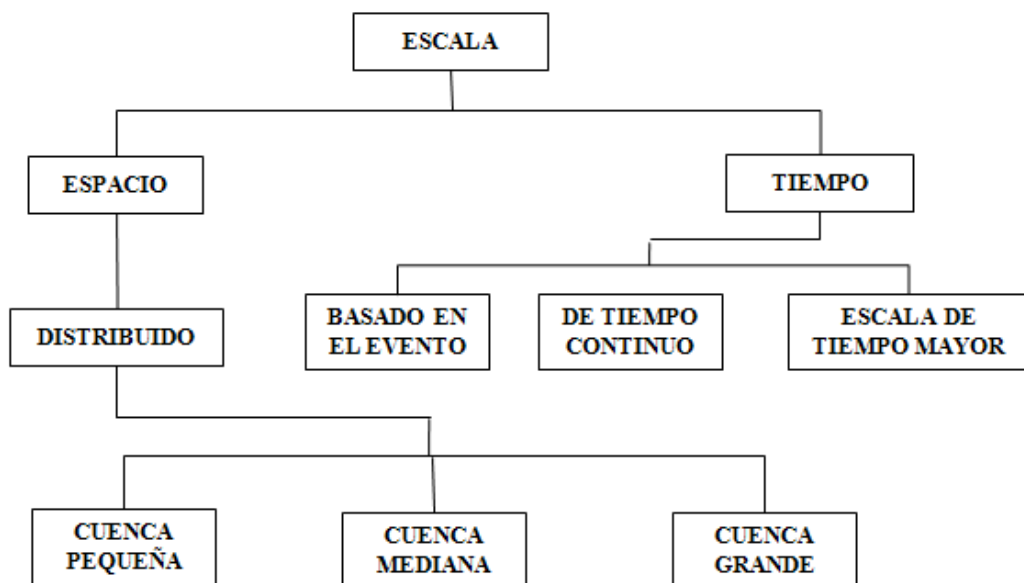
Los modelos hidrológicos constituyen una representación de un sistema (en este caso una cuenca hidrográfica) con el fin de reproducir sus respuestas frente a diferentes entradas: precipitaciones extremas, modificaciones en el uso del suelo, cambio climático, etc. Todos los modelos hidrológicos son aproximaciones de la realidad, por lo que resulta muy difícil conseguir la consideración simultánea de las cinco fuentes de variación: aleatoriedad, tres dimensiones espaciales y tiempo (Mendoza, 2010).

Como se muestra en la Figura 3-2 los modelos de simulación hidrológica pueden ser clasificados de acuerdo con tres criterios (Singh, 1995): Descripción de procesos, Escala y Técnica de solución.

(a) Clasificación de modelos en base a la descripción de procesos



(b) Clasificación de modelos en base a la escala espacio - temporal



(c) Clasificación de modelos en base a la técnica de resolución

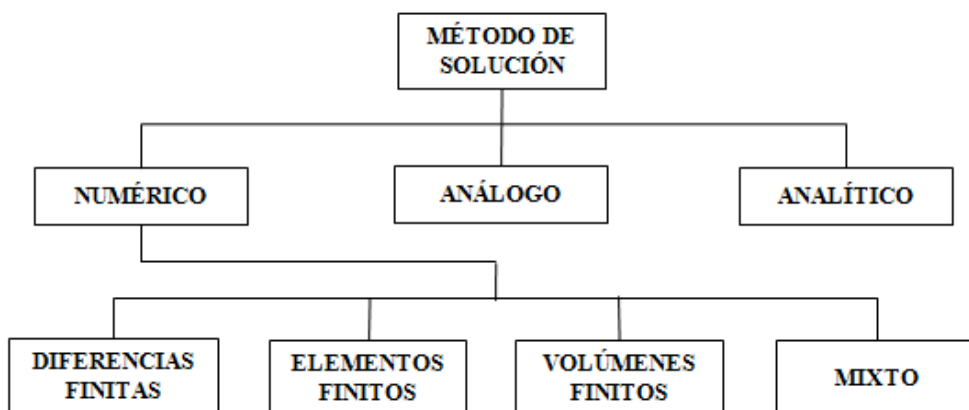


Figura 3-2: Clasificación de modelos hidrológicos según descripción de procesos, escala espacio-temporal y técnica de resolución. Modificado de (Singh, 1995).

En la literatura se pueden encontrar una gran cantidad de modelos hidrológicos, cuya aplicabilidad dependerá del régimen de escurrimiento imperante, el objetivo de la modelación y de la información disponible.

Los modelos hidrológicos tienen un papel fundamental en el desarrollo de sistemas de alerta de crecidas. Los estudios para pronosticar la ocurrencia de crecidas en cuencas chilenas contemplan principalmente modelos de función de transferencia (Pacheco, 2004; Sandoval, 1988; Vargas et al., 1988) y modelos de redes neuronales artificiales (González, 2003). Dichos modelos comparten dos características esenciales: son espacialmente agregados y tienen una base conceptual matemática.

Sin embargo, a pesar de que dichos modelos no requieren de grandes cantidades de información además de datos de precipitación, caudales y ocasionalmente temperatura, presentan como carencia importante la poca o nula representatividad de los procesos hidrológicos que ocurren en las cuencas estudiadas. De manera adicional, el carácter agregado de las modelaciones, así como la falta de datos hidrometeorológicos en distintos puntos de la cuenca, resulta un impedimento para investigar cómo se distribuyen en ella los distintos flujos involucrados en el ciclo hidrológico (Mendoza, 2010)

- Estadísticos:

Dentro de los modelos estadísticos, se encuentra el modelo SARIMAX el cual se basa en datos y es utilizado en esta memoria. El modelo SARIMAX (modelos estacionales autorregresivos integrados de media móviles) corresponde a una extensión de los modelos ARIMA (modelos autorregresivos integrados de media móviles) incluyendo la estacionalidad presente en los datos de la serie temporal, en este caso, del caudal y variables externas. El modelo SARIMAX queda definido por los siguientes parámetros:

$$(p, d, q) (P, D, Q, s) \text{ con } X_t$$

Donde:

- p: Coeficiente autorregresivo (AR).
- d: Diferenciación.
- q: Coeficiente de medias móviles (MA).
- P: Coeficiente autorregresivo estacional (AR_s)
- D: Diferenciación estacional.
- Q: Coeficiente de medias móviles estacional (MA_s)
- s: Estacionalidad.
- X_t : Representa las variables exógenas que se utilizan en el modelo para ayudar en la predicción.

El coeficiente autorregresivo (ya sea p o P) representa el número de valores anteriores que se utilizan como predictores en el modelo. La diferenciación (ya sea d o D), indica cuantas veces se diferencian los datos para hacer la serie estacionaria. A su vez, el coeficiente de medias móviles (ya sea q o Q), representa el número de errores pasados que se utilizan en la predicción. Por último, la estacionalidad representa el periodo estacional de la serie temporal.

El modelo SARIMAX cuenta con una ecuación que lo define, la cual se presenta a continuación:

$$\phi_{is}(B_s) \cdot \phi_i(B) \cdot (1 - B_s)^D \cdot (1 - B)^d \cdot Z_t = \theta_{qs}(B_s) \cdot \theta_q(B) \cdot u_t + \beta \cdot X_t \quad (1)$$

Donde:

- B: operador de retardo. $B \cdot Z_t = Z_{t-1}$

- B_s : operador de rezago estacional. $B_s \cdot Z_t = Z_{t-s}$
- $(1 - B)^d$: Diferenciación no estacional la cual elimina tendencias no estacionarias en la serie de tiempo.
- $(1 - B_s)^D$: Diferenciación estacional la cual elimina tendencias no estacionarias en la serie de tiempo.
- $\phi_i(B)$: Polinomio autorregresivo no estacional de orden p.

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p ; 0 \leq i \leq p \quad (2)$$

- $\phi_{is}(B_s)$: Polinomio autorregresivo estacional de orden P.

$$\phi_p(B_s) = 1 - \phi_{1,s} B_s - \phi_{2,s} B_s^2 - \dots - \phi_{p,s} B_s^p ; 0 \leq i \leq P \quad (3)$$

- $\theta_i(B)$: Polinomio de media móvil no estacional de orden q.

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q ; 0 \leq i \leq q \quad (4)$$

- $\theta_i(B)$: Polinomio de media móvil estacional de orden Q.

$$\theta_Q(B_s) = 1 - \theta_{1,s} B_s - \theta_{2,s} B_s^2 - \dots - \theta_{Q,s} B_s^Q ; 0 \leq i \leq Q \quad (5)$$

- Z_t : Caudal pronosticado en el tiempo t.
- u_t : Error residual o ruido blanco.
- β : es el coeficiente asociado a la variable exógena X_t .
- X_t : es la variable exógena que se incorpora en el modelo como predictor externo.

El modelo SARIMA ha demostrado ser eficaz para pronosticar caudales mensuales en estudios como el realizado en la cuenca del Río Cuyaguaje, Cuba, donde se destacó por su capacidad para pronósticos a largo plazo, especialmente durante la temporada lluviosa (Alonso et al., 2021). Similarmente, en la baja Cuenca del Plata, un estudio utilizó SARIMA para predecir el caudal mensual en varias estaciones de los ríos Paraná y Uruguay, obteniendo buenos resultados a pesar de algunas desviaciones en fenómenos extremos (Meis et al., 2017). En el caso del río Magdalena, un estudio realizado en la estación Limnigráfica de El Banco, Colombia, mostró que el modelo SARIMA (0, 1, 2, 0, 1, 1) logro representar de buena manera los caudales diarios observados (Gutierrez et al., n.d.).

• Híbridos:

Los modelos híbridos suelen mezclar métodos estadísticos como ARIMA (Auto-Regressive Integrated Moving Average) y modelos de aprendizaje automático como redes neuronales (por ejemplo, LSTM o ANN). ARIMA es efectivo para modelar series temporales con componentes lineales, pero tiene limitaciones cuando los datos presentan patrones no lineales. Por otro lado, las redes neuronales, como LSTM (Long Short-Term Memory), son capaces de modelar relaciones no lineales, pero pueden no ser tan eficaces cuando se trata de datos con estacionalidad o tendencia lineal (Peirano et al., 2021; Sina et al., 2023; Sun et al., 2020).

La combinación de estos modelos, por ejemplo, ARIMA-LSTM o SARIMA-LSTM, busca utilizar lo mejor de ambos enfoques: la capacidad de los modelos estadísticos para manejar datos lineales y la habilidad de las redes neuronales para modelar relaciones complejas y no lineales (Peirano et al., 2021; Sina et al., 2023; Sun et al., 2020).

3.4 Tiempos de adelanto (Lead Times)

El tiempo de adelanto en el pronóstico de caudales se refiere al período de anticipación con el que es posible prever el caudal de un río o cuerpo de agua. A mayor tiempo de adelanto en el pronóstico, mayor será la capacidad de respuesta ante un evento hidrológico.

Sin embargo, una de las principales dificultades en este tipo de modelos es la capacidad de predecir con precisión múltiples pasos temporales hacia adelante. A medida que aumenta el lead time, los errores en el pronóstico tienden a incrementarse debido a la incertidumbre en las condiciones meteorológicas y otros factores que afectan la hidrología de la cuenca.

En este estudio, se desarrollarán modelos de pronóstico con horizontes de predicción que van desde 1 hasta 6 horas de adelanto. Se espera que, preliminarmente, el pronóstico con un lead time de 6 horas presente una mayor incertidumbre y error en comparación con el pronóstico de 1 hora de adelanto.

El uso de modelos para el pronóstico de caudales con diferentes tiempos de adelanto ha sido ampliamente estudiado, tanto en enfoques hidrológicos como estadísticos. En el caso de un modelo hidrológico aplicado en el Valle del Alto Cachapoal, se utilizó redes neuronales recurrentes (LSTM) para prever los caudales con adelantos de 1, 3 y 5 horas. Los resultados mostraron que el pronóstico con un adelanto de una hora fue altamente preciso, pero la precisión disminuyó a medida que aumentaba el tiempo de adelanto (Pascal, 2023). Por otro lado, en la Cuenca del Plata, se implementó un modelo SARIMA para pronosticar el caudal mensual en varias estaciones, con pronósticos de hasta 32 meses. Aunque los modelos SARIMA mostraron un buen rendimiento en algunos subperíodos, los resultados a largo plazo fueron menos precisos debido a la estacionalidad y eventos climáticos extremos (Meis et al., 2017).

Otra de las grandes dificultades es que los modelos solamente tendrán salidas de Caudal, pero no de la variable predictora, por lo que para resolver esta falta de inputs de predictores se implementa la siguiente red para generar los tiempos de adelanto:

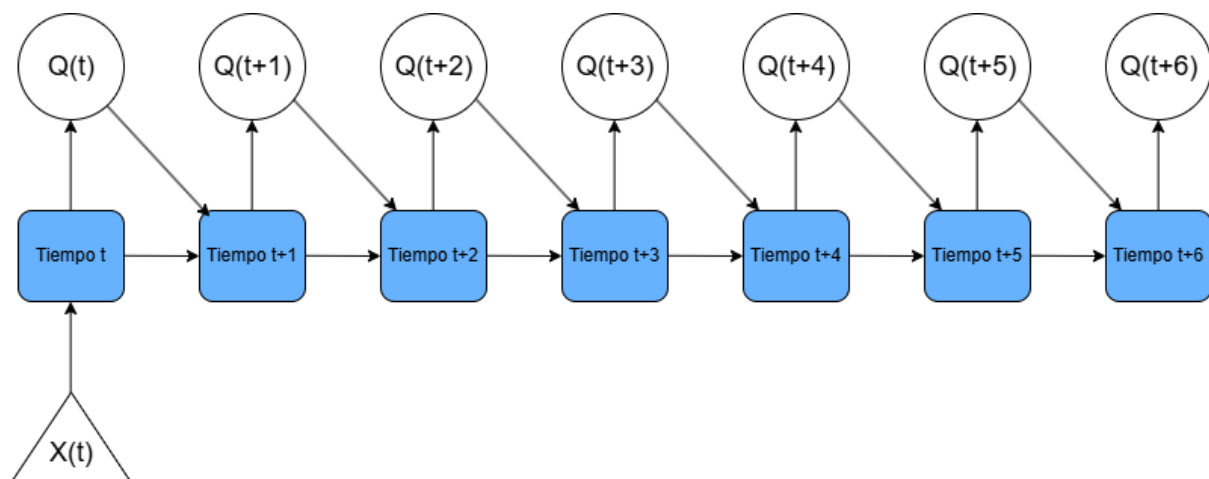


Figura 3-3: Red para generar tiempos de adelanto.

En la Figura 3-3 se pueden identificar tres secciones principales:

1. Sección rectangular (azul): Representa los distintos instantes de tiempo en la serie temporal, desde el tiempo actual (t) hasta los tiempos futuros ($t+1$, $t+2$, ..., $t+6$).
2. Sección circular (blanca): Representa el caudal pronosticado (Q) para cada uno de los tiempos de adelanto. Cada nodo de esta sección está vinculado a un instante de tiempo específico.

3. Sección triangular (blanca): Representa la(s) variable(s) predictora(s) utilizadas como entrada del modelo ($X(t)$). Notar que esta variable también podría ir al pasado, la cual sería considerada como $X(t-1)$.

4 Zona de estudio

La investigación se desarrolla en la Cuenca del Río Maipo, la cual está ubicada en la zona central de Chile, entre las latitudes $32^{\circ}55'$ – $34^{\circ}15'$ y las coordenadas $69^{\circ}46'$ - $71^{\circ}43'$ estando contenida prácticamente en su totalidad en la región metropolitana, y parte de la quinta y sexta región. (Gobierno de Chile, 2004; Videla Giering, 2013).

El Río Maipo es el principal cuerpo hídrico de la cuenca, nace en las laderas del volcán Maipo, a una altitud aproximada de 5.623 msnm, y luego de un recorrido de 250 km a través de montañas y llanos, entre comunidades humanas y ecosistemas, desemboca en el Océano Pacífico en las inmediaciones de la localidad de Lolleo, al sur del puerto de San Antonio. Además, es la principal fuente de agua de la Región Metropolitana ya que proporciona alrededor del 70% de la demanda actual de agua potable y cerca de un 90% de las demandas de regadío. Además, otro uso importante de este río es la generación de energía hidroeléctrica (Gobierno de Chile, 2004; Videla Giering, 2013). A continuación, se presenta el mapa de la Cuenca del Río Maipo, en donde se destaca toda su extensión junto con la Zona Alta de la Cuenca, la cual corresponde a la zona de estudio de este documento.

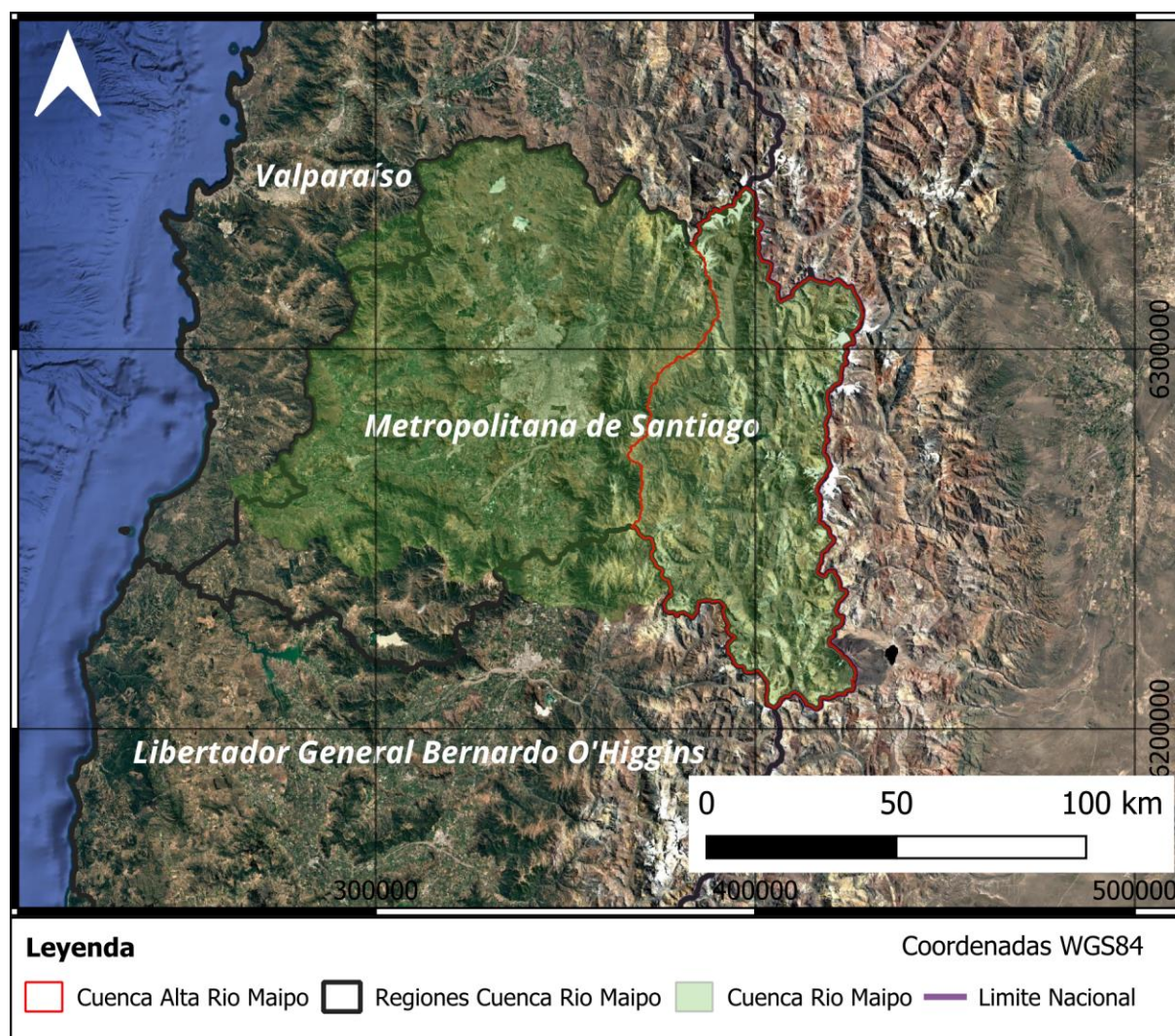


Figura 4-1: Cuenca del Río Maipo: Divisiones Regionales y zona de estudio.

Dentro de la cuenca del Río Maipo, se registran más de 20 estaciones fluviométricas y más de 60 estaciones pluviométricas (Herramienta Evaluación Rápida Agua (HERA), 2025). Para el desarrollo de este trabajo, se seleccionaron las siguientes estaciones fluviométricas para analizar el comportamiento del nivel del caudal, las cuales se encuentran en la cuenca alta del río Maipo: Río Maipo en el Manzano, Río Maipo en las Hualtatas, Río Maipo en las Melosas, Río Maipo en San Alfonso y Río Volcán en Queltehues. Para las estaciones antes mencionadas, se presenta la Tabla 4-1 la cual incorpora el periodo de registro, las coordenadas y el porcentaje de completitud para cada una de las estaciones. Esta última columna, indica el porcentaje de registros diarios que efectivamente se disponen para la estación.

Tabla 4-1: Estaciones Fluviométricas utilizadas en este estudio. Fuente (Dirección General de Aguas, 2024).

Estaciones Fluviométricas	Período de Registro	Coordenadas		Altitud (metros)	Porcentaje de completitud
		Latitud	Longitud		
Río Maipo en el Manzano	1946-2020	-33.5975°	-70.3794°	850	90%
Río Maipo en San Alfonso	1942-2017	-33.9781°	-70.1456°	1092	87%
Río Volcán en Queltehues	1946-2020	-33.8519°	-70.1983°	1365	79%
Río Maipo en Las Melosas	1962-1993	-33.7349°	-70.3011°	1527	46%
Río Maipo en Las Hualtatas	1979-2013	-33.8094°	-70.2103°	1820	91%

La zona de estudio de este documento se desarrolla en la Cuenca Alta del Río Maipo, ubicada en la región central de Chile. Esta cuenca, de gran relevancia tanto para el abastecimiento hídrico como para la gestión ambiental, se extiende a lo largo de un complejo territorio montañoso. A continuación, en la Figura 4-2 se presenta un mapa de la cuenca zona de estudio, que muestra las diferentes altitudes de la zona, así como la localización de estaciones clave como lo son Río Maipo en El Manzano, Río Maipo en San Alfonso, San Gabriel y el Río Volcán en Queltehues.

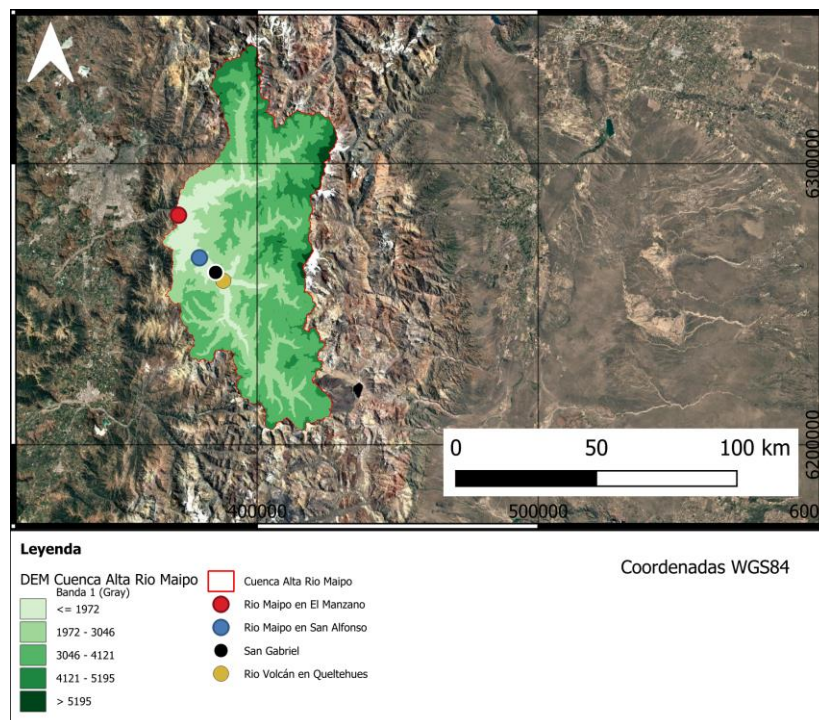


Figura 4-2: Mapa de la Cuenca Alta del Río Maipo con Clasificación de Altitudes y Ubicación de estaciones.

La cuenca del Río Maipo se caracteriza por presentar un ciclo anual de precipitación bien definido, caracterizado por un máximo invernal y un período estival con un monto significativamente menor que va aumentando hacia el sur. Este régimen se conoce como clima templado mediterráneo y abarca desde los 30°S hasta los 40°S (Videla Giering, 2013).

