

**PROTOCOLO SISTEMÁTICO PARA EL MONITOREO DE
CALIDAD DE AGUA SUBTERRÁNEA BASADO EN LA
APLICACIÓN DE TECNOLOGÍAS EMERGENTES**

Liz Verónica Báez Lovera

Orientadores: Prof. Cynthia Villalba, Dra.
Prof. Juan Pablo Nogués, Dr.

Tesis presentada a la Facultad Politécnica, Universidad Nacional de Asunción,
como requisito para la obtención del Grado de Doctora en Ciencias de la
Computación.

SAN LORENZO - PARAGUAY

Octubre - 2023

*Dedicado a mis padres, Verónica y Melanio, y a mis abuelos,
Petrona y Leoncio(+).*

*Los formadores de mi carácter y educación, mis maestros sobre el valor del
esfuerzo, el trabajo, el amor y la unión familiar.
Para Ustedes, mi eterna gratitud y amor.*

Liz Verónica

Agradecimientos

Agradezco a Dios y a la Virgen María, por bendecir mi vida y por sostenerme muy fuerte en todo tiempo.

Agradezco al amor de mi vida Ariel, por ser el mejor compañero de vida y acompañarme en todos mis momentos, felices y difíciles. ¡Este logro es de los dos!

Agradezco a toda mi gran Familia, por apoyar siempre mis decisiones y comprender mis ausencias. Ustedes son los pilares que me sostienen y motivan día a día.

Agradezco a mis orientadores Prof. Cynthia V. y Prof. Juan Pablo N., por ser excelentes mentores y profesionales con gran calidad humana, por todas sus enseñanzas y el valioso tiempo que dedicaron en mi formación profesional. Sin el apoyo de ustedes este proyecto de investigación no hubiera sido posible.

Agradezco a los Profesores, Compañeros y Amigos que me acompañaron durante mi formación.

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT, Paraguay) mediante el Proyecto POSG17-53/BNAC02 y por la beca otorgada para cursar el programa.

Liz Verónica

PROTOCOLO SISTEMÁTICO PARA EL MONITOREO DE CALIDAD DE AGUA SUBTERRÁNEA BASADO EN LA APLICACIÓN DE TECNOLOGÍAS EMERGENTES

Autor: Liz Verónica Báez Lovera

Orientadores: Cynthia Villalba, Dra.

Juan Pablo Nogués, Dr.

RESUMEN

Los monitoreos de calidad de aguas subterráneas son sumamente necesarios para la caracterización, el control de los acuíferos y el desarrollo sostenible, especialmente en entornos urbanos donde la contaminación tiene un origen antropogénico debido a diferentes prácticas desarrolladas desde la superficie. Los programas de monitoreo incluyen la identificación de ubicaciones de los pozos, la detección de contaminantes y la gestión sostenible. La selección de los pozos de monitoreo y parámetros de calidad de agua subterránea se basan principalmente en diferentes criterios subjetivos como la experiencia del tomador de decisiones, la experiencia en campo y los antecedentes de contaminación. A su vez, conducen a la generación de grandes conjuntos de datos que comprenden varios parámetros de calidad de agua, distribuidos tanto espacial como temporalmente, lo que los hace ricos en datos, pero pobres en información y dificultan su interpretación. En esta tesis doctoral se presentó un protocolo sistemático para el monitoreo de calidad de agua subterránea y la reducción de parámetros, aplicado a un caso de estudio real y replicable a otras áreas de interés. El aporte consistió en proponer un protocolo de monitoreo integral y de utilidad para la evaluación objetiva de la calidad del agua subterránea, basado en la aplicación de tecnologías emergentes. En la selección de pozos se aplicó un algoritmo evolutivo multi-objetivo vinculado con el uso de datos y herramientas geoespaciales, sobre el riesgo de contaminación, la ubicación de los pozos y la accesibilidad a los mismos para maximizar las posibilidades de encontrar contaminación en las aguas subterráneas. En la reducción de parámetros se aplicaron técnicas de reducción multivariada y algoritmos de selección de atributos vinculados con la indexación de la calidad del agua. Dichas metodologías se aplicaron a un caso de estudio de un acuífero local, utilizando datos geoespaciales reales, se implementó el monitoreo en 66 pozos, se evaluaron 21 parámetros de calidad de agua, se analizaron los resultados de calidad de agua

y finalmente, se redujo el número de parámetros a monitorear. En el 83% de los pozos se encontraron valores fuera del límite permitido, según la normativa considerada. Además, se detectó la presencia de concentraciones de nitrato por encima de los límites permisibles en el 42% de los pozos. El protocolo de selección de pozos ejecutado tuvo una eficiencia del 50% a la hora de encontrar muestras fuera de los límites permisibles, mientras que dos campañas anteriores sin un protocolo de selección optimizado mostraron una eficiencia del 21% y del 37%. Además, en la validación considerando otros escenarios de estudios se obtuvo una eficiencia del 90%.

Apéndice General

Lista de Figuras	g
Lista de Tablas	h
Lista de Acrónimos	i
1 Introducción	1
1.1 Problemas de Investigación	4
1.2 Objetivos de la Investigación	6
1.2.1 Objetivo General	6
1.2.2 Objetivos Específicos	6
1.3 Contribuciones	6
1.3.1 Publicación en revista	6
1.4 Organización de la Tesis	8
2 Revisión de la Literatura	9
3 Metodología	14
3.1 Caso de estudio: Acuífero Patiño	14
3.2 Selección óptima de pozos de monitoreo	16
3.2.1 Definición de las funciones objetivo	19
3.3 Evaluación de los resultados de calidad de agua del monitoreo	21
3.4 Reducción y selección de parámetros de calidad de agua	24
3.5 Validación del protocolo	26
3.5.1 Escenario I con datos reales geoespaciales	26
3.5.2 Escenario II con datos sintéticos	28
4 Resultados y Análisis	30
4.1 Selección óptima de pozos de monitoreo	30
4.2 Resultados de calidad de agua del monitoreo	32

4.2.1	Evaluación del protocolo mediante una comparación con otras campañas de muestreo	34
4.2.2	Mapeo de la calidad de agua subterránea	35
4.3	Reducción y selección de parámetros de calidad de agua	36
4.4	Validación del protocolo	40
4.4.1	Selección de pozos	40
4.4.2	Selección de parámetros	41
5	Conclusiones y Trabajos Futuros	43
5.1	Conclusiones	43
5.2	Trabajos Futuros	45
	Referencias Bibliográficas	47

Lista de Figuras

3.1	Mapa de ubicación del acuífero Patiño, Departamentos Central y Paraguari, Paraguay.	15
3.2	Datos geoespaciales considerados para la selección óptima de pozos de monitoreo.	17
3.3	Diagrama de flujo del método de resolución computacional implementado para la selección óptima de pozos. El diagrama de flujo muestra el proceso de evaluación de cada solución basado en el algoritmo NSGA-II.	18
3.4	Datos geoespaciales considerados para el Escenario I (567 km ²), obtenidos de [19].	27
4.1	Gráfico de coordenadas paralelas del frente Pareto con 61 soluciones óptimas. La solución seleccionada aparece resaltada. Cada eje representa la normalización de cada objetivo.	31
4.2	Distribución de los 66 pozos seleccionados con el NSGA-II y PO. Los pozos están identificados según la prioridad: pozos de acceso público (aguaterías, juntas de saneamiento, piezómetros) y pozos privados.	31
4.3	Gráfico circular con el porcentaje de los 8 parámetros fuera de los límites permisibles de calidad de agua.	32
4.4	Mapa de los puntos críticos, considerando la densidad de parámetros fuera de los límites admisibles por pozo.	33
4.5	Mapa de la distribución espacial de la calidad de agua subterránea del acuífero Patiño.	36
4.6	Mapa de la distribución espacial de los 70 pozos óptimos seleccionados con NSGA-II PO para el Escenario I.	40

Lista de Tablas

3.1	Pesos unitarios y pesos relativos para cada parámetro WQI.	23
4.1	Comparación de las campañas de muestreo sobre el acuífero Patiño. Sólo se consideran los pozos muestreados en el AMA.	35
4.2	Selección de variables de calidad de agua subterránea con PCA-VSC.	37
4.3	Parámetros identificados por los algoritmos de selección de atributos.	38
4.4	Comparación de resultados de los algoritmos clasificadores en función al WQI inicial con los parámetros relevantes de los Subconjuntos 1, 2 y 3.	39
4.5	Comparación de la eficiencia entre las campañas.	41
4.6	Resumen de los subconjuntos evaluados por escenario con el algo- ritmo de clasificación J48.	42

Lista de Acrónimos

GIS: Sistemas de Informacion Geografica.

NSGA-II: Algoritmo Genetico de Clasificacion No Dominada II.

MOEA: Algoritmo Evolutivo Multi-Objetivo.

PCA: Analisis de Componentes Principales.

PO: Orden de Preferencia.

WQI: Indice de Calidad de Agua.

VSC: *Variable Selection Criteria*.

VC: *Variance Combination criterion*.

M1: Criterios de Selección M1.

M2: Criterios de Selección M2.

M3: Criterios de Selección M3.

Capítulo 1

Introducción

El agua subterránea se encuentra debajo de la superficie terrestre, en los acuíferos, y es un recurso vital para el suministro de agua potable. La importancia del agua subterránea radica en su capacidad para abastecer de agua dulce a áreas geográficas que no tienen acceso a fuentes superficiales, como ríos o lagos [14]. Por lo tanto, es un recurso valioso económicamente porque es esencial para la supervivencia humana y la salud, lo que lo convierte en un bien fundamental para el bienestar de las personas. Sin acceso a agua potable, las comunidades y las economías no pueden funcionar adecuadamente [41, 42, 63].

Además de ser vital para el consumo humano, el agua subterránea también es un recurso indispensable en la agricultura para el regadío de los cultivos, especialmente en regiones donde las precipitaciones son limitadas. También, es muy importante en la ganadería porque permite mantener la productividad de las tierras, asegurar la alimentación de la población y contribuir a la economía agrícola. Otro aspecto destacado es su papel fundamental en los ecosistemas terrestres y acuáticos. Los acuíferos ocupan grandes extensiones de terreno, alimentan arroyos y ríos, contribuyendo a mantener caudales de agua constantes. Sin embargo, no es un recurso ilimitado y se enfrenta a importantes amenazas debido a la expansión urbana y los cambios en el uso del suelo [41].

A medida que las áreas urbanas crecen y se desarrollan, las extensiones de los acuíferos se ven amenazadas. La expansión urbana implica cambios significativos en el uso del suelo, donde las superficies que antes estaban permeables se convierten en áreas impermeables debido a la construcción de edificios, carreteras y otras infraestructuras [18]. Además, los cambios en el uso del suelo en áreas urbanas a menudo implican la contaminación de los acuíferos. Las actividades humanas, como la industria, la agricultura intensiva y la disposición inadecuada de desechos, pueden introducir contaminantes en el suelo que eventualmente se in-

filtran en las aguas subterráneas, ocasionando graves consecuencias para la salud y el medio ambiente [27, 41, 49].

Por tanto, es crucial reconocer la importancia de proteger el agua subterránea, y realizar investigaciones y evaluaciones sobre su calidad para proteger la salud, gestionar el recurso y garantizar el cumplimiento de las normas o regulaciones ambientales. Dichas acciones permiten tomar decisiones acertadas, y promover el uso responsable y seguro de tan importante recurso. La investigación sobre la calidad del agua subterránea se centra en evaluar y determinar las concentraciones de diferentes parámetros a través del monitoreo, y analizar los procesos hidrogeoquímicos que controlan la calidad del agua subterránea [41]. Por lo tanto, establecer estrategias de monitoreo constante es fundamental para resguardar su calidad y el desarrollo sostenible.

Los programas de monitoreo de aguas subterráneas incluyen la identificación de ubicaciones de los pozos, la detección de contaminantes y la gestión sostenible. De acuerdo con [49], se debe diseñar una red efectiva de monitoreo de agua subterránea considerando una cantidad mínima de sitios de monitoreo, mientras se maximiza la detección de contaminación. Este último criterio de diseño se asocia a la selección de parámetros a monitorear y su posterior análisis cuantitativo en el laboratorio. La selección de los parámetros de calidad de agua subterránea, así como de los pozos de monitoreo, se realizan principalmente mediante diferentes criterios de los expertos en el área, basados en la experiencia subjetiva del tomador de decisiones, sugerencias propuestas por diferentes organismos relacionados a la gestión del agua, antecedentes de contaminación o experiencia en campo, y dependiendo de las fuentes potenciales de contaminación existentes y los parámetros consecuentes asociados. Por otro lado, los programas de monitoreo conducen a la generación de grandes conjuntos de datos que comprenden varios parámetros de calidad de agua, distribuidos tanto espacial como temporalmente, lo que los hace ricos en datos, pero pobres en información y dificultando su interpretación [20].

La literatura sobre el monitoreo de calidad de agua subterránea no reporta la utilización de protocolos en todas las fases del monitoreo. Diferentes autores [17, 51, 55, 56, 61], propusieron métodos estandarizados que involucran solo la toma de muestras en campo mediante criterios establecidos en guías o protocolo de monitoreo. Otros esfuerzos [5, 17, 43, 47], se enfocaron en los métodos para desarrollar protocolos de análisis de datos, relacionados a los resultados de calidad de agua e indicando diferentes técnicas estadísticas para evaluar los datos.

De acuerdo con [10], la utilización de un protocolo en todas las fases del monitoreo de calidad de agua permite minimizar y eliminar errores, garantizar la

generación de información consistente y confiable, determinar la línea de partida y mejorar la gestión de control de calidad de agua. Es posible mejorar la efectividad del monitoreo cuando se utiliza información para orientar el diseño del monitoreo y el análisis de datos [17]. Sin embargo, es necesario plantear varios pasos viables para el control, la evaluación y la conservación de la calidad del agua subterránea, siguiendo un método o sistema innovador para realizar un monitoreo racional y no subjetivo. A su vez, aplicar métodos que permitan que el resultado del monitoreo de calidad de agua, con una disminución de parámetros, se pueda comunicar fácilmente sin perder ninguna información significativa.

El presente trabajo de tesis doctoral se enmarca en investigaciones relacionadas a las tecnologías emergentes integradas con herramientas y técnicas avanzadas de información, con la ausencia de evaluaciones subjetivas, para diseñar un protocolo sistemático para monitorear la calidad del agua subterránea. Por lo tanto, se propone una metodología novedosa, sobre datos reales para: i) seleccionar los pozos óptimos de monitoreo, ii) evaluar los resultados del monitoreo, iii) seleccionar y reducir el número de parámetros.

El uso de tecnologías emergentes como la aplicación de algoritmos multi-objetivos vinculados con el uso de datos y herramientas geoespaciales; la aplicación de técnicas de reducción multivariada y algoritmos de selección de atributos vinculados con la indexación de la calidad del agua; pueden resultar beneficiosas en el dominio del monitoreo de calidad de agua subterránea. En la actualidad, la aplicabilidad de dichas técnicas combinadas se encuentra en su etapa primitiva de investigación [20]. Por lo tanto, las modificaciones adecuadas en las técnicas, así como la integración de más de una herramienta, aportarán a la metodología para desarrollar una propuesta novedosa de protocolo de monitoreo de calidad de agua subterránea.

En la primera parte del trabajo, se encara el problema de la selección de pozos de monitoreo utilizando un enfoque multi-objetivo, dado un caso de estudio de un acuífero local, utilizando datos geoespaciales reales, para su posterior muestreo y trabajo de campo. Seguido, se incluye el análisis de los resultados del muestreo como la distribución geoespacial de los parámetros de calidad de agua, la evaluación y el mapeo de calidad de agua subterránea basado en el método del Índice de calidad de agua (WQI, por sus siglas en inglés).

En la siguiente parte, se aborda el problema de la selección y reducción de parámetros de calidad de agua, aplicando diferentes criterios de selección de variables en el análisis de componentes principales, utilizando algoritmos de selección de atributos y el análisis predictivo considerando la clasificación en función al

WQI. Finalmente, se realiza una validación del protocolo propuesto en otros escenarios.

1.1 Problemas de Investigación

La presente tesis doctoral aborda desafíos de investigación científica relacionados al diseño de una estrategia de monitoreo de agua subterránea, que proporcione una guía de pasos a seguir mediante la utilización de tecnologías emergentes integradas con herramientas y técnicas avanzadas de información, y de esta manera, dejar de depender solo de evaluaciones subjetivas para resguardar la calidad del agua subterránea.

