



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA CIVIL

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.

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

DANIEL CRISTÓBAL PASCAL CHADWICK

PROFESOR GUÍA:
MIGUEL LAGOS ZÚÑIGA

MIEMBROS DE LA COMISIÓN:
DANIEL CÉSPEDES CHÁVES
PABLO MENDOZA ZÚÑIGA

SANTIAGO DE CHILE
2023

Resumen ejecutivo

El pronóstico de caudales de crecida y su utilidad para la prevención de catástrofes ha llevado a la comunidad internacional desde los inicios del siglo XXI, con la Organización de Naciones Unidas (ONU) como líder, a impulsar la creación de Sistemas de Alertas Tempranas de crecidas en diversos países propensos a riesgo de eventos meteorológicos extremos.

Muchos enfoques que se han tomado por parte de la comunidad científica experta en hidrología para fines de pronóstico de caudales ha sido el uso de la inteligencia artificial y su gran poder predictivo.

En esta investigación se generan, entrenan y testean seis modelos de Redes Neuronales Recurrentes *Long-Short Term Memory* (LSTM) con el propósito de impulsar un Sistema de Alertas Tempranas (SAT), que sirva de ayuda para pronosticar crecidas en los ríos de Pangal y Cachapoal en el valle del Alto Cachapoal de la VI Región. Para lograr esto se utilizan los datos horarios de precipitación y caudal de las estaciones Pangal en Pangal y Cachapoal 5 km aguas abajo junta Cortaderal y se seleccionan las crecidas con un periodo de retorno mayor a un año para utilizar como base de información. Los modelos están ensamblados para que tengan un tiempo de predicción definido de uno, tres y cinco horas de adelanto para cada cuenca.

Los resultados indican que los pronósticos de una hora de tiempo de adelanto logran reproducir de manera exitosa las dinámicas de caudal para las crecidas de testeo, llegando a tener indicadores R^2 de 0,98 y 0,82 para los sets de testeo de Cachapoal y Pangal respectivamente. A medida que se aumentan los tiempos de adelanto los resultados disminuyen su calidad y se produce un fenómeno de límite superior de predicción en donde el modelo no logra predecir caudales de más de 200 y 150 (m^3/s) para cada cuenca, respectivamente.

Las pruebas realizadas para evaluar la capacidad de los modelos hidrológicos utilizando datos de otras cuencas arrojaron resultados prometedores en cuanto a su capacidad de generalización. En particular, se encontró que el modelo de Cachapoal presentó un ajuste de mayor R^2 que el modelo específico de la cuenca de Pangal. Este resultado sugiere que el modelo de Cachapoal tiene una mayor capacidad de adaptación a diferentes condiciones hidrológicas y una mayor precisión en la predicción de crecidas en cuencas distintas a la que fue ajustado originalmente.

Finalmente, se explicitan los pasos necesarios para continuar con el desarrollo del proyecto, se comentan líneas de investigación afines y útiles para el complemento del proyecto, se entregan directrices para la utilización de los modelos y se recomienda la incorporación de una nueva estación fluviométrica para la cuenca del Río Cipreses.

Agradecimientos

En primer lugar, quiero agradecer a mi profesor guía, Miguel Lagos, por su orientación, guía, confianza y apoyo constante durante todo el proceso. Su experiencia y dedicación han sido fundamentales para la realización de este trabajo.

Quiero agradecer de manera especial a Pacific Hydro Chile, por el apoyo brindado como empresa y por la confianza depositada en mi persona. quiero expresar mi agradecimiento a Daniel Céspedes, profesor co-guia, por sus comentarios y sugerencias constructivas, que me permitieron mejorar la calidad de este trabajo.

Agradezco además la disposición del profesor Pablo Mendoza de formar parte de la comisión evaluadora.

Asimismo, deseo agradecer a Max y Werner por su colaboración y disposición para brindarme información valiosa para guiar el proceso de aprendizaje del campo de redes neuronales y la programación. Su aporte ha sido fundamental para el éxito de este proyecto.

Además, no puedo dejar de agradecer a mi familia y amigos por su constante apoyo y motivación. En especial, quiero mencionar a mi madre Karen y mis abuelos Ingrid y Carlos, por su incondicional apoyo moral y su aliento constante en los momentos más difíciles.

Gracias a mi padre Cristóbal que me hizo elegir la Universidad de Chile como casa de estudios.

Por último, quiero expresar mi más profundo agradecimiento a mi novia Catalina, quien ha estado a mi lado en todo momento, brindándome su amor, paciencia y comprensión. Su apoyo incondicional ha sido un gran pilar en mi vida y en el desarrollo de este trabajo.

Gracias a todos por su contribución y colaboración en la consecución de este objetivo. Sin su apoyo, este trabajo no hubiera sido posible.

Tabla de Contenido

Capítulo 1	1
1.1 Motivación	1
1.2 Objetivo General	3
1.2.1 Objetivos específicos	3
1.3 Alcances del estudio	3
1.4 Organización de la memoria	4
Capítulo 2 : Revisión de antecedentes	5
2.1 Sistema de alertas temprana SAT	5
2.2 Modelación hidrológica	6
2.2.1 Pronóstico de crecidas con redes neuronales	7
2.3 Marco teórico Redes Neuronales Artificiales (RNA)	8
2.3.1 Idea y estructura RNA.....	8
2.3.2 Redes propagación hacia adelante.....	9
2.3.3 Metodología de descenso de gradiente	10
2.3.4 Función objetivo	11
2.4 Redes recurrentes	12
2.5 Red LSTM	14
Capítulo 3 : Zona de estudio	17
3.1 Caracterización zona de estudio	17
3.2 Hidroclima y relieve	19
3.3 Disponibilidad de datos	21
3.4 Selección de crecidas	23
3.4.1 Crecidas Pangal.....	24
3.4.2 Crecidas Cachapoal 5km.....	26
Capítulo 4 : Metodología	28
4.1 Metodología	28
4.2 Recolección y análisis de datos	28
4.3 Programación del modelo	29
4.3.1 Tiempos de adelanto	31
4.4 Obtención de resultados	31
4.4.1 Combinación de <i>DataFrames</i> de crecidas	32
4.4.2 Optimización parámetros del modelo	33
4.5 Consideraciones finales	34
4.5.1 Indicadores de ajuste.....	34

4.5.2	Umbrales de alerta	35
4.5.3	Recomendaciones Sistema de Alertas Tempranas (SAT).....	35
Capítulo 5 : Resultados		36
5.1	Pronóstico de caudales en la cuenca del río Cachapoal	36
5.1.1	Pronóstico hidrológico con tiempo de adelanto de 1 hora	36
5.1.2	Pronóstico hidrológico con tiempo de adelanto de 3 horas.....	39
5.1.3	Pronóstico hidrológico con tiempo de adelanto de 5 horas.....	41
5.2	Pronóstico de caudales en la cuenca del río Pangal.....	43
5.2.1	Pronóstico hidrológico con tiempo de adelanto de 1 hora	43
5.2.2	Pronóstico hidrológico con tiempo de adelanto de 3 horas.....	44
5.2.3	Pronóstico hidrológico con tiempo de adelanto de 5 horas.....	46
5.3	Pruebas de pronóstico cruzado de cuencas.....	46
5.4	Umbrales de alertas.....	47
Capítulo 6 : Discusión.....		49
6.1.1	Resultados modelación.....	49
6.1.2	Requerimientos y recomendaciones implementación SAT	51
6.1.1	Trabajos futuros.....	52
Capítulo 7 : Conclusiones		53
Anexos		59
Anexo A: Funciones de arreglo de base de datos		59
Anexo B: Consideraciones para correr modelos		60
Anexo C: Figuras extras de modelación		62

Índice de Tablas

Tabla 3.1: Estaciones Consultadas.	21
Tabla 3.2: Períodos de retorno estación Pangal en Pangal (PHC,2017	23
Tabla 3.3: Períodos de retorno estación Cachapoal 5km Aguas abajo junta Cortaderal (PHC,2017)	23
Tabla 3.4: Correlación Precipitación Caudal Crecidas Pangal y período de retorno	24
Tabla 3.5: Correlación entre datos de precipitación y caudal Cachapoal	26
Tabla 4.1: Crecidas de validación y testeo modelo Cachapoal	33
Tabla 4.2: Crecidas de validación y testeo modelo Pangal	33
Tabla 5.1: Hiperparámetros de modelo óptimo Cachapoal 1 hora de adelanto	36
Tabla 5.2: Hiperparámetros de modelo óptimo Cachapoal 3 horas de adelanto	39
Tabla 5.3: Hiperparámetros de modelo óptimo Cachapoal 5 horas de adelanto	41
Tabla 5.4: Hiperparámetros de modelo óptimo Pangal 1 hora de adelanto	43
Tabla 5.5: Hiperparámetros de modelo óptimo Pangal 3 horas de adelanto	44
Tabla 5.6: Umbrales de alerta DGA	48

Índice de Figuras

Figura 2.1: Diagrama SAT (PNUD,2020)	5
Figura 2.2: estructura de red <i>feedforward</i> (Sai,2020).....	9
Figura 2.3: diferencias entre red <i>feedforward</i> y RNN (Radhakrishnan, 2017).....	12
Figura 2.4: Red recurrente desenrollada.....	13
Figura 2.5: Celda LSTM	15
Figura 3.1: Cartografía Cuenca Río Cachapoal En Pte Termas De Cauquenes con bocatomas de interés.....	17
Figura 3.2: Diagrama unifilar sistema de generación hidroeléctrica (PHC,2018)	18
Figura 3.3: Ubicación cuencas de estudio y estaciones de medición.	19
Figura 3.4: Curvas de Variación Estacional cuencas Pangal y Cachapoal.....	20
Figura 3.5: Mapas de calor de información por año para las estaciones seleccionadas.	22
Figura 3.6: Set de crecidas cuenca Pangal.....	25
Figura 3.7: Set crecidas cuenca Cachapoal.....	27
Figura 4.1: Esquema metodología de trabajo.	28
Figura 4.2: Esquema de pasos programación del modelo.	30
Figura 4.3: Desfase de red para generar tiempos de adelanto.....	31
Figura 4.4: Esquema de obtención de resultados	31
Figura 5.1: Crecida de testeo (izquierda) y de validación (derecha)	36
Figura 5.2: Crecidas de entrenamiento	37
Figura 5.3: Gráficos de dispersión crecidas de entrenamiento.....	37
Figura 5.4: Boxplot indicadores RMSE, MAE, AIC, R^2 y Pbias de pronóstico Cachapoal 1 hora.	38
Figura 5.5: Crecida de testeo (izquierda) y de validación (derecha).....	39
Figura 5.6: Crecidas de entrenamiento.	40
Figura 5.7: Boxplot indicadores RMSE, MAE, AIC, R^2 y Pbias de pronóstico Cachapoal 3 hora.	40
Figura 5.8: Dispersión de datos en crecidas de mayor magnitud.	41
Figura 5.9: Crecidas de entrenamiento (izquierda), validación (centro) y testeo (derecha).....	42
Figura 5.10: Boxplot indicadores RMSE, MAE, AIC, R^2 y Pbias de pronóstico Cachapoal 5 hora.	42
Figura 5.11: Crecida de testeo (izquierda) y validación (derecha).....	43
Figura 5.12: Boxplot indicadores RMSE, MAE, AIC, R^2 y Pbias de pronóstico Pangal 1 hora.	44
Figura 5.13: Crecida de testeo (izquierda) y validación (derecha).....	45
Figura 5.14: Boxplot indicadores RMSE, MAE, AIC, R^2 y Pbias de pronóstico Pangal 3 horas.	45
Figura 5.15: Mayores crecidas de Pangal testeadas en modelo Cachapoal 1 hora	46
Figura 5.16: Crecidas de Cachapoal testeadas en modelo Pangal 1 hora.....	47
Figura 8.1: Serie Posterior a la aplicación de <i>TimeFix</i>	59

Figura 8.2: Serie previa a la aplicación de <i>TimeFix</i>	59
Figura 8.3: Scatterplot conjunto testeo.....	62
Figura 8.4: Scatterplot conjunto entrenamiento	63
Figura 8.5: Scatterplot conjunto testeo.....	64
Figura 8.6: Scatterplots conjunto entrenamiento.....	65
Figura 8.7: scatterplot conjunto testeo.....	66
Figura 8.8:Scatterplot conjunto entrenamiento	67
Figura 8.9: Scatterplots conjunto de entrenamiento Pangal 1H.....	68
Figura 8.10: Scaterplots de testeo (izquierda) y validación (derecha).....	69
Figura 8.11: Scaterplots de testeo (izquierda) y validación (derecha) Pangal 3 horas	69
Figura 8.12: Scatterplots conjunto de entrenamiento Pangal 1H.....	70
Figura 8.13: Crecidas de testeo (izquierda) y validación (derecha)	71
Figura 8.14: Crecidas de entrenamiento.	71
Figura 8.15: Boxplot Modelo Pangal 5 horas de pronóstico	72
Figura 8.16: Scatterplot testeo y validación Pangal 5 horas	73
Figura 8.17: Scatterplot conjunto entrenamiento Pangal 5 horas.....	73

Capítulo 1: Introducción

1.1 Motivación

A lo largo de la historia de la humanidad, los ríos han sido fuentes de agua y alimento, por ende, gran parte de ciudades y poblados se han asentado en sus cercanías. Sin embargo, esta práctica no solo conlleva beneficios, sino que también entra en juego un grado alto de peligrosidad por eventos de crecidas e inundaciones, a los cuales los humanos, desde los más antiguos, hasta las sociedades más desarrolladas de hoy en día se enfrentan. En respuesta de lo anterior y enfocado en la correcta precaución de catástrofes por daños en infraestructura hidráulica causada por subidas repentinas de caudal, es que se recurre a la modelación hidrológica de pronóstico y Sistemas de Alertas Tempranas (SAT) con el fin principal de reducir las condiciones de riesgo de las personas ante un evento de crecida (Vargas et al., 2003).

En vista de los escenarios de emisiones de CO₂ y la creciente incertidumbre climática por el aumento de temperatura de la tierra (Loaiciga et al., 1996), se esperan alteraciones en los comportamientos hidrológicos producto del cambio climático. De hecho, en Chile se proyectan aumentos en la intensidad de precipitación de eventos extremos y temperatura (Santibáñez, 2016).

Internacionalmente se han realizado grandes esfuerzos por parte de la oficina de Estrategia Internacional para la Reducción de Desastres de las Naciones Unidas (UNISDR) para concientizar sobre las posibles amenazas que existen en los distintos territorios, estas pueden ser de índole geológica, hidrometeorológica, biológica, natural, sociocultural o tecnológica. El sistema de alertas temprana debe moldearse a las amenazas y las principales comunidades afectadas.

El Programa de las Naciones Unidas para el Desarrollo (PNUD) ha apoyado y promovido la implementación de SAT de crecidas en varios países, especialmente aquellos que son vulnerables a eventos climáticos extremos como inundaciones y deslizamientos de tierra (PNUD,2020).

Un sistema de alerta temprana de crecidas es un conjunto de herramientas y procedimientos que permiten la detección temprana de la probabilidad de una crecida en una determinada área geográfica. Esto puede incluir la medición de los niveles de agua, la observación de las condiciones climáticas y la evaluación de diversos parámetros físicos que influyen en el ciclo hidrológico. El objetivo de un SAT de crecidas es proporcionar información oportuna y precisa a las autoridades y población en general, para que se puedan tomar medidas preventivas. Esto puede incluir la evacuación de áreas vulnerables, la activación de planes de emergencia y la preparación de recursos y equipos de rescate.

En Chile, el Coordinador Eléctrico Nacional cuenta con un Sistema de Pronóstico de Caudales (SPC), que corresponde a un modelo predictivo a 14 días de las variables caudal, lluvia, caudal de deshielo y temperatura, para distintas estaciones fluviométricas y meteorológicas, ubicadas en cuencas de interés para la generación eléctrica del centro y sur del país (CEN,2021). Están implementados además dos proyectos que comparten el objetivo de predecir, cuantificar y/o identificar una amenaza con el fin de alertar a los posibles damnificados. En primer lugar, se publicó por parte del Servicio Nacional de Geología y Minería (SERNAGEOMIN) la plataforma: MAPA DE PELIGRO DE REMOCIONES EN MASA DE TIPO FLUJO (Marín,2020) en dónde se exponen las áreas expuestas a estos peligros en cuencas del Río Maipo. En segundo lugar, el sistema de monitoreo y alerta temprana en la quebrada San Ramón, región Metropolitana de Santiago (DICTUC, 2016; Gironás y Sandoval, 2017) es un proyecto que busca dar seguimiento de la precipitación, caudal e isoterma 0 en la quebrada de Macul, sentando las bases para la realización de un modelo hidrológico que se ajuste a la disponibilidad de los datos recopilados.

La presente memoria de título, busca realizar el primer paso para la alerta de crecidas en los ríos Pangal y Cachapoal a los operadores de las bocatomas Pangal y Chacayes, para que, en caso de fuertes lluvias y alertas del modelo, se tomen las acciones preventivas de operación en las compuertas hidráulicas, y a la par se genere una cadena de alarmas entre los propios trabajadores y se genere un seguimiento de lo sucedido, que en caso de conllevar una amenaza mayor será conversado con la Oficina Nacional de Emergencia del Ministerio del Interior (ONEMI) que se encarga día a día de monitorear los reportes meteorológicos en todas las regiones del país.

En el trabajo se estudia la zona del valle Alto del Cachapoal en la VI Región, donde existe un sistema de generación de energía mediante centrales hidroeléctricas de pasada, construidas en los inicios del siglo XX como solución eléctrica para la minería de la zona (BCN, 2014), en específico la tubería de madera de Pangal, construida en 1919, es declarada monumento nacional en 2014 (Decreto n°129, 2014). Hoy en día las instalaciones pertenecen a la empresa Pacific Hydro Chile (PHC), que cuenta con un sistema compuesto por cinco bocatomas, tres casas de máquinas y un método de aprovechamiento de energía hidroeléctrica mediante trasvase de caudal por canales, acueductos y túneles entre las casas de máquinas. Las instalaciones y sus operadores son los primeros en recibir aumentos súbitos de caudal producto de la precipitación y el rápido actuar es crucial para proteger el sistema hidráulico en caso de caudales peligrosos para las capacidades de las centrales. Es importante mencionar que estos no representan las mismas magnitudes que los caudales peligrosos para las localidades cercanas al río Cachapoal y lo propuesto en esta memoria toma en cuenta los dos tipos de umbrales.

En el desarrollo de este estudio, se busca aplicar un modelo de redes neuronales destinado al pronóstico de crecidas, basado en datos de caudal y precipitación a nivel horario. Para esto, se seleccionan todas las crecidas pluviales disponibles para las estaciones cercanas desde el año 2000, con el fin de implementar algoritmos de inteligencia artificial, en

específico, redes neuronales recurrentes *Long-Short Term Memory* (LSTM) a una resolución horaria.

Previo a la elección de un modelo corresponde la evaluación de disponibilidad de datos, distribución espacial y escalas de tiempo de las mediciones. El catastro entrega la información necesaria para definir los parámetros de entrada y limitar las opciones, ya que la oferta de modelos y herramientas computacionales es extensa y se encuentra ampliamente estudiada, por lo tanto, lo complejo es la selección de la herramienta adecuada para lograr el objetivo. Toth y Brath (2007), plantean que en la elección de un modelo de pronóstico precipitación-caudal hay dos puntos cruciales, primero cuantificar y comprender la cantidad y calidad de datos necesarios para obtener una correcta calibración; y, en segundo lugar, la fiabilidad de las escalas de tiempo predichas por la modelación.

1.2 Objetivo General

- Desarrollar un modelo de redes neuronales artificiales con datos históricos para los ríos Pangal y Cachapoal, con el fin de apoyar la toma de decisiones para preservar la integridad de las personas, ecosistemas e infraestructura, mediante el desarrollo de la primera etapa de un Sistema de Alertas Tempranas (SAT).

1.2.1 Objetivos específicos

- Identificar la predictibilidad de los sistemas hidrológicos analizados mediante el análisis de la memoria hidrológica y correlación estadística con predictores meteorológicas.
- Desarrollar modelos de pronóstico de crecidas a resolución temporal horaria para las bocatomas de los ríos Pangal y Chacayes, considerando distintas combinaciones de predictores y tiempos de adelanto.
- Análisis comparativo entre modelos de pronóstico en base a su desempeño en eventos históricos, para propuesta de implementación y puesta en marcha, así como la determinación de umbrales de alerta basado en las condiciones pronosticadas por el modelo.
- Sentar los pasos a seguir y recomendaciones para la implementación del SAT.

1.3 Alcances del estudio

El Alcance del trabajo se enmarca en el desarrollo de un modelo de Redes Neuronales Recurrentes (RNR) de pronóstico de crecidas pluviales para las bocatomas Pangal y

Chacayes, junto con una propuesta de puesta en marcha y requerimientos futuros para el desarrollo completo de un Sistema de Alertas Tempranas (SAT) en el valle alto del Río Cachapoal.

Para lograr el objetivo se estudia la información horaria de las forzantes climáticas precipitación, temperatura y caudal, obtenidas de portales de la Dirección General de Aguas (DGA). Estos datos son, a priori, los *input* para entrenar el modelo RNR de arquitectura específica LSTM, identificando alrededor de 40 eventos de crecidas.

1.4 Organización de la memoria

En el Capítulo 2 se plantea un relato de la historia de la modelación hidrológica y de los Sistemas de Alertas Tempranas. Además, se habla sobre el funcionamiento de las redes recurrentes estilo LSTM y el porqué es la herramienta escogida por sobre otro tipo de modelación, adentrándose en los algoritmos que rigen en las redes neuronales.

En el Capítulo 3 se caracteriza la zona de estudio, presentando las cuencas de estudio, diagrama unifilar de los ríos y centrales junto a una descripción hidroclimática general del territorio. Se finaliza la sección exponiendo las estaciones, su disponibilidad de datos y el análisis de las crecidas pluviales seleccionadas en cantidad, magnitud, extensión y período de retorno.

En el Capítulo 4 se presentan las técnicas de preprocesamiento de datos en base a la técnica *Exploratory Data Analysis* (EDA). Se explica el paso a paso de la programación del modelo y las técnicas utilizadas para la obtención y optimización de los resultados.

En el Capítulo 5 se exponen los resultados en forma gráfica y son discutidos en base a la literatura revisada y la experiencia ganada en las pruebas de modelación. Adicionalmente se exponen los pasos a seguir para lograr la implementación del SAT, los trabajos futuros afines para complementar el proyecto y se entregan consideraciones para la utilización de los modelos. Finalmente, en el Capítulo 7 se presentan las principales conclusiones del trabajo en respuesta de los objetivos planteados.

Capítulo 2: Revisión de antecedentes

2.1 Sistema de alertas temprana SAT

La motivación de un SAT es la protección de las personas. Para lograr el objetivo, es necesario una estructura que organice las acciones, actores, partes y requerimientos de la generación y divulgación de alertas. En la Figura 2.1 se presenta un diagrama de las partes fundamentales que debe incorporar un SAT.

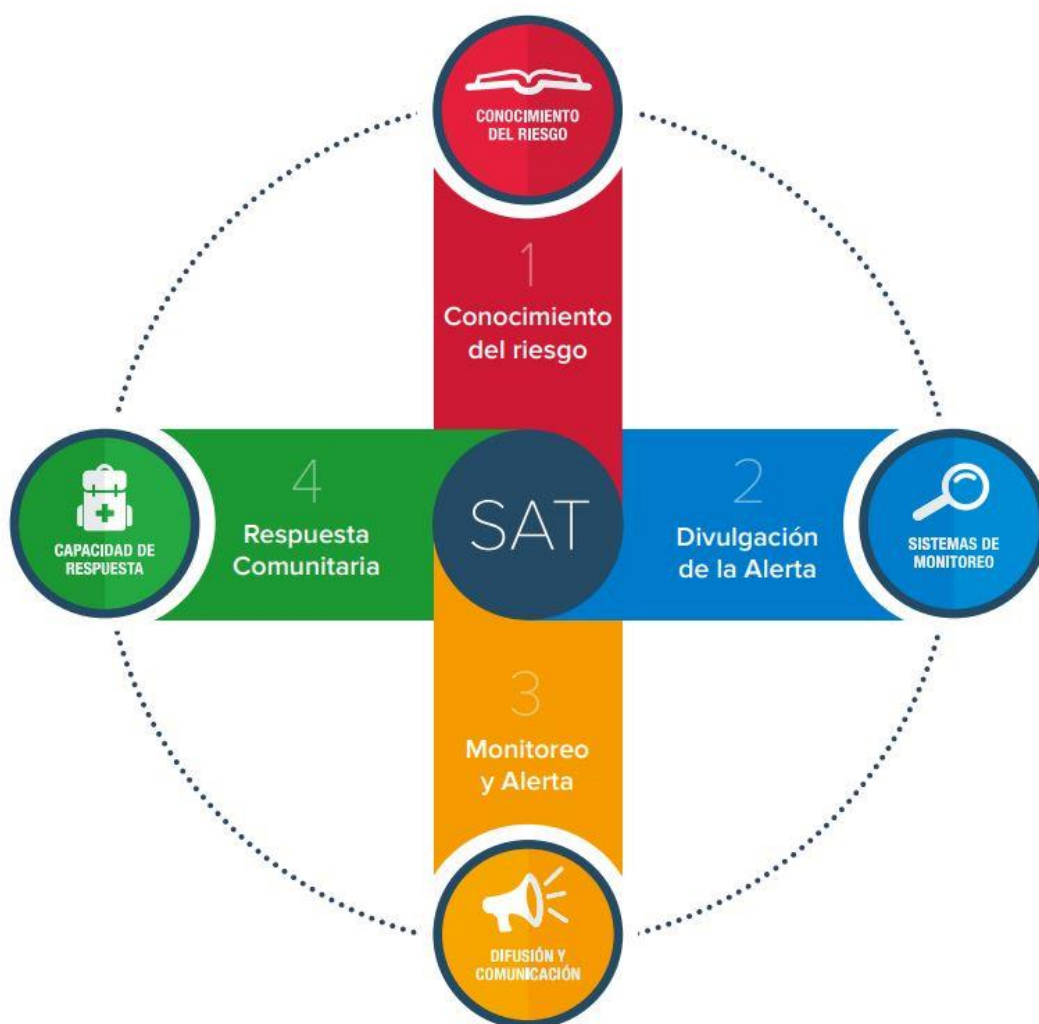


Figura 2.1: Diagrama SAT (PNUD,2020)

El primer paso contempla el estudio de los riesgos asociados a la amenaza y establece un proceso sistemático para recopilar, evaluar y compartir información sobre las amenazas y vulnerabilidades (PNUD,2020). Por parte de Pacific Hydro Chile, la motivación de impulsar el proyecto SAT, mediante esta memoria, es debido al conocimiento y experiencia frente al

riesgo de crecidas y lo que ha implicado en el pasado, es más, eventos como la crecida de mayo de 2016, que alcanzó caudales máximos instantáneos de 283 (m³/s) (22 años período de retorno) en la estación Pangal en Pangal, dejó damnificados en Doñihue (Emol, 2016) y generó indisponibilidad en las bocatomas

El segundo paso tiene el objetivo de establecer servicios efectivos de monitoreo y alerta de amenazas basados en sólidos fundamentos científicos y tecnológicos (PNUD, 2020). Como parte de este eslabón, y siendo un requerimiento clave en caso de amenazas de crecidas, se encuentra la generación de un modelo de predicción de crecidas que se encuentre funcionando en tiempo real y tenga incorporada la función de generar una alerta.

La parte 3 corresponde a la operación propiamente tal del sistema implementado y forma parte de la continuación del proyecto SAT, en dónde el modelo debe pasar por una fase de automatización e interconexión en tiempo real con datos DGA para luego testear los resultados obtenidos por un período de marcha blanca de una temporada de invierno. En lo que respecta a la respuesta comunitaria esta se debe regir por los protocolos de comunicación entre los trabajadores de operaciones, la empresa y autoridades son parte del funcionamiento interno de PHC.

2.2 Modelación hidrológica

El inminente avance en la tecnología y computación en la década de los 60s dieron cabida a las primeras modelaciones hidrológicas. Freeze & Harlan (1969) fueron los primeros en darse cuenta de cómo estos avances podrían beneficiar el futuro de la hidrología, entendiendo que la combinación de los avances en hidrología física y la tecnología informática digital debería impulsar la necesidad de una reorientación en la investigación de simulación hidrológica. Tanto fue su impacto que Simmons et al. (2020) conmemoraron los 50 años del trabajo realizado por Freeze y Harlan, destacando la visión futurista que tuvieron para la época, al buscar integrar y comprender los procesos dentro del ciclo hidrológico. Los avances en modelación estuvieron limitados por lo que sucedía con la tecnología geoespacial y computacional, los avances en percepción remota, uso de plataformas SIG y en herramientas de medición.

