UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE OBRAS CIVILES

VALPARAÍSO - CHILE

METODOLOGÍA DE PROYECCIÓN DE CONSUMO DE AGUA POTABLE PARA EL ABASTECIMIENTO DE LA SERENA Y COQUIMBO

CRISTIAN ROBERTO SOTO MAUREIRA

Memoria para optar al Título de

INGENIERO CIVIL

Profesor Guía

Danilo Kusanovic

Profesor Correferente

Alvaro Ossandón

Marzo 2025

AGRADECIMIENTOS

Quisiera partir agradeciendo a todos quienes me motivaron a cerrar esta etapa durante muchos años, en especial a mi esposa Daniela, mi madre Irene y mi tía One.

Además, quería agradecer el apoyo y motivación del profesor Kusanovic en guiar este proceso y dar la mayor de las facilidades para quienes buscamos cerrar esta etapa.

Este trabajo es resultado de la motivación del equipo de Inteligencia Operacional de ESVAL, en particular del ingeniero Iván Gonzalez, quien implementó la metodología que se presenta, la cual ha servido como base para generar este estudio.

DEDICATORIA

A mi hija Gabrielita y su amor que inspira a ser mejor cada día…

RESUMEN

La memoria aborda la crisis hídrica en la región de Coquimbo, Chile, centrándose en La Serena y Coquimbo, donde la escasez de agua ha afectado gravemente la agricultura, la economía y la calidad de vida. El objetivo principal fue desarrollar un modelo de predicción de la demanda operativa diaria de agua potable para optimizar su gestión. Se utilizaron datos de caudal del sistema SCADA (SMARTSCADA) de Aguas del Valle y variables meteorológicas de Open-Meteo, abarcando de enero de 2022 a diciembre de 2024. La metodología integró el Análisis de Componentes Principales (PCA) para reducir la dimensionalidad de los datos de consumo de agua potable y el modelo ARIMAX para predecir el consumo, considerando factores como temperatura, precipitación y cobertura de nubes. Los resultados muestran un modelo preciso, con un Error Absoluto Medio (MAE) de 38.9 l/s y un Error Porcentual Absoluto Medio (MAPE) de 2.97%, destacando la temperatura como la variable más influyente. Implementado en Excel y usado diariamente por Aguas del Valle, el modelo permite anticipar la demanda con siete días de horizonte, optimizando el uso de fuentes superficiales y subterráneas, y mejorando la eficiencia hídrica en un contexto de escasez.

ABSTRACT

The report addresses the water crisis in the Coquimbo region of Chile, focusing on La Serena and Coquimbo, where water scarcity has severely affected agriculture, the economy, and quality of life. The main objective was to develop a model to predict the daily operational demand for drinking water in order to optimize its management. Flow data from the Aguas del Valle SCADA system (SMARTSCADA) and meteorological variables from Open-Meteo were used, covering the period from January 2022 to december 2024. The methodology integrated Principal Component Analysis (PCA) to reduce the dimensionality of drinking water consumption data and the ARIMAX model to predict consumption, considering factors such as temperature, precipitation, and cloud cover. The results show an accurate model, with a Mean Absolute Error (MAE) of 38.9 l/s and a Mean Absolute Percentage Error (MAPE) of 2.97%, highlighting temperature as the most influential variable. Implemented in Excel and used daily by Aguas del Valle, the model allows for anticipating demand with a seven-day horizon, optimizing the use of surface and groundwater sources, and improving water efficiency in a context of scarcity.

GLOSARIO

1. **Acueducto**: Conducto artificial, generalmente de gran longitud, diseñado para transportar agua desde una fuente (e.g., río Elqui) hasta centros de distribución, como los que conectan Las Rojas con La Serena y Coquimbo.
2. **ARIMAX**: Modelo estadístico que extiende el ARIMA al incorporar variables exógenas (e.g., meteorológicas), utilizado en la tesis para predecir la demanda de agua con base en datos temporales y climáticos.
3. **Captación**: Proceso o infraestructura para extraer agua de una fuente natural, como la captación superficial del río Elqui en Las Rojas o los pozos subterráneos del acuífero Elqui Bajo.
4. **Caudal**: Flujo de agua medido en litros por segundo (l/s), representa la cantidad de agua que circula por un sistema en un tiempo dado; en este estudio, es el indicador principal del consumo.
5. **Estanque de Regulación**: Depósito utilizado para almacenar agua y estabilizar el suministro en la red de distribución, como los 38 estanques con capacidad total de 55,000 m^3 en el sistema de La Serena-Coquimbo.
6. **PCA (Análisis de Componentes Principales)**: Técnica estadística que transforma datos correlacionados en componentes principales no correlacionados, empleada para reducir la dimensionalidad de los datos horarios de caudal y meteorológicos.
7. **POSIXct**: Formato en R que representa fechas y horas como segundos desde el 1 de enero de 1970 UTC, usado para alinear temporalmente los datos de caudal y meteorológicos en el preprocesamiento.
8. **SCADA**: Acrónimo de *Supervisory Control and Data Acquisition* (Control de Supervisión y Adquisición de Datos), sistema que monitorea y controla en tiempo real variables operativas como caudales y niveles en la infraestructura hídrica.
9. **Series Temporales**: Conjunto de datos ordenados cronológicamente, como los registros horarios de caudal de SCADA, analizados para identificar patrones y realizar predicciones de demanda.

10. **Varianza Explicada**: Proporción de la variabilidad total de los datos capturada por los componentes principales en el PCA, criterio para seleccionar componentes que expliquen al menos el 90% de la varianza en la tesis.

ÍNDICE

AGRADECIMIENTOS 2

DEDICATORIA 3

RESUMEN 4

ABSTRACT 5

GLOSARIO 6

ÍNDICE DE TABLAS 11

ÍNDICE DE FIGURAS 11

ÍNDICE DE GRÁFICOS 12

1 Introducción 13

1.1 Motivación 13

1.2 Objetivos 13

1.3 Alcance 14

1.4 Metodología 14

1.5 Justificación 14

2 Marco Teórico 16

2.1 Antecedentes 16

2.2 Descripción de la Zona Abastecida 17

2.3 Fundamentos de PCA 34

2.4 Modelado Predictivo con ARIMAX 35

2.5 Proceso de proyección y reconstrucción de la serie original 39

3 Descripción de los Datos 41

3.1 Datos de Caudal de Agua 41

3.2 Datos Meteorológicos 42

3.3 Análisis Exploratorio 43

4 Metodología 52

4.1 Enfoque General 52

4.2 Justificación de Análisis de Componentes Principales (PCA) 52

4.3 Recopilación y Preprocesamiento de Datos 52

4.4 Cálculo del caudal de salida 54

4.5 Aplicación del PCA 56

4.6 Modelado Predictivo con ARIMAX 57

4.7 Resumen 58

5 Resultados 59

5.1 Desempeño del Modelo Predictivo 59

5.2 Métricas de error agregadas 60

6 Gestión Operativa 61

6.1 Aplicación del modelo en la Gestión Operativa. 61

6.2 Descripción de la Herramienta Operativa 62

7 Conclusiones y Recomendaciones 64

7.1 Conclusiones 64

7.2 Recomendaciones y Futuras Líneas de Investigación 66

8 Bibliografía 67

9 Anexos 69

9.1 POSIXct 69

9.2 Preparación de los Datos 70

ÍNDICE DE TABLAS

TABLA 1 CANTIDAD DE SENSORES NIVEL Y CAUDAL INSTALADOS EN LA SERENA Y COQUIMBO 24

TABLA 2 RESUMEN ESTADÍSTICAS DESCRIPTIVAS 49

TABLA 3 RESUMEN COMPARACIÓN ESTACIONAL 50

TABLA 4 CAUDAL DE CONSUMO AÑO NUEVO 51

ÍNDICE DE FIGURAS

ILUSTRACIÓN 1 INFRAESTRUCTURA GEOREFERENCIADA LA SERENA COQUIMBO 20

ILUSTRACIÓN 2 ESQUEMA SERVIDORES SMARTSCADA 22

ILUSTRACIÓN 3 HERRAMIENTA SMARTSCADA 23

ILUSTRACIÓN 4 ESQUEMA SPAP LAS ROJAS 25

ILUSTRACIÓN 5 FILTROS SPAP LAS ROJAS 26

ILUSTRACIÓN 6 PANTALLA SCADA Y TELECONTROL DE SONDAJE 27

ILUSTRACIÓN 7 ADUCCIÓN TRAMO LAS ROJAS – LAZARETO 28

ILUSTRACIÓN 8 ADUCCIÓN - TRAMO LAZARETO - BALMACEDA 29

ILUSTRACIÓN 9 ACUEDUCTO ANTIGUO 30

ILUSTRACIÓN 10 ACUEDUCTO NUEVO 30

ILUSTRACIÓN 11 ACUEDUCTO LAZARETO – COQUIMBO 32

Ilustración 12 Ubicación Estanques de Regulación 33

Ilustración 13 Estanque Semi Enterrado San Ramón 33

Ilustración 14 Caudal de Salida (Semana de 10-02-25 a 18-02-25) 50

Ilustración 15 Caudal de Salida (Semana de 02-06-24 a 10-06-24) 50

Ilustración 16 Caudal de Salida del 31-12-24 a 01-01-25 51

Ilustración 17 Procesos del Modelo de predicción de consumo de agua potable 59

Ilustración 18 Resultados de Proyección – Datos observados versus Predicción 60

Ilustración 19 Coordinación diaria para aplicación del modelo 62

ÍNDICE DE GRÁFICOS

Gráfico 1 Caudal de salida diario en el Elqui 43

Gráfico 2 Caudal de salida semanal en La Serena - Coquimbo 44

Gráfico 3 Descomposición de la Serie Temporal 45

Gráfico 4 Correlación Caudal vs Temperatura 46

Gráfico 5 Correlación Caudal vs Precipitación 47

Gráfico 6 Correlación Caudal vs Cobertura de Nubes 48

Gráfico 7 Matriz de Correlación de Pearson 48

1 Introducción

1.1 Motivación

La gestión eficiente del agua en La Serena y Coquimbo es vital para enfrentar una crisis hídrica que compromete el abastecimiento de más de medio millón de habitantes y el sostenimiento de actividades económicas clave, como la agricultura y el turismo, en un entorno de recursos cada vez más escasos. La importancia de esta tarea radica en su capacidad para garantizar la disponibilidad de agua potable, mitigar los impactos de la sequía prolongada y preservar la calidad de vida en una región semiárida golpeada por condiciones climáticas adversas. Sin embargo, predecir la demanda de agua presenta desafíos significativos, como la variabilidad estacional, la influencia de factores externos impredecibles y la falta de herramientas precisas que integren datos operativos y climáticos en tiempo real. Este trabajo aborda estas problemáticas proponiendo una solución que busca mejorar la gestión hídrica, contribuyendo a superar las limitaciones actuales y a fortalecer la resiliencia ante un futuro incierto. Los capítulos siguientes explorarán el enfoque adoptado y los resultados alcanzados en este esfuerzo por responder a una necesidad crítica de nuestra sociedad.

1.2 Objetivos

1.2.1 Objetivo General

Desarrollar un modelo de predicción de la demanda operativa diaria de agua potable para optimizar la gestión del recurso hídrico en La Serena y Coquimbo, en un contexto de escasez hídrica prolongada.

1.2.2 Objetivos Específicos

- Desarrollar un modelo ARIMAX que integre variables meteorológicas como factores exógenos para predecir la demanda de agua a corto plazo (7 días).
- Validar la precisión del modelo mediante métricas estadísticas y análisis visual, comparando las predicciones con datos históricos.

- Integrar el modelo en la operación diaria de Aguas del Valle para optimizar la gestión de fuentes superficiales y subterráneas.

1.3 Alcance

El estudio abarca datos desde el 1 de enero de 2022 hasta diciembre de 2024, seleccionados para evitar distorsiones por la pandemia de COVID-19. La investigación se centra en las ciudades de La Serena y Coquimbo, ubicadas en la región de Coquimbo, Chile. El modelo incluye datos de caudal de agua (l/s) registrados por SmartSCADA y variables meteorológicas (temperatura, precipitación y cobertura de nubes) obtenidas de Open-Meteo.

1.4 Metodología

El enfoque metodológico combina la recopilación de datos históricos y proyectados, análisis estadístico y técnicas de modelación avanzada:

- **Recopilación de Datos**: Datos de caudal y meteorológicos de Open-Meteo.
- **Preprocesamiento**: Alineación temporal, filtrado de días completos y selección de horas relevantes.
- **Análisis Exploratorio**: Identificación de patrones temporales y correlaciones entre variables.
- **Modelación**: Uso de PCA para reducir dimensionalidad y ARIMAX para modelar la serie temporal con variables exógenas.
- **Validación**: Evaluación del modelo con métricas como RMSE, MAE y MAPE, y análisis visual de las predicciones.

1.5 Justificación

La región de Coquimbo, en Chile, atraviesa una de las crisis hídricas más severas de su historia. Esta situación, agravada por un déficit de precipitaciones del 59% en el último año, ha puesto en jaque la disponibilidad de agua para más de 500.000 habitantes de La Serena y Coquimbo,

así como para actividades económicas esenciales como la agricultura y el turismo. En este contexto, la gestión eficiente del agua potable no es solo una necesidad operativa, sino un imperativo para la sostenibilidad y la calidad de vida en la región.

El desarrollo de un modelo predictivo permite a Aguas del Valle, la empresa encargada del abastecimiento de agua potable, anticipar la demanda diaria de agua con un horizonte de planificación de 7 días. Esta capacidad de predicción optimiza el uso de fuentes hídricas disponibles, como el río Elqui (superficial) y el acuífero Elqui Bajo (subterráneo), priorizando aquellas de menor costo operativo y reduciendo la dependencia de recursos más costosos o vulnerables. Por ejemplo, al prever valores máximos de demanda durante períodos cálidos o eventos masivos como las celebraciones de Año Nuevo, la empresa puede ajustar la producción y distribución, evitando las pérdidas de agua y minimizando el riesgo de desabastecimiento.

Además, la integración del modelo en la operación diaria de Aguas del Valle, mediante reuniones de coordinación y ajustes en tiempo real basados en las predicciones, demuestra su viabilidad práctica. Este enfoque no solo mejora la eficiencia hídrica, sino que también genera ahorros económicos significativos al reducir los costos asociados a la sobreproducción o a la activación innecesaria de fuentes alternativas. La herramienta también tiene el potencial de servir como un sistema de alerta temprana ante condiciones extremas, como sequías prolongadas o incrementos inesperados en el consumo, fortaleciendo la resiliencia del sistema de abastecimiento frente al cambio climático y la creciente presión demográfica.

2 Marco Teórico

2.1 Antecedentes

2.1.1 Escenario Hídrico

La Región de Coquimbo enfrenta una crisis hídrica extrema que se ha prolongado por más de una década. Esta situación se ha agravado en los últimos años, afectando dramáticamente a la agricultura, la economía y la vida cotidiana de sus habitantes[12]. Los embalses de la región a mayo 2024, estaban en niveles críticos: el Embalse Cogotí se encuentraba en un 0% de su capacidad, mientras que el Embalse La Paloma, el más grande de Chile, está al 1%. Esto ha llevado a los diferentes entes a generar acciones, como:

1. Gestión de Recursos Hídricos: Se han implementado medidas para optimizar el uso del agua disponible, incluyendo la modernización de sistemas de riego y la promoción de técnicas de agricultura sostenible[1].

2. Inversión en Infraestructura: Se están construyendo y mejorando embalses y sistemas de distribución de agua para asegurar un suministro más eficiente[1].

3. Programas de Apoyo a Agricultores: Se han establecido programas de apoyo financiero y técnico para pequeños y medianos agricultores, ayudándolos a enfrentar la sequía mediante la construcción de pozos y la adopción de tecnologías de riego más eficientes[1].

[1] Severa crisis hídrica en la Región de Coquimbo: medidas, propuestas y .... https://www.paiscircular.cl/agua/severa-crisis-hidrica-en-la-region-de-coquimbo-medidas-propuestas-y-reflexiones-ante-una-carrera-contra-el-tiempo/.