En particular, los siguientes problemas claves de investigación son encarados:

- **Problema de investigación 1 - Selección de pozos de monitoreo considerando múltiples objetivos**

En ocasiones, se pretende identificar los diferentes lugares en donde se deberán construir los pozos de monitoreo de agua subterránea. Sin embargo, cuando se cuenta con un inventario de pozos existentes; estos pueden ser considerados para realizar el monitoreo y evitar los gastos de inversión que conlleva la instalación de cada pozo nuevo.

Para seleccionar los pozos de monitoreo a partir de un conjunto de pozos existentes, se deberá considerar una alternativa de solución dentro de las tecnologías emergentes, mediante la utilización de un algoritmo que permita evaluar múltiples objetivos en conflicto simultáneamente como: maximizar las concentraciones de Nitrógeno Total (NT), maximizar las concentraciones de Coliformes Totales (CT), maximizar el área de cobertura y maximizar la selección de pozos de acceso público, como piezómetros, aguaterias y juntas de saneamiento.

- **Problema de investigación 2 - Evaluación y mapeo de la calidad del agua subterránea**

Los estándares de calidad de agua son normativas que describen los requisitos mínimos razonables (límites permisibles) de los componentes del agua y están definidos por valores numéricos, que deben cumplirse para proteger la salud de los consumidores [62]. Al definir dichas normativas, se consideran las condiciones ambientales, económicas y culturales locales. Es decir, cada

jurisdicción local puede reglamentar diferentes niveles permisibles dependiendo de las limitaciones socioeconómicas y ambientales. Por lo tanto, lo que puede ser adecuado en una región puede no serlo para otra [11].

En este contexto, es importante considerar otro enfoque para la planificación y gestión de la calidad del agua subterránea, mediante modelos estadísticos y matemáticos que ayudan a determinar el estado general del cuerpo de agua, considerando la influencia compuesta de un conjunto de parámetros de calidad de agua en comparación con ciertos estándares.

- **Problema de investigación 3 - Reducción del número de parámetros a monitorear**

Por lo general, en el monitoreo se consideran tantos parámetros como sea posible. Dicha práctica es común debido a que, en una etapa inicial o en ausencia de cualquier conocimiento previo, se desconocen qué parámetros son los más importantes para el muestreo. Como resultado se obtiene un gran conjunto de datos, con variables numerosas y resulta difícil de analizar e interpretar. En dichos casos, una reducción de parámetros es deseable.

Es necesario plantear la reducción de la dimensionalidad del conjunto de parámetros sin alterar las características principales, mediante un espacio de baja dimensión que contenga la mayor parte de la información del espacio original de alta dimensión. Para ello, se podrán plantear diferentes técnicas para reducir el sesgo y tomar decisiones más objetivas, a su vez, reducir el tiempo, el esfuerzo y los costos requeridos para monitorear un gran número de variables.

- **Problema de investigación 4 - Evaluación del protocolo propuesto**

La metodología propuesta involucra la utilización de un conjunto de técnicas emergentes para un protocolo sistemático para el monitoreo de calidad de agua subterránea. En nuestro mejor conocimiento, no se ha utilizado dicho protocolo con enfoque de inteligencia artificial en otros estudios. Por lo tanto, se debe evaluar el desempeño del método propuesto con los datos de otra área de interés y/o simular otros escenarios posibles, de manera a comprender las limitaciones, si necesita ser mejorado o si será eficiente, independientemente a cualquier área de interés.

1.2 Objetivos de la Investigación

A efectos de abordar dichos desafíos, los siguientes objetivos son delineados:

1.2.1 Objetivo General

Diseñar y validar un protocolo sistemático para el monitoreo de calidad de agua subterránea y la reducción de parámetros, utilizando un caso de estudio real y replicable a otras áreas de interés, basado en la aplicación de tecnologías emergentes.

1.2.2 Objetivos Específicos

1. Seleccionar los pozos de monitoreo utilizando un enfoque multi-objetivo.
2. Evaluar y mapear la calidad del agua subterránea basada en el método del Índice de calidad de agua (WQI).
3. Reducir el número de parámetros de calidad de agua aplicando diferentes criterios de selección de variables en el análisis de componentes principales (PCA) y aplicando algoritmos de selección de atributos.
4. Validar el protocolo propuesto considerando otros escenarios.

1.3 Contribuciones

La principal contribución de la presente tesis doctoral se presenta a continuación.

1.3.1 Publicación en revista

Protocolo sistemático para diseñar una campaña de muestreo de calidad de agua basada en un enfoque multi-objetivo para la selección de pozos, seguido de los resultados de la aplicación del protocolo, un análisis de los resultados del muestreo, la aplicación del cálculo WQI y una evaluación de la propuesta de selección de pozos (Objetivos 1-2). Dicha contribución se presenta en el artículo:

- **Liz Báez**, Claudia Ávalos, Christian von Lücken, Cynthia Villalba y Juan Pablo Nogués (2021). Designing and validating a groundwater sampling campaign in an unmonitored aquifer: Patiño aquifer case. *Environmental Earth Sciences*, 80(11), 1-16. DOI: 10.1007/s12665-021-09706-3

Publicación relacionada:

- **Liz Báez**, Cynthia Villalba y Juan Pablo Nogués (2019). Comparison of contaminant-specific risk maps for an urban aquifer: Patiño aquifer case. *Environmental Earth Sciences*, 78(5), 1-15. DOI: 10.1007/s12665-019-8141-3

1.4 Organización de la Tesis

El trabajo se divide en cinco capítulos para la comprensión y aplicación del protocolo sistemático desarrollado. En el Capítulo 1 se introduce al tema de investigación. En el cual se destaca la importancia de la calidad del agua subterránea y se abordan los antecedentes del problema. Además, se presenta una visión general de la metodología, los objetivos y principales aportes del trabajo.

En el Capítulo 2 se realiza una revisión exhaustiva de la literatura relacionada con el protocolo de monitoreo de calidad de agua subterránea, las tecnologías emergentes de monitoreo y las metodologías existentes. Se identifican brechas en el conocimiento y se analizan investigaciones previas relevantes que respaldan la necesidad de desarrollar un protocolo integral.

En el Capítulo 3 se describe la metodología utilizada para desarrollar el protocolo sistemático para el monitoreo. Además, se indica el área de estudio considerado y los datos utilizados para la aplicación del protocolo. El capítulo se divide en secciones que abordan los siguientes aspectos: selección óptima de pozos de monitoreo basado en un enfoque de selección multi-objetivo; se describe cómo evaluar los resultados de la calidad de agua, basado en el cálculo del índice de calidad de agua subterránea (WQI); se explica el proceso de reducción y selección de parámetros más relevantes para el monitoreo de calidad de agua subterránea. Posteriormente, se presentan las evaluaciones realizadas para validar el protocolo, considerando dos escenarios: utilizando datos reales geoespaciales y datos sintéticos creados de manera algorítmica.

En el Capítulo 4 se presentan y analizan los resultados obtenidos en la implementación del protocolo. Cada sección corresponde a cada uno de los aspectos abordados en la metodología: la selección de pozos óptimos, la evaluación de resultados con WQI, los parámetros relevantes y la validación del protocolo.

Finalmente, en el Capítulo 5 se presentan las conclusiones generales del trabajo, destacando los hallazgos más importantes y se discuten sus implicaciones. Además, se proponen áreas de investigación futura, relacionadas con el monitoreo de calidad de agua subterránea y la mejora continua del protocolo desarrollado.

Capítulo 2

Revisión de la Literatura

El monitoreo de calidad de agua consiste en recopilar información en lugares establecidos e intervalos regulares, permite evaluar la aceptabilidad del recurso vital para un uso particular de acuerdo a los parámetros físicos, químicos y bacteriológicos [30, 65]. Los objetivos del monitoreo consisten en caracterizar la condición ambiental, e identificar las tendencias espaciales o temporales de los parámetros medidos [21]. En este sentido, la planificación del monitoreo permite complementar y validar el procedimiento de evaluación de riesgo, conocer los cambios de la calidad del agua subterránea y evaluar tendencias [14].

De acuerdo con [65], el método tradicional para realizar el monitoreo de calidad de agua subterránea conlleva una serie de pasos. La persona encargada de la calidad del agua visita el pozo a monitorear y obtiene manualmente las muestras de agua. La muestra se lleva al laboratorio para conocer los valores de los parámetros de calidad de agua y posteriormente, establecer las medidas de control adecuadas, basadas en el análisis de los resultados del monitoreo. Los pasos mencionados son largos, complejos, costosos y muy deficientes, debido a todo los procesos involucrados en la selección de los pozos, la identificación de los contaminantes, el nivel de contaminación y la fuente de contaminante.

El agua potable segura es una necesidad y su demanda es creciente. En consecuencia, es necesario enfocar los esfuerzos en la planificación del monitoreo y la evaluación de calidad de agua subterránea basado sobre el conocimiento científico para gestionarlo de manera sostenible y eficiente [47]. [14] reconocen que el conocimiento científico disponible sobre las necesidades de la calidad del agua subterránea es inferior, en comparación a los ecosistemas acuáticos superficiales. Por una parte, se presentan métodos poco objetivos o difíciles de replicar, como lo mencionan en [15] y [64]. Por otra parte, se realizan esfuerzos para mejorar la manera en que se realiza el monitoreo de calidad de agua subterránea, cómo se

comparte la información y cómo se toman las decisiones en base a los resultados del monitoreo [27, 28, 32, 58].

En este contexto, se realizó una revisión de la literatura de los artículos científicos publicados en los últimos 10 años (2013 - 2023); el enfoque principal en los criterios de inclusión es el protocolo de monitoreo de calidad de agua subterránea, en uno o más de los siguientes aspectos: revisión o encuesta (survey), desarrollo de un marco o técnica para el monitoreo de calidad de agua subterránea, estudios empíricos o experimentales para evaluar la calidad del agua subterránea.

Los trabajos relevantes identificados se agrupan en: i) monitoreos de calidad de agua subterránea enfocados en atender problemas en un área de interés local, ii) revisiones de los estudios de calidad de agua subterránea y los métodos utilizados para evaluar los resultados del monitoreo, iii) tecnologías para el monitoreo automático de calidad de agua.

Los monitoreos enfocados en atender problemas en una área de interés local y diseñados según sus necesidades particulares [7, 23, 34, 39, 40], indican un número determinado de pozos y los parámetros monitoreados, sin embargo, no justifican la selección de los parámetros, no especifican los métodos utilizados para la selección de pozos o indican una selección aleatoria [7].

Los trabajos de revisiones bibliográficas identificados [14, 18, 41, 48], tienen como objetivo una evaluación exhaustiva de la literatura sobre la calidad del agua subterránea. En [48] identifican los métodos utilizados para comprender los resultados del monitoreo en los estudios realizados en los últimos años: técnicas estadísticas multivariadas como el análisis factorial, el análisis de componentes principales (PCA, por sus siglas en inglés), métodos gráficos para la hidrogeoquímica y sistemas de información geográfica (GIS, por sus siglas en inglés). Los trabajos identifican la necesidad de mejorar la determinación de los indicadores de monitoreo de agua subterránea y la frecuencia de monitoreo, mejorar el diseño de pozos de monitoreo, la precisión de los datos de monitoreo y la mejora de las técnicas de análisis de los resultados de calidad de agua subterránea [14, 18, 41].

Otros trabajos [9, 32, 37, 47], se enfocan en las técnicas actuales de interpretación para la evaluación de calidad de agua subterránea. En [47] proporcionan una revisión general de los principales Índices de Calidad del Agua subterránea (WQI) desarrollados a lo largo de los años y empleados para la evaluación de la calidad del agua subterránea en todo el mundo. En [32] presentan una revisión donde discuten los cinco pasos principales del desarrollo del Índice de Calidad del Agua Potable (DWQI, por sus siglas en inglés) para ayudar a los tomadores de

decisión a seleccionar los métodos de cálculo más adecuados para una aplicación local. En [37] evalúan la calidad del agua subterránea basada en GIS y WQI. En los trabajos mencionados concluyen que no existe un método universalmente aceptable para la generación de índices de calidad de agua subterránea, en los pasos para calcular los índices existe subjetividad e incertidumbre, en muchos casos de estudios no se proporcionan las razones de la selección de parámetros. Además, resaltan la necesidad de realizar más estudios destinados a mejorar el método tradicional de monitoreo para evitar los inconvenientes de los costos elevados, el diseño complejo y la dificultad logística de monitorear todos los componentes del agua potable.

Los trabajos sobre el monitoreo automático de calidad de agua son revisiones bibliográficas [27, 30, 65], aunque no tratan exclusivamente sobre las aguas subterráneas. Se identificaron las tecnologías existentes para la implementación de redes de monitoreo considerando la recopilación de datos, la comunicación, el análisis de datos y la alerta temprana.

En [65] y [30] presentan una evaluación exhaustiva de la literatura sobre la tecnología de la Internet de las cosas (IoT, por sus siglas en inglés) para la gestión de la calidad del agua. En [65], también identifican los algoritmos de aprendizaje automático utilizados para predecir la calidad del agua en base a la información recolectada, indican las limitaciones en las técnicas de tratamiento para los datos faltantes del monitoreo y la técnica inexistente sobre la cantidad óptima de sitios a monitorear. En [27] describen las tecnologías disponibles asociadas a la redes inalámbricas (WSN, por sus siglas en inglés) y los sensores remotos (RS, por sus siglas en inglés) para la calidad del agua, la arquitectura de la red de transmisión de datos, identifican el número de parámetros limitados para monitorear de forma automática y en tiempo real. Los desafíos identificados son la inexistencia de métodos estándar internacionales reconocidos universalmente y aceptados para la selección de parámetros de calidad de agua, la necesidad de la investigación de modelos de optimización de las frecuencias de muestreo y los diseños de redes de monitoreo adecuados para aplicar de forma global.

Los trabajos mencionados [27, 30, 65], coinciden en que todavía existen puntos importantes que obstaculizan el avance en esta área de investigación; por ejemplo, se requieren diseños electrónicos especializados para la implementación del monitoreo automático, mejorar la facilidad de implementación, utilizar tecnologías menos costosas y más confiables. Además, los costos de operación del sistema considerando el factor del terreno, la distancia, el medio ambiente, el clima, entre otros [27].

Las soluciones presentadas sobre el monitoreo y la evaluación de calidad de agua subterránea, utilizan una variedad de métodos, desde las convencionales hasta las más modernas, con amplias consideraciones subjetivas. Como lo señalan [27], no existe un programa modelo universal para el monitoreo de calidad de agua, por lo tanto, se están produciendo diferentes resultados de evaluación para el mismo objetivo, limitando la interpretación correcta. La revisión realizada por [47], evidenció la necesidad de desarrollar algún tipo de protocolo para monitorear la calidad de las aguas subterráneas, a fin de tener una evaluación comparativa del grado de vulnerabilidad del acuífero en diferentes partes del mundo.

Las necesidades reveladas en esta área de investigación conducen a establecer un monitoreo integral y de utilidad para la evaluación objetiva de la calidad del agua subterránea. Para ello, es fundamental optimizar la distribución espacial de los puntos de monitoreo, mejorar la manera de determinar los indicadores o parámetros a considerar y mejorar las técnicas del análisis de resultados de calidad de agua subterránea [14, 18, 27, 41].

Los puntos de monitoreo seleccionados correctamente mejoran la precisión y la integridad de la calidad del agua, permiten reducir los costos de los programas de monitoreo; por ello, surge la necesidad de utilizar técnicas avanzadas de selección [27, 47]. En [27] resaltan la inexistencia de métodos estándar internacionales reconocidos universalmente y aceptados para la selección de parámetros de calidad de agua. El monitoreo de todos los parámetros de calidad de agua aumenta la carga de trabajo y diluye los factores que determinan la calidad del agua. Generalmente, los parámetros son seleccionados considerando los estándares establecidos por los entes regulatorios de cada país u organizaciones internacionales, los parámetros relacionados con la salud o según el juicio de los tomadores de decisiones. Sin embargo, surge la necesidad de establecer una técnica avanzada de selección.

También, en los trabajos revisados se identifica la importancia de mejorar las capacidades de interpretación para la evaluación de la calidad del agua subterránea en diferentes escalas espaciales. En este sentido, la integración de la tecnología del sistema de información geográfica (GIS) con la inteligencia artificial, y las técnicas multivariadas de estadísticas y geoestadísticas, son herramientas más poderosas que los métodos tradicionales para una mejor evaluación de la calidad del agua subterránea [47].