En la actualidad, existe a disposición herramientas muy potentes en el ámbito computacional y de calidad de mediciones, por lo que hay una vasta oferta de modelos hidrológicos que aprovechan y utilizan las nuevas tecnologías. Estos se basan en hipótesis y parámetros que buscan simular, numéricamente, los diversos procesos hidrológicos del pasado, presente y futuro. Uno de los enfoques más utilizados es el de los modelos basados en procesos. Estos modelos se fundan en la comprensión de los procesos hidrológicos, como la infiltración, la evapotranspiración, el flujo de agua subterránea y la escorrentía superficial mediante las ecuaciones que rigen el transporte de masa, energía y momentum entre los distintos actores involucrados (principalmente el agua en sus distintas fases, la radiación, temperatura y propiedades cuantificables de las cuencas). Ejemplos de modelos

con este enfoque son los modelos hidrológicos SWAT¹ y NWSRFS² (Larson, 2002; William & Arnold, 1993).

Otro enfoque comúnmente utilizado es el de los modelos empíricos basados en datos históricos. Estos modelos se sustentan en la correlación estadística entre las variables hidrológicas de entrada y salida. Chang y Chen (2001) proponen un modelo empírico basado en redes neuronales y lógica difusa para predecir el caudal en tiempo real en ríos. El modelo utiliza datos históricos de caudal y datos meteorológicos como entradas para entrenar una red neuronal de retropropagación y una red neuronal de propagación hacia atrás. Luego, se utiliza lógica difusa para fusionar las salidas de ambas redes y obtener la predicción final de caudal en tiempo real.

Finalmente, los modelos conceptuales de pronóstico de crecidas son modelos hidrológicos que se basan en la descripción conceptual de los procesos hidrológicos que ocurren en una cuenca. Estos modelos suelen ser más simples que los modelos numéricos de base física y se utilizan principalmente en situaciones en las que no se dispone de datos suficientes para calibrar modelos más complejos (Nash & Sutcliffe, 1970).

2.2.1 Pronóstico de crecidas con redes neuronales

Las crecidas son eventos hidrológicos extremos que pueden causar daños significativos a la infraestructura y a la población cercana a los ríos. Históricamente, los métodos para predecir estas crecidas se han basado en técnicas empíricas y modelos hidrológicos sencillos. Sin embargo, en los últimos años, las redes neuronales han entrado al mundo de la modelación hidrológica como una alternativa prometedora para mejorar la precisión de los pronósticos de crecidas (Vargas et al., 2003).

Las redes neuronales son modelos matemáticos que simulan el comportamiento del cerebro humano y se utilizan para resolver problemas complejos de predicción y clasificación. En hidrología, las redes neuronales se han empleado para predecir variables hidrológicas como caudales y niveles de agua en ríos y lagos. Los modelos basados en redes neuronales han demostrado ser útiles para mejorar la precisión de los pronósticos de crecidas, en comparación con los modelos hidrológicos tradicionales (Ávila Mercado, 2018). A continuación, se presentarán los principales resultados de distintos estudios que respaldan lo expuesto.

Primeramente, en el estudio de Connor et al. (1994), se examinó el uso de Redes Neuronales Recurrentes (RNN) para predecir series de tiempo. Los resultados mostraron que la RNN superó a los métodos tradicionales de análisis de series de tiempo y que la

¹ Soil and Water Assessment Tool

² National Weather Service River Forecast System

arquitectura de la RNN tenía un gran impacto en su capacidad de previsión. Además, se encontró que la red era robusta frente a la presencia de ruido en los datos de entrada, dando a entender que es una herramienta útil en hidrología.

En el artículo "*Artificial intelligence based models for stream-flow forecasting: 2000-2015*" (Yaseen et al., 2015) los autores revisan los principales estudios que han utilizado técnicas de inteligencia artificial desde el año 2000 hasta el 2015, examinando técnicas como redes neuronales artificiales, máquinas de vectores de soporte y sistemas difusos en el pronóstico de caudales en diferentes cuencas hidrográficas del mundo. Los resultados muestran que los modelos basados en inteligencia artificial han mejorado significativamente la precisión de los pronósticos de caudales en comparación con los modelos hidrológicos tradicionales. Además, los autores destacan que se necesitan más estudios que comparen diferentes modelos de inteligencia artificial para determinar cuál es el más efectivo para cada situación hidrológica específica.

Wagena et al. (2020) evaluaron la eficacia de diferentes modelos para la predicción de caudales a corto plazo en una cuenca en Estados Unidos. Se utilizaron cuatro enfoques diferentes: modelos estocásticos de series temporales, redes neuronales, modelos basados en procesos y modelos bayesianos. Los resultados mostraron que RNA superaron a los otros modelos en términos de precisión de pronóstico. Además, se encontró que la combinación de RNA con modelos hidrológicos basados en procesos puede mejorar aún más la precisión de la predicción. También se destacó la importancia de seleccionar la cantidad adecuada de datos históricos para el entrenamiento del modelo, así como la influencia de las condiciones meteorológicas y de los datos de precipitación en la fidelidad de la predicción.

Finalmente, Sahoo et al. (2019) se enfocaron en el uso de redes neuronales recurrentes de memoria a largo plazo (LSTM) para predecir caudales bajos en series de tiempo hidrológicas. Los autores comparan el desempeño de la LSTM con otros modelos de pronóstico, incluyendo una red neuronal de retropropagación (BPNN) y modelos lineales de autorregresión (AR) y media móvil (MA). Los resultados muestran que la LSTM supera significativamente a los otros modelos en la predicción de caudales bajos, lo que sugiere su potencial como herramienta para la gestión de recursos hídricos.

Teniendo el estado del arte del tema en el cual se enmarca esta memoria, se procede a adentrarse en el funcionamiento de las redes neuronales y el porqué de la elección de una red LSTM para el propósito de predicción de caudales de crecidas.

2.3 Marco teórico Redes Neuronales Artificiales (RNA)

2.3.1 Idea y estructura RNA

Las redes neuronales artificiales (RNA) son un modelo computacional que se basa en el funcionamiento de las neuronas biológicas, y se interconectan de manera densa para procesar información mediante la recepción, procesamiento y transmisión de señales. Las

RNA se utilizan ampliamente en aplicaciones de inteligencia artificial y aprendizaje automático, como reconocimiento de patrones, clasificación, predicción y control. Estas redes se entrenan con algoritmos de aprendizaje para ajustar los pesos de las conexiones entre las neuronas, mejorando su precisión y capacidad para generalizar a nuevos datos. Desde su Génesis (Turing & Haugeland, 1950), las RNA han evolucionado y se han vuelto cada vez más sofisticadas, lo que ha permitido su aplicación en diversas áreas, como la medicina, la robótica y la gestión de recursos hídricos.

2.3.2 Redes propagación hacia adelante

Las redes *feedforward*, también conocidas como redes de propagación hacia adelante, son un tipo de red neuronal artificial en la que las señales fluyen solo en una dirección, desde las entradas hasta las salidas, sin retroalimentación. Las unidades de procesamiento se organizan en capas, donde cada capa está compuesta por un conjunto de neuronas. Las señales de entrada se propagan a través de la primera capa y se procesan sucesivamente en las siguientes hasta antes de alcanzar la salida final de la red. La Figura 2.2 muestra tres capas: una capa de entrada, una capa oculta y una capa de salida. La capa oculta consta de cuatro nodos y los pesos se representan por líneas con flechas y valores numéricos. Cada flecha representa una conexión entre dos nodos, y el número en la línea representa el peso de esa conexión.

La figura también muestra cómo se realiza el cálculo de la salida de la red para una entrada específica. Las entradas se propagan hacia adelante a través de la red, multiplicándose por los pesos de las conexiones y pasando por las funciones de activación en cada capa. Finalmente, se obtiene una salida de dos valores en la capa de salida, que representa la respuesta de la red para la entrada dada.

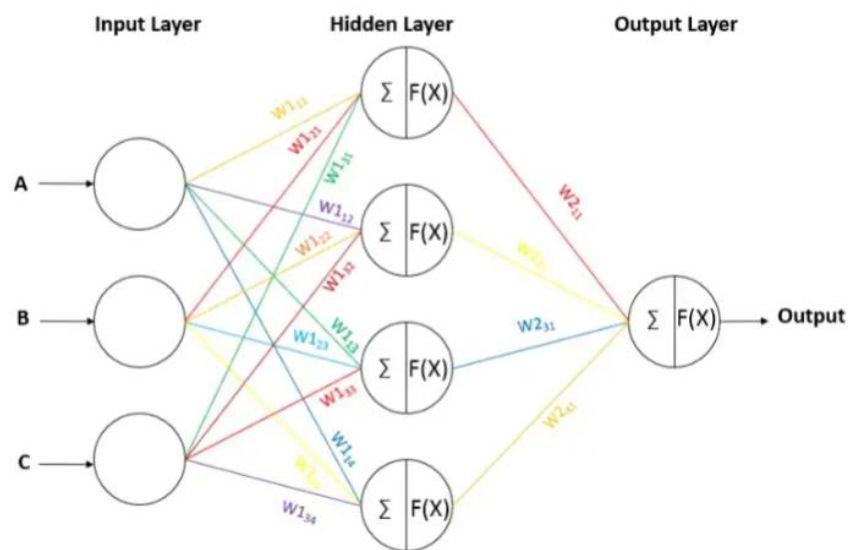


Figura 2.2: estructura de red *feedforward* (Sai,2020)

La ecuación de una red feedforward se puede expresar matemáticamente como la composición de varias funciones lineales y no lineales aplicadas a las entradas de la red. En cada capa de la red, se realiza una transformación lineal de las entradas, seguida de una función de activación no lineal, y el resultado se utiliza como entrada para la siguiente capa. La 1 presenta lo que sucede en cada neurona al recibir un *input* x .

$$y_i = f \left(\sum_{j=1}^n w_{ij} x_j + b_i \right) \quad 1.$$

donde:

x_j es el dato de entrada.

w_{ij} es el peso de la conexión entre las neuronas.

f es la función de activación de la neurona.

b_i es el sesgo (bias) de la neurona.

y_i Es la salida de la neurona.

2.3.3 Metodología de descenso de gradiente

Para partir explicando el funcionamiento de la actualización de parámetros de una red se define el aprendizaje supervisado. Es una técnica de aprendizaje automático que implica aprender a partir de datos etiquetados. En este enfoque, el modelo aprende a mapear las entradas a las salidas deseables a partir de ejemplos de entrenamiento, es decir, un conjunto de pares de entrada-salida. El objetivo de esta metodología es entrenar un modelo que pueda generalizar a nuevos ejemplos que no se han visto durante el entrenamiento. En el caso de esta memoria la salida deseada del modelo son los caudales de crecida previamente seleccionados.

La metodología aplicada por Rumelhart et al. (1986) que revolucionó la concepción de las redes neuronales y dio pie a todos los avances posteriores es la retropropagación (*backpropagation*), un algoritmo utilizado en el aprendizaje supervisado para calcular los gradientes de los pesos de una red neuronal. El objetivo es actualizar los pesos de la red de manera que la función objetivo se minimice. El backpropagation se realiza en dos fases: propagación hacia adelante y propagación hacia atrás. Durante la primera, se aplica una entrada a la red y se calcula la salida de la red. Durante la propagación hacia atrás, se calcula el error de la red y se propagan los gradientes hacia atrás a través de la red para ajustar los pesos de las conexiones.

La backpropagation utiliza el algoritmo de descenso de gradiente para ajustar los pesos de las conexiones. En cada iteración, se calcula el gradiente de la función objetivo (o de error)

con respecto a los pesos de las conexiones y se ajustan los pesos en la dirección opuesta al gradiente para minimizar la función de error.

La tasa de aprendizaje es un hiperparámetro crítico en el algoritmo de *backpropagation* ya que determina el tamaño de los pasos que se dan en la dirección opuesta al gradiente durante la actualización de los pesos en la red neuronal. En otras palabras, la tasa de aprendizaje controla la rapidez con la que el modelo se ajusta a los datos.

Si la tasa de aprendizaje es demasiado baja, el modelo tardará mucho tiempo en converger y, por lo tanto, en ajustarse a los datos. Por otro lado, si la tasa de aprendizaje es demasiado alta, el modelo puede saltar alrededor del mínimo global en lugar de converger, o incluso divergir y no ajustarse bien a los datos. En consecuencia, la tasa de aprendizaje debe ser cuidadosamente ajustada para obtener un buen rendimiento del modelo (Stevens et al., 2020).

El documento "*An overview of gradient descent optimization algorithms*" (Ruder, 2016) es un artículo de revisión que presenta una visión general de los algoritmos de optimización de descenso de gradiente. En el artículo se describen diferentes tipos de *backpropagation*, incluyendo el descenso de gradiente estándar, el descenso de gradiente estocástico, el descenso de gradiente con momento, el descenso de gradiente con Adagrad (Duchi et al., 2011), el descenso de gradiente con Adam (Kingma & Ba, 2014) y otros optimizadores globales. También se discuten las ventajas y desventajas de cada algoritmo y se presentan algunas recomendaciones para elegir el mejor algoritmo de descenso de gradiente para una tarea específica de aprendizaje automático el cuál, al igual que las funciones de activación no tiene una única solución y depende de cada caso (Wani et al., 2020).

2.3.4 Función objetivo

Para el aprendizaje supervisado la función objetivo o *loss function* determina la métrica de error para la cual se calculará el desempeño de la red, posteriormente se generará la optimización propiamente tal, por medio de la actualización de parámetros con el descenso de gradiente. A medida que el entrenamiento avanza se puede apreciar como esta función se aproxima a valores menores que los iniciales y eventualmente converge.

La familia de funciones de pérdida es amplia y se puede usar cualquier métrica de error entre dos series de datos. Entre los más conocidos se encuentra: error cuadrático medio o RMSE, error absoluto medio o MAE y coeficiente R^2 . Sus fórmulas matemáticas se presentan a continuación.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad 2.$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad 3.$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad 4.$$

En todas las ecuaciones n es el número de elementos demuestras, y_i es la etiqueta verdadera de la muestra i y \hat{y}_i es la salida predicha por el modelo para el instante de tiempo i . Para el coeficiente R^2 \bar{y} es la media de las etiquetas verdaderas y_i .

2.4 Redes recurrentes

En la Figura 2.3 se muestra la diferencia en la estructura y la conectividad entre una red neuronal normal y una RNN. Mientras que una red neuronal normal procesa una entrada de manera independiente en cada paso, una RNN procesa la entrada secuencialmente y utiliza la información anterior para informar el procesamiento en pasos posteriores. Esto hace que las RNN sean³-especialmente adecuadas para tareas que implican el procesamiento de datos secuenciales como series de tiempo, cadenas de texto, etc. (Pérez Guerrero, 2020). Las ecuaciones de RNN se exponen a continuación.

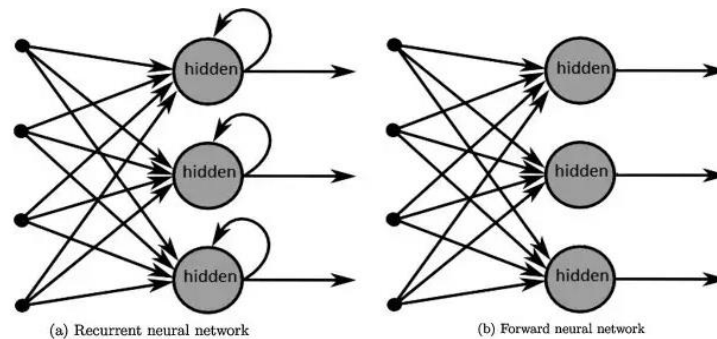


Figura 2.3: diferencias entre red *feedforward* y RNN (Radhakrishnan, 2017).

$$h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad 5.$$

$$y_t = W_{hy}h_t + b_y \quad 6.$$

En dónde:

W_{xh} es la matriz de pesos que conecta la entrada x_t al estado oculto h_t .

W_{hh} es la matriz de pesos que conecta el estado oculto anterior h_{t-1} al estado oculto actual h_t .

W_{hy} es la matriz de pesos que conecta el estado oculto h_t con la salida y_t .

b_h es el vector de sesgo del estado oculto.

b_y el vector de sesgo de la salida.

σ es la función de activación.

La función de activación es una función matemática que se utiliza en las redes neuronales para introducir no linealidades en los datos y permitir que la red pueda aprender patrones más complejos. En otras palabras, las funciones de activación permiten que la red neuronal pueda modelar relaciones no lineales entre las entradas y las salidas, lo cual es fundamental para resolver problemas de aprendizaje más complejos.

Cabe destacar que el estado anterior de h se utiliza para actualizar el estado actual y que la matriz de pesos W se comparte para todos los instantes de tiempo en cada celda, actualizándose con el descenso de gradiente en cada iteración. Este enfoque de compartir parámetros es una de las características clave de las RNN, ya que les permite modelar de manera efectiva las secuencias de entrada de longitud variable e identificar patrones o eventos en las secuencias de datos independiente de la posición en donde ocurran, esto es muy importante para el objetivo que se persigue, ya que se busca que el modelo sea capaz de identificar crecidas sin importar si estas ocurren en el inicio de la secuencia de datos o en cualquier punto intermedio. Además, esta propiedad de compartición de pesos también permite que las RNN sean entrenadas con eficiencia en comparación con las redes neuronales tradicionales, ya que la cantidad de parámetros en la red no aumenta con la longitud de la secuencia de entrada. En la Figura 2.4 se muestra una representación

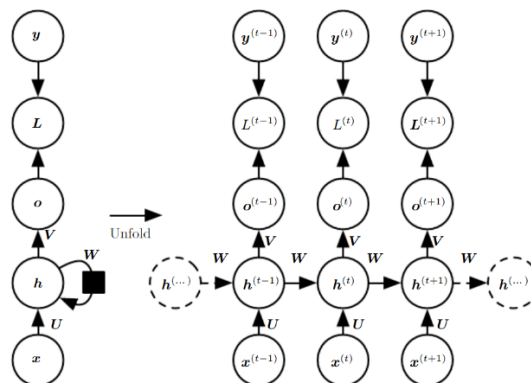


Figura 2.4: Red recurrente desenrollada

simplificada de una red recurrente desenrollada, que consta de una capa de entrada, una capa oculta y una capa de salida. En cada instante de tiempo, se presenta una entrada x_t a la red, y la salida correspondiente y_t se calcula utilizando la información de la entrada actual

y los estados ocultos generados en el instante de tiempo anterior. La red está conectada consigo misma a través del tiempo, lo que significa que los estados ocultos se retroalimentan desde una iteración a la siguiente.

Si bien esta arquitectura permite el procesamiento de datos secuenciales, sufren un problema conocido como el "problema de desvanecimiento de gradiente". Este inconveniente se produce cuando los gradientes (es decir, las derivadas parciales de la función de pérdida con respecto a los pesos de la red) se hacen muy pequeños a medida que se propagan hacia atrás a través de la red. Esto puede generar que los pesos de las conexiones no se actualicen correctamente, lo que a su vez puede afectar negativamente el rendimiento de la red.

El problema de desvanecimiento de gradiente se debe a la naturaleza recurrente de las RNN, que hace que los gradientes se multipliquen repetidamente a medida que se propagan hacia atrás en el tiempo. En otras palabras, cuanto más tiempo haya pasado desde que se introdujo la información inicial en la red, más pequeños serán los gradientes y más difícil será para la red aprender y hacer predicciones precisas.

2.5 Red LSTM

Las redes neuronales LSTM (Long Short-Term Memory) son una variante de las redes neuronales recurrentes (RNN) que a diferencia de las RNN convencionales pueden aprender a mantener información a largo plazo y evitar el problema de desvanecimiento de gradientes. Para lograr esto, las LSTM incorporan mecanismos de "olvido" de memoria llamados celdas, las cuales contienen compuertas que les permiten descartar información antigua y centrarse en la información más reciente y relevante. El olvido de memoria en las RNN permite que la red sea más eficiente en la utilización de los recursos computacionales y reduce la cantidad de información redundante que se almacena. Esto mejora la capacidad de la red para procesar datos secuenciales y contribuye a mejorar su capacidad para generalizar a datos nuevos y no vistos previamente.

Su arquitectura presenta un diseño basado en cuatro puntos importantes:

- Memoria a largo plazo (o celda de memoria, denotado por C_t).
- Compuertas (*forget*, *input* y *output*, denotadas por f_t , i_t y o_t respectivamente).
- Emisión de valores candidatos (información candidata a quedarse en memoria, denotada por \tilde{c}_t).
- Un estado oculto h_t

La celda LSTM se presenta en la Figura 2.5, cada letra en el diagrama representa una operación matemática diferente. La operación "+" representa la suma de dos matrices y la operación "x" representa la multiplicación de dos matrices. La operación "sig" es la función de activación sigmoide y la operación "tanh" es la función de activación tangente hiperbólica.

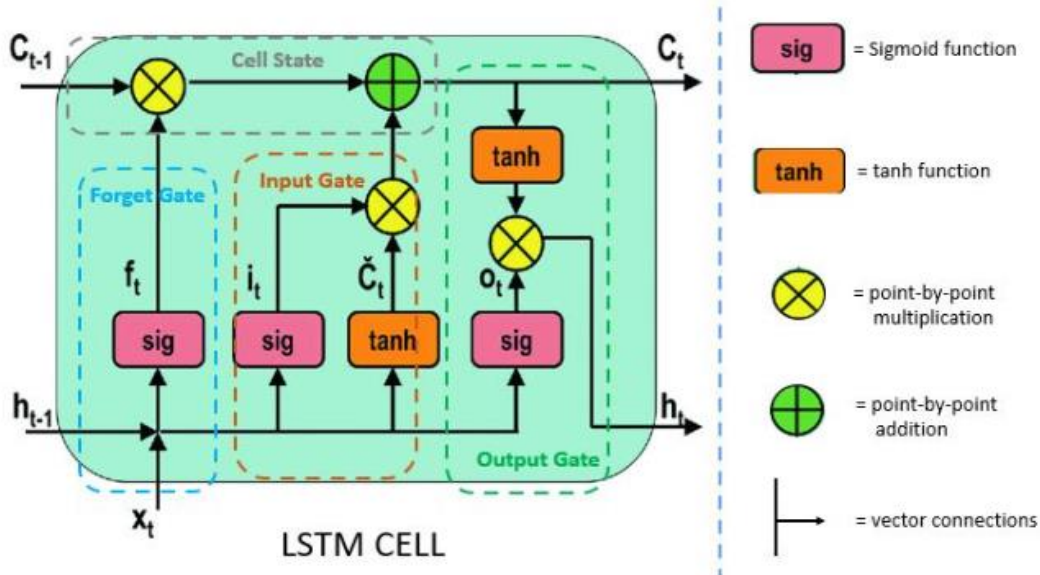


Figura 2.5: Celda LSTM

De forma similar a las RNN, las LSTM generan estados ocultos h_t , con la diferencia de que en las LSTM estos dependen de las compuertas, las cuales se describen a continuación:

1. Compuerta de olvido (*forget gate*): Esta compuerta determina cuánta información se debe olvidar de la celda de memoria. Toma como entrada la información de la entrada anterior y la salida anterior y produce un número entre 0 y 1 para cada unidad de la celda de memoria. Un valor de 0 indica que la información debe olvidarse completamente, mientras que un valor de 1 indica que la información debe mantenerse completamente.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad 7.$$

2. Compuerta de entrada (*input gate*): Esta compuerta decide cuánta información debe agregarse a la celda de memoria. Toma como entrada la información de la entrada y la salida anteriores y produce un número entre 0 y 1 para cada unidad de la celda de memoria.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad 8.$$

3. Compuerta de salida (*output gate*): Esta compuerta decide cuánta información se debe emitir desde la celda de memoria. Toma como entrada la información de la entrada anterior y la salida anterior y produce un número entre 0 y 1 para cada unidad de la celda de memoria. Este valor se utiliza para multiplicar la salida de la celda de memoria antes de enviarla a la siguiente capa de la red.

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad 9.$$

Donde x_t es la entrada en el tiempo t , h_{t-1} es la salida de la celda en el tiempo anterior, c_{t-1} es el estado de la celda LSTM en el tiempo anterior, i_t es la puerta de entrada, f_t es la puerta de olvido y o_t es la puerta de salida. σ es la función sigmoide y W y b son los pesos y sesgos respectivamente.

De esta manera, la ecuación para actualizar el estado de la celda es una combinación de la información que fluye las compuertas de entrada y olvido, tal como se muestra en la ecuación a continuación.

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad 10.$$

Donde \tanh es la función tangente hiperbólica. Finalmente, el estado oculto h_t se obtiene a través de la compuerta de salida, siguiendo la siguiente ecuación.

$$h_t = o_t \tanh(c_t) \quad 11.$$

Capítulo 3: Zona de estudio

3.1 Caracterización zona de estudio

El territorio analizado en este estudio pertenece a la sexta región del Libertador Bernardo O'Higgins, la cual tiene una superficie de 16.387(km²) y sus límites territoriales son por el norte la Región Metropolitana y de Valparaíso, a través del río Rapel y la hoya del alto río Maipo; por el este la República de Argentina; por el sur la Región del Maule, por la hoya del río Tinguiririca; y por el Oeste el Océano Pacífico. La ubicación espacial se presenta el mapa 2 de la Figura 3.1.