[2] El lado feo de la sequía en Chile: el caso de la región de Coquimbo .... https://tiemporeal.periodismoudec.cl/2021/09/03/el-lado-feo-de-la-sequia-en-chile-el-caso-de-la-region-de-coquimbo/.

4. Educación y Conciencia Pública: Campañas de sensibilización para promover el uso responsable del agua entre la población[2].

Sin embargo, estas acciones no han sido efectivas, provocando impactos en la comunidad. En la agricultura, la pérdida de miles de hectáreas de cultivos, cierre de empresas agrícolas y un aumento en el desempleo en el sector, generando un endeudamiento significativo entre los agricultores y ha afectado negativamente la economía regional[1]. En el ámbito, de "Salud y Saneamiento", la sequía ha generado problemas sanitarios en algunas comunidades rurales, donde el acceso al agua potable es limitado[2] y la falta de agua ha afectado la biodiversidad local, con una disminución en la vegetación y la fauna autóctona[2].

En el caso particular de la cuenca del Rio Elqui, de cual se abastecen las ciudades de Coquimbo y La Serena, la situación no es diferente. Esto se refleja en el agua acumulada en Embalse Puclaro, la que a mayo 2025 representa un 16% equivalente a 32.4 [Hm^3] de 209 [Hm^3].

2.2 Descripción de la Zona Abastecida

Como objeto de estudio, la zona abastecida corresponde a las ciudades de La Serena y Coquimbo, perteneciente a la región de Coquimbo y parte de la cuenca del Rio Elqui. Sobre estas se describen los principales aspectos:

2.2.1 Características Climáticas.

Para las ciudades de La Serena y Coquimbo, el clima presenta una segmentación clara por estaciones. Veranos sin precipitaciones, con nubosidad matinal y lloviznas que se disipan al mediodía, temperaturas entre 20 y 24 °C. Inviernos con temperaturas entre 6 y 12 °C, influenciadas por la corriente de Humboldt. Las precipitaciones anuales son aproximadamente 120 mm, concentradas en los meses de junio a agosto.

2.2.2 Características demográficas

La conurbación La Serena-Coquimbo, conocida como Gran La Serena, es la cuarta área metropolitana más poblada de Chile, después de Gran Santiago, Gran Concepción y Gran Valparaíso. A continuación, se detallan las características demográficas actualizadas:

Según el Censo Nacional de 2017, la población de la conurbación La Serena-Coquimbo era de **534,239 habitantes**, desglosada en:

- **La Serena**: 249,656 habitantes.
- **Coquimbo**: 256,735 habitantes (zona urbana).

Actividad Económica y Ocupación consta de los siguientes sectores principales:

- **Turismo**: La Serena es un destino clave en verano, duplicando su población temporalmente por visitantes.
- **Minería**: Coquimbo destaca por su actividad portuaria y exportaciones de mineral de hierro (96% de las exportaciones según DataChile 2016).
- **Agricultura**: Producción de uvas y otros cultivos en valles cercanos.
- **Servicios**: Comercio y educación (e.g., Universidad de La Serena).

Fuerza Laboral: En 2016, DataChile reportó 321,380 personas empleadas en la región, con un desempleo del 8%. Para 2025, se estima una fuerza laboral de aproximadamente 350,000 personas en la conurbación, ajustada al crecimiento poblacional.

2.2.3 Características del Abastecimiento Agua

La principal fuente de abastecimiento del sistema Coquimbo-La Serena corresponde a la captación superficial del río Elqui, aguas abajo del embalse Puclaro, ubicada en el sector Las Rojas, a través de una canalización construida a un costado del río en su ribera sur, desde donde las aguas son trasladadas por un canal abierto hasta la planta de tratamiento. La captación cuenta con un caudal de explotación de 750 l/s.La alimentación de La Serena y Coquimbo desde el sector Las Rojas

se realiza por medio de dos acueductos y una tubería en presión, cuyo trayecto se extiende hasta el sector Lazareto. Además, es de alta relevancia comentar que en planta Las Rojas existe un tranque de acumulación, de 47.000 m3, utilizado como infraestructura de apoyo ante crecidas que generen una indisponibilidad del aporte superficial. Desde este punto nacen dos alimentadoras que abastecen el sector Las Compañías y dos alimentadoras que abastecen todo el sector sur de La Serena y la localidad de Coquimbo. Adicionalmente, se cuenta con captaciones subterráneas en el acuífero Elqui Bajo:

- Pozos Sector Las Rojas
- Pozos Sector Punta de Piedra
- Dren Punta de Piedra
- Pozos Saturno y Gabriela Mistral
- Pozos Alfalfares
- Pozos Las Compañías
- Pozos Sector Urbano La Serena – Coquimbo.

Se tiene un total de 63 Pozos, equivalentes al 50% del volumen de producción anual.

2.2.4 Características del Sistema de Distribución

El sistema Coquimbo-La Serena dispone de 38 estanques de regulación con un total de 55.000 m3 de regulación, que abastecen a los distintos sectores de cada localidad. La mayoría de los estanques son del tipo semienterrado. El abastecimiento de los clientes de Coquimbo y La Serena se realiza por medio de una extensa red de distribución que cubre prácticamente la totalidad del territorio operacional de la empresa. La red de distribución de La Serena y Coquimbo consiste de aproximadamente 1.315 km de tuberías. El sistema posee plantas elevadoras y presurizadoras, válvulas de entrada de estanques y válvulas reguladoras de presión. Además, se cuenta con infraestructura de apoyo, como equipos de monitoreo, grifos, válvulas de corte, cámaras y arranques domiciliarios. A continuación, se presenta un esquema general de la infraestructura georreferenciada

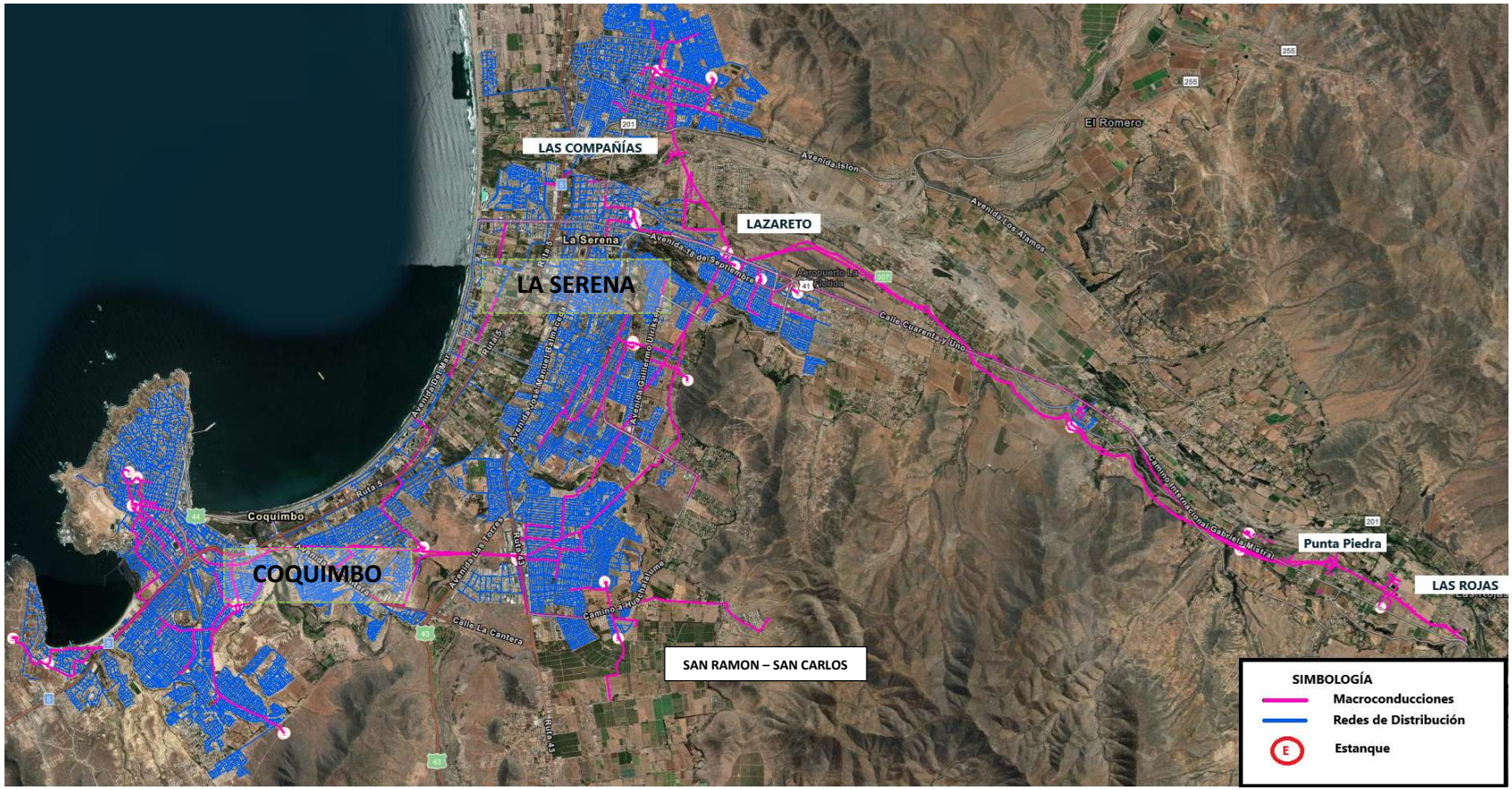


Ilustración 1 Infraestructura Georeferenciada La Serena Coquimbo

2.2.5 Demanda de Agua Potable en la zona de Estudio

Según los datos del Plan de Desarrollo de Aguas del Valle para Coquimbo y La Serena[iiiiv], la población abastecida dentro del territorio operacional de La Serena y Coquimbo, por ADV, es de 563.557 habitantes, equivalente a más de 188.000 clientes con una Demanda Máxima Diaria de 1805 l/s.

2.2.6 Sistemas de telemetría y telecontrol

El sistema se administra mediante el sistema **SCADA**, es un acrónimo que significa Supervisory Control and Data Acquisition (control de supervisión y adquisición de datos), donde se mantienen datos operacionales de infraestructura como estanques, sondajes, Plantas de Tratamiento de Agua Potable (PTAP), Plantas Elevadoras de Agua Potable (PEAP), Plantas Elevadoras de Agua Servidas (PEAS), Plantas Tratamiento de Aguas Servidas (PTAS), así como partida, parada de bombas, caudales, presiones, etc.

Para Agua Potable en las localidades La Serena, Coquimbo se controla y monitorea utilizando SCADA, través de la unidad de CIO (centro de Inteligencia operacional), compuesta por un equipo de personas disponibles las 24 hrs del día. El nivel de cobertura, recintos monitoreados de Agua Potable corresponde al 93%.

El sistema de comunicación de telemetría que se despliega en la empresa Aguas del Valle, cuenta con infraestructura de comunicación propia que se compone de radios de banda de 900MHz/400Mhz/5GHz u 4G dependiendo de la situación geográfica, que interconectan los distintos recintos operativos de la empresa aprovechando la infraestructura de la compañía (por ejemplo estanques o plantas de producción). Estos radios enlaces se van conectando a la red de

[iii] https://www.aguasdelvalle.cl/empresas/nosotros/plan-de-desarrollo-vigente/zonal-elqui/

[iv] Plan de Desarrollo: documento público que establece las inversiones y obras que se realizarán para prestar servicios de agua potable y alcantarillado.

comunicación hasta llegar a un recinto que cuente con conexión a la MPLS (Multiprotocol Label Switching), en los cuales se ha tomado medidas para generar sistemas de respaldo mediante soluciones Satelitales, estos servicios son prestados por una empresa de telecomunicaciones (TELCO).

El sistema SCADA actual (Xylem Vue - SmartScada), tiene como objetivo asegurar la continuidad y calidad del servicio, mejora continua, resultados operacionales y disponibilidad de información para la toma de decisiones. La plataforma SmartSCADA se compone de arquitectura en la plataforma Microsoft Azure[v], dentro de los cuales encontramos servidores de servicios.

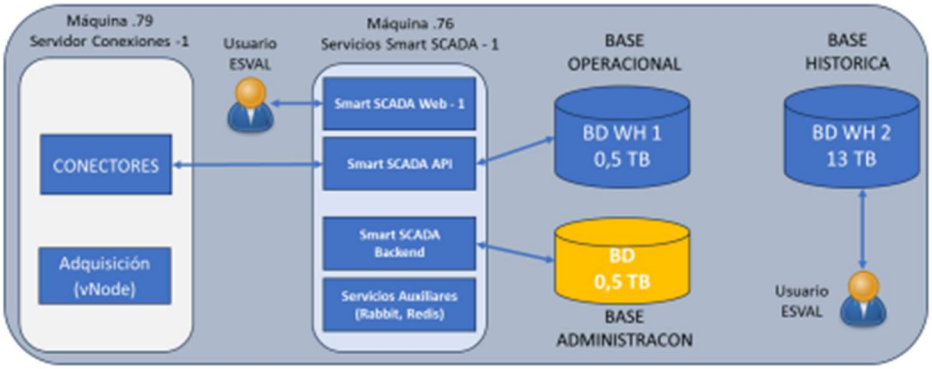


Ilustración 2 Esquema Servidores SMARTSCADA

La plataforma HA o máquinas de respaldo son una copia, la BBDD Admin (base de dato administrativa) es sincronizada a la Zona redundante, en caso de que la plataforma principal fallara, automáticamente se realiza el levantamiento de la plataforma HA o máquinas de respaldo.

[v] Microsoft Azure es un conjunto de servicios en la nube desarrollado por Microsoft para ayudar a los clientes a crear, implementar y gestionar aplicaciones en la nube

Los Operadores del Centro de Inteligencia Operacional se conectan a la aplicación mediante su explorador web, https://smartscada.esval.cl/Portal/LoadPortal. En el siguiente esquema se observa un ejemplo de una visualización del sistema SCADA:

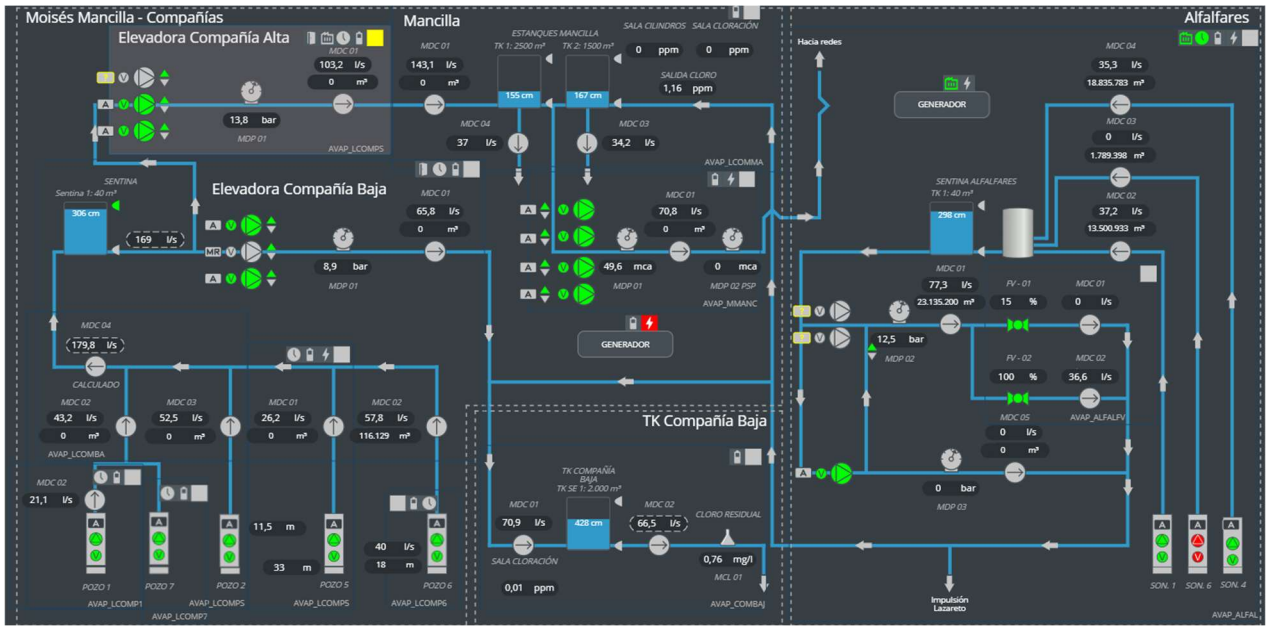


Ilustración 3 Herramienta SMARTSCADA

Las principales ventajas de estos sistemas son:

- Permiten monitorear en tiempo real variables clave como flujos, presiones y calidad del agua en diferentes puntos de la red.
- Permiten operar remotamente elementos de control como bombas, válvulas, etc.
- Detección y localización de fugas y eventos de calidad del agua.
- Validación y reconstrucción de datos de sensores.
- Diagnóstico de fallos en sensores y actuadores.
- Combinar actividades operacionales y de gestión empresarial.
- Aprovechar sinergias para mejorar eficiencia, costos y calidad del servicio.

