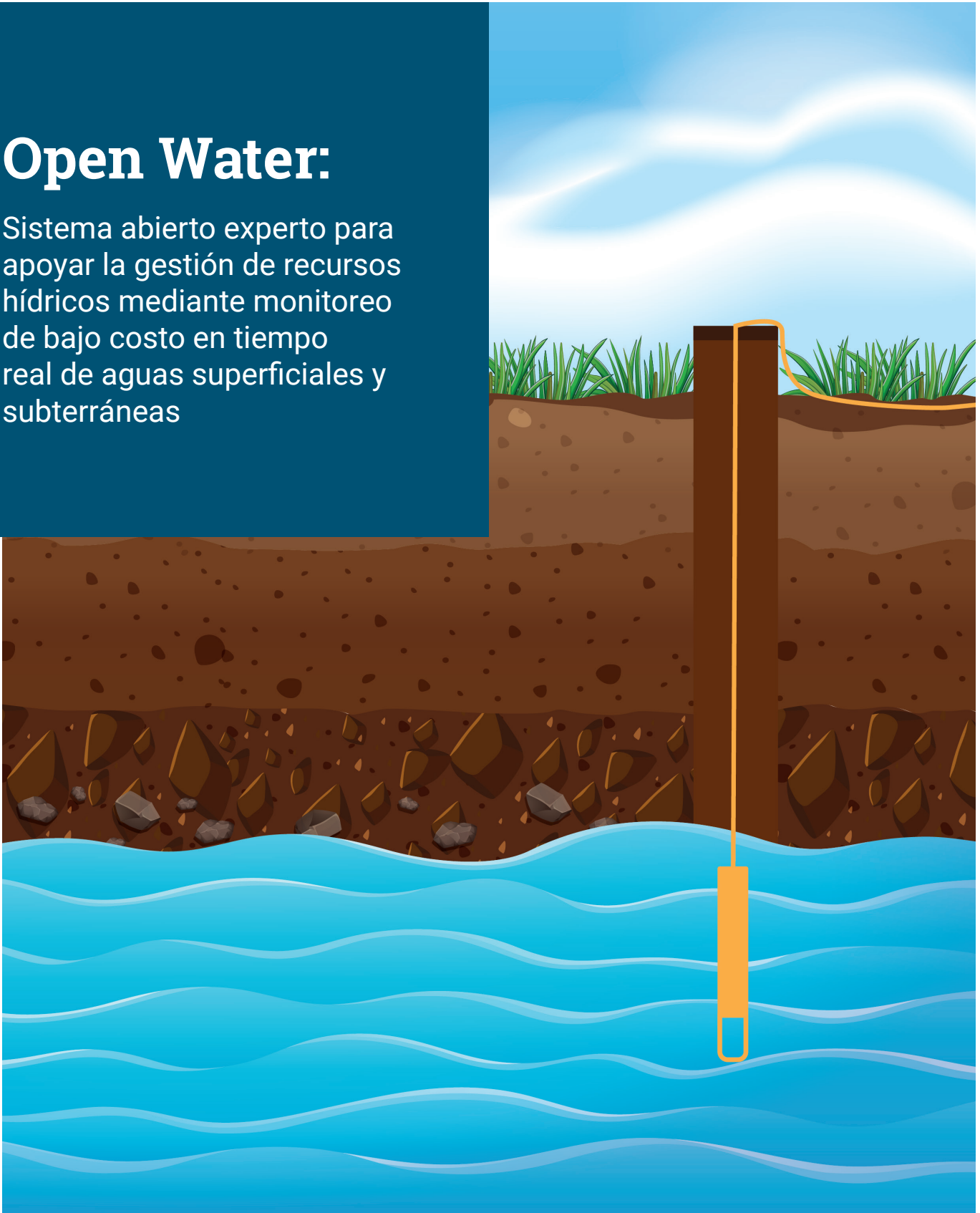




Open Water:

Sistema abierto experto para apoyar la gestión de recursos hídricos mediante monitoreo de bajo costo en tiempo real de aguas superficiales y subterráneas





SANDRA CÉSPEDES

Profesora Asistente del Departamento de Ciencias de la Computación e Ingeniería de Software, Concordia University (Canadá). Profesora Asociada, jornada parcial, del Departamento de Ingeniería Eléctrica, Universidad de Chile. Investigadora Asociada de NIC Labs Chile y del Centro Avanzado de Ingeniería Eléctrica y Electrónica (AC3E).