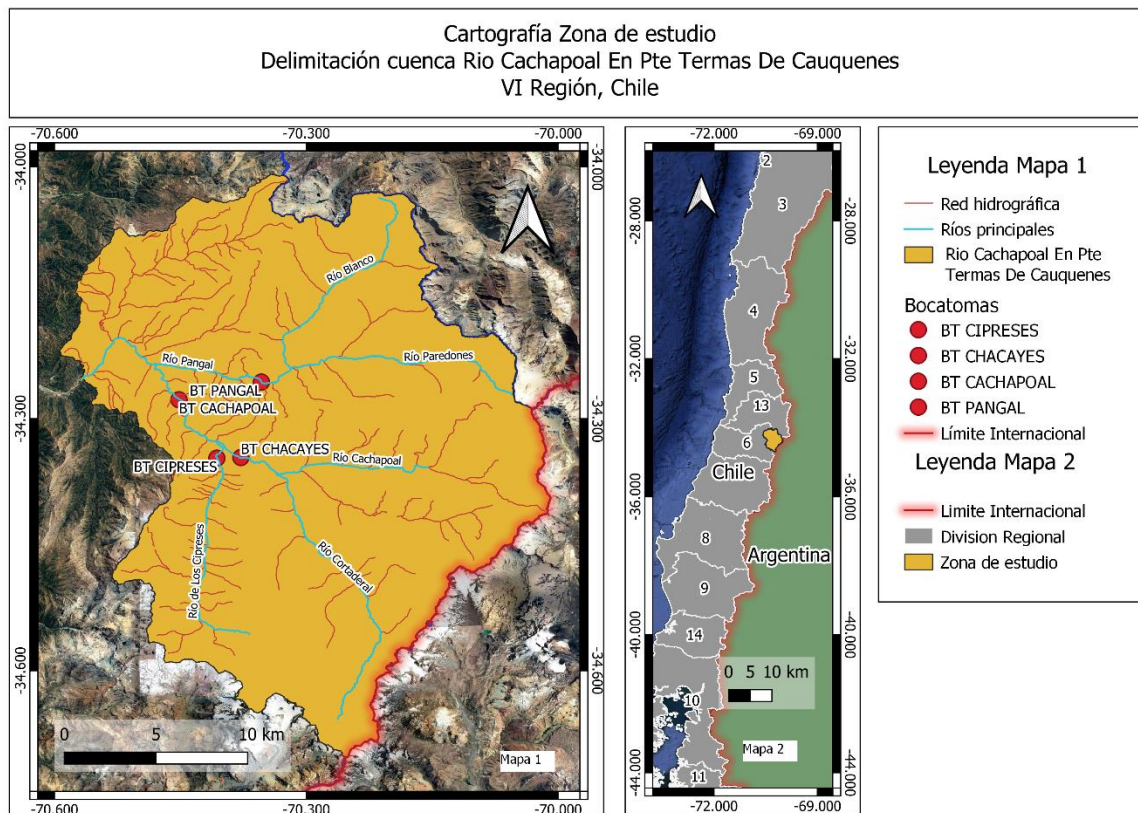
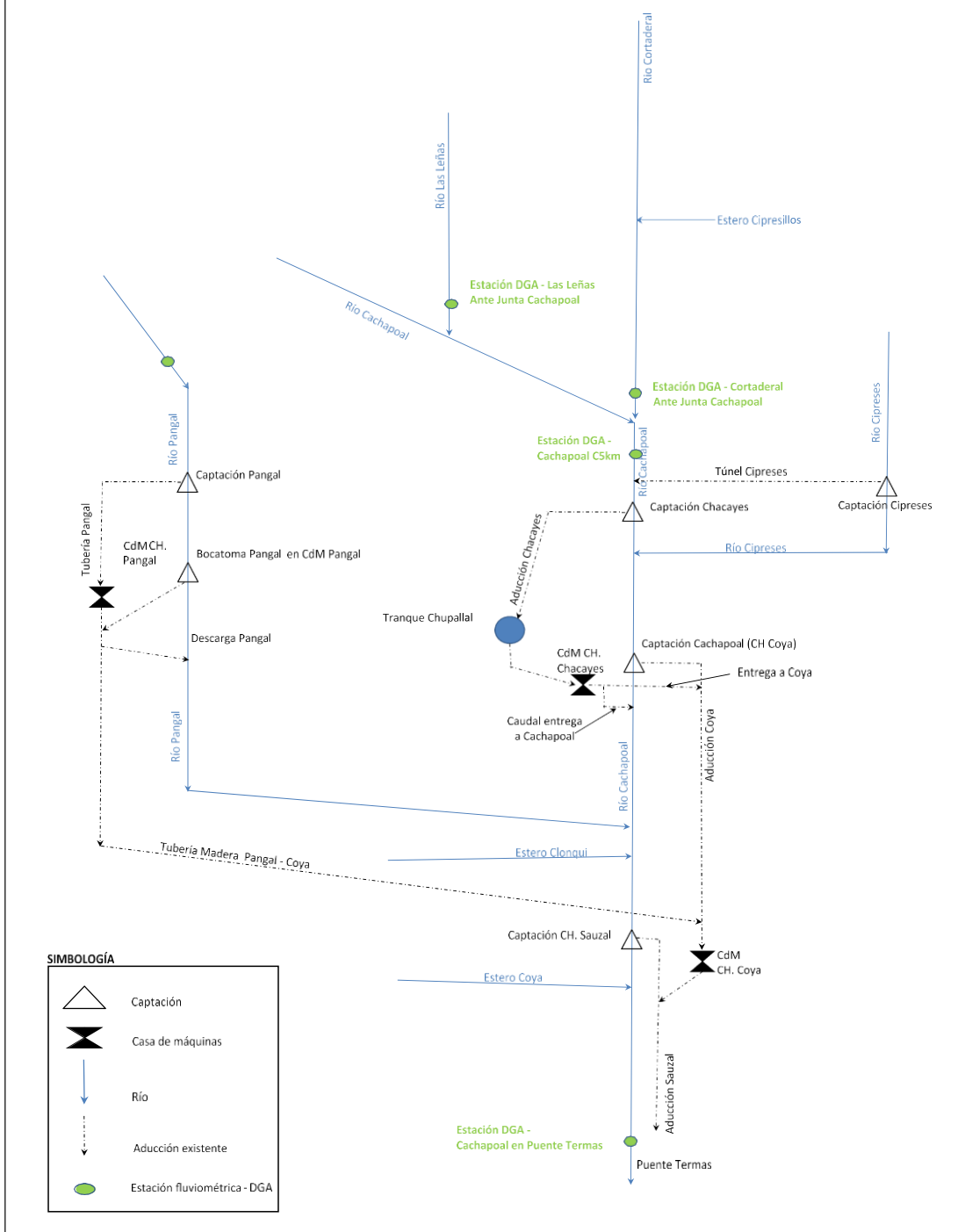


Figura 3.1: Cartografía Cuenca Río Cachapoal En Pte Termas De Cauquenes con bocatomas de interés

La cuenca Río Cachapoal En Pte Termas De Cauquenes definida por la estación fluviométrica del mismo nombre posee tres subcuencas en la zona cordillerana definidas por los ríos Cachapoal (cauce de salida), Pangal y Cipreses. Como se observa en el mapa de la Figura 3.1 en cada tributario existen bocatomas pertenecientes a un sistema interconectado de generación hidroeléctrica, que data de más de 100 años para la bocatoma Cachapoal y Pangal construidas en 1911 y 1921 respectivamente. La bocatoma Chacayes y Cipreses son parte de un proyecto conjunto, puesto en marcha el mes de noviembre del año 2011. En la Figura 3.2 se presenta el diagrama unifilar de la zona de estudio.



Elaboro: DC / Dic 2018

Figura 3.2: Diagrama unifilar sistema de generación hidroeléctrica (PHC,2018)

La Figura 3.3 representa espacialmente la ubicación de las subcuencas de estudio junto con las estaciones de medición presentes.

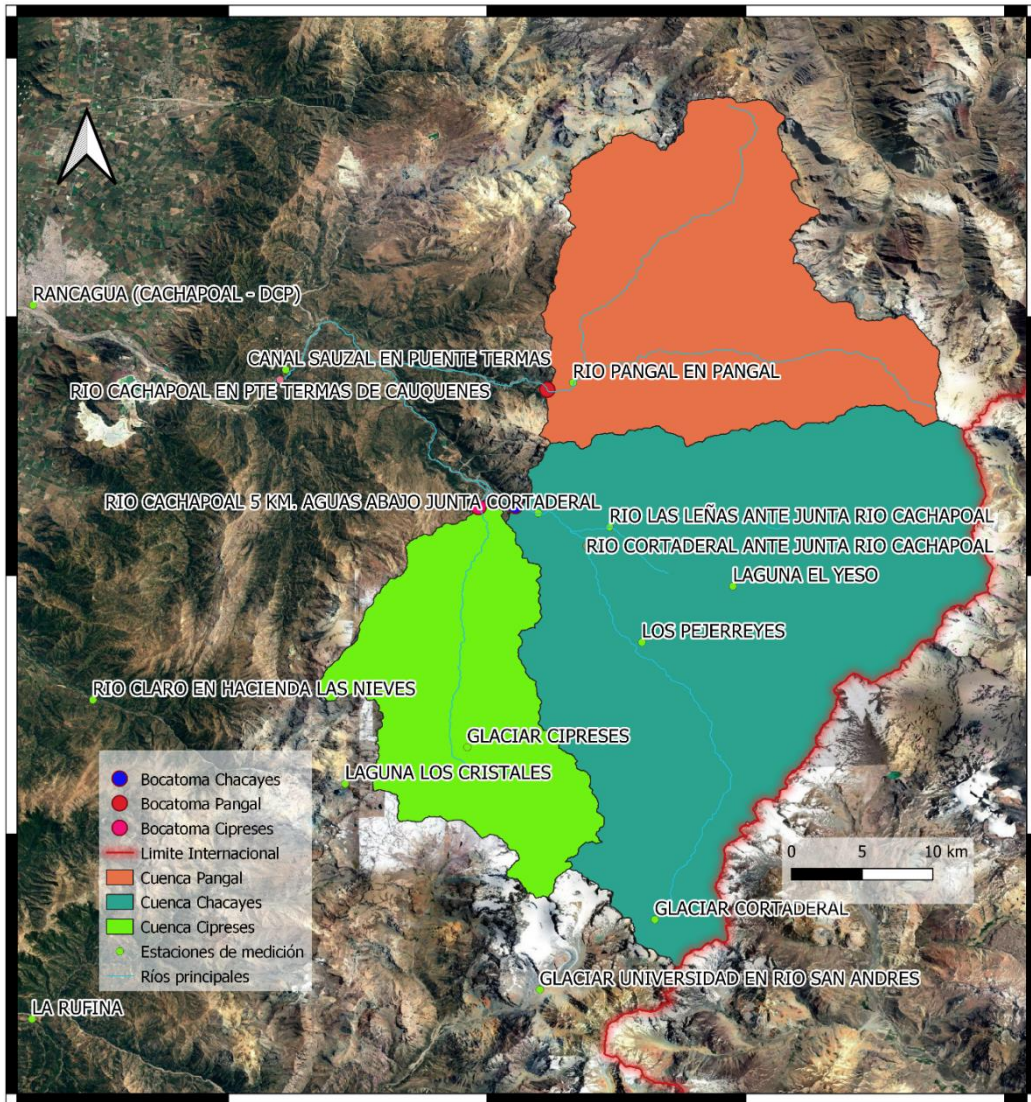


Figura 3.3: Ubicación cuencas de estudio y estaciones de medición.

3.2 Hidroclima y relieve

Las características del relieve más importantes en la región son la presencia de cuatro fajas longitudinales: Cordillera de los Andes, depresión intermedia, Cordillera de la Costa y planicies costeras. Cada una de estas moldea los procesos hidrológicos como precipitación, escorrentía, acumulación nival e infiltración según sus características de vegetación, pendientes de terreno, tipo de suelo, tipo de uso de suelo, entre otros.

Para las cuencas de interés se presenta el clima específico de la Cordillera de los Andes. El paulatino ascenso del relieve en la región de la cordillera provoca cambios de temperatura y pluviométricos de gran importancia. Por lo tanto, al aumentar la altitud, la temperatura desciende hasta los 0 °C, y el agua caída, en estado líquido y sólido, aumenta de 600 mm a 1.500 mm a una altitud superior a los 3.500 m sobre el nivel del mar. La severidad del clima aumenta hasta llegar a tener temperaturas de frío extremo, y es por lo que esta zona tiene alrededor del 40% de los glaciares del centro de Chile (IPG,2022). La mayor parte de la precipitación en esta zona cae sobre los nevados más altos, lo que lo convierte en un reservorio de agua en invierno, la cual se derrite en primavera y verano y se utiliza para el riego.

Debido al clima específico cordillerano es que el régimen de la cuenca es nivo-pluvial, con caudales máximos en el mes de diciembre. En la Figura 3.4 se exponen las curvas de variación estacional para ambas cuencas en base a los datos de caudales medios mensuales obtenidos de la DGA entre los años 1990 y 2022. En base a la experiencia de la gente de operación de bocatomas se sabe cuándo empieza la temporada de deshielos debido a que la turbiedad con la que llega el agua a la captación es mayor debido al arrastre de sedimento producto del derretimiento de nieve y glaciares.

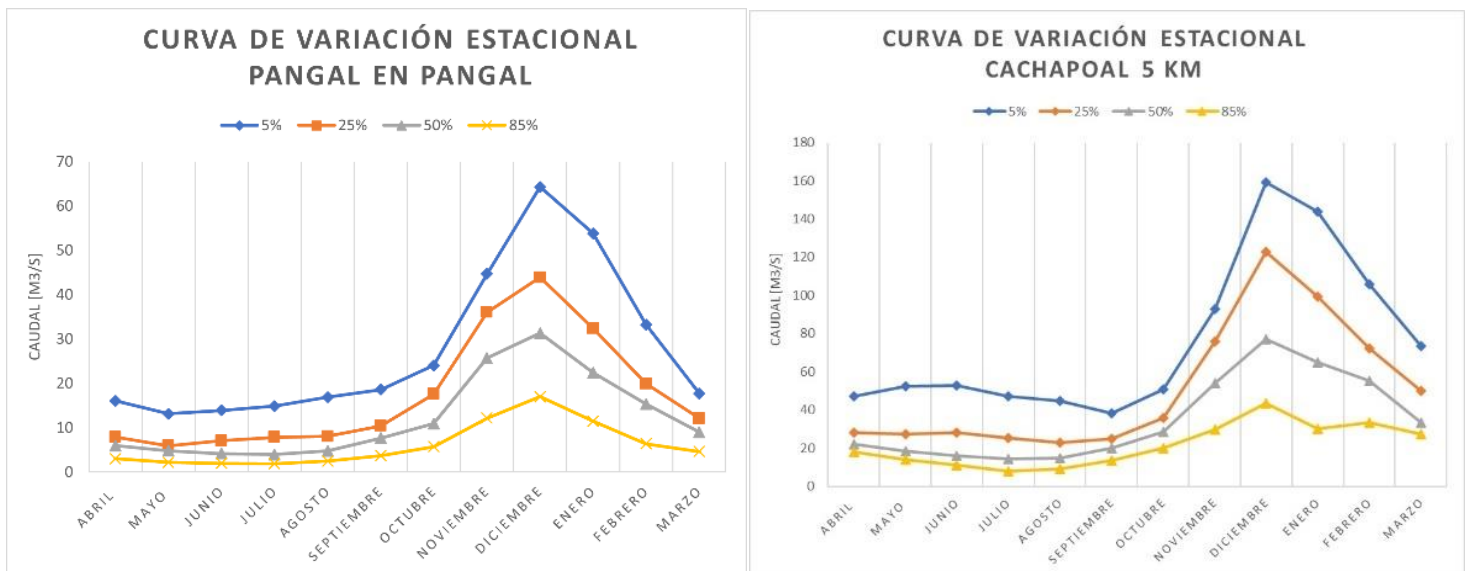


Figura 3.4: Curvas de Variación Estacional cuencas Pangal y Cachapoal

3.3 Disponibilidad de datos

Como primer paso en el desarrollo de un modelo hidrológico se encuentra el catastro de la información meteorológica existente, junto con el chequeo de consistencia y extensión de la serie temporal de medición para cada estación de estudio. En Chile, la información recopilada por la red hidrométrica pública es distribuida oficialmente por la DGA a través de la plataforma del sistema de información ambiental del Ministerio de Obras Públicas (DGA, 2022). Se analizaron todas las estaciones disponibles en la zona presentando los datos de estas en la Tabla 3.1.

Tabla 3.1: Estaciones Consultadas.

Estación	Código BNA	UTM Este	UTM Norte	Elevación	Fecha Inicio	Variables medidas
Canal sauzal en puente termas.	06008009-7	356748	6210031	750	01-07-2005	Caudal y precipitación
Laguna los cristales.	06013008-6	361491	6174592	2343	09-05-2012	Precipitación y temperatura
Río Cachapoal 5 km. Aguas abajo junta cortaderal.	06003001-4	374870	6197990	1127	01-12-1989	Precipitación, temperatura y caudal
Río las Leñas ante junta río Cachapoal.	06000003-4	379964	6196870	1296	23-11-2006	Precipitación, temperatura y caudal
Río Pangal en Pangal.	06006001-0	377203	6209222	1500	01-12-1985	Caudal y precipitación
Laguna el yeso.	06000004-2	388788	6191905	2140	20-05-2012	Precipitación y temperatura
Glaciar Cortaderal.	06001000-5	383606	6163250	3156	01-04-2018	Precipitación y temperatura
Río Cortaderal ante junta río Cachapoal.	06002001-9	378334	6195231	1200	01-01-1985	Caudal
Río Cachapoal en pte termas de cauquenes.	06008005-4	356342	6209147	700	01-10-1974	Caudal

Se aplican metodologías iniciales en las series de precipitación y temperatura, se cuenta los datos horarios de cada forzante por año. Para caudales, se usa la información diaria para generar un mapa de calor. La Figura 3.5 muestra para cada año la densidad de información,

la escala normaliza la cantidad de datos disponibles de horas en el año implicando que si está de color azul posee información completa.

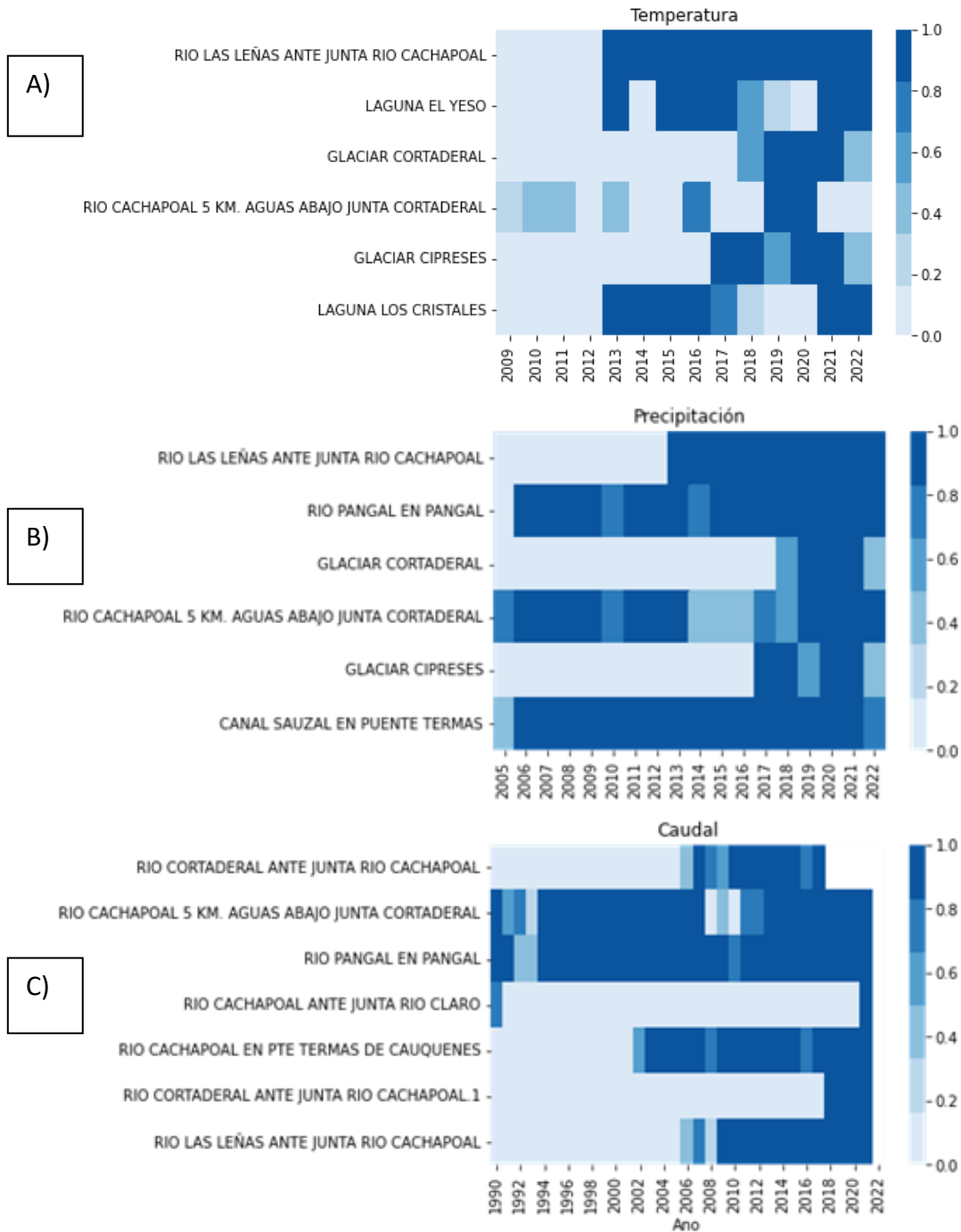


Figura 3.5: Mapas de calor de información por año para las estaciones seleccionadas.

A) Temperatura horaria, B) Precipitación horaria y C) Caudales diarios.

3.4 Selección de crecidas

Para seleccionar las crecidas que alimentarán el modelo se realiza un análisis de estudios pasados de crecidas (PHC,2017) en donde se identifican las más importantes y además se realiza una búsqueda manual en toda la extensión de la serie de precipitación y caudal para identificar subidas repentinas de caudal como producto de un evento de precipitación. También se cuantifican los períodos de retorno de estas en base al mismo estudio. En la Tabla 3.2 y Tabla 3.3 se exponen los períodos de retorno adoptados para las cuencas Pangal y Cachapoal (PHC,2017).

Tabla 3.2: Períodos de retorno estación Pangal en Pangal (PHC,2017)

Período de retorno en años	Caudal (m ³ /s)
2	81
5	149
10	206
25	290
50	361
100	441
250	559
500	660
1000	770
10000	1220

Tabla 3.3: Períodos de retorno estación Cachapoal 5km Aguas abajo junta Cortaderal (PHC,2017)

Período de retorno en años	Caudal (m ³ /s)
2	142
5	261
10	372
25	535
50	669
100	811
250	1009
500	1167
1000	1330
10000	1911

3.4.1 Crecidas Pangal

Para la estación Pangal en Pangal se selecciona un total de 27 crecidas de caudal producto de precipitaciones a las cuales se les aplica un filtro por magnitud y calidad de la información. Aquellas que no presentaran un caudal mayor a 50 m³/s representando aproximadamente un año de período de retorno se les descartó debido a que se busca que el modelo pueda aprender las dinámicas de los datos en los eventos que tienen potencial peligro. Para Pangal los períodos de retorno contemplan desde 2 a 25 años teniendo una buena variedad de crecidas medianas y una de mayor magnitud que se destina al testeado de los resultados. En la Tabla 3.4: Correlación Precipitación Caudal Crecidas Pangal se exponen los coeficientes R² obtenidos a partir de las series horarias de crecidas en la estación Pangal en Pangal. Los valores fluctúan entre 0,3 y 0,8 indicando que existen correlaciones medias y altas entre ambas variables en el tiempo. En la Figura 3.6 se exponen gráficamente las crecidas, en forma de línea se plasma el caudal y como gráfico de barra con el eje invertido la precipitación.

Tabla 3.4: Correlación Precipitación Caudal Crecidas Pangal y período de retorno

Crecidas Pangal	Período de retorno aproximado	Correlación R² PP-Caudal
2002-05-25	3 años	0,70
2002-08-24	4 años	0,53
2005-06-27	1,5 años	0,41
2005-08-27	2 años	0,87
2012-05-27	15 años	0,66
2016-04-16	22 años	0,45
2017-12-31	8 años	0,36
2021-01-30	12 años	0,71

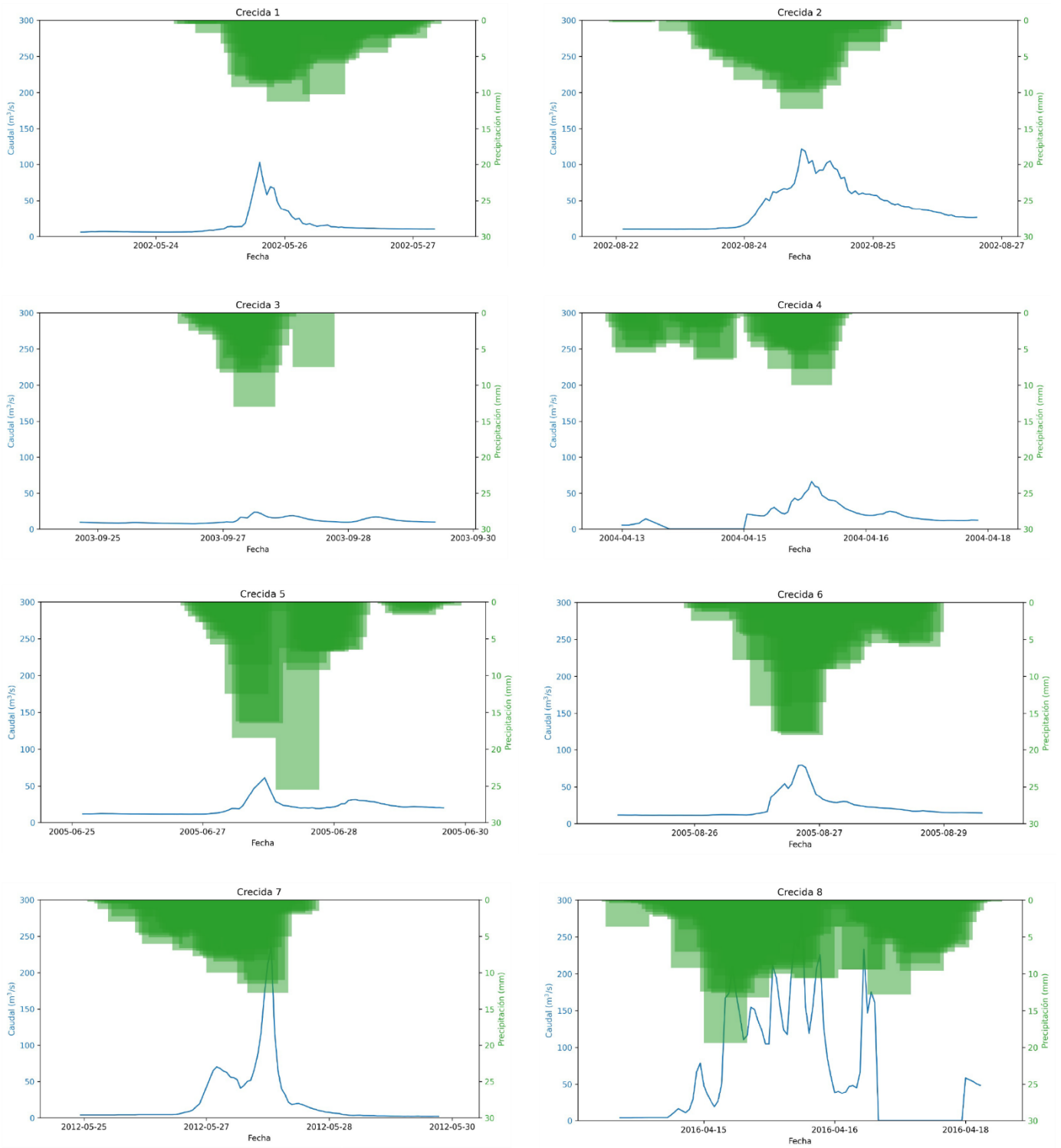


Figura 3.6: Set de crecidas cuenca Pangal

3.4.2 Crecidas Cachapoal 5km

Para la estación Cachapoal se realiza el mismo procedimiento de selección de crecidas en dónde se eligen primeramente 20 crecidas y luego son reducidas a once para fines de la modelación. En la Figura 3.7 se presentan las crecidas seleccionadas para la estación de Cachapoal 5 km. Es importante señalar que la mayoría de estas se sitúan antes del año 2005, tiempo en el que no se tenían mediciones de otras estaciones consultadas. Otro punto por recalcar es la falta de información que existe entre los años 2009 y 2012 siendo una pérdida importante de datos.

En la Tabla 3.5 se exponen los coeficientes R^2 obtenidos a partir de las series horarias de crecidas en la estación Cachapoal 5km bajo junta Cortaderal. Los valores fluctúan entre 0 y 0,5 indicando que existen correlaciones bajas y medias entre ambas variables en el tiempo. Además, se exponen los valores de período de retorno en años para cada crecida en donde la mayor contempla 12 años y todo el resto siendo menores a 10 años.