Las condiciones climáticas dependen de la zona de la cuenca debido a que el clima de la zona centro y baja difiere al de la zona alta. Para este estudio, se utiliza la zona alta de la cuenca la cual presenta un clima Frio de Altura, donde la principal característica son las bajas temperaturas y las precipitaciones sólidas caracterizan un clima Frio de Altura, permitiendo la acumulación de nieve y mantención de cuerpos de hielo en cumbres y quebradas de la alta Cordillera (Gobierno de Chile, 2004; Videla Giering, 2013). En relación con la precipitación media anual de la cuenca, las zonas más elevadas presentan un valor de 536 milímetros medios anuales (Gobierno de Chile, 2004).

Hasta el año 2002, el 40% de la población de Chile se encontraba dentro de la Cuenca del Río Maipo, siendo aproximadamente 6.061.185 habitantes. En la cuenca, hay un total de 163 localidades, siendo el gran Santiago la que presenta mayor población con 4.668.473 habitantes (Gobierno de Chile, 2004). Más recientemente, se estima que cerca del 50% de la población de Chile vive en la cuenca del Río Maipo (Urrutia et al., 2021). Particularmente, la Cuenca Alta del Río Maipo se caracteriza por ser una zona montañosa y rural. La zona urbana más cercana y con mayor población es San José de Maipo con poco menos de 20000 habitantes (Instituto Nacional de Estadísticas (INE), 2018).

La superficie total de la Cuenca del Río Maipo es de 1.521.340 hectáreas. Dentro de los usos que se le da al suelo de la cuenca del Río Maipo, en el año 2004 se realizó un estudio por parte del Gobierno de Chile en el cual se muestra que los 2 usos más relevantes son para terrenos agrícolas y áreas urbanas e industriales con un 16,1% (246.447 [Ha]) y 4% (58.220 Ha) respectivamente. Dentro del porcentaje restante (cerca del 80%), encontramos praderas (4,2%), plantaciones forestales (0,6%), minería industrial (0,1%), bosque nativo y bosque mixto (4,8%), áreas sin vegetación (28,9%) y otros usos (41,4%) entre los que se encuentran matorrales, áreas no reconocidas, entre otras. (Gobierno de Chile, 2004).

A lo largo de los años, ha existido una variabilidad en el uso de este suelo, ya que según un estudio realizado por el Centro de Información de Recursos Naturales (CIREN) (CIREN, 2021) un 14,1% a terrenos agrícolas,

un 7,6% a áreas urbanas e industriales y un restante de 78.3%. Se presenta una tabla resumen comparando ambos estudios y observando la variabilidad con el paso el paso de los años:

Se observa en la Tabla 4-2 un aumento significativo en la superficie destinada a áreas urbanas e industriales, con un incremento del 3,6%, lo que indica un proceso de urbanización y expansión industrial en la región. Por otro lado, la superficie de terrenos agrícolas ha disminuido en un 2,0. Finalmente, la categoría de otros usos también presenta una leve reducción del 1,6%, reflejando ajustes en la distribución territorial de la cuenca. Todo lo anterior, refleja la variabilidad en el uso del suelo que ha tenido la Cuenca del Río Maipo en relación del año 2004 con el 2021.

Tabla 4-2: Variabilidad del uso del suelo en la Cuenca del Río Maipo.

Uso del suelo	Año 2004 (%)	Año 2021 (%)	Variabilidad (%)
Terrenos agrícolas	16,1	14,1	-2,0
áreas urbanas e industriales	4,0	7,6	3,6
Otros usos	79,9	78,3	-1,6

A continuación, se presenta una gráfica que muestra la distribución del uso de suelo en la Cuenca del Río Maipo para el año 2021.

Distribución del Uso de Suelo en la Cuenca del Río Maipo

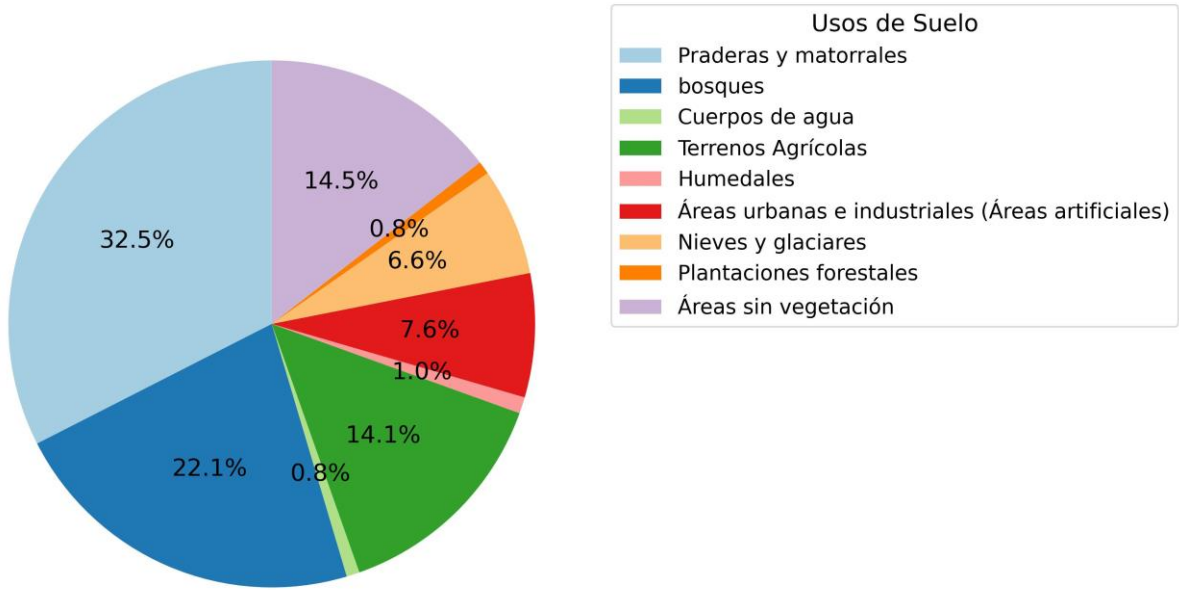


Figura 4-3: Distribución del uso de suelo en la Cuenca del Río Maipo para el año 2021. Fuente: (CIREN, 2021).

Finalmente, se cuantifican los periodos de retorno asociados al máximo caudal medio diario desde 1946 hasta 2024 para la estación Río Maipo en el manzano. El periodo de retorno es un concepto utilizado en hidrología para representar la recurrencia de un evento hidrológico, como un caudal máximo. En este caso, los periodos de retorno se calculan a partir de una serie histórica de caudales máximos anuales. El cálculo sigue el principio general de que el periodo de retorno es inversamente proporcional a la probabilidad de que ocurra un caudal determinado dentro de un año. La fórmula para determinar el periodo de retorno (T) es la siguiente:

$$T = \frac{N + 1}{n} \quad (6)$$

Donde N es el número total de años en la serie histórica (es decir, el total de observaciones disponibles) y n es el orden de ocurrencia del caudal máximo en ese año, es decir, el lugar que ocupa el valor dentro del ranking de caudales más altos, de mayor a menor.

La probabilidad de no excedencia asociada a un caudal máximo de un año se determina a partir de la fórmula del periodo de retorno. Esta probabilidad es la inversa del periodo de retorno, y es representada como:

$$\text{probabilidad de no excedencia} = 1 - \frac{1}{T} \quad (7)$$

El periodo de retorno a utilizar en este estudio se presenta en la en la Tabla 5-3 y Figura 4-4 los cuales se presentan a continuación:

Tabla 4-3: Periodo de retorno adoptado estación Río Maipo en el Manzano.

Periodo de retorno en años	Probabilidad de no excedencia	Caudal [m³/s]
1	0,000	100
1,5	0,333	240
2	0,500	305
3	0,667	399
4	0,750	443
5	0,800	489
8	0,875	565
10	0,900	633
20	0,950	775
40	0,975	1044
80	0,988	1135

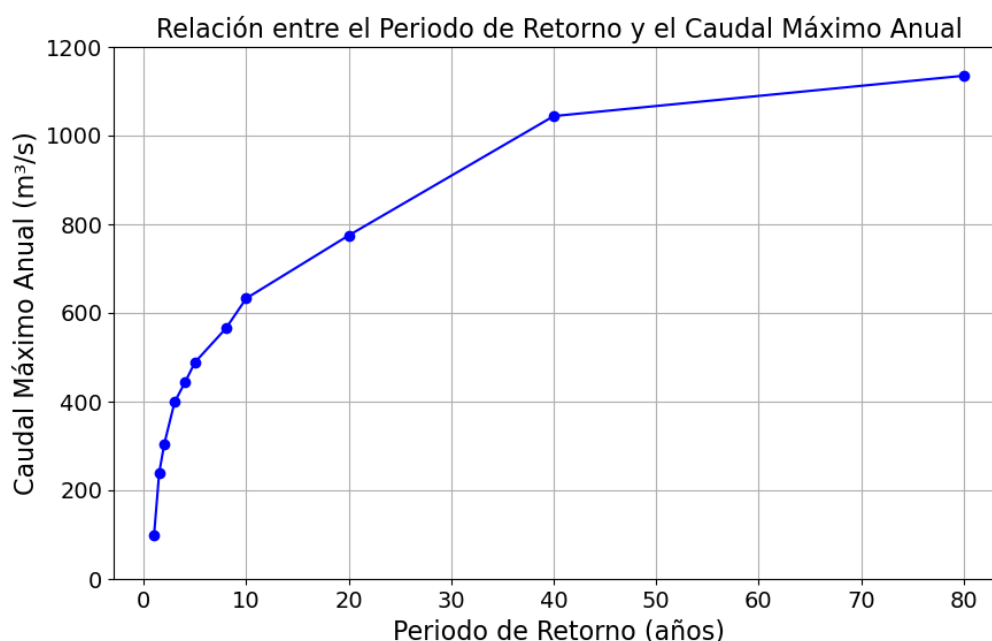


Figura 4-4: Relación entre el Periodo de Retorno y el Caudal Máximo Anual.

En la Tabla 5-3 se observa que el valor máximo de caudal diario registrado en la cuenca se espera que ocurra aproximadamente cada 80 años. Por otro lado, un caudal máximo diario de 100 m³/s se espera que ocurra de manera anual. Es importante mencionar que el periodo de retorno asociado a cada uno de los caudales máximos anuales analizados desde el año 1946 hasta el 2024 en la Cuenca del Río Maipo corresponden al Anexo 10.1 de este documento.

Finalmente, se estima la isoterma 0 utilizando datos de temperatura registrados en la Cuenca Alta del Río Maipo. Para ello, se aplicó un enfoque de interpolación espacial de las series temporales de temperatura obtenidas de estaciones meteorológicas distribuidas a lo largo de la cuenca, específicamente de las estaciones de Río Maipo en El Manzano y San Gabriel. Utilizando el gradiente térmico promedio de 6,5°C por cada kilómetro de ascenso, estimamos la temperatura en diversas elevaciones mediante la siguiente ecuación de corrección de altitud:

$$Altitud_{0^{\circ}C} = \frac{T_{ecuación}}{6,5} * 1000 [m] \quad (8)$$

5 Metodología

La metodología se divide en actividades destinadas a cumplir con el objetivo general y los objetivos específicos propuestos. Se presentan actividades transversales que responderán a los objetivos específicos planteados. A continuación, se presenta un diagrama de la metodología a seguir:

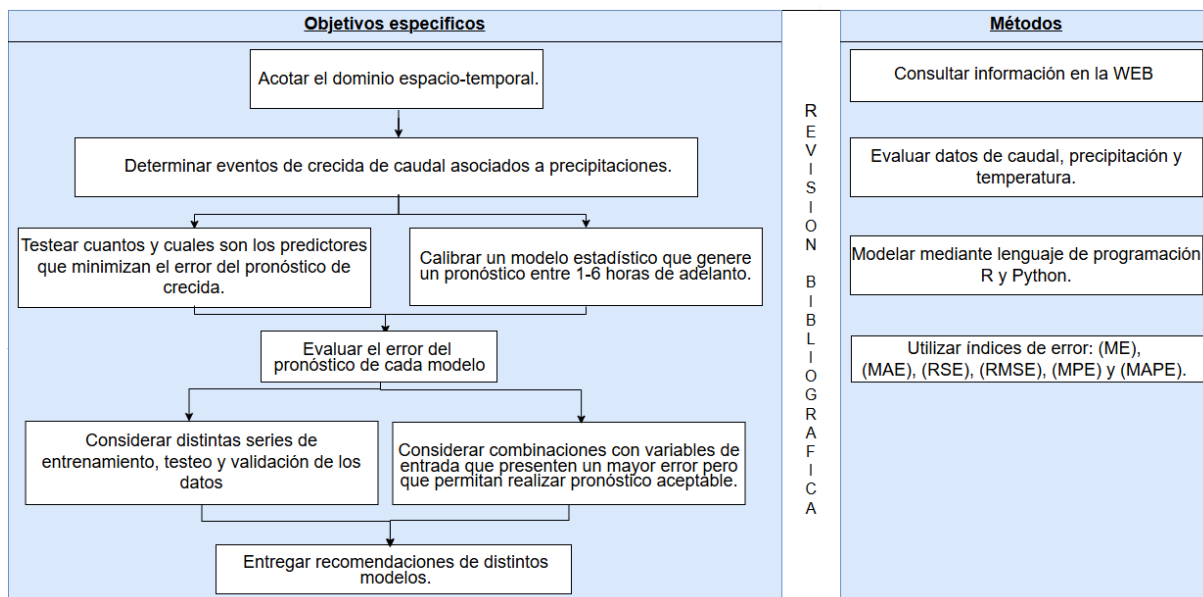


Figura 5-1: Metodología.

5.1 Base de datos

En primer lugar, se busca acotar el dominio espacio-temporal de las variables de entrada disponibles, lo cual corresponde a la primera actividad a realizar de las indicadas en la figura 1. Esta actividad implica definir cuál será la estación patrón de la cuenca e identificar las estaciones ubicadas aguas arriba y aguas abajo que cuenten con datos de precipitación, temperatura, entre otros. Para ello, se consulta la información disponible en la web (DGA, DMC, etc.).

La información de las estaciones utilizadas para la realización del presente estudio ha sido proporcionada por la Dirección General de Aguas (Dirección General de Aguas, 2024) . La ubicación de las estaciones pluviométricas y fluviométricas analizadas de la Cuenca del Río Maipo se pueden observar en la **¡Error! No se encuentra el origen de la referencia..** Además, a continuación, se presenta el detalle del periodo de registro de cada una de las estaciones analizadas:

Tabla 5-1: Estaciones Fluviométricas analizadas. Fuente: Elaboración propia en base a los datos de (Centro de Ciencia del Clima y la Resiliencia (CR2), 2024)

Estaciones Fluviométricas	Período de Registro	Coordenadas		Altitud (metros)	Porcentaje de completitud
		Latitud	Longitud		
Río Maipo en el Manzano	1946-2020	-33.5975°	-70.3794°	850	90%
Río Maipo en San Alfonso	1942-2017	-33.9781°	-70.1456°	1092	87%
Río Volcán en Queltehues	1946-2020	-33.8519°	-70.1983°	1365	79%

Estaciones Fluviométricas	Período de Registro	Coordenadas		Altitud (metros)	Porcentaje de completitud
		Latitud	Longitud		
Río Maipo en Las Melosas	1962-1993	-33.7349°	-70.3011°	1527	46%
Río Maipo en Las Hualtatas	1979-2013	-33.8094°	-70.2103°	1820	91%

Tabla 5-2: Estaciones Pluviométricas analizadas. Fuente: Elaboración propia en base a los datos de (Centro de Ciencia del Clima y la Resiliencia (CR2), 2024)

Estaciones pluviométricas	Período de Registro	Coordenadas		Altitud (metros)	Porcentaje de completitud
		Latitud	Longitud		
Río Maipo en el Manzano	2010-2024	-33.5975°	-70.3794°	850	16%
Río Maipo en San Alfonso	1962-1973	-33.9781°	-70.1456°	1092	13%
Río Volcán en Queltehues	2014-2024	-33.8519°	-70.1983°	1365	11%
Río Maipo en Las Melosas	1962-2004	-33.7349°	-70.3011°	1527	45%
Río Maipo en San Gabriel	1977-2019	-33.7825°	-70.2394°	1266	46%

Se observa en la Tabla 5-1, que la estación “Río Maipo en el Manzano” es la que presenta mayor periodo con datos, por lo cual esta es la que se utilizara como la estación patrón del estudio. Además, se observa que la estación “Río Volcán en Queltehues” es la que presenta la mayor cantidad de datos en el período analizado, por lo que esta estación será utilizada como auxiliar para el presente estudio.

En la Tabla 5-2, se observa que para la estación patrón, se cuentan con datos de precipitación desde el año 2010 hasta el 2024, por lo que será en dicho periodo en el cual se analizaran los datos de caudal para determinar crecidas.

En segundo lugar, se determina qué eventos de crecida de caudal están asociados a fenómenos extremos de precipitación, con el fin de generar una base de datos a nivel horario. Esta actividad requiere la evaluación de los datos de caudal, precipitación y temperatura. Para este estudio, se considera que una crecida corresponde a un incremento inusual en los valores de caudal instantáneo. Para seleccionar las crecidas que alimentarán el modelo se realiza un barrido histórico de los datos de caudal desde el año 1946 hasta el 2024 en la Cuenca del Río Maipo, en donde se identifican las más importantes por año en base a su magnitud. Además, se realiza una búsqueda manual en toda la extensión de la serie de precipitación y caudal para identificar el aumento repentino de caudal como producto de un evento de precipitación. Finalmente, se cuantifican los periodos de retorno para la estación Río Maipo en el manzano, los cuales se presentan en la Tabla 5-3:

Tabla 5-3: Periodo de retorno adoptado estación Río Maipo en el Manzano.

Periodo de retorno en años	Probabilidad de no excedencia	Caudal [m³/s]
1	0,000	100
1,5	0,333	240
2	0,500	305
3	0,667	399
4	0,750	443
5	0,800	489
8	0,875	565
10	0,900	633
20	0,950	775
40	0,975	1044
80	0,988	1135

En la Tabla 5-3 se observa que el valor máximo de caudal registrado en la cuenca se espera que ocurra aproximadamente cada 80 años. Por otro lado, un caudal de 100 m³/s se espera que ocurra de manera anual.

Es importante mencionar que el periodo de retorno asociado a cada uno de los caudales analizados desde el año 1946 hasta el 2024 en la Cuenca del Río Maipo corresponden al Anexo 10.1 de este documento.

A modo de barrido en general se evaluaron todas las crecidas de caudal desde el año 2010 en adelante, las cuales se presentan a continuación y hay que tener en cuenta que para cada evento se registra en la columna “Fecha” el día en el que se registra el caudal máximo instantáneo para el evento, el cual es registrado en la columna “Caudal [m³/s]”. Además, se incluye en la Tabla 5-4, el periodo de retorno para cada una de las crecidas junto con la precipitación acumulada la cual es obtenida de la mediciones de las estaciones Pluviométricas Río Maipo en el manzano y Río Maipo en San Gabriel desde la Dirección General de Aguas (Dirección General de Aguas., n.d.). Para estimar la Isoterma 0, se utilizan datos de temperatura de la estación Río Maipo en el manzano y Río Volcán en Queltehués, los cuales se obtienen desde el Explorador Climático (Centro de Ciencia del Clima y la Resiliencia (CR2), 2024).

Tabla 5-4: Eventos de crecida de caudal. Fuente: (Dirección General de Aguas., n.d.)

Evento	Fecha	Q_{max} instantáneo [m³/s]	Periodo de Retorno (T)	Pp acumulada [mm]	Temperatura media diaria [°C]	Isoterma 0 [m]
Evento 1	24-06-2010	75,3	~1,0	52,0	13,7	2108
Evento 2	22-08-2010	61,2	~1,0	11,0	6,6	1015
Evento 3	14-04-2012	79,0	~1,0	27,0	16,0	2462
Evento 4	27-05-2012	305,0	2,0	91,7	11,7	1800
Evento 5	13-06-2012	70,2	~1,0	35,5	9,0	1385
Evento 6	30-06-2012	116,0	1,1	45,2	10,0	1538
Evento 7	28-05-2013	57,3	~1,0	120,0	6,4	985
Evento 8	12-08-2013	64,5	~1,0	24,0	7,1	1092
Evento 9	07-09-2013	97,9	~1,0	6,0	14,7	2262
Evento 10	19-06-2014	58,2	~1,0	0	15,0	2308
Evento 11	01-08-2014	52,8	~1,0	0	12,8	1969
Evento 12	17-04-2016	1044	40,0	102,0	9,3	1431
Evento 13	13-06-2016	75,0	~1,0	12,6	11,7	1800

Evento	Fecha	Q_{max} instantáneo [m³/s]	Periodo de Retorno (T)	Pp acumulada [mm]	Temperatura media diaria [°C]	Isoterma 0 [m]
Evento 14	20-04-2017	150,0	1,2	12,5	12,9	1985
Evento 15	11-05-2017	76,0	~1,0	72,2	8,3	1277
Evento 16	16-06-2017	90,1	~1,0	70,0	8,7	1338
Evento 17	04-04-2018	65,9	~1,0	0	16,8	2585
Evento 18	06-07-2018	49,3	~1,0	54,7	4,7	723
Evento 19	28-08-2018	55,9	~1,0	0	3,6	554
Evento 20	01-09-2019	46,8	~1,0	0	10,7	1646
Evento 21	27-09-2019	65,3	~1,0	0	13,2	2031

En la Tabla 5-4 se destaca que el evento 12, ocurrido el 17 de abril de 2016, registró el mayor caudal con 1037,8 m³/s, representando un evento de gran magnitud. En contraste, el evento 20, ocurrido el 1 de septiembre de 2019, registró el menor caudal con 46,8 m³/s.

Asimismo, se observa en la Tabla 5-4 que todos los eventos de crecida ocurrieron entre los meses de abril y septiembre, lo que indica una fuerte estacionalidad en la ocurrencia de estos fenómenos. Este patrón coincide con los meses de otoño e invierno, períodos en los que se presentan mayores precipitaciones, factores que pueden contribuir al aumento del caudal en la cuenca.

A partir de las crecidas observadas Tabla 5-4, se analizan los datos de precipitación y temperatura en la estación Río Maipo en el manzano para la misma fecha con el fin de establecer las crecidas que corresponden a un evento extremo de precipitación.

A modo de ejemplo, se presentan 2 eventos de los registrados en la Tabla 5-4 junto con sus datos de precipitación y temperatura para definir cuál es la razón de la crecida de caudal. Los eventos corresponden al Evento 8 y al Evento 10.

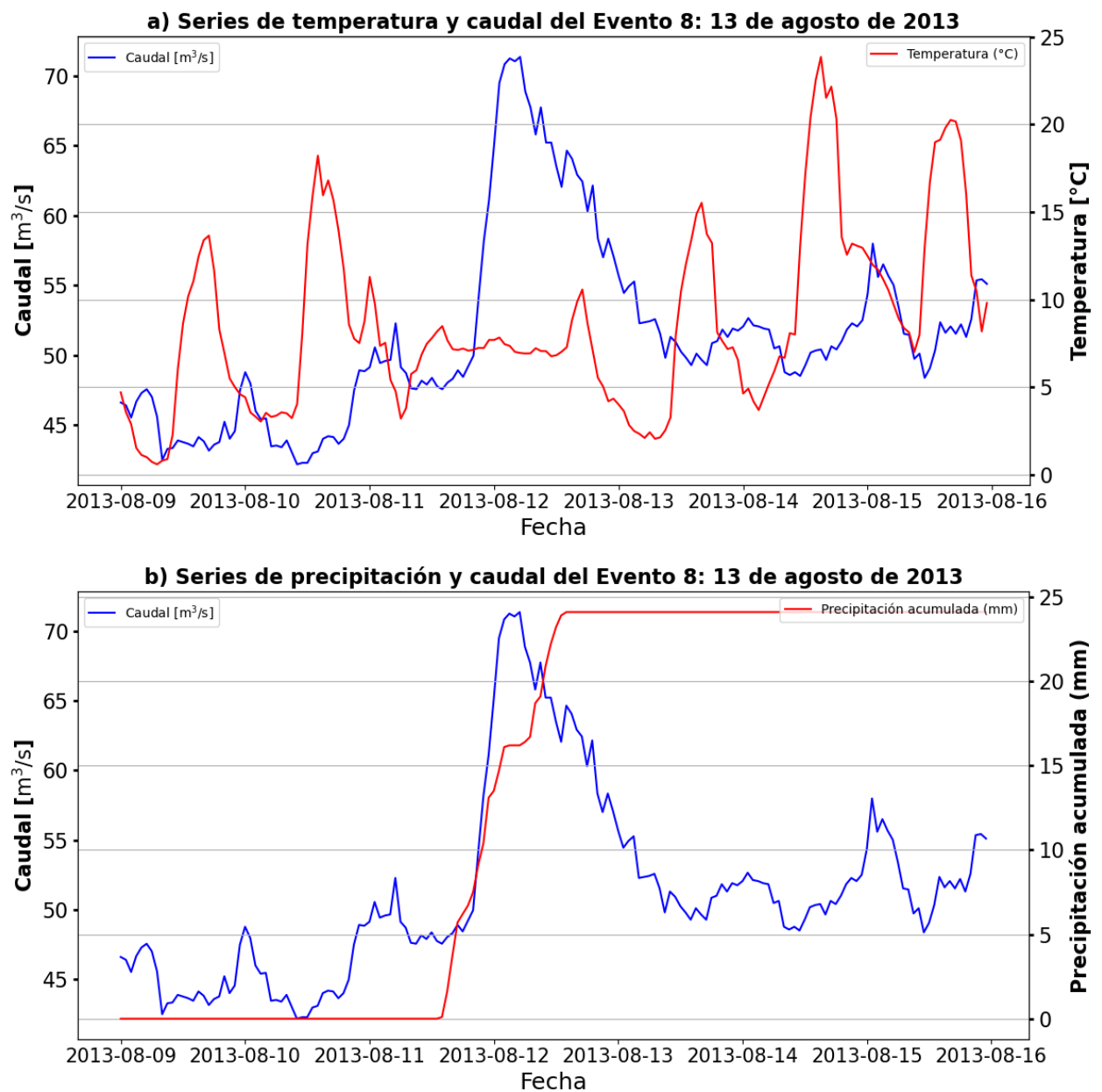


Figura 5-2: Caudal, precipitación y temperatura Evento 8.