✉ sandra.cespedes@concordia.ca



LINDA DANIELE

Profesora Asociada del Departamento de Geología, Universidad de Chile. Investigadora del Centro de Excelencia en Geotermia de Los Andes - CEGA, del Núcleo Milenio Trazadores de Metales y Directora alterna del Centro Avanzado para Tecnologías del Agua (Consortio CAPTA).

✉ ldaniele@ing.uchile.cl



DORIS SÁEZ

Profesora Titular del Departamento de Ingeniería Eléctrica, Universidad de Chile. Directora del Programa de Pueblos Indígenas de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.

✉ dsaez@ing.uchile.cl



JAVIER BUSTOS

Director de NIC Labs, laboratorio de investigación de NIC Chile. Profesor Adjunto del Departamento de Ciencias de la Computación (DCC), Universidad de Chile.

✉ jbustos@niclabs.cl

RESUMEN. En enero del 2020, bajo el marco del proyecto FONDEF IDeA ID19110363, se unieron NIC Chile Research Labs, el Departamento de Ingeniería Eléctrica y el Departamento de Geología de la Universidad de Chile, junto con las entidades asociadas Arcadis SPA, University of Illinois at Urbana-Champaign, Esva y la Dirección General de Aguas para diseñar y desarrollar un prototipo experimental de un sistema de medición multiparamétrico abierto de bajo costo, con el cual se pueda masificar el monitoreo de la calidad de agua tanto superficial como subterránea. Este artículo da cuenta de los resultados finales del proyecto, los aciertos y desafíos en su implementación y los siguientes pasos para llevar esta tecnología hacia un producto final.

Dadas las evidencias del cambio climático, el crecimiento de la población mundial y un mayor consumo generalizado de agua, la disponibilidad y calidad de los recursos hídricos han tomado una relevancia vital. Ejemplos a nivel global muestran que la disponibilidad de agua ha disminuido a niveles críticos o reducido su calidad. En el caso de Chile, existen reportes de escasez de agua y ha sido identificado que la mitad del país no tiene los niveles mínimos de agua para un desarrollo sostenible. Un primer paso hacia ese desarrollo sostenible está en la implementación de sistemas de monitoreo de las aguas, tanto subterráneas como superficiales, que permita a través de los datos recolectados comprender un sistema hidrogeológico, con el fin de proponer una planificación y el manejo sustentable de los recursos hídricos.

Si bien en Chile existe una Red Hidrométrica Nacional que entrega datos de niveles y de calidad química básica del agua, esta todavía no logra cubrir a cabalidad las cuencas mayores identificadas en el Banco Nacional de Aguas, ni en las áreas



declaradas de restricción. En la red actual, muchos de los datos se toman manualmente, causando variaciones en la periodicidad de las lecturas y en su consiguiente publicación, sumado a que los puntos de medición son muy discretos espacialmente, lo cual en muchos casos no permite observar y detectar a tiempo eventuales alteraciones de las aguas. La Contraloría General de la República confirmó en su Informe de Auditoría 265/2021 [1], que la ausencia de una red extensa de monitoreo conlleva al desconocimiento de los acuíferos que están en explotación, a la falta de datos que permitan caracterizar la disponibilidad y calidad del agua, y a la imposibilidad de generar información y estadísticas relevantes para la gestión del recurso hídrico. Los problemas de acceso y calidad de información sobre aguas ya asignadas también impiden la implementación de estrategias medioambientales que permitan adaptarse al cambio climático [2].

Sin embargo, las deficiencias que se encuentran en la obtención y análisis de información de recursos hídricos no se deben a la inexistencia de tecnologías avanzadas de monitoreo. Los equipos existentes, altamente especializados, permiten integrar medición en continuo y el almacenamiento local de las mediciones. El mayor inconveniente es su alto costo, el cual oscila entre USD \$1.000 y USD \$8.000 por equipo. A esto se suma la imposibilidad de tener

acceso a los datos en tiempo real puesto que el difícil o retirado acceso a algunos puntos de medición, implica que la recuperación de los datos termine con altas dispersiones temporales y retraso en su procesamiento de entre 1 hasta 6 meses.