Tabla 3.5: Correlación entre datos de precipitación y caudal Cachapoal

Crecida Cachapoal	Período de retorno aproximado	Correlación R^2 PP-Caudal
2000-06-30	6 años	0,01
2001-07-19	2 años	0,11
2002-08-24	4 años	0,27
2003-01-21	4 años	0,43
2003-06-20	1 año	0,47
2004-04-14	12 años	0,58
2016-04-15	2 años	0,40
2017-02-26	2 años	0,28

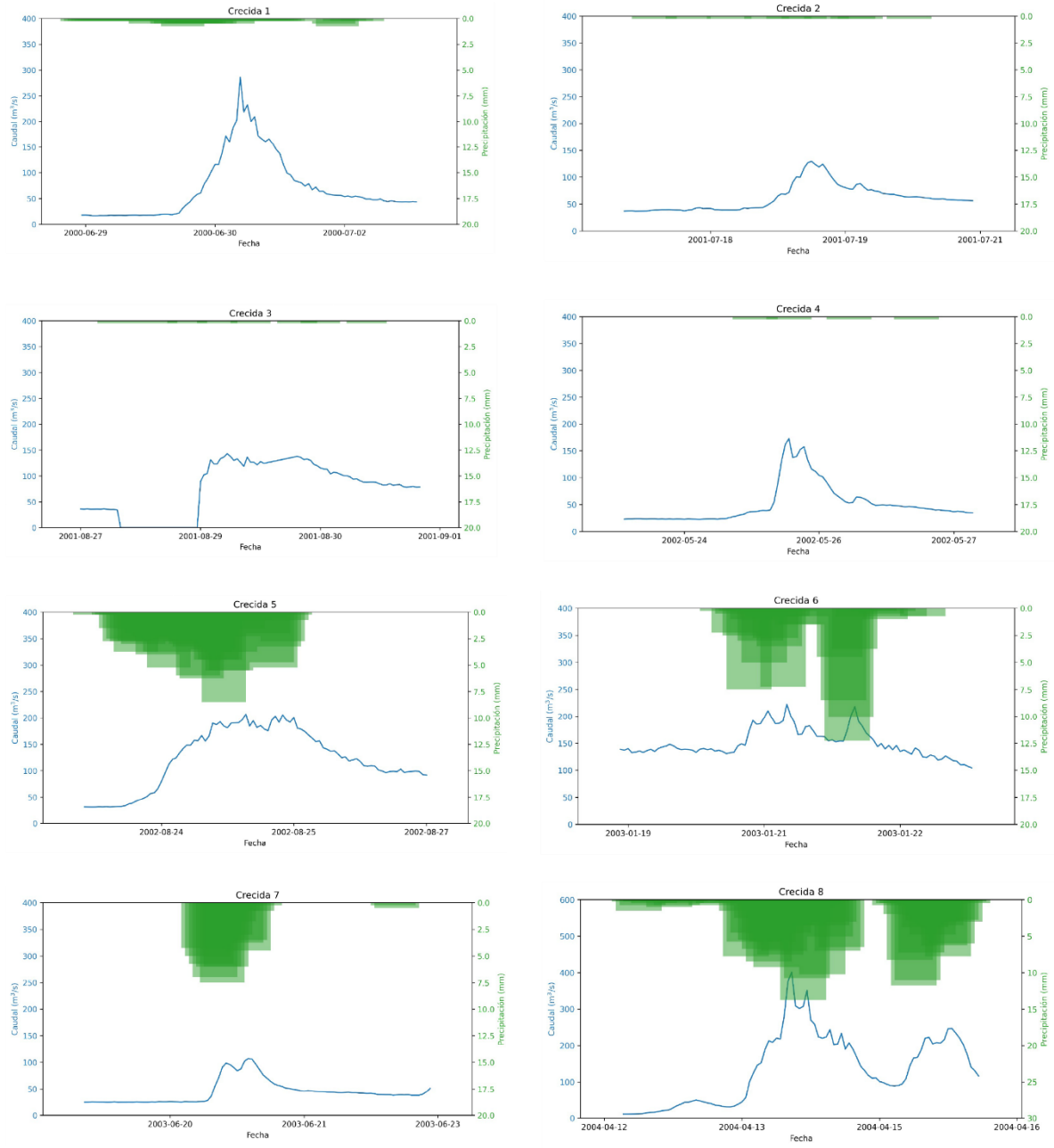


Figura 3.7: Set crecidas cuenca Cachapoal

Capítulo 4: Metodología

4.1 Metodología

La metodología de trabajo para lograr los objetivos planteados contempla seis partes fundamentales que están presentes en el esquema presentado en la Figura 4.1: Esquema metodología de trabajo

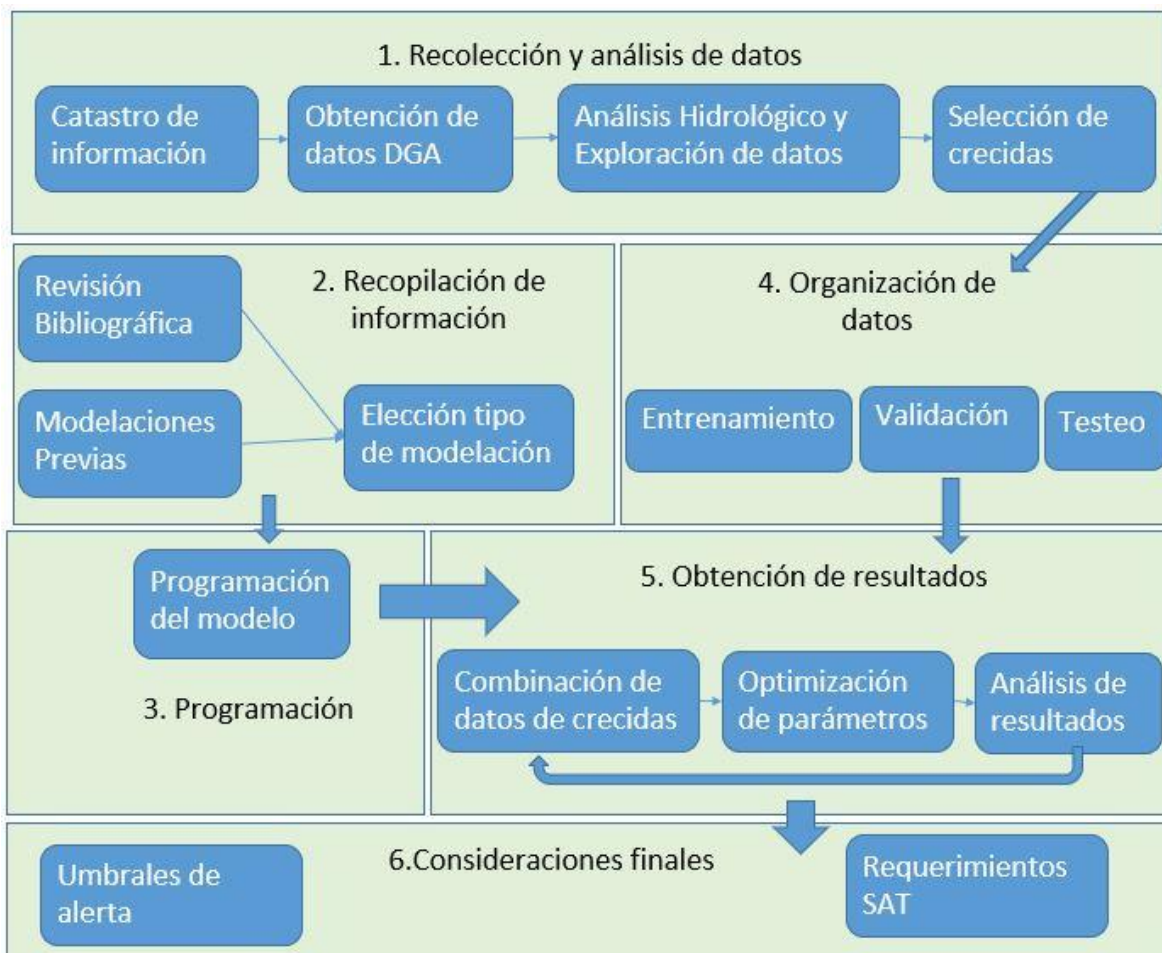


Figura 4.1: Esquema metodología de trabajo.

4.2 Recolección y análisis de datos

El paso inicial de la modelación, luego de haber recopilado la información en bruto, es realizar la metodología *Exploratory Data Analysis* (EDA), con el objetivo de explorar, investigar y aprender de los datos en base a gráficos y visualizaciones. Para esto se aplican

distintas funciones y manejo de *DataFrames* con la ayuda de la librería Pandas de Python (McKinney, 2010).

Principalmente lo que se realiza en esta parte es el manejo de datos de precipitación y temperatura a nivel horario, cuantificando los valores *NaN* (Not a Number), valores atípicos, frecuencias de medición diarias y el estado específico de los datos para los rangos en dónde existen subidas de caudal, ya que la forma de operar la información es teniendo como base períodos de 100 horas de caudales que son concatenadas por fecha con las series completas de precipitación y temperatura.

A partir de este análisis de datos es posible identificar los problemas locales de manera asertiva. En las estaciones Cachapoal 5 km y Las Leñas existen más de 2.000 días aproximadamente que poseen más de un dato de precipitación por hora de medición, de hecho, hasta 4 datos en una hora, lo que supone una dificultad para la entrega de la información al modelo LSTM, por el contrario, en la estación Pangal la precipitación posee información consistente de una medición por hora. Complementando el análisis, se desprende que dentro de las estaciones con frecuencias más altas de mediciones por hora generalmente hay valores atípicos.

Cabe destacar que el problema global que tienen los datos es significativo y para poder cumplir los objetivos del trabajo se necesita entrenar una red con series de este largo y sin datos *NAN* en ningún *DataFrame*. Se decide descartar todas las estaciones consultadas exceptuando Pangal en Pangal y Cachapoal 5km debido a la gran pérdida de información que generan al ser combinadas con las crecidas identificadas para las estaciones mencionadas. Esto principalmente debido a que casi la mitad de las crecidas utilizadas son previas a 2005 y no hay mediciones de otras estaciones en esas fechas. Además, al utilizar temperatura (que comienza desde 2013) en los rangos de datos importantes no se encuentra una serie consistente por lo que no aporta información adicional de valor, por lo que se descarta su uso debido a la falta de información de las crecidas con mayores períodos de retorno. Finalmente, se decide realizar dos modelos precipitación escorrentía para las bocatomas Pangal y Chacayes con la información exclusiva de las estaciones previamente mencionadas.

4.3 Programación del modelo

El trabajo de programación del modelo difiere de lo que se realiza en Pandas con los *DataFrames* y se adentra en la librería de *Deep Learning Pytorch* (Paszke et al., 2019).

Para los efectos de este informe los códigos generados no serán expuestos debido a que corresponde a material intelectual de PHC, es por esto por lo que la metodología de programación del modelo se presenta de manera descriptiva, escrita y, en la Figura 4.2 de forma esquemática.

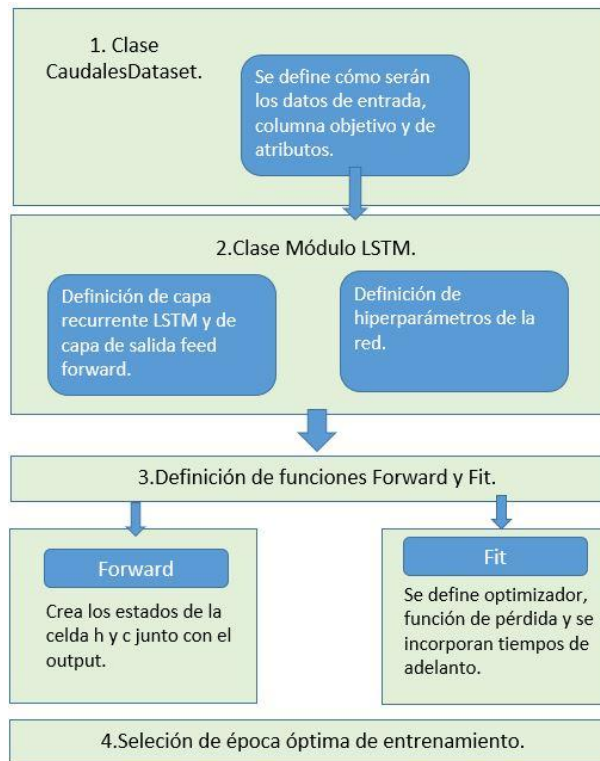


Figura 4.2: Esquema de pasos programación del modelo.

- Se genera la clase CaudalesDataset que define los batches, columnas de atributos y columnas objetivo.
- Los DataFrames se convierten en tensores mediante la incorporación de una nueva dimensión y se requiere de un largo de serie igual para la dimensión que contiene los datos.
- Se importan los complementos nn.Module para la generación de la clase ModelLSTM.
- Dentro de ModelLSTM se definen las capas recurrentes LSTM y la capa de salida de función lineal.
- Se seleccionan los hiperparámetros base de la red como la tasa de aprendizaje, cantidad de atributos de entrada, número de capas ocultas, tamaños de batch de entrenamiento, tamaños de batch de validación y el número de épocas o iteraciones.
- La función forward crea los estados de la celda h y c junto con el output propiamente tal. Se genera un tensor $z = \text{self.FC}(\text{output})$ como retorno de forward.
- La función Fit es dónde ocurre el entrenamiento, la optimización y el backpropagation según la función de pérdida elegida.
- Se utiliza el optimizador Adam y la función de pérdida MSELoss incorporadas en los módulos de torch.
- El modelo está operativo para ser entrenado con los datos recolectados de las cuencas de Pangal y Chacayes.

4.3.1 Tiempos de adelanto

Una dificultad presente en este tipo de modelo es la capacidad de predecir a más de un paso temporal hacia el futuro (una hora en este caso), debido a la falta en el *input* de precipitación, pues el modelo predice caudales y pueden ser utilizados como *input* para aumentar la distancia de predicción, no así la precipitación. Para resolver este desafío se implementa un desfase en la función *Fit*, “engañando” a la red para lograr predicciones en pasos temporales mayores a una hora. La Figura 4.3 muestra esquemáticamente lo que ocurre al aplicar el desfase de la red respecto de los parámetros internos “ h_t ” y las salidas “ y ” en que las celdas reciben *inputs* con un desfase llamado “ dt ” el cuál será los pasos temporales hacia adelante que predecirá sus resultados. En esta memoria se plantearán modelos para pronósticos de 1,3 y 5 horas de adelanto.

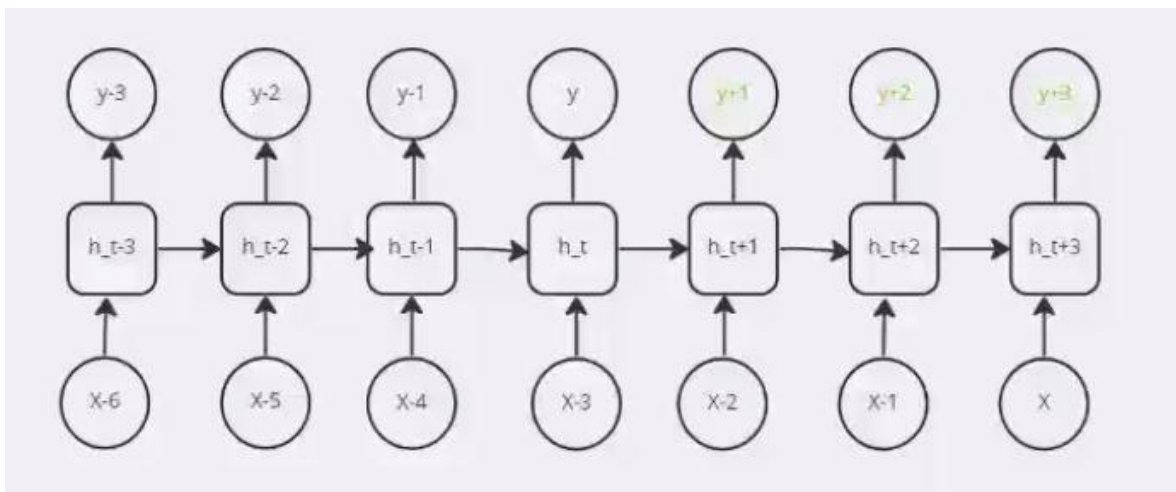


Figura 4.3: Desfase de red para generar tiempos de adelanto

4.4 Obtención de resultados

La obtención de resultados es un proceso de prueba y error en el que se deben seguir los tres pasos mencionados en la Figura 4.4 y repetir el proceso hasta dar con los mejores desempeños del modelo.

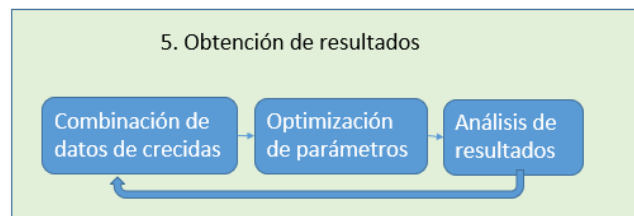


Figura 4.4: Esquema de obtención de resultados

4.4.1 Combinación de *DataFrames* de crecidas

Para lograr el objetivo principal de alertar en caso de crecida es de suma importancia asegurarse de que el modelo va a poder predecir la magnitud de los caudales de períodos de retorno de más de 10 años, y es esta la principal razón de por qué no se incluyen más columnas con información de otras estaciones dado que se pierden las crecidas de mayores caudales, limitando las posibilidades de predecir las crecidas.

Primero es importante entender la función de los sets de entrenamiento, validación y testeo.

- **Entrenamiento:** Son los *DataFrames* que utilizará la red para aprender de los datos y sus correlaciones mediante la iteración de sus parámetros internos con el algoritmo de *backpropagation*. Estos deben contener por lo menos una crecida de período de retorno mayor a 10 años para que sea posible que la red prediga una crecida de magnitud, por lo menos igual, a la más grande de entrenamiento.
- **Validación:** Es una parte clave del entrenamiento dado que el set de validación rige cuál es el modelo óptimo dependiendo de la función de pérdida de validación, es decir, el set de validación se utiliza en el entrenamiento, pero no para actualizar los parámetros de la red sino que juega un papel de testeo dentro del entrenamiento en donde el *DataFrame* es pasado por el módulo LSTM cada 5 iteraciones y se calcula su función de pérdida arrojando un número llamado *val loss*. Este es crucial para el desempeño del modelo y es recomendable tener más de un set de datos de validación, en este caso la información es escasa y se decide priorizar el entrenamiento por lo que solamente se escoge un *DataFrame* de esta clase por modelo. Con la experiencia obtenida de los numerosos entrenamientos realizados se decide que no se utilizarán las dos crecidas más grandes en validación ya que para el objetivo final es requerido poder entrenar y testear con esos sets.
- **Testeo:** Son los datos que no fueron entregados a la red en ningún momento y se utilizan netamente para evaluar el desempeño de esta. En Pangal se utiliza la crecida mayor para testeo debido a que se tienen otras que alcanzan caudales máximos relativamente cercanos por lo que es posible para la red llegar a valores cercanos a los máximos registrados en todo el rango de tiempo de estudio. Por otro lado, en Cachapoal se tiene una crecida que supera con creces a las demás y es requerida en los sets de entrenamiento dado que sino la red no supera los máximos de entrenamiento y no sería de utilidad para predicción de crecidas peligrosas.

Con los *DataFrames* de entrenamiento, validación y testeo (conjuntos con intersección vacía) escogidos llega el momento de optimizar los hiperparámetros de la red y así encontrar la mejor configuración para esa elección de set de datos. En la Tabla 4.1 y Tabla 4.2 se presentan las crecidas de testeo y validación para cada modelo.

Tabla 4.1: Crecidas de validación y testeo modelo Cachapoal

Fecha crecida	Caudal máximo instantáneo	Período de retorno	Utilización
2000-06-30	286 (m ³ /s)	6 años	Testeo
2002-08-24	207 (m ³ /s)	4 años	Validación

Tabla 4.2: Crecidas de validación y testeo modelo Pangal

Fecha crecida	Caudal máximo instantáneo	Período de retorno	Utilización
2016-04-16	281 (m ³ /s)	22 años	Testeo
2021-01-30	216 (m ³ /s)	12 años	Validación

4.4.2 Optimización parámetros del modelo

Para finalizar la obtención de resultados se optimizan las variables internas: *Learning rate*, *hidden dimensions* y *num layers* de los modelos mediante un paquete de optimización llamado Optuna (Akiba et al., 2019).

El proceso de optimización de hiperparámetros en Optuna se realiza a través de la búsqueda automática, que consiste en un conjunto de técnicas para explorar el espacio de hiperparámetros de un modelo y encontrar la mejor combinación de ellos. Optuna utiliza algoritmos de búsqueda como TPE (*Tree-structured Parzen Estimator*), CMA-ES (*Covariance Matrix Adaptation Evolution Strategy*), entre otros.

El proceso de optimización en Optuna se divide en tres etapas:

- **Definición del espacio de búsqueda:** En esta etapa, se define el espacio de búsqueda para los hiperparámetros. Optuna proporciona diferentes tipos de variables como "discretas", "continuas" o "categóricas" para definir el espacio de búsqueda.
- **Ejecución de los ensayos:** En esta etapa, se ejecutan los ensayos de los modelos de *Machine Learning* con diferentes combinaciones de hiperparámetros en función del espacio de búsqueda definido en la primera etapa. Cada ensayo se evalúa en función de una métrica de evaluación, siendo el coeficiente R^2 para los *DataFrames* de testeo en cada cuenca definida.
- **Análisis de resultados:** En esta etapa, se analizan los resultados de los ensayos para identificar la mejor combinación de hiperparámetros para el modelo. Optuna proporciona diferentes métodos de análisis de resultados, como la visualización de gráficos y estadísticas descriptivas.

Se plantea un muestreo de los valores a testear en los hiperparámetros y un número de intentos para obtener el mejor modelo. Optuna requiere de una métrica para evaluar el desempeño de las predicciones, siendo el coeficiente R^2 para los *DataFrames* de testeo en cada cuenca.

La optimización ocurre en dos oportunidades, para los sets de validación en cada iteración del modelo donde se encuentra la época óptima y para cada modelo óptimo por validación se busca el que mejor coeficiente R^2 obtuvo en el testeo, siendo este el mejor modelo elegido.

4.5 Consideraciones finales

4.5.1 Indicadores de ajuste

En conjunto con los indicadores RMSE y R^2 presentados en la 2 y 2. respectivamente se utilizan las métricas de porcentaje de sesgo (Pbias), Criterio de información de Akaike (AIC) y Error Absoluto Medio (MAE) para el análisis y comprensión de los resultados.

$$AIC = 2k - 2 \ln(\hat{L}) \quad 12.$$

donde k es el número de parámetros en el modelo y \hat{L} es la máxima verosimilitud del modelo ajustado. El término $2k$ penaliza modelos con un número excesivo de parámetros y el término $2 \ln(\hat{L})$ mide la adecuación del modelo a los datos.

$$Pbias = 100\% \times \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{\sum_{i=1}^n y_i} \quad 13.$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad 14.$$

donde n es el tamaño de la muestra, y_i es el valor verdadero de la muestra i , \hat{y}_i es la predicción del modelo para la muestra i , $\sum_{i=1}^n (\hat{y}_i - y_i)$ es la suma de las diferencias entre las predicciones del modelo y los valores verdaderos y $\sum_{i=1}^n |y_i - \hat{y}_i|$ es la suma de los valores absolutos de las diferencias entre las predicciones del modelo y los valores observados.

4.5.2 Umbrales de alerta

La definición de umbrales de alerta tiene dos aristas a considerar, en primer lugar, los protocolos internos de PHC en dónde se dictaminan rangos de caudales que puedan afectar la operación normal de las Centrales (PHC, 2012) y, en segundo lugar, los rangos establecidos como peligrosos por la DGA.

Al no tener más información de testeo para los modelos no se puede asegurar que para crecidas mayores a las de entrenamiento se generen predicciones acertadas, es por esto por lo que los umbrales deben considerar este punto para poder generar alertas en caso de presencia de riesgo por crecidas debido a eventos de precipitación.

4.5.3 Recomendaciones Sistema de Alertas Tempranas (SAT)

Las recomendaciones y pasos por seguir para la implementación del SAT en el valle del Alto Cachapoal son sustentadas por los diversos documentos internacionales destinados al manejo de amenazas y requerimientos de implementación de acciones y actores para que el sistema funcione y pueda obtener los resultados esperados de protección de la vida de las personas mediante medidas preventivas y difusión asertiva y eficaz de la información de alertas generadas por el modelo. Algunos de estas recomendaciones vienen de:

- OMM (Organización Meteorológica Mundial) - Directrices técnicas y de gestión para servicios meteorológicos e hidrológicos eficaces: Guía para el establecimiento y funcionamiento de servicios hidrológicos nacionales. (2006)
- Conferencia Internacional Sobre Alerta Temprana - Marco de acción de Hyogo para 2005-2015: Aumento de la resiliencia de las naciones y las comunidades ante los desastres. (2006)
- RALCEA (Red de Alerta de Centros de Pronóstico Hidrológico de América Latina y el Caribe) - Recomendaciones para la implementación de sistemas de alerta hidrológica temprana. (2015)
- Saalmueller, N. (2006). EWS (Sistemas de Alerta Temprana) en la planificación y gestión del riesgo de desastres: una introducción. Programa de las Naciones Unidas para el Desarrollo (PNUD).

Capítulo 5: Resultados

El capítulo presenta los resultados obtenidos para cada cuenca y tiempos de adelanto uno, tres y cinco horas. Se exponen los valores de coeficiente R^2 , RMSE, MAE, AIC y Pbias de las predicciones en formato *boxplot* para los sets de entrenamiento y validación-testeo, junto con las figuras de comparación de curva real y pronóstico simulado y gráfico de dispersión en crecidas de interés. La organización de los resultados está separada para las cuencas del Río Cachapoal y el Río Pangal.

5.1 Pronóstico de caudales en la cuenca del río Cachapoal

Se utilizaron 12 crecidas para realizar todo el proceso de entrenamiento, validación y testeo. A continuación, se exponen las figuras respectivas de cada *DataFrame* luego de ser pasado por la red ya entrenada, con énfasis en las predicciones de crecidas mayores.

5.1.1 Pronóstico hidrológico con tiempo de adelanto de 1 hora

En primer lugar, se presentan las configuraciones óptimas del modelo para Cachapoal con adelanto de una hora obtenidas en base a la optimización de hiperparámetros con Optuna en la Tabla 5.1.

Tabla 5.1: Hiperparámetros de modelo óptimo Cachapoal 1 hora de adelanto

Hiperparámetro	Valor
Tasa de aprendizaje	0.008
Número de capas ocultas	35
Épocas de entrenamiento	2950
Número de capas LSTM	1

Los gráficos de ajuste de los subconjuntos de validación/testeo y entrenamiento del modelo se presentan en Figura 5.1 ¡Error! No se encuentra el origen de la referencia. y Figura 5.2 respectivamente. Se presentan los valores de R^2 de cada evento de crecida simulado. Se presentan en Anexos los eventos faltantes.

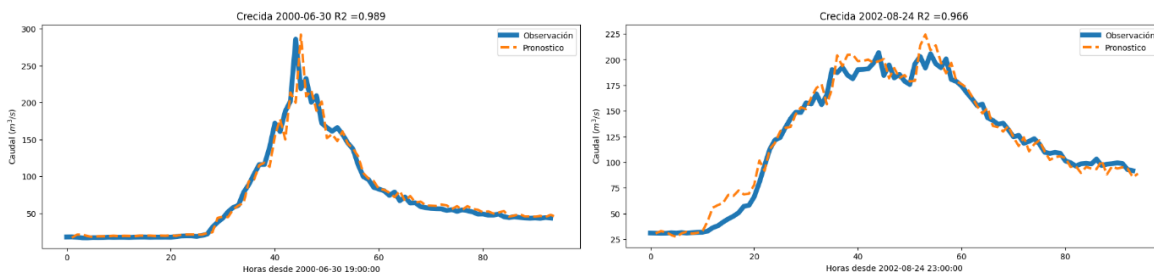


Figura 5.1: Crecida de testeo (izquierda) y de validación (derecha)

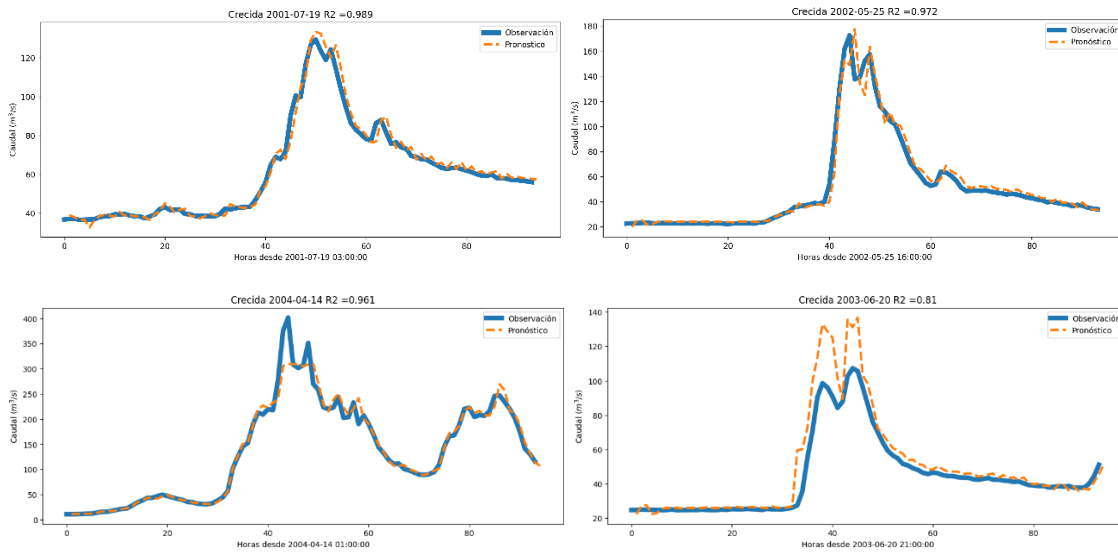


Figura 5.2: Crecidas de entrenamiento

En la Figura 5.3 se presentan los gráficos de dispersión para los eventos de entrenamiento de 2003 y 2004 entre las variables reales y predichas. Se busca apreciar el límite superior generado por el modelo en la crecida mayor, en donde los valores al pasar 300 (m³/s) se escapan de la línea de tendencia. también se pretende observar la sobreestimación que ocurre en la crecida de 2003, en la que se supera el caudal observado por casi 40 (m³/s) en el peak. Para enriquecer el análisis se presentan los indicadores MAE, RMSE, Pbias, R² y AIC como parte del título de las figuras. En La Figura 5.4 se muestran de forma gráfica los indicadores obtenidos para el conjunto de datos de entrenamiento y de validación-testeo.

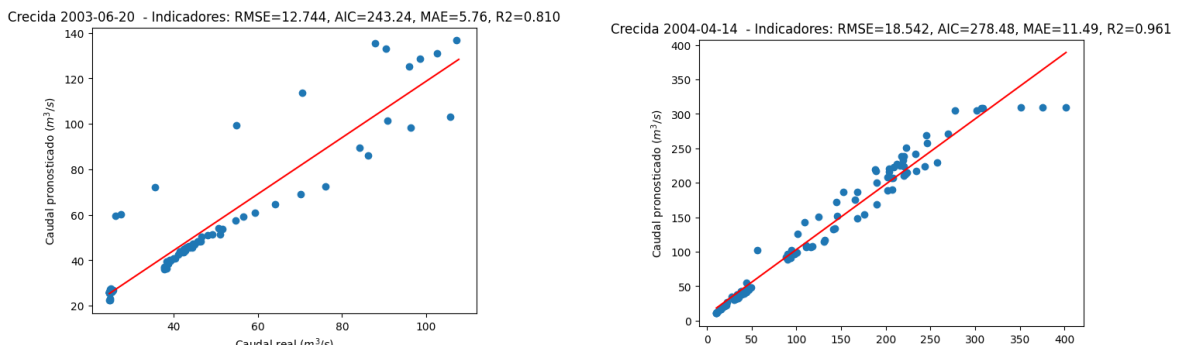


Figura 5.3: Gráficos de dispersión crecidas de entrenamiento.