En la siguiente tabla se resume la cantidad de información desplegada en el sistema de SCADA, asociada a la infraestructura de Agua Potable (producción y distribución) de las localidades de La Serena y Coquimbo:

Tabla 1 Cantidad de sensores nivel y caudal instalados en La Serena y Coquimbo

Tipo	Cantidad
Sensores de Nivel	35
Sensores de Caudal	50

Fuente: Aguas del Valle

2.2.7 Descripción de la Gestión Operativa

El control operacional del sistema se realiza principalmente a través de tres modalidades:

2.2.7.1 Operación Local

Se gestiona principalmente la captación desde el Río Elqui, única fuente superficial del sistema, ubicada aguas abajo del embalse Puclaro. Esta fuente tiene el mayor impacto económico debido a:

- Costos de compra de agua (tarifa que Aguas del Valle debe cancelar a Junta de Vigilancia del Rio Elqui).
- Restricciones de la Junta de Vigilancia del Río Elqui (limitaciones de oferta).
- Otros impactos operacionales (debido a que este sistema posee más del 50% del volumen necesario para cubrir la demanda).

El control y operación se realiza localmente en el Sistema Productivo de Agua Potable (SPAP) Las Rojas, mediante operadores en terreno.

2.2.7.2 Sistema de Producción de Agua Potable Las Rojas.

El Sistema de Producción de Agua Potable (SPAP) Las Rojas constituye la principal instalación de producción del sistema, utilizando principalmente agua superficial proveniente del río Elqui. En

su máxima capacidad operativa, la planta de filtros convencional puede procesar hasta 650 l/s, complementándose con una red de ocho fuentes subterráneas que aportan conjuntamente 60 l/s adicionales. Adicionalmente, cuando las condiciones de calidad del río Elqui lo permiten, es posible aumentar la capacidad productiva mediante la recarga artificial del "Dren Las Rojas", un filtro lento que puede aportar hasta 250 l/s extras.

En respuesta a los eventos críticos de turbiedad extrema ocurridos en el Río Elqui durante 2015 y 2017, que provocaron interrupciones masivas del servicio, se implementó en 2018 un tranque de respaldo que entró en operación en 2019. Esta infraestructura, con una capacidad de almacenamiento de 47.000 m³ de agua cruda, garantiza una autonomía operacional de 28 horas a un flujo de 420 L/s, sirviendo como soporte fundamental para la PTAP Las Rojas. El diseño del tranque permite una operación a máxima capacidad de 650 L/s, manteniendo una autonomía de 20 horas bajo estas condiciones. A continuación, se presenta un esquema georreferenciado de la infraestructura del sistema productivo:



Ilustración 4 Esquema SPAP Las Rojas



Ilustración 5 Filtros SPAP Las Rojas

2.2.7.3 Operación de Fuentes Subterráneas

Comprende 33 pozos con sistemas automatizados**.** El monitoreo y control se realiza mediante el Sistema de telemetría y Sistema SCADA y su operación se centraliza en el Centro de Inteligencia Operacional, ubicado en Viña del Mar (sede central de ESVAL y Aguas Del Valle).

Se prioriza mantener estas fuentes operativas por sus menores costos de operación, y solo su detención responde a variación de parámetros de calidad, fallas en los sistemas de bombeo o mantenciones programadas de la infraestructura.

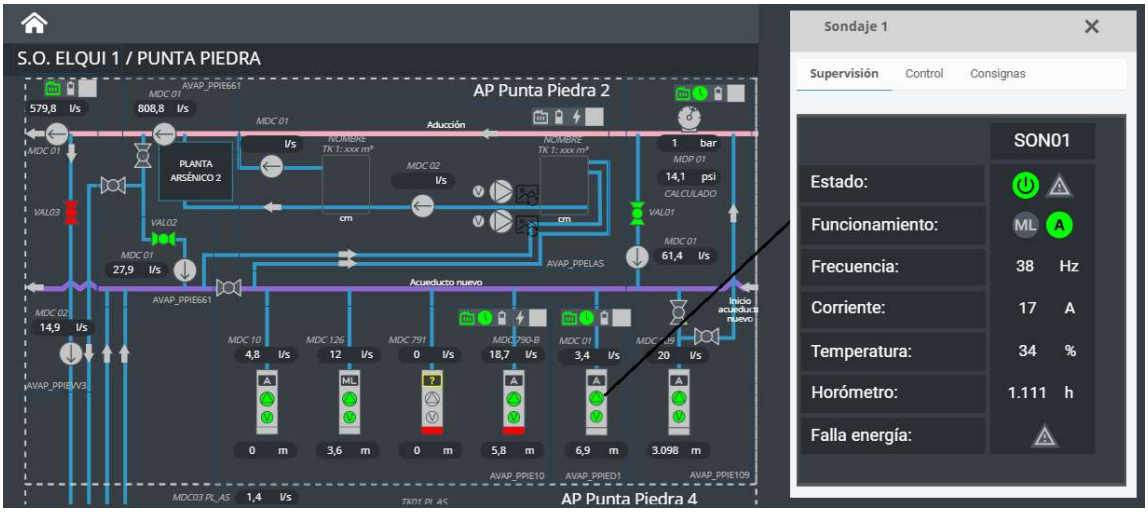


Ilustración 6 Pantalla SCADA y Telecontrol de Sondaje

2.2.7.4 Sistema de Distribución

2.2.7.4.1 a) Macroconducciones y Transferencias

Corresponde a las conducciones utilizada para transporte de agua desde fuentes productiva hasta estanques. Su cobertura abarca todo territorio operacional de Aguas del Valle (La Serena y Coquimbo), con una extensión de más de 90 kilómetros, con una operación principalmente gravitacional, tanto en régimen de canal o presurizado. Entre las principales conducciones destacan:

- Aducción Las Rojas: Construida en 1984, esta aducción es la principal infraestructura de distribución de agua potable en el sistema, la "aducción" transporta el caudal proveniente del sistema productivo Las Rojas" ubicado camino al Valle del Elqui, a 21 (km) de la ciudad de La Serena. Las características técnicas de esta conducción son:
    - Longitud total: 32,3 kilometros metros
    - Diámetros: entre 400 mm y 900 mm
    - Materiales: Combinación de tuberías de acero, asbesto cemento y PVC de diferentes clases

- Tramo principal: Desde la Cámara de Carga (SPAP Las Rojas) hasta el recinto de estanque La Herradura Baja, Coquimbo.
- Centros de consumo abastecidos:
    1. La Florida
    2. Pedro Aguirre Cerda
    3. San Joaquín
    4. Cerro Grande
    5. Edmundo Gálvez (Tierras Blancas)
    6. Peñuelas
    7. Julio González (San Juan) – Punta Mira
    8. Balmaceda
    9. Herradura Baja

Esta estructura es fundamental para el sistema de distribución de agua potable de la zona, garantizando el suministro directo a ocho importantes centros poblados. En las siguientes imágenes, se presenta el trazado general de la aducción en Google Earth, dividida en dos tramos, “Aducción tramo Las Rojas – Lazareto” y “Aducción - Tramo Lazareto - Balmaceda:



Ilustración 7 Aducción tramo Las Rojas – Lazareto



Ilustración 8 Aducción - Tramo Lazareto - Balmaceda

- Acueducto Antiguo “Antiguo Punta Piedra – La Serena”: comienza en el Dren Antiguo Punta Piedra ubicado a un costado del rio “El Elqui”, en la localidad de Punta Piedra ubicada a unos 25 (km) al oeste de la ciudad de La Serena y termina su recorrido en el recinto Lazareto donde se encuentran los estanques de distribución Lazareto 1 y 2 de 200 (m3) cada uno. En su recorrido de 15,2 (km) existía una derivación al estanque Algarrobito con capacidad volumétrica de 200 (m3) que abastece la localidad del mismo nombre. Este acueducto tiene un diámetro de 600 (mm) en casi su totalidad con materialidad de cemento comprimido. Existen tramos de acero (generalmente en atraviesos) y reparaciones con materialidad de HDPE. El acueducto obtiene sus recursos desde un dren llamado “Antiguo Punta Piedra”. En la actualidad, el dren tiene un caudal de explotación de 70 (l/s). A continuación, el trazado general del acueducto en Google Earth:



Ilustración 9 Acueducto Antiguo

- Acueducto Nuevo: Este acueducto, construido en 1960, transporta agua desde los sondajes Punta de Piedra hasta los estanques de acumulación del recinto Lazareto. Con una extensión de 13.690 metros de tubería de Hormigón de 700 mm de diámetro, abastece a los recintos de Santa Lucía, Compañía Baja y Balmaceda. a continuación, el trazado general del acueducto en Google Earth:

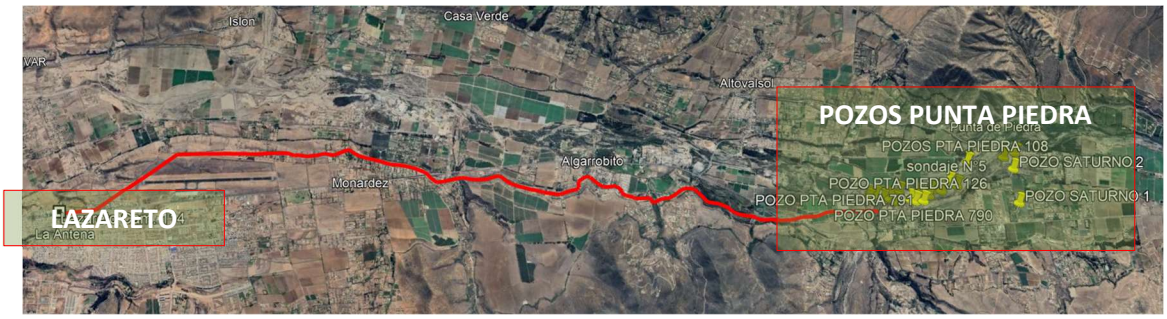


Ilustración 10 Acueducto Nuevo

- Acueducto Lazareto Coquimbo: Esta infraestructura constituye la extensión de los Acueductos Antiguo y Nuevo, iniciando su recorrido en los estanques de acumulación del recinto Lazareto. Su infraestructura se compone de:
    - 8,572 metros de tubería de Acero
    - 8,947 metros de tubería de Asbesto Cemento
    - 246 metros de tubería de Hormigón Totalizando 17,765 metros de longitud, con diámetros que varían entre 450 mm y 700 mm.

  Su función principal es abastecer al recinto de estanques Balmaceda en Coquimbo. Además, sirve como sistema alternativo de emergencia para:

- El resto del sistema en caso de fallas en la Aducción
- Los recintos de estanques de Tierras Blancas y Peñuelas, que normalmente se abastecen desde la Aducción La Serena – Coquimbo, mediante conexión directa o plantas re-elevadoras.

A continuación, el trazado general del acueducto en Google Earth:

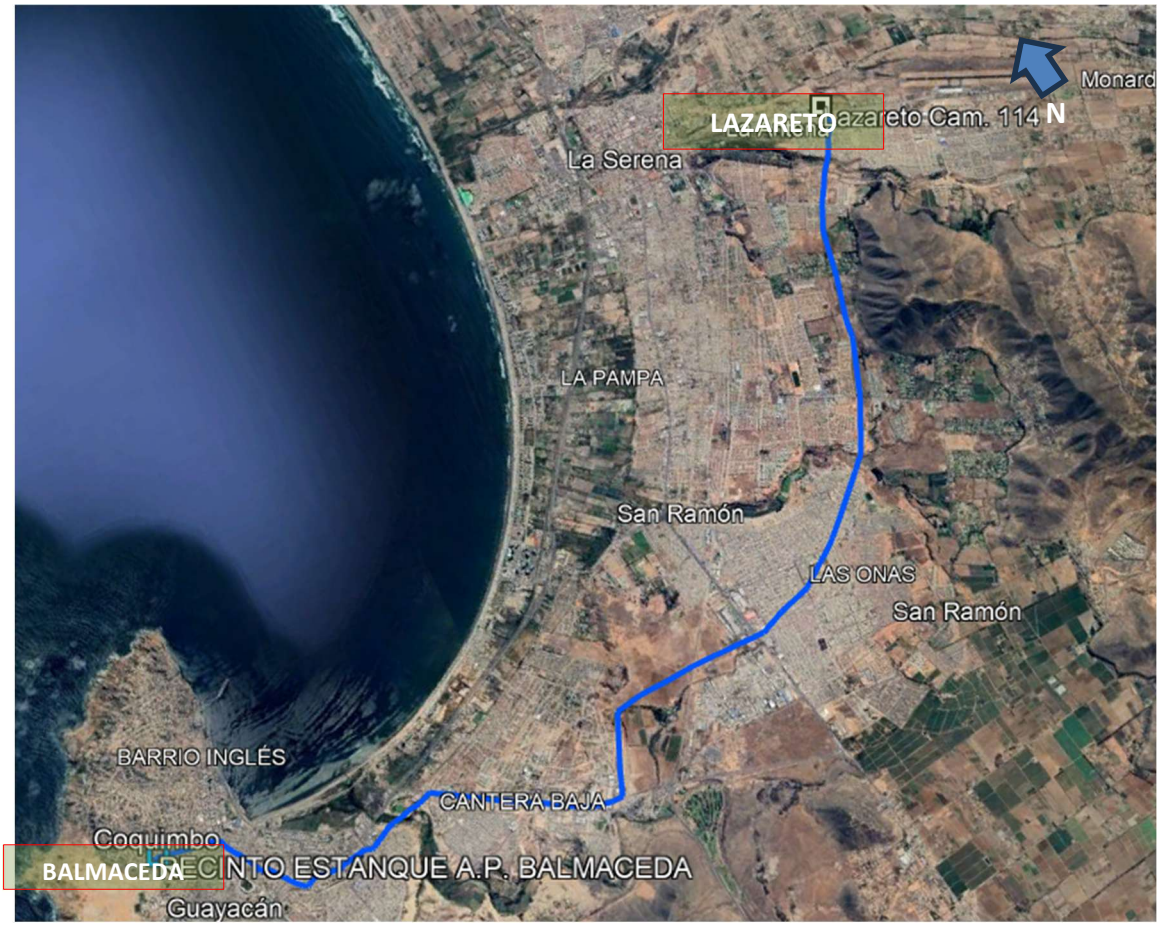


Ilustración 11 Acueducto Lazareto – Coquimbo

2.2.7.4.2 Estanques de Regulación

El sistema de distribución posee 38 estanques, con volumen de regulación total de 55000 m³. El Control de llenado mediante válvulas automatizadas o estaciones elevadoras, las cuales son monitoreadas y controladas remotamente a través de telemetría y sistema SMARTSCADA.