Finalmente, los trabajos coinciden en la necesidad de la investigación sobre la calidad del agua subterránea; abogar por mejorar los marcos de monitoreo de agua subterránea, aumentar las inversiones en investigación por parte de los gobiernos y fortalecer la colaboración entre las instituciones del gobierno, universidades, individuos, industrias, organizaciones internacionales y nacionales [14, 18, 41, 47].

Capítulo 3

Metodología

A continuación, se presenta la metodología utilizada para proponer un protocolo sistemático para el monitoreo de calidad de agua subterránea, utilizando una aplicación integral de tecnologías emergentes con herramientas y técnicas avanzadas de información. El protocolo propuesto se realiza sobre un caso de estudio del acuífero Patiño; se utilizan datos reales, desarrollados y obtenidos en el marco de un proyecto de investigación ¹.

Las tecnologías emergentes son un conjunto de técnicas y conocimientos que permiten el aprovechamiento práctico del conocimiento científico, caracterizadas por ser soluciones novedosas que pueden ser aplicadas en diferentes dominios, poseen un crecimiento relativamente rápido y cierto grado de coherencia que persiste a lo largo del tiempo [57].

El uso de tecnologías emergentes como (i) la aplicación de algoritmos multi-objetivos vinculados con el uso de datos y herramientas geoespaciales; (ii) la aplicación de técnicas de reducción multivariada y algoritmos de selección de atributos vinculados con la indexación de la calidad del agua; pueden resultar beneficiosas en el dominio del monitoreo de calidad de agua subterránea.

3.1 Caso de estudio: Acuífero Patiño

El acuífero Patiño se encuentra localizado en el Departamento Central del Paraguay, correspondiente a la zona urbana más grande y densamente poblada que abarca la capital del país, Asunción, y las demás ciudades aledañas (Figura 3.1). El acuífero posee una extensión total de 1173 km² y es una de las principales fuentes de suministro de agua de la zona, abastece a gran parte del sector comercial, industrial

¹Proyecto 14-INV-190 “Monitoreo y simulación de transporte de contaminantes en zonas urbanas del acuífero Patiño”, cofinanciado por el Consejo Nacional de Ciencia y Tecnología - CONACYT con recursos del FEEI

y agropecuario.

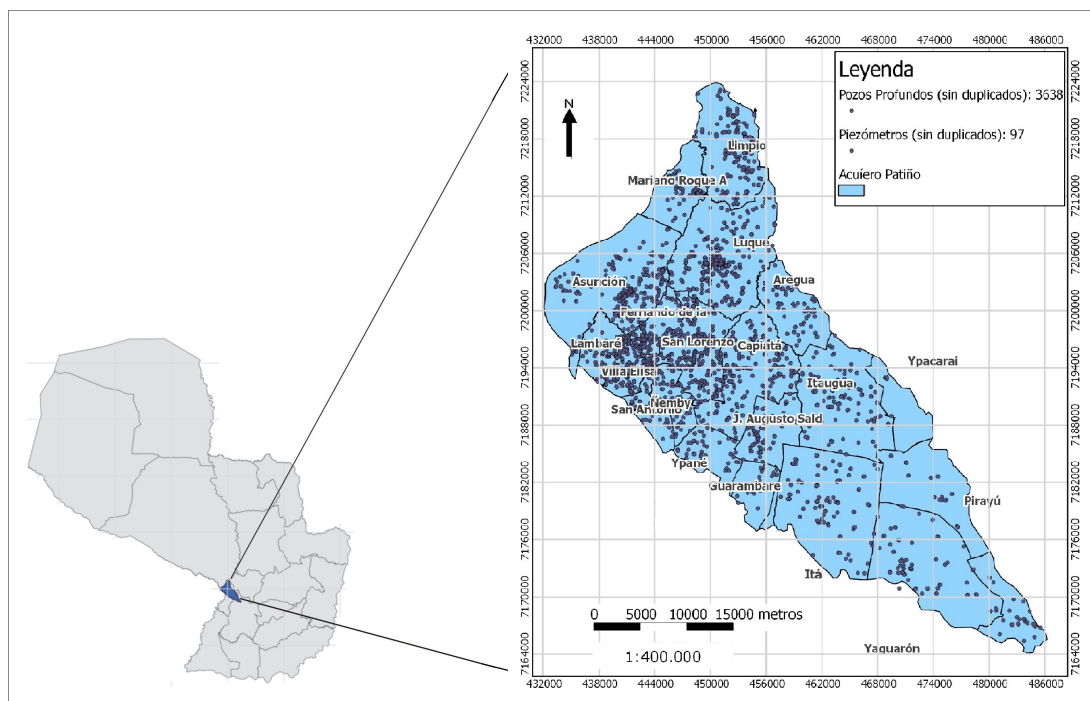


Figura 3.1: Mapa de ubicación del acuífero Patiño, Departamentos Central y Paraguarí, Paraguay.

El área metropolitana de Asunción (AMA) ha experimentado un crecimiento urbano exponencial en las últimas décadas, produciendo un profundo impacto sobre la recarga y la calidad del agua subterránea. En dicha zona, el acuífero Patiño se extiende en unos 567 km², abastece de agua potable a cerca de 2,1 millones de habitantes. Se conocen más de 2800 pozos de extracción, se estima que existen más de 8000 fuentes potenciales de contaminación [24], y el 79% de las viviendas disponen de pozos ciegos o fosa sépticas como método de evacuación de las aguas residuales [24].

Por ello, los vertidos urbanos y la sobreexplotación del acuífero Patiño son una amenaza constante para la calidad de sus aguas. Además, la escasa información sobre su calidad, dado que no existe un seguimiento constante, dificulta cualquier tipo de gestión proactiva. Los pocos estudios existentes proporcionan un conocimiento limitado de la calidad y cantidad de agua del acuífero. En general, estos estudios se han centrado en la búsqueda de focos de contaminación sin otro protocolo claro sobre el muestreo de los pozos disponibles.

Por ejemplo, en [29] tomaron muestras de 90 pozos existentes, que cubrían toda el área del acuífero en zonas cercanas a las estaciones de servicios, para analizar los niveles de metil tert-butil éter (MTBE) y productos relacionados.

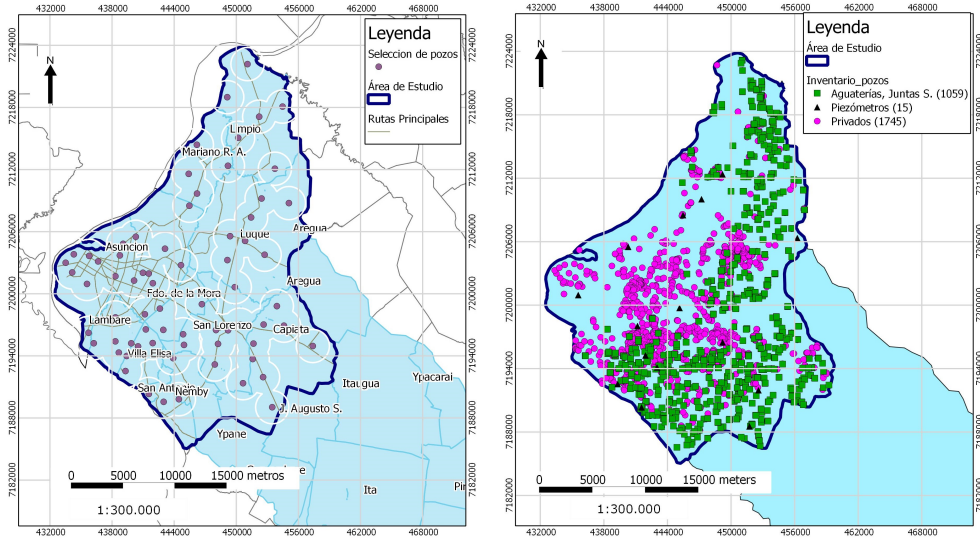
Detectaron MTBE en el 44% de los pozos, alcohol tert-butílico (TBA) en el 21% de los pozos y formiato de tert-butilo (TBF) en el 13% de los pozos. Otro estudio, realizado por la consultora INCLAM-HQA [35], muestreó 36 piezómetros existentes, cubriendo toda la zona del acuífero, los parámetros evaluados fueron nivel estático, y 19 parámetros fisicoquímicos y bacteriológicos. Los valores de nitrato medidos superaron el límite máximo en el 25% de las muestras recogidas. Además, se observó un aumento de la salinidad en algunos puntos de control. El objetivo de estas campañas de muestreo era categorizar la calidad del agua del acuífero. En [33] muestrearon 65 pozos existentes, cubriendo 79 km², en la cuenca hídrica del arroyo San Lorenzo, donde buscaron la presencia de parámetros físico-químicos, bacteriológicos y metales pesados. Detectaron coliformes totales en el 38% de las muestras, E. Coli en el 9% de las muestras y nitratos en el 17% de las muestras, lo que evidenciaba prácticas de saneamiento inadecuadas. Estos tres estudios se centraron en caracterizar la zona, pero sin un protocolo claro sobre cómo se seleccionaron los pozos, aparte de la heurística general de la facilidad de acceso a los pozos. La premisa que surge a partir de estos estudios es, si se eligieron las ubicaciones óptimas de los pozos para cumplir los objetivos propuestos o si se podría haber hecho un uso más eficiente de los recursos para llevar a cabo la campaña de muestreo.

El objetivo de este trabajo es diseñar un protocolo sistemático para el monitoreo de calidad de agua subterránea basado en un enfoque multi-objetivo, utilizando el caso de estudio indicado para su posterior uso en otras áreas de interés.

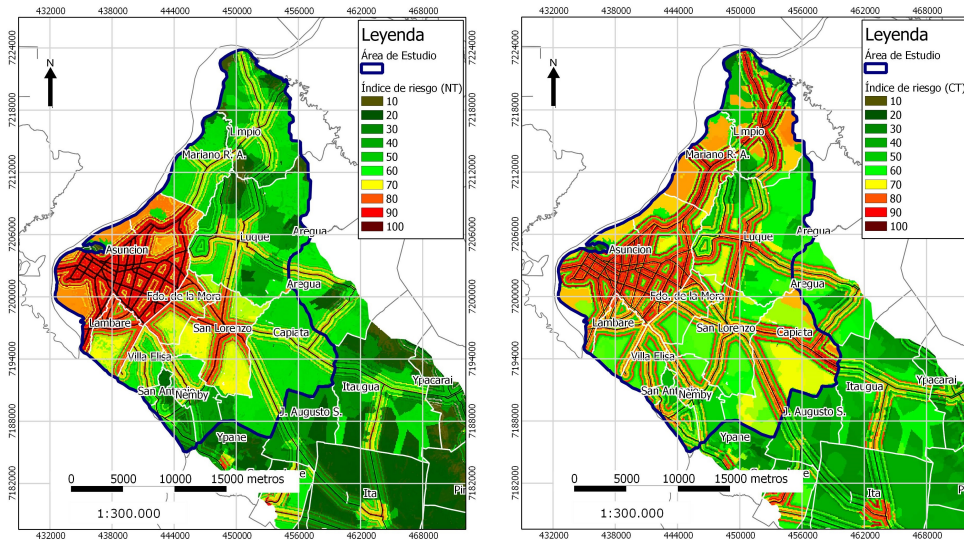
Los datos geoespaciales utilizados en este trabajo, corresponden a datos reales y existentes sobre el acuífero Patiño, específicamente en el área urbana (AMA). Los datos geoespaciales se visualizan en los mapas de la Figura 3.2, corresponden al área de cobertura en el AMA (567 km²) (Figura 3.2a), el inventario de pozos con 2819 puntos (Figura 3.2b), y los índices de riesgos de contaminación por las concentraciones de Nitrógeno Total (INT) (Figura 3.2c) y Coliformes Totales (ICT) (Figura 3.2c).

3.2 Selección óptima de pozos de monitoreo

La selección multi-objetivo de los pozos de monitoreo se basa en la aptitud de cuatro funciones objetivo, evaluadas a través del Algoritmo Genético de Clasificación No Dominada II (NSGA-II, por sus siglas en inglés), propuesto por [22] y la adición de un orden de preferencia (PO, por sus siglas en inglés), propuesto por [26].



(a) Mapa del área de cobertura a considerar (567 km²). (b) Mapa del inventario de pozos realizado en [12].



(c) Mapa del Índice de riesgo por Nitrógeno total (INT) realizado en [13]. (d) Mapa del Índice de riesgo por fósforo total (ICT) realizado en [13].

Figura 3.2: Datos geoespaciales considerados para la selección óptima de pozos de monitoreo.

El algoritmo NSGA-II es ampliamente utilizado y validado para diseñar redes de monitoreo [8, 16, 25]. El PO, que otorga una mayor especificidad en la selección de las posibles soluciones, no se ha utilizado antes en este tipo de aplicaciones. Una de nuestras contribuciones es la adición de este proceso para ayudar a lidiar con el problema de tener múltiples objetivos en conflicto, donde la mejora en un objetivo disminuye el desempeño en algunos otros objetivos [38].

El algoritmo asigna los valores de aptitud a cada posible solución, dada una

población de soluciones. Además, posee una funcionalidad de preservación de la diversidad mediante el cálculo de la distancia de *crowding* entre soluciones. Su objetivo principal es encontrar un conjunto de soluciones óptimas llamado “frente pareto óptimo”. Sin embargo, el problema del proceso de la clasificación del NSGA-II ocurre cuando se consideran más de tres objetivos. El número de soluciones que se clasifican en el frente óptimo sólo depende de la distancia de *crowding*, es decir la diversidad.

La distancia de *crowding* pierde su eficiencia al aumentar las dimensiones del problema ya que se basa en medir cuboides con respecto a los individuos más cercanos, lo que solo funciona bien con dos objetivos [3]. Para evitar este problema, en este trabajo se implementó una modificación del NSGA-II, basada en el método de [26], denominada Orden de Preferencia (PO). Con la implementación del método PO se comparan dos soluciones para establecer en cuántos objetivos es mejor una solución frente a la otra. Por ejemplo, la solución A se clasifica por encima de la solución B, si se desempeña mejor en más objetivos que la solución B y viceversa. El esquema general del algoritmo aplicado se muestra en la Figura 3.3.

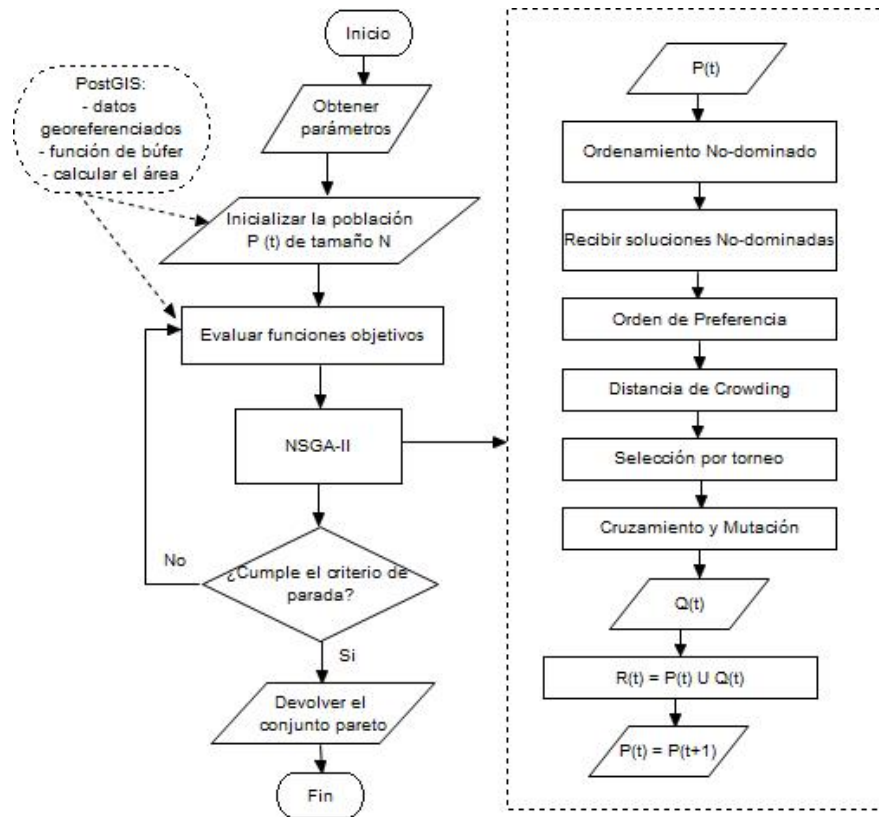


Figura 3.3: Diagrama de flujo del método de resolución computacional implementado para la selección óptima de pozos. El diagrama de flujo muestra el proceso de evaluación de cada solución basado en el algoritmo NSGA-II.