Plataforma Open Water

Para abordar esta problemática, el equipo del proyecto Open Water en el año 2019 se propuso como meta construir un prototipo para la recolección de datos en tiempo real a partir de dispositivos de monitoreo de bajo costo; una vez recibidos los datos, se debía contar con un sistema de información que incorporara conocimiento experto para detectar eventos y anomalías de forma automática, además de contar con una interfaz de visualización que permitiera la (des)agregación espacial y temporal de los datos y eventos, para facilitar la toma de decisiones. El componente “open” del proyecto consistiría en que la guía de construcción del dispositivo quedaría de uso libre para fines no comerciales, también conocido como *open hardware*. Por otro lado, la base de datos para almacenamiento y extracción posterior de los mismos, así como de los eventos detectados, también serían de uso público, en busca de mejorar el acceso a información de nuestras fuentes hídricas.

A partir de enero de 2020 se dio inicio oficial al proyecto apoyado por el programa FONDEF IDeA 2019, con la participación de NIC Chile Research Labs, el Departamento de Ingeniería Eléctrica y el Departamento de Geología de la Universidad de Chile, y en asociación con las entidades Arcadis SPA, the University of Illinois at Urbana-Champaign, Esval y la Dirección General de Aguas. Las etapas principales que se definieron para el prototipo fueron:

1. La construcción de nodos sensores (módulo de captura) usando o adaptando tecnologías existentes de bajo costo, para medición de nivel y cuatro parámetros fisicoquímicos del agua: conductividad, pH, temperatura y turbiedad.
2. La integración de un sistema de comunicaciones (módulo de comunicación) con tecnologías para Internet de las Cosas (IoT).
3. La creación de un sistema con conocimiento experto que combine técnicas de procesamiento de datos y entrenamiento, para la detección de eventos y anomalías.
4. El diseño de un componente de visualización de datos, en la forma de un tablero de control interactivo que permita la interpretación, filtrado y lectura intuitiva de los datos.

En la Figura 1 se observa el diagrama de bloques de la solución.

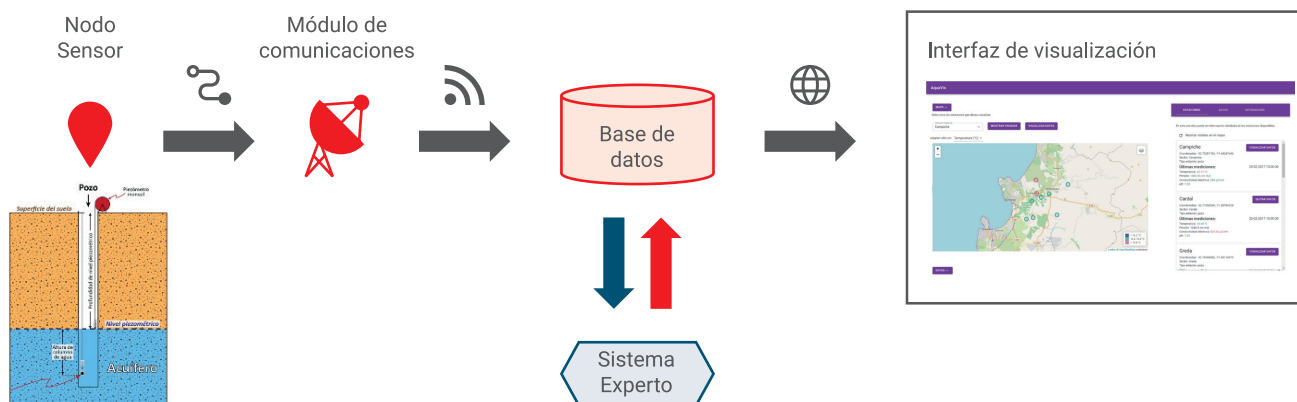


Figura 1. Diagrama de bloques Plataforma Open Water.



Construcción del Nodo Sensor

Para el desarrollo del dispositivo de monitoreo se realizó una preselección de los sensores de bajo costo que cumplirían un conjunto de requisitos mínimos, como profundidad máxima de operación, temperatura máxima, entre otros. Con este conjunto de sensores se realizó un estudio para verificar su precisión mediante comparaciones con sensores de rango alto, de uso comercial y científico. En la Figura 2 se observan algunos de los laboratorios de pruebas, que en periodo de pandemia se tuvieron que improvisar en las casas de los miembros del equipo.