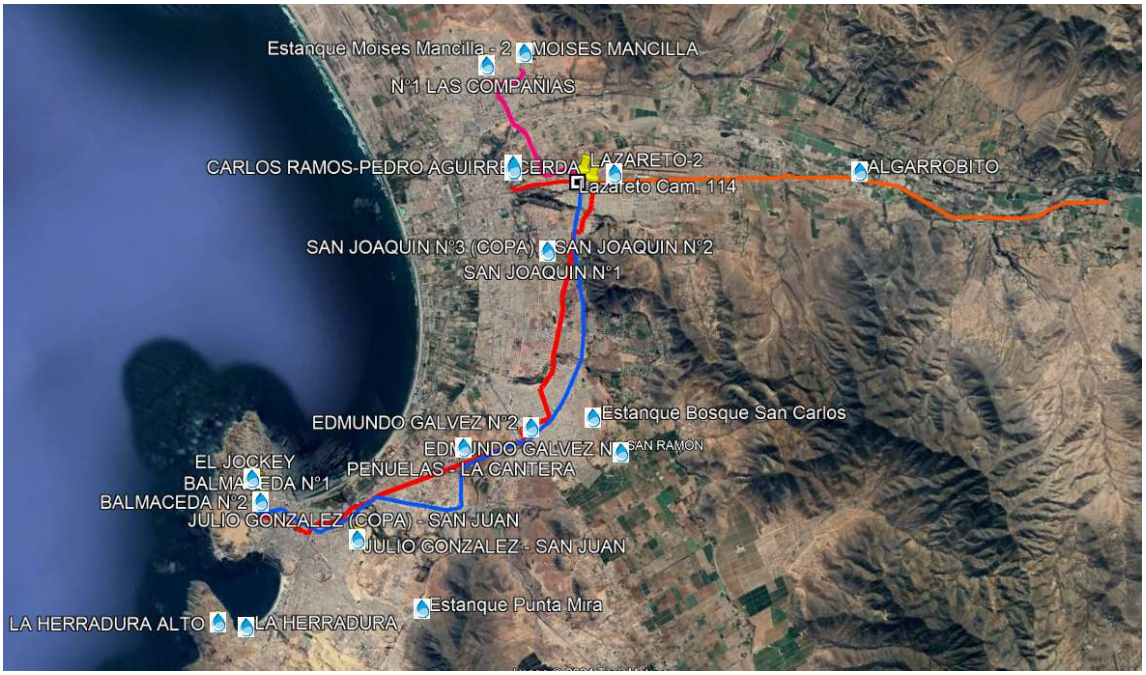


Ilustración 12 Ubicación Estanques de Regulación



Ilustración 13 Estanque Semi Enterrado San Ramón

2.3 Fundamentos de PCA

2.3.1 Fundamento teórico del PCA

El PCA (Análisis de componentes principales) se basa en la descomposición de la matriz de covarianza de los datos. Dados unos datos centrados $X \in \mathbb{R}^{n \times p}$ (con n observaciones y p variables), se calcula la matriz de covarianza:

$$\sum = \frac{1}{n-1} X^T X$$

La descomposición espectral de Σ permite obtener los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_p$ y los correspondientes vectores propios $v_1, v_2, \ldots, v_p$. Los vectores propios determinan las direcciones en el espacio de las variables en las cuales la varianza de los datos es máxima, y cada valor propio indica la cantidad de varianza explicada por la componente principal asociada. La proporción de varianza explicada por la i-ésima componente se puede expresar como:

$$Varianza\ Explicada_i = \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}$$

La selección de un número adecuado de componentes se realiza generalmente mediante un criterio de varianza acumulada, en este caso, se opta por retener aquellas componentes que, en conjunto, expliquen al menos el 90% de la varianza total.

2.3.2 Selección de componentes y justificación

La elección de retener componentes que expliquen el 90% de la varianza responde a un compromiso entre simplicidad y fidelidad de la representación. En términos prácticos:

- **Reducción del ruido:** Al eliminar componentes que representan poca varianza, se filtra el ruido inherente a los datos.

- **Eficiencia computacional:** Se reduce la complejidad del modelo, lo que es esencial para la calibración y proyección en tiempo real.
- **Mantenimiento de la información:** El umbral del 90% asegura que la mayor parte de la variabilidad y los patrones subyacentes se mantengan intactos.

Técnicamente, la decisión se fundamenta en la comparación entre la varianza explicada acumulada y el umbral establecido:

$$\text{Número de componentes k tal que } \sum_{i=1}^{k} \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j} \geq 0.90$$

Esta metodología se aplica tanto a los datos, en caso de este estudio a las variables de caudal como a cada una de las variables meteorológicas, y los scores resultantes (la proyección de los datos originales en el espacio de las componentes principales) son posteriormente utilizados en el modelado ARIMAX para la proyección del consumo de agua.

2.4 Modelado Predictivo con ARIMAX

El modelo ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variables) es una extensión del clásico modelo ARIMA que permite la inclusión de variables exógenas, lo que lo hace especialmente adecuado para contextos en los que factores externos inciden en la serie temporal. En este caso, se utiliza ARIMAX para modelar y proyectar el consumo de agua, integrando las componentes principales obtenidas del PCA tanto de los datos de caudal como de las variables meteorológicas

2.4.1 Descripción del Modelo ARIMAX

El modelo ARIMAX combina tres componentes fundamentales:

- Autoregresivo (AR): Representa la dependencia de la serie actual respecto de sus valores pasados. Matemáticamente, se expresa como:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \epsilon_t,$$

donde φ_i son los coeficientes autoregresivos y ϵ_t es el error en el tiempo t.

- Integrado (I): Hace referencia al proceso de diferenciación para estabilizar la media de la serie. Si la serie es no estacionaria, se diferencia d veces para obtener una serie estacionaria.
- Media Móvil (MA): Modela el error de la serie como una combinación lineal de errores pasados:

$$\epsilon_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \eta_t,$$

donde θi son los coeficientes de la parte de media móvil y η_t es el término de ruido blanco.

- Variables Exógenas (X): En el modelo ARIMAX se incorporan predictores externos. La formulación del modelo se extiende a:

$$y_t = \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \beta' X_t + \vartheta_1 \epsilon_{t-1} + \cdots + \vartheta_q \epsilon_{t-q} + \eta_t,$$

donde X_t representa el vector de variables exógenas en el tiempo t y β los coeficientes asociados. En nuestro caso, las variables exógenas corresponden a los puntajes obtenidos del PCA aplicado a las variables meteorológicas.

2.4.2 Ajuste del modelo utilizando componentes principales

Para el modelado del caudal, la serie temporal original se reestructura en una matriz diaria, a la cual se le aplica PCA para extraer los patrones dominantes, tal como se mencionó anteriormente. En el proceso de ajuste del ARIMAX, cada una de las componentes principales derivadas de la matriz del caudal se modela individualmente. Los scores de PCA de la serie de caudal se consideran como la variable dependiente, mientras que los scores correspondientes a las variables meteorológicas (obtenidos a partir del PCA aplicado a cada variable meteorológica) se incorporan como predictores exógenos. Esta estrategia permite que el modelo capture no solo la

dinámica intrínseca del caudal, sino también el impacto de las condiciones meteorológicas en el consumo de agua.

Una formulación matemática típica del modelo ARIMAX, que integra tanto la serie de caudal como las variables meteorológicas (a través de sus scores obtenidos del PCA), es la siguiente:

$$\varphi(B)(1 - B)^d y_t = c + \theta(B)\epsilon_t + \sum_{k=1}^{K} \beta_k X_{k,t} + \epsilon_t$$

donde:

- y_t representa la variable de caudal (consumo de agua) en el tiempo t (o, en nuestro caso, podría ser uno de los scores derivados del PCA de la serie de caudal).
- B es el operador de retroceso (backshift), tal que $B_{yt} = y_{t-1}$.
- $(1 - B)^d$ es el operador de diferenciación de orden d , que se utiliza para estabilizar la media de la serie (integración).
- $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$ representa la parte autoregresiva (AR) del modelo.
- $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$ representa la parte de media móvil (MA).
- c es la constante del modelo.
- $X_{k,t}$ denota el k-ésimo componente exógeno en el tiempo t , derivado de los scores del PCA aplicado a las variables meteorológicas (por ejemplo, temperatura, lluvia y cobertura de nubes), y K es el número de componentes retenidas.
- β_k es el coeficiente asociado a la k-ésima variable exógena.
- ϵ_t es el término de error (ruido blanco) en el tiempo t.

Esta ecuación resume la idea central del ARIMAX: la serie y_t se modela mediante su propia dinámica (a través de los términos autoregresivos y de media móvil), junto con el efecto de factores externos $X_{k,t}$ que se supone influyen en el comportamiento del consumo de agua.

2.4.3 Mecanismo de reutilización del modelo para nuevas proyecciones

Un aspecto clave en el enfoque implementado es la capacidad de reutilizar y actualizar los modelos ARIMAX sin necesidad de recalibrarlos completamente en cada ejecución, con el fin de reducir el costo computacional. Para ello, se implementa el siguiente mecanismo:

- Guardado del modelo: En la primera ejecución, se utiliza la función auto.arima() para seleccionar automáticamente la mejor estructura (valores de p, d y q) y se ajusta el modelo ARIMAX. El modelo ajustado se guarda en un archivo (por ejemplo, en formato .rds).
- Actualización del modelo: En ejecuciones posteriores, el código verifica la existencia del modelo guardado. Si el modelo ya existe, se carga y se actualiza utilizando la nueva data de entrenamiento, trabajando en el mismo espacio, con las mismas transformaciones, manteniendo la estructura identificada inicialmente. Esto garantiza consistencia y reduce el tiempo computacional requerido para el ajuste.
- Proyección para el horizonte deseado: Una vez actualizado, el modelo se utiliza para generar proyecciones a 7 días, incorporando las variables exógenas futuras (los scores derivados del PCA de las variables meteorológicas para el periodo de proyección).

Este mecanismo de reutilización del modelo es especialmente valioso en entornos operativos, ya que permite mantener un modelo actualizado y eficiente sin la sobrecarga de un reentrenamiento completo cada vez que se dispone de nuevos datos.

En resumen, el uso de ARIMAX en este contexto permite:

- Integrar información histórica del consumo de agua y su dinámica intradía.
- Incorporar el impacto de variables meteorológicas como predictores exógenos, lo que mejora la precisión de la proyección.
- Reutilizar y actualizar el modelo de forma eficiente, garantizando que las proyecciones futuras se beneficien de los modelos ya calibrados.

2.5 Proceso de proyección y reconstrucción de la serie original

La fase final del proceso consiste en la generación de predicciones a 7 días y la reconstrucción de la serie original a partir de las componentes principales. En este contexto, se asume que los datos meteorológicos futuros están disponibles, lo cual permite incorporar los efectos exógenos en el modelo ARIMAX para mejorar la precisión de las proyecciones.

2.5.1 Generación de predicciones a 7 días

Una vez ajustado el modelo ARIMAX sobre las componentes principales derivadas del PCA, se procede a generar predicciones para cada una de estas componentes a lo largo de un horizonte de 7 días. Sea $y_t^{(i)}$ el score de la i -ésima componente en el tiempo t , el modelo ARIMAX genera las predicciones:

$$\hat{y}_{t+h}^{(i)} \; para \; h = 1,2, \ldots \ldots . H$$

donde H equivale al número total de períodos en 7 días (por ejemplo, 7 días × 24 horas). En este proceso se utiliza la información exógena, es decir, los scores correspondientes a las variables meteorológicas proyectadas para el mismo horizonte, lo que permite que el modelo capte la influencia de las condiciones climáticas futuras sobre el consumo de agua.

2.5.2 Reconstrucción de la serie original

Para recuperar la señal en el espacio original, se utiliza la matriz de pesos P obtenida del PCA. Si la matriz de pesos es $\mathbf{P} \in \mathbf{R}^{p \times k}$, donde p es la dimensión original (por ejemplo, 24 horas) y k es el número de componentes retenidas, la reconstrucción de la serie se realiza mediante:

$$\hat{y}_{t+h} = \sum_{i=1}^{k} \hat{y}_{t+h}^{(i)} p_i$$

donde pi representa el i-ésimo vector de carga.

Es importante aplicar las transformaciones inversas de escalamiento y centrado utilizadas en el PCA para recuperar los valores en sus unidades originales. Esta reconstrucción permite transformar las predicciones del espacio reducido de PCA al dominio original, facilitando la interpretación y el uso operativo de los pronósticos.

2.5.3 Disponibilidad de datos meteorológicos futuros

El proceso de proyección se fundamenta en la hipótesis de que se dispone de datos meteorológicos para los próximos 7 días. Esta suposición es crítica, ya que:

- Permite calcular, de forma análoga a la preparación de los datos históricos, los scores de las variables meteorológicas (temperatura, lluvia y cobertura de nubes) para el horizonte de proyección.
- Asegura que el modelo ARIMAX pueda incorporar los efectos exógenos futuros, ajustando sus predicciones en función de las condiciones climáticas esperadas.

En esencia, se presupone que las condiciones meteorológicas futuras están disponibles o pueden ser estimadas con suficiente precisión, lo que facilita la integración de estas variables en el proceso de proyección.

3 Descripción de los Datos

Este capítulo describe las fuentes de datos y los pasos de preprocesamiento empleados para construir un modelo de previsión del consumo de agua en La Serena y Coquimbo. Los datos se obtienen de sistemas avanzados de monitoreo (SCADA) y plataformas meteorológicas(Open-Meteo), proporcionando una base robusta para el análisis y la predicción.

3.1 Datos de Caudal de Agua

Los datos de caudal de agua son esenciales para medir la demanda y establecer un historial que sustente las predicciones. Estos se recolectan mediante SmartSCADA, un sistema de monitoreo y control automatizado implementado por Aguas del Valle. Este sistema registra en tiempo real el caudal en puntos estratégicos de los sistemas productivos, macroconducciones y transferencias, red de distribución y estanques, destacando la captación superficial del río Elqui en el sector Las Rojas y los pozos subterráneos del acuífero Elqui Bajo.

- **Frecuencia y Período**: Los registros son horarios y abarcan desde el 1 de enero de 2022 hasta octubre de 2024. Este período, de aproximadamente 2 años y 10 meses, se eligió para evitar alteraciones en los patrones de consumo causadas por la pandemia de COVID-19, asegurando datos representativos.
- **Variable Objetivo**: El caudal de salida, medido en litros por segundo (l/s), refleja el consumo de agua en la zona. La demanda máxima diaria alcanza 1805 l/s, atendiendo a una población de más de 500.000 habitantes
- **Preprocesamiento**:
    - **Conversión Temporal**: Las fechas se transformaron al formato POSIXct con zona horaria UTC para mantener una secuencia temporal precisa y facilitar la integración con otras fuentes de datos.
    - **Filtrado**: Se seleccionaron únicamente días con 24 registros horarios completos, eliminando días incompletos para evitar sesgos en el análisis y la modelación.

Estos datos capturan la dinámica del consumo de agua y reflejan las necesidades operativas de Aguas del Valle para la gestión eficiente de los recursos hídricos.

3.2 Datos Meteorológicos

Las variables meteorológicas se incluyen como factores exógenos clave que influyen en el consumo de agua. Estos datos provienen de **Open-Meteo(https://open-meteo.com)**, una plataforma que ofrece información histórica y proyectada de alta resolución, permitiendo un análisis detallado de las condiciones climáticas.

- **Frecuencia y Período**: Los datos son horarios y cubren el mismo período que los datos de caudal (enero de 2022 a octubre de 2024), con proyecciones adicionales para los 7 días siguientes a la fecha actual.
- **Variables Consideradas**:
    - **Temperatura (°C)**: Afecta directamente la demanda, con un aumento notable en días cálidos.
    - **Precipitación (mm)**: Puede reducir la demanda al cubrir necesidades como el riego, aunque su efecto es limitado en esta región.
    - **Cobertura de Nubes (%)**: Influye indirectamente a través de su relación con la temperatura y la evaporación.
- **Preprocesamiento**:
    - **Conversión Temporal**: Las fechas se ajustaron a POSIXct en UTC, alineándolas con los datos de caudal.
    - **Selección Horaria**: Se priorizaron las horas de 9:00 a 17:00 para captar las condiciones climáticas más relevantes durante el período de mayor consumo diurno.
    - **Integración de Proyecciones**: Se incorporaron proyecciones meteorológicas de 7 días para anticipar el consumo futuro, fortaleciendo la capacidad predictiva del modelo.