Para la implementación, los datos de entrada se almacenaron en una base de datos con extensión espacial PostGIS [53], dedicada a la manipulación de datos georreferenciados.

Los datos de entrada del algoritmo consisten en información de 2819 registros de pozos, que contienen los siguientes valores: índice de riesgo de contaminación por nitrógeno (INT), índice de riesgo de contaminación por coliformes totales (ICT), tipo de pozo (aguatería, junta de saneamiento, piezómetros o pozo privado) y las coordenadas del pozo (X, Y). Se establece una cantidad de 66 pozos para la selección óptima.

Las cuatro funciones objetivo consideradas son: (i) maximizar las concentraciones de INT, (ii) maximizar las concentraciones de ICT, (iii) maximizar el área de cobertura y (iv) maximizar la selección de pozos de acceso público, como piezómetros, aguaterías y juntas de saneamiento.

NSGA-II se implementa siguiendo la propuesta de [22]. El algoritmo inicia con una población inicial de N soluciones posibles, cada solución contiene 66 ubicaciones de pozos seleccionadas aleatoriamente con sus respectivos datos de INT, ICT, tipo de pozo y área de cobertura. Estas soluciones se evalúan por su aptitud (*fitness*), se clasifican en orden y se utilizan para crear una nueva solución de población hijo, a través de operadores de cruzamiento y mutación. Las poblaciones de padres e hijos se combinan y las N soluciones más aptas pasan a la siguiente generación. Para seleccionar las soluciones más aptas, el orden de no dominancia y un ranking PO son asignados a cada solución siguiendo la propuesta de [26].

3.2.1 Definición de las funciones objetivo

El problema multi-objetivo para la selección de pozos de monitoreo de calidad de agua subterránea se define de la siguiente manera:

$$F(x) = (f_1(x), f_2(x), f_3(x), f_4(x)) \quad (3.1)$$

donde,

- i. $f_1(x)$: Maximizar los pozos con alto riesgo de contaminación por Nitrógeno Total.
- ii. $f_2(x)$: Maximizar los pozos con alto riesgo de contaminación por Coliformes Totales.
- iii. $f_3(x)$: Maximizar el área de cobertura de los 66 pozos seleccionados.

iv. $f_4(x)$: Maximizar los pozos de acceso público.

donde x es un vector de n posibles ubicaciones de pozos que son considerados. En este trabajo $n = 66$. Las distintas funciones objetivo se definen como sigue:

i. Maximizar el índice de riesgo de contaminación por la concentración de Nitrógeno Total: este objetivo maximiza el total acumulado de los índices de riesgo de contaminación por la concentración de NT de la siguiente manera:

$$f_1(x) = \sum_{i=1}^n INT_i \quad (3.2)$$

donde INT_i es el valor del índice de riesgo NT localizado en el pozo i y n es el número total de pozos que se seleccionarán como posible solución.

ii. Maximizar el índice de riesgo de contaminación por la concentración de Coliformes Totales: este objetivo maximiza el total acumulado de los índices de riesgo de contaminación por la concentración de CT de la siguiente manera:

$$f_2(x) = \sum_{i=1}^n ICT_i \quad (3.3)$$

donde ICT_i es el valor del índice de riesgo CT localizado en el pozo i y n es el número total de pozos que se seleccionarán como posible solución.

iii. Maximizar el área de cobertura: este objetivo maximiza el área de cobertura total, suponiendo que cada pozo tiene un radio de influencia de 1600 m, que se obtiene dividiendo el área total (567 km²) por el número de pozos que se pueden muestrear (66 pozos). El cálculo de esta función tiene en cuenta las zonas de cobertura que pueden solaparse para evitar un doble recuento. La función de optimización se define como

$$f_3(x) = \sum_{i=1}^n A_i \quad (3.4)$$

donde A_i es el área ajustada para el pozo i y n es el número total de pozos a seleccionar como posible solución.

iv. Maximizar los pozos de acceso público: este objetivo considera maximizar la selección de pozos que son de dominio público, como las aguaterías, las juntas de saneamiento, los piezómetros o los pozos exploratorios. Estos tienen mayor prioridad que los pozos de propiedad privada. El modelo considera maxi-

mizar la prioridad de los pozos a seleccionar, de la siguiente manera:

$$f_4(x) = \sum_{i=1}^n Ip_i \quad (3.5)$$

$$Ip_i = \begin{cases} 1 \\ 0 \end{cases} \quad (3.6)$$

donde Ip_i es el nivel de prioridad del pozo. La prioridad 1 es para los pozos de acceso público y 0 para los de propiedad privada.

3.3 Evaluación de los resultados de calidad de agua del monitoreo

Luego de seleccionar los pozos de monitoreo, se procede al trabajo de campo para el muestreo de los pozos seleccionados, siguiendo las buenas prácticas para la toma de muestras de agua subterránea y su posterior transporte al laboratorio de CEMIT². Inicialmente, son considerados un total de 21 parámetros analizados en laboratorio y 66 registros de pozos.

Los 21 parámetros analizados en el laboratorio son los siguientes: alcalinidad total, arsénico, bicarbonato, calcio, carbonato, cloruro, cobre, conductividad, cromo total, sulfato, coliformes fecales, magnesio, manganeso, materia orgánica, mercurio, nitratos, nitritos, nitrógeno amoniacal, pH, potasio y sodio.

Luego de obtener los resultados de los análisis de laboratorio de calidad de agua, se consideran las evaluaciones detalladas a continuación.

En primer lugar, los parámetros analizados se comparan con tres normas o estándares de agua potable³ para tener una perspectiva más amplia de la idoneidad de las aguas subterráneas.

Según [62], las normas de potabilidad son directrices que describen requisitos mínimos razonables (límites permisibles) y se definen mediante valores numéricos de los componentes del agua, que deben cumplirse para proteger la salud de los consumidores. A la hora de definir los estándares, se tienen en cuenta las condiciones medioambientales, económicas y culturales locales. Es decir, cada

²CEMIT - Centro Multidisciplinario de Investigaciones Tecnológicas

³ERSAN (2000): Ley N 1614/2000 “Ley general del marco regulatorio y tarifario del servicio de agua potable y alcantarillado Sanitario de la República del Paraguay”. WHO (2017): “*Guidelines for drinking water quality: fourth edition incorporating first addendum*”. Geneva: World Health Organization. EU (1998): *Directive related with quality of water intended for human consumption. European Directive 98/83/EC*.

regulador local puede tener niveles permisibles diferentes en función de las limitaciones socio-económicas y medioambientales. Por lo tanto, lo que puede ser adecuado en una región puede no serlo en otra.

En cambio, la calidad del agua se refiere a un concepto global que permite un formato estandarizado para comparar distintas regiones. En este contexto, el Índice de Calidad del Agua (WQI), propuesto por [45], responde a la aptitud general del agua potable, independientemente de la normativa. El WQI también otorga una indicación del nivel de tratamiento que puede ser necesario o de la fragilidad del agua. En cambio, las normas de calidad del agua son binarias y sólo ofrecen una etiqueta de apta o no apta.

El WQI es el método más conveniente para determinar el estado de un cuerpo de agua en particular porque permite comprender el impacto de todos los parámetros en su conjunto [20]. De todas las técnicas de indexación, es la más eficiente y de mucha utilidad para la gestión de los recursos hídricos [4, 44, 50, 52]. Por lo tanto, la metodología propuesta para evaluar los resultados de calidad de agua considera el cálculo WQI, indicado por [46], e integrado con las técnicas de reducción multivariada y algoritmos de selección de atributos, para la modificación de los parámetros evaluados en el cálculo.

El cálculo WQI [46], considera los diez parámetros siguientes: pH, sólidos totales disueltos, cloruro, sulfato, sodio, potasio, calcio, magnesio, dureza total y nitratos.

En primer lugar, se asigna el peso unitario (w_i) a cada parámetro considerado, varía entre 2 y 5, basado en sus efectos sobre la salud (Tabla 3.1). A continuación, el peso relativo para parámetro se calcula mediante la Ecuación (3.7):

$$W_i = \frac{w_i}{\sum_{i=1}^n w_i} \quad (3.7)$$

donde W_i es el peso relativo, w_i es el peso de cada parámetro y n es el número de parámetros. El peso w_i , los valores calculados del peso relativo (W_i) y el estándar para cada parámetro se indican en la Tabla 3.1.

El siguiente paso consistió en calcular la escala de valoración de la calidad de cada parámetro (q_i) mediante la Ecuación (3.8):

$$q_i = \frac{|C_i - I_i|}{S_i - I_i} \times 100 \quad (3.8)$$

donde q_i es la calificación de calidad, C_i es la concentración que corresponde al i -ésimo parámetro en un sitio de muestreo determinado, I_i es el valor ideal del i -ésimo parámetro en el agua potable y S_i es la norma de potabilidad para el

Tabla 3.1: Pesos unitarios y pesos relativos para cada parámetro WQI.

Parámetros	Estándar	Peso unit. (w_i)	Peso relativo (W_i)
pH	6.5 - 8.5*	4	0.105
STD (mg/L)	1000*	5	0.132
Cloruro (mg/L)	250	5	0.132
Sulfato (mg/L)	250	5	0.132
Sodio (mg/L)	200	4	0.105
Potasio (mg/L)	12*	2	0.053
Calcio (mg/L)	100*	3	0.079
Magnesio (mg/L)	50*	3	0.079
Dureza total (mg/L)	400*	2	0.053
Nitrato (mg/L)	45*	5	0.132
		$\sum w_i = 38$	$\sum W_i = 1$
Valores de pesos considerados por la metodología propuesta por [46] y complementada con [60].			
(*) Valores establecidos en la norma ERSSAN (2000).			

i-ésimo parámetro. Cabe señalar que el pH es el único parámetro que tiene un valor ideal y que para los demás parámetros el valor ideal se fija en cero (es decir, $I_{pH} = 7$, e $I_i = 0$ para todos los demás parámetros).

Para calcular el WQI, se determina el subíndice (S_i) para cada parámetro mediante la Ecuación (3.9), que se utiliza a continuación para determinar el WQI según la Ecuación (3.10):

$$SI_i = W_i \times q_i \quad (3.9)$$

$$q_i = \frac{|C_i - I_i|}{S_i - I_i} \times 100 \quad (3.10)$$

donde SI_i es el subíndice del i-ésimo parámetro, W_i es el peso relativo de cada parámetro, q_i es la calificación de calidad, y n es el número de parámetros.

Los resultados de WQI se clasifican en cinco categorías para el consumo humano: 0-50 (excelente), 50-100 (buena), 100-200 (deficiente), 200-300 (muy deficiente), +300 (no apta para beber). El índice se calcula y mapea con las herramientas GIS para visualizar la variabilidad espacial.

3.4 Reducción y selección de parámetros de calidad de agua

La reducción se aplica a los parámetros del monitoreo, con el objetivo de disminuir el número de evaluaciones, reducir el sesgo y tomar decisiones más objetivas, a su vez, reducir el tiempo, el esfuerzo y los costos requeridos para monitorear un gran número de variables.

El primer paso consiste en la recolección y preprocesamiento de datos. El conjunto de datos inicial o variables originales, corresponden a los parámetros y las concentraciones de calidad de agua, de campañas o muestreos de monitoreos pasados. Además, es necesario realizar el preprocesamiento de datos para eliminar los atributos redundantes, constantes, identificadores; y normalizar los atributos, debido a que poseen diferentes rangos de valores y unidades de medidas.

El segundo paso consiste en generar varios subconjuntos de parámetros (atributos) obtenidos del conjunto inicial, con la finalidad de considerar otros atributos también relevantes, capaces de generar un mejor modelo. Por lo tanto, es necesario reducir la dimensionalidad del conjunto inicial sin alterar las características principales, mediante un espacio de baja dimensión que contenga la mayor parte de la información del espacio original de alta dimensión. El método considerado es el enfoque integrado de múltiples técnicas como el Análisis de Componentes Principales (PCA), los algoritmos de selección de atributos y el análisis predictivo considerando la clasificación en función al WQI.

El PCA es una de las técnicas más utilizadas para identificar los parámetros significativos de un gran conjunto con múltiples variables, se analiza en detalle en [36]. Otros autores [6, 6, 15, 59], valoran su eficiencia en la utilización de datos medioambientales. El PCA transforma linealmente las variables originales a nuevas variables no correlacionadas conocidas como componentes principales (PC, por sus siglas en inglés). Para ello, primero se calcula la matriz de covarianza, que indica la medida de dispersión conjunta entre las variables y se obtienen los autovectores y autovalores. La varianza de la PC es igual al autovalor más grande de la matriz de covarianza, indicando la variación total explicada por las primeras PC.

Seguidamente, los subconjuntos de parámetros son seleccionados. En PCA, no solo se reduce la dimensión del conjunto de datos al seleccionar un número de PC, sino también se reduce el número de variables que se consideran en el conjunto de datos al retener solo x variables; para lo cual existen diversos criterios de selección de variables (VSC, por sus siglas en inglés) para la interpretación de

las covarianzas de PC. En este trabajo, se consideran los VSC indicados en [6], y también se propone una nueva selección que considera las variables más repetidas en los subconjuntos obtenidos. Las implementaciones son realizadas mediante las librerías scikit-learn y scipy de Python.

Además, otros subconjuntos de parámetros son seleccionados, basados en algoritmos de Selección de atributos. Estos reducen la dimensionalidad, al presentar un ranking de características en función a su relevancia predictiva, creando modelos que describen mejor el vector de clases observado. Los algoritmos de selección de atributos se aplican utilizando Weka (Waikato Environment for Knowledge Analysis) [31], software libre de aprendizaje automático y minería de datos. Los modelos se realizaron con los algoritmos: Symmetrical Uncert Attribute Eval, Cfs Subset Eval, Info Gain Attribute Eval, Correlation Attribute Eval, Gain Ratio Attribute Eval, OneR Attribute Eval y ReliefF Attribute Eval; evaluados con validación cruzada sobre $k=10$.

El tercer paso consiste en verificar la efectividad de cada subconjunto seleccionado, realizando el análisis predictivo considerando una clasificación supervisada en función al WQI. Los algoritmos clasificadores considerados son: árboles de decisión J48 [31, 54] y Random forest, son modelos predictivos basados en reglas; Naive Bayes [31] es un modelo predictivo basado en probabilidades; y el algoritmo KNN (IBk) [31] es un modelo predictivo basado en similitudes. Cada modelo se evalúa con validación cruzada, $k=10$. Las evaluaciones mencionadas, también son realizadas con el software Weka [31].

Por otra parte, para poder analizar con solvencia los algoritmos se tienen en cuenta el porcentaje de aciertos del clasificador y los niveles de:

- Precisión: mide la relación de positivos reales identificados por los algoritmos, en relación con la población total de positivos reales y erróneos identificados por los algoritmos.
- Sensibilidad o *recall*: mide la relación de positivos reales identificados por los algoritmos, en relación con la población total real de positivos el conjunto de datos.

Los modelos realizados aportan un nuevo conocimiento y conducen a un mejor entendimiento de la distribución de los datos. El modelo con mejor desempeño, analizando los factores objetivos, es seleccionado y se identifican los parámetros más significativos a ser propuestos en el muestreo de la calidad del agua subterránea y para el cálculo WQI.

3.5 Validación del protocolo

La validación consiste en aplicar el protocolo sistemático propuesto considerando diferentes escenarios para:

- i. La selección óptima de pozos de monitoreo y la validación comparando dos campañas distintas de muestreo de un mismo escenario.
- ii. El cálculo WQI inicial y la validación de los parámetros relevantes en función a la clase.

La información inicial necesaria son los datos de entrada: el índice de riesgo de contaminación por NT, índice de riesgo de contaminación por CT, los puntos georreferenciados de los pozos con los atributos: coordenadas del pozo (X, Y), tipo de pozo (aguatería, junta de saneamiento, piezómetros o pozo privado), y los parámetros o atributos de calidad de agua subterránea.

El protocolo sistemático para el monitoreo de calidad de agua subterránea se aplica a los escenarios presentados a continuación.

3.5.1 Escenario I con datos reales geoespaciales

Los datos reales geoespaciales utilizados corresponden al acuífero Barreiras-Brasil, obtenidos del Catálogo de Metadatos ANA (Agencia Nacional de Aguas y Saneamiento Básico, Brasil) [19]. En este escenario se considera un área de estudio de 567 km², de igual dimensión al área urbana del acuífero Patiño, presentada en la Sección 3.2.