En la Figura 3 se muestran ejemplos de los resultados de mediciones de los sensores de temperatura y presión y su comparación con el sensor de referencia. Además de corroborar su precisión, se realizaron experimentos que permitieran observar su comportamiento en el tiempo, para detectar posibles descalibraciones al sumergir los sensores de bajo costo por largos periodos. A partir de los experimentos se logró determinar que los sensores de temperatura, presión y pH cumplían con los requisitos de operación, arrojando errores menores a 0.5°C , 1 cm , y 1 , respectivamente, en comparación con los sensores comerciales de referencia. Por su parte, los resultados de los sensores de conductividad eléctrica y turbidez dieron cuenta de la necesidad de estudios adicionales por periodos más extendidos y con condiciones particulares para la toma de los datos.

De forma paralela se inició el desarrollo del componente electrónico que permitiría la lectura de los sensores, el empaquetamiento de los datos con su marca temporal y su envío a través de una red inalámbrica para su posterior almacenamiento en la nube. Se trató de la selección de microcontroladores y de componentes de fácil adquisición, así como el diseño de

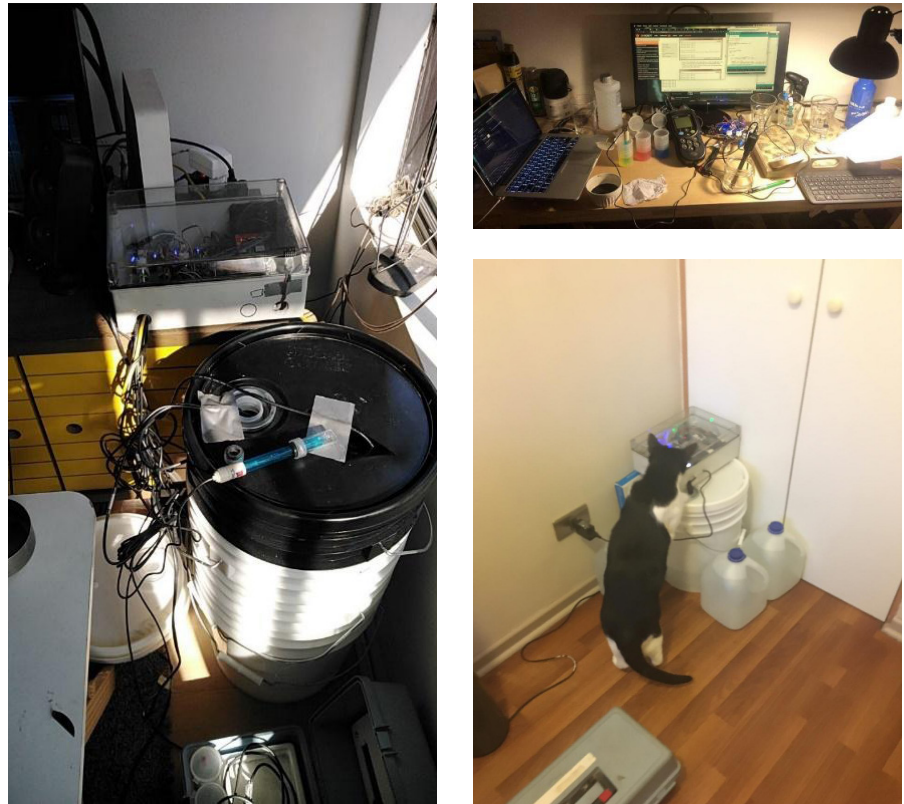


Figura 2. Laboratorios de prueba de sensores de bajo costo.