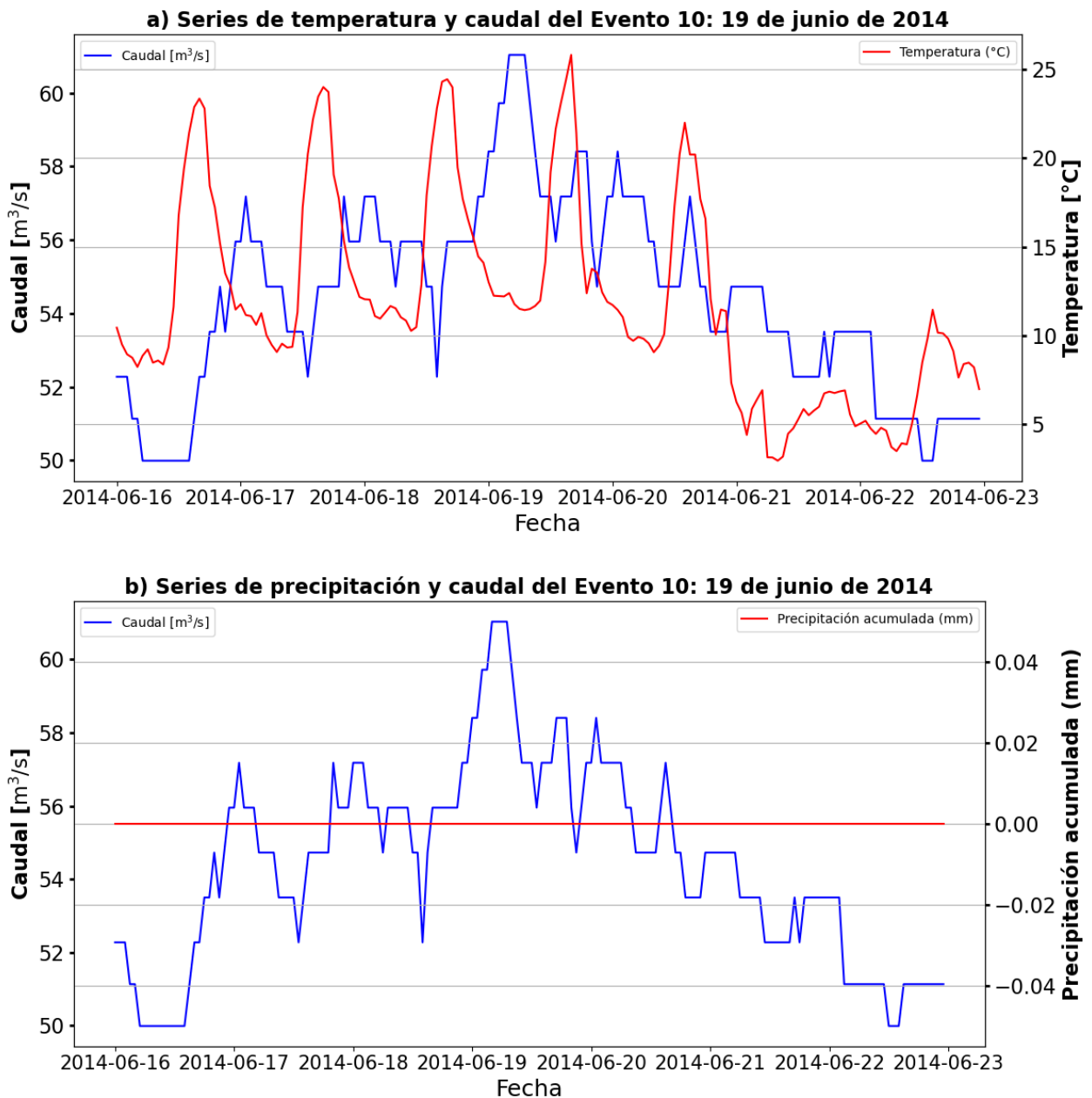


Figura 5-3: Caudal, precipitación y temperatura Evento 10.

El comportamiento de caudal, precipitación y temperatura para el Evento 8 queda representado en la Figura 5-2 en la cual se observa que durante los días de crecida de caudal existe un evento extremo de precipitación y un comportamiento sin anomalías para la temperatura. Dado lo anterior, se puede concluir que el evento 8 fue generado por Precipitación y no por Temperatura.

El comportamiento de caudal, precipitación y temperatura para el Evento 1 queda representado en la Figura 5-3, en la cual se observa que durante los días de crecida de caudal existen alzas en la temperatura lo cual contribuye a un deshielo, lo que favorece a un aumento de caudal. Además, se observa que, durante esos días, no hubo registros de precipitación en la cuenca. Dado lo anterior, se puede concluir que el evento 10 fue generado por deshielo, como consecuencia de un alza en las temperaturas y no por Precipitación.

El análisis realizado para los eventos 8 y 10, se replica para cada uno de los eventos registrado en la Tabla 5-4, lo cual va disminuyendo la cantidad total de eventos de crecida de caudal, dejando únicamente los eventos asociados a precipitación.

5.2 Evaluación de la Memoria Hidrológica del Río Maipo mediante ACF y PACF

Tal como se ha visto anteriormente, la estación patrón corresponde a Río Maipo en el Manzano debido a su gran disponibilidad de datos de caudal tanto a nivel diario como horario. Se lleva a cabo el análisis de la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) para el caudal del Río Maipo a nivel horario y diario, con el fin de comprender mejor la memoria de la cuenca en distintos períodos del año. El análisis se realizó utilizando datos de caudal diarios y horarios desde el 1 de enero de 2020 hasta el 30 de septiembre de 2024. Los datos fueron segmentados en tres períodos distintos para evaluar posibles variaciones estacionales en la dinámica de autocorrelación: (i) el año completo, (ii) el período pluvial, y (iii) el período de deshielo. Se aplicó la función ACF para evaluar la dependencia temporal en los distintos períodos, mientras que la PACF se utilizó para examinar la dependencia residual después de eliminar la correlación de rezagos anteriores.

5.3 Combinación de predictores

Una vez conformada la base de datos, se observa en la Figura 5-1 que hay dos actividades para el desarrollo de los modelos: la primera consiste en testear cuántos y cuáles predictores minimizan el error del pronóstico de crecida (Objetivo Específico 1); el segundo se enfoca en predecir crecidas con un tiempo de adelanto de 1 a 6 horas (Objetivo Específico 2). Para llevar a cabo estas actividades, se utiliza los softwares R y Python (Python Software Foundation, 2024; R Core Team, 2024). A continuación, se presentan las principales librerías utilizadas en el código para el ajuste y evaluación del modelo SARIMAX en series temporales:

1. **pandas**: Utilizada para la manipulación y análisis de datos, incluyendo la lectura de archivos Excel, el manejo de fechas y la realización de operaciones sobre series temporales.
2. **statsmodels.tsa.statespace.sarimax**: Se utiliza para crear y ajustar el modelo SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors), que es esencial para la predicción de series temporales con componentes estacionales y regresores exógenos.
3. **numpy**: Usada para realizar operaciones matemáticas y estadísticas, como el cálculo de la raíz cuadrada, y la manipulación de arrays, especialmente en la evaluación de modelos, como el cálculo del RMSE (Root Mean Squared Error).
4. **sklearn.metrics**: Proporciona funciones para calcular métricas de evaluación del modelo de predicción, como el RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), y el R^2 (coeficiente de determinación).
5. **itertools**: Utilizada para generar combinaciones de parámetros del modelo SARIMAX, lo que permite explorar diversas configuraciones del modelo de manera sistemática.
6. **matplotlib.pyplot**: Usada para la visualización de los resultados, generando gráficos que comparan las predicciones con los valores reales de la serie temporal.
7. **warnings**: Se emplea para suprimir las advertencias generadas durante el ajuste del modelo, garantizando que el proceso de ejecución no se vea interrumpido por mensajes no relevantes.

Para evaluar de los predictores que minimizan el error en el pronóstico de crecidas, se han definido distintas combinaciones de variables para la generación de modelos. Estas combinaciones incluyen la precipitación y la temperatura de la estación patrón (EP), correspondiente a la estación Río Maipo en El Manzano, así como el caudal y la precipitación de la estación auxiliar aguas arriba (AA), ubicada en la estación Río Volcán en Queltehues. A continuación, se presentan las combinaciones de predictores a testear.

En la Tabla 5-5, se observa que las combinaciones de predictores presentan variaciones en la cantidad de eventos registrados. Esto implica que el número de modelos de pronóstico generados por combinación será variable, dependiendo de la disponibilidad de datos en cada caso.

Tabla 5-5: Combinación de predictores a testear.

Nº de combinación	Combinación de predictores	Numero de eventos con registro	Eventos disponibles
1	Precipitación + temperatura	11	3 al 13
2	Precipitación	11	3 al 13
3	Temperatura	13	1 al 13
4	Precipitación + temperatura + caudal AA	5	9 al 13
5	Precipitación + caudal AA	5	9 al 13
6	Temperatura + caudal AA	5	9 al 13
7	Precipitación + temperatura + caudal AA + precipitación AA	5	9 al 13
8	Precipitación + caudal AA + precipitación AA	5	9 al 13
9	Temperatura + caudal AA + precipitación AA	5	9 al 13
10	Caudal AA	5	9 al 13

Para este estudio, cada modelo de pronóstico se genera a partir de los datos disponibles de cada evento. Como ejemplo, si una combinación cuenta con 10 eventos registrados, se desarrollarán 10 modelos de pronóstico, uno por cada evento. Cada modelo generado se evalúa utilizando los otros 9 eventos de la misma combinación para su validación.

Para estandarizar la identificación de los modelos, se empleará la siguiente nomenclatura:

$$\begin{aligned}
 \text{Modelo 1} &= \theta_1 \\
 \text{Modelo 2} &= \theta_2 \\
 &\vdots \\
 \text{Modelo 13} &= \theta_{13}
 \end{aligned}$$

Para determinar el modelo que mejor representa la serie temporal, se procede a efectuar la actividad que evalúa el error del pronóstico de caudal de cada modelo, para ello se utilizan distintos índices de error, tales como El criterio de información de Akaike (AIC), Coeficiente de determinación (R^2), raíz del error cuadrático medio (RMSE), y error absoluto porcentual medio (MAPE) (Diego Martínez et al., 2009; Meis et al., 2017). A continuación, se detallan cada uno de los errores mencionados anteriormente:

- AIC: El criterio de información de Akaike (*An Information Criterion*, AIC) proporciona un método simple y objetivo que selecciona el modelo más parsimonioso para caracterizar los datos experimentales (Diego Martínez et al., 2009). Este criterio, que se enmarca en el campo de la teoría de la información, se define como:

$$AIC = -2 \cdot \log(\mathcal{L}(\hat{\theta})) + 2 \cdot K \quad (9)$$

donde $\log(\mathcal{L}(\hat{\theta}))$ es el logaritmo de la máxima verosimilitud el cual corresponde a la probabilidad de que el modelo haya generado los valores observados y que además permite determinar los valores de los parámetros libres de un modelo estadístico, y K es el número de parámetros libres del modelo.

- Coeficiente de determinación (R^2): Es una medida que te representa la variación de los datos reales con los observados. También, representa que tan bien describe el modelo de regresión los datos observados. A continuación, se presenta la ecuación que define al coeficiente de determinación en la cual el subíndice i en la ecuación hace referencia a cada uno de los datos en el conjunto de datos utilizado. Es decir, para cada valor en el conjunto de datos, el subíndice i representa la posición de ese valor en la secuencia de datos.

$$R^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (10)$$

Donde:

Y_i = Valores reales de la variable dependiente.

$\hat{Y}_i$ = Valores predichos por el modelo.

$\bar{Y}$ = Media de los valores observados.

- Raíz del error cuadrático medio (RMSE): Corresponde a la raíz cuadrada del error cuadrático medio. Además, este índice se conoce como la desviación estándar de los errores. A continuación, se presenta la fórmula para determinar el error descrito:

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (y_i - f_i)^2} \quad (11)$$

- Error absoluto porcentual medio (MAPE): Representa el promedio de los errores absolutos en porcentaje, calculando la diferencia entre los valores predichos y los valores reales, normalizada por los valores reales. A continuación, se presenta la fórmula para determinar el error descrito:

$$MAPE = \frac{1}{n} \cdot \sum_{i=1}^n 100 \cdot \left| \frac{y_i - f_i}{y_i} \right| \quad (12)$$

- Error en el Peak de Caudal: Este indicador muestra la variación porcentual entre el caudal máximo instantáneo real de una crecida y el valor del pico de caudal pronosticado por un modelo. Si el valor del error es positivo, significa que el pico del caudal simulado es superior al pico del caudal real. En cambio, si el valor del error es negativo, indica que el pico de caudal observado es mayor que el pico simulado.

$$\varepsilon (\%) = \frac{(Q_{max}^{simulado} - Q_{max}^{observado})}{Q_{max}^{observado}} \cdot 100 \quad (13)$$

5.4 Validación Modelo Optimo

Una vez determinado cual es el modelo que mejor representa el caudal en la cuenca, se procede a validar el modelo. Para ello, se utilizan 3 eventos extras denominados Evento 15, Evento 16 y Evento 17, los cuales se pronostican utilizando el modelo y se analiza su comportamiento. A continuación, se presentan los eventos de validación:

Tabla 5-6: Eventos de validación del modelo optimo. Fuente: (Dirección General de Aguas., n.d.)

Evento	Fecha	Q_{max} instantáneo [m ³ /s]	Periodo de Retorno (T)	Pp acumulada [mm]
Evento 14	19-08-2021	107,4	1,1	8,9
Evento 15	23-06-2023	1037,8	27,0	145
Evento 16	22-08-2023	378,9	2,7	237

5.5 Tiempos de adelanto

Una vez validado el modelo optimo, se genera un pronóstico con tiempos de adelanto de 1 a 6 horas para los dos mejores modelos. Cada lead time obtenido es analizado con los mismos índices de error utilizados en la evaluación de los predictores, con el objetivo de determinar cuál o cuáles presentan un mejor desempeño. Lo anterior busca responder al Objetivo Específico 2 de este estudio.

Finalmente, en la Figura 5-1, se observa que se siguen dos enfoques para entregar recomendaciones de modelos: por un lado, se emplean diferentes series de entrenamiento, testeo y validación; por otro, se prueban distintas combinaciones de variables de entrada. Aunque estas combinaciones pueden presentar un error mayor, este puede ser considerado aceptable según las métricas de error. El objetivo es generar recomendaciones de modelos válidos en caso de fallos o errores, así como proporcionar alternativas a los modelos con menor error. Estas recomendaciones buscan responder al Objetivo Específico 3.

5.6 Validación Manual Tiempos de adelanto

Para validar los resultados obtenidos en los diferentes tiempos de adelanto, se seleccionó el modelo óptimo correspondiente a los tiempos de adelanto de 1, 3 y 5 horas. El desempeño de cada modelo se comparó con 100 combinaciones generadas en Excel, las cuales se obtuvieron utilizando la función INV.NORM para calcular un ruido blanco específico para cada combinación. Esta función requiere conocer la media y la desviación estándar del modelo, valores que fueron proporcionados por el modelo de Python (Python Software Foundation, 2024).

Para llevar a cabo la comparación, se replicó la ecuación del modelo óptimo correspondiente al tiempo de adelanto que se estaba evaluando. Una vez replicada la ecuación, se calculó el coeficiente de determinación R^2 para cada combinación. Finalmente, se generaron dos gráficos: un boxplot que muestra la distribución de los valores de R^2 y un gráfico temporal del evento de crecida, en el cual se comparan los datos reales, las predicciones del modelo y las 100 validaciones realizadas. Estos gráficos permiten evaluar de manera visual y cuantitativa la precisión y consistencia de las predicciones del modelo en comparación con los datos observados.

6 Análisis y Resultados

En esta sección se presentan y analizan los resultados obtenidos en la evaluación de cada uno de los modelos de pronóstico de caudal para la Cuenca del Río Maipo. En primer lugar, se determina la base de datos de crecida de caudal que están asociados a un evento de precipitación, la cual se presenta a continuación:

Tabla 6-1: Base de datos de crecida de caudal en la estación “Río Maipo en el Manzano” asociados a un evento de Precipitación.

Evento	Fecha	Q_{max} instantáneo [m ³ /s]	Periodo de Retorno (T)	Pp acumulada [mm]	Temperatura media diaria [°C]	Isoterma 0 [m]
Evento 1	24-06-2010	75,3	~1,0	52,0	13,7	2108
Evento 2	22-08-2010	61,2	~1,0	11,0	6,6	1015
Evento 3	14-04-2012	79,0	~1,0	27,0	16,0	2462
Evento 4	13-06-2012	70,2	~1,0	35,5	9,0	1385
Evento 5	30-06-2012	116,0	1,1	45,2	10,0	1538
Evento 6	28-05-2013	57,3	~1,0	120,0	6,4	985
Evento 7	12-08-2013	64,5	~1,0	24,0	7,1	1092
Evento 8	07-09-2013	97,9	~1,0	6,0	14,7	2262
Evento 9	17-04-2016	1044	40	102,0	9,3	1431
Evento 10	13-06-2016	75,0	~1,0	12,6	11,7	1800
Evento 11	20-04-2017	150,0	1,2	12,5	12,9	1985
Evento 12	11-05-2017	76,0	~1,0	72,2	8,3	1277
Evento 13	16-06-2017	90,1	~1,0	70,0	8,7	1338

Se observa en la Tabla 6-1, que la cantidad de eventos de crecida de caudal difiere a la cantidad total (21) presentada en la Tabla 5-4. Lo anterior se debe a que 8 crecidas iniciales se generaron por un aumento de la temperatura. Además, es importante mencionar que los eventos registrados en la Tabla 6-1, no siguen la misma numeración que en la Tabla 5-4, sino que se vuelven a enumerar desde el Evento 1 al Evento 13.

Luego, se generan y calibran cada uno de los modelos para las combinaciones detalladas en la Tabla 5-5. Es relevante mencionar que los parámetros (p,d,q) (P,D,Q,s) tienen un rango de 0 a 2 cada uno, exceptuando la estacionalidad (s) para la cual se adopta un valor de 12 o 24 horas ya que se considera un ciclo diario en los datos. Por consiguiente, se obtiene un modelo óptimo por evento, el cual se denomina “Theta X”, donde X corresponde al número de evento con el que fue calibrado

6.1 Función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF)

Se genera, en primer lugar, la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) para el caudal a nivel horario y diario para comprender la memoria que presenta la cuenca y, en particular, el Río Maipo. Para lo anterior, se disponen de datos de Caudal a nivel diario y horario desde el 01 de enero de 2020 hasta el 30 de septiembre de 2024. Además, dichas funciones se evalúan en 3 aristas:

- Año completo: Datos de caudal desde el 01 de enero al 31 de diciembre.
- Periodo Pluvial: Datos desde el 01 de mayo hasta el 31 de agosto.
- Periodo Deshielo: Datos desde el 01 de septiembre hasta el 31 de marzo.

A continuación, se presentan los resultados obtenidos para el análisis de la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF). En la Figura 6-1 se presentan los resultados de la función de autocorrelación (ACF) para el caudal del Río Maipo, analizados a nivel horario y diario. Los gráficos muestran cómo la autocorrelación disminuye a medida que aumenta el número de rezagos. Se comparan tres escenarios distintos: "Completo", "Pluvial" y "Deshielo", cada uno representado con diferentes colores y líneas. La gráfica horaria se encuentra en el gráfico de la izquierda, mientras que la diaria se encuentra en el de la derecha. Por su parte, En la Figura 6-2 se ilustra la función de autocorrelación parcial (PACF) para el caudal del Río Maipo, también a nivel horario y diario. La PACF muestra la relación entre el caudal y su valor en rezagos sucesivos, corrigiendo la influencia de rezagos intermedios. Al igual que en la figura anterior, se analizan tres escenarios: "Completo", "Pluvial" y "Deshielo", con las líneas representando cada condición en diferentes colores. La gráfica horaria está ubicada a la izquierda y la diaria a la derecha.

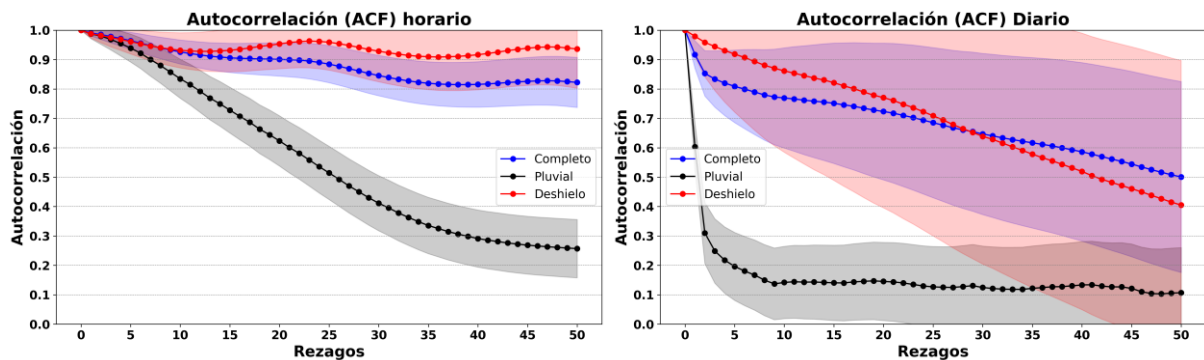


Figura 6-1: Función de Autocorrelación (ACF) para el caudal del Río Maipo a nivel horario y diario.

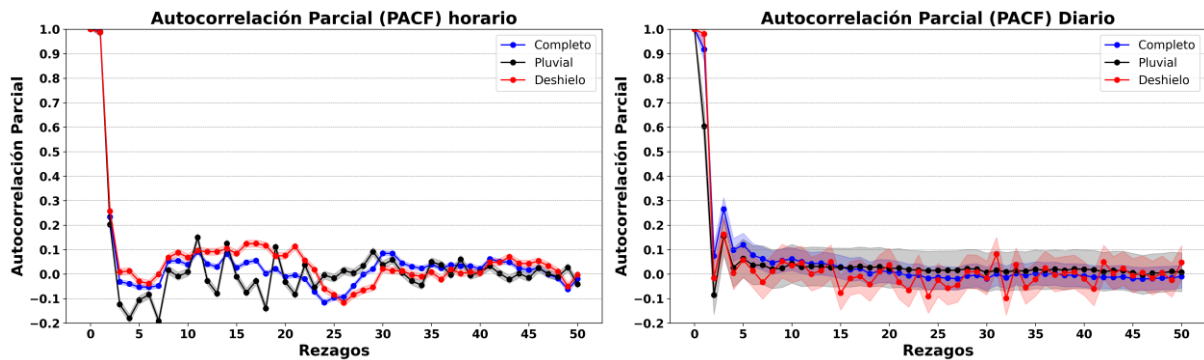


Figura 6-2: Función de Autocorrelación Parcial (PACF) para el caudal del Río Maipo a nivel horario y diario.