Los datos geoespaciales utilizados del acuífero Barreiras se muestran en la Figura 3.4, correspondientes al área de 567 km² y la distribución de 2657 pozos, correspondientes a 294 públicos y 2363 privados (Figura 3.4a), el área de cobertura máxima a cubrir sobre el acuífero (Figura 3.4b), y los índices de riesgos de contaminación por las concentraciones de Nitrógeno Total (INT) (Figura 3.4c) y Coliformes Totales (ICT) (Figura 3.4c).

Los parámetros de calidad de agua del acuífero Barreiras corresponden a dos campañas de muestreos realizados en los años 2017 y 2018, también obtenidos de [10]. En la “campaña 1” del 2017 fueron muestreados 72 pozos y en la “campaña 2” del 2018 fueron muestreados 70 pozos. En total se consideran 20 parámetros de calidad de agua: aluminio, arsénico, bicarbonato, calcio, carbonato, cloruro, coliforme fecal, conductividad hidráulica, dureza total, hierro, magnesio, manganeso, nitrato, nitrito, nitrógeno amoniacal, pH, potasio, sodio, sólidos totales disueltos y sulfato.

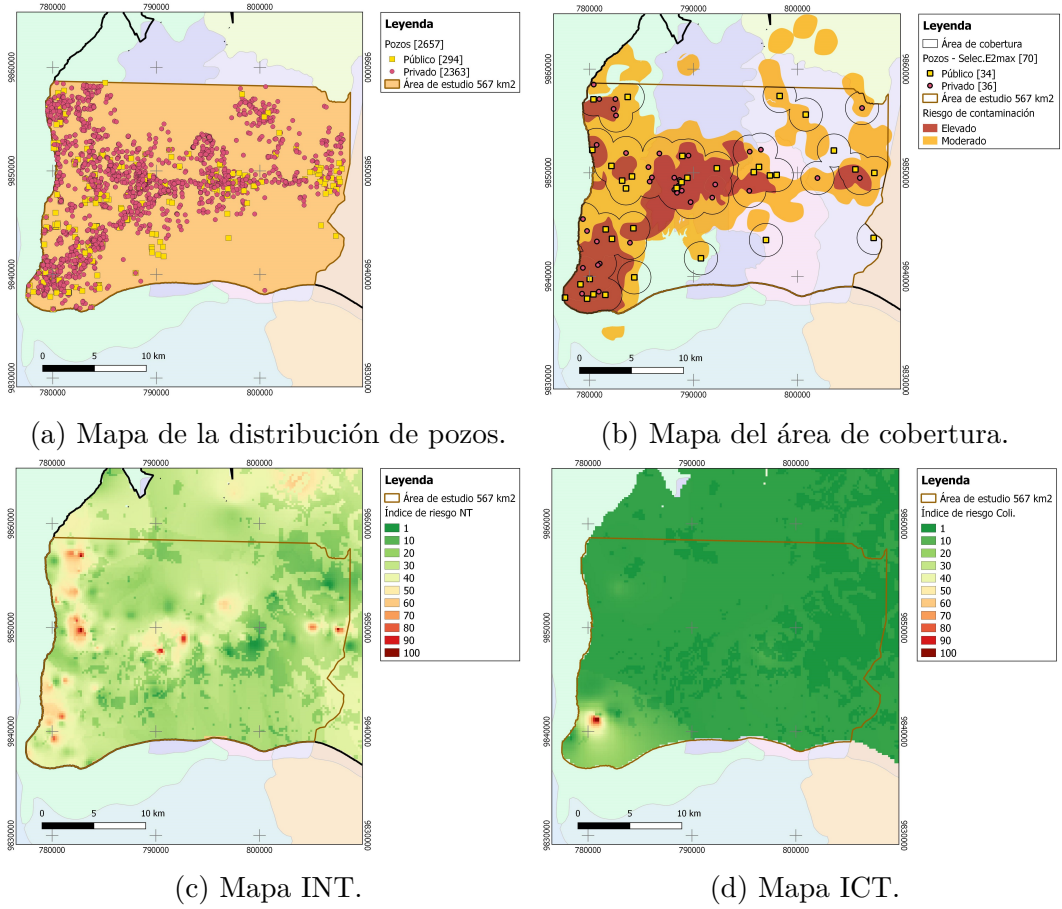


Figura 3.4: Datos geospaciales considerados para el Escenario I (567 km²), obtenidos de [19].

Validación de la selección óptima de pozos de monitoreo

Primeramente, se realiza la selección óptima de pozos de monitoreo del Escenario I, aplicando el NSGA-II con PO, siguiendo la metodología presentada en la Sección 3.2 y obtener un conjunto de 70 “pozos óptimos” seleccionados.

Posteriormente, la validación de la selección de pozos consiste en comparar la eficiencia de dos campañas distintas. Una campaña en donde se considera la implementación de la selección óptima de pozos de monitoreo, frente a otra campaña donde no se implementa un protocolo específico para la selección de pozos de monitoreo. Las campañas comparadas y sus consideraciones son las siguientes:

- NSGA-II PO en Escenario I: se considera la distribución del conjunto de 70 “pozos óptimos” y se simula un muestreo posterior al protocolo de selección de pozos, utilizando la “campaña 2” existente [10]. Los datos de cada parámetro son interpolados. De esta manera, se consiguen la información

de parámetros de calidad de agua en los “70 pozos óptimos”.

- Barreiras Campaña 1: se considera la “campaña 1” y su selección de 72 pozos [10], es decir, sin la implementación o simulación de un protocolo de selección óptima de pozos.

Finalmente, para comparar la efectividad de ambas campañas y sus respectivas selecciones de pozos, se ejecuta la selección de pozos de la “campaña 1” a través del algoritmo y se calcula la puntuación para cada función objetivo. La efectividad de cada campaña se calcula, considerando cuántos de los pozos muestreados poseen parámetros fuera de los límites permisibles de potabilidad del agua. La eficiencia del protocolo se define como la relación entre el número de pozos con parámetros fuera del rango permisible y el número total de pozos muestreados.

Validación de los parámetros relevantes en función al WQI

El cálculo WQI inicial se realiza con los datos de la “campaña 1” y siguiendo toda la metodología presentada en la Sección 3.3. Posteriormente, se validan los diferentes conjuntos de parámetros relevantes en función a la clase WQI inicial.

3.5.2 Escenario II con datos sintéticos

Correspondiente a datos sintéticos⁴ de una “muestra artificial” de parámetros de calidad de agua, de 100 instancias extendiendo la tabla XOR con 15 atributos de entrada de la siguiente manera: atributos aleatorios a partir de una distribución gaussiana, aplicando un filtro de rangos diferentes y considerando la correlación entre algunos atributos. Se consideran rangos de concentraciones por encima de los límites permisibles de los estándares de calidad de agua; de esta manera, se simula una contaminación de las aguas subterráneas.

La herramienta utilizada para crear los datos sintéticos es NumPy, una biblioteca de código abierto para el lenguaje de programación Python.

Los datos sintéticos creados corresponden a los siguientes 15 parámetros de calidad de agua: bicarbonato, calcio, carbonato, cloruro, coliforme fecal, conductividad hidráulica, dureza total, magnesio, nitrato, nitrógeno amoniacal, pH, potasio, sodio, sólidos totales disueltos y sulfato.

⁴Información generada de manera algorítmica en lugar de ser producida por eventos del mundo real. Esta información se personaliza según necesidades específicas.

Validación de los parámetros relevantes en función al WQI

El cálculo WQI inicial se realiza con los datos sintéticos y siguiendo toda la metodología presentada en la Sección 3.3. Posteriormente, se validan los diferentes conjuntos de parámetros relevantes en función a la clase WQI inicial.

Capítulo 4

Resultados y Análisis

A continuación se presentan los resultados obtenidos al implementar el protocolo sistemático propuesto para el monitoreo de la calidad de agua subterránea del acuífero Patiño, es su área urbana (567 km²). Estos resultados representan el fruto de un meticuloso proceso de trabajo de campo, recolección de datos y análisis que tiene como objetivo evaluar con objetividad y precisión la condición de los recursos hídricos subterráneos. A través de un enfoque riguroso y científico, esta sección proporciona una visión detallada de los hallazgos claves que arroja el protocolo, permitiendo una comprensión más profunda de la calidad del agua subterránea en la zona de estudio y sus implicaciones para la gestión y preservación del recurso vital.

4.1 Selección óptima de pozos de monitoreo

El algoritmo se aplicó para seleccionar 66 pozos de monitoreo de los 2819 posibles. Se realizaron múltiples ejecuciones del algoritmo, presentando una mejora en el frente a medida que aumentaba el número de generaciones. Se generó un total de 1000 generaciones y se obtuvo un frente de pareto con 61 soluciones.

La Figura 4.1 muestra la última generación de soluciones tras una normalización mín-máx, donde se resalta la solución seleccionada. La solución seleccionada fue la que tenía la suma más alta de los cuatro objetivos normalizados. El mapa de la Figura 4.2 muestra la ubicación espacial de los 66 pozos seleccionados del acuífero Patiño.

A partir del conjunto de pozos óptimos seleccionados con el algoritmo NSGA-II y PO, se procedió a realizar el trabajo de campo, representando la materialización del muestreo en un área de cuidado fundamental como es el acuífero Patiño. Este proceso se benefició en gran medida de la utilización de meta-

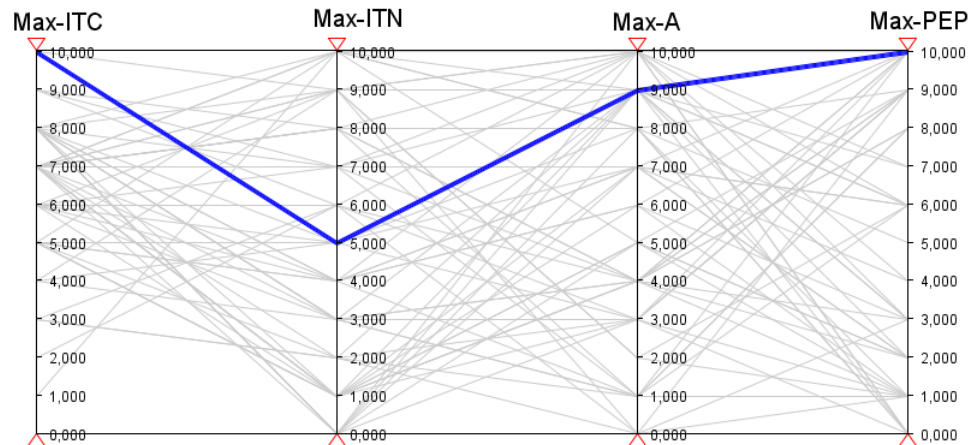


Figura 4.1: Gráfico de coordenadas paralelas del frente Pareto con 61 soluciones óptimas. La solución seleccionada aparece resaltada. Cada eje representa la normalización de cada objetivo.

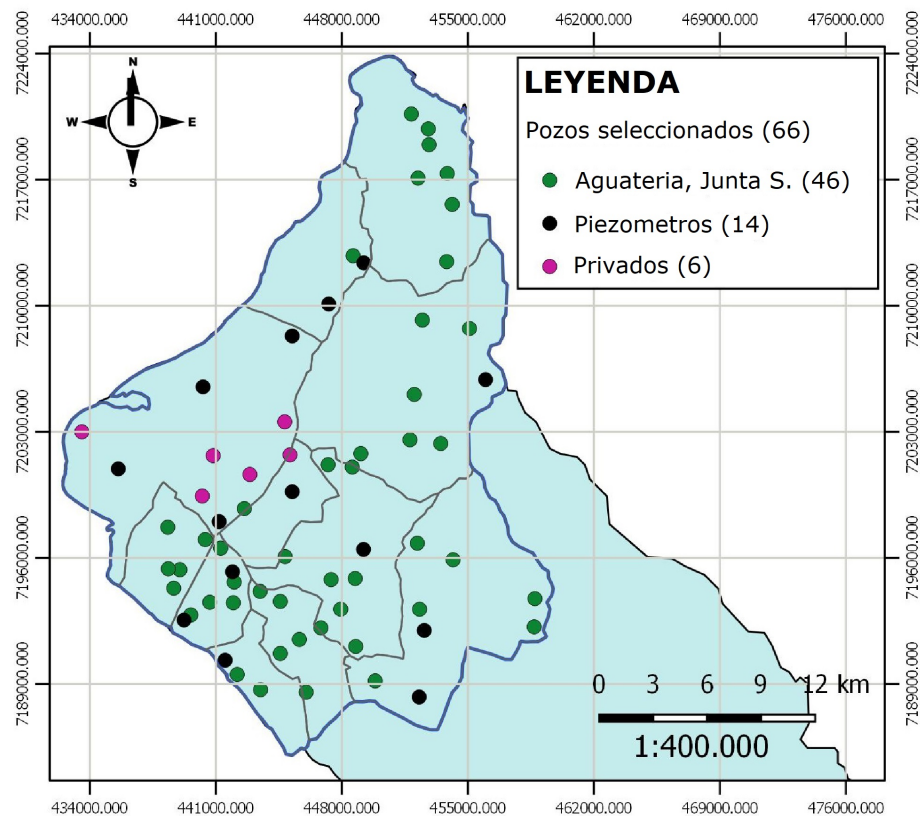


Figura 4.2: Distribución de los 66 pozos seleccionados con el NSGA-II y PO. Los pozos están identificados según la prioridad: pozos de acceso público (aguaterías, juntas de saneamiento, piezómetros) y pozos privados.

heurísticas en la selección de pozos, ya que estas técnicas permitieron explorar de manera eficiente un vasto espacio de soluciones, teniendo en cuenta múltiples ob-

jetivos y las limitaciones específicas del proyecto. La flexibilidad y adaptabilidad de las metaheurísticas también jugaron un papel crucial al permitir ajustar la selección de pozos según las necesidades del monitoreo y la información geoespacial disponible. De esta manera, se logró optimizar la distribución de los pozos de monitoreo en el acuífero Patiño, mejorando la eficacia de la gestión y preservación de este recurso vital.

4.2 Resultados de calidad de agua del monitoreo

Las 66 muestras fueron evaluadas en laboratorio. De los 21 parámetros analizados, nueve (43%) presentaron concentraciones por encima de las normas de agua potable para el consumo humano. En la Figura 4.3, se muestra la distribución porcentual de los parámetros que se encontraron fuera de rango según los valores límites permisibles. Se puede identificar que el valor del pH del agua es el que representa el mayor porcentaje (45%) de muestras con valores fuera de los límites permisibles para agua de consumo humano, seguido de los valores de nitratos con un 29% de las muestras que presentan concentraciones por encima de los límites. Se puede observar una correspondencia entre las elevadas concentraciones de nitrato y las zonas con ocupación urbana más antigua y más densamente poblada, por lo tanto con mayor posibilidad de contaminación por filtración de aguas negras. Las concentraciones de magnesio fuera de los límites admisibles podrían estar relacionados a los valores de pH bajos, ya que en estas condiciones se facilita la disolución de minerales y rocas que contienen estos iones.

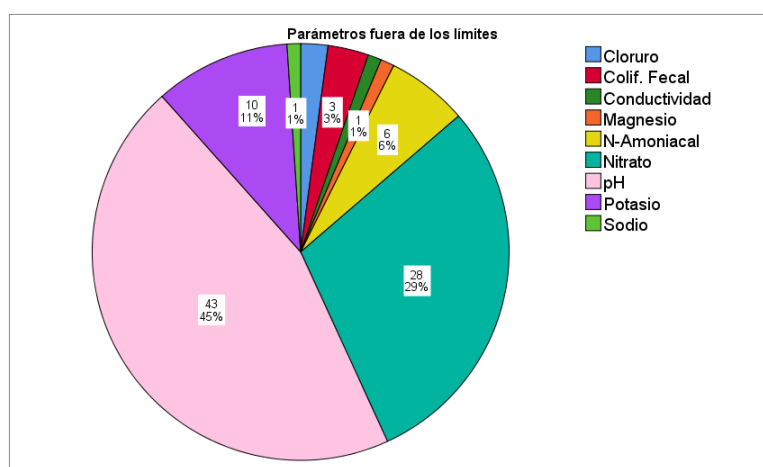


Figura 4.3: Gráfico circular con el porcentaje de los 8 parámetros fuera de los límites permisibles de calidad de agua.