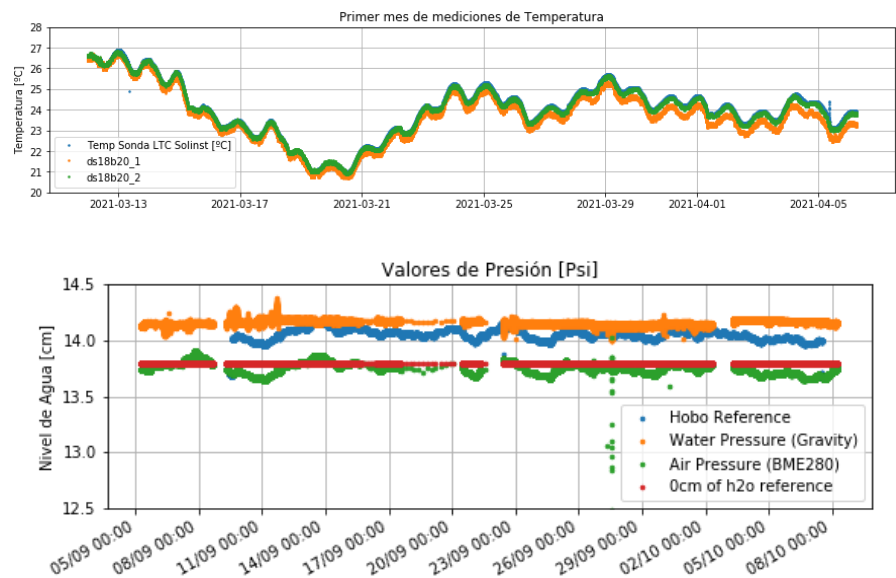
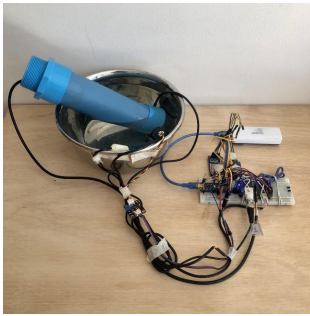
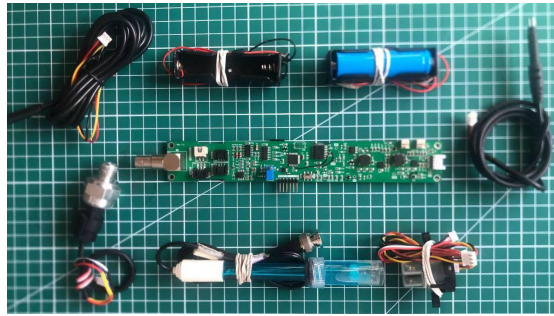


Figura 3. Ejemplo de estudios comparativos de sensores de bajo costo versus sensores de referencia.



a. Integración de componentes en tarjeta de desarrollo.



b. Versión final del prototipo con integración en PCB.

Figura 4. Evolución de la electrónica del dispositivo sensor.



a. Versión final prototipo carcasa y cable de datos.



b. Pruebas de inmersión en terreno.

Figura 5. Dispositivo sensor con carcasa y cable de datos sumergibles.

una tarjeta de circuito impreso (PCB) para la integración de todos los componentes. El diseño electrónico fue mejorando en varias iteraciones, en búsqueda de reducir su consumo energético, aligerar su carga de procesamiento y mantener los requisitos de tamaño, dadas las restricciones que se tenían en cuanto a las dimensiones del dispositivo para hacer posible su inmersión en pozos de observación. En la Figura 4 se muestra la evolución del prototipo desde su primera integración en una tarjeta de prototipado hasta su versión final con la PCB y los sensores conectados.

Para el empaquetamiento de la electrónica se trabajó en el diseño de una carcasa sumergible que mantuviera el

dispositivo en condiciones correctas de operación y soportara la presión de estar a 40 metros de profundidad. El diseño debía cumplir dos exigencias: usar materiales de fabricación de bajo costo y fácil acceso, y tener las dimensiones apropiadas para inmersión en pozos donde hubieran no más de 7 centímetros de diámetro disponibles para introducir el dispositivo. La carcasa también pasó por varias iteraciones en su diseño, selección de materiales y construcción. Finalmente se llegó al diseño mostrado en la Figura 5-a, el cual se sometió a validaciones exitosas en laboratorio a 5 metros de profundidad y pruebas iniciales en terreno (ver Figura 5-b), con inmersiones hasta 40 metros de profundidad.

Las deficiencias [...] en la obtención y análisis de información de recursos hídricos se deben al alto costo de los equipos de monitoreo y a la imposibilidad de tener acceso a los datos en tiempo real.