La combinación de datos históricos y proyectados permite reflejar tanto las condiciones pasadas como las esperadas, mejorando la precisión de las predicciones.

3.3 Análisis Exploratorio

El análisis exploratorio examina patrones temporales, relaciones entre variables y estadísticas descriptivas, ofreciendo información valiosa para la selección de variables y la optimización del modelo.

3.3.1 Gráficas Descriptivas del Consumo de Agua en el Tiempo

- **Caudal Diario**: Un gráfico de líneas muestra una alta variabilidad en el caudal de salida, con valores entre 0 y 2500 l/s. Los picos ocasionales podrían estar relacionados con eventos climáticos extremos o fallas en la infraestructura de distribución.

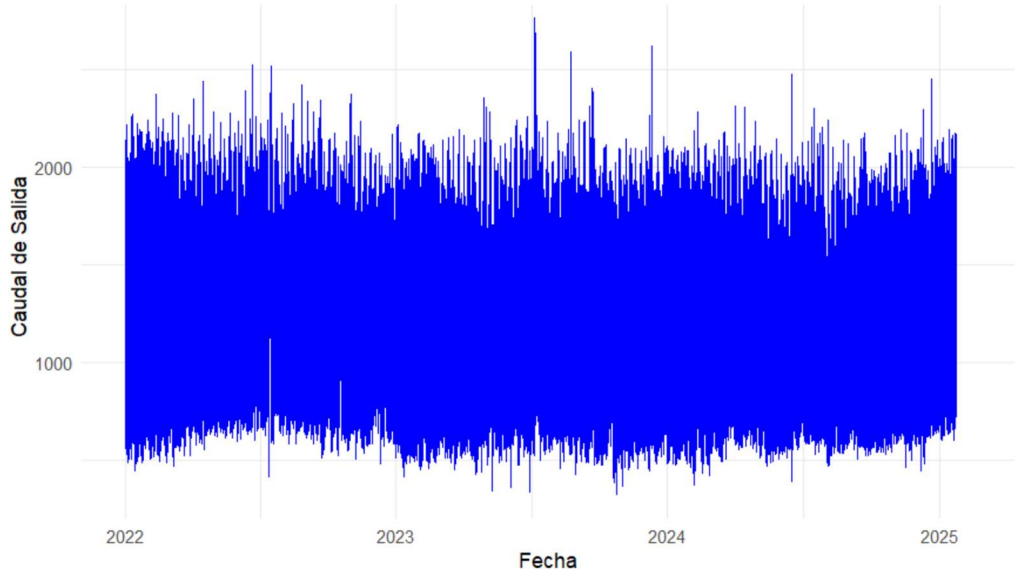


Gráfico 1 Caudal de salida diario en el Elqui

- **Caudal Semanal**: El promedio semanal, que oscila entre 1200 y 1400 l/s, suaviza las fluctuaciones diarias. Se observa una tendencia descendente hasta 2024, seguida de un repunte reciente, posiblemente ligado a cambios en la gestión hídrica o la demanda.

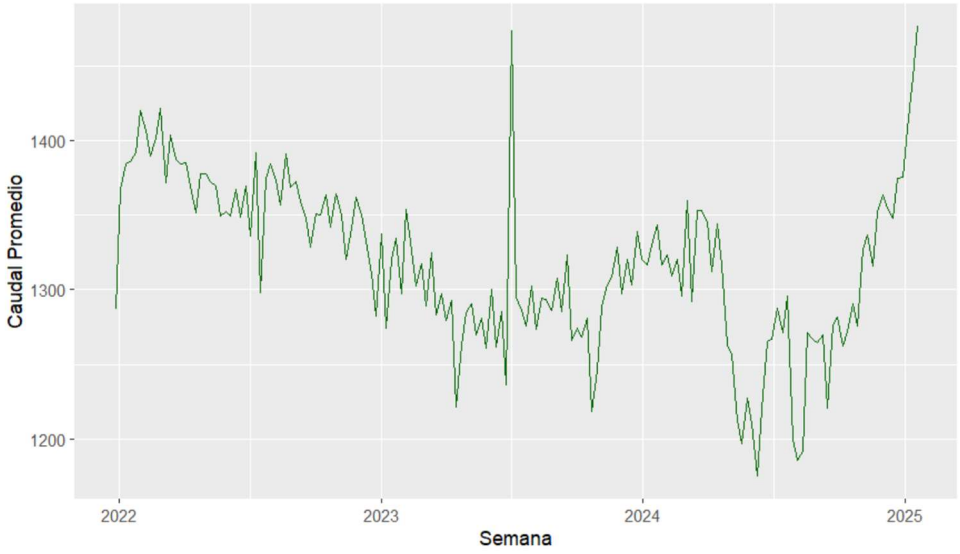


Gráfico 2 Caudal de salida semanal en La Serena - Coquimbo

3.3.2 Descomposición de la Serie Temporal

La descomposición del caudal separa sus componentes:

- **Tendencia**: Revela una disminución inicial, estabilización y un repunte final, reflejando la crisis hídrica y las medidas de mitigación en la región.
- **Estacionalidad**: Muestra un patrón constante, probablemente asociado a ciclos climáticos o variaciones estacionales en la demanda.
- **Residuo**: Presenta variabilidad constante con picos extremos, indicando eventos atípicos que deben considerarse en la modelación.

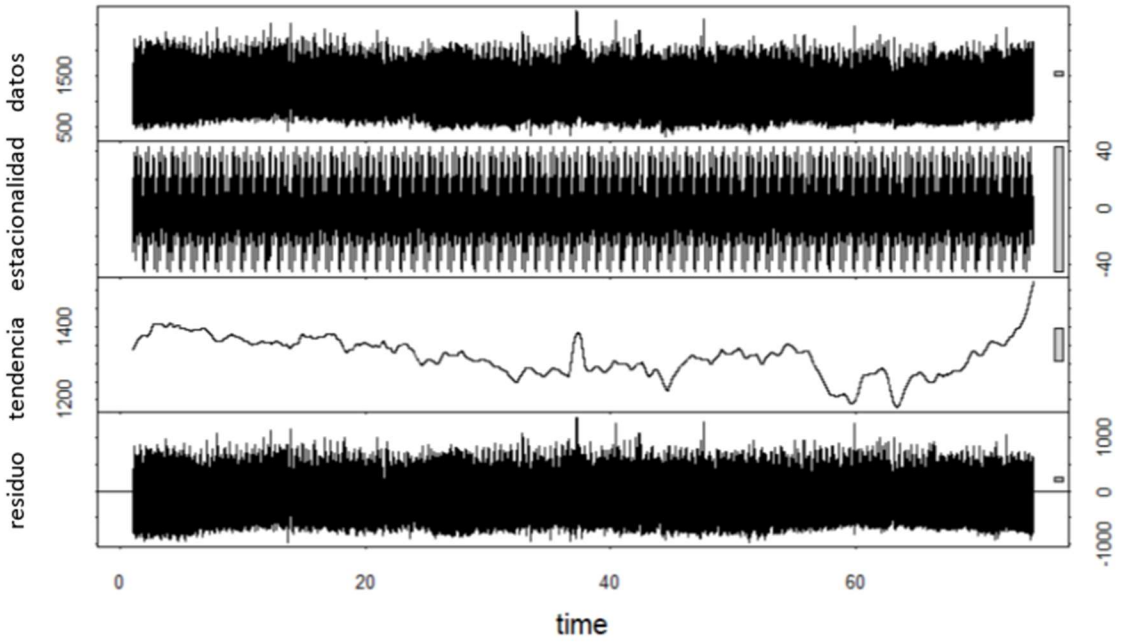


Gráfico 3 Descomposición de la Serie Temporal

3.3.3 Análisis de Correlación

La matriz de correlación de Pearson entre el caudal y las variables meteorológicas indica:

- **Temperatura**: Correlación positiva moderada (0.53), sugiriendo mayor consumo en días cálidos.

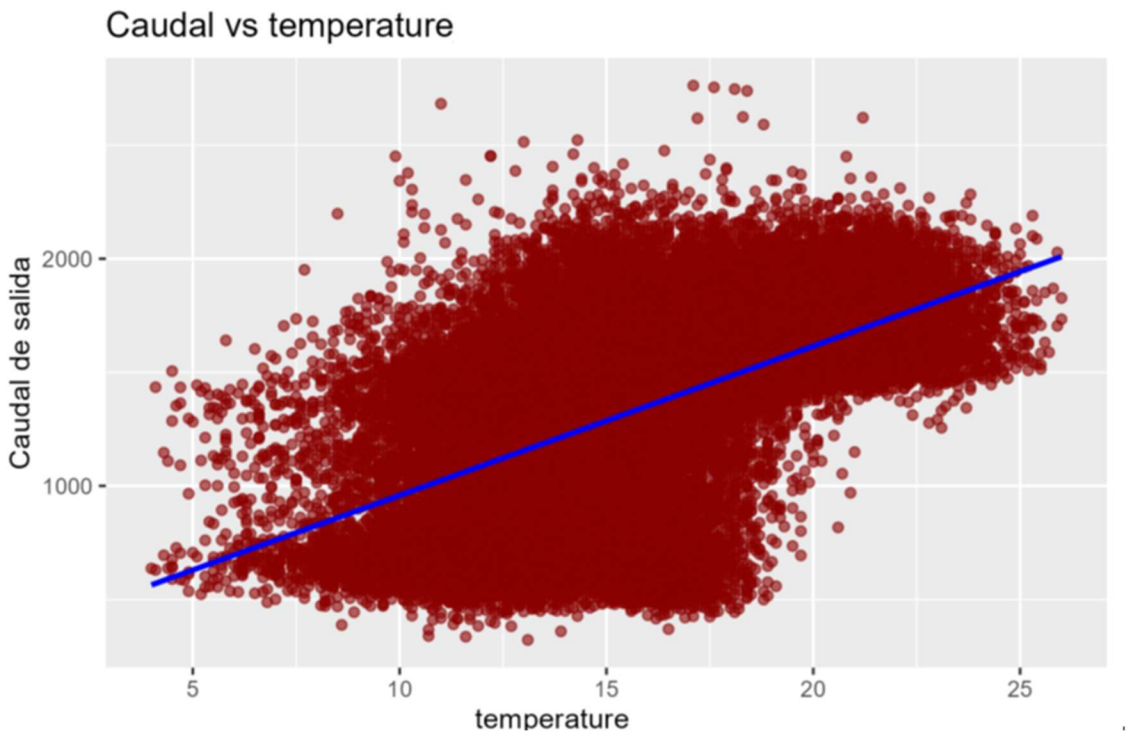


Gráfico 4 Correlación Caudal vs Temperatura

- **Precipitación**: Correlación nula (-0.01), debido a la baja frecuencia de lluvias en la región. Las lluvias no son lo suficientemente frecuentes para sustituir el uso de agua potable, a diferencia de regiones húmedas, Arbués et al. (2003).

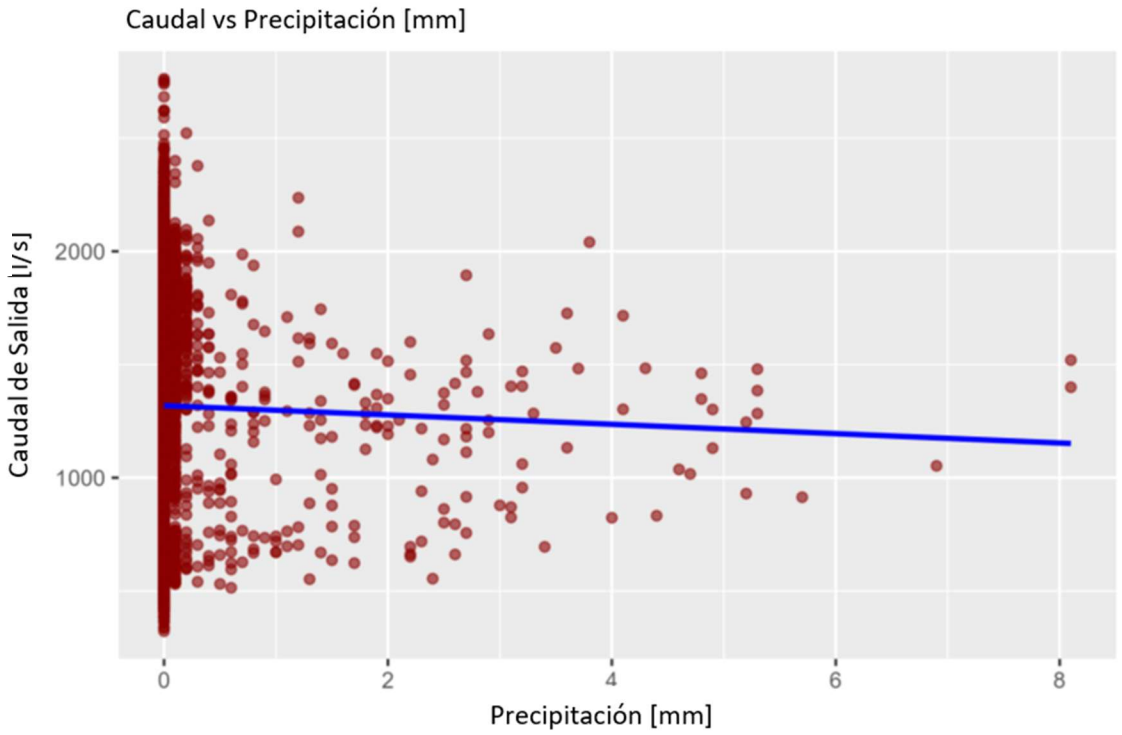


Gráfico 5 Correlación Caudal vs Precipitación

- **Cobertura de Nubes**: Correlación negativa débil (-0.05), relacionada indirectamente con la temperatura y la evaporación. La cobertura de nubes, al reducir la radiación solar, puede disminuir ligeramente la temperatura y la evaporación, pero su impacto en el consumo de agua es mínimo debido a patrones de uso dominados por necesidades urbanas y climáticas estables, correspondiente a regiones semiáridas (Abulibdeh et al. 2024).