En la Figura 6-1, se observa que para la ACF horaria, las series presentan autocorrelaciones iniciales muy altas (cercanas a 1), lo que refleja una fuerte dependencia temporal a corto plazo. Esto indica que la cuenca posee una alta memoria temporal en el caudal. Entre las tres curvas representadas, la serie Pluvial muestra una disminución más pronunciada de la autocorrelación a medida que aumentan los rezagos. Esta caída es particularmente notable a partir de las 10 horas de retardo, donde la

autocorrelación deja de ser significativa. En contraste, la serie Deshielo mantiene valores elevados de autocorrelación a lo largo de los rezagos, evidenciando una mayor persistencia temporal, con una autocorrelación significativa incluso después de 30 horas de retardo. La autocorrelación del caudal considerando el período Completo se sitúa entre las curvas de Pluvial y Deshielo, con valores relativamente altos y sostenidos en el tiempo.

Al aumentar la escala temporal de horario a diario, también visible en la Figura 6-1, se aprecia un descenso en los niveles de autocorrelación para las tres series, siendo este decaimiento especialmente notable en la serie Pluvial.

Con base en lo anterior, se anticipa que para los Lead Times considerados en este estudio (1 a 6 horas), se presenten valores buenos en las métricas de error, debido a la buena memoria temporal observada en los datos horarios.

En la Figura 6-2, se observa que, tanto a nivel horario como diario, todas las series presentan una autocorrelación parcial cercana a 1 para el primer rezago. Además, en ambas escalas temporales, a partir del rezago 2, las autocorrelaciones parciales disminuyen significativamente y fluctúan en torno a cero. Esto indica que el caudal presenta una fuerte dependencia con el valor anterior. En el caso de la serie horaria, las autocorrelaciones dejan de ser significativas a partir de las 2 horas de retardo, mientras que, en la serie diaria, la autocorrelación sigue siendo significativa hasta las 6 horas. Esto sugiere que los modelos que utilicen el caudal en el tiempo $t-1$ como predictor pueden generar pronósticos con buenos niveles de precisión, especialmente en el corto plazo.

6.2 Combinación Precipitación EP + Temperatura EP

A continuación, se presentan los modelos SARIMA óptimos por cada evento de la combinación Precipitación EP + Temperatura EP, donde las columnas “Modelo” y “Ecuación del Modelo” corresponden al modelo asociado al número evento, a la ecuación que define al modelo respectivamente. Para la Tabla 6-2, considerar lo siguiente:

- $\hat{Q}_t$: Valor de caudal pronosticado para el tiempo t .
- Q_{t-1} : Valor de caudal en el tiempo anterior (1 hora antes).
- Q_{t-24} : Valor de caudal en el periodo anterior (24 horas antes).
- Q_{t-25} : Valor de caudal en el tiempo anterior del periodo anterior (25 horas antes).
- u_t : Ruido blanco o error en el tiempo t .
- u_{t-24} : Ruido blanco o error en el periodo anterior (24 horas antes).
- $X_{1,t-1}$: Precipitación en el tiempo $t-1$ [mm].
- $X_{2,t-1}$: Temperatura en el tiempo $t-1$ [°C].

Tabla 6-2: Modelos óptimos con su ecuación para la combinación Precipitación EP + Temperatura EP.

Modelo	Ecuación del Modelo
Theta 3	$\hat{Q}_t = 1,227 \cdot Q_{t-1} + Q_{t-24} - Q_{t-25} + u_t + 0,999 \cdot u_{t-24} - 1,058 \cdot X_{1,t-1} + 0,344 \cdot X_{2,t-1}$
Theta 4	$\hat{Q}_t = 1,174 \cdot Q_{t-1} + Q_{t-24} - Q_{t-25} + u_t - 0,999 \cdot u_{t-24} - 1,145 \cdot X_{1,t-1} + 0,384 \cdot X_{2,t-1}$
Theta 5	$\hat{Q}_t = 1,707 \cdot Q_{t-1} + Q_{t-24} - Q_{t-25} + u_t - 0,489 \cdot u_{t-1} - 0,999 \cdot u_{t-24} - 1,747 \cdot X_{1,t-1} + 0,548 \cdot X_{2,t-1}$
Theta 6	$\hat{Q}_t = Q_{t-1} + Q_{t-24} - Q_{t-25} + u_t - 0,998 \cdot u_{t-24} - 0,063 \cdot X_{1,t-1} + 0,030 \cdot X_{2,t-1}$
Theta 7	$\hat{Q}_t = 1,848 \cdot Q_{t-1} + Q_{t-24} - Q_{t-25} + u_t - 0,741 \cdot u_{t-1} - 0,998 \cdot u_{t-24} - 0,328 \cdot X_{1,t-1} + 0,057 \cdot X_{2,t-1}$
Theta 8	$\hat{Q}_t = 1,348 \cdot Q_{t-1} + Q_{t-24} - Q_{t-25} + u_t - 0,999 \cdot u_{t-24} - 0,172 \cdot X_{1,t-1} + 0,299 \cdot X_{2,t-1}$
Theta 9	$\hat{Q}_t = Q_{t-1} + 0,868 \cdot Q_{t-24} - Q_{t-25} + u_t + 0,676 \cdot u_{t-1} - 0,997 \cdot u_{t-24} - 2,652 \cdot X_{1,t-1} + 1,371 \cdot X_{2,t-1}$
Theta 10	$\hat{Q}_t = 1,774 \cdot Q_{t-1} + 1,157 \cdot Q_{t-24} - Q_{t-25} + u_t - 0,600 \cdot u_{t-1} - 0,999 \cdot u_{t-24} - 7,642 \cdot X_{1,t-1} + 0,626 \cdot X_{2,t-1}$
Theta 11	$\hat{Q}_t = 1,774 \cdot Q_{t-1} + 0,847 \cdot Q_{t-24} - Q_{t-25} + u_t - 0,382 \cdot u_{t-1} - u_{t-24} - 12,526 \cdot X_{1,t-1} + 0,133 \cdot X_{2,t-1}$
Theta 12	$\hat{Q}_t = Q_{t-1} + Q_{t-24} - Q_{t-25} + u_t - 0,989 \cdot u_{t-24} + 1,746 \cdot X_{1,t-1} + 0,203 \cdot X_{2,t-1}$
Theta 13	$\hat{Q}_t = 1,747 \cdot Q_{t-1} + Q_{t-24} - Q_{t-25} + u_t - 0,397 \cdot u_{t-1} - 0,999 \cdot u_{t-24} - 1,0189 \cdot X_{1,t-1} - 0,067 \cdot X_{2,t-1}$

Se observa en la Tabla 6-2 El modelo Theta 12 es el único que presenta un coeficiente positivo para la precipitación, lo que sugiere que, en este caso, un incremento en la precipitación se asocia con un aumento en el caudal pronosticado. Este comportamiento contrasta con el de los demás modelos, en los cuales la precipitación presenta coeficientes negativos, indicando una relación inversa entre ambas variables.

Cada uno de los modelos presentados en la Tabla 6-2, ha sido evaluado utilizando los índices de error descritos en el capítulo de

Metodología. A continuación, se presentan gráficos de barras que ilustran los valores obtenidos para el Coeficiente de determinación (R^2), la Raíz del error cuadrático medio (RMSE) y la Raíz del error cuadrático medio (RMSE) en cada uno de los modelos analizados. Estos indicadores permiten comparar el desempeño de los modelos y determinar cuál ofrece una mejor precisión en el pronóstico del caudal.

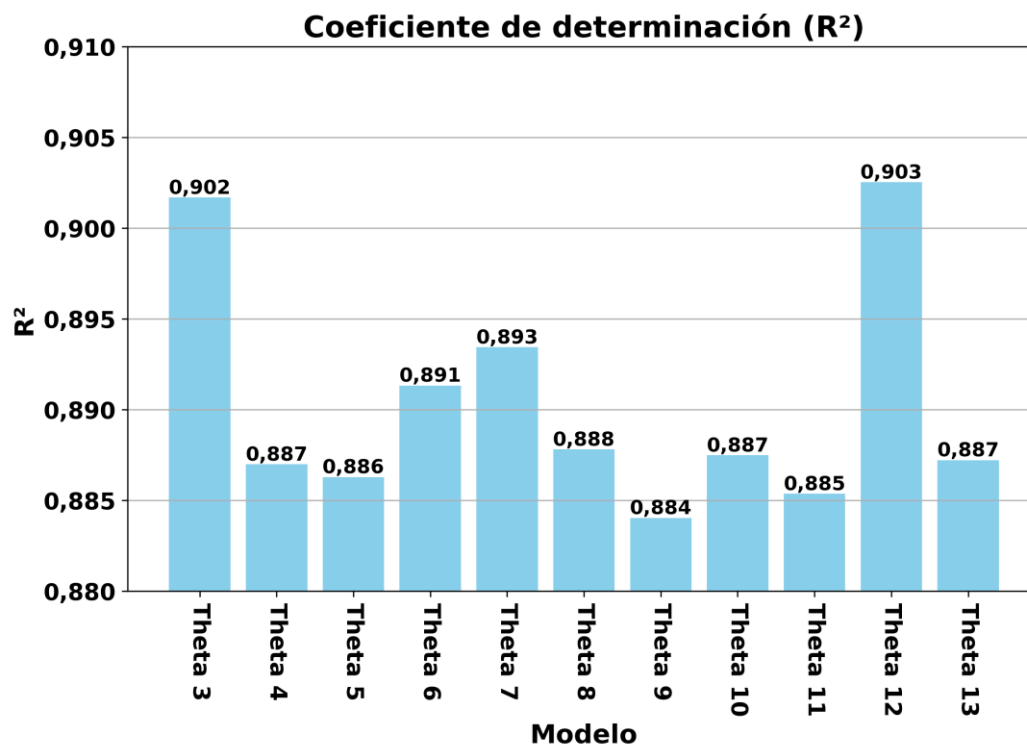


Figura 6-3: Coeficiente de Determinación (R^2) de los modelos óptimos de la combinación Precipitación EP + Temperatura EP.

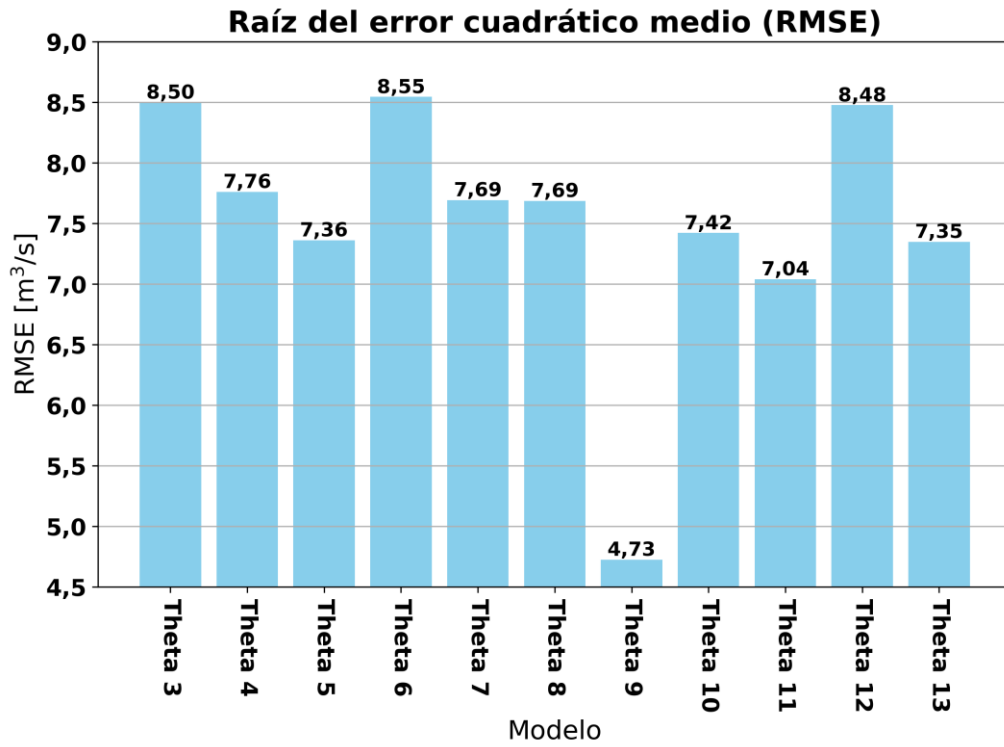


Figura 6-4: Raíz del error cuadrático medio (RMSE) de los modelos óptimos de la combinación Precipitación EP + Temperatura EP.

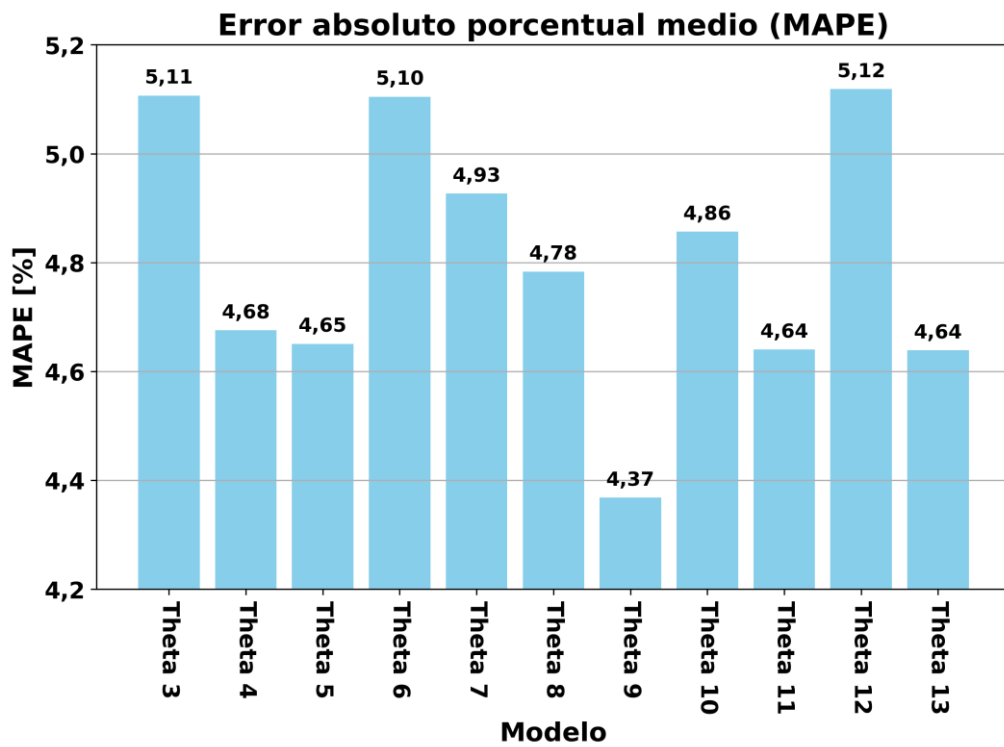


Figura 6-5: Error absoluto porcentual medio (MAPE) de los modelos óptimos de la combinación
Precipitación EP + Temperatura EP.

En la Figura 6-3, se observa que los modelos Theta 3 y Theta 12 presentan los valores más altos del Coeficiente de determinación R^2 lo que indica que estos modelos explican mejor la variabilidad de los datos. Sin embargo, la variación entre los valores de R^2 es mínima, siendo el modelo Theta 9 el que presenta el menor valor.

Por otro lado, en la Figura 6-4;**Error! No se encuentra el origen de la referencia.** y Figura 6-5, se evidencia que los modelos Theta 3, Theta 6 y Theta 12 presentan los valores más altos de Raíz del Error Cuadrático Medio (RMSE) y de Error Absoluto Porcentual Medio (MAPE). Esto indica que estos modelos presentan un mayor error en las predicciones, lo que implica un menor ajuste a los datos observados. En cambio, el modelo Theta 9 exhibe los valores más bajos de RMSE y MAPE, lo que sugiere que este modelo tiene el mejor ajuste a los datos y la mayor precisión en el pronóstico de caudal horario.

Por lo tanto, para la combinación Precipitación EP + Temperatura EP, el modelo Theta 9 es el que mejor desempeño tiene en base a la representatividad de la variabilidad de los datos y en la precisión en las predicciones del caudal.

A modo de ejemplo en el desempeño del modelo Theta 9, en la siguiente gráfica se presenta la comparación entre los valores reales del caudal y el pronóstico generado con el modelo Theta 9 para el Evento 5. Esta evaluación permite analizar la capacidad del modelo para capturar las variaciones del caudal en función del tiempo y determinar su precisión a través de los índices de error, como el MAPE, que en este caso fue de 5,11%, tal como se describe en el capítulo

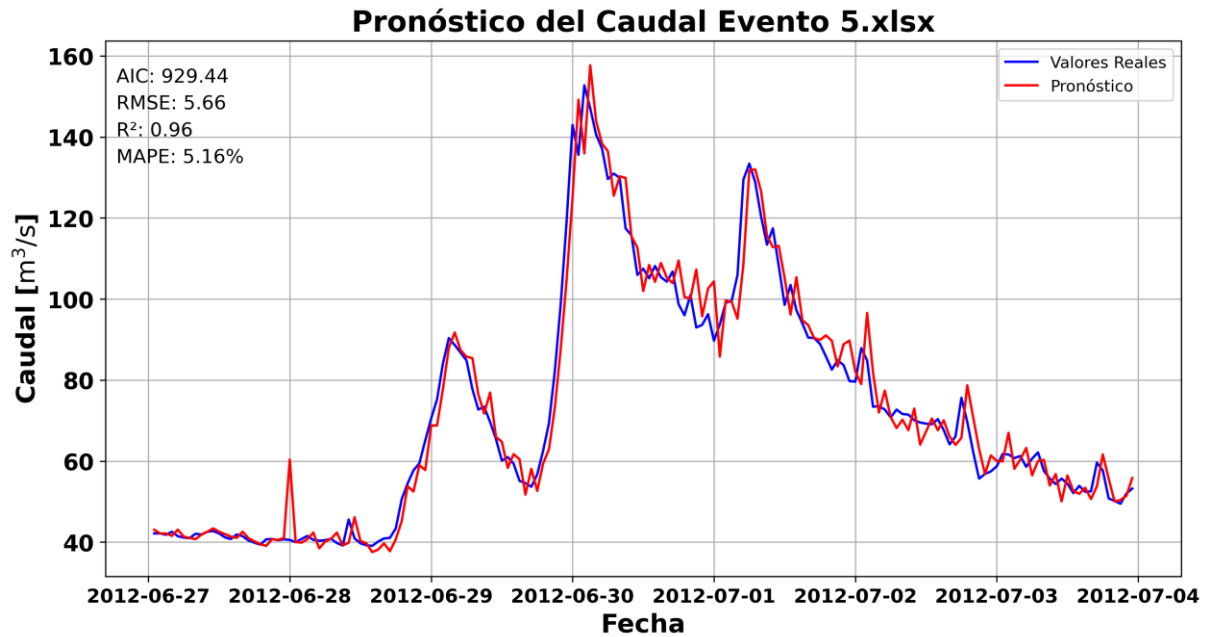


Figura 6-6: Pronóstico de caudal para el Evento 5 utilizando el modelo Theta 9 de la combinación Precipitación EP + Temperatura EP.

Se observa en la Figura 6-6 que existe un alto grado de correlación entre los valores reales y los pronosticados, lo cual se refleja con un valor de 0,96 del Coeficiente de Determinación (R^2). Además, se observa un valor de Raíz del error cuadrático medio (RMSE) y Error absoluto porcentual medio (MAPE) de 5,66 y 5,16% respectivamente.

6.3 Combinaciones de predictores

El procedimiento descrito en la sección Combinación Precipitación EP + Temperatura EP se aplica de manera similar a cada una de las combinaciones de predictores con el objetivo de determinar el modelo que mejor representa el caudal de la cuenca del río Maipo.

A continuación, se presentan los Boxplots del Coeficiente de determinación (R^2), la Raíz del Error Cuadrático Medio (RMSE) y el Error Absoluto Porcentual Medio (MAPE) para cada uno de los modelos analizados.

Es importante destacar que los resultados generales se dividen en dos grupos, debido a que la inclusión de variables provenientes de estaciones aguas arriba introduce variaciones significativas en los valores obtenidos. Para facilitar la interpretación de los resultados, se han establecido los siguientes grupos:

- **Grupo 1 de combinaciones:** Combinaciones que incluyen solo precipitación y temperatura de la estación Patrón. Dichas combinaciones corresponden desde la combinación 1 hasta la 3 de la Tabla 5-5.
- **Grupo 2 de combinaciones:** Combinaciones que incluyen variables predictoras de la estación aguas arriba. Dichas combinaciones corresponden desde la combinación 4 hasta la 10 de la Tabla 5-5.

6.3.1 Resultados Coeficiente de determinación (R^2):

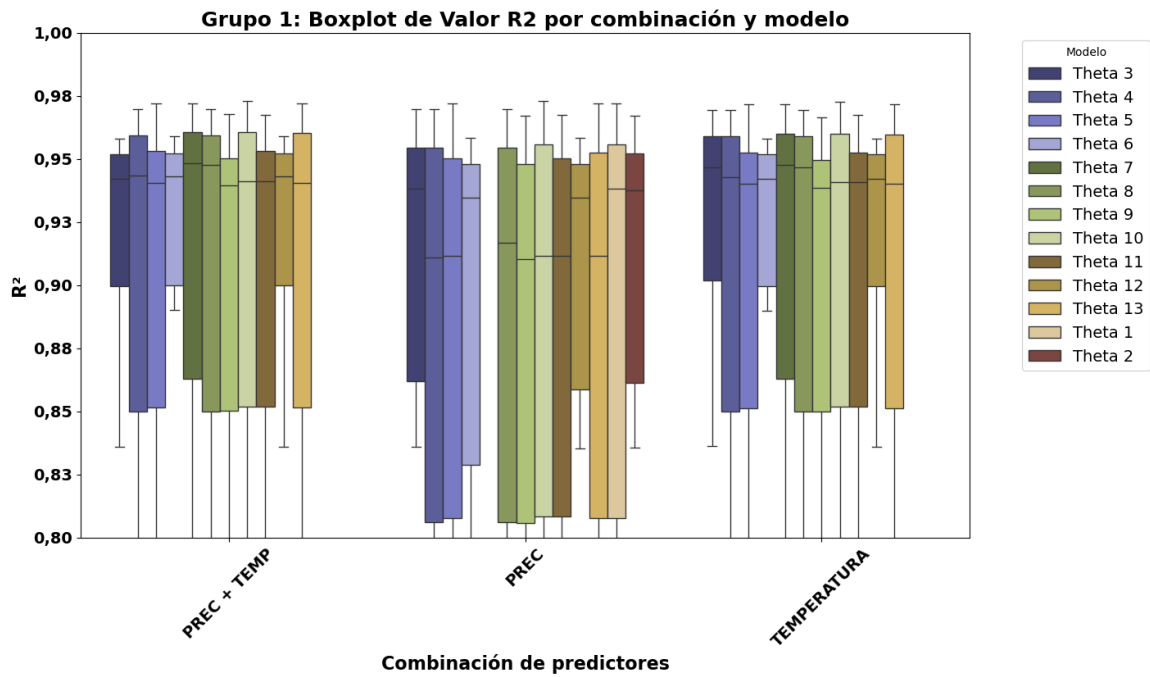


Figura 6-7: Valores del Coeficiente de Determinación (R^2) de cada modelo para el grupo 1 de combinaciones de predictores.

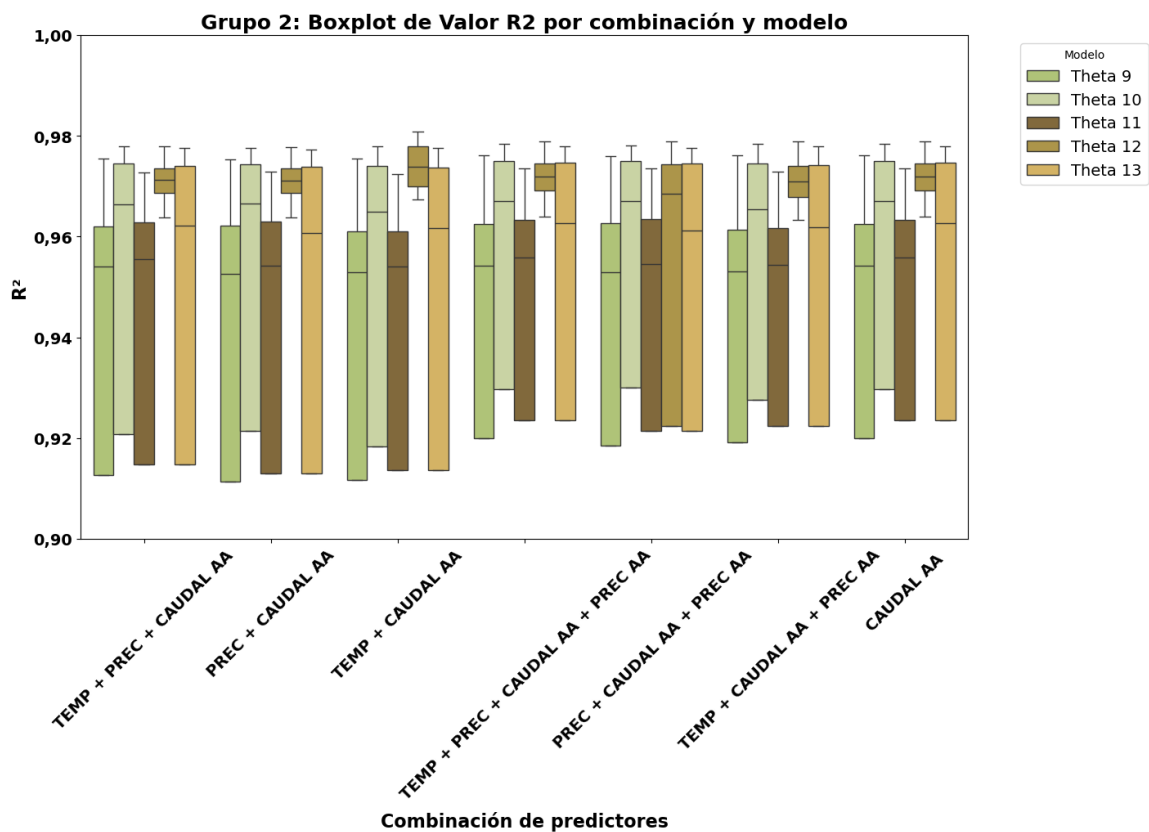


Figura 6-8: Valores del Coeficiente de Determinación (R^2) de cada modelo para el grupo 2 de combinaciones de predictores.