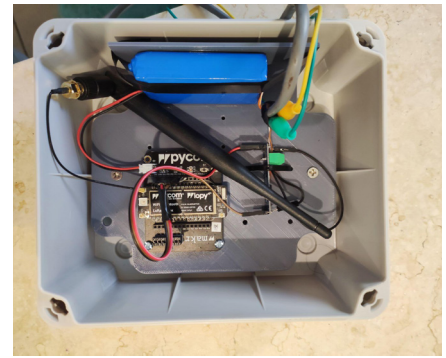


Figura 6. Módulo de conectividad del prototipo con tecnología LoRaWAN.

En cuanto al módulo de comunicación del dispositivo sensor, se realizaron estudios de tecnologías de bajo costo para verificar su cobertura y conectividad exitosa a largas distancias. Finalmente se seleccionó un módulo de prototipado con tecnología LoRaWAN que permitía distancias de más de 5 kilómetros y un despliegue de bajo costo con un punto de acceso que puede conectarse a Internet mediante diversas opciones: WiFi, LTE, Ethernet, etc. En la Figura 6 se observa la versión final del módulo de conectividad a instalar en la superficie y que se conecta al dispositivo sensor mediante un cable de datos sumergible desarrollado por el equipo de ingeniería y diseño del proyecto.

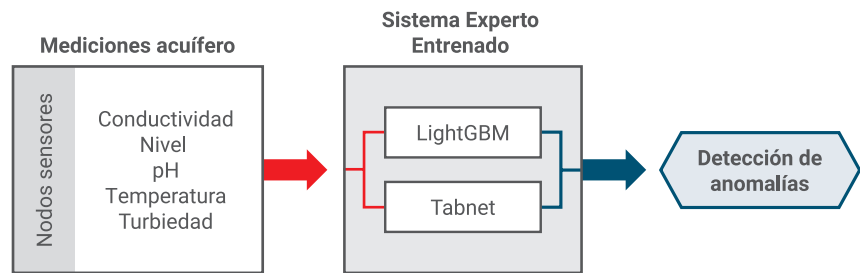


Construcción del Sistema Experto

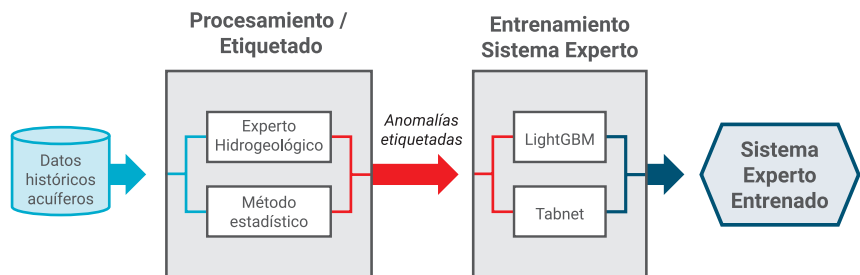
El monitoreo continuo es esencial para comprender el funcionamiento de un sistema hidrogeológico. El análisis de los datos recolectados permite que experto/as en recursos hídricos logren entender el funcionamiento natural de dicho sistema y distinguirlo de aquellos comportamientos anómalos. Algunos ejemplos de anomalías en los sistemas hidrogeológicos pueden ser los periodos de sequía, donde la disponibilidad, así como también la calidad del recurso hídrico, pueden verse afectados notablemente. Otro ejemplo también son los eventos de contaminación, donde la composición química del agua se ve afectada. De este modo, una rápida detección ante instancias anómalas es crítica para la planificación y toma de decisiones, pero esta tarea se torna compleja para el/la experto/a en sistemas de monitoreo continuo donde hay una gran cantidad de datos almacenados. Debido a esto, se propuso la implementación del *sistema experto*, que incorpora un *software* capaz de analizar los datos obtenidos desde los acuíferos de forma automática.

El diseño consta de la integración de dos herramientas (o modelos) de *machine learning* entrenadas para trabajar en paralelo, y cuyas salidas son fusionadas mediante una estrategia determinada. Las herramientas de *machine learning* escogidas para el prototipo fueron LightGBM [3] y TabNet [4]. La Figura 7-a ilustra los componentes del sistema con sus entradas y sus salidas. Es importante mencionar que el Sistema Experto no se limita solamente a utilizar los modelos LightGBM y TabNet, ya que por su diseño flexible se pueden incorporar otros tipos de modelos para trabajar en paralelo juntos a los ya propuestos.