Los pozos que poseen al menos un valor fuera del límite permisible son en total

55 pozos, correspondientes al 83% de las muestras y se visualizan en la Figura 4.4. La zona central y sur-oeste del área de estudio es la de mayor preocupación, identificando hasta 2 parámetros fuera del límite por pozo, principalmente con parámetros como N-amoniaco y N-nitratos o N-nitratos y coliformes fecales.

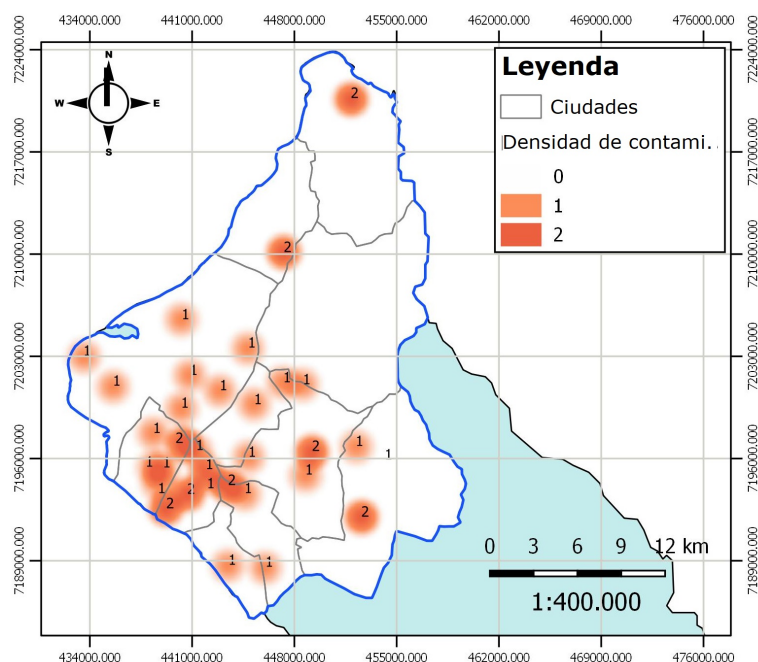


Figura 4.4: Mapa de los puntos críticos, considerando la densidad de parámetros fuera de los límites admisibles por pozo.

La ingesta de agua con altas concentraciones de nitrato se reduce a nitrito en el estómago y puede oxidar la hemoglobina a metahemoglobina, bloqueando el transporte de oxígeno a través del cuerpo, causando cianosis y asfixia [2]. La presencia de nitrato en el agua está asociada a prácticas agrícolas, fertilizantes, aguas residuales y saneamiento in situ [1] -en el caso del acuífero Patiño se hipotetiza que esta última sea la principal causa-.

La presencia de nitrógeno amoniacal se asocia a aguas residuales domésticas o industriales, residuos animales y vertidos de drenaje recientes; su ingesta no tiene un impacto inmediato en la salud [1]. Una hipótesis similar a la anterior sigue siendo válida, dada la presencia de pozos negros, a lo que se añaden algunos posibles emplazamientos industriales que podrían ser las fuentes de estos contaminantes.

Las concentraciones de coliformes fecales (>0) se identificaron en tres pozos. Su ingesta produce enfermedades infecciosas y parasitarias del aparato digestivo como diarrea, cólera, disentería; y su presencia está asociada a pozos contaminados o abandonados, fosas sépticas, letrinas y materia fecal derivada de la

ganadería intensiva [1].

Un completo análisis estadístico descriptivo de los resultados del muestreo fueron presentados en la contribución [11].

4.2.1 Evaluación del protocolo mediante una comparación con otras campañas de muestreo

Para evaluar la efectividad del protocolo implementado en el acuífero Patiño, se comparó el protocolo de selección y resultados con otras dos campañas realizadas por [33, 35] en el mismo acuífero. Como se menciona en la metodología 3.1, el objetivo de cada estudio comparativo fue caracterizar la contaminación potencial.

En [35] muestrearon 19 pozos en 567 km², priorizaron el muestreo en 11 piezómetros ya construidos y diseñados para monitoreo y otros 8 pozos profundos, actualmente en uso y dirigidos a grandes consumidores de agua como instituciones públicas, gasolineras, lavaderos de coches, obras hidráulicas y juntas de saneamiento. Los pozos seleccionados se consideraron importantes porque estaban distribuidos en zonas relacionadas con la contaminación por nitratos y la presencia de alta salinidad, identificadas mediante una campaña de muestreo previa. En [33] muestrearon 65 pozos en un área de 79 km² dentro del acuífero Patiño, concretamente en la cuenca hidrográfica del arroyo San Lorenzo, dentro de una zona urbana con diferentes tipos de industrias, comercios y pocos servicios sanitarios instalados. El objetivo fue evaluar de forma inmediata el estado del acuífero, centrándose en una zona más pequeña con gran presión antrópica con una tipología muy variada de posibles fuentes de contaminación.

En comparación, en este trabajo se desarrolló una forma sistemática de llevar a cabo una campaña de muestreo teniendo en cuenta cuatro objetivos, como se muestra en la Sección 3.2.1. Como indicador de eficacia, los pozos seleccionados en los otros estudios se evaluaron utilizando los cuatro objetivos que se desarrollaron en el protocolo propuesto para comprender cómo se comparaban con la presente selección. Como segunda medida de la eficacia, se comparó el número de pozos con parámetros fuera de los límites permisibles (pozos preocupantes) frente al número total de pozos muestreados.

Según la Tabla 4.1, el protocolo de selección de pozos propuesto tiene un mejor rendimiento en los cuatro objetivos evaluados: mayor suma de INT, mayor suma de ICT, mayor área de cobertura y mayor número de pozos con prioridad. Además, evaluando la eficacia del protocolo como pozos fuera de los límites sobre el total de pozos muestreados, el presente trabajo obtuvo un mejor rendimiento en comparación a las otras dos campañas.

Tabla 4.1: Comparación de las campañas de muestreo sobre el acuífero Patiño. Sólo se consideran los pozos muestreados en el AMA.

Estudio	Área de Estudio (km ²)	Parámetros monit.	Pozos muestreados	Objetivos				Pozos preocupantes	Eficiencia del protocolo
				1. ICT Sum	2. INT Sum	3. Área Cobert.	4. Pozos Prioridad		
Presente trabajo	567	21	66	10	5.2	9.0	10.0	33	0.50
[35] INCLAM-HQA (2017)	567	19	19	1.6	0.0	9.7	10.0	7	0.37
[33] Houben et al. (2012)	79	31	63	5.5	0.0	10.0	0.0	13	0.21

Los parámetros químicos o bacteriológicos fuera de los límites, en [33] fueron sodio, nitratos y coliformes fecales; en [35] los parámetros fuera de rango fueron sodio, nitrato, cloruro, potasio y E. coli; mientras que en el presente trabajo se identificaron sodio, nitratos, N-amoniaco, cloruro y coliformes fecales. En ese sentido, el protocolo propuesto posee un mejor o igual desempeño al categorizar el número de parámetros identificados fuera de los límites permisibles.

4.2.2 Mapeo de la calidad de agua subterránea

El siguiente paso del protocolo consistió en calcular el índice de calidad del agua (WQI). Los resultados indicaron que el 75% de los pozos muestreados poseen la clase 1 o “excelente”, el 23% de los pozos muestreados poseen la clase 2 o “buena” y el 2% de los pozos presentan la clase 3 o “deficiente”. Dicho cálculo es el resultado de la combinación de 10 parámetros (pH, sólidos totales disueltos, cloruro, sulfato, sodio, potasio, calcio, magnesio, dureza total y nitratos) y proporciona información integrada sobre la calidad global del agua.

La Figura 4.5 muestra el mapa de distribución espacial del WQI e indica que las muestras con aguas de calidad “buenas a malas” se encuentran en zonas con concentraciones altas de nitratos (promedio de 93 mg/L), además existe una mayor densidad poblacional (promedio de 4134 habitantes por km²), con una densidad alta de pozos ciegos por hectárea (promedio de 6,29); y al norte del acuífero existe una influencia importante de aguas saladas, con altos valores de sodio (hasta 284 mg/L).

Los resultados obtenidos con el WQI indicaron que la mayoría de las muestras son de calidad “excelente” o “buena”, pero al evaluar parámetros específicos, se identificaron 9 parámetros (cloruro, coliformes fecales, conductividad, magnesio, nitratos, nitrógeno amoniacal, pH, potasio y sodio) por encima de los valores permisibles y aceptables. Por lo tanto, el WQI indica una calificación general de la aptitud del agua potable, que podría ser útil solo para comparar diferentes campañas de muestreo.

En este contexto, se propone la selección de parámetros relevantes para reducir el número de parámetros a monitorear y, a su vez, considerarlos en un nuevo cálculo de WQI.

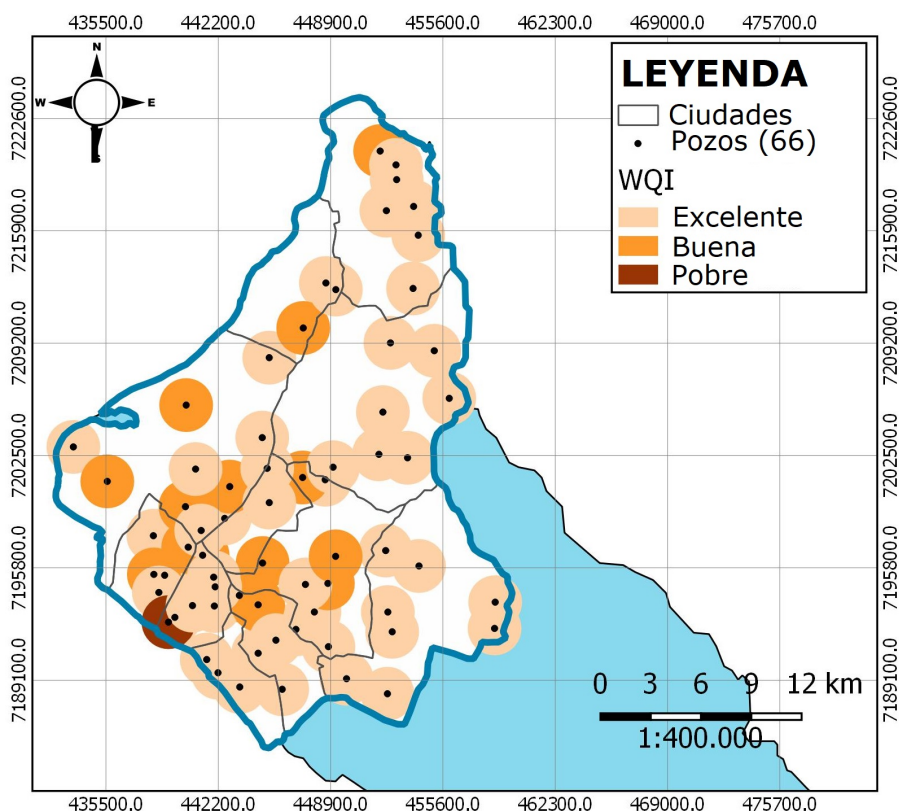


Figura 4.5: Mapa de la distribución espacial de la calidad de agua subterránea del acuífero Patiño.

4.3 Reducción y selección de parámetros de calidad de agua

A continuación se indican los pasos seguidos para reducir y seleccionar los parámetros de calidad de agua más significativos, a partir del conjunto de parámetros realizados en el monitoreo indicado en las secciones previas. En total se cuenta con 21 parámetros del conjunto de datos inicial: alcalinidad total, arsénico, bicarbonato, calcio, carbonato, cloruro, cobre, conductividad, cromo total, coliformes fecales, magnesio, manganeso, materia orgánica, mercurio, nitratos, nitritos, nitrógeno amoniacal, pH, potasio y sodio, sulfato.

Primero, se realizó un preprocesamiento del conjunto de datos inicial. En el mismo se descartaron los parámetros con valores constantes (arsénico, cobre,

cromo total, mercurio y manganeso) y se analizaron los coeficientes de correlaciones iniciales entre todas las variables. Donde se identificaron correlaciones directas y fuertes (>0.8) entre las variables: alcalinidad-bicarbonato, conductividad-sodio-cloruro. Los coeficientes muy altos indican que se están midiendo lo mismo o algún fenómeno muy cercano entre ellos; por lo tanto se continuó eliminando las variables: alcalinidad total, sodio y cloruro.

Se continúa la reducción y selección de parámetros con las siguientes 13 variables: bicarbonato, calcio, carbonato, conductividad, coliformes fecales, magnesio, materia orgánica, nitratos, nitritos, nitrógeno amoniacal, pH, potasio y sulfato. Se tiene una dimensión de (66, 13).

Seguidamente, se calculó el PCA y se aplicaron los diferentes criterios de selección de variables (VSC) indicados en la metodología. En la Tabla 4.2 se muestran las variables seleccionadas (1) por cada criterio de selección considerado. La columna “nueva selección” considera las variables más repetidas en los subconjuntos obtenidos.

El Subconjunto 1 de parámetros seleccionados con PCA fueron los siguientes: bicarbonato, calcio, carbonato, magnesio, nitrato y n-amoniacal.

Tabla 4.2: Selección de variables de calidad de agua subterránea con PCA-VSC.

#	Variables	VC	M1	M2	M3	Nueva Selec.	
1	Bicarbonato	1	1	1	1	4/4	1
2	Calcio	1	1	1	1	4/4	1
3	Carbonato	1	1	1	1	4/4	1
4	Conductividad	1	-	-	-	1/4	-
5	Colif. Fecal	-	-	-	-	0/4	-
6	Magnesio	1	1	1	1	4/4	1
7	Materia orgánica	-	1	1	1	3/4	-
8	Nitratos	1	1	1	1	4/4	1
9	Nitritos	1	-	-	-	1/4	-
10	N-amoniacal	1	1	1	1	4/4	1
11	pH	-	1	1	1	3/4	-
12	Potasio	-	1	1	1	3/4	-
13	Sulfato	-	-	-	1	1/4	-

Continuando con los algoritmos de selección de atributos, fueron aplicados los algoritmos presentados en la Tabla 4.3, donde también se indica el orden, entre 1 y 13, en que cada atributo fue seleccionado en cada modelo. El atributo más relevante posee el valor 1. Total columna presenta el grado de relevancia para cada atributo, refleja el número total de atributos (13) multiplicado por el número de algoritmos analizados (7), menos la suma de los valores de cada columna.

Tabla 4.3: Parámetros identificados por los algoritmos de selección de atributos.

		N-Amoniaca	Nitritos	Nitratos	MatOrganica	Conductividad	pH	Bicarbonato	Carbonato	Sulfato	Magnesio	Calcio	Potasio	ColiFecal
	Total columna	47	23	83	13	60	17	3	22	21	71	53	51	61
1	SymmetricalUncert-AttributeEval	5	8	1	10	4	12	13	9	11	2	6	7	3
2	CfsSubsetEval	4	13	1	13	5	13	13	13	13	3	6	13	2
3	InfoGainAttributeEval	4	9	1	10	3	12	13	8	11	2	7	6	5
4	CorrelationAttributeEval	7	11	1	12	4	9	13	10	6	2	5	3	8
5	GainRatioAttributeEval	5	9	2	10	4	12	13	8	11	3	7	6	1
6	OneRAttributeEval	12	7	1	11	6	10	13	8	9	4	5	2	3
7	ReliefFAttributeEval	7	11	1	12	5	6	10	13	9	4	2	3	8

Como se observa en la Tabla 4.3, no hay ningún valor cero puesto que no hay ningún atributo que para todos los algoritmos esté en el orden 13, que sería el caso más desfavorable. Se estableció un punto de corte o selección en 51 puntos, ya que visualmente selecciona las variables más significativas.

El Subconjunto 2 de parámetros seleccionados con los algoritmos fueron los siguientes: calcio, coliformes totales, conductividad, magnesio, nitrato, potasio.

El Subconjunto 3 considera: Subconjunto 1 $\cup$ Subconjunto 2. Bicarbonato, calcio, carbono, coliformes totales, conductividad, magnesio, nitrato, n-amoniaca y potasio.