A partir de la Figura 6-7 y Figura 6-8, se observa que las combinaciones pertenecientes al Grupo 2 presentan valores más elevados del Coeficiente de Determinación (R^2), en comparación con las del Grupo 1. Esto indica que los modelos logran una mejor representación del caudal cuando se incluyen variables predictoras provenientes de la estación aguas arriba.

En particular, el modelo Theta 12 se destaca por presentar el mejor desempeño entre todos los modelos evaluados. Además, la combinación "Temperatura + Caudal AA " alcanza el valor más alto de R^2 promedio = 0,974, lo que indica una excelente capacidad de ajuste a los datos observados.

6.3.2 Resultados Raíz del Error Cuadrático Medio (RMSE):

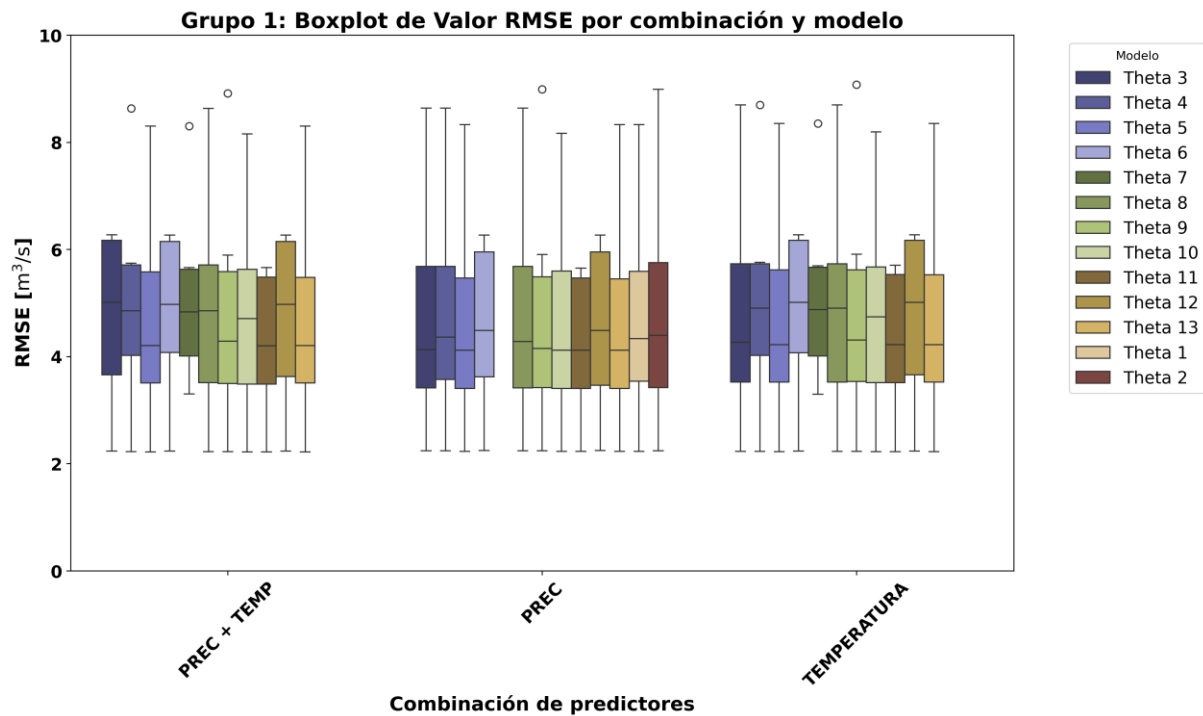


Figura 6-9: Valores de la Raíz del error cuadrático medio (RMSE) de cada modelo para el grupo 1 de combinaciones de predictores.

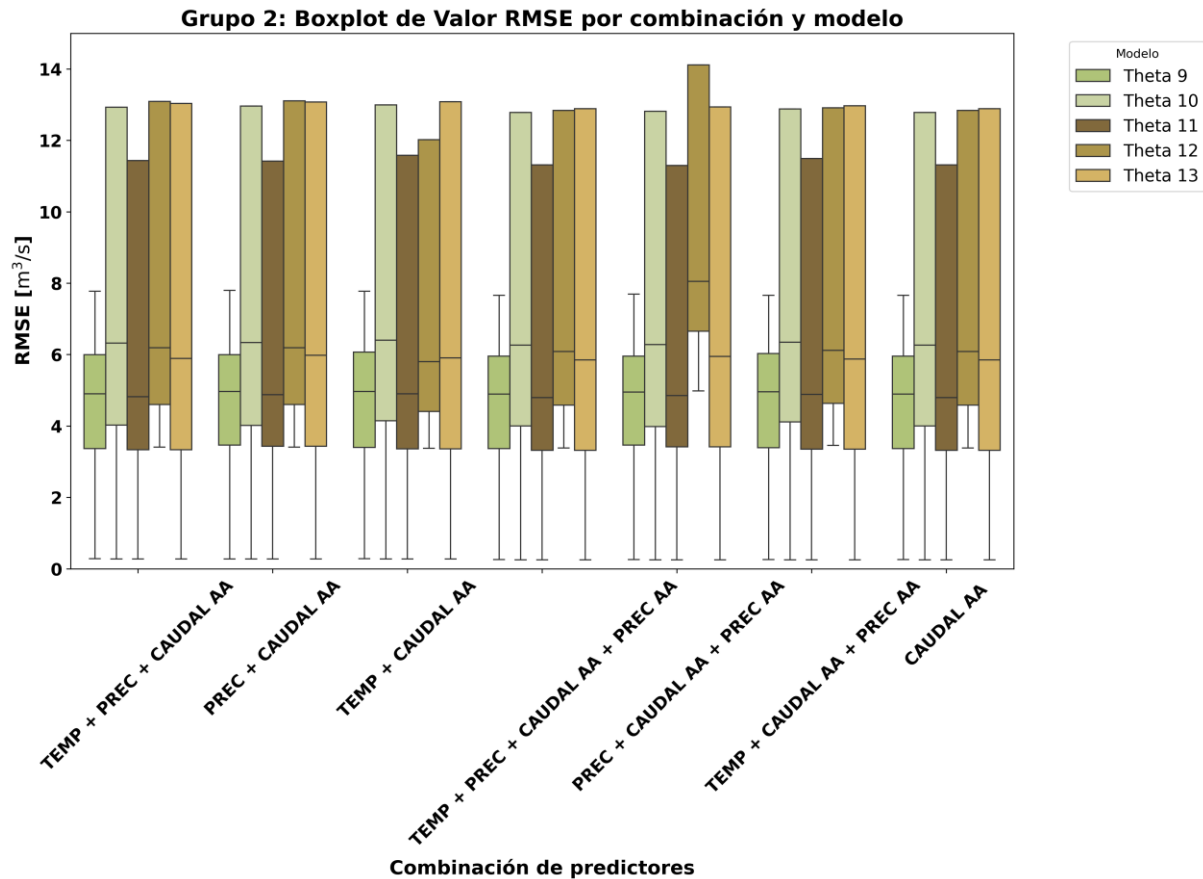


Figura 6-10: Valores de la Raíz del error cuadrático medio (RMSE) de cada modelo para el grupo 2 de combinaciones de predictores.

A partir de la Figura 6-9 y la Figura 6-10, se observa que las combinaciones pertenecientes al Grupo 2 presentan valores más elevados de la Raíz del Error Cuadrático Medio (RMSE) en comparación con las del Grupo 1, lo que indica un mayor error en el pronóstico cuando se incluyen variables predictoras provenientes de la estación aguas arriba.

Sin embargo, se destaca que el modelo Theta 9 en el Grupo 2 exhibe un comportamiento similar al de los modelos del Grupo 1, con valores de RMSE comparables. Esto sugiere que, aunque la incorporación de caudales aguas arriba tiende a aumentar la magnitud del error absoluto en la predicción, el modelo Theta 9 logra mantener una estabilidad en su desempeño.

6.3.3 Resultados Error Absoluto Porcentual Medio (MAPE):

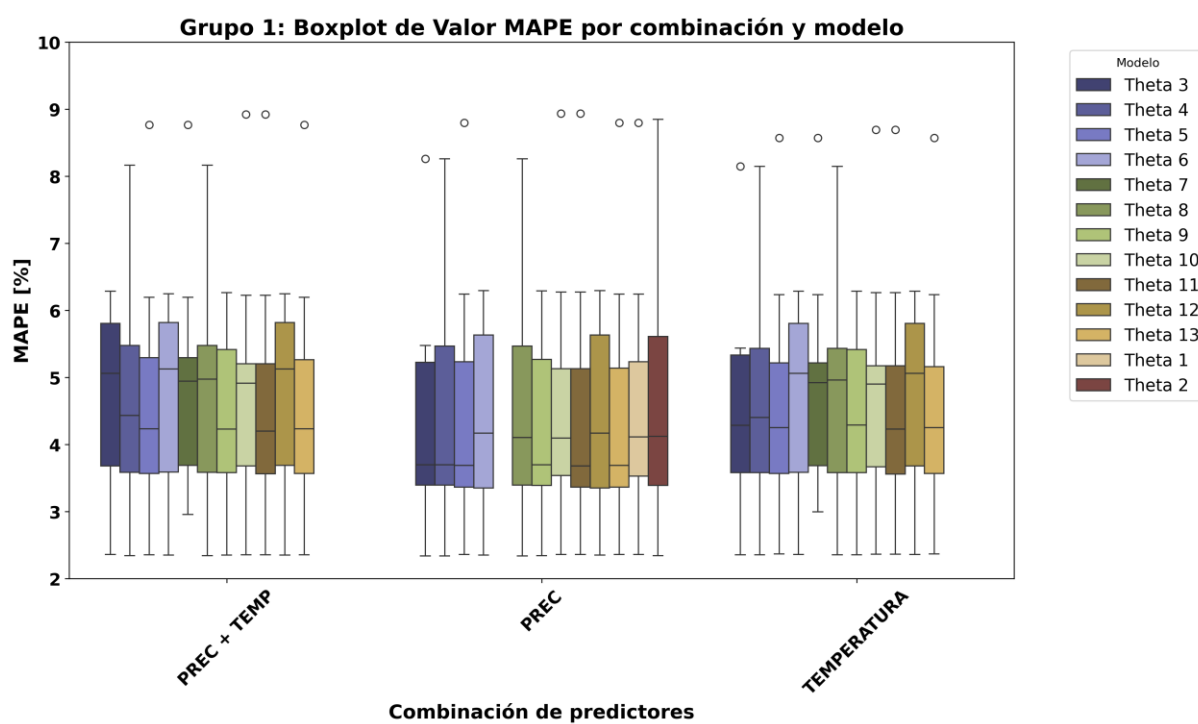


Figura 6-11: Valores Error absoluto porcentual medio (MAPE) de cada modelo para el grupo 1 de combinaciones de predictores.

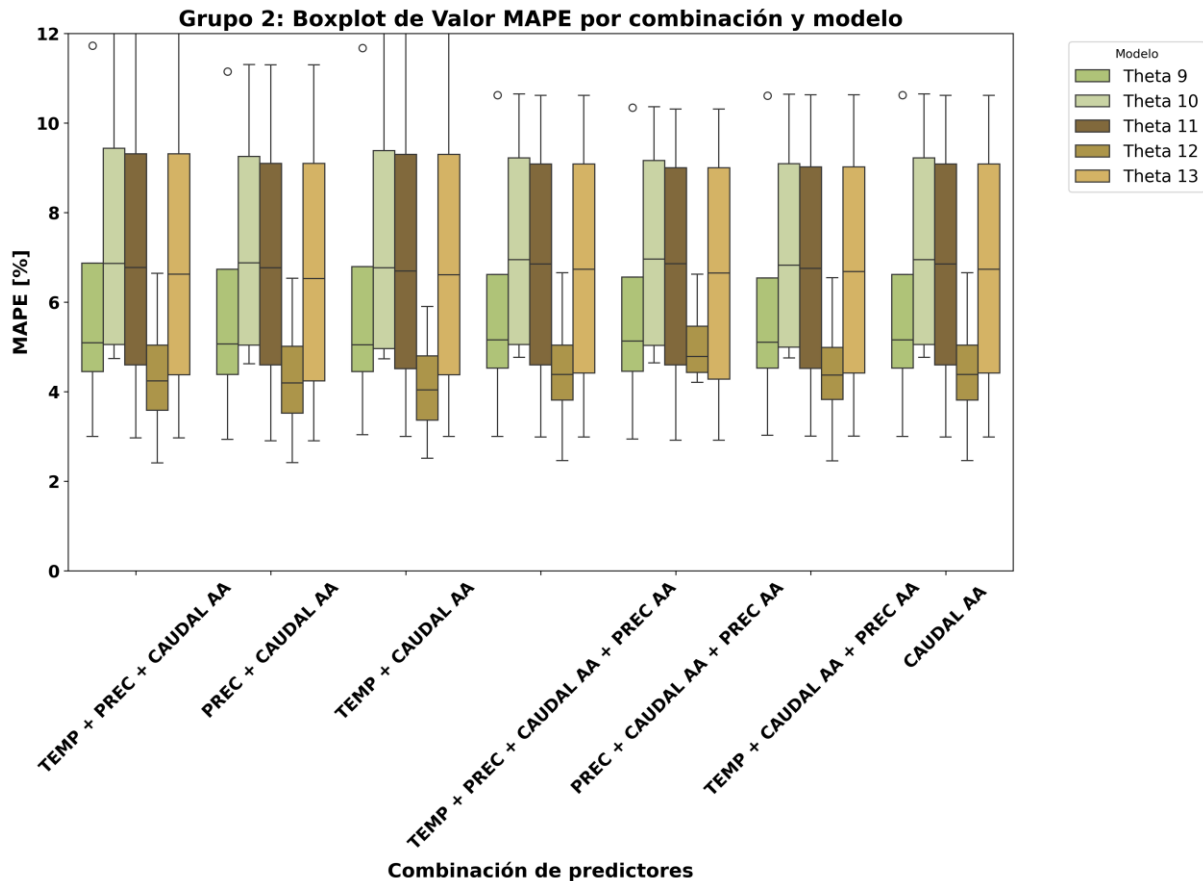


Figura 6-12: Valores Error absoluto porcentual medio (MAPE) de cada modelo para el grupo 2 de combinaciones de predictores.

A partir de la Figura 6-11 y Figura 6-12, se observa que las combinaciones que involucran predictores de la estación aguas arriba (Grupo 2 de combinaciones) presentan valores más elevados del Error absoluto porcentual medio (MAPE) en comparación con las del Grupo 1, lo que indica un mayor error en el pronóstico cuando se incluyen variables predictoras provenientes de la estación aguas arriba. Lo anterior, sigue el mismo comportamiento que para la Raíz del error cuadrático medio (RMSE).

Sin embargo, se destaca que el modelo Theta 12 en el Grupo 2 exhibe un comportamiento similar al de los modelos del Grupo 1, con valores de RMSE comparables. Esto sugiere que, aunque la incorporación de caudales aguas arriba tiende a aumentar la magnitud del error absoluto en la predicción, el modelo Theta 12 logra mantener una estabilidad en su desempeño.

En particular, el modelo Theta 12 de la combinación “Temperatura + Caudal AA” es el que presenta el menor valor de la Raíz del error cuadrático medio con $RMSE = 4,1 \%$. Dicho valor corresponde al promedio del RMSE de cada una de las crecidas pronosticadas.

A partir de los resultados obtenidos, se concluye que el modelo Theta 12, correspondiente a la combinación de variables “Temperatura + Caudal AA”, es el que mejor representa el comportamiento del caudal en el río Maipo. En función de su desempeño, este modelo es considerado como el óptimo para la predicción del caudal. La ecuación y parámetros que lo define son los siguientes:

$$Theta\ 12 = \theta_{12} = (1, 1, 1)(0, 1, 1, 24)$$

$$\hat{Q}_t = 1.772 \cdot Q_{t-1} + Q_{t-24} - Q_{t-25} + u_t - 0.523u_{t-1} - 0.999u_{t-24} + 6.679 \cdot X_{1,t-1} - 0.140 \cdot X_{2,t-1} \quad (11)$$

Donde:

- $\hat{Q}_t$: Valor de caudal pronosticado para el tiempo t.
- Q_{t-1} : Valor de caudal en el tiempo anterior (1 hora antes).
- Q_{t-24} : Valor de caudal en el periodo anterior (24 horas antes).
- Q_{t-25} : Valor de caudal en el tiempo anterior del periodo anterior (25 horas antes).
- u_t : Ruido blanco o error en el tiempo t.
- u_{t-24} : Ruido blanco o error en el periodo anterior (24 horas antes).
- $X_{1,t-1}$: Caudal AA en el tiempo t - 1 [m^3/s].
- $X_{2,t-1}$: Temperatura en el tiempo t - 1 [$^{\circ}C$].

Los pronósticos realizados por el modelo optimo Theta 12 junto con sus índices de error se encuentran en el Anexo 10.2 de este documento.

6.4 Validación Modelo

La validación de modelos predictivos es un proceso fundamental para asegurar la precisión y fiabilidad de las proyecciones generadas. En este caso, se valida el modelo óptimo Theta 12 correspondiente a la combinación “Temperatura EP + Caudal AA” mediante la comparación de los pronósticos con los valores reales. A través de un análisis exhaustivo de los eventos 14, 15 y 16, se busca evaluar la capacidad del modelo para predecir correctamente los máximos registrados y las variaciones del caudal a lo largo del tiempo. Se presenta una evaluación de desempeño mediante indicadores estadísticos detallados en la sección

Metodología de este documento. Estos indicadores son cruciales para determinar la robustez del modelo en distintos escenarios y su capacidad para replicar las dinámicas de caudal observadas en los ríos de la cuenca bajo condiciones variables.

A continuación, se detallan los resultados de la validación utilizando los datos correspondientes a los eventos 14, 15 y 16, los cuales se encuentran registrados en la Tabla 5-6 y se ilustran en las gráficas de los pronósticos de caudal.

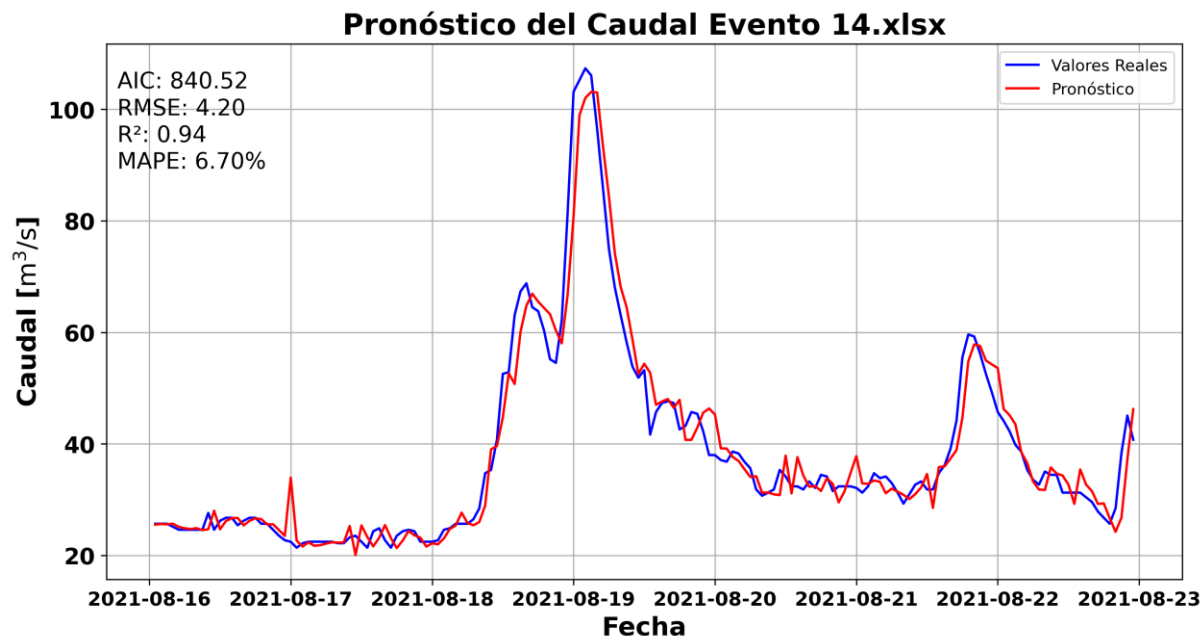


Figura 6-13: Validación Modelo Theta 12 de la combinación Temperatura + Caudal AA con evento 14.

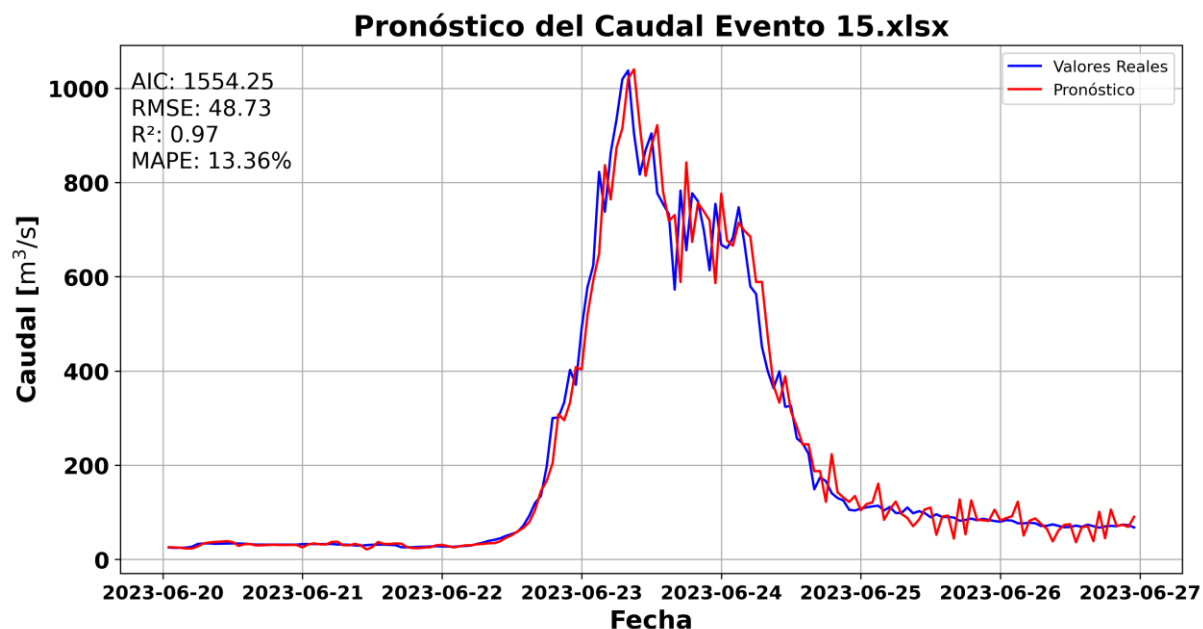


Figura 6-14: Validación Modelo Theta 12 de la combinación Temperatura + Caudal AA con evento 15.