Para la incorporación de conocimiento experto humano en el sistema, este se



a. Entradas y salidas del sistema en línea.



b. Esquema de entrenamiento del sistema.

Figura 7. Componentes del sistema experto en operación.

incluyó en la etapa de preprocesamiento de los datos históricos para ser utilizados en el entrenamiento de los modelos de *machine learning*, proceso que es ilustrado en la Figura 7-b. Este preprocesamiento corresponde al etiquetado de datos realizado por el/la experto/a y a la selección del conjunto de datos históricos más adecuado para el entrenamiento de los modelos de detección de anomalías. De manera adicional, y con el fin de tener un etiquetado robusto, el preprocesamiento de los datos incluye un método estadístico que permite obtener datos etiquetados mediante otro criterio, y que pueden utilizarse como complemento a los datos previamente etiquetados por el/la experto/a.

Al realizar la evaluación del sistema implementado con una base de datos histórica y debidamente etiquetada por expertos/as, se determinó que los dos modelos, por separado, tienen un desempeño por sobre un 75% en prácticamente todas las métricas usadas: *Accuracy*, *F1-Score*, *Precision*, *Recall*, ex-

cepto el *Recall* que registró un 73% para Tabnet. Sin embargo, al fusionar ambos modelos, todas las métricas propuestas sobrepasan el 75% y, considerando solo el *Accuracy*, este logró alcanzar cerca del 95%. La Figura 8 muestra de forma gráfica la efectividad de la detección usando una base de datos de validación y comparando las anomalías reales (etiquetadas) a las anomalías detectadas automáticamente mediante la fusión de los modelos LightGBM y TabNet a través de la función lógica OR.

La última parte del sistema de información corresponde a la interfaz de visualización. El diseño de la interfaz gráfica se delineó bajo tres ejes principales: estructura de la interfaz gráfica, usabilidad, y experiencia de usuario. Estos ejes apuntan a desplegar la mayor información de interés para el usuario, permitiéndole interactuar con los datos medidos desde los acuíferos de una manera intuitiva y similar a otras plataformas existentes que cumplan los mismos propósitos. El diseño fue propuesto por el investigador



Juan Salamanca de la Universidad de Illinois (Urbana, Champaign, Estados Unidos), institución asociada del proyecto. El diseño se adaptó en diversas iteraciones con clientes potenciales del sistema. Finalmente, la funcionalidad de la interfaz fue desarrollada 100% *in-house* por el equipo de desarrollo del proyecto.

El *front end* fue desarrollado en React, biblioteca de JavaScript que ayuda a crear interfaces de usuario interactivas. Se utilizaron las librerías Leaflet para visualizar e interactuar con el mapa, y Plotly.js para mostrar los gráficos con sus interacciones. React se renderizó usando Node. Por otro lado, en el *back end*, la obtención de los datos de los sensores instalados en los acuíferos se realizó a través de una *interfaz de programación de aplicaciones* (API, por sus siglas en inglés) desarrollada en Flask. La API se conecta con las bases de datos: influx-DB para las series temporales y MySQL para todo lo demás, y genera rutas de acceso que son posteriormente utilizadas en la interfaz web para obtener los datos solicitados. Esto permite que el *front end* pueda tener acceso a los datos cuando sea necesario, y de este modo, dejarlos a disposición del cliente.

Para verificar el funcionamiento de la interfaz gráfica implementada, a la cual se le llamó AquaVis (Aqua Visualization), se generaron datos sintéticos que se relacionaron a nueve acuíferos artificiales (puntos geográficos que representaban de manera ficticia la existencia de un acuífero en ese lugar). Esto permitió realizar una simulación de la interacción de la interfaz gráfica con la base de datos que contiene los datos medidos en los pozos monitoreados.