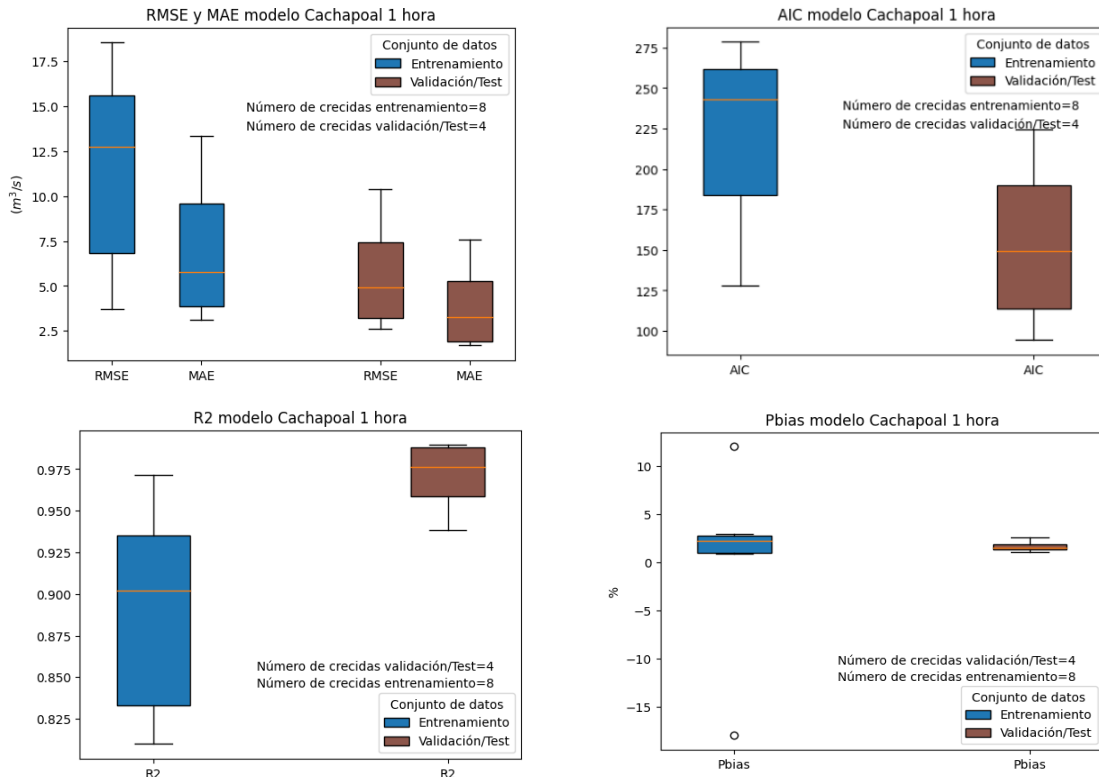


Figura 5.4: Boxplot indicadores RMSE, MAE, AIC, R^2 y Pbias de pronóstico Cachapoal 1 hora.

Respecto a los indicadores RMSE y MAE se observan en un rango acotado que no supera los 10 (m^3/seg) de error para la validación y testeo y los 20 (m^3/seg) para entrenamiento. Para el coeficiente de correlación R^2 se presentan valores mayores a 0,8 en todas las crecidas utilizadas, indicando que bajo esta métrica se poseen resultados de calidad.

Finalmente, respecto a los indicadores AIC y Pbias se observa un rango de mayor amplitud en ambos para los sets de entrenamiento por sobre validación/testeo, lo que a priori podría ser contraintuitivo debido a que el modelo está entrenado en base a ciertos datos y debería ser los que mejor pueda reproducir. Como explicación de este fenómeno se podría pensar en la brecha existente entre las magnitudes de las crecidas de entrenamiento, donde solo un set de datos supera los 400 (m^3/seg) y los demás rondan entre 100 y 200 (m^3/seg) por lo que un ajuste satisfactorio es difícil de lograr, aunque haya sido entrenado con esas crecidas. Por otro lado, el que suceda así indica que no hay un sobreajuste ni un ajuste insuficiente en los datos de entrenamiento pues el modelo generaliza y obtiene resultados de alta precisión en las crecidas de testeo.

5.1.2 Pronóstico hidrológico con tiempo de adelanto de 3 horas

Al aumentar el tiempo de adelanto del modelo se le está induciendo una tarea de alta complejidad dado que no se le entrega un *input* de precipitación para las horas de adelanto y la metodología para lograr estas predicciones hace que los resultados bajen su calidad.

Se presentan las configuraciones óptimas del modelo para Cachapoal con adelanto de tres horas obtenidas en base a la optimización de hiperparámetros con Optuna en la Tabla 5.2.

Tabla 5.2: Hiperparámetros de modelo óptimo Cachapoal 3 horas de adelanto

Hiperparámetro	Valor
Tasa de aprendizaje	0.008
Número de capas ocultas	15
Épocas de entrenamiento	2950
Número de capas LSTM	1

Al analizar la Figura 5.5 se observa la disminución de la precisión en los pronósticos de validación y testeo junto con la presencia de un límite superior para el testeo. Es claro que el aumento de tiempo de adelanto tuvo secuelas en las optimizaciones internas de la red, dando pie a peores ajustes en todos los data sets utilizados.

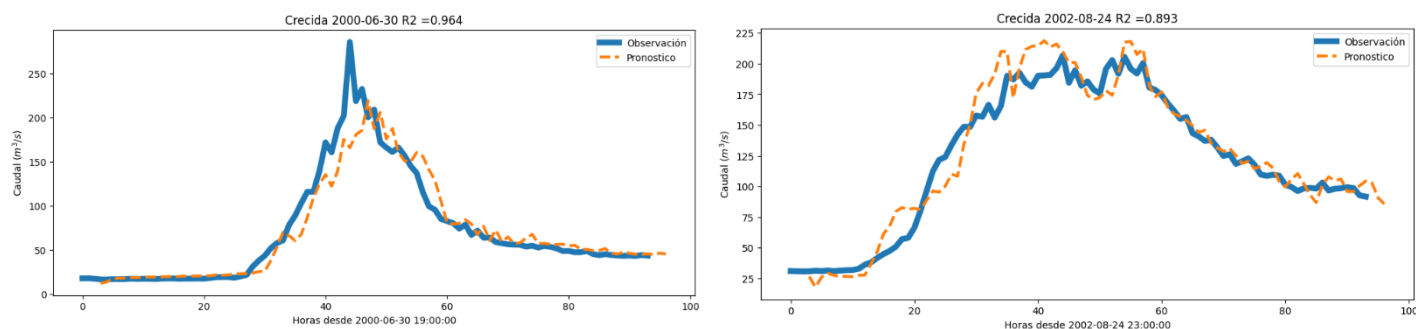


Figura 5.5: Crecida de testeo (izquierda) y de validación (derecha)

En el pronóstico realizado a la crecida mayor de entrenamiento se aprecia claramente que el límite superior previamente definido en $300 \text{ (m}^3\text{/s)}$ desciende a $200 \text{ (m}^3\text{/s)}$ en la Figura 5.6. Debido a la poca información disponible de crecidas horarias y la dificultad mayor que implica para el modelo adelantar la predicción de la manera en la que se realiza la baja la calidad de los resultados frente al modelo de una hora era de esperar. Por parte del modelo Cachapoal 3 horas de pronóstico, en validación-testeo se encuentran los valores de 0,93, 8,2 y 6 presentándose resultados de buena calidad en términos de R2, RMSE y MAE.

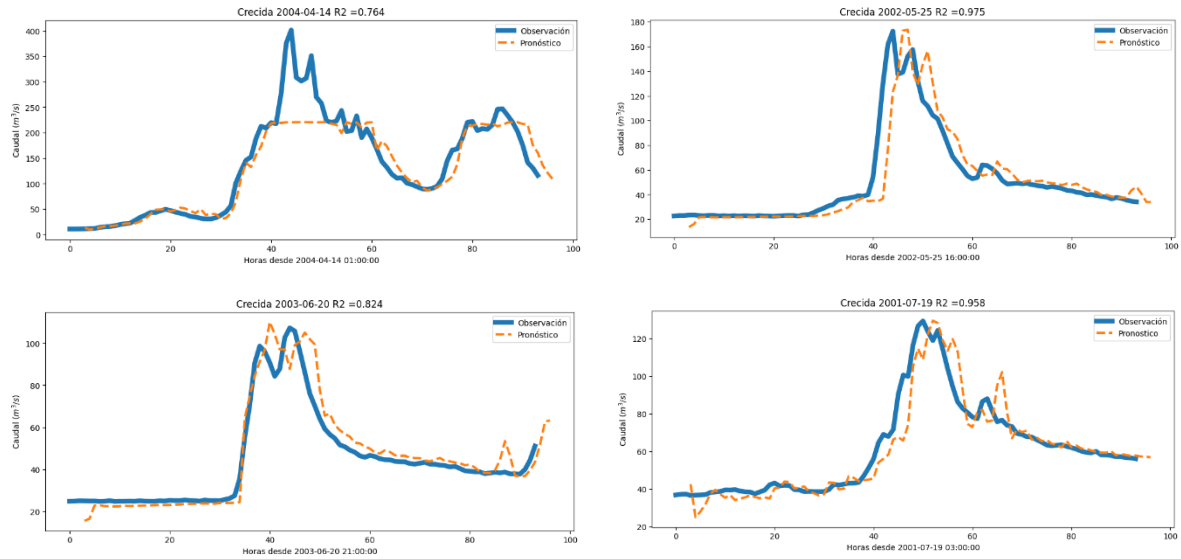


Figura 5.6: Crecidas de entrenamiento.

Se aprecia en la Figura 5.7 que se mantiene la línea de lo ocurrido con el modelo a 1 hora y el entrenamiento obtiene peores resultados que el testeo al realizar el adelanto de tiempo. Al revisar cada crecida utilizada para el entrenamiento se puede observar que la naturaleza de ellas es distinta y poseen patrones de caudal que se podrían considerar difíciles de aprender debido a la variación que hay entre los tipos de crecida por época del año (variable que no se toma en cuenta en la modelación) y pequeñas fluctuaciones de caudal y precipitación.

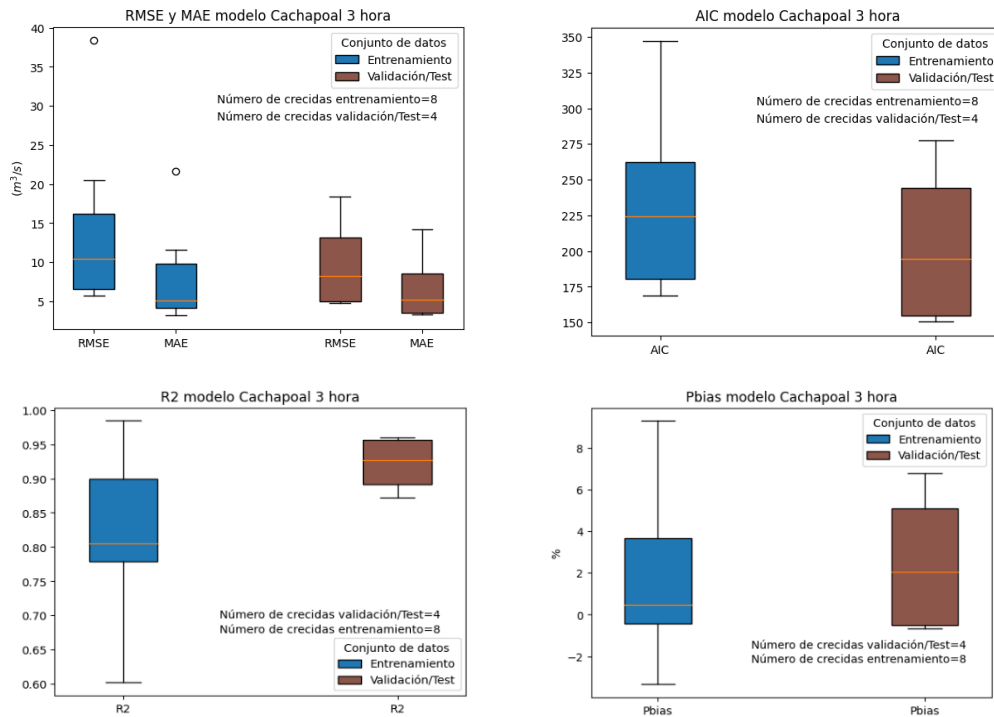


Figura 5.7: Boxplot indicadores RMSE, MAE, AIC, R² y Pbias de pronóstico Cachapoal 3 hora.

A pesar de que los ajustes ya no son tan buenos como para una hora, los indicadores AIC entre ambos modelos y sets de datos no han variado en gran magnitud, indicando que la calidad y complejidad de los modelos utilizados no ha mutado en gran cantidad, resultado que era de esperarse debido a que la arquitectura base LSTM es la misma para todos los modelos que se optimizan a través de la combinación de hiperparámetros, tasa de aprendizaje, número de capas LSTM y número de capas ocultas. La dispersión de los datos en las dos crecidas mayores muestra un patrón de límite superior y se presenta en la Figura 5.8.

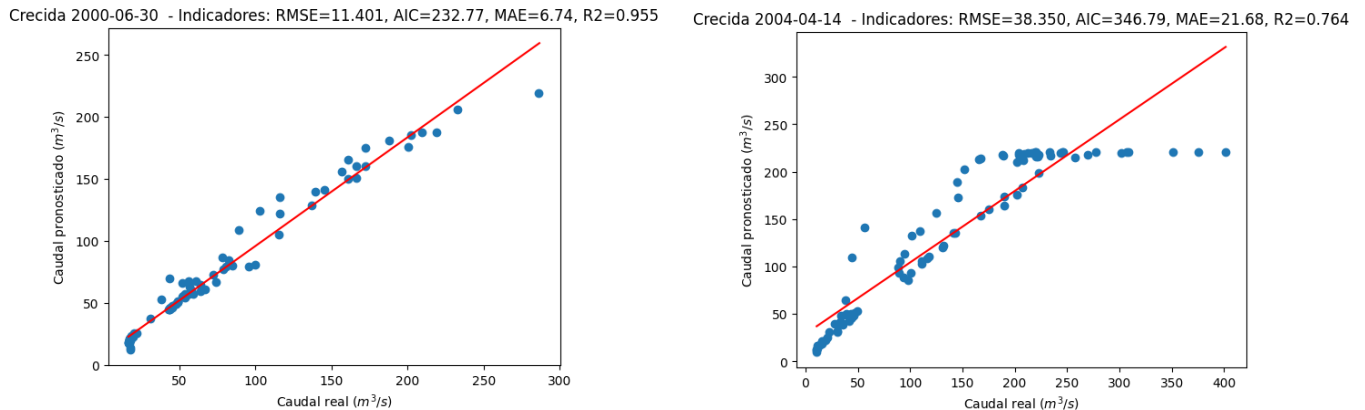


Figura 5.8: Dispersión de datos en crecidas de mayor magnitud.

5.1.3 Pronóstico hidrológico con tiempo de adelanto de 5 horas

Se presentan las configuraciones óptimas del modelo para Cachapoal con adelanto de una hora obtenidas en base a la optimización de hiperparámetros con Optuna en la Tabla 5.3.

Tabla 5.3: Hiperparámetros de modelo óptimo Cachapoal 5 horas de adelanto

Hiperparámetro	Valor
Tasa de aprendizaje	0.006
Número de capas ocultas	10
Épocas de entrenamiento	2950
Número de capas LSTM	2

El análisis para este caso no difiere mucho de los dos anteriores en dónde se vuelve a percibir un piso superior de 200 (m³/s) para los caudales máximos pronosticados. Los resultados de validación bajan su calidad considerablemente no así el testeo y entrenamiento que no varían mucho en comparación con los adelantos de tres horas.

Finalmente, se aprecia al comparar la que los indicadores de entrenamiento en general tienen rangos y promedios de mejor calidad que el set de validación-testeo. Esto claramente debido a los bajos niveles de ajuste que se obtuvieron en el set de validación lo que genera

entonces estas diferencias contrarias a los otros dos casos, pero ligado con lo esperable que sucediera en un modelo de este tipo. Respecto del indicador AIC sistemáticamente el mejor resultado lo obtiene la crecida de testeo ocurrida en el año 2000. Indicando que el modelo sí es capaz de extrapolar lo aprendido en entrenamiento hacia nuevos datos. Se presentan las figuras Figura 5.9: Crecidas de entrenamiento (izquierda), validación (centro) y testeo (derecha) Figura 5.9 Figura 5.10 con las crecidas y boxplots, los gráficos de dispersión se presentan en anexos.

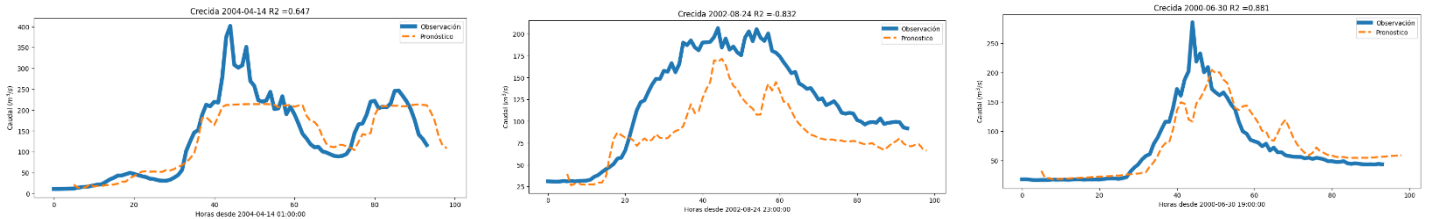


Figura 5.9: Crecidas de entrenamiento (izquierda), validación (centro) y testeo (derecha)

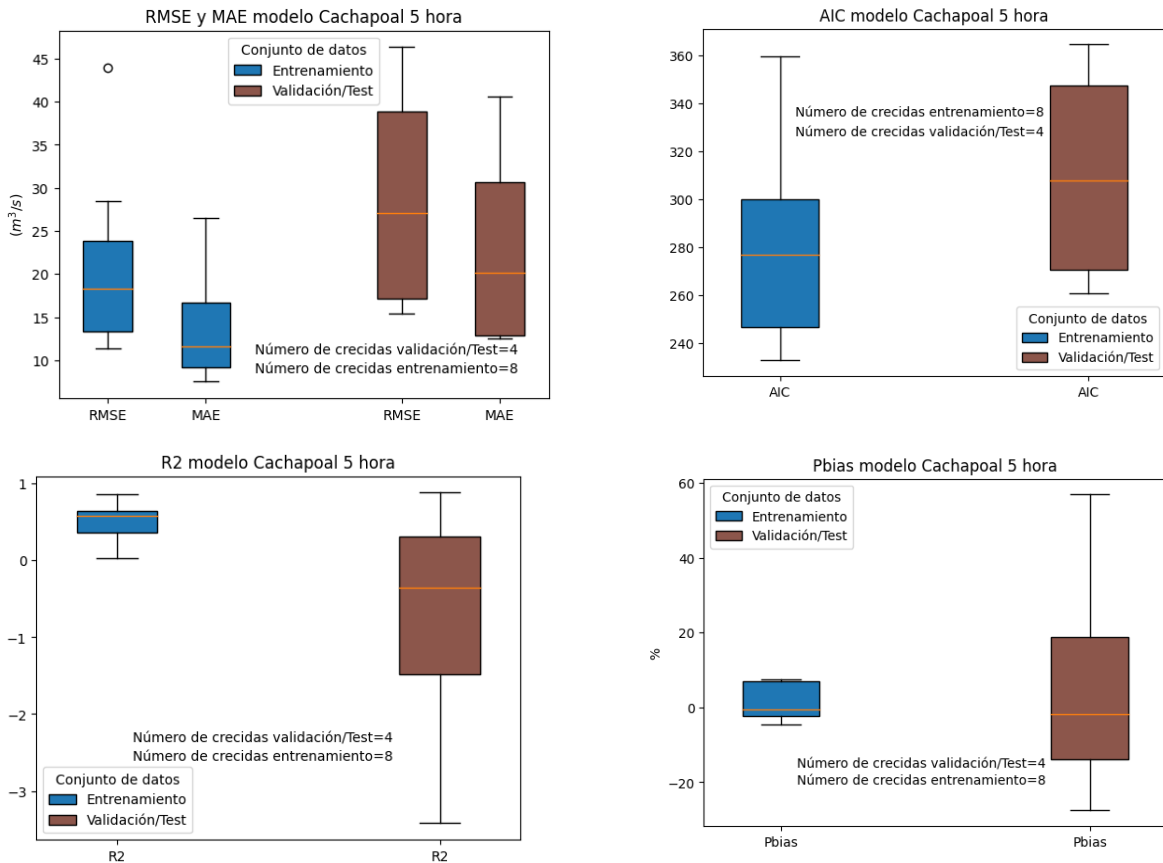


Figura 5.10: Boxplot indicadores RMSE, MAE, AIC, R^2 y Pbias de pronóstico Cachapoal 5 hora.

5.2 Pronóstico de caudales en la cuenca del río Pangal

Se utilizaron nueve crecidas para realizar todo el proceso de entrenamiento, validación y testeo. A continuación, se presentan las figuras respectivas de cada *DataFrame* luego de ser pasado por la red ya entrenada, con énfasis en las predicciones de crecidas mayores.

5.2.1 Pronóstico hidrológico con tiempo de adelanto de 1 hora

Se presentan las configuraciones óptimas del modelo para Pangal con adelanto de una hora obtenidas en base a la optimización de hiperparámetros con Optuna en la Tabla 5.3.

Tabla 5.4: Hiperparámetros de modelo óptimo Pangal 1 hora de adelanto

Hiperparámetro	Valor
Tasa de aprendizaje	0.008
Número de capas ocultas	50
Épocas de entrenamiento	5000
Número de capas LSTM	1

Para esta cuenca los ajustes del tiempo de adelanto una hora obtienen resultados de correcto ajuste, teniendo coeficientes de R^2 superiores a 0,82 para validación y testeo y de 0,95 para la mayor crecida de entrenamiento. Las crecidas utilizadas en esta modelación poseen diferentes tipos de hidrogramas, en específico la fluctuación de caudal de la crecida de testeo hace que su predicción no sea simple para el modelo, en especial porque no hay otra que lleve esa forma. A pesar de esto la predicción logra ajustarse a la forma de la crecida identificando los máximos locales y prediciendo sus magnitudes en la Figura 5.11 se presentan las crecidas de validación y testeo para Pangal. Crecidas de entrenamiento y scatterplots se presentan en anexos.

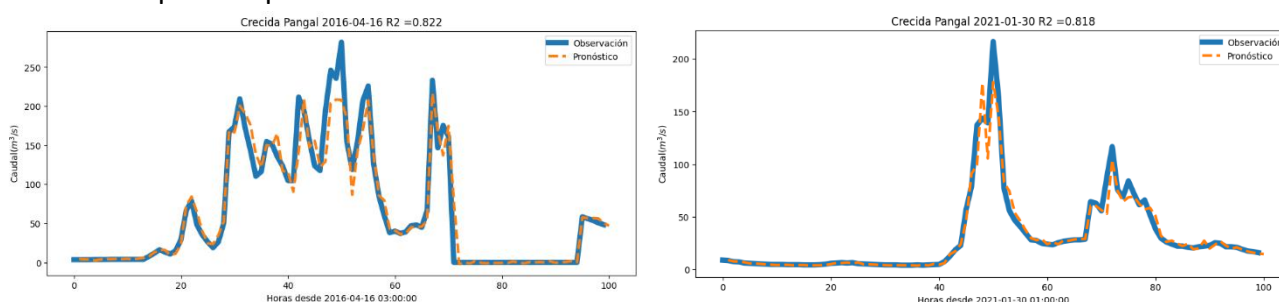


Figura 5.11: Crecida de testeo (izquierda) y validación (derecha)

Al comparar los valores de cada indicador de la Figura 5.12 se obtienen diferencias esperables en todos los indicadores en que el entrenamiento entrega métricas de mejor ajuste que los resultados en los campos estudiados, siguiendo el camino lógico en el que el modelo se ajusta bien al conjunto de datos de entrenamiento.

Finalmente comparando los rangos de valores de AIC entre Cachapoal una hora y Pangal una hora se obtienen diferencias de casi 100 unidades, indicando que el modelo implementado en Cachapoal sería un mejor modelo en base al AIC.

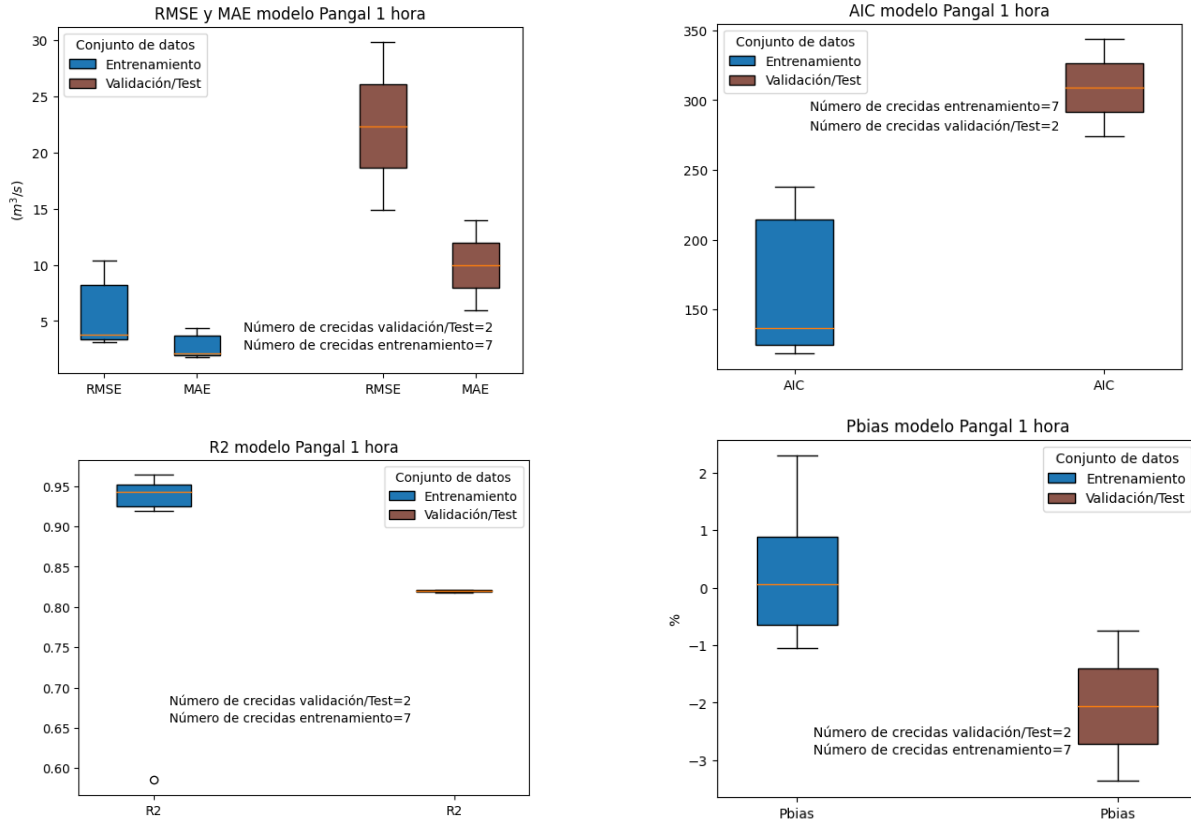


Figura 5.12: Boxplot indicadores RMSE, MAE, AIC, R² y Pbias de pronóstico Pangal 1 hora.

5.2.2 Pronóstico hidrológico con tiempo de adelanto de 3 horas

Se presentan las configuraciones óptimas del modelo para Pangal con adelanto de una hora obtenidas en base a la optimización de hiperparámetros con Optuna en la Tabla 5.5.