El tercer paso fue verificar la efectividad de cada subconjunto de parámetros seleccionados, realizando el análisis predictivo considerando una clasificación supervisada en función al resultado WQI inicial. La clase WQI se evalúa con los valores de índices 1 “excelente”, 2 “buena” y 3 “deficiente”. En la Tabla 4.4 se muestra en detalle la verificación de los Subconjuntos 1, 2 y 3, se listan los algoritmos clasificadores considerados, juntos con los algoritmos evaluadores de selección. Además, se indican para cada uno su porcentaje de aciertos y los resultados de cada valor de la clase (*precision* y *recall*). La clasificación con el algoritmo J48 arrojó una mejor precisión con mayor porcentaje de aciertos; el Subconjunto 1 fue 92,42% y los Subconjuntos 2 y 3 fue 93,93%.

Tabla 4.4: Comparación de resultados de los algoritmos clasificadores en función al WQI inicial con los parámetros relevantes de los Subconjuntos 1, 2 y 3.

Clasificador	Evaluador	Subconjunto 1						
		Acier- tos	Resultado por cada clase					
			Precision			Recall		
			1	2	3	1	2	3
J48	-	92.42	0.94	0.85	0.00	0.98	0.80	0.00
J48	Symmetrical Uncert Attribute Eval	92.42	0.94	0.85	0.00	0.98	0.80	0.00
J48	Cfs Subset Eval	92.42	0.94	0.85	0.00	0.98	0.80	0.00
J48	Info Gain Attribute Eval	92.42	0.94	0.85	0.00	0.98	0.80	0.00
IBk	-	87.87	0.92	0.73	0.00	0.94	0.73	0.00
IBk	Symmetrical Uncert Attribute Eval	87.87	0.92	0.73	0.00	0.94	0.73	0.00
IBk	Cfs Subset Eval	84.84	0.90	0.66	0.00	0.92	0.66	0.00
IBk	Info Gain Attribute Eval	87.87	0.92	0.73	0.00	0.94	0.73	0.00
NaiveBayes	-	84.84	0.91	0.64	0.00	0.90	0.73	0.00
NaiveBayes	Symmetrical Uncert Attribute Eval	84.84	0.91	0.64	0.00	0.90	0.73	0.00
NaiveBayes	Cfs Subset Eval	87.87	0.93	0.70	0.00	0.92	0.80	0.00
NaiveBayes	Info Gain Attribute Eval	84.84	0.91	0.94	0.00	0.90	0.73	0.00
RandomForest	-	87.87	0.92	0.73	0.00	0.94	0.73	0.00
RandomForest	Symmetrical Uncert Attribute Eval	87.87	0.92	0.73	0.00	0.94	0.73	0.00
RandomForest	Cfs Subset Eval	86.36	0.92	0.68	0.00	0.92	0.73	0.00
RandomForest	Info Gain Attribute Eval	87.87	0.92	0.73	0.00	0.94	0.73	0.00

Clasificador	Evaluador	Acier- tos	Subconjunto 2					
			Resultado por cada clase					
			Precision			Recall		
			1	2	3	1	2	3
J48	-	93.93	1.00	0.78	0.00	0.94	1.00	0.00
J48	Symmetrical Uncert Attribute Eval	93.93	1.00	0.78	0.00	0.94	1.00	0.00
J48	Cfs Subset Eval	92.42	0.96	0.81	0.00	0.96	0.86	0.00
J48	Info Gain Attribute Eval	93.93	1.00	0.78	0.00	0.94	1.00	0.00
IBk	-	89.39	0.92	0.78	0.00	0.96	0.73	0.00
IBk	Symmetrical Uncert Attribute Eval	89.39	0.92	0.78	0.00	0.96	0.73	0.00
IBk	Cfs Subset Eval	87.87	0.93	0.70	0.00	0.92	0.80	0.00
IBk	Info Gain Attribute Eval	89.39	0.92	0.78	0.00	0.96	0.73	0.00
NaiveBayes	-	89.39	0.97	0.70	0.00	0.90	0.93	0.00
NaiveBayes	Symmetrical Uncert Attribute Eval	89.39	0.97	0.70	0.00	0.90	0.93	0.00
NaiveBayes	Cfs Subset Eval	89.39	0.95	0.72	0.00	0.92	0.86	0.00
NaiveBayes	Info Gain Attribute Eval	89.39	0.97	0.70	0.00	0.90	0.93	0.00
RandomForest	-	89.39	0.92	0.78	0.00	0.96	0.73	0.00
RandomForest	Symmetrical Uncert Attribute Eval	89.39	0.92	0.78	0.00	0.96	0.73	0.00
RandomForest	Cfs Subset Eval	90.90	0.94	0.80	0.00	0.96	0.80	0.00
RandomForest	Info Gain Attribute Eval	89.39	0.92	0.78	0.00	0.96	0.73	0.00

Clasificador	Evaluador	Acier- tos	Subconjunto 3					
			Resultado por cada clase					
			Precision			Recall		
			1	2	3	1	2	3
J48	-	93.93	1.00	0.78	0.00	0.94	1.00	0.00
J48	Symmetrical Uncert Attribute Eval	93.93	1.00	0.78	0.00	0.94	1.00	0.00
J48	Cfs Subset Eval	92.42	0.96	0.05	0.00	0.96	0.83	0.00
J48	Info Gain Attribute Eval	93.93	1.00	0.78	0.00	0.94	1.00	0.00
IBk	-	90.90	0.92	0.84	0.00	0.98	0.73	0.00
IBk	Symmetrical Uncert Attribute Eval	90.90	0.92	0.84	0.00	0.98	0.73	0.00
IBk	Cfs Subset Eval	87.87	0.93	0.70	0.00	0.92	0.80	0.00
IBk	Info Gain Attribute Eval	90.90	0.92	0.84	0.00	0.98	0.73	0.00
NaiveBayes	-	89.39	0.94	0.75	0.00	0.94	0.80	0.00
NaiveBayes	Symmetrical Uncert Attribute Eval	89.39	0.94	0.75	0.00	0.94	0.80	0.00
NaiveBayes	Cfs Subset Eval	89.39	0.95	0.72	0.00	0.92	0.86	0.00
NaiveBayes	Info Gain Attribute Eval	89.39	0.94	0.75	0.00	0.94	0.80	0.00
RandomForest	-	89.39	0.90	0.83	0.00	0.98	0.66	0.00
RandomForest	Symmetrical Uncert Attribute Eval	89.39	0.90	0.83	0.00	0.98	0.66	0.00
RandomForest	Cfs Subset Eval	90.90	0.94	0.80	0.00	0.95	0.80	0.00
RandomForest	Info Gain Attribute Eval	89.39	0.90	0.83	0.00	0.98	0.66	0.00

Los modelos evaluados en esta sección, conducen a un mejor entendimiento de la distribución de los datos y aportan un nuevo conocimiento, relacionados a la identificación de los parámetros más significativos. Los subconjuntos de parámetros seleccionados se utilizan a continuación para su validación frente a otros escenarios y de esta manera, proponer aquellos más significativos para un nuevo cálculo WQI.

4.4 Validación del protocolo

4.4.1 Selección de pozos

La validación consistió en la comparación de la eficiencia de la selección de pozos de monitoreo del Escenario I. La Figura 4.6 muestra la distribución de los 70 pozos seleccionados en dicho escenario. Los pozos seleccionados cubren un área de cobertura de 291 km², maximizando el ICT, INT y pozos con prioridad (Tabla 4.5).

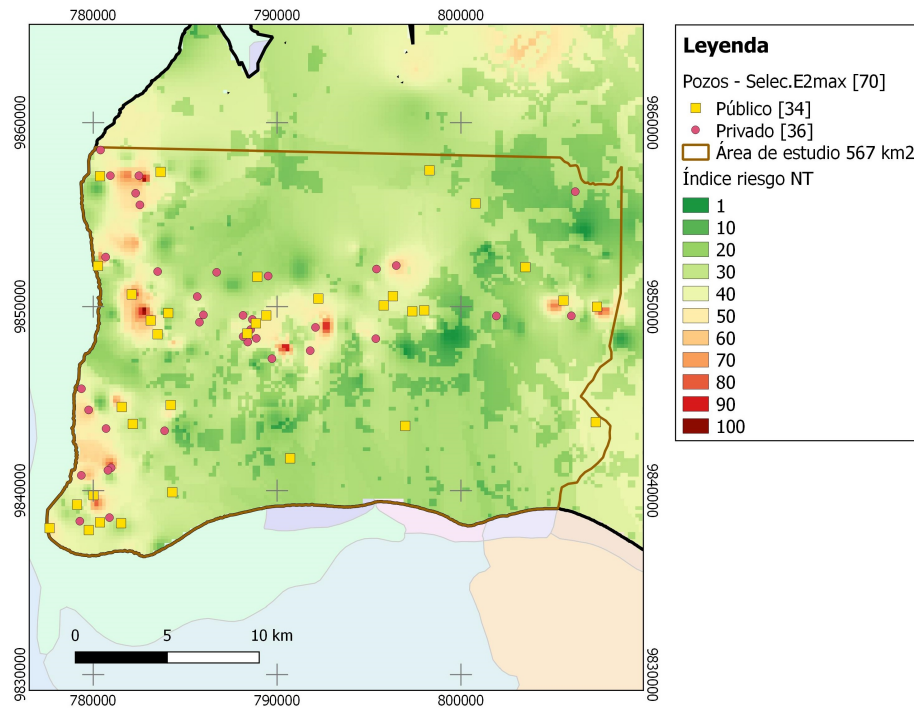


Figura 4.6: Mapa de la distribución espacial de los 70 pozos óptimos seleccionados con NSGA-II PO para el Escenario I.

Como indicador de eficiencia, los pozos seleccionados en la Campaña 1 se evaluaron utilizando los cuatro objetivos para poder compararlos con los resultados de la campaña con selección de pozos óptimos con NSGA-II PO para el

Escenario I. Además, se comparó el número de pozos con parámetros fuera de los límites permisible de potabilidad de agua, dividido el número total de pozos muestreados.

Como se visualiza en la Tabla 4.5, el Escenario I tiene un mejor rendimiento en los cuatro objetivos evaluados: mayor suma ICT, mayor suma INT, mayor área cubierta y mayor número de pozos con prioridad. Además, resultó superior al evaluar el indicador de eficiencia.

Tabla 4.5: Comparación de la eficiencia entre las campañas.

Campañas	Nro. Parámetros	Nro. Pozos	ICT Sum	INT Sum	Área cubierta	Pozos con prioridad	Pozos fuera de límites	Eficiencia
NSGA-II PO Escenario I	20	70	682.9	2782.8	291	34	66	0.9
Barreiras Campaña 1	20	72	506.8	2475.5	261	5	53	0.7

4.4.2 Selección de parámetros

Los subconjuntos identificados en la Sección 4.3 fueron validados, realizando el análisis predictivo considerando la clasificación supervisada en función al resultado WQI inicial de cada escenario. El Escenario I con datos de concentraciones reales geoespaciales del acuífero Barreiras-Brasil [19]; y el Escenario II con datos sintéticos, generados siguiendo la metodología indicada en la Sección 3.5.2.

El WQI inicial de cada escenario fue calculado siguiendo la metodología indicada en la Sección 3.3. El WQI del Escenario I posee 46 pozos con la clase 1, 21 pozos con la clase 2 y 5 pozos con la clase 3. El WQI del Escenario II posee 28 pozos con la clase 1, 54 pozos con la clase 2 y 18 pozos con la clase 3; como era de esperarse, debido al muestreo artificial con mayor número de concentraciones fuera de rango, se obtuvo una mayor cantidad de pozos en la clase “deficiente”.

Posteriormente, se realizó la clasificación supervisada en función a los resultados obtenidos, con la finalidad de determinar si se mantienen los mismos porcentajes de aciertos y si mejoran los resultados por cada valor de la clase (*precision* y *recall*). El algoritmo clasificador J48 se ejecutó por ser el de mayor precisión en la evaluación presentada en la Tabla 4.4.

Para la validación se consideraron las mismas selecciones de parámetros significativos en los Escenarios I y II. El desempeño de las selecciones de parámetros, utilizando el algoritmo J48 se resume en la siguiente Tabla 4.6.

Como se observa en la Tabla 4.6, los subconjuntos seleccionados presentaron buenos desempeños en los escenarios considerados.

Tabla 4.6: Resumen de los subconjuntos evaluados por escenario con el algoritmo de clasificación J48.

	Subconjunto 1			Subconjunto 2			Subconjunto 3		
	Aciertos	Precisión	Sensibilidad	Aciertos	Precisión	Sensibilidad	Aciertos	Precisión	Sensibilidad
Escenario I	94.4%	0.94	0.94	91.6%	0.91	0.91	93.1%	0.92	0.93
Escenario II	65.0%	0.61	0.65	73.0%	0.75	0.73	72.0%	0.75	0.72

Los subconjuntos del Escenario I, presentaron un buen rendimiento con el algoritmo J48, se obtuvieron porcentajes de aciertos superiores al 90%, alta precisión y sensibilidad; por lo tanto, los modelos son confiables para clasificar la mayoría de las instancias de manera correcta en función al WQI inicial, el cual considera un conjunto de parámetros distintos.

En el Escenario 2, los porcentajes de aciertos oscilan entre el 65% y el 73%, se mantienen en un rango razonable, pero con mayor dificultad para recuperar todas las instancias de la clase positiva. Lo que sugiere que el modelo podría estar limitado en la clasificación de datos en los subconjuntos de parámetros seleccionados. Por lo tanto, sería importante reevaluar la consideración de la muestra artificial con tantos parámetros fuera de rango e indicando contaminación elevada. También, podría estar ocurriendo que los pesos unitarios en el cálculo WQI se deberían estimar a partir de la reducción de parámetros y posteriormente reevaluarlos con la clasificación del WQI modificado.

En general, los resultados de la validación del protocolo realizado son prometedores. El protocolo de selección de pozos propuesto con el algoritmos NSGA-II y PO resultó con una alta eficiencia al ser aplicado en otro escenario. La metodología del protocolo propuesto se podría aplicar a otras áreas de interés, siguiendo cada uno de los pasos de (i) selección óptima de pozos de monitoreo utilizando un enfoque multi-objetivo aplicando el algoritmo NSGA-II modificado con PO, (ii) evaluación de los resultados de calidad de agua del monitoreo aplicando el WQI inicial y, (iii) la reducción y selección de parámetros de calidad de agua.

Capítulo 5

Conclusiones y Trabajos Futuros

5.1 Conclusiones

En este trabajo se abordó el diseño de un protocolo sistemático para el monitoreo de calidad de agua subterránea y la reducción de parámetros, utilizando un caso de estudio real y replicable a otras áreas de interés, basado en la aplicación de tecnologías emergentes.

Se dieron respuestas a las preguntas de investigación planteadas. Se exploró el estado del arte, identificando que la literatura sobre el monitoreo de calidad de agua subterránea no reporta la utilización de protocolos en todas las fases del monitoreo. La utilización de un protocolo en todas las fases del monitoreo de calidad de agua permite minimizar y eliminar errores, garantizar la generación de información consistente y confiable, determinar la línea de partida y mejorar la gestión de control de calidad de agua. Motivados con lo indicado, este trabajo propuso un protocolo sistemático siguiendo varios pasos viables para el control, la evaluación y la conservación de la calidad del agua subterránea.

El protocolo propuesto inició con la selección óptima de pozos de monitoreo de calidad de agua subterránea. Se consideró una alternativa de solución dentro de las tecnologías emergentes, utilizando el algoritmo NSGA-II con PO, para evaluar cuatro funciones objetivos: maximizar las concentraciones de Nitrógeno Total (NT), maximizar las concentraciones de Coliformes Totales (CT), maximizar el área de cobertura y maximizar la selección de pozos de acceso público. Se demostró su aplicación seleccionando 66 pozos en el área de estudio del acuífero Patiño, en su zona urbana, donde se analizaron 21 parámetros de calidad de agua. Los valores de pH se hallaron fuera de rango en el 65% de los pozos, seguidos de las concentraciones de nitrato, donde el 42% de los pozos muestreados tenían valores fuera del rango permisible.

El siguiente paso propuesto en el protocolo fue aplicar el cálculo del índice de calidad de agua subterránea (WQI). Los resultados indicaron que el 75% de los pozos muestreados tienen el tipo de agua categorizado como “excelente”, el 23% de los pozos muestreados tienen el tipo de agua categorizado como “bueno” y el 2% de los pozos tienen el agua categorizada como “deficiente”. Aunque esta calificación podría indicar una evaluación positiva del agua, en realidad reduce el efecto de ciertos parámetros que podrían estar fuera de rango. Además, se podría estar omitiendo en la evaluación, otros parámetros relevantes por presentar concentraciones fuera de rango.