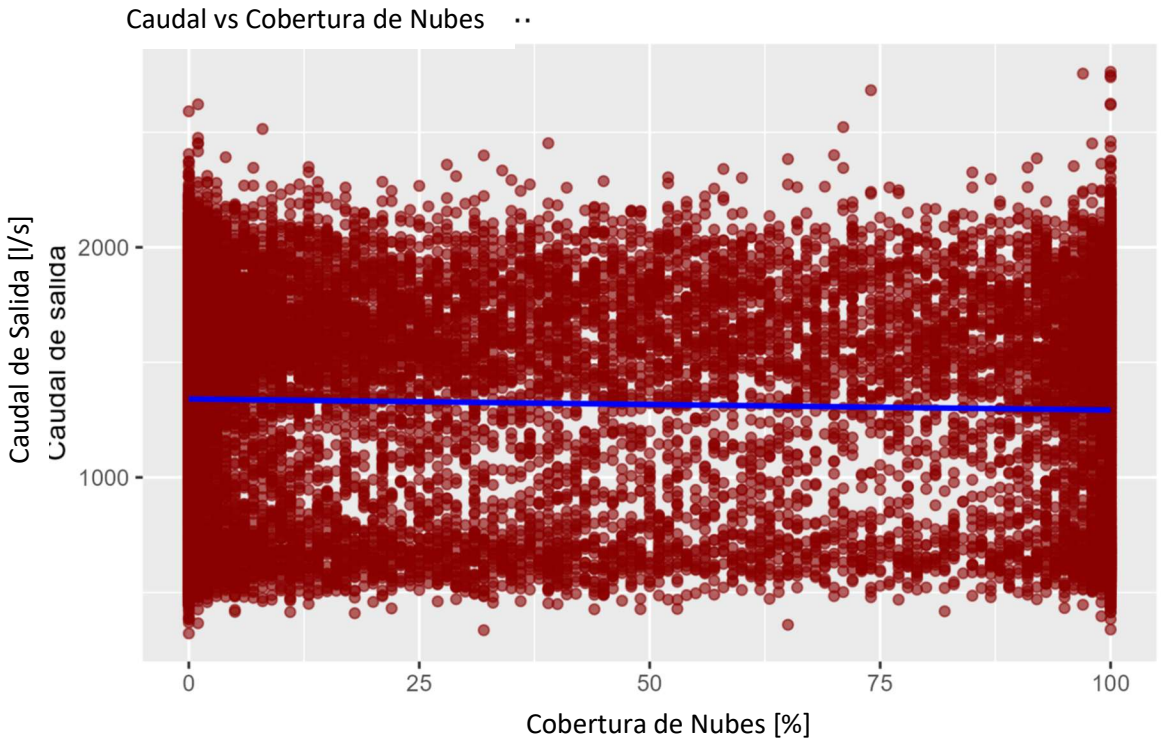


Gráfico 6 Correlación Caudal vs Cobertura de Nubes

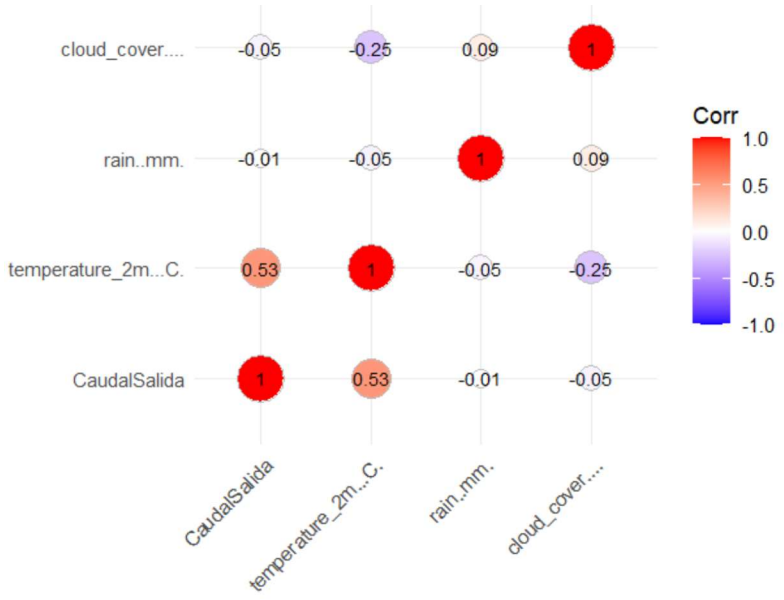


Gráfico 7 Matriz de Correlación de Pearson

- **3.3.4 Análisis Estadístico Descriptivo**

La siguiente tabla resume las estadísticas descriptivas de las variables clave:

Tabla 2 Resumen estadísticas descriptivas

Variable	Mínimo	1er Cuartil	Mediana	Media	3er Cuartil	Máximo
Caudal Salida (l/s)	321.8	849.2	1427.0	1318.4	1683.0	2762.6
Temperatura (°C)	4.0	12.7	15.4	15.55	18.2	26.0
Precipitación (mm)	0.0	0.0	0.0	0.0161	0.0	8.1
Cobertura Nubes (%)	0.0	0.0	33.0	46.42	100.0	100.0

- **Caudal de salida** : Alta variabilidad con una media de 1318.4 l/s, acorde con la capacidad operativa del sistema y la infraestructura.
- **Temperatura**: Distribución estable, típica del clima moderado de la región.
- **Precipitación**: Predominan días secos, con lluvias raras pero intensas, explicando su baja correlación con el consumo.

3.3.4 Estacionalidad

El consumo aumenta significativamente en verano (diciembre a febrero) debido a las altas temperaturas y disminuye en invierno (junio a agosto), lo que demuestra la correlación existente entre la temperatura y el caudal. Igualmente existe un factor de movilidad de la población debido a actividad turística de la zona, la cual aumenta el consumo en verano (diciembre a febrero).

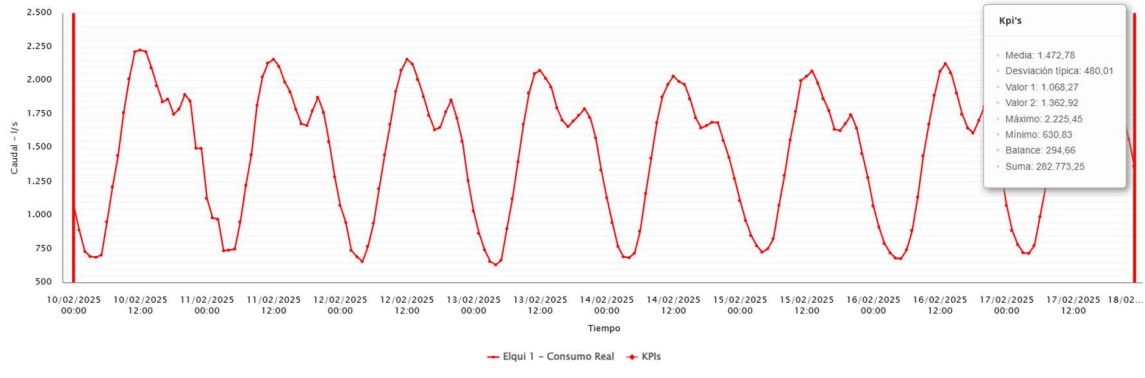


Ilustración 14 Caudal de Salida (Semana de 10-02-25 a 18-02-25)

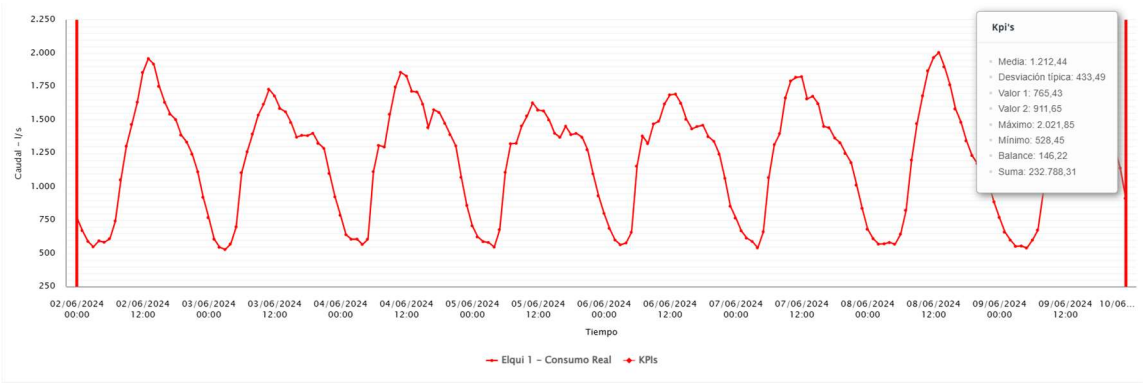


Ilustración 15 Caudal de Salida (Semana de 02-06-24 a 10-06-24)

Tabla 3 Resumen Comparación Estacional

Caudal de salida (l/s)	Media	Desviación	Máximo	Mínimo
Invierno	1212.4	433.5	2021.9	528.5
Verano	1472.8	480.0	2225.5	630.8
variación estacional [l/s]	260.3	46.5	203.6	102.4
variación estacional [%]	21.5%	10.7%	10.1%	19.4%

3.3.5 Eventos masivos de la población

Existen eventos o festivos, que provocan variaciones de la demanda producto de actividades masivas de la población, y por lo tanto, generan cambios en la operación habitual. Se destaca el día año nuevo, donde la variación de consumo es marcada entre el día 31-12-2024 y el día posterior 01-01-2025. En la siguiente grafica se observa el caudal y su cambio.

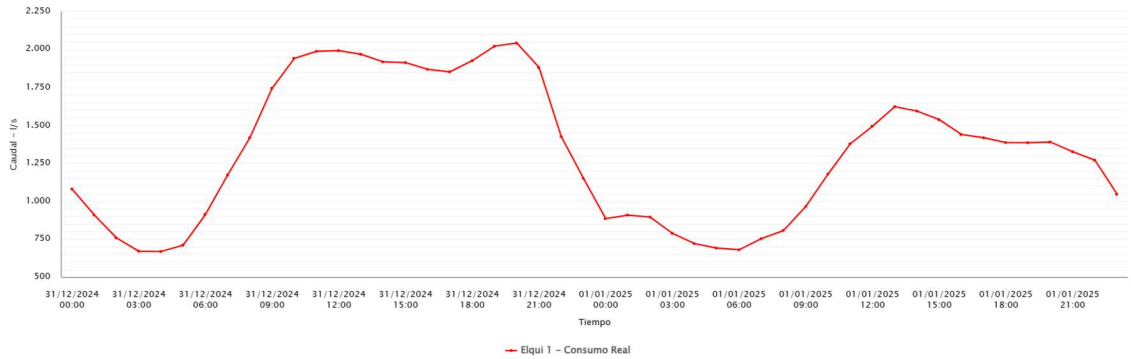


Ilustración 16 Caudal de Salida del 31-12-24 a 01-01-25

En la siguiente tabla se resume la variación destacando que los consumos máximos pueden disminuir de 2041.7 l/s a 1622.2 l/s, entre 31-12-24 y 01-01-25.

Tabla 4 Caudal de Consumo Año Nuevo

Caudal de salida (l/s)	Media	Dev. Std	máximo	mínimo
día 31-12-2024	1472.3	509.0	2041.7	667.9
día 01-01-2025	1147.8	314.4	1622.2	679.1
variación [l/s]	-324.6	-194.6	-419.5	11.2
variación [%]	-22.0%	-38.2%	-20.5%	1.7%

4 Metodología

Este capítulo explica el enfoque utilizado para predecir el consumo de agua, combinando técnicas avanzadas de reducción de dimensionalidad y modelado de series temporales para capturar patrones históricos e influencias meteorológicas.

4.1 Enfoque General

La metodología integra el Análisis de Componentes Principales (PCA) para reducir la dimensionalidad de los datos y el modelo ARIMAX para modelar la serie temporal del consumo, incluyendo variables meteorológicas como factores exógenos. Este enfoque permite prever la demanda a corto plazo (7 días), optimizando la gestión hídrica en un contexto de escasez. El modelo ARIMAX y Análisis de Componentes Principales (PCA) es resultado de un análisis previo aplicado a otro sistema, de similares características que arrojó resultados.

4.2 Justificación de Análisis de Componentes Principales (PCA)

La justificación del PCA (Análisis de Componentes Principales) en este estudio se basa en su capacidad para simplificar datos complejos y mejorar la predicción del consumo de agua. Esta técnica reduce las numerosas mediciones horarias de caudal (24 por día) a un pequeño conjunto de componentes que resumen los patrones principales de consumo, haciendo más manejable el análisis de series temporales. Asimismo, al aplicarlo a variables meteorológicas como temperatura, precipitación y cobertura de nubes, elimina redundancias y correlaciones entre ellas, creando variables más útiles para el modelo predictivo ARIMAX. Con esto, el PCA no solo agiliza los cálculos, sino que también evita problemas como el sobreajuste, logrando predicciones más confiables y eficientes para la gestión del agua en La Serena y Coquimbo.

4.3 Recopilación y Preprocesamiento de Datos

La metodología en detalle se encuentra en los anexos “9.19.1 POSIXct

POSIXct es una clase en el lenguaje de programación R diseñada para representar fechas y horas como un valor numérico continuo, específicamente el número de segundos transcurridos desde el 1 de enero de 1970, 00:00:00 UTC, conocida como la "época Unix" (Hyndman & Athanasopoulos, 2021). Según Hyndman y Athanasopoulos, esta clase forma parte del sistema POSIX y es una de las dos formas principales de manejar datos temporales en R, junto con POSIXlt. En POSIXct, el "ct" significa "calendar time" (tiempo calendario), lo que indica que almacena la información temporal en un formato compacto y eficiente, ideal para cálculos y almacenamiento en análisis de series temporales.

En el contexto del trabajo, POSIXct se utiliza para convertir y gestionar las fechas de los datos de caudal (SmartSCADA) y meteorológicos (Open-Meteo). Hyndman y Athanasopoulos (2021) destacan que este formato es especialmente útil para trabajar con series temporales porque permite especificar una zona horaria (como UTC) y asegura que las observaciones estén ordenadas cronológicamente, lo que es esencial para modelar datos horarios como los del consumo de agua. Por ejemplo, al transformar una fecha como "2022-01-01 00:00:00" a POSIXct con as.POSIXct(), R la convierte en un valor numérico (e.g., 1640995200 segundos desde la época), facilitando operaciones como diferencias temporales o sincronización entre conjuntos de datos.

A diferencia de POSIXlt, que descompone las fechas en una lista de componentes (año, mes, día, etc.), POSIXct ofrece una representación más eficiente para grandes volúmenes de datos, como los registros horarios de la tesis (Hyndman & Athanasopoulos, 2021). Esta eficiencia es clave para el preprocesamiento descrito en el estudio, donde se requiere alinear datos de múltiples fuentes y filtrar días completos de 24 horas, garantizando la consistencia temporal necesaria para el modelo ARIMAX.

Preparación de los Datos". A continuación, se resumen los principales pasos:

- **Fuentes de Datos**:
    - **Caudal de Agua**: Registrado por el sistema SmartSCADA de Aguas del Valle.

- **Datos Meteorológicos**: Obtenidos de la plataforma Open-Meteo.

- **Preprocesamiento**:
  - **Alineación Temporal**: Todas las fechas se convirtieron a POSIXct en UTC para consistencia.
  - **Filtrado de Caudal**: Se seleccionaron días con 24 registros horarios completos.
  - **Selección Meteorológica**: Se priorizaron las horas de 9:00 a 17:00 para condiciones diurnas relevantes.
  - **Proyecciones**: Se integraron predicciones meteorológicas de 7 días para mejorar las proyecciones.

4.4 Cálculo del caudal de salida

Para cada estanque de regulación y elementos de almacenamiento (tranques), se calcula el volumen almacenado utilizando la relación lineal:

$$V(t) = \frac{h(t)}{h_{max}} * Vmax$$

donde:

- $h(t)$ es el nivel medido en el instante t.
- h_{max} es el nivel máximo permitido (en cm).
- V_{max} es el volumen máximo del depósito (en m^3).

La proporcionalidad se basa en la hipótesis de que el volumen varía linealmente con el nivel, lo cual es válido en sistemas con geometría regular, lo que corresponde a la totalidad de estanques de regulación de Aguas del Valle (geometría cilíndrica).

Luego se calcula la tasa de cambio del volumen total en el tiempo para cada estanque, sumando todos en un instante “t”:

$$V_{total}(t) = \sum_{i} V_i(t)$$

La variación del volumen se calcula mediante:

$$\Delta V(t) = V_{total}(t) - V_{total}(t - \Delta t)$$

con un intervalo temporal fijo Δt = 3600segundos (1 hora). Así, la tasa de cambio del volumen es:

$$\frac{\Delta V(t)}{\Delta t} = \frac{V_{total}(t) - V_{total}(t - 3600s)}{3600}$$

Este cálculo es crucial ya que la tasa de cambio representa la variación del volumen en el tiempo, un indicador directo del flujo de salida o entrada en el sistema

El caudal neto de entrada, denotado por $Q_{Entrada}(t)$, se obtiene de las mediciones. Para determinar el caudal de salida, se emplea la siguiente relación de balance:

$$Q_{salida}(t) = Q_{Entrada}(t) - \frac{\Delta V(t)}{\Delta t}$$

La ecuación describe que el caudal de salida en un sistema de almacenamiento de agua se calcula como el caudal de entrada menos la tasa de cambio del volumen almacenado. Esto implica que:

- Si el volumen almacenado aumenta (tasa de acumulación positiva), el caudal de salida será menor que el caudal de entrada.
- Si el volumen disminuye (tasa de acumulación negativa), el caudal de salida será mayor que el caudal de entrada.

El proceso implementado utiliza una combinación de técnicas para modelar el balance hídrico de forma eficiente:

- Interpolación lineal: Asegura la continuidad de las series temporales de datos.
- Método de diferencias finitas: Estima la derivada del volumen, clave para determinar el caudal de salida.
- Conversión de unidades y agrupamiento de datos: Garantizan precisión y consistencia en los resultados.