Tabla 5.5: Hiperparámetros de modelo óptimo Pangal 3 horas de adelanto

Hiperparámetro	Valor
Tasa de aprendizaje	0.006
Número de capas ocultas	20
Épocas de entrenamiento	4000
Número de capas LSTM	3

Al observar las predicciones realizadas y la dispersión de los datos se puede ver claramente la baja en la calidad de resultados y la baja del umbral superior de modelación a 150 (m³/s). Por lo que es claro que al realizar la metodología de aumento de tiempo de adelanto se induce a peores resultados, sin embargo, los valores obtenidos igual pueden ser de utilidad como guía para la toma de decisiones, pero nunca siendo la herramienta única los modelos adelantados. La Figura 5.13 presenta el desempeño del pronóstico en testeo y validación

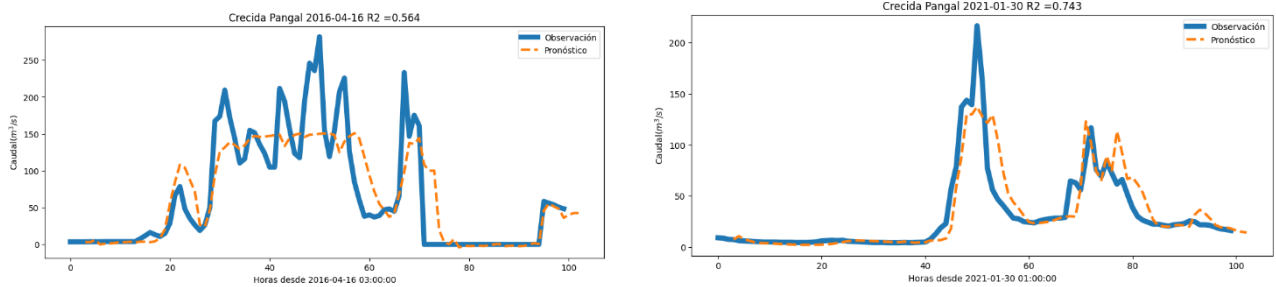


Figura 5.13: Crecida de testeo (izquierda) y validación (derecha)

La comparación de indicadores en base a los boxplots presentados en la Figura 5.14 sigue la misma tendencia que para una hora de adelanto en términos de las diferencias entre los sets de entrenamiento y testeo-validación pero con la baja sistemática de desempeño en cada uno debido a lo mencionado anteriormente.

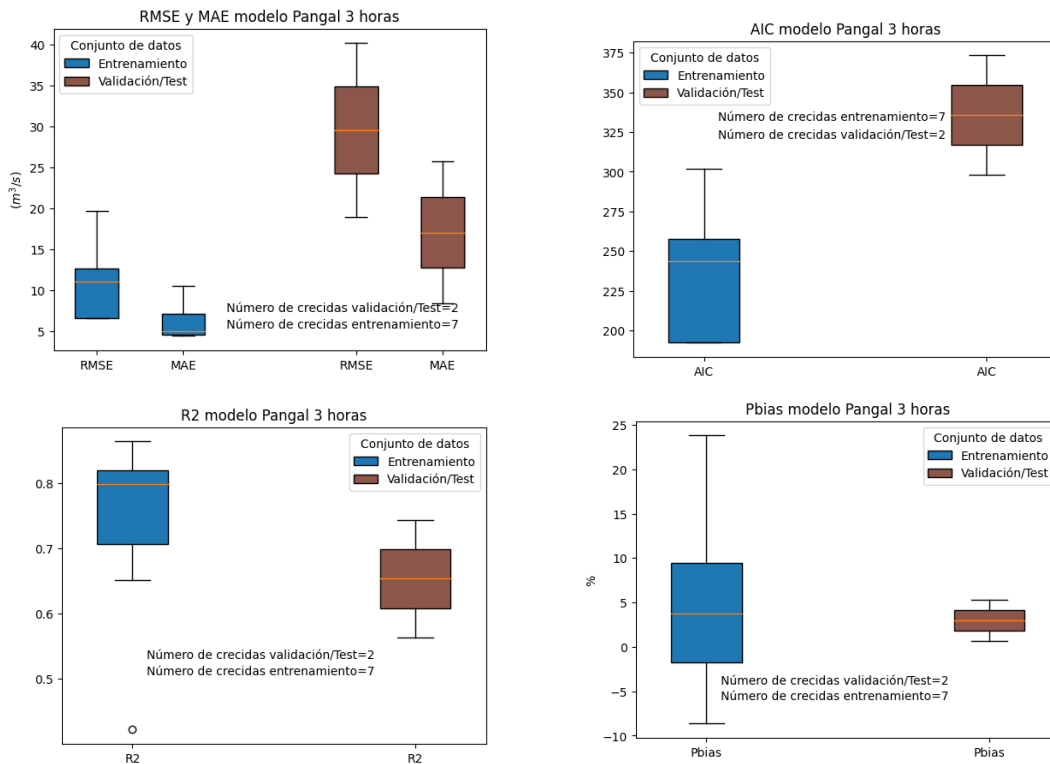


Figura 5.14: Boxplot indicadores RMSE, MAE, AIC, R² y Pbias de pronóstico Pangal 3 horas.

5.2.3 Pronóstico hidrológico con tiempo de adelanto de 5 horas

Para el modelo Pangal a 5 horas los resultados no son concluyentes entregando indicadores R^2 menores que 0,3 en todas las crecidas y ajustándose de mala manera a las formas que toma el caudal en el evento. Es por esto por lo que se presenta toda la información para este modelo en anexos dado que no aporta a los objetivos finales del trabajo ni tiene una utilización práctica para la generación del sistema de alertas tempranas y umbrales.

5.3 Pruebas de pronóstico cruzado de cuencas

La prueba cruzada de modelos entre cuencas se emplea para verificar si un modelo hidrológico puede simular con precisión la respuesta hidrológica en una cuenca no calibrada previamente. Si el modelo puede hacerlo, se considera que tiene la capacidad de generalización. El proceso de prueba cruzada de modelos entre cuencas involucra ajustar el modelo a una cuenca, donde se determinan los valores de los parámetros, y luego aplicar el modelo ajustado en otra cuenca, evaluando su capacidad de generalización. Para determinar la precisión del modelo, se comparan los resultados de la simulación con los datos observados. En este caso se prueba el modelo pronóstico de una hora para Cachapoyal con los datos de crecidas de Pangal y viceversa. Los resultados se presentan gráficamente en la Figura 5.15.

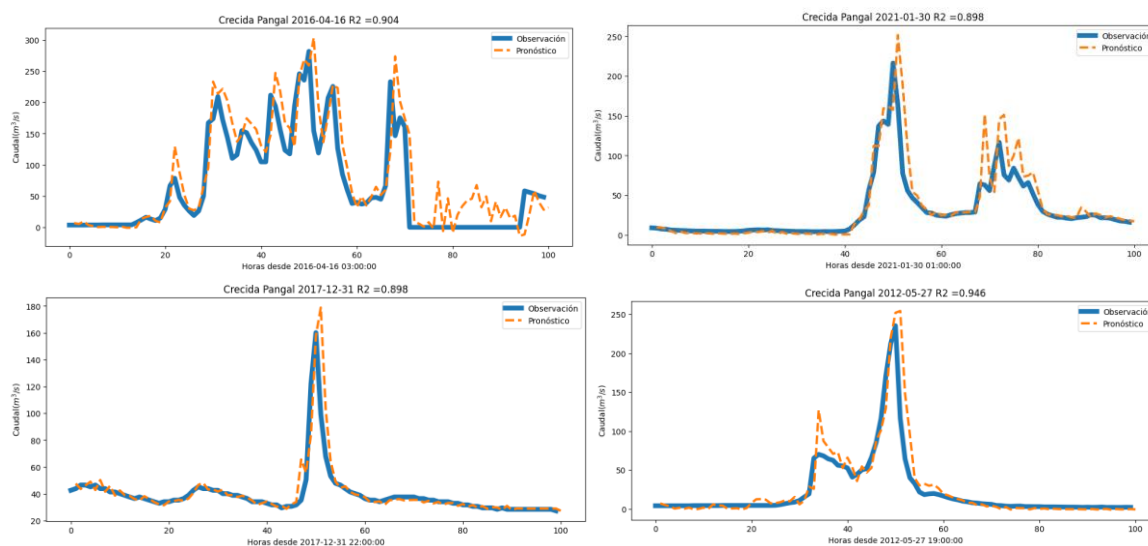


Figura 5.15: Mayores crecidas de Pangal testeadas en modelo Cachapoyal 1 hora

A pesar de obtener muy buenos ajustes en las mayores crecidas se aprecia una tendencia a la sobreestimación del modelo en las crecidas de menor magnitud de la cuenca Pangal.

Al observar los ajuste e indicadores R^2 es claro como el modelo Cachapoal $dt=1$ se ajusta bien a lo que ocurre en Pangal, pero no para el caso contrario. Se puede explicar en base a las respuestas hidrológicas que tiene cada cuenca, al ser Pangal más pequeña los caudales base, tiempos de respuesta y áreas aportantes son menores y tratar de predecir lo que ocurre con el Cachapoal en base a solo lo aprendido en Pangal se puede evidenciar que no funciona del todo bien para eventos extremos. En cambio, al pasar de la cuenca más grande hacia la cuenca vecina con menor régimen natural de caudales y similares características topográficas y climáticas se evidencia un ajuste de gran calidad, incluso con mejor indicador R^2 que para el modelo entrenado para Pangal en el set de testeo en las mayores crecidas. Para crecidas medias y bajas el modelo de Cachapoal tiende a sobreestimar.

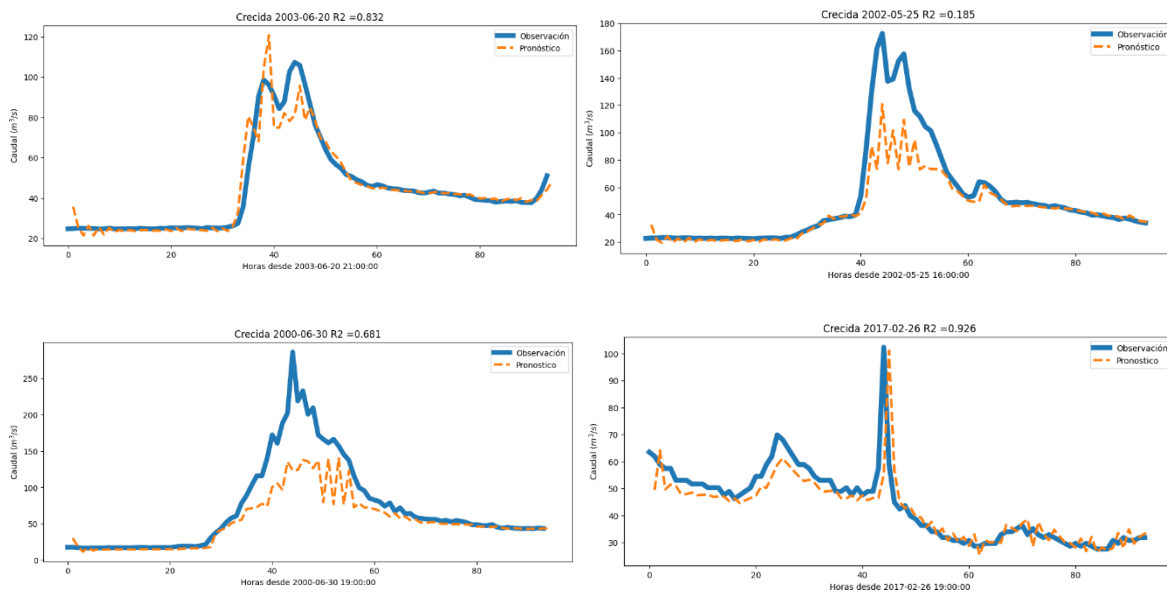


Figura 5.16: Crecidas de Cachapoal testeadas en modelo Pangal 1 hora.

5.4 Umbrales de alertas

La definición de umbrales de alerta se realiza en base a la información previamente disponible de los informes técnicos de la DGA (DGA, 2010) y los manuales de emergencia y prevención de PHC (PHC, 2012). En ambos documentos se presentan alertas para las cuencas definidas por la estación Cachapoal 5 km aguas abajo junta Cortaderal. Los umbrales de alerta pretenden servir de apoyo a la toma de decisiones para resguardar la infraestructura existente, así como salvaguardar a la población frente a una eventual crecida. De esta manera los criterios de umbrales podrían ser más estrictos entre una entidad pública o privada dependiendo del usuario que está siendo afectado, por ejemplo, una bocatoma puede tener un umbral de magnitud de caudal por alerta amarilla por aumento de caudales producto de aumento de la turbiedad y carga de sedimentos, pero podría no ser el mismo nivel de alerta para producir desbordes.

A modo de referencia, La Tabla 5.6 expone umbrales de alerta que, según (DGA, 2010), comprenden un peligro para las zonas afectadas por desbordes del río Cachapoal. En el documento no se habla del río Pangal por lo que se aplica que la alerta roja será para los caudales de igual período de retorno que los de Cachapoal llegando a los valores mostrados.

Tabla 5.6: Umbrales de alerta DGA

Tipo de alerta	Pangal en Pangal	Cachapoal 5km
Amarilla	140 (m ³ /s) T=5 años	240 (m ³ /s) T=5 años
Roja	190 (m ³ /s) T=9 años	300 (m ³ /s) T=9 años

Capítulo 6: Discusión

6.1.1 Resultados modelación

Sin lugar a duda las predicciones con mejores indicadores y representación de las crecidas son las definidas para una hora de adelanto. En Cachapoal se obtuvieron resultados prometedores y acertados para las cuencas de testeo y validación. Para la crecida mayor (Figura 5.2) de 2004, que estuvo presente en los *DataFrame* de entrenamiento, se aprecia que no reproduce el *peak* de caudal, pero si predice la forma de la crecida, quedando casi 80 (m³/s) bajo el valor real. Para los sets de validación y testeo (Figura 5.1) las crecidas son reproducidas de manera precisa obteniendo coeficientes R² muy cercanos a 1. Para crecidas de menor magnitud se puede observar una tendencia a la sobreestimación de la crecida, no obstante, el pronóstico nunca supera por más de 40 (m³/s) a la observación. Respecto a los indicadores calculados se mantienen en rangos bajos, por ejemplo el Pbias no supera los 10% de sesgo en los datos para los modelos de Cachapoal a una y tres horas de adelanto.

Específicamente en los datos de entrenamiento se puede observar un límite superior de caudal en los 300 (m³/s) para el *peak* de la crecida mayor. Respecto a este fenómeno la respuesta no es del todo clara y puede llevar relación con que la cantidad de datos usados en el entrenamiento es baja y que la diferencia entre la crecida 2004 y las demás utilizadas para el entrenamiento es de magnitud importante. Lo que llevaría al modelo aprender de mejor manera las dinámicas de los datos de caudales menores y por ende no responder de manera acertada para ocurrencias de caudales mayores al piso superior generado.

La tónica general que se aprecia en los resultados es la presencia de un límite superior entre 150 y 300 (m³/s) que restringe las predicciones del modelo y que va bajando su magnitud conforme sube el tiempo de adelanto, siendo un problema para las predicciones de más de una hora. A pesar de esto los umbrales definidos DGA son superados, pero solo en los modelos de una y tres horas de adelanto.

Es importante destacar que la modelación realizada para Cachapoal fue extrapolable hacia Pangal con resultados de buen ajuste en una y tres horas de adelanto, hasta mejores que los definidos por las crecidas de la propia cuenca en términos de R². Esto da indicios de que posiblemente se podría aplicar el modelo Cachapoal una hora en los datos de la cuenca Cipreses.

Todos los resultados e indicadores obtenidos son comparados con el estudio “*Deep learning with a long short-term memory networks approach for rainfall-runoff simulation*” (Hu et al., 2018) que aplica una red LSTM para predicción de crecidas en China. La comparativa es adecuada debido a que se utiliza el mismo enfoque de modelación y los resultados en general se ajustan de igual o mejor manera para la cuenca Cachapoal y de peor manera para la cuenca Pangal. Es interesante destacar que los datos utilizados por el estudio mencionado

toman en cuenta 14 estaciones distribuidas por la cuenca y una serie total de más de 30 años, identificando cerca de 100 eventos de crecida. Es por esto por lo que el hecho de que la modelación realizada en la tesis, con las limitaciones existentes de datos, tuviese los resultados mostrados indica que la predicción de caudales de crecida con LSTM tiene un gran potencial y la aplicación de estos métodos contando con una mayor cantidad de información y/o la combinación con datos de modelaciones de otras variables es una opción prometedora para trabajos de investigación en modelación hidrológica de pronóstico de caudales.

El modelo propuesto por Chang y Chen en 2001 es un punto base para comparar los resultados obtenidos por este trabajo. Los pronósticos realizados con el enfoque de redes neuronales y lógica difusa entrega resultados de MAE de entre 1 y 8 (m^3/s) para pronósticos de una hora de antelación lo cual es del orden de los resultados obtenidos por los modelos de pronóstico de una hora propuestos en este trabajo.

Al hacer una comparación con el trabajo realizado por Milena Muños en 2008 se nota que las metodologías empleadas para encontrar el modelo óptimo son similares, pero el proceso totalmente distinto debido al avance tecnológico. En la memoria de título mencionada se busca encontrar la mejor configuración de la red en base a la prueba de cantidad de neuronas, con los indicadores Error Cuadrático Medio de validación y R^2 de Testeo. Las configuraciones son probadas una a una y se elige la mejor configuración en base al valor de R^2 . En el caso de la presente memoria se realiza el mismo procedimiento con la ayuda de la librería Optuna, siendo un proceso automatizado que sigue el mismo principio. Al comparar los resultados de R^2 de pronóstico se encuentran valores mayores que 0,8 para los pronósticos entre una y tres horas en ambos estudios.

Es de sumo interés la comparación que se tiene con la memoria de Claudio Ávila en 2018, en esta se asevera que los pronósticos de la red sin tener input de precipitación pronosticada por otro modelo son aceptables hasta con 3 horas de anticipación, y los valores de los indicadores R^2 para redes neuronales en la en la cuenca del río Curanilahue en Curanilahue son de 0,978 para una hora, 0,818 para 3 horas y 0,672 para 5 horas de adelanto. La comparativa de este indicador es sumamente parecida a los resultados obtenidos para los modelos de Cachapoal, indicando que los resultados están dentro del espectro de los estudios de redes neuronales en Chile y el Mundo.

6.1.2 Requerimientos y recomendaciones implementación SAT

Se presentan en esta sección los requerimientos y recomendaciones para la implementación de un Sistema de Alertas Tempranas utilizando como base de predicción a los modelos previamente analizados. Como base teórica para esta sección se utilizan los documentos OMM (2006); GWP (2006); Tercera Conferencia Internacional sobre Alerta Temprana (2006); RALCEA (2015) y Saalmueller (2006)

Como primer paso es importante establecer una conexión en tiempo real con las estaciones de monitoreo de la Dirección General de Aguas (DGA) para obtener datos precisos y actualizados sobre los niveles de los ríos y las precipitaciones en la zona de estudio. Esto se logra mediante la implementación de un sistema de comunicación y transmisión desde estaciones meteorológicas, lo que permitirá obtener información confiable y en tiempo real para la toma de decisiones. Para esto se requiere Gestionar un manejo integrado con la DGA que controla el funcionamiento y flujo de información de cada estación.

Luego es requerido desarrollar un servidor que permita almacenar los datos meteorológicos y utilizarlos en tiempo real con los modelos hidrológicos. Esto permitirá mejorar la eficacia del sistema de alertas tempranas, ya que los modelos hidrológicos necesitan datos meteorológicos precisos y actualizados para generar pronósticos confiables. En esta parte se deberá hacer un análisis EDA de los datos de entrada en tiempo real para comprobar que no tienen errores, datos atípicos o negativos.

Sigue el ejecutar todos los modelos a la vez y comparar resultados para los distintos tiempos de adelanto y desempeño con datos de ambas cuencas. Es importante correr diferentes modelos hidrológicos y comparar sus resultados para determinar cuál de ellos es el más adecuado para las condiciones específicas de la zona de estudio. Además, se debe evaluar el desempeño de los modelos con datos de ambas cuencas para tener una perspectiva más completa y precisa de la situación hidrológica. Este sería un período crítico para el proyecto dado que la marcha blanca del modelo dictaminará su factibilidad.

Una vez que se han definido los umbrales de alerta y se ha seleccionado el modelo hidrológico más adecuado, se debe implementar la automatización de alertas para los resultados generados por el modelo. Esto permitirá que las alertas se emitan de manera oportuna y eficiente, lo que aumentará la capacidad de respuesta y la eficacia del sistema de alertas tempranas. Para la redefinición de los umbrales de alerta, es importante realizar una evaluación cuidadosa de los resultados de los modelos y de la experiencia práctica en la gestión de emergencias por crecidas. Se debería considerar la inclusión de técnicas de retroalimentación en el sistema de alerta temprana, que permitan ajustar los umbrales de alerta de forma dinámica en función de la experiencia acumulada y los resultados de los modelos.

6.1.1 Trabajos futuros

Dentro de los trabajos futuros a realizar se encuentra el mejoramiento de predicciones a más de una hora aplicando los resultados de un modelo climatológico de precipitación como input de los modelos de 1 hora. La idea sería que en el pronóstico se utilicen datos de precipitación simulados para la estación en cuestión de cada modelo y los resultados de predicción de caudal sumado a la precipitación simulada sean el input para el siguiente paso temporal. Con esta metodología podrían mejorar los resultados de más de una hora considerablemente.

La incorporación de la nueva estación fluviométricas en Río Cipreses. Para esto, es necesaria la evaluación previa de su ubicación, características hidrológicas de la cuenca y calidad de los datos que proporciona, con el fin de asegurar que su inclusión mejorará la capacidad de predicción del sistema de alerta temprana. Además, se debería definir un protocolo de mantenimiento y calibración de la estación para asegurar la calidad de los datos en el largo plazo.

Para complementar el sistema y encontrar umbrales basados realmente en su peligrosidad se puede plantear un modelo hidráulico tipo HEC-HMS que mapee las áreas de inundación según rangos de caudales y sea la base para las alertas generadas.

Hoy en día con la gran capacidad de toma de datos hidrológicos estáticos (atributos físicos e indicadores en base a estudio de imágenes satelitales) y dinámicos (series de tiempo), y con la aparición de las plataformas Camels (Addor et al., 2017; Alvarez-Garretón et al., 2018) la modelación hidrológica está dando un paso hacia la posibilidad de extrapolar modelos para más de una cuenca debido al entrenamiento mixto, es decir, por cuenca entregar al modelo toda la información recopilada por Camels y generar un aprendizaje en base a las características de ella con la posibilidad de encontrar interdependencias entre las respuestas hidrológicas de una cuenca según sus características físicas. Esto es realizado por Kratzert et al. (2022) en el entorno de programación NeuralHydrology.

Capítulo 7: Conclusiones

Como resultado de esta memoria de título se logró implementar seis modelos para pronóstico de caudales de crecida en las cuencas Pangal y Cachapoal para tiempos de adelanto de una tres y cinco horas, cumpliendo así con el objetivo principal del trabajo. Los mejores resultados para cada cuenca fueron obtenidos por los pronósticos de una hora de adelanto, demostrando incluso la capacidad del modelo de Cachapoal de pronosticar el caudal de la cuenca Pangal.

Según los resultados obtenidos de las pruebas realizadas entre las cuencas de Cachapoal y Pangal, se ha demostrado que el modelo utilizado puede extrapolarse con gran precisión ($R^2 = 0,9$) para predecir crecidas en el Río Pangal, que tiene un área aportante de 520 km², utilizando como base los datos de la cuenca de Cachapoal, con un área aportante de 965 km². Estos hallazgos destacan una alta correlación entre los predictores de caudal y precipitación en ambas cuencas.

Sin embargo, al intentar utilizar el modelo en sentido inverso, es decir, desde Pangal hacia Cachapoal, los resultados no fueron satisfactorios. Aunque las causas no están completamente claras, se plantea la hipótesis de que esto se debe a la configuración del caudal base y la distribución de precipitación dentro de la cuenca, lo cual dificulta adaptar el modelo a una respuesta hidrológica de mayor magnitud. Es posible que otras variables estén influyendo en la forma en que los modelos responden a diferentes tipos de crecidas y cuencas. La combinación de problemas de escala, procesos hidrológicos y calidad de los datos utilizados para entrenar y validar cada modelo puede ser la causa de un pronóstico deficiente.

Es importante hablar de las deficiencias y debilidades de los modelos frente a su aplicación real. En los modelos de tiempos de adelanto de más de una hora se genera un fenómeno de límite superior de caudal. Se aprecia que los valores máximos alcanzados por las modelaciones de más de una hora alcanzan a estar en rangos de alerta amarilla por lo que no se pierde el objetivo de ganar tiempo para alertar a la población en una potencial situación de riesgo. Por otra parte, se pueden observar casos de sobreestimación de caudales de crecida en eventos de baja o mediana intensidad para estos casos no se alcanza a sobrepasar los umbrales de alerta definidos.

Se presentan los requerimientos de implementación y se brindan recomendaciones para lograr el desarrollo del Sistema de Alertas Temprana (SAT). Además, es importante destacar que el proceso de automatización abre oportunidades para los científicos de datos, permitiendo la posibilidad mejora continua de los modelos existentes. No obstante, esto no implica que los modelos actuales sean inadecuados, ya que han sido desarrollados con base en los conocimientos y datos disponibles hasta la fecha.

BIBLIOGRAFÍA

- Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10), 5293–5313. <https://doi.org/10.5194/hess-21-5293-2017>
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631.
- Alvarez-Garretón, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., & Ayala, A. (2018). The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset. *Hydrology and Earth System Sciences*, 22(11), 5817–5846. <https://doi.org/10.5194/hess-22-5817-2018>
- Ávila Mercado, C. A. (2018). *Propuesta de actualización del sistema de alerta de crecidas en Chile usando redes neuronales e hidrogramas unitarios*.
- Chang, F.-J., & Chen, Y.-C. (2001). A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. *Journal of Hydrology*, 245(1–4), 153–164.
- Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, 5(2), 240–254. <https://doi.org/10.1109/72.279188>
- DE CRECIDAS, P. D. E. G. (2006). *Aspectos sociales y participación de los interesados en la gestión integrada de crecidas*.
- de Crecientes, P. A. de G. (2006). Gestión integrada de crecientes, caso de estudio México: Río Grijalva. *Organización Meteorológica Mundial y Global Water Partnership*.
- Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(7).
- Freeze, R. A., & Harlan, R. L. (1969). Blueprint for a physically-based, digitally-simulated hydrologic response model. *Journal of Hydrology*, 9(3), 237–258. [https://doi.org/10.1016/0022-1694\(69\)90020-1](https://doi.org/10.1016/0022-1694(69)90020-1)

- Hu, C., Wu, Q., Li, H., Jian, S., Li, N., & Lou, Z. (2018). Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water*, 10(11), 1543.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*.
- Kratzert, F., Gauch, M., Nearing, G., & Klotz, D. (2022). NeuralHydrology — A Python library for Deep Learning research in hydrology. *Journal of Open Source Software*, 7(71), 4050. <https://doi.org/10.21105/joss.04050>
- Larson, L. (2002). National Weather Service River Forecast System (NWSRFS). *Mathematical Models of Small Watershed Hydrology and Applications*, 657–703.
- Loaiciga, H. A., Valdes, J. B., Vogel, R., Garvey, J., & Schwarz, H. (1996). Global warming and the hydrologic cycle. *Journal of Hydrology*, 174(1–2), 83–127. [https://doi.org/10.1016/0022-1694\(95\)02753-X](https://doi.org/10.1016/0022-1694(95)02753-X)
- McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 445(1), 51–56.
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, 10(3), 282–290.
- para las Américas, U. O. R. (2006). Tercera Conferencia Internacional sobre Alerta Temprana. *Bonn, Alemania*.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., & Antiga, L. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32.
- Pérez Guerrero, J. (2020). *Redes recurrentes*.
- RALCEA, C. D. E. L. E. J. E. T., & DE, M. D. E. A. Y. D. (2015). *GESTIÓN INTEGRADA DE CRECIDAS*.
- Ruder, S. (2016). *An overview of gradient descent optimization algorithms*.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Saalmueller, J. (2006). *Aspectos jurídicos e institucionales de la gestión integrada de crecidas*.