En la Figura 9 se observa la pantalla principal de la interfaz gráfica implementada. Además, se visualizan los tres paneles principales que componen la interfaz. En (1) aparece el mapa interactivo para que el usuario pueda buscar los acuíferos en los cuales se encuentra un dispositivo de monitoreo. En (2) se

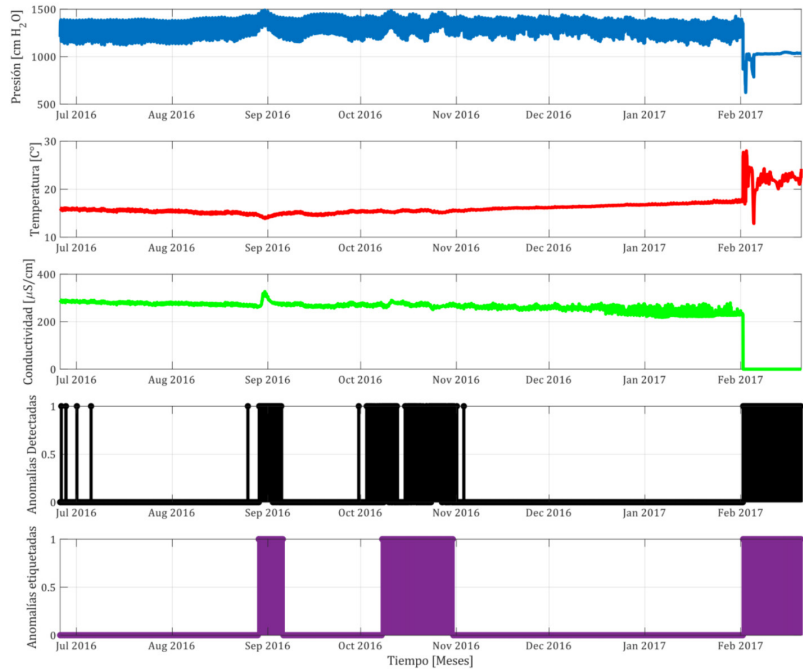


Figura 8. Base de datos de validación junto con las anomalías detectadas por la fusión de LightGBM y TabNet (color negro) y las anomalías previamente etiquetadas (color morado).

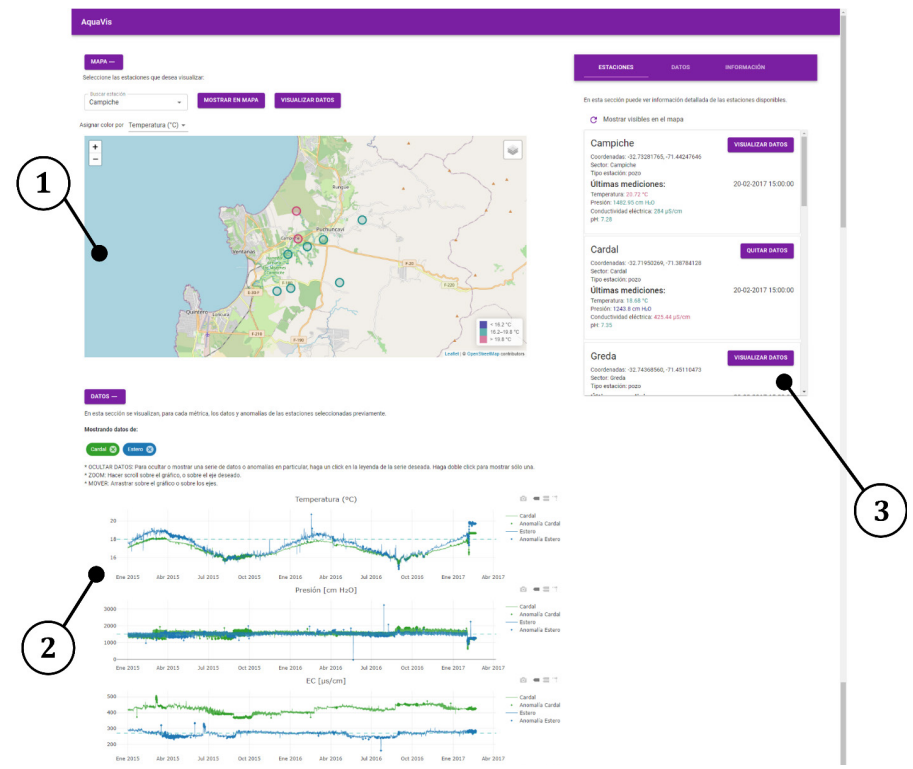


Figura 9. Pantalla principal de AquaVis.



El diseño electrónico [del nodo sensor] fue mejorando