4.5 Aplicación del PCA

El PCA transforma variables correlacionadas en componentes principales no correlacionados, simplificando los datos sin perder información esencial.

4.5.1 Aplicación al Caudal de Salida

- **Datos**: Matriz diaria de caudales horarios (24 horas por día).
- **Proceso**: Se aplica PCA a la matriz escalada y centrada, seleccionando componentes que expliquen al menos el 90% de la varianza (generalmente 3 a 5 componentes).
- **Resultado**: Los scores de los componentes representan la dinámica diaria del consumo y se usan como serie temporal en el modelo.

4.5.2 Aplicación a Variables Meteorológicas

- **Datos**: Matrices diarias de temperatura, precipitación y cobertura de nubes (9:00-17:00).
- **Proceso**: Se aplica PCA por variable, seleccionando componentes hasta el 90% de la varianza. El número máximo de componentes (e.g., 3) se usa para todas.
- **Resultado**: Los scores combinados de los componentes se emplean como variables exógenas en el modelo ARIMAX.

4.6 Modelado Predictivo con ARIMAX

4.6.1 Descripción del Modelo ARIMAX

El modelo ARIMAX extiende el ARIMA tradicional al incluir variables exógenas (scores meteorológicos). Sus componentes son:

- **AR (Autorregresivo)**: Modela la dependencia de valores pasados.
- **I (Integrado)**: Diferencia la serie para lograr estacionariedad.
- **MA (Media Móvil)**: Incorpora errores pasados.
- **X (Exógenas)**: Añade la influencia de las condiciones climáticas proyectadas.

4.6.2 Ajuste del Modelo

4.6.2.1 Entrenamiento Inicial

Se usa auto.arima() para determinar los parámetros óptimos (p, d, q), guardando el modelo base.

La función **auto.arima()** del paquete forecast en R automatiza el proceso de selección del modelo ARIMAX más adecuado para una serie temporal. Ajustar un modelo ARIMA manualmente implica probar diferentes combinaciones de parámetros (**p**, **d**, **q**, y sus equivalentes estacionales si aplica).

Cuando se trabaja con modelos ARIMAX, auto.arima() permite incorporar variables exógenas a través del argumento xreg, integrando información adicional en el análisis.

4.6.2.2 Actualización

Con nuevos datos, el modelo se reentrena manteniendo su estructura, asegurando eficiencia y adaptabilidad.

4.6.3 Proceso de proyección y reconstrucción de la serie original

Una vez entrenado el modelo ARIMAX utilizando los *scores* de los componentes principales del caudal y las variables meteorológicas como exógenas, se procede a realizar la proyección de la serie temporal para un horizonte de 7 días (168 horas). Este proceso implica los siguientes pasos:

4.6.3.1 Proyección de los scores:

Utilizando el modelo ARIMAX ajustado, se proyectan los *scores* de los componentes principales del caudal para los próximos 7 días. Para ello, se emplean tanto los *scores* históricos como las proyecciones de las variables meteorológicas (obtenidas de fuentes como Open-Meteo) como inputs exógenos.

4.6.3.2 Reconstrucción de la serie original:

Los *scores* pronosticados se transforman de vuelta a la serie original de caudales horarios utilizando los *loadings* (pesos) obtenidos del PCA aplicado a los datos de caudal. Este paso permite obtener predicciones detalladas del consumo de agua para cada hora de los próximos 7 días. Este enfoque garantiza que las predicciones no solo capturen las tendencias y patrones históricos del consumo, sino que también incorporen la influencia de las condiciones climáticas proyectadas.

4.7 Resumen

En el siguiente esquema se resume el proceso para generar un pronóstico de consumo de agua potable:

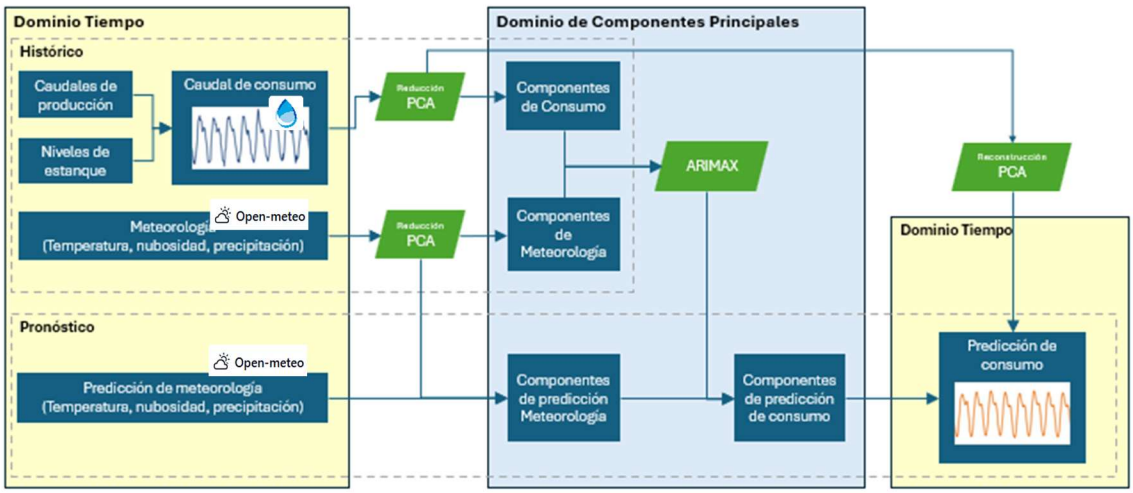


Ilustración 17 Procesos del Modelo de predicción de consumo de agua potable

Fuente: Elaborado por ESVAL-ADV

5 Resultados

Este capítulo presenta los resultados de la aplicación de modelo la predicción del consumo de agua en La Serena y Coquimbo, basado en el análisis de datos de caudal y variables meteorológicas, aplicando las técnicas descritas en capítulos anteriores, como el Análisis de Componentes Principales (PCA) y el modelo ARIMAX para desarrollar predicciones precisas y útiles para la gestión.

5.1 Desempeño del Modelo Predictivo

5.1.1 Análisis visual

La comparación gráfica entre la serie histórica y la serie reconstruida a partir de las predicciones permite identificar si se capturan correctamente las tendencias y patrones diarios. Para ilustrar el desempeño del modelo, se realizó una evaluación del consumo de agua proyectado para

el período del 13 de febrero al 20 de febrero de 2025, representado en “Ilustración 18 Resultados de Proyección – Datos observados versus Predicción”.

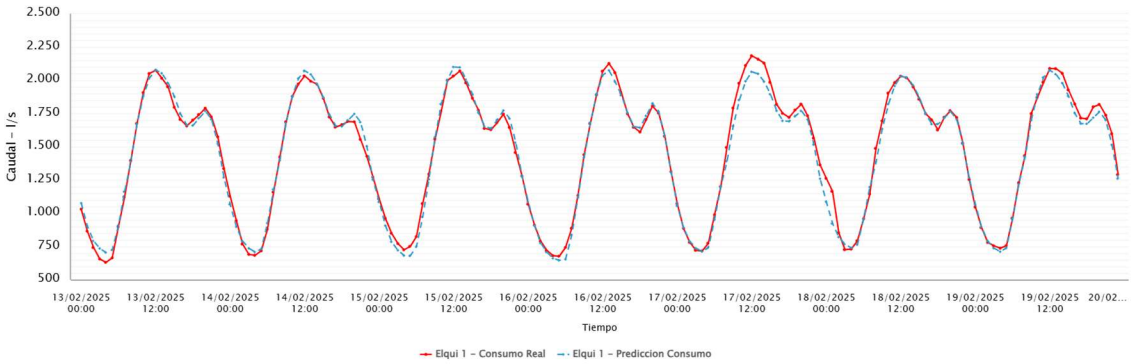


Ilustración 18 Resultados de Proyección – Datos observados versus Predicción

Según el análisis visual, las proyecciones horarias muestran que el modelo es capaz de reproducir los picos y valles típicos del consumo diario. Se observa, por ejemplo, un incremento en la demanda durante las horas punta y una disminución en períodos de menor actividad, lo que concuerda con la estructura intradía del consumo registrada en el historial

5.2 Métricas de error agregadas

El modelo ARIMAX, combinado con PCA, mostró un desempeño sobresaliente en la predicción del consumo de agua a corto plazo (7 días). En este horizonte de proyección se calcularon las siguientes métricas en el conjunto de prueba. Las métricas obtenidas fueron:

- **Error Cuadrático Medio (RMSE)**: 53.3 l/s
- **Error Absoluto Medio (MAE)**: 38.9 l/s
- **Error Porcentual Absoluto Medio (MAPE):** 2.97%
- **Error Porcentual Medio (MPE):** 1.53%
- **Coeficiente de correlación (R^2)**: 0.988

El análisis de estas métricas indica que el RMSE de 53.3 L/s y el MAE de 38.9 (l/s) sugieren que, en promedio, el error absoluto de las predicciones se mantiene en rangos aceptables, considerando el nivel general de consumo es en promedio 1318 (l/s). Un MAPE de 2.97% denota

que el error medio porcentual es relativamente bajo, lo que indica una alta precisión relativa en las predicciones. La MPE de 1,53% revela un ligero sesgo positivo, señalando que, en promedio, las predicciones tienden a sobreestimar marginalmente el consumo real.

6 Gestión Operativa

6.1 Aplicación del modelo en la Gestión Operativa.

El modelo se utiliza mediante una gestión que involucra a diferentes actores en la operación del sistema. Se aplica de una forma práctica para definir la operación de los sistemas productivos, optimizando el uso de fuentes superficiales y de mayor costo operativo.

6.1.1 Implementación del Modelo de Predicción

El modelo se encuentra implementado en un archivo EXCEL, en el cual a través de un complemento se de descargan los datos y los resultados del modelo de predicción provenientes de la base de datos de SmartSCADA. Los resultados del modelo de predicción son inyectados como una variable a SMARTSCADA de forma automática a través de máquinas virtuales. El acceso a esta herramienta se realiza a través de conexiones remotas y la gobernanza de la herramienta es liderada por el equipo de Inteligencia Operacional de ESVAL y Aguas del Valle.

6.1.2 Reuniones de coordinación

Para revisar los resultados y programar la operación del sistema, se realiza una reunión diaria a las 16:30, con el objetivo de que la participación de equipos multidisciplinarios asegure correctamente la implementación de la metodología. El uso de este modelo está asociado solo a los miembros del equipo de Inteligencia Operacional, quienes dan soporte y utilizan la herramienta para definir el caudal de operación necesario para las próximas 24 horas según el pronóstico de consumo arrojado por el modelo de predicción. Los resultados son validados y ajustado, si así lo amerita, según el criterio experto de otras áreas. Un actor complementario importante en el proceso de validación es el “Equipo territorial de producción”, llamado “Departamento de Producción de Elqui”, quienes son los encargados de la operación en terreno de la infraestructura.

Su rol en la reunión es validar las condiciones operativas de los sistemas de producción. En el caso de existir observaciones, este equipo levanta las observaciones para ajustar el modelo y los caudales de operación. Este equipo debe validar y considerar las gestiones necesarias para cumplir la operación de los sistemas de producción.

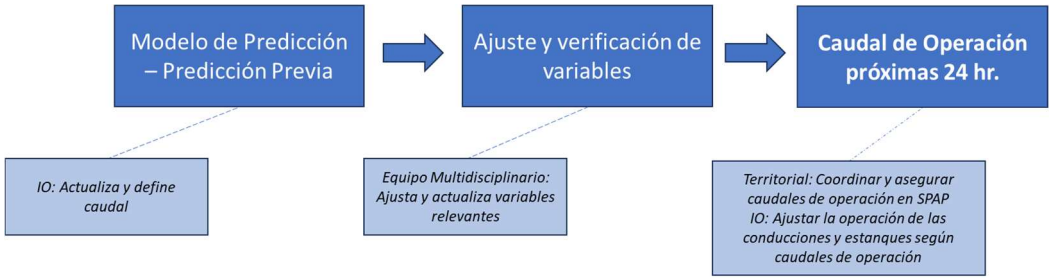


Ilustración 19 Coordinación diaria para aplicación del modelo

6.2 Descripción de la Herramienta Operativa

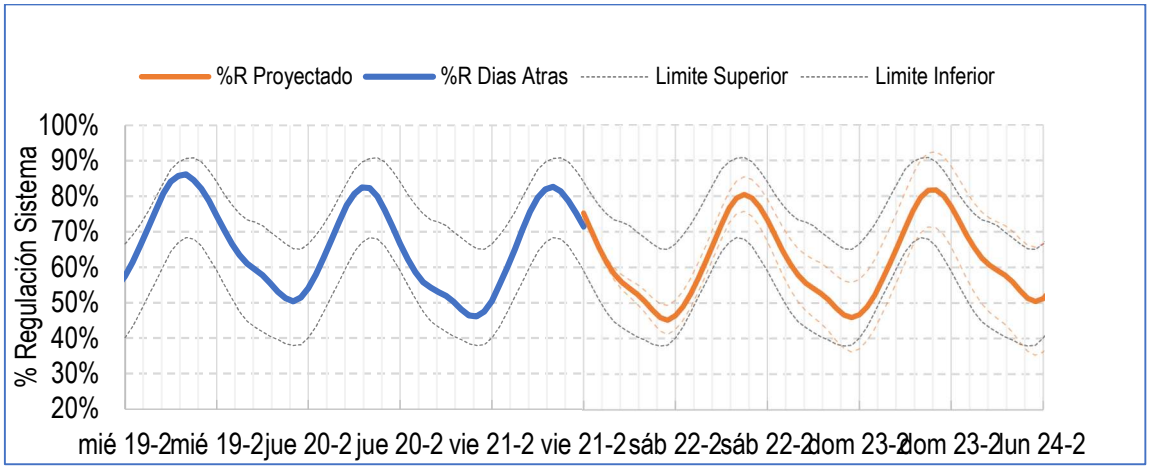
En la herramienta operativa, se cargan los datos del modelo de proyección de consumo (caudal de salida). En base a esta proyección, se define como operaran las fuentes y se ajustaran las fuentes variables con el objetivo de que el volumen almacenado a las 7:00 AM de la mañana del día siguiente sea del 82%. Como metodología de control, ya que corresponde a la operación del sistema, lo que se controla es el porcentaje del volumen regulación total del sistema, es decir el volumen proyectado que se almacenará versus el volumen total del sistema (55.000 m3).

En la planilla operativa se realizan los siguientes pasos para proyectar el volumen de regulación:

- Definición de la operación de fuentes: en base a la información operativa se define como operaran las fuentes subterráneas. También se define la operación de aporte o recarga del tranque de regulación en SPAP Las Rojas. Esta fuente permite amortiguar los cambios de demanda.

	Fuentes (l/s)											
	Las Rojas (Sub)	Las Rojas (Sup)	Las Rojas (Río)	Las Rojas (Salida Tranque)	Las Rojas (Entrada Tranque)	Compra de Agua	Saturno	Punta Piedra	Compañias	Alfalfares	Tierras Blancas	San Ramon + San Carlos + Mandarinos
Promedio Ultimas 24 Horas	78	701	739	0	38		9	210	148	64	59	76
Minimo Diario	64						8	167	135	55	47	63
Máximo Diario	80						10	214	168	72	58	95
sáb. 22-mar-25 17:00	78	693		0	0		10	227	166	73	59	75
sáb. 22-mar-25 18:00	78	705		0	0		11	231	166	72	58	75
sáb. 22-mar-25 19:00	78	694		0	0		9	228	166	71	58	76
sáb. 22-mar-25 20:00	77	691		0	0		10	226	166	71	59	75
sáb. 22-mar-25 21:00	78	695		0	0		10	226	166	68	59	76
sáb. 22-mar-25 22:00	77	428	428	0	30	262	10	210	148	64	59	76
sáb. 22-mar-25 23:00	77	428	428	0	30	262	10	210	148	64	59	76
dom. 23-mar-25 0:00	77	428	428	0	30	262	10	210	148	64	59	76

- Ajuste de la banda de Regulación: Para ajustar la regulación al target definido, se ajusta el caudal de aporte superficial del rio Elqui. Este caudal se define constante por 24 Hrs (criterio definido por Junta de Vigilancia de Rio Elqui). Dependiendo la disponibilidad de agua, se ajusta el llenado de tranque las rojas



- Factor de consumo: Por hora se pueden ajustar los factores de consumo. Normalmente no se ajustan, pero debido a eventualidades (mantenciones programadas, trabajos, emergencias, etc.) este factor se ajusta, aumentando la demanda.