El protocolo considera dos métodos para la reducción y selección de parámetros, utilizando técnicas avanzadas como el análisis de componentes principales (PCA) y la aplicación de algoritmos de selección de atributos (algoritmo *J48*, *Random forest*, *Naive Bayes* e *IBk*). Los modelos fueron evaluados en función a la clase WQI calculada. Los resultados indicaron tres subconjuntos de parámetros relevantes, contemplan los parámetros físico-químicos, bacteriológicos e iones mayoritarios. El Subconjunto 1 de parámetros seleccionados con PCA fue: bicarbonato, calcio, carbono, magnesio, nitrato y n-amoniaco. El Subconjunto 2 de parámetros seleccionados con los algoritmos fue: calcio, coliformes totales, conductividad, magnesio, nitratos, potasio. El Subconjunto 3 considera los parámetros del Subconjunto 1 y el Subconjunto 2.

Los pasos propuestos en el protocolo sistemático fueron validados considerando diferentes escenarios. Escenario I con datos reales geoespaciales y el Escenario II con datos sintéticos.

La validación de la selección de pozos consistió en la comparación de la eficiencia del Escenario I frente a otra campaña de monitoreo. El Escenario I obtuvo un mejor rendimiento en los cuatro objetivos evaluados: mayor suma ICT, mayor suma INT, mayor área cubierta y mayor número de pozos con prioridad. Además, resultó superior al evaluar el indicador de eficiencia (0,9).

Los subconjuntos de parámetros relevantes fueron validados, realizando el análisis predictivo considerando la clasificación supervisada en función al resultado WQI inicial de cada escenario. Los subconjuntos seleccionados presentaron buenos desempeños en los escenarios considerados y con los datos del acuífero Patiño. Los promedios de aciertos fueron: 92% en el Subconjunto 1, 92% en el Subconjunto 2 y 86% en el Subconjunto 3.

Creemos que nuestro trabajo es un buen argumento para incorporar un protocolo sistemático de muestreo y determinar los pozos óptimos que deben evaluarse. Es comprensible que las limitaciones de tiempo dicten a veces la selección de los

pozos, pero hemos demostrado que un simple conjunto de cuatro funciones objetivo lineales, se pueden aportar ideas y mejores resultados a la hora de caracterizar la contaminación de los acuíferos. A su vez, aplicar métodos que permitan que los resultados del monitoreo de calidad del agua, con una disminución de parámetros, se pueda comunicar fácilmente sin perder información significativa.

5.2 Trabajos Futuros

- i. Estimar cada uno de los pesos unitarios del cálculo WQI, a partir del método de reducción y selección de parámetros presentados.
- ii. Evaluar los atributos redundantes mediante la aplicación de algoritmos.
- iii. Utilizar un *framework* unificado como *ensemble learning* y/o *ensemble feature selection* para combinar los resultados de los métodos de selección de atributos.
- iv. Calcular el nuevo WQI con los parámetros propuestos y presentar un Índice de Calidad de Agua Subterránea para el acuífero Patiño (ICAS-AP).
- v. Escribir el artículo sobre la metodología de selección de atributos para el desarrollo del cálculo de Índice de Calidad de Agua Subterránea: caso de estudio del acuífero Patiño.
- vi. Implementar el cálculo ICAS-AP como herramienta de QGIS de escritorio.
- vii. Considerar el uso del atributo “profundidad de pozo” en la selección óptima de pozos. Asignar un peso de prioridad a los pozos que poseen dicha información e incluirlo como función objetivo del algoritmo.
- viii. Implementar métodos avanzados de inteligencia artificial y estadística en la herramienta GIS para permitir análisis avanzados de estas técnicas de manera espacial.
- ix. Modelado de series de temporales en los estudios de calidad del agua subterránea.
- x. Ampliar el protocolo incluyendo sistemas optimizados de monitoreo de aguas subterráneas y el intercambio de datos de monitoreo de aguas subterráneas.

- xi. Aplicar la ciencia ciudadana en el diseño de la investigación, la recopilación de datos y el proceso de interpretación de la calidad de agua subterránea.
- xii. Evaluar los modelos predictivos que describan mejor la distribución del transporte de contaminantes de forma dinámica.

Referencias Bibliográficas

- [1] Guías para la calidad del agua potable, 3ra edición: Vol. 1 - Recomendaciones. Ginebra, author =.
- [2] Guidelines for drinking-water quality: 4th edition incorporating first addendum. Ginebra, author =.
- [3] S. A., D. B., N. A.J., and F. H. Un nuevo método de estimación de densidad para problemas de optimización con tres objetivos usando metaheurísticas. *XVII Conferencia de la Asociacion Espanola para la Inteligencia Artificial*, pages 345–354, 2016.
- [4] S. A.D., M. N., Y. A.G., and P. B. Development of river water quality indices - a review. *Environ. Monit. Assess.*, 188:58, 2016.
- [5] N. C. Adkins. A framework for development of data analysis protocols for groundwater quality monitoring systems. 1993.
- [6] N. M. Al-Kandari and J. I. T. Variable selection and interpretation of covariance principal components. *Communications in Statistics - Simulation and Computation*, 30(2):339–354, 2001.
- [7] M. Ali Molla, S. Mohiuddin Abdus Salam, and N. Saha. Surface and groundwater quality assessment based on multivariate statistical techniques in the vicinity of Mohanpur, Bangladesh. *International Journal of Environmental Health Engineering*, (4):1–9, 2015.
- [8] A. H. Alzraiee, D. A. Bau, and L. A. Garcia. Multiobjective design of aquifer monitoring networks for optimal spatial prediction and geostatistical parameter estimation. *Water Resources Research*, 49(6):3670–3684, 2013.
- [9] V. Amiri, M. Rezaei, and N. Sohrabi. Groundwater quality assessment using entropy weighted water quality index (EWQI) in Lenjanat, Iran. *Environmental Earth Sciences*, 72:3479–3490, 2014.
- [10] A. N. d. A. ANA. Protocolo nacional para el monitoreo de la calidad de los recursos hídricos superficiales, 2016.

- [11] L. Báez, C. Ávalos, and von L. Designing and validating a groundwater sampling campaign in an unmonitored aquifer: Patiño aquifer case. *Environmental Earth Sciences*, 80(11):406, 2021.
- [12] L. Báez, C. Villalba, and J. P. Nogués. Protocolo técnico-científico: Monitoreo y simulación de transporte de contaminantes en zonas urbanas del acuífero Patiño, 2018.
- [13] L. Báez, C. Villalba, and N. J. P. Comparison of contaminant-specific risk maps for an urban aquifer: Patiño aquifer case. *Environmental Earth Sciences*, 78(5):1–15, 2019.
- [14] M. Balderacchi, P. Benoit, P. Cambier, O. M. Eklo, A. Gargini, A. and Gemitzi, ..., and M. Trevisan. Groundwater pollution and quality monitoring approaches at the European level. *Critical reviews in environmental science and technology*, 43(4):323–408, 2013.
- [15] J. Barbosa Filho and I. B. de Oliveira. Development of a groundwater quality index: GWQI, for the aquifers of the state of Bahia, Brazil using multivariable analyses. *Scientific Reports*, 11(1):1–22, 2021.
- [16] S. N. Bashi-Azghadi and R. Kerachian. Locating monitoring wells in groundwater systems using embedded optimization and simulation models. *Science of the Total Environment*, 408(10):2189–2198, 2010.
- [17] H. P. Broers. Strategies for regional groundwater quality monitoring.
- [18] N. M. Burri, R. Weatherl, C. Moeck, and M. Schirmer. A review of threats to groundwater quality in the anthropocene. *Science of the Total Environment*, 684:136–154, 2019.
- [19] Catálogo de Metadatos ANA. Estudos hidrogeológicos para a gestão das águas subterrâneas da região de belém, pa, 16 de enero de 2023.
- [20] S. Dash and A. S. Kalamdhad. Science mapping approach to critical reviewing of published literature on water quality indexing. *Ecological Indicators*, 128:107862, 2012.
- [21] C. J. Daughney, M.-F. M. Raiber, M., U. Morgenstern, and R. van der Raaij. Use of hierarchical cluster analysis to assess the representativeness of a baseline groundwater quality monitoring network: comparison of New Zealand’s national and regional groundwater monitoring programs. *Hydrogeology Journal*, 20(1):185, 2012.

- [22] K. Deb, A. Pratap, S. Agarwal, and T. A. M. T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput*, 6(2):182–197, 2002.
- [23] G. Devic, D. Djordjevic, and S. Sakan. Natural and anthropogenic factors affecting the groundwater quality in Serbia. *Science of the Total Environment*, 468:933–942, 2014.
- [24] DGEEC. Compendio Estadístico Año 2012, 27 de marzo de 2020.
- [25] A. Dhar. Geostatistics-based design of regional groundwater monitoring framework. *ISH Journal of Hydraulic Engineering*, 19(2):80–87, 2013.
- [26] F. Di Pierro. Many-objective evolutionary algorithms and applications to water resources engineering, 2006.
- [27] J. Dong, G. Wang, H. Yan, J. Xu, and X. Zhang. A survey of smart water quality monitoring system. *Environmental Science and Pollution Research*, 22:4893–4906, 2015.
- [28] U. S. E. P. A. EPA. Monitoring and Assessing Water Quality, 22 de diciembre de 2022.
- [29] J. Facetti, R. Nunez, C. Gomez, J. Ojeda, C. Bernal, R. Leon-Ovelar, and F. Carvalho. Methyl tert-butyl ether (MtBE) in deepwells of the Patiño Aquifer, Paraguay: a preliminary characterization. *Sci Total Environ*, 647:1640–1650, 2019.
- [30] S. Geetha and S. J. S. W. Gouthami. Internet of things enabled real time water quality monitoring system. *Smart Water*, 2(1):1–19, 2016.
- [31] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
- [32] X. Han, L. X., D. Gao, B. Ma, X. Gao, and M. Cheng. Costs and benefits of the development methods of drinking water quality index: A systematic review. *Ecological Indicators*, 144:109501, 2022.
- [33] G. Houben, C. Rojas, A. Romero, N. Cabral, A. Olavarrieta, C. Enciso, M. L. Brunelli, F. Colman, J. Silvero, M. Almada, and G. Cruzans. Informe técnico: Investigación de la calidad del agua. Cuenca hídrica del arroyo

- San Lorenzo, Departamento Central. Proyecto "Manejo sostenible y protección de las aguas subterráneas en Paraguay"(PAS-PY), 2012.
- [34] Z. Y. Hseu and Z. S. Chen. Experiences of mass pig carcass disposal related to groundwater quality monitoring in Taiwan. *Sustainability*, 9(1):46, 2016.
- [35] INCLAM-HQA. Diagnóstico del Acuífero Patiño, en Proyecto BID PR-T 1207 "Estudio de Recursos Hídricos y Vulnerabilidad Climática del Acuífero Patiño", 2017.
- [36] E. J. Jackson. A Users Guide to Principal Components, 1991.
- [37] M. K. Jha, A. Shekhar, and M. A. Jenifer. Assessing groundwater quality for drinking water supply using hybrid fuzzy-GIS-based water quality index. *Water Research*, 179:115867, 2020.
- [38] D. K. Multi-objective optimisation using evolutionary algorithms: an introduction. In: Multi-objective evolutionary optimisation for product design and manufacturing. *Springer, London*, pages 3–34, 2011.
- [39] T. Kaur, R. Bhardwaj, and S. Arora. Assessment of groundwater quality for drinking and irrigation purposes using hydrochemical studies in Malwa region, southwestern part of Punjab, India. *Applied water science*, 7:3301–3316, 2017.
- [40] P. Li, S. He, N. Yang, and G. Xiang. Groundwater quality assessment for domestic and agricultural purposes in Yan-an City, northwest China: implications to sustainable groundwater quality management on the Loess Plateau. *Environmental Earth Sciences*, 77:1–16, 2018.
- [41] P. Li, R. Tian, C. Xue, and J. Wu. Progress, opportunities, and key fields for groundwater quality research under the impacts of human activities in China with a special focus on western China. *Environmental Science and Pollution Research*, 24(15):13224–13234, 2017.
- [42] E. Lioubimtseva. A multi-scale assessment of human vulnerability to climate change in the Aral Sea basin. *Environmental Earth Sciences*, 73:719, 2015.
- [43] J. C. Loftis. Trends in groundwater quality. *Hydrological processes*, 10(2):335–355, 1996.

- [44] S. T. Lumb A. and B. J. F. A review of genesis and evolution of water quality index (WQI) and some future directions. *Water Qual. Exposure Health*, 3:11–24, 2011.
- [45] B. R. M., M. N. I., D. R. A., and O. M. F. A water quality index-crashing the psychological barrier, 1972.
- [46] S.-N. S. M., E. K., and L. A. M. Groundwater quality assessment using the Water Quality Index and GIS in Saveh-Nobaran aquifer, Iran. *Environmental Earth Sciences*, 71(9):3827–3843, 2014.
- [47] D. Machiwal, V. Cloutier, and G. A review of GIS-integrated statistical techniques for groundwater quality evaluation and protection. *Environmental Earth Sciences*, 77(19):1–30, 2018.
- [48] J. Mallick, C. K. Singh, M. K. AlMesfer, V. P. Singh, and M. Alsubih. Groundwater quality studies in the kingdom of Saudi Arabia: Prevalent research and management dimensions. *Water*, 13(9):1266, 2021.
- [49] A. M. Motlagh, Z. Yang, and H. Saba. Groundwater quality. *Water Environment Research*, 92(10):28–34, 2020.
- [50] M. M.R., S. R., M. A., V. K. A., L. S., O. S., A. M., and M. A. Assessment of water quality in groundwater resources of Iran using a modified drinking water quality index (DWQI). *Ecol. Ind.*, 30:28–34, 2013.
- [51] D. C. Nedvidek. Evaluating the effectiveness of regulatory stormwater monitoring protocols on groundwater quality in urbanized karst regions. Master theses, 2014.
- [52] D. P., F. R., U. R., B. R., and N. X. Evaluation of water quality in the Chillán River (Central Chile) using physicochemical parameters and a modified water quality index. *Environ. Monitor. Assessment*, 110:301–322, 2005.
- [53] PostGIS. Spatial and Geographic objects for PostgreSQL. Open source geospatial foundation project, 2015.
- [54] J. R. Quinlan. C4.5: Programs for Machine Learning, 1993.
- [55] P. Ravenscroft and L. Lytton. Practical Manual on Groundwater Quality Monitoring, 2022.

- [56] M. O. Rivett, A. V. Miller, D. J. MacAllister, A. Fallas, G. J. Wanangwa, P. Mleta, ..., and R. M. Kalin. A conceptual model based framework for pragmatic groundwater-quality monitoring network design in the developing world: Application to the Chikwawa District, Malawi. *Groundwater for Sustainable Development*, 6:213–226, 2018.
- [57] D. Rotolo, D. Diana Hicks, and M. B.R. What Is an Emerging Technology? *Research Policy*, 2016.
- [58] J. Saraceno, J. T. Kulongoski, and T. M. Mathany. A novel high-frequency groundwater quality monitoring system. *Environmental monitoring and assessment*, 190(8):477, 2018.
- [59] M. I. e. a. Silva. Assessment of groundwater quality in a Brazilian semiarid basin using an integration of GIS, water quality index and multivariate statistical techniques. *J Hydrol*, 2021.
- [60] S. Varol and A. Davraz. Evaluation of the groundwater quality with WQI (Water Quality Index) and multivariate analysis: a case study of the Tefenni plain (Burdur, Turkey). *Environ Earth Sci*, 73(4):1725–1744, 2015.
- [61] R. S. Ward, M. J. Streetly, A. J. Singleton, and R. Sears. A framework for monitoring regional groundwater quality. *Quarterly Journal of Engineering Geology and Hydrogeology*, 37(4):271–281, 2004.
- [62] WHO. Guidelines for drinking-water quality: 4th edition incorporating first addendum, 2017.
- [63] J. Wu and Z. Sun. Evaluation of shallow groundwater contamination and associated human health risk in an alluvial plain impacted by agricultural and industrial activities, mid-west China. *Expo Health*, 8(3):311–329, 2016.
- [64] W. Wu, G. C. Dandy, and H. R. Maier. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. *Environmental Modelling and Software*, 54:108–127, 2014.
- [65] C. Z. Zulkifli, S. Garfan, M. Talal, A. H. Alamoodi, A. Alamleh, I. Y. Ahmaro, ..., and H. H. Chiang. IoT-Based Water Monitoring Systems: A Systematic Review. *Water*, 14(22):3621, 2022.