- Sahoo, B. B., Jha, R., Singh, A., & Kumar, D. (2019). Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophysica*, 67(5), 1471–1481. <https://doi.org/10.1007/s11600-019-00330-1>
- Santibáñez, F. (2016). El cambio climático y los recursos hídricos de Chile. *Santiago, ODEPA*.
- Simmons, C. T., Brunner, P., Therrien, R., & Sudicky, E. A. (2020). Commemorating the 50th anniversary of the Freeze and Harlan (1969) Blueprint for a physically-based, digitally-simulated hydrologic response model. *Journal of Hydrology*, 584, 124309. <https://doi.org/10.1016/j.jhydrol.2019.124309>
- Stevens, E., Antiga, L., & Viehmann, T. (2020). *Deep learning with PyTorch*. Manning Publications.
- Turing, A. M., & Haugeland, J. (1950). Computing machinery and intelligence. *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*, 29–56.
- Vargas, X., Gonzblez, R., Valck, P., & Ntiiiez, C. (2003). A flood warning system using meteorological predictions coupled to a neural network and transfer function hydrological model. *Transactions on Ecology and the Environment*, 61. www.witpress.com,
- Wagena, M. B., Goering, D., Collick, A. S., Bock, E., Fuka, D. R., Buda, A., & Easton, Z. M. (2020). Comparison of short-term streamflow forecasting using stochastic time series, neural networks, process-based, and Bayesian models. *Environmental Modelling and Software*, 126(August 2019), 104669. <https://doi.org/10.1016/j.envsoft.2020.104669>
- Wani, M. A., Bhat, F. A., Afzal, S., & Khan, A. I. (2020). *Advances in deep learning*. Springer.
- William, J. R., & Arnold, J. G. (1993). A system of erosion/sediment yield models. *Reunion Internacional Sobre Procesos de Erosion En Tierras de Altas Pendientes: Evaluacion y Modelaje, Merida (Venezuela), 16-20 May 1993*.
- Yaseen, Z. M., El-shafie, A., Jaafar, O., Afan, H. A., & Sayl, K. N. (2015). Artificial intelligence based models for stream-flow forecasting: 2000-2015. *Journal of Hydrology*, 530, 829–844. <https://doi.org/10.1016/j.jhydrol.2015.10.038>

Sitios Web

DGA MOP. (s/f). Gob.cl. Recuperado el 26 de febrero de 2023, de <https://dga.mop.gob.cl/Paginas/InventarioGlaciares.aspx>

Emol TV - Las imágenes han sido publicadas en las redes sociales y por Bomberos. (s/f). Emol.com. Recuperado el 27 de febrero de 2023, de <https://mtv.emol.com/#/detail/20160417202430817/crecida-del-rio-cachapoal-genera-problemas-en-donihue>

Mónica Marín, A. U. y. L. E. (2020, septiembre 21). *MAPA DE PELIGRO DE REMOCIONES EN MASA DE TIPO FLUJO.* Esri. <https://storymaps.arcgis.com/stories/0cc9e3dd3042455f839e85ed610bbcd5>

Sai. (2020, julio 20). *Math behind Artificial Neural Networks.* Analytics Vidhya. <https://medium.com/analytics-vidhya/math-behind-artificial-neural-networks-42f260fc1b25>

Tubería de madera de Pangal. (s/f). Gob.cl. Recuperado el 21 de febrero de 2023, de <https://www.monumentos.gob.cl/monumentos/monumentos-historicos/tuberia-madera-pangal>

(28), I. G., (64), E., (66), S., & (66), P. (2019, enero 11). *Breve Historia de las Redes Neuronales Artificiales (Artículo 1).* Steemit. <https://steemit.com/spanish/@iars.geo/breve-historias-de-las-redes-neuronales-artificiales-articulo-1>

A Practical Introduction to Deep Learning with Caffe and Python // Adil Moujahid // Data Analytics and more. (s/f). Adilmoujahid.com. Recuperado el 30 de diciembre de 2022, de <http://adilmoujahid.com/posts/2016/06/introduction-deep-learning-python-caffe/>

Abril, R. R. (2021, mayo 7). *Redes Neuronales Recurrentes.* La Máquina Oráculo. <https://lamaquinaoraculo.com/computacion/redes-neuronales-recurrentes/>

MOP - Chile. (s/f). Gob.cl. Recuperado el 27 de diciembre de 2022, de <https://snia.mop.gob.cl/BNAConsultas/reportes>

MOP DGA División de Hidrología. (s/f). Gob.cl. Recuperado el 27 de diciembre de 2022, de https://snia.mop.gob.cl/dgasat/pages/dgasat_param/dgasat_param_1.jsp

¿Qué es una Red Neuronal? (s/f). Codificando Bits. Recuperado el 31 de diciembre de 2022, de <https://www.codificandobits.com/blog/que-es-una-red-neuronal/>

Radhakrishnan, P. (2017, agosto 20). *Introduction to recurrent neural network*. Towards Data Science. <https://towardsdatascience.com/introduction-to-recurrent-neural-network-27202c3945f3>

Sahay, M. (2020, junio 6). *Neural Networks and the universal approximation theorem*. Towards Data Science. <https://towardsdatascience.com/neural-networks-and-the-universal-approximation-theorem-8a389a33d30a>

Sai. (2020, julio 20). *Math behind Artificial Neural Networks*. Analytics Vidhya. <https://medium.com/analytics-vidhya/math-behind-artificial-neural-networks-42f260fc1b25>

Anexos

Anexo A: Funciones de arreglo de base de datos

Este es el último paso del trabajo en donde se seleccionan y acondicionan las crecidas encontradas para ser usadas por la red con el fin de aprender los patrones de los datos. Para estandarizar la información horaria fue necesario generar dos funciones para atacar el problema de más de una medición por hora y la eliminación de datos erróneos.

Para estandarizar los tiempos a horas se crea *TimeFix*, una función que toma los datos dentro de los 30 minutos anteriores y posteriores de una hora redondeada y los promedia para generar series indexadas por los mismos tiempos. Esta función está hecha de dos maneras distintas para utilizar los formatos de DGA-SAT y de los datos de transparencia.

En la Figura 0.1 y Figura 0.2 se presenta la transformación que realiza la función *TimeFix* a los datos, quedando estos indexados por las fechas y con las horas redondeadas.

	Fecha-Hora de Medicion	Caudal (m3/seg)	Ppt Instantanea (mm)
1	09/06/2020 00:27	1.422	0
2	09/06/2020 01:27	1.422	0
3	09/06/2020 02:27	1.422	0
4	09/06/2020 03:27	1.422	0
5	09/06/2020 04:27	1.422	0
...
739	09/07/2020 19:27	2.193	0
740	09/07/2020 20:27	2.053	0
741	09/07/2020 21:27	2.193	0
742	09/07/2020 22:27	2.053	0
743	09/07/2020 23:27	2.053	0

743 rows x 3 columns

Figura 0.2: Serie previa a la aplicación de *TimeFix*

	Caudal (m3/seg)	Ppt Instantanea (mm)
tiempo		
2020-06-09 00:00:00	1.422	0.0
2020-06-09 01:00:00	1.422	0.0
2020-06-09 02:00:00	1.422	0.0
2020-06-09 03:00:00	1.422	0.0
2020-06-09 04:00:00	1.422	0.0
...
2020-07-09 19:00:00	2.193	0.0
2020-07-09 20:00:00	2.053	0.0
2020-07-09 21:00:00	2.193	0.0
2020-07-09 22:00:00	2.053	0.0
2020-07-09 23:00:00	2.053	0.0

743 rows x 2 columns

Figura 0.1: Serie Posterior a la aplicación de *TimeFix*

Anexo B: Consideraciones para correr modelos

Para poder correr los modelos se deberá tener en cuenta las siguientes directrices:

- El modelo fue optimizado por la biblioteca Optuna por lo que los archivos *.pickle* deberán ser leídos teniendo en cuenta esto. Es necesario tener instalada esta biblioteca en el entorno de trabajo en Colab.
- Es preciso utilizar el entorno de ejecución GPU para correr el modelo dado que todo está realizado en *Pytorch cuda GPU* y no funcionará de no ser así.
- Para los archivos *.pt* simplemente se requiere del código `torch.load('Directorio\del\modelo')`
- Para cargar el archivo *.pickle* y reconstruir el estudio de optimización en Colab se puede utilizar la siguiente secuencia de código:

```
import optuna

import pickle

# Carga el archivo pickle desde Google Drive
with open('/content/gdrive/My Drive/mi_estudio.pickle', 'rb') as f:
    study = pickle.load(f)

# Muestra el mejor valor de prueba
print('Mejor valor de prueba: ', study.best_trial.value)

# Muestra los mejores hiperparámetros
print('Mejores hiperparámetros: ')

for key, value in study.best_params.items():
    print(f' {key}: {value}')
```

- Para entrenar y obtener resultados los modelos reciben tensores por lo que los datos deberán ser moldeados a tensores para ser utilizados. Para esto se puede aplicar la siguiente secuencia de código.

```
import pandas as pd

import torch

# Cargar el DataFrame con los datos de entrada
data = pd.read_csv('datos.csv')

# Cargar el modelo entrenado
modelo = torch.load('modelo_entrenado.pt')
```

```
# Convertir el DataFrame a un tensor de PyTorch
tensor = torch.tensor(np.array(data))

#generar el device
device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')

# Hacer las predicciones del modelo
Predicciones, _ = modelo(tensor.to(device))

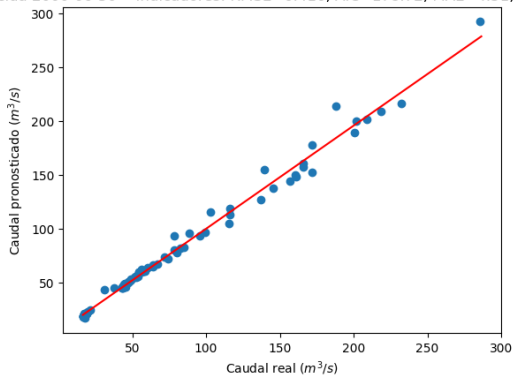
#obtener las predicciones en numpy
Predicciones.detach().cpu().numpy()
```

Anexo C: Figuras extras de modelación

Se presentan en esta sección las gráficas de dispersión de las crecidas no incluidas en la sección de resultados para cada modelo. Se dividen por modelos y tiempos de adelanto, mostrando para Cachapoal todos los *scatterplots* de los conjuntos de entrenamiento y validación/testeo. Para Pangal se presenta lo mismo en las secciones de 1 hora y 3 horas de adelanto. Finalmente en el pronóstico de Pangal a 5 horas se presentan los gráficos de línea y boxplots de los indicadores RMSE, MAE, R^2 , Pbias y AIC.

Cachapoal 1 hora

Crecida 2000-06-30 - Indicadores: RMSE=6.416, AIC=178.72, MAE=4.51, $R^2=0.988$



Crecida 2003-01-21 - Indicadores: RMSE=7.624, AIC=194.95, MAE=4.62, $R^2=0.902$

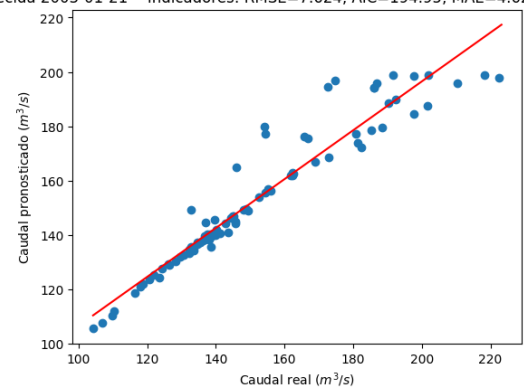
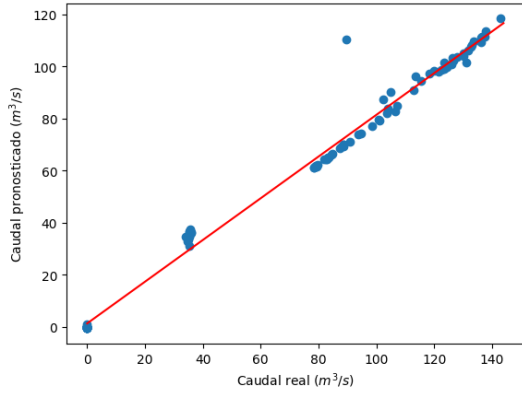
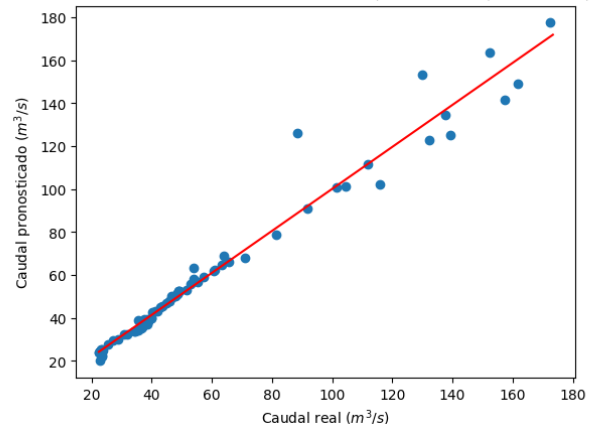


Figura 0.3: Scatterplot conjunto testeo

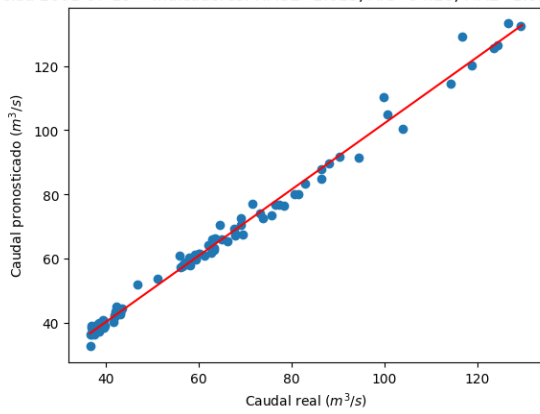
Crecida 2001-08-29 - Indicadores: RMSE=16.942, AIC=270.00, MAE=13.37, R2=0.827



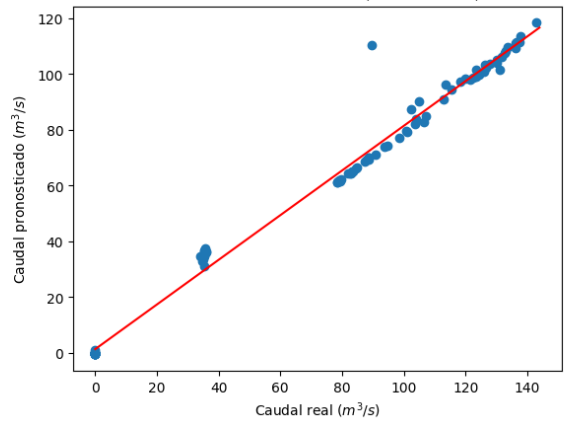
Crecida 2002-05-25 - Indicadores: RMSE=6.049, AIC=173.19, MAE=3.13, R2=0.972



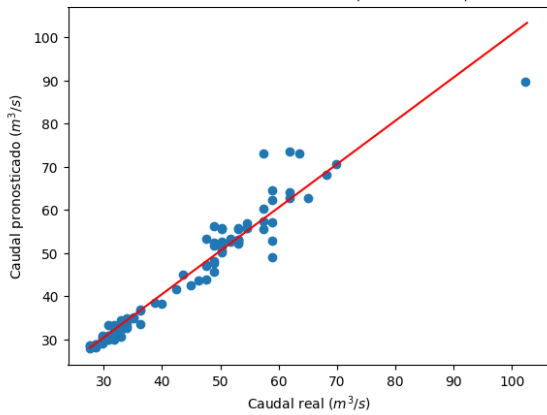
Crecida 2001-07-19 - Indicadores: RMSE=2.613, AIC=94.28, MAE=1.69, R2=0.989



Crecida 2001-08-29 - Indicadores: RMSE=16.942, AIC=270.00, MAE=13.37, R2=0.827



Crecida 2017-02-26 - Indicadores: RMSE=3.443, AIC=120.21, MAE=1.99, R2=0.938



Crecida 2002-08-24 - Indicadores: RMSE=10.410, AIC=224.22, MAE=7.58, R2=0.965

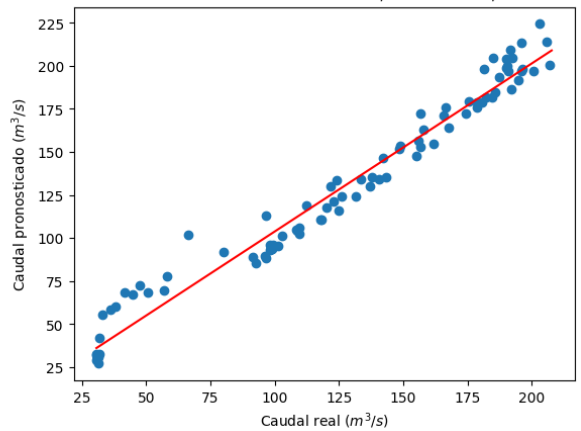
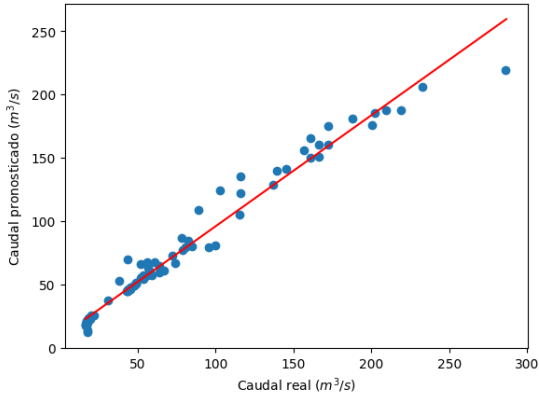


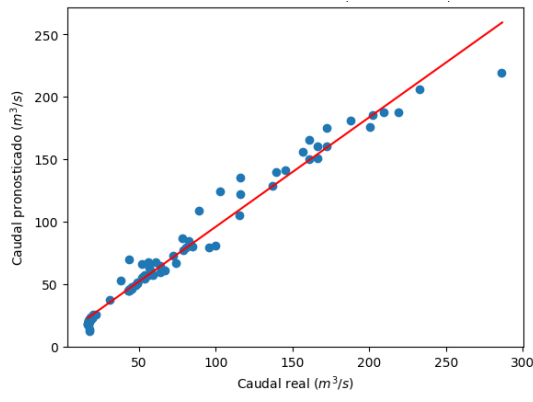
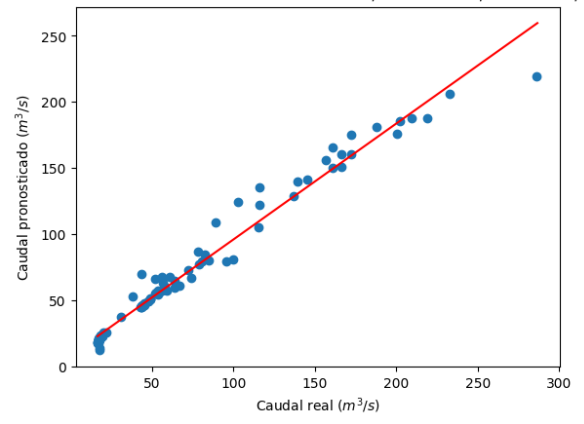
Figura 0.4: Scatterplot conjunto entrenamiento

Cachapoa 3 horas

Crecida 2000-06-30 - Indicadores: RMSE=11.401, AIC=232.77, MAE=6.74, R2=0.955



Crecida 2000-06-30 - Indicadores: RMSE=11.401, AIC=232.77, MAE=6.74, R2=0.955



Crecida 2000-06-30 - Indicadores: RMSE=11.401, AIC=232.77, MAE=6.74, R2=0.955

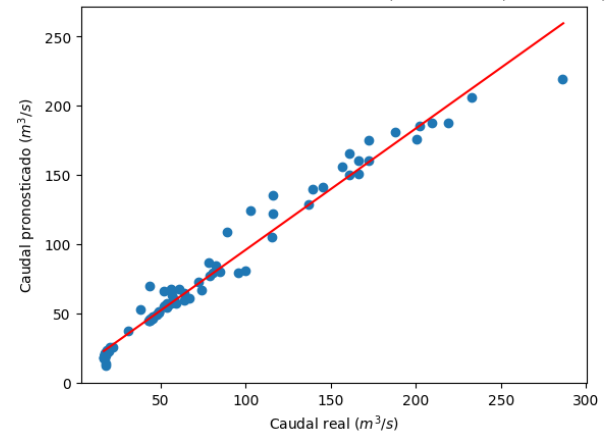
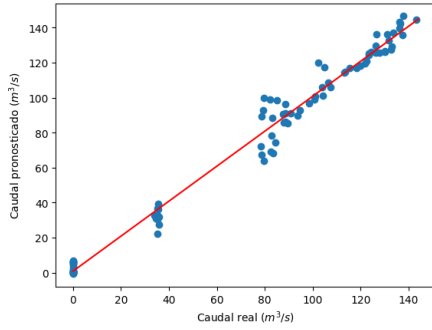
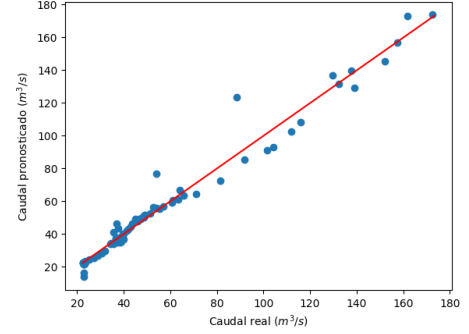


Figura 0.5: Scatter

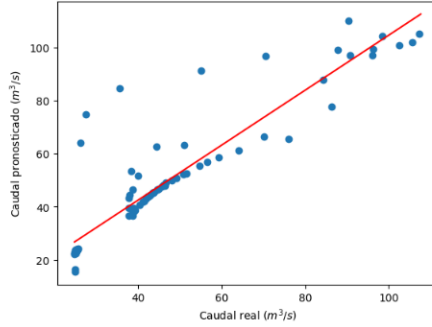
Crecida 2001-08-29 - Indicadores: RMSE=6.375, AIC=178.12, MAE=4.46, R2=0.984



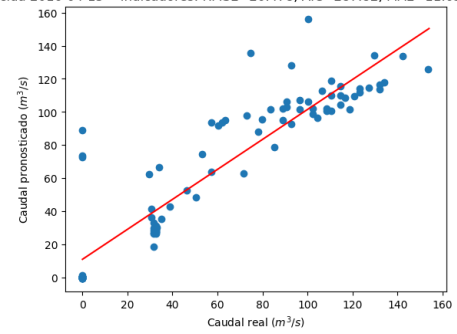
Crecida 2002-05-25 - Indicadores: RMSE=5.759, AIC=168.58, MAE=3.17, R2=0.975



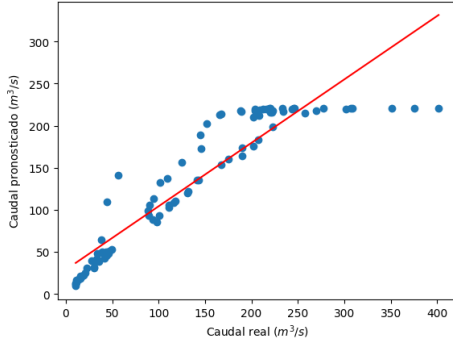
Crecida 2003-06-20 - Indicadores: RMSE=10.423, AIC=224.33, MAE=5.09, R2=0.824



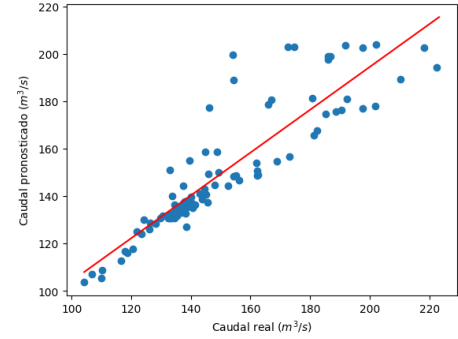
Crecida 2016-04-15 - Indicadores: RMSE=20.478, AIC=287.82, MAE=11.63, R2=0.805



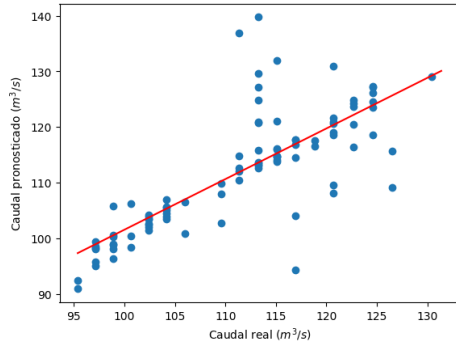
Crecida 2004-04-14 - Indicadores: RMSE=38.350, AIC=346.79, MAE=21.68, R2=0.764



Crecida 2003-01-21 - Indicadores: RMSE=11.886, AIC=236.69, MAE=8.03, R2=0.794



Crecida 2012-12-19 - Indicadores: RMSE=6.742, AIC=183.38, MAE=3.77, R2=0.602



Crecida 2017-02-26 - Indicadores: RMSE=4.749, AIC=150.45, MAE=3.66, R2=0.873

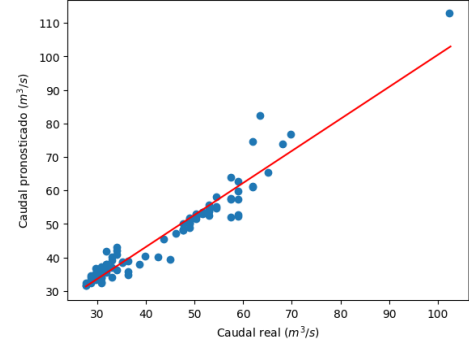
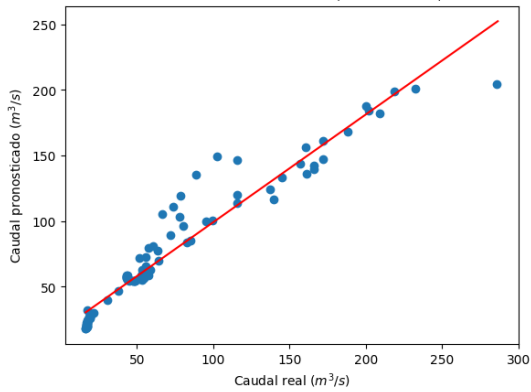


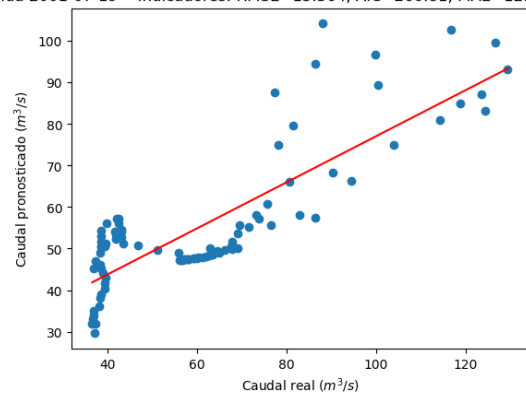
Figura 0.6: Scatterplots conjunto entrenamiento.

Cachapal 5 horas

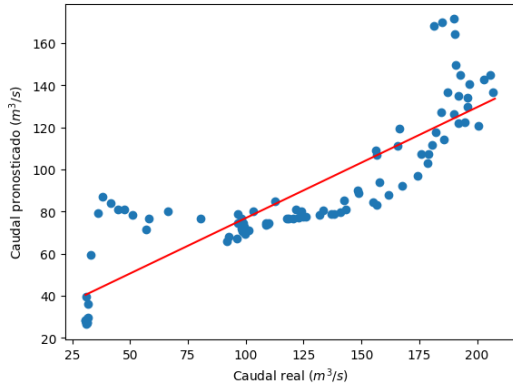
Crecida 2000-06-30 - Indicadores: RMSE=17.705, AIC=274.14, MAE=12.50, R2=0.881



Crecida 2001-07-19 - Indicadores: RMSE=15.364, AIC=260.81, MAE=12.98, R2=0.120



Crecida 2002-08-24 - Indicadores: RMSE=46.301, AIC=364.51, MAE=40.54, R2=-0.832



Crecida 2017-02-26 - Indicadores: RMSE=36.358, AIC=341.78, MAE=27.37, R2=-3.422

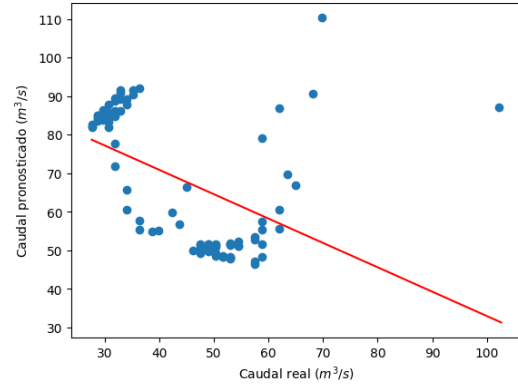
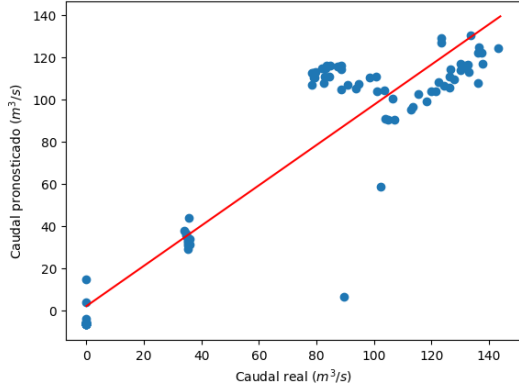
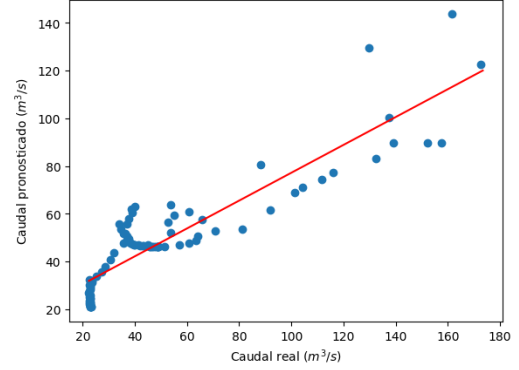


Figura 0.7: scatterplot conjunto testeo.