Día	Regulacion 8:00 am	Factor Consumo	Consumo diario	Oferta sin Las Rojas	Sub	Río (Base)	Tranque (Salida)	Tranque (Entrada)	Dren / Directo a Planta	Compra de agua	Total Las Rojas	Disponibilidad Fuentes	Oferta Total Elqui 1	Regulación 8:00 am Proyectada	Nivel Tranque (m) 8 AM día siguiente
							l/s	l/s	l/s	l/s				Target 82%	
vie. 21-02-25 08:00	83%	1,05	1539	618	77	428	0	50	790	412	741	1123	1359	81%	2,84
sáb. 22-02-25 08:00	79%	1,00	1428	650	77	428	0	100	710	382	837	1155	1487	80%	3,65
dom. 23-02-25 08:00	80%	1,00	1376	650	77	428	0	100	710	382	787	1155	1437	80%	4,43
lun. 24-02-25 08:00	79%	1,00	1468	650	77	428	0	0	810	382	887	1155	1537	84%	4,42
mar. 25-02-25 08:00	84%	1,00	1468	650	77	428	0	0	810	382	887	1155	1537	81%	4,42

7 Conclusiones y Recomendaciones

7.1 Conclusiones

Tras el análisis exhaustivo de los datos de consumo de agua y su modelación predictiva en el sistema de La Serena y Coquimbo operado por Aguas del Valle, se han obtenido las siguientes conclusiones:

1. El modelo ARIMAX integrado con PCA ha demostrado una alta eficacia predictiva, alcanzando un MAPE del 2.97% para un horizonte de 7 días, lo que lo convierte en una herramienta operativa confiable para la planificación a corto plazo.
2. La temperatura se confirma como el factor meteorológico más influyente en el consumo de agua, con una correlación de 0.53, evidenciando la estrecha relación entre las condiciones climáticas y la demanda hídrica en la región.
3. Las variables de precipitación y cobertura de nubes presentan correlaciones débiles (-0.01 y -0.05 respectivamente), coherente con el clima semiárido característico de la zona, donde los eventos de lluvia son escasos e irregulares.
4. Se ha identificado una marcada estacionalidad en el consumo, con incrementos significativos durante el período estival (diciembre-febrero), presentando una variación promedio del 21.5% respecto a los meses invernales, atribuible tanto a factores climáticos como a la actividad turística.
5. La tendencia general al alza en el consumo, a pesar de la crisis hídrica actual, refleja el crecimiento poblacional y el desarrollo económico sostenido de la región, planteando desafíos importantes para la gestión futura del recurso.

6. La implementación del modelo ha permitido optimizar los procesos de producción y distribución, priorizando fuentes subterráneas de menor costo operativo y reduciendo la dependencia del río Elqui, especialmente por la disponibilidad de agua embalsada en embalse Puclaro.
7. El sistema de distribución, compuesto por 1,315 km de tuberías y 38 estanques con capacidad total de 55,000 m^3, demuestra robustez para soportar la demanda actual, pero requiere una gestión predictiva precisa para su operación eficiente.
8. La metodología de reducción dimensional mediante PCA ha resultado altamente efectiva, simplificando la complejidad de los datos horarios mientras mantiene el 90% de la varianza, lo que ha mejorado significativamente la calidad del modelo.
9. La integración del sistema SmartSCADA para la recolección de datos en tiempo real ha sido determinante para el éxito del modelo, subrayando la importancia de la digitalización en la gestión moderna de recursos hídricos.
10. Las reuniones diarias de coordinación entre los equipos de inteligencia operacional y producción territorial han garantizado la validación continua del modelo y su ajuste a las condiciones reales de operación.
11. Se han identificado variaciones significativas durante eventos masivos, como la reducción del 22% en el consumo durante el 1 de enero, señalando la necesidad de incorporar ajustes específicos para estas situaciones excepcionales.
12. El horizonte predictivo de 7 días ha demostrado ser efectivo para las necesidades operativas actuales, aunque existe potencial para extenderlo mediante la integración de proyecciones climáticas de mayor plazo.
13. La implementación práctica en entornos Excel y su uso cotidiano por parte del personal operativo confirman la viabilidad y adopción efectiva del modelo en condiciones reales de trabajo.
14. El modelo desarrollado presenta alta adaptabilidad para su aplicación en otras zonas semiáridas que cuenten con sistemas adecuados de recolección y monitoreo de datos, ampliando su potencial impacto más allá del ámbito local.

15. La incorporación de variables socioeconómicas o información sobre eventos masivos representa una oportunidad para reducir aún más el error medio absoluto (MAE), actualmente en 38.9 l/s.
16. Desde una perspectiva económica, la optimización en el uso de fuentes de agua ha permitido reducir costos operativos significativos, especialmente frente a las restricciones impuestas por la Junta de Vigilancia del Río Elqui.
17. El enfoque desarrollado contribuye significativamente a la sostenibilidad hídrica en un contexto de cambio climático, alineándose con los objetivos de desarrollo sostenible y gestión responsable de recursos escasos.

7.2 Recomendaciones y Futuras Líneas de Investigación

La investigación sugiere que la incorporación de técnicas avanzadas de machine learning podría capturar relaciones no lineales entre variables, mejorando aún más la precisión del modelo en escenarios complejos y variables. Además, se debe generar una exploración de variables meteorológicas adicionales, ya que, aunque la temperatura fue el factor meteorológico más influyente, otras variables como la humedad, la radiación solar o la velocidad del viento podrían impactar el consumo en un clima semiárido. Investigar su correlación y posible inclusión en el modelo podría refinar las predicciones, especialmente ante la baja influencia de la precipitación observada

Otra línea de investigación que se sugiere es la optimización multiobjetivo de fuentes de Agua, ya que la metodología optimiza el uso de fuentes superficiales y subterráneas, pero una investigación más profunda podría explorar la optimización considerando costos, sostenibilidad y calidad del agua. Modelos de decisión multiobjetivo podrían mejorar la gestión integral de los recursos.

8 Bibliografía

(1) Severa crisis hídrica en la Región de Coquimbo: medidas, propuestas y .... https://www.paiscircular.cl/agua/severa-crisis-hidrica-en-la-region-de-coquimbo-medidas-propuestas-y-reflexiones-ante-una-carrera-contra-el-tiempo/.

(2) El lado feo de la sequía en Chile: el caso de la región de Coquimbo .... https://tiemporeal.periodismoudec.cl/2021/09/03/el-lado-feo-de-la-sequia-en-chile-el-caso-de-la-region-de-coquimbo/.

(3) Se acaba el agua en la Región de Coquimbo: "Esto ya es una catástrofe .... https://bing.com/search?q=escenario+hidrico+de+sequ%c3%ada+de+la+region+de+coquimbo.

(4) Se acaba el agua en la Región de Coquimbo: "Esto ya es una catástrofe .... https://www.elmostrador.cl/cultura/2024/04/03/sequia-en-la-region-de-coquimbo-esto-ya-es-una-catastrofe-humanitaria/.

(5) Crítica situación de sequía en Región de Coquimbo: Embalses comienzan a .... https://radio.uchile.cl/2024/03/10/critica-situacion-de-sequia-en-region-de-coquimbo-embalses-comienzan-a-secarse/.

(6) Por qué persiste la sequía en la Región de Coquimbo | INIA. https://www.inia.cl/2024/07/04/por-que-persiste-la-sequia-en-la-region-de-coquimbo/.

(7) The R Project for Statistical Computing. (n.d.). R documentation: Date-time classes (POSIXct and POSIXlt). https://stat.ethz.ch/R-manual/R-devel/library/base/html/DateTimeClasses.html

(8) Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. https://otexts.com/fpp3/

(9) The R Project for Statistical Computing. (n.d.). R documentation: Date-time classes (POSIXct and POSIXlt).

https://stat.ethz.ch/R-manual/R-devel/library/base/html/DateTimeClasses.html

(10) Aguas del Valle. (2024). *Plan de desarrollo vigente - Zonal Elqui*. https://www.aguasdelvalle.cl/empresas/nosotros/plan-de-desarrollo-vigente/zonal-elqui/

(11) iAgua. (s. f.). ¿Cómo se aplica la inteligencia artificial en la gestión del agua? *iAgua*. Recuperado de https://www.iagua.es/respuestas/como-se-aplica-inteligencia-artificial-gestion-agua

(12) Superintendencia de Servicios Sanitarios. (2021). *Estudio tarifario Empresa Servicios Sanitarios Aguas del Valle 2021-2026: Informe final*. Capítulo 6: Definición de la empresa modelo (Proyecto de reposición). Repositorio Dirplan, Ministerio de Obras Públicas. https://repositoriodirplan.mop.gob.cl/biblioteca/handle/20.500.12140/208220

(13) Arbués, F., García-Valiñas, M. Á., & Martínez-Espiñeira, R. (2003). Estimation of residential water demand: A state-of-the-art review. *The Journal of Socio-Economics, 32*(1), 81-102. https://doi.org/10.1016/S1053-5357(03)00005-2

(14) Abulibdeh, A., Pirasteh, S., Mafi-Gholami, D., Kucukvar, M., & Onat, N. C. (2024). Unveiling the nexus between land use, land surface temperature, and carbon footprint: A multi-scale analysis of building energy consumption in arid urban areas. *Earth Systems and Environment, 8*(3), 873-894. https://doi.org/10.1007/s41748-024-00550-6

9 Anexos

9.1 POSIXct

POSIXct es una clase en el lenguaje de programación R diseñada para representar fechas y horas como un valor numérico continuo, específicamente el número de segundos transcurridos desde el 1 de enero de 1970, 00:00:00 UTC, conocida como la "época Unix" (Hyndman & Athanasopoulos, 2021). Según Hyndman y Athanasopoulos, esta clase forma parte del sistema POSIX y es una de las dos formas principales de manejar datos temporales en R, junto con POSIXlt. En POSIXct, el "ct" significa "calendar time" (tiempo calendario), lo que indica que almacena la información temporal en un formato compacto y eficiente, ideal para cálculos y almacenamiento en análisis de series temporales.

En el contexto del trabajo, POSIXct se utiliza para convertir y gestionar las fechas de los datos de caudal (SmartSCADA) y meteorológicos (Open-Meteo). Hyndman y Athanasopoulos (2021) destacan que este formato es especialmente útil para trabajar con series temporales porque permite especificar una zona horaria (como UTC) y asegura que las observaciones estén ordenadas cronológicamente, lo que es esencial para modelar datos horarios como los del consumo de agua. Por ejemplo, al transformar una fecha como "2022-01-01 00:00:00" a POSIXct con as.POSIXct(), R la convierte en un valor numérico (e.g., 1640995200 segundos desde la época), facilitando operaciones como diferencias temporales o sincronización entre conjuntos de datos.

A diferencia de POSIXlt, que descompone las fechas en una lista de componentes (año, mes, día, etc.), POSIXct ofrece una representación más eficiente para grandes volúmenes de datos, como los registros horarios de la tesis (Hyndman & Athanasopoulos, 2021). Esta eficiencia es clave para el preprocesamiento descrito en el estudio, donde se requiere alinear datos de múltiples fuentes y filtrar días completos de 24 horas, garantizando la consistencia temporal necesaria para el modelo ARIMAX.

9.2 Preparación de los Datos

En esta etapa se busca estructurar y depurar la información histórica necesaria para la proyección del consumo de agua, garantizando la consistencia temporal y la integración de las variables exógenas. Es importante destacar que los datos vienen limpios (sin outliers, ni valores nulos). A continuación, se detallan los pasos realizados en el código:

9.2.1 Conversión y ordenamiento de fechas

Para asegurar una correcta manipulación temporal, la columna que contiene las fechas en el archivo de datos de caudal se convierte al formato POSIXct con la zona horaria "UTC" utilizando el formato "%Y-%m-%d %H:%M:%S", pues esta conversión permite realizar operaciones de ordenamiento cronológico de los registros y extraer componentes temporales (día y hora) para posteriores agrupaciones.

Una vez convertidas las fechas, se ordenan los datos de forma ascendente para garantizar la continuidad en la secuencia de tiempo.

9.2.2 Selección de días completos y segmentación de la serie

Dado que el análisis se fundamenta en la estructura diaria (24 horas), se verifica que cada día presente en el histórico cuente con exactamente 24 observaciones. Para ello, se agrupan los registros por día y se filtran aquellos en los que la cantidad de observaciones es igual al tamaño de la ventana definida (24 horas). Esta estrategia es crucial para:

- Eliminar días incompletos que podrían introducir inconsistencias en la modelación.
- Facilitar la estructuración de los datos en una matriz diaria, donde cada fila representa un día y las columnas corresponden a las 24 horas (denominadas de H0 a H23).

Posteriormente, se determina el último día completo disponible, el cual se utiliza para delimitar el período de entrenamiento.

Los datos se segmentan en dos conjuntos:

1. **Conjunto de entrenamiento:** Comprende todos los días completos hasta el último día identificado.
2. **Horizonte de proyección:** Se define la proyección para los 7 días siguientes, asumiendo que se disponen de los datos meteorológicos correspondientes para ese período.

9.2.3 Integración con variables meteorológicas

Para mejorar la capacidad predictiva del modelo se incorporan las variables meteorológicas (temperatura, lluvia y cobertura de nubes) como factores exógenos. El proceso incluye:

- Conversión del campo de fecha en el archivo meteorológico al formato POSIXct (zona horaria UTC).
- Ordenamiento de los datos y extracción de los componentes day y hour.
- Filtrado de la información relevante, concentrándose en el intervalo de las 09 : 00 a las 17 : 00, que se considera representativo para el consumo de agua.
- Realización de un cruce (merge) entre los datos de caudal y los meteorológicos, asegurando que las observaciones de ambos conjuntos estén sincronizadas temporalmente.

9.2.4 Preparación de los datos de caudal de salida en SmartSCADA

Para este punto, se ha señalado que se dispone de los datos de caudal de salida, y los datos meteorológicos, no obstante, es importante hacer un paréntesis y destacar que dichos datos del caudal de salida, como se indicó en un comienzo, son datos descargados desde SmartSCADA, sin embargo, también se ha desarrollado un código el cual fue implementado en SmartSCADA, el cual se encarga de obtener los valores de caudal de salida (consumo de agua) de forma diaria y horaria.

Es por ello, que a continuación se presenta un análisis exhaustivo del proceso computacional empleado para determinar el consumo de agua a partir del cálculo del caudal de salida. Se describen en detalle los métodos numéricos utilizados, tales como la interpolación y el cálculo mediante diferencias finitas, así como las conversiones de unidades y fórmulas aplicadas. Todo ello se sustenta en fundamentos matemáticos sólidos, proporcionando una visión profesional y rigurosa del algoritmo implementado en Python.

9.2.5 Extracción y preprocesamiento de datos

Se accede a dos fuentes de datos principales:

1. **Base de datos SQLite:** Se extraen dos conjuntos de datos:
    - datos_nivel: Contiene mediciones de nivel de agua, identificadas por una *descripción*, un *tag*, una marca temporal (timeStamp y localTime) y el valor medido.
    - datos_caudal: Contiene mediciones de caudal.

2. **Archivo Excel:** Proporciona parámetros geométricos, en particular el Volumen máximo (en m3) y el Nivel máximo (en cm) para cada descripción.

Los datos se ordenan cronológicamente y, para garantizar la continuidad de las series temporales, se aplica una interpolación lineal sobre los valores nulos, lo cual permite la estimación de valores faltantes de manera coherente.