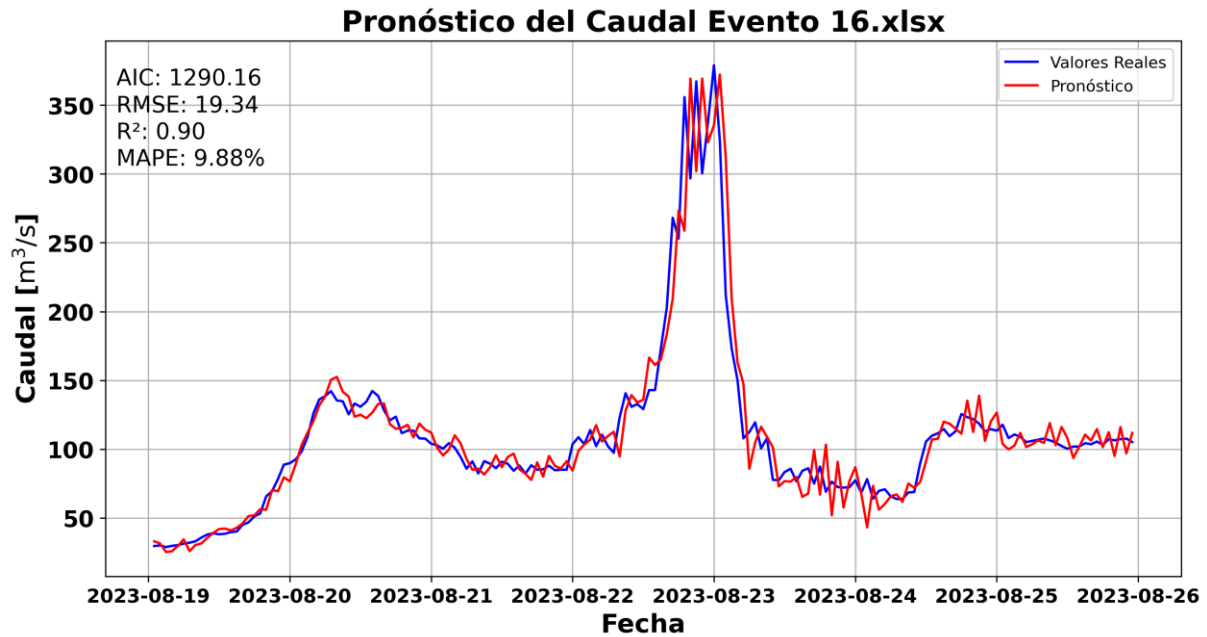


Figura 6-15: Validación Modelo Theta 12 de la combinación Temperatura + Caudal AA con evento 16.

Se observa en la Figura 6-13, Figura 6-14 y Figura 6-15 que el modelo Theta 12 de la combinación Temperatura + Caudal AA, presenta un valor de coeficiente de determinación (R^2) de 0,96, 0,97 y 0,91 para los Eventos 14, 15 y 16 respectivamente. Además, se aprecia que el modelo es capaz de representar correctamente los picos de caudal en su tiempo y forma. Dado lo anterior se indica que el modelo es capaz de reproducir el caudal observado en el río Maipo en el Manzano.

6.5 Tiempos de Adelanto

A continuación, se presentan las métricas de error correspondientes a cada uno de los Tiempos de adelanto (1 a 6 horas). Se generan graficas de línea y sombra para comprender la evolución de los índices de error a medida que se aumenta el tiempo de adelanto en el pronóstico. Cabe destacar que el análisis se realiza para los Eventos 9 al 13, ya que la combinación que presenta el modelo óptimo “Caudal AA + Temperatura” dispone de datos únicamente para dichos eventos.

Las Figura 6-16, Figura 6-17, Figura 6-18 y Figura 6-19 presentan la evolución de diversas métricas de desempeño del modelo de pronóstico en función del Lead Time para los eventos 9 al 13. La Figura 6-18 muestra la evolución del MAPE, destacando la estabilidad en los primeros tiempos de adelanto, mientras que la Figura 6-19 analiza el error en la predicción de los peaks de caudal. Figura 6-16 ilustra la evolución del coeficiente de determinación R^2 , que se mantiene bastante constante, y la Figura 6-17 presenta el RMSE, con algunas fluctuaciones en los primeros tiempos de adelanto. En todas las figuras se incluyen tanto el rango de error (mínimo y máximo) como el valor promedio, proporcionando una visión integral del desempeño del modelo.

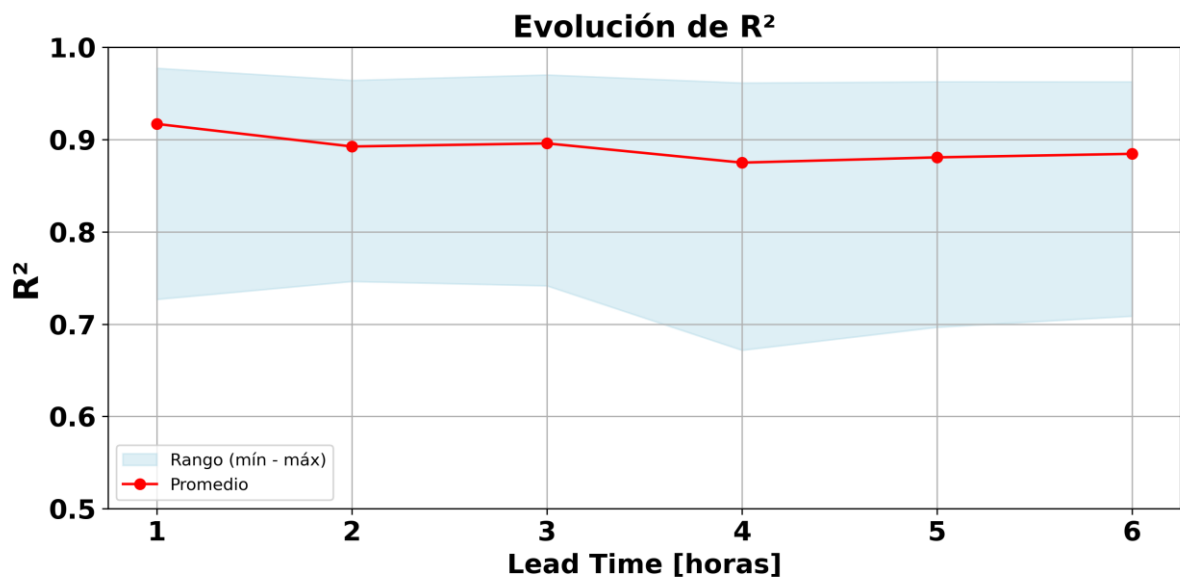


Figura 6-16: Evolución de R^2 en función del Lead Time para los Eventos 9 al 13.

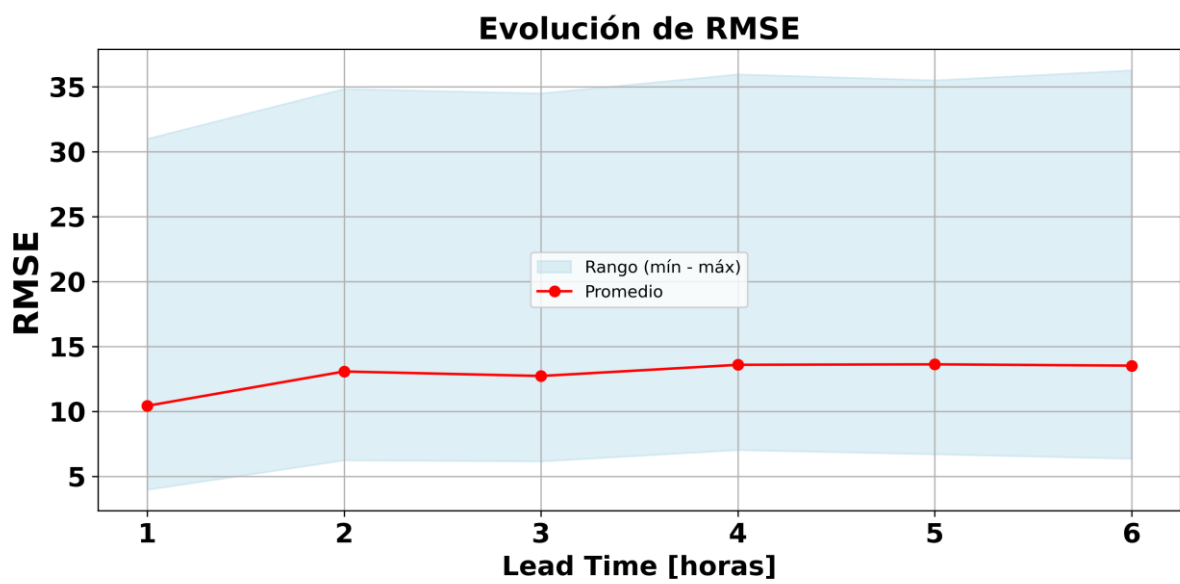


Figura 6-17: Evolución de RMSE en función del Lead Time para los Eventos 9 al 13.

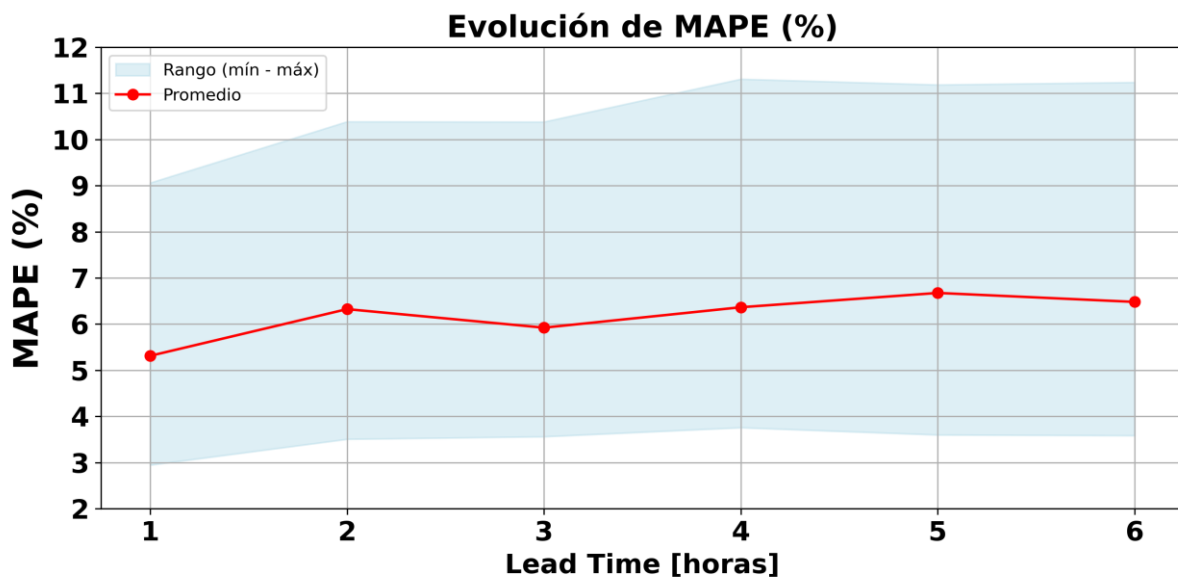


Figura 6-18: Evolución de MAPE en función del Lead Time para los Eventos 9 al 13.

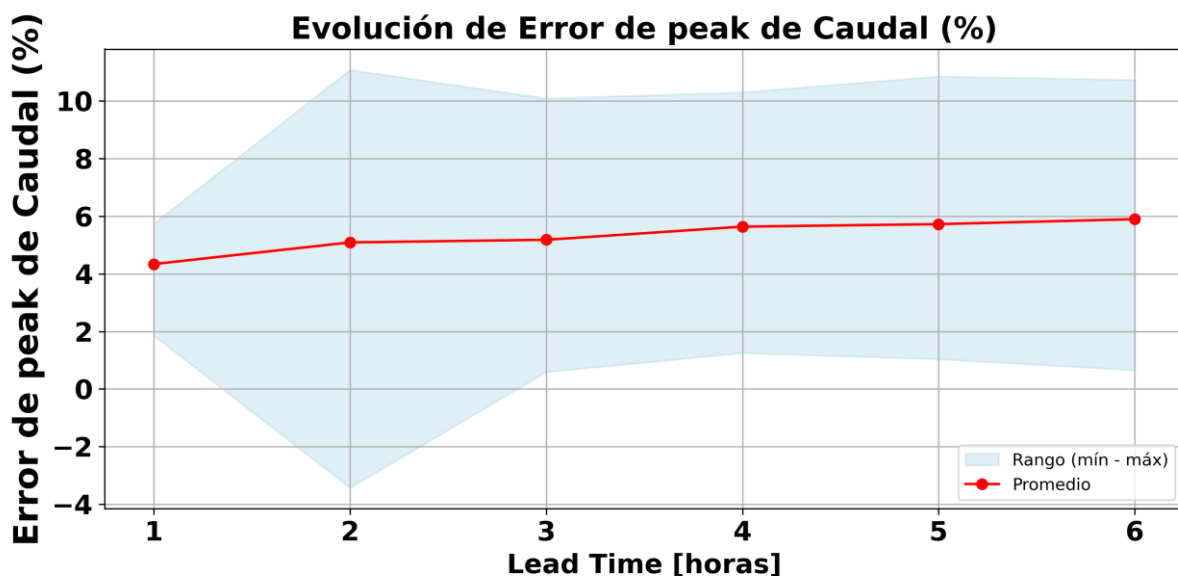


Figura 6-19: Evolución de Error de peak de caudal en función del Lead Time para los Eventos 9 al 13.

En la Figura 6-16 se observa que el coeficiente de determinación (R^2) se mantiene alto en todos los lead times, con valores promedio superiores a 0,8. Aunque se aprecia una leve disminución a medida que aumenta el lead time, los valores siguen siendo elevados, lo que indica que el modelo conserva una buena capacidad explicativa incluso hasta 6 horas de pronóstico.

Asimismo, la raíz del error cuadrático medio (RMSE) presenta un aumento progresivo con el lead time, con una pendiente moderada. No obstante, el valor promedio se mantiene por debajo del umbral de 15, lo que refleja una adecuada precisión del modelo en términos absolutos.

El error absoluto porcentual medio (MAPE) permanece prácticamente constante a lo largo del tiempo, con valores promedio entre 5 % y 8 %. Este comportamiento indica que el modelo mantiene un buen

desempeño relativo conforme aumentan los lead times. Se observa, además, que los valores mínimos y máximos del MAPE tienden a aumentar en el desempeño de los pronósticos para cierto tipo de tormentas.

Finalmente, el error en el valor del peak de caudal muestra que los errores promedio se mantienen entre 4 % y 6 %, con un leve aumento conforme se incrementa el lead time, aunque sin incrementos significativos. Además, se observa que para el Lead Times 2 presenta un valor mínimo de -3 aproximadamente, lo cual indica que el caudal real es mayor al caudal simulado. Dicho pronóstico corresponde al realizado utilizando el Evento 12 y el cual queda registrado en el Anexo 10.3 de este informe.

Este buen desempeño del modelo, reflejado en los altos valores de R^2 y en los errores relativamente bajos y estables a lo largo de los lead times, se encuentra en concordancia con lo observado en los análisis de autocorrelación (ACF) y autocorrelación parcial (PACF) presentados previamente (Figura 6-1 y Figura 6-2 respectivamente). En dichos análisis se evidenció una fuerte dependencia temporal de corto plazo y una alta memoria en los datos horarios. Esta memoria temporal favorece la capacidad del modelo para realizar pronósticos precisos en horizontes de hasta 6 horas, ya que la información contenida en los valores recientes del caudal proporciona una base sólida para generar pronósticos con buenos índices de error.

Dado el buen desempeño observado en el pronóstico, se decidió validar los lead times 1, 3 y 5 utilizando, a modo de ejemplo, el Evento 9. Esta validación se llevó a cabo replicando manualmente el pronóstico mediante la ecuación correspondiente a cada uno de los modelos SARIMAX en su respectivo lead time. Para ello, se emplearon 100 combinaciones de ruido blanco (u_t) con el fin de probar distintas alternativas y evaluar la robustez de los resultados.

6.5.1 Validación Lead Times 1:

Para el lead time de 1 hora, el modelo seleccionado corresponde a un SARIMAX con parámetros (p, d, q) = (1, 2, 2) y (P, D, Q, s) = (0, 1, 1, 24). El desempeño del modelo, calculado durante la validación con el Evento 9, arrojó los siguientes resultados:

- R^2 : 0,972
- RMSE: 30,9994
- MAPE: 9,06 %
- AIC: 1150,53
- Error en el peak de caudal: 4,33 %

A continuación, se presenta la ecuación que define al modelo:

$$Q_t = 0,54 \cdot Q_{t-1} - 1,12 \cdot u_{t-1} + 0,11 \cdot u_{t-2} - 0,99 \cdot u_{t-24} - 0,03 \cdot X_{1,t-1} + 3,95 \cdot X_{2,t-1} + u_t$$

Donde:

- Q_t : Valor de caudal pronosticado para el tiempo t.
- Q_{t-1} : Valor de caudal real en el tiempo anterior (1 hora antes).
- u_t : Ruido blanco o error en el tiempo t.
- u_{t-1} : Ruido blanco o error en el tiempo anterior anterior (1 hora antes).
- u_{t-2} : Ruido blanco o error 2 horas antes.
- u_{t-24} : Ruido blanco o error en el periodo anterior (24 horas antes).

- $X_{1,t-1}$: Caudal AA en el tiempo anterior [m^3/s].
- $X_{2,t-1}$: Temperatura en el tiempo anterior [$^{\circ}\text{C}$].

A continuación, se presentan los resultados obtenidos de la validación para el Lead Times 1:

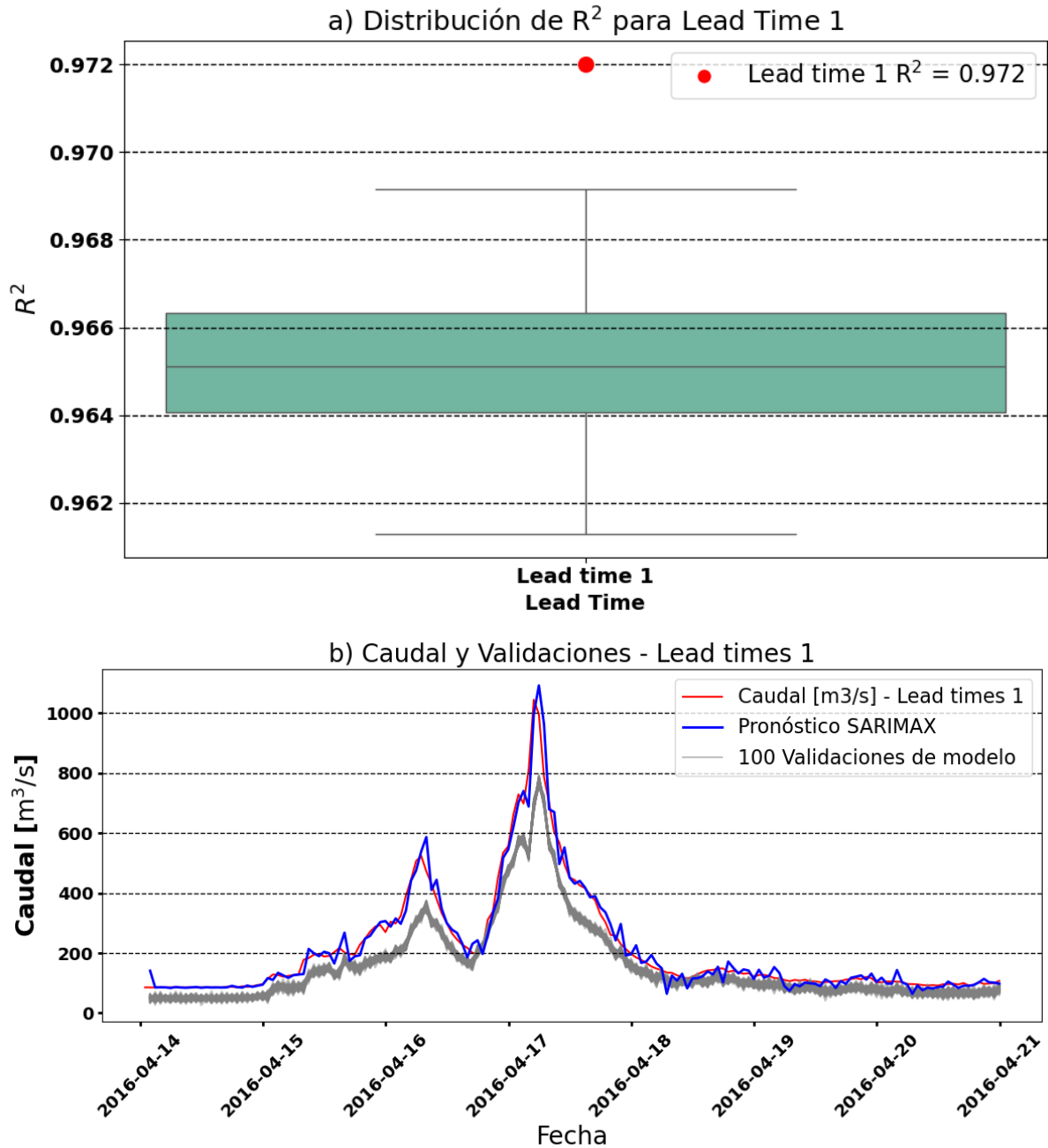


Figura 6-20: Validación del modelo SARIMAX para Lead Time 1: distribución de R^2 y comparación temporal de caudal observado y pronosticado.

En la Figura 6-20, se observa que el Boxplot muestra que los valores de R^2 para las 100 validaciones se sitúan entre 0,964 y 0,966, con valores mínimos y máximos cercanos a 0,962 y 0,970, respectivamente. El valor de R^2 correspondiente al modelo SARIMAX (0,972) se encuentra fuera del rango intercuartílico, lo que indica que el modelo presenta un mejor desempeño en comparación con las

validaciones. Este comportamiento podría estar asociado a distintos factores, entre los cuales destaca la posibilidad de que la librería utilizada en Python emplee valores optimizados para el ruido blanco, lo que resulta en un modelo de pronóstico muy preciso.

Además, al analizar las 100 validaciones (área gris), se observa que el modelo logra capturar adecuadamente la forma y magnitud de los picos de caudal, lo que resalta su capacidad para predecir eventos extremos de manera efectiva.

6.5.2 Validación Lead Times 3:

Para el lead time de 3 horas, el modelo seleccionado corresponde a un SARIMAX con parámetros $(p, d, q) = (0, 2, 2)$ y $(P, D, Q, s) = (0, 1, 1, 24)$. El desempeño del modelo, calculado durante la validación con el Evento 9, arrojó los siguientes resultados:

- R^2 : 0,964
- RMSE: 35,9666
- MAPE: 11,3102 %
- AIC: 1096,73
- Error en el peak de caudal: 10,298 %

A continuación, se presenta la ecuación que define al modelo:

$$Q_t = 0,0005 \cdot u_{t-1} - 0,9995 \cdot u_{t-2} - 0,9989 \cdot u_{t-24} - 0,14 \cdot \hat{Q}_{t-1} + u_t$$

Donde:

- Q_t : Valor de caudal pronosticado para el tiempo t.
- $\hat{Q}_{t-1}$: Valor de caudal pronosticado en el tiempo anterior (1 hora antes).
- u_t : Ruido blanco o error en el tiempo t.
- u_{t-1} : Ruido blanco o error en el tiempo anterior (1 hora antes).
- u_{t-2} : Ruido blanco o error 2 horas antes.
- u_{t-24} : Ruido blanco o error en el periodo anterior (24 horas antes).

A continuación, se presentan los resultados obtenidos de la validación para el Lead Times 3:

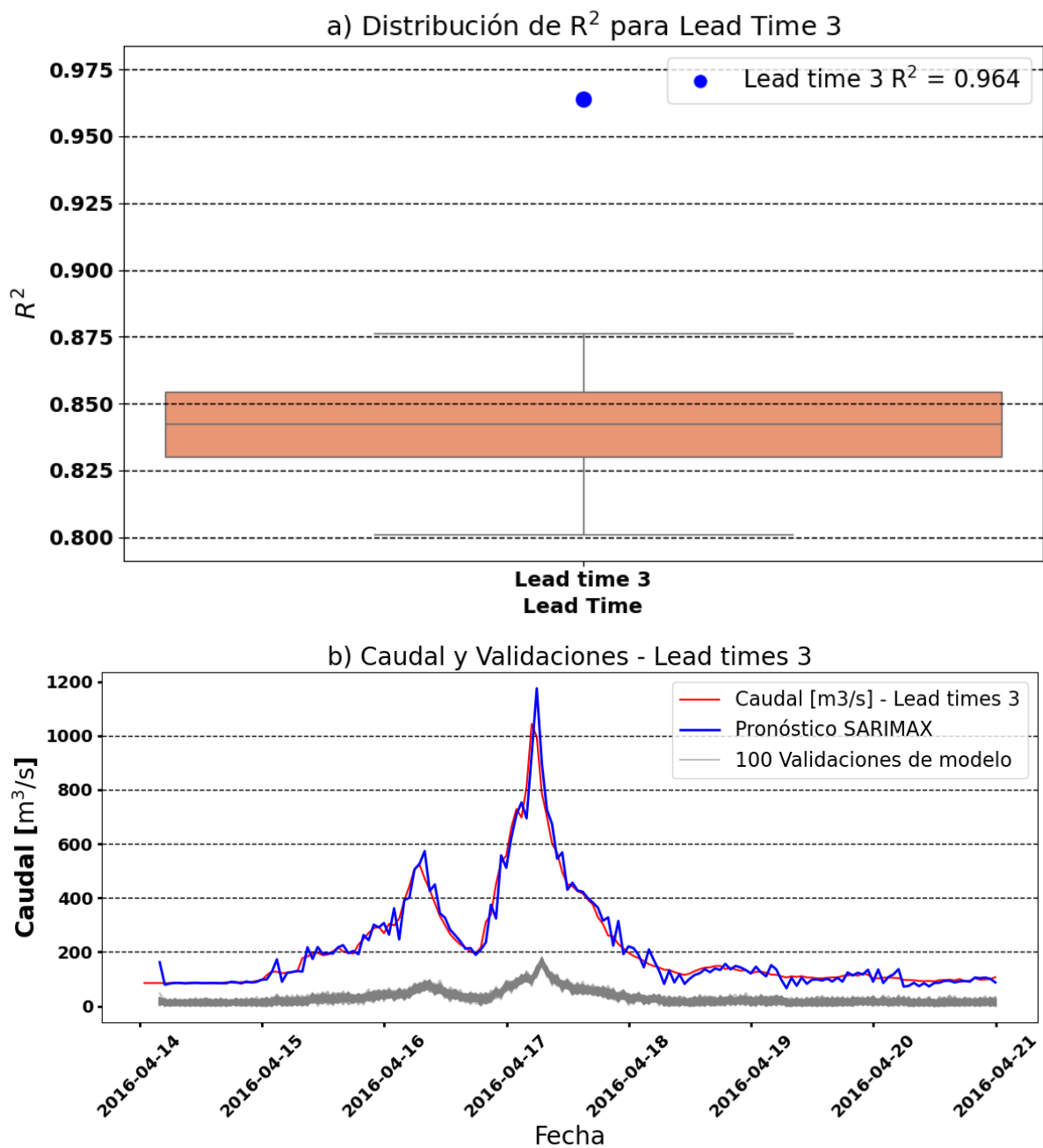


Figura 6-21: Validación del modelo SARIMAX para Lead Time 3: distribución de R^2 y comparación temporal de caudal observado y pronosticado.

En la Figura 6-21, se observa que el Boxplot para el lead time 3 presenta una distribución con valores de R^2 más bajos en comparación con el lead time 1, con un rango que varía aproximadamente entre 0,825 y 0,875.

El valor de R^2 para el modelo SARIMAX es de 0,964, lo que se muestra como un punto aislado fuera del rango intercuartílico. Esto indica que, en el modelo implementado en Python, se obtienen mejores valores de R^2 en comparación con las validaciones manuales.

Además, el rango de las 100 validaciones de modelo (área gris) muestra una mayor dispersión en comparación con el lead time 1. Aunque los picos de caudal son reflejados correctamente, se observa

un mayor desajuste en el valor del peak de caudal. Esto sugiere que, si bien el modelo sigue capturando la tendencia general del caudal, presenta dificultades para predecir con precisión los valores extremos.

6.5.3 Validación Lead Times 5:

Para el lead time de 5 horas, el modelo seleccionado corresponde a un SARIMAX con parámetros $(p, d, q) = (0, 2, 2)$ y $(P, D, Q, s) = (0, 1, 1, 24)$. El desempeño del modelo, calculado durante la validación con el Evento 9, arrojó los siguientes resultados:

- R^2 : 0,962
- RMSE: 36,2758
- MAPE: 11,2402 %
- AIC: 1057,49
- Error en el peak de caudal: 10,7322 %

A continuación, se presenta la ecuación que define al modelo:

$$Q_t = -0,0009 \cdot u_{t-1} - 0,9988 \cdot u_{t-2} - 0,9989 \cdot u_{t-24} - 0,1471 \cdot \hat{Q}_{t-1} + u_t$$

Donde:

- Q_t : Valor de caudal pronosticado para el tiempo t.
- $\hat{Q}_{t-1}$: Valor de caudal pronosticado en el tiempo anterior (1 hora antes).
- u_t : Ruido blanco o error en el tiempo t.
- u_{t-1} : Ruido blanco o error en el tiempo anterior (1 hora antes).
- u_{t-2} : Ruido blanco o error 2 horas antes.
- u_{t-24} : Ruido blanco o error en el periodo anterior (24 horas antes).

A continuación, se presentan los resultados obtenidos de la validación para el Lead Times 5:

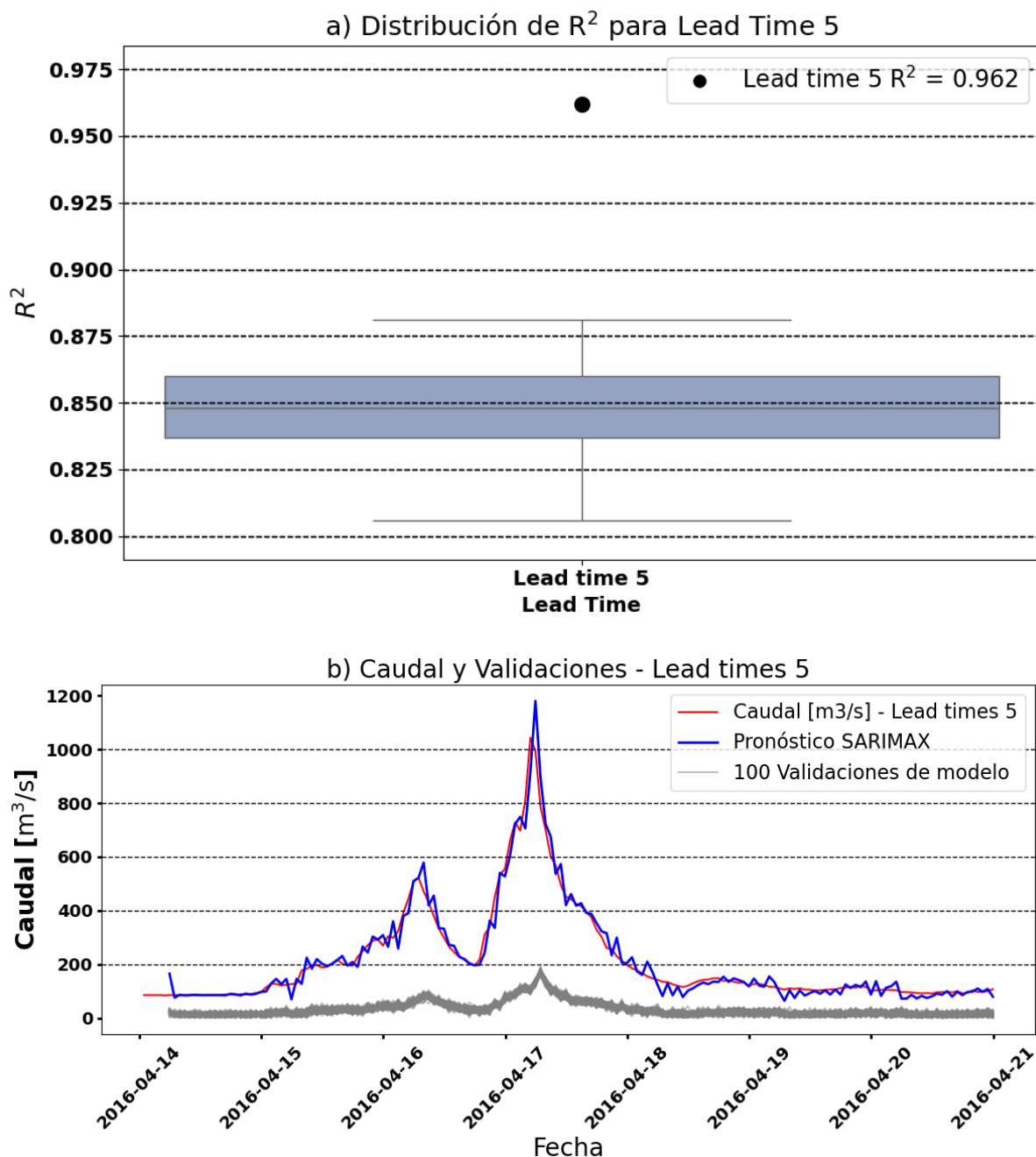


Figura 6-22: Validación del modelo SARIMAX para Lead Time 5: distribución de R^2 y comparación temporal de caudal observado y pronosticado.

En la Figura 6-22 se observa que el Boxplot para el lead time 5 presenta un rango más estrecho de R^2 en comparación con los lead times 1 y 3. Al igual que en los tiempos anteriores, el valor de R^2 se encuentra por encima del rango intercuartílico, con un valor de 0,962. Sin embargo, el R^2 promedio de las 100 validaciones se aproxima a 0,85, lo que es levemente más bajo en este lead time que en los anteriores, aunque sigue siendo un valor altamente aceptable.

Además, se observa que las validaciones no logran representar correctamente los valores reales de caudal, sin embargo, si lo logran replicar el momento en donde se producción los aumentos y descenso de caudal. Lo anterior indica que las validaciones tienen dificultades para capturar con precisión los valores extremos en el horizonte temporal de 5 horas. A pesar de esto, el modelo pronosticado presenta

un alto grado de semejanza con los valores reales. Este comportamiento puede estar relacionado con diversos factores que afectan las validaciones manuales, lo que explica el mejor rendimiento del modelo SARIMAX implementado en Python en comparación con las validaciones basadas en combinaciones aleatorias de ruido blanco.

7 Discusión

El modelo estadístico SARIMAX utilizado en este estudio ha demostrado ser eficaz para la predicción de caudales de crecida en la Cuenca del Río Maipo, específicamente en la estación Río Maipo en El Manzano. Una de las principales fortalezas del modelo SARIMAX es su capacidad para integrar variables exógenas, como la precipitación y el caudal aguas arriba, en su proceso de predicción. Este enfoque ha sido fundamental para mejorar la precisión de los pronósticos, ya que las variables externas tienen un impacto significativo en la dinámica hidrológica de la cuenca. La incorporación de estos predictores ha permitido al modelo capturar de manera efectiva las variaciones climáticas que influyen directamente en los eventos de crecida.

En este sentido, los resultados obtenidos son consistentes con investigaciones previas en otras cuencas, como en el estudio realizado en la cuenca del Río Cuyaguateje (Alonso et al., 2021), donde el modelo SARIMAX también demostró una alta capacidad para integrar información meteorológica y de caudal para mejorar las predicciones. Este enfoque ha sido similar a los estudios realizados en la Cuenca del Plata (Meis et al., 2017), donde se ha empleado el modelo SARIMA para pronosticar el caudal mensual en varias estaciones, obteniendo buenos resultados en la validación de los pronósticos, aunque con algunas desviaciones en eventos extremos. Estos estudios resaltan la efectividad de los modelos estadísticos en el pronóstico de caudales y su utilidad para la gestión de recursos hídricos y la mitigación de inundaciones.

Se obtuvieron valores altamente satisfactorios en cuanto a los coeficientes de determinación (R^2) y errores de pronóstico (RMSE, MAPE), lo que valida la robustez del modelo en el contexto de predicción de crecidas. No obstante, a pesar de los buenos resultados obtenidos en los pronósticos a corto plazo (con tiempos de adelanto de 1 a 3 horas), los resultados muestran un decrecimiento en la precisión a medida que aumentan los tiempos de adelanto, especialmente en las predicciones a 5 y 6 horas. Este fenómeno es esperado en la mayoría de los modelos estadísticos utilizados en predicciones de series temporales, ya que la incertidumbre aumenta con el tiempo. Factores como las fluctuaciones impredecibles en las condiciones climáticas, el cambio de patrones hidrológicos y la incertidumbre inherente en la medición de las variables de entrada contribuyen a la disminución de la precisión. En este sentido, es importante considerar que, aunque la precisión disminuye con el tiempo, los resultados aún se mantienen dentro de un rango aceptable, lo que permite su uso en aplicaciones prácticas, como los sistemas de alerta temprana.

En cuanto a la comparación con los resultados obtenidos por Daniel Pascal en su estudio en el Valle Del Alto Cachapoal (Pascal, 2023) se observa que ambos estudios muestran una disminución de la precisión en los pronósticos a medida que aumenta el tiempo de adelanto (lead time). En el estudio de Pascal, esta disminución es más pronunciada, ya que, a medida que los lead times aumentan, especialmente para los intervalos de 3 y 5 horas, la incertidumbre en las predicciones se incrementa, lo que reduce la capacidad predictiva del modelo.

En contraste con los resultados de Pascal, donde la incertidumbre creció debido a la falta de predictores meteorológicos en los tiempos futuros, este estudio destaca la importancia de las variables predictoras, que pudieron haber mitigado la disminución de la precisión en los tiempos de adelanto más largos (5 y 6 horas). La inclusión de estas variables podría explicar por qué el modelo en este estudio mostró una mayor capacidad de adaptación y una mayor precisión frente a los eventos extremos de crecida, en comparación con los modelos utilizados en el estudio de Pascal.

En resumen, aunque ambos estudios coinciden en que la precisión disminuye con los tiempos de adelanto, este trabajo sugiere que el modelo SARIMAX, al incorporar variables exógenas, tiene una mayor capacidad para mantener la precisión en tiempos de adelanto más largos, lo que refuerza la utilidad del modelo para pronósticos en escenarios de alerta temprana.

En el contexto de los sistemas de alerta temprana, la disminución de precisión a medida que aumenta el tiempo de adelanto es una cuestión crítica, ya que la efectividad de estos sistemas depende de la capacidad para predecir eventos extremos con suficiente antelación. En términos prácticos, mantener una alta precisión en los pronósticos a corto plazo, especialmente en tiempos de adelanto de 1 a 3 horas,

es crucial para la toma de decisiones inmediatas, como la activación de alertas de evacuación o la movilización de recursos. Sin embargo, a medida que el lead time se extiende a 5 o 6 horas, aunque la precisión puede disminuir, sigue siendo útil para proporcionar una previsión general de los posibles eventos, lo que permite a los responsables de la gestión del riesgo tomar decisiones preparatorias, como la actualización de planes de contingencia o la alerta temprana de las autoridades locales.

Una característica importante que se destaca de este modelo es la capacidad para reproducir las tendencias generales del caudal, incluso en situaciones extremas de aumento o descenso abrupto de los valores. Sin embargo, es posible que el modelo SARIMAX implementado en Python sobreestime algunos parámetros, especialmente aquellos relacionados con las condiciones extremas, con el fin de mejorar la precisión de las predicciones. Esta posible sobreestimación podría contribuir a los buenos resultados obtenidos en comparación con los modelos de validación manual, como los que se generan a partir de combinaciones aleatorias de ruido blanco. A pesar de esta posible sobreestimación, el modelo ha mostrado un rendimiento superior, lo que refuerza la idea de que es una herramienta potente capaz de replicar con alta precisión los comportamientos hidrológicos observados en la cuenca. No obstante, sería útil realizar una validación más profunda para evaluar la estabilidad y consistencia de los parámetros en diferentes condiciones y escenarios.

Es relevante señalar que la memoria hidrológica de la cuenca, tal como lo evidencian los análisis de autocorrelación (ACF), juega un papel fundamental en el desempeño del modelo. La fuerte dependencia temporal observada en las series históricas refuerza la importancia de variables como el caudal aguas arriba como predictores robustos para la estimación de futuros caudales, un hallazgo consistente con trabajos previos en otras cuencas. En la Cuenca del Plata (Meis et al., 2017), el estudio ha demostrado que la incorporación de variables como el caudal en estaciones adyacentes mejora significativamente la precisión del pronóstico de caudales mensuales. Los modelos SARIMA empleados en esa región también han destacado la dependencia temporal de los caudales y la utilidad de los datos históricos para predecir eventos extremos, lo que refuerza la relevancia de la memoria hidrológica en los modelos estadísticos para cuencas con características hidrológicas complejas.

8 Conclusiones y recomendaciones

El modelo estadístico SARIMAX ha demostrado ser altamente efectivo en el pronóstico de caudales de crecida en la Cuenca del Río Maipo, específicamente en la estación Río Maipo en El Manzano. La incorporación de variables exógenas ha mejorado significativamente la precisión de los pronósticos, particularmente a corto plazo (1 a 3 horas). Este estudio refuerza la importancia de las variables predictoras al integrar información externa en los modelos estadísticos, lo que permite una mejor representación de la dinámica hidrológica de la cuenca. El modelo óptimo pertenece a la combinación Temperatura y Caudal Aguas Arriba (AA), alcanzando un R^2 promedio de 0,974, un RMSE de 4,12 m³/s y un MAPE de 5,11% en los eventos calibrados, lo que evidencia un excelente ajuste. Lo anterior, marca notoriamente que, a pesar de pronosticar crecidas de caudal debido a precipitaciones intensas, el mejor modelo no utiliza dicha variable en el pronóstico.

Para los tiempos de adelanto (Lead Times) de 1 a 6 horas, el modelo Theta 12 mantuvo su desempeño con una variación moderada de error conforme aumentaba el Lead Time. Por ejemplo, el valor promedio del coeficiente de determinación (R^2) disminuyó de 0,92 en el primer tiempo de adelanto a 0,88 en el sexto tiempo de adelanto, lo que respalda la utilidad del modelo en sistemas de alerta temprana. Este comportamiento refleja la capacidad del modelo para replicar el caudal en la cuenca, lo cual se confirma con la alta memoria de la cuenca, que permite mantener una buena capacidad predictiva. Además, la inclusión del caudal aguas arriba (Volcán en Queltehués) demostró ser una alternativa robusta en la generación de pronósticos, ya que, al integrar esta variable, el modelo puede compensar la falta de datos de precipitación actualizados o mitigar los errores derivados de fallos en el modelo principal. Esta incorporación mejora la precisión del pronóstico, dado que el caudal aguas arriba está fuertemente correlacionado con los flujos aguas abajo, lo que permite una estimación más precisa de los caudales futuros, incluso cuando los datos meteorológicos directos no están disponibles o son incompletos.

La memoria hidrológica observada en la cuenca, evaluada mediante funciones de autocorrelación (ACF), respalda la capacidad predictiva de los modelos SARIMAX. Los análisis de ACF muestran que la autocorrelación se mantiene por sobre 0,8 hasta rezagos de 50 unidades. Esta alta autocorrelación indica que el comportamiento pasado de la serie aporta información relevante para pronosticar valores futuros, lo que justifica el buen desempeño de los modelos SARIMAX aplicados. Además, la elevada memoria de la cuenca valida el uso del caudal aguas arriba (Caudal AA) como variable predictora principal, dado que su comportamiento pasado se traduce en una fuerte capacidad explicativa para estimar los caudales futuros en la estación de interés. Para fortalecer la aplicabilidad del modelo desarrollado, se recomienda aplicar el modelo en cuencas con menor memoria hidrológica para evaluar su desempeño en escenarios donde la persistencia temporal de los caudales es reducida, lo cual permitirá validar la adaptabilidad del enfoque SARIMAX en distintos tipos de cuencas y regímenes hidrológicos.

Finalmente, se recomienda explorar el uso de modelos de inteligencia artificial, como redes neuronales recurrentes (RNN) o Long Short-Term Memory (LSTM), para manejar mejor las no linealidades y estacionalidades complejas que los modelos SARIMAX pueden no capturar completamente. También es crucial investigar más a fondo el comportamiento del término de ruido blanco en los modelos, ya que una calibración excesiva podría estar subestimando los errores reales. Las perspectivas futuras incluyen la aplicación de estos modelos en cuencas con menor memoria hidrológica para evaluar su adaptabilidad y desempeño en diferentes escenarios, así como la posibilidad de optimizar los parámetros del modelo para mejorar el ajuste en eventos extremos.

9 Referencias

- ALNAP. (2010). *Inundaciones: Aprendiendo de anteriores operaciones de emergencia y recuperación*. (S. (Cabir, Trans.). Madrid. Retrieved from <https://iecah.org/inundaciones-aprendiendo-de-anteriores-operaciones-de-emergencia-y-recuperacion/>
- Alonso, G., Rivero, A., Lau, A., & Ruíz, M. (2021). Comparison between SARIMA and Holt–Winters models for forecasting monthly streamflow in the western region of Cuba. *SN Applied Sciences*, 3(6), 1–12. doi: 10.1007/S42452-021-04667-5
- BBC News Mundo. (2023, August 22). Inundaciones en Chile: imágenes de los efectos de las lluvias en el sur y centro del país que llevaron al presidente Boric a declarar “estado de catástrofe.” *BBC News Mundo*. Retrieved from <https://www.bbc.com/mundo/articles/c3glez6268qo>
- Bustos, D. (2011). *Cambio climático y eventos de emergencia en el suministro de agua potable en el Gran Santiago*. Santiago de Chile. Retrieved from <https://repositorio.uchile.cl/handle/2250/104252>
- Cabral, V., Reis, F., Veloso, V., Correa, C., Kuhn, C., & Zarfl, C. (2023). The consequences of debris flows in Brazil: A historical analysis based on recorded events in the last 100 years. *Landslides*, 20(3), 511–529. doi: 10.1007/s10346-022-01984-7
- Centro de Ciencia del Clima y la Resiliencia (CR2). (2024, December 1). *Explorador Climático*. Retrieved from <https://explorador.cr2.cl/>
- Chavasse, D., & Seoane, R. (1997). Asociación determinístico-estocástica para predicción de caudales. *Ingeniería Del Agua*, 4(2), 55–64. doi: <https://doi.org/10.22325/fes/res.2025.256>
- CIREN. (2021, May 31). *Determinan caracterización de uso de suelo para la cuenca del río Maipo*. Retrieved from <https://www.ciren.cl/noticias/determinan-caracterizacion-de-uso-de-suelo-para-la-cuenca-del-rio-maipo/>
- Court, S. (2008). *Pronóstico de caudales de las cuencas de Betania y Bío-bío utilizando métodos estadísticos*. Santiago. Retrieved from <https://repositorio.uchile.cl/handle/2250/103288>
- Diego Martínez, Julio Albín, José Cabaleiro, Tomás Pena, Francisco Rivera, & Vicente Blanco. (2009). El criterio de información de Akaike en la obtención de modelos estadísticos de Rendimiento. *XX Jornadas de Paralelismo, A Coruña, España*, 439–444. Retrieved from https://www.researchgate.net/profile/Tomas-Pena/publication/236279245_El_criterio_de_informacion_de_Akaike_en_la_obtencion_de_modelos_estadisticos_de_Rendimiento/links/58904fa3aca272bc14be3600/El-criterio-de-informacion-de-Akaike-en-la-obtencion-de-modelos-estadisticos-de-Rendimiento.pdf
- Dirección General de Aguas. (n.d.). *Estadísticas estaciones DGA*. Retrieved from <https://dga.mop.gob.cl/estadisticas-estaciones-dga/>
- Dirección General de Aguas, M. de O. P. (2024). *Dirección General de Aguas – Ministerio de Obras Públicas*. Retrieved from <https://dga.mop.gob.cl>
- EL PAÍS Chile. (2024, June 24). *Las inundaciones y lluvias en Chile dañan un centenar de casas y afectan a miles de personas*. Retrieved from <https://elpais.com/chile/2024-06-12/las-inundaciones-y-lluvias-en-chile-danan-un-centenar-de-casas-y-afectan-a-miles-de-personas.html>
- García, E. (2025). After the Barrancà, the large flash flood in Valencia, Spain: some tasks for sociology. *Revista Espanola de Sociologia*, 34(1). doi: 10.22325/fes/res.2025.256

- García, R. (2008). *Gestión de cuencas y Cambio Climático: El caso del Maipo*. Santiago, Chile. Retrieved from <https://cdn.plataformaurbana.cl/wp-content/uploads/2016/05/cambio-clim%C3%A1tico-y-cuenca-r%C3%ADo-maipo.pdf>
- Gobierno de Chile, M. del M. A. D. G. de Aguas. (2004). *Diagnóstico y clasificación de los cursos y cuerpos de agua según objetivos de calidad: Cuenca del Río Maipo*. Santiago, Chile. Retrieved from <https://mma.gob.cl/wp-content/uploads/2017/12/Maipo.pdf>
- González, R. (2003). *Pronóstico de crecidas en tiempo real: uso de redes neuronales*.
- Gutierrez, Y., & Sandoval, W. (n.d.). *ESTIMACIÓN DE PRONOSTICOS DE CAUDALES DEL RÍO MAGDALENA A LA ALTURA DEL MUNICIPIO DE EL BANCO, DEPARTAMENTO DEL MAGDALENA TITULO EN INGLÉS*.
- Hasan, M. M. (2024). Effectiveness of flood early warning for the Jamuna char-dwellers at Sirajganj District in Bangladesh. *Progress in Disaster Science*, 24, 9. doi: 10.1016/j.pdisas.2024.100373
- Herramienta Evaluación Rápida Agua (HERA). (2025). *La cuenca hidrográfica del río Maipo*. Retrieved from <https://www.cuencadelmaipo.cl/cuenca/>
- Instituto Nacional de Estadísticas (INE). (2018). *Síntesis de resultados: Censo de Población y Vivienda 2017*. Santiago, Chile. Retrieved from <https://www.ine.gob.cl>
- Kalogeropoulos, K., Tsanakas, K., Stathopoulos, N., Tsesmelis, D. E., & Tsatsaris, A. (2023). Cultural heritage in the light of flood hazard: The case of the “Ancient” Olympia, Greece. *Hydrology*, 10(3). doi: 10.3390/hydrology10030061
- Khedhiri, S. (2022). Comparison of SARIMA and LSTM methods to model and to forecast Canadian temperature. *Regional Statistics*, 12(2), 177–194. doi: 10.15196/RS120204
- Kumar Dubey, A., Kumar, A., García-Díaz, V., Kumar Sharma, A., & Kanhaiya, K. (2021). Study and analysis of SARIMA and LSTM in forecasting time series data. *Sustainable Energy Technologies and Assessments*, 47. doi: 10.1016/J.SETA.2021.101474
- La Nación. (1953, August 21). Santiago, centro del temporal. *La Nación*, s.n. Retrieved from <https://culturadigital.udp.cl/index.php/lanacion/la-nacion-12403/>
- Lee, D. Y., Centanaro, D. E., Titulo, E. L., & Maestría, D. ". (2016). *Diseño de un modelo estadístico de pronóstico de caudales promedios mensuales para la cuenca hidrográfica 108 del río Chiriquí, años 1970-2012*. Ciudad de Panamá. Retrieved from <https://up-rid.up.ac.pa/61/>
- Martínez, C., Villarroes, M., Pérez, N., Peña, P., Zúñiga, J., Acuña, V., Guerrero, N., Inzunza, S., Qüense, J., Bonet, M., & Merino, Y. (2023). *Inundaciones 21-26 junio 2023 Cuenca Río Maipo*. Santiago: CIGIDEN. Retrieved from <https://www.cigiden.cl/informe-post-desastres-cigiden-inundaciones-21-26-junio-2023-cuenca-rio-maipo/>
- Meis, M., & Llano, M. P. (2017). *Modelado estadístico del caudal mensual en la baja cuenca Del Plata.: Vol. N. Buenos Aires*. Retrieved from https://www.scielo.org.ar/scielo.php?pid=S1850-468X2018000200004&script=sci_abstract
- Mendoza, P. (2010). *Pronóstico de Caudales de Crecida Mediante un Modelo Hidrológico Distribuido y la Asimilación de Datos Observados*. Santiago de Chile. Retrieved from <https://www.researchgate.net/publication/47804769>
- Ministerio del Medio Ambiente. (2021). *capítulo 17: Eventos extremos y desastres*. Santiago de Chile. Retrieved from <https://sinia.mma.gob.cl/wp-content/uploads/2021/04/17-eventos-extremos-y-desastres.pdf>
- Oberli, C., Gironás, J., Escauriaza, C., & Cienfuegos, R. (2021a). *Sistemas de monitoreo y alerta temprana (SMAT), un elemento esencial en la gestión de desastres de origen*

- hidrometeorológico*. Santiago, Chile: CIGIDEN. Retrieved from https://www.cigiden.cl/wp-content/uploads/2021/04/PP_SMAT_v06-con-ISBN-DIGITAL.pdf
- Oberli, C., Gironás, J., Escauriaza, C., & Cienfuegos, R. (2021b). *Sistemas de monitoreo y alerta temprana (SMAT), un elemento esencial en la gestión de desastres de origen hidrometeorológico*.
- OCHA. (2024). *Brasil - Emergencia de Inundaciones Rio Grande do Sul - Nota Resumen #3 al 10 de junio 2024*. Retrieved from <https://droughtclp.unccd.int/system/files/2024-06/Brasil%20-%20Emergencia%20de%20Inundaciones%20Rio%20Grande%20do%20Sul%20-%20Nota%20Resumen%20%233%20al%2010%20de%20junio%202024.pdf>
- Oficina de las Naciones Unidas para la Reducción del Riesgo de Desastres (UNDRR). (2020). *Inundaciones*. Retrieved from <https://www.eird.org/fulltext/ABCDesastres/teoria/inundacion.htm>
- Olcina-Cantos, J., & Díez-Herrero, A. (2025). Floods in Spain: the role of spatial planning. *Ciudad y Territorio Estudios Territoriales*, 57(223), 2659–3254. doi: 10.37230/CyTET.2025.223.1
- Ortega, J. C., Bendezu, M. A., Del Savio, A. A., & Canales, F. A. (2014). A tecnologia de remoção de fósforo: Gerenciamento do elemento em resíduos industriais. *Revista Ambiente e Água*, 9(3), 445–458. doi: 10.4136/1980-993X
- Pacheco, F. (2004). *Modelos de pronóstico de crecidas en cuencas de la Sexta Región*.
- Palummo Lantes, J. (2025). *IMPACTOS DE LAS INUNDACIONES EN RIO GRANDE DO SUL: OBSERVACIONES Y RECOMENDACIONES PARA GARANTIZAR LOS DERECHOS ECONÓMICOS, SOCIALES, CULTURALES Y AMBIENTALES*. Comisión Interamericana de Derechos Humanos.
- Pascal, D. (2023). *Aplicación de un Modelo Hidrológico de Pronóstico para el Desarrollo Inicial de un Sistema de Alerta Temprana (SAT) en el Valle del Alto Cachapoal, VI Región*. Retrieved from <https://repositorio.uchile.cl/handle/2250/194603>
- Peirano, R., Kristjanpoller, W., & Minutolo, M. (2021). *Forecasting inflation in Latin American countries using a SARIMA-LSTM combination*.
- Python Software Foundation. (2024). *Python Language Reference, version 3.12*. Retrieved from <https://www.python.org/>
- R Core Team. (2024). *R: The R Project for Statistical Computing*. Retrieved from <https://www.r-project.org/>
- Ríos Insua, D., Ceppi, C., Meléndez, K., José Molero, J., Academia de Ciencias Exactas, R., Naturales, F., de Bari, P., & Rey Juan Carlos, U. (2013). Cambio climático, fenómenos meteorológicos extremos y análisis de riesgos. *Real Academia de Ciencias Exactas, Físicas y Naturales. (Esp)*, 106(2), 147–156. Retrieved from <https://rac.es/ficheros/doc/01106.pdf>
- Roldán, F., Salazar, I., González, G., Roldán, W., & Toro, N. (2022). Flow-Type Landslides Analysis in Arid Zones: Application in La Chimba Basin in Antofagasta, Atacama Desert (Chile). *Water (Switzerland)*, 14(14). doi: 10.3390/w14142225
- Sandoval, G. (1988). *Simulación del proceso lluvia – escorrentía a través de modelos de función de transferencia*.
- SENAPRED. (n.d.). *Inundaciones*. Santiago, Chile. Retrieved from <https://web.senapred.cl/inundaciones/>
- Sina, L. B., Secco, C. A., Blazevic, M., & Nazemi, K. (2023). Hybrid Forecasting Methods—A Systematic Review. In *Electronics (Switzerland)* (Vol. 12, Issue 9). MDPI. doi: 10.3390/electronics12092019

- Singh, V. (1995). *Computer Models of Watershed Hydrology*. Highlands Ranch, CO. Retrieved from <https://www.tucson.ars.ag.gov/unit/publications/PDFfiles/1091.pdf>
- Stillwell, H. D. (1992). Natural hazards and disasters in Latin America. *Natural Hazards*, 6(2), 131–159. doi: 10.1007/BF00124620
- Suescún, A. (2019). *Modelo SARIMA para el pronóstico del nivel del río Magdalena a la altura del municipio de Barrancabermeja*. Bogotá. Retrieved from <http://hdl.handle.net/11371/2811>
- Sun, Q., Wan, J., & Liu, S. (2020). Estimation of Sea Level Variability in the China Sea and Its Vicinity Using the SARIMA and LSTM Models. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 3317–3326. doi: 10.1109/JSTARS.2020.2997817
- Unidad Nacional para la Gestión del Riesgo de Desastres (UNGRD). (2021). *Sistema Nacional de Gestión del Riesgo de Desastres. Guía para el Desarrollo de Sistemas de Alerta Temprana (SAT)* (Unidad Nacional para la Gestión del Riesgo de Desastres (UNGRD), Ed.). Bogotá, Colombia. Retrieved from www.gestiondelriesgo.gov.co
- United Nations Office for Disaster Risk Reduction (UNDRR). (2021, April 21). *Día de la tierra: en 2020 hubo un gran aumento en inundaciones y tormentas*. United Nations Office for Disaster Risk Reduction (UNDRR). Retrieved from <https://www.undrr.org/es/news/dia-de-la-tierra-en-2020-hubo-un-gran-aumento-en-inundaciones-y-tormentas>
- Urrutia, P., Benohr, J., & Peña, L. (2021). *Somos cuenca, habitantes del Maipo*. Santiago, Chile. Retrieved from <https://www.bestiasdelsursalvaje.cl/portfolio/somos-cuenca-habitantes-del-maipo/#:~:text=Somos%20cuenca%3B%20habitantes%20del%20Maipo%20%E2%86%92%20Bestias%20del%20Sur%20Salvaje&text=Este%20proyecto%20es%20un%20esfuerzo,experiencias%20realizadas%20en%20el%20territorio.>
- Vargas, X., & Brown, E. (1988). Modelos de función de transferencia de múltiples entradas para el pronóstico de crecidas en tiempo real. *XIII Congreso Latinoamericano de Hidráulica*.
- Videla Giering, Y. (2013). *MODELACIÓN HIDROLÓGICA DE LA CUENCA ALTA DEL RÍO MAIPO MEDIANTE LA APLICACIÓN DEL SOFTWARE “COLD REGION HYDROLOGICAL MODEL.”* Santiago, Chile. Retrieved from <https://repositorio.uchile.cl/handle/2250/114889>

10 Anexos

10.1 Anexo 1: Tabla con caudales máximos anuales ordenados y cálculo de periodos de retorno (T) con asignación de probabilidades de excedencia

Año	Caudal máximo instantáneo año	Orden mayor a menor	Orden de ocurrencia	Probabilidad	Periodo de Retorno T
1946	259	1135	1	1,25%	80,0
1947	371	1044	2	2,50%	40,0
1948	299	1038	3	3,75%	26,7
1949	261	775	4	5,00%	20,0
1950	209	773	5	6,25%	16,0
1951	214	722	6	7,50%	13,3
1952	183	639	7	8,75%	11,4
1953	388	633	8	10,00%	10,0
1954	312	595	9	11,25%	8,9
1955	183	565	10	12,50%	8,0
1956	157	557	11	13,75%	7,3
1957	189	525	12	15,00%	6,7
1958	211	523	13	16,25%	6,2
1959	286	517	14	17,50%	5,7
1960	302	492	15	18,75%	5,3
1961	492	489	16	20,00%	5,0
1962	274	485	17	21,25%	4,7
1963	489	474	18	22,50%	4,4
1964	399	472	19	23,75%	4,2
1965	296	443	20	25,00%	4,0
1966	313	441	21	26,25%	3,8
1967	221	434	22	27,50%	3,6
1968	127	429	23	28,75%	3,5
1969	240	425	24	30,00%	3,3
1970	218	417	25	31,25%	3,2
1971	226	408	26	32,50%	3,1
1972	384	399	27	33,75%	3,0
1973	633	398	28	35,00%	2,9
1974	252	388	29	36,25%	2,8
1975	246	384	30	37,50%	2,7
1976	169	376	31	38,75%	2,6
1977	434	371	32	40,00%	2,5
1978	557	365	33	41,25%	2,4
1979	517	350	34	42,50%	2,4

Año	Caudal máximo instantáneo año
1980	485
1981	305
1982	773
1983	775
1984	376
1985	443
1986	441
1987	595
1988	196
1989	350
1990	212
1991	1135
1992	340
1993	722
1994	338
1995	262
1996	183
1997	398
1998	523
1999	195
2000	474
2001	525
2002	408
2003	429
2004	202
2005	639
2006	565
2007	310
2008	417
2009	365
2010	223
2011	118
2012	305
2013	193
2014	172
2015	278
2016	1044
2017	208
2018	160
2019	130
2020	150

Orden mayor a menor	Orden de ocurrencia	Probabilidad	Periodo de Retorno T
340	35	43,75%	2,3
338	36	45,00%	2,2
313	37	46,25%	2,2
312	38	47,50%	2,1
310	39	48,75%	2,1
305	40	50,00%	2,0
305	41	51,25%	2,0
302	42	52,50%	1,9
299	43	53,75%	1,9
296	44	55,00%	1,8
286	45	56,25%	1,8
278	46	57,50%	1,7
274	47	58,75%	1,7
262	48	60,00%	1,7
261	49	61,25%	1,6
259	50	62,50%	1,6
252	51	63,75%	1,6
246	52	65,00%	1,5
240	53	66,25%	1,5
226	54	67,50%	1,5
223	55	68,75%	1,5
221	56	70,00%	1,4
218	57	71,25%	1,4
214	58	72,50%	1,4
212	59	73,75%	1,4
211	60	75,00%	1,3
209	61	76,25%	1,3
208	62	77,50%	1,3
202	63	78,75%	1,3
196	64	80,00%	1,3
195	65	81,25%	1,2
193	66	82,50%	1,2
189	67	83,75%	1,2
183	68	85,00%	1,2
183	69	86,25%	1,2
183	70	87,50%	1,1
172	71	88,75%	1,1
169	72	90,00%	1,1
160	73	91,25%	1,1
157	74	92,50%	1,1
150	75	93,75%	1,1

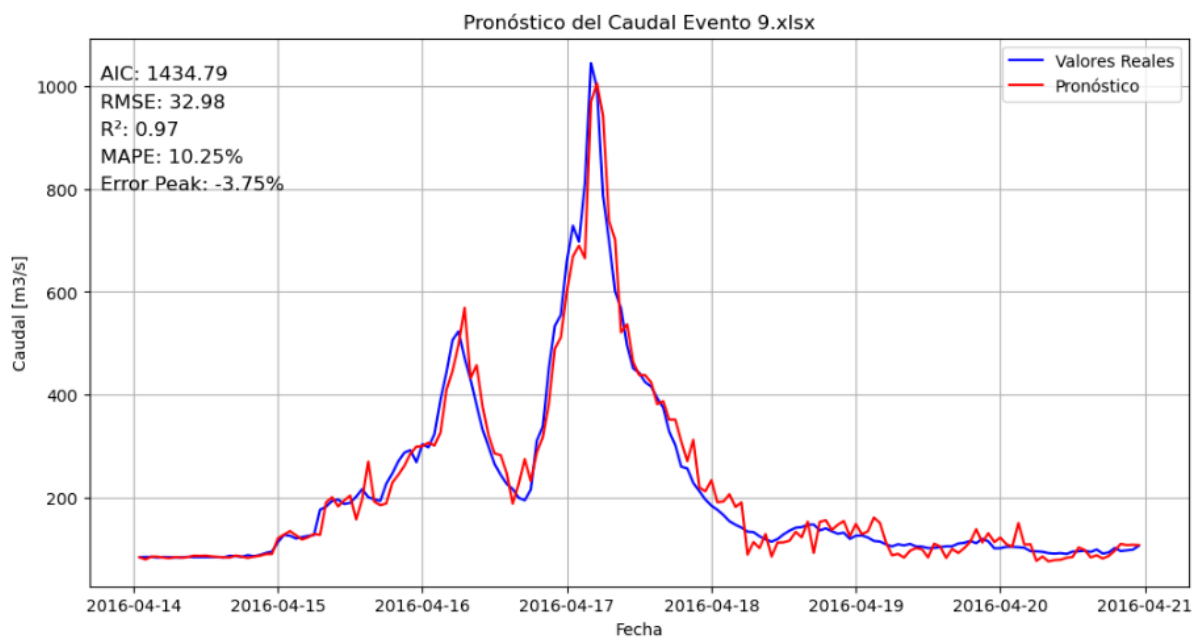
Año	Caudal máximo instantáneo año
2021	472
2022	136
2023	1038
2024	425

Orden mayor a menor	Orden de ocurrencia	Probabilidad	Periodo de Retorno T
136	76	95,00%	1,1
130	77	96,25%	1,0
127	78	97,50%	1,0
118	79	98,75%	1,0

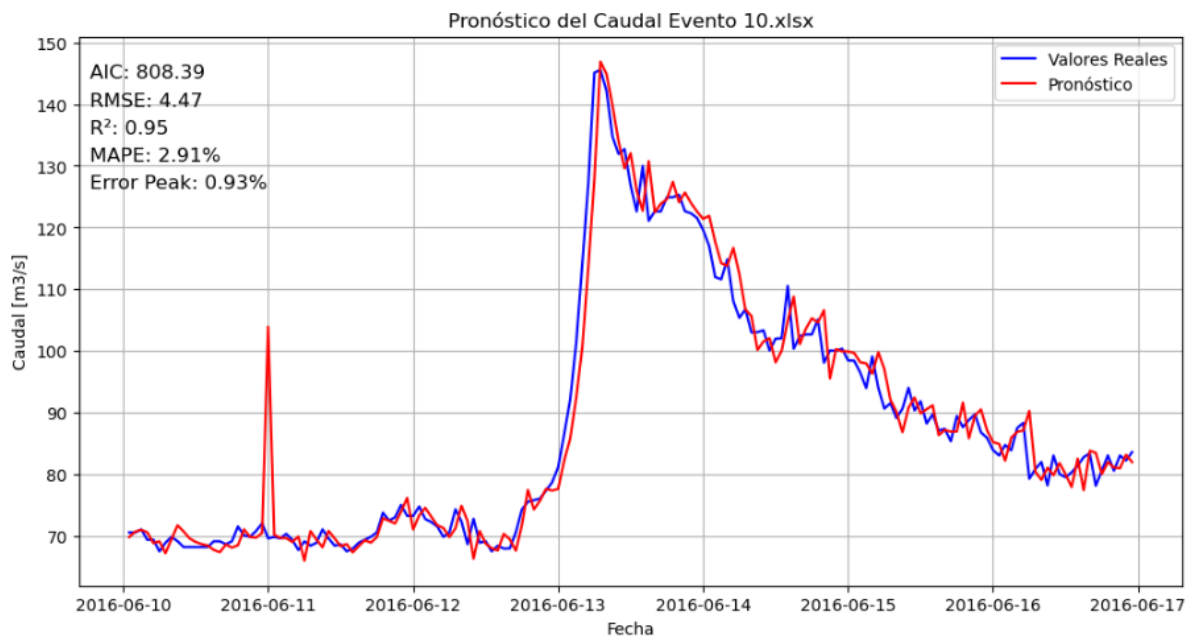
10.2 Anexo 2: Pronósticos de caudal utilizando el modelo Theta 12.

Las gráficas presentadas a continuación muestran los pronósticos de caudal generados para los eventos 9, 10, 11 y 13 utilizando el modelo Theta 12. Estos pronósticos fueron elaborados a partir de variables exógenas, como la temperatura y el caudal en estaciones aguas arriba (Caudal AA). Cada conjunto de gráficos corresponde a un evento específico, donde se comparan los valores pronosticados con los valores reales de caudal en función del tiempo. Los indicadores de desempeño del modelo, como AIC, RMSE, R^2 , MAPE y Error Peak, se incluyen para evaluar la precisión de las predicciones en cada evento, reflejando la capacidad del modelo para capturar las variaciones del caudal.

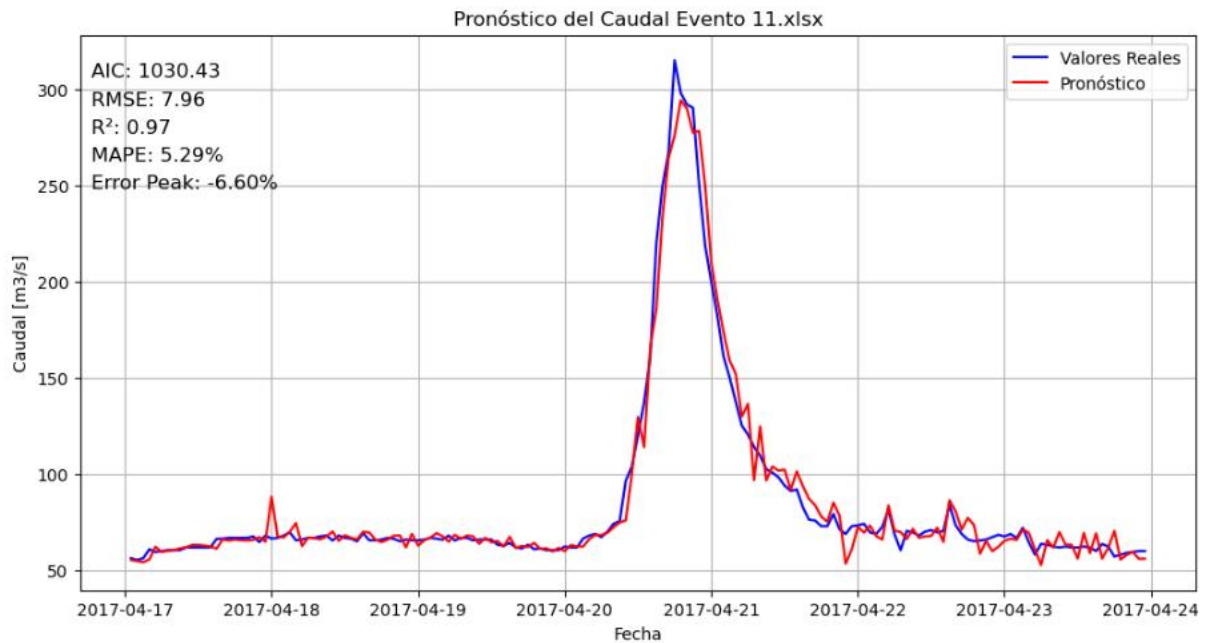
10.2.1 Pronóstico del Caudal para el Evento 9 utilizando el modelo Theta 12 con variables exógenas: Temperatura y Caudal AA



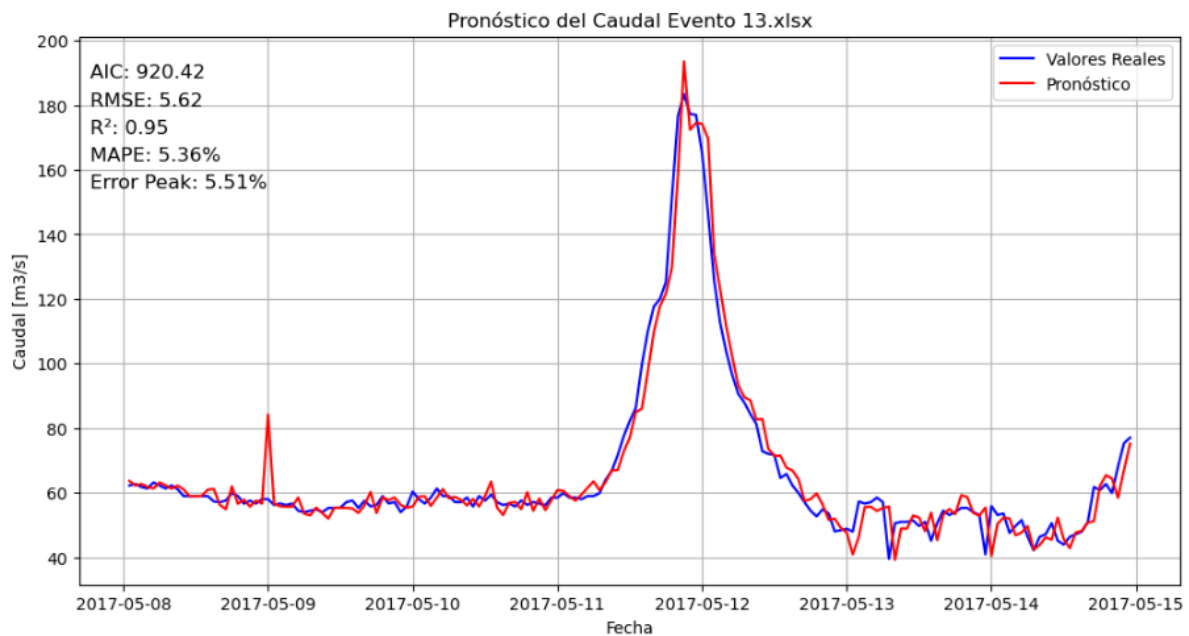
10.2.2 Pronóstico del Caudal para el Evento 10 utilizando el modelo Theta 12 con variables exógenas: Temperatura y Caudal AA



10.2.3 Pronóstico del Caudal para el Evento 11 utilizando el modelo Theta 12 con variables exógenas: Temperatura y Caudal AA



10.2.4 Pronóstico del Caudal para el Evento 13 utilizando el modelo Theta 12 con variables exógenas: Temperatura y Caudal AA



10.3 Pronóstico de caudal a Lead Time 2 con el modelo Theta 12 y comparación con datos observados