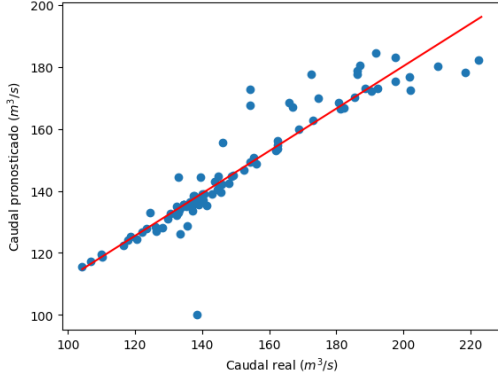
Crecida 2001-08-29 - Indicadores: RMSE=19.129, AIC=281.42, MAE=14.69, R2=0.863



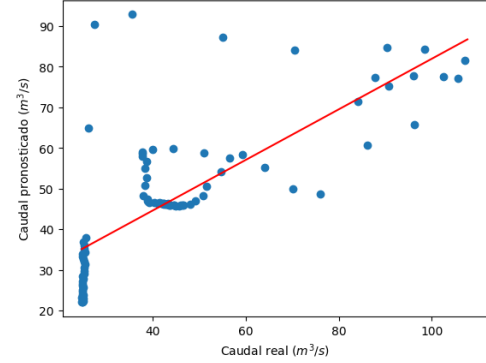
Crecida 2002-05-25 - Indicadores: RMSE=18.255, AIC=277.02, MAE=11.63, R2=0.390



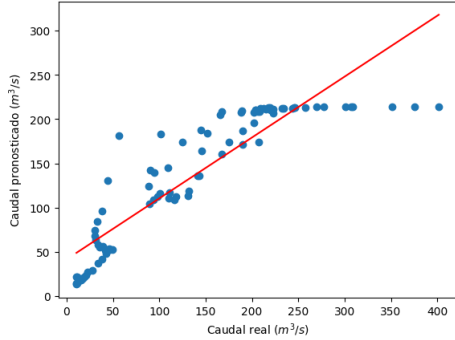
Crecida 2003-01-21 - Indicadores: RMSE=11.408, AIC=232.83, MAE=7.52, R2=0.641



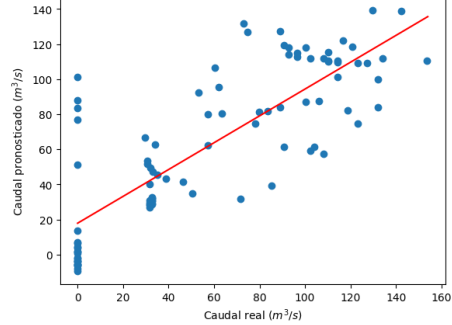
Crecida 2003-06-20 - Indicadores: RMSE=14.774, AIC=257.13, MAE=9.47, R2=0.333



Crecida 2004-04-14 - Indicadores: RMSE=43.959, AIC=359.63, MAE=26.50, R2=0.647



Crecida 2016-04-15 - Indicadores: RMSE=28.470, AIC=318.79, MAE=18.80, R2=0.583



Crecida 2012-12-19 - Indicadores: RMSE=11.837, AIC=236.29, MAE=8.97, R2=0.024

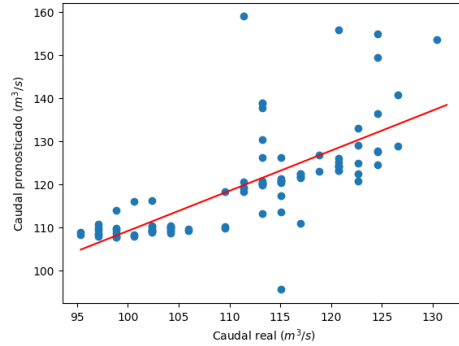
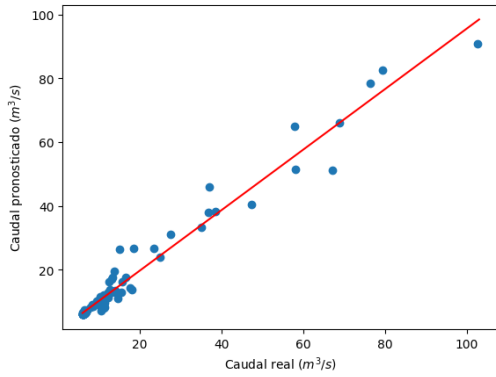


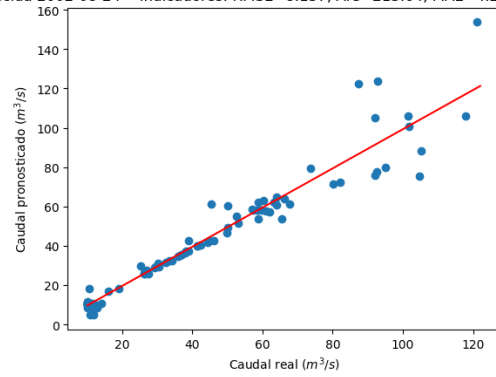
Figura 0.8: Scatterplot conjunto entrenamiento

Pangal 1 hora

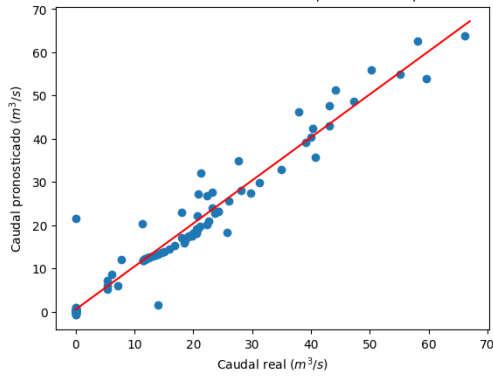
Crecida 2002-05-25 - Indicadores: RMSE=3.254, AIC=121.99, MAE=1.73, R2=0.964



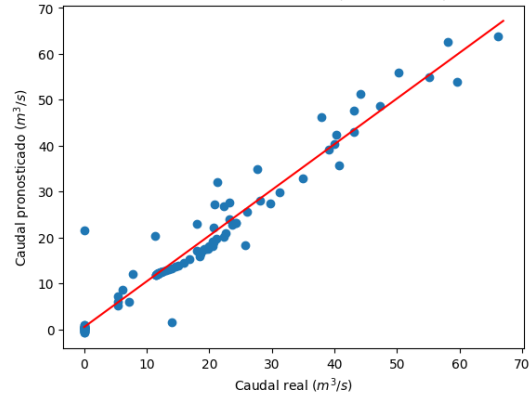
Crecida 2002-08-24 - Indicadores: RMSE=8.137, AIC=213.64, MAE=4.24, R2=0.931



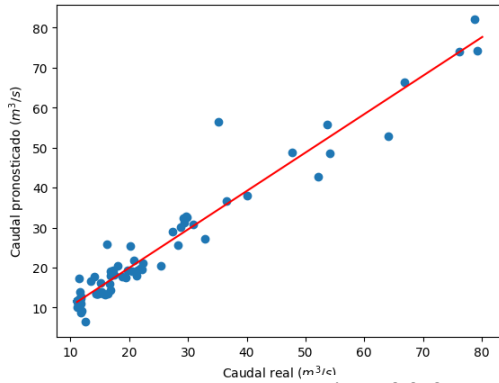
Crecida 2004-04-15 - Indicadores: RMSE=3.756, AIC=136.34, MAE=2.08, R2=0.943



Crecida 2004-04-15 - Indicadores: RMSE=3.756, AIC=136.34, MAE=2.08, R2=0.943



Crecida 2005-08-27 - Indicadores: RMSE=3.448, AIC=127.78, MAE=2.02, R2=0.948



Crecida 2005-08-27 - Indicadores: RMSE=3.448, AIC=127.78, MAE=2.02, R2=0.948

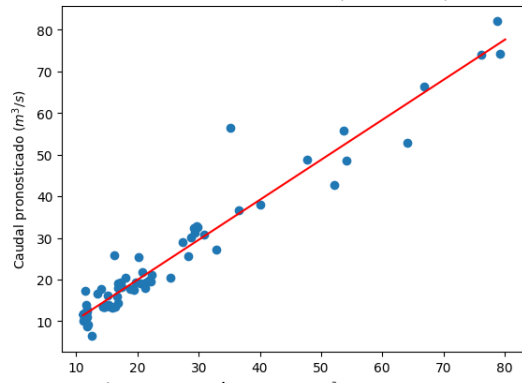
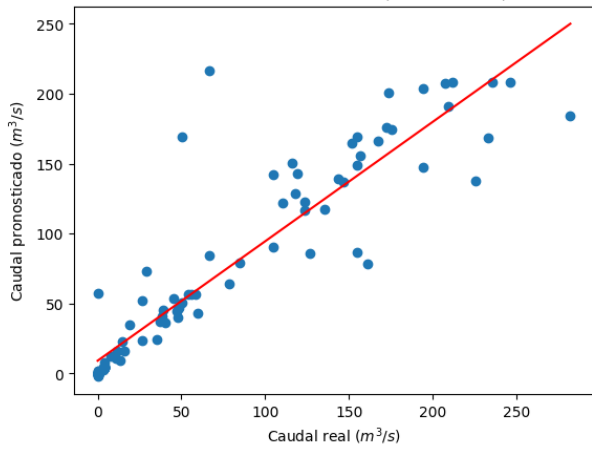


Figura 0.9: Scatterplots conjunto de entrenamiento Pangal 1H real (m³/s)

Crecida 2016-04-16 - Indicadores: RMSE=29.820, AIC=343.52, MAE=13.95, R2=0.822



Crecida 2021-01-30 - Indicadores: RMSE=14.907, AIC=274.19, MAE=5.95, R2=0.818

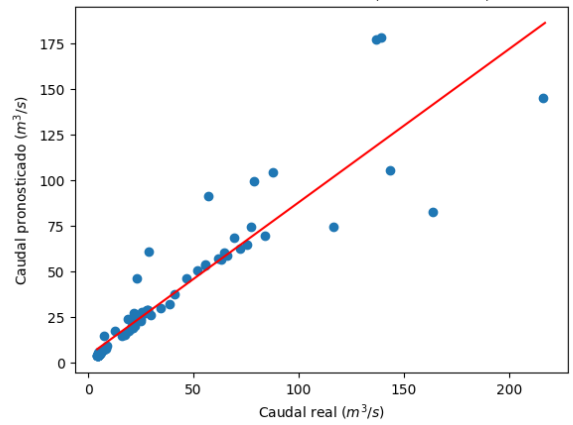
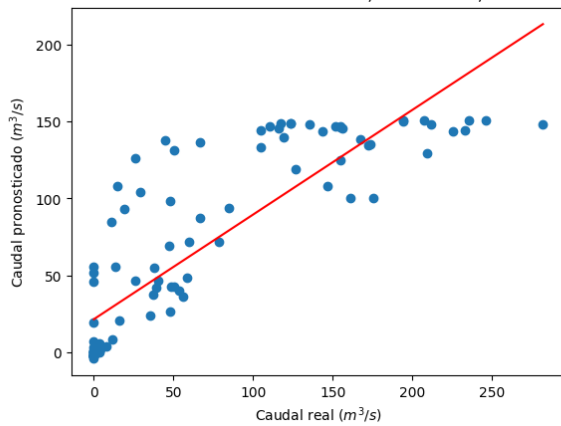


Figura 0.10: Scaterplots de testeo (izquierda) y validación (derecha)

Pangal 3 horas

Crecida 2016-04-16 - Indicadores: RMSE=40.129, AIC=373.21, MAE=25.74, R2=0.564



Crecida 2021-01-30 - Indicadores: RMSE=18.885, AIC=297.84, MAE=8.36, R2=0.743

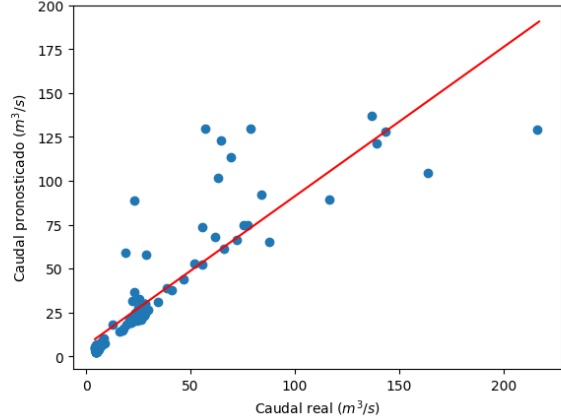
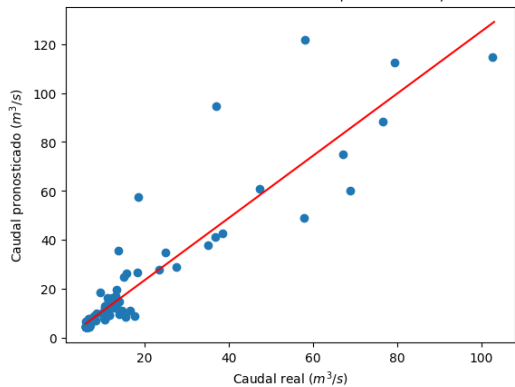
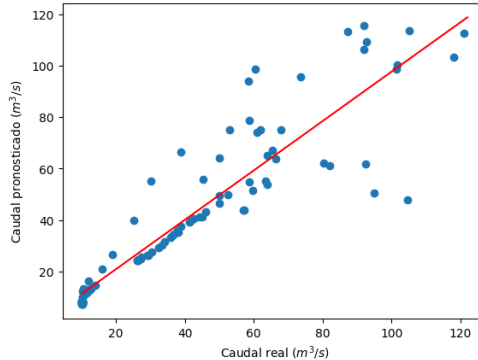


Figura 0.11: Scaterplots de testeo (izquierda) y validación (derecha) Pangal 3 horas

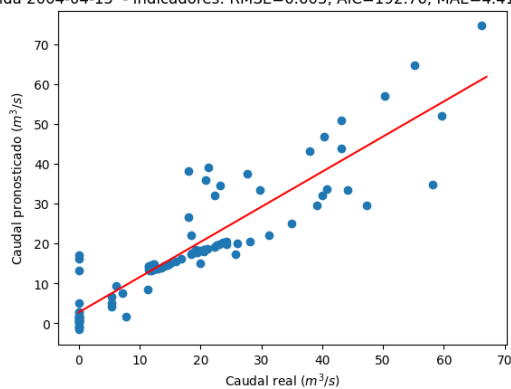
Crecida 2002-05-25 - Indicadores: RMSE=11.018, AIC=243.95, MAE=4.99, R2=0.798



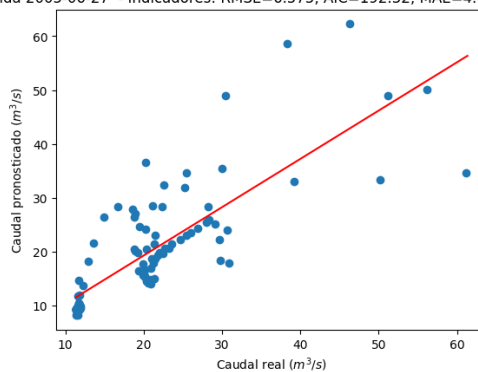
Crecida 2002-08-24 - Indicadores: RMSE=12.813, AIC=259.04, MAE=7.51, R2=0.834



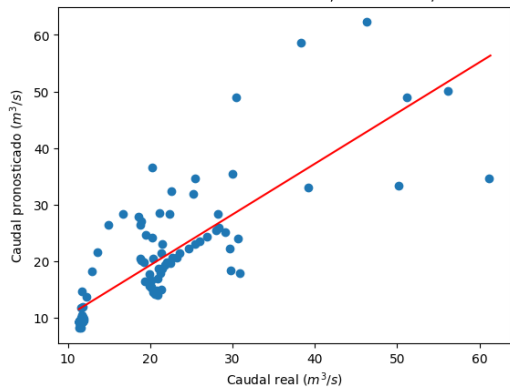
Crecida 2004-04-15 - Indicadores: RMSE=6.603, AIC=192.76, MAE=4.41, R2=0.804



Crecida 2005-06-27 - Indicadores: RMSE=6.575, AIC=192.32, MAE=4.68, R2=0.651



Crecida 2005-06-27 - Indicadores: RMSE=6.575, AIC=192.32, MAE=4.68, R2=0.651



Crecida 2012-05-27 - Indicadores: RMSE=19.636, AIC=301.74, MAE=10.57, R2=0.761

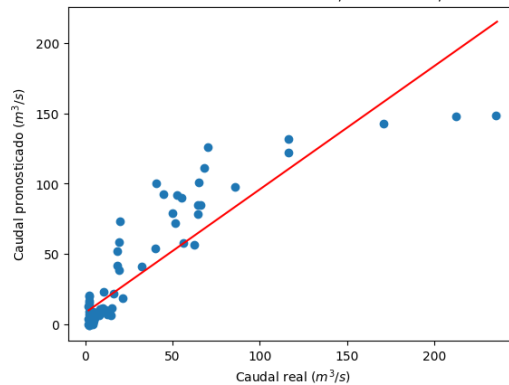


Figura 0.12: Scatterplots conjunto de entrenamiento Pangal 1H

Pangal 5 horas

En el presente anexo se presentan los resultados del modelo Pangal con tiempo de adelanto 5 horas.

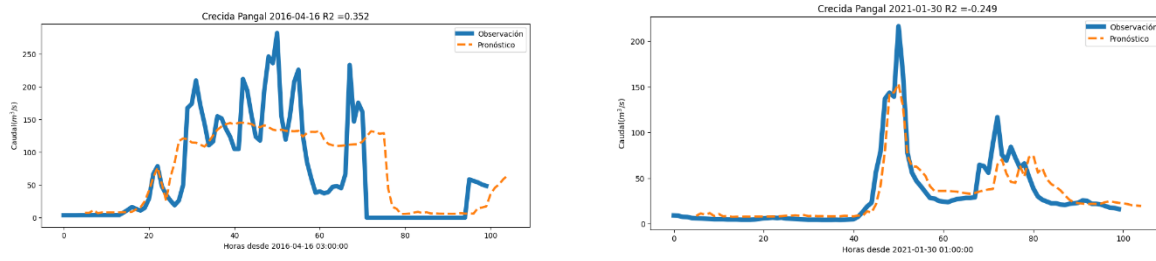


Figura 0.13: Crecidas de testeo (izquierda) y validación (derecha)

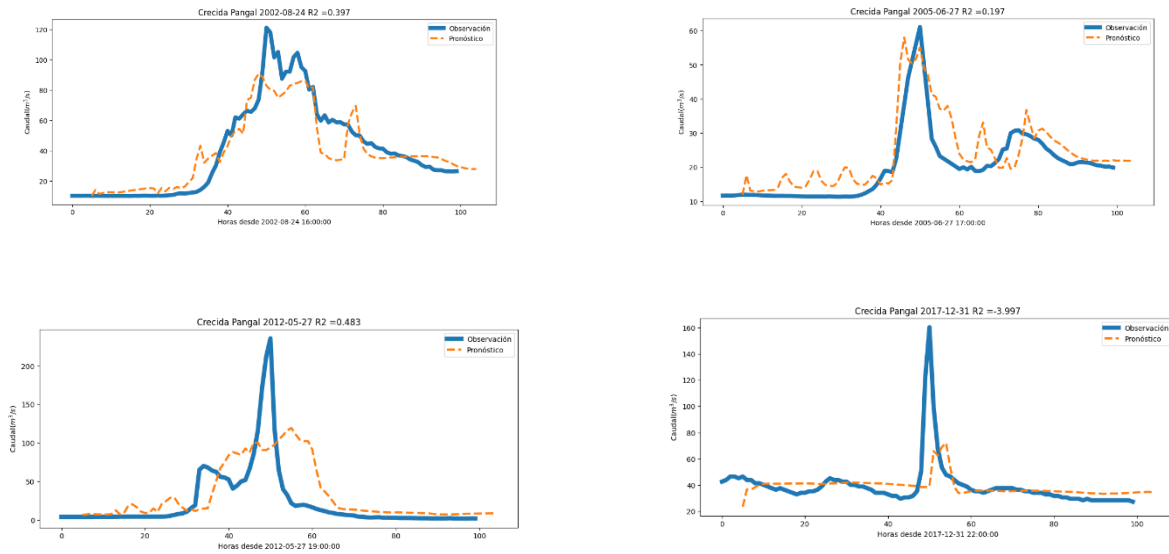


Figura 0.14: Crecidas de entrenamiento.

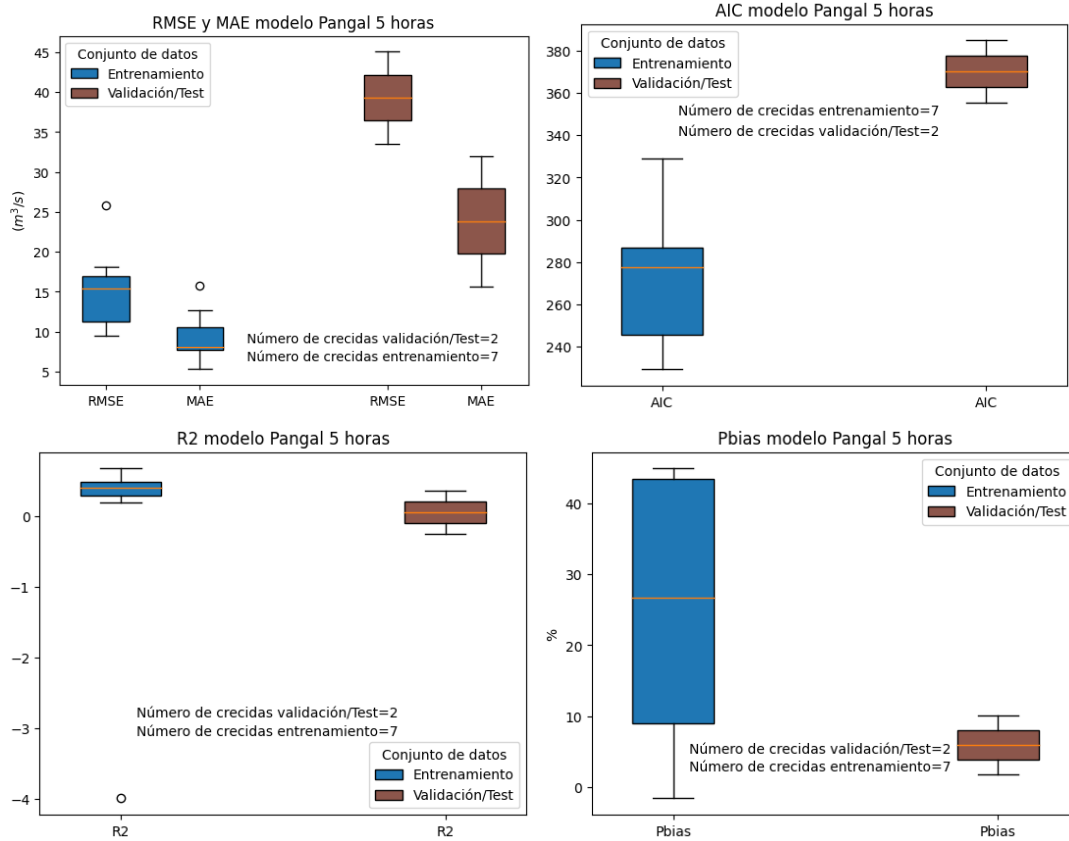
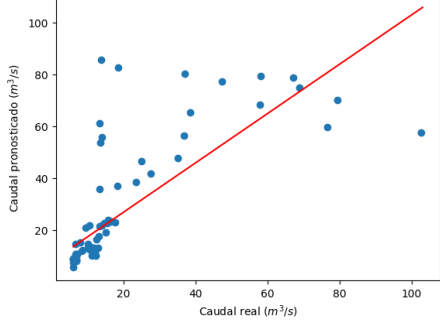
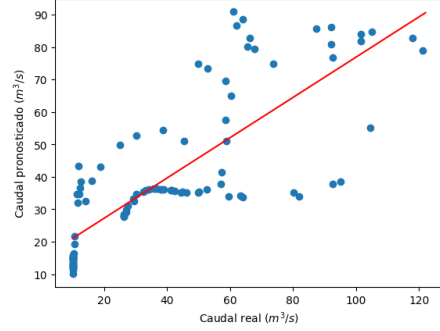


Figura 0.15: Boxplot Modelo Pangal 5 horas de pronóstico

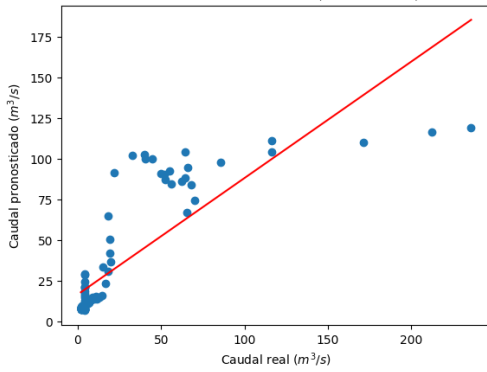
Crecida 2002-05-25 - Indicadores: RMSE=15.817, AIC=280.11, MAE=8.54, R2=0.485



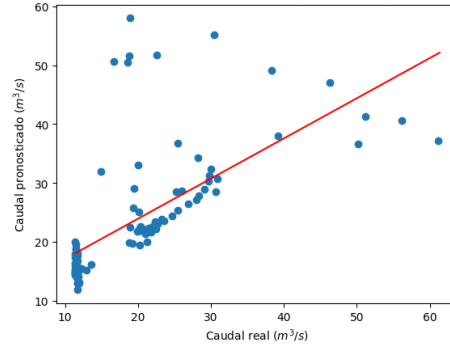
Crecida 2002-08-24 - Indicadores: RMSE=18.132, AIC=293.77, MAE=12.63, R2=0.397



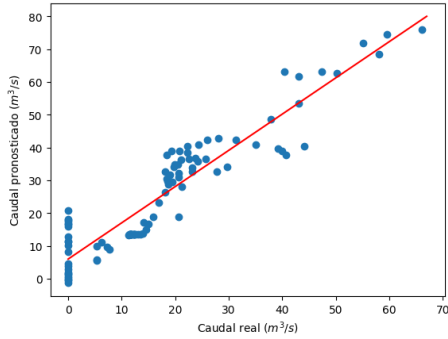
Crecida 2012-05-27 - Indicadores: RMSE=25.831, AIC=329.16, MAE=15.79, R2=0.483



Crecida 2005-06-27 - Indicadores: RMSE=9.509, AIC=229.23, MAE=5.28, R2=0.197



Crecida 2004-04-15 - Indicadores: RMSE=10.399, AIC=238.17, MAE=8.09, R2=0.672



Crecida 2017-12-31 - Indicadores: RMSE=15.455, AIC=277.80, MAE=7.41, R2=-3.997

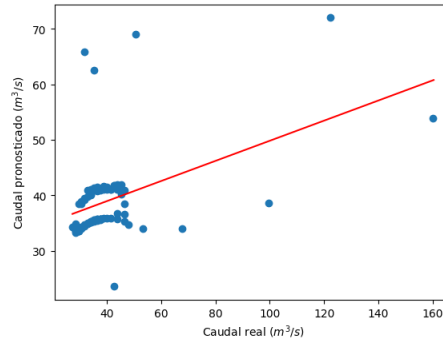
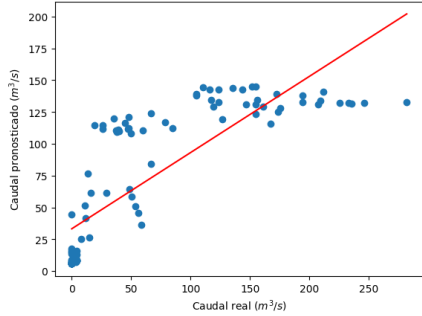


Figura 0.17: Scatterplot conjunto entrenamiento Pangal 5 horas

Crecida 2016-04-16 - Indicadores: RMSE=45.068, AIC=384.82, MAE=31.99, R2=0.352



Crecida 2021-01-30 - Indicadores: RMSE=33.543, AIC=355.28, MAE=15.69, R2=-0.249

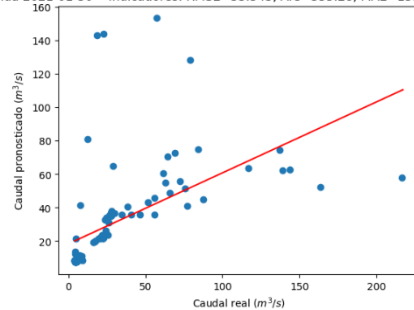


Figura 0.16: Scatterplot testeo y validación Pangal 5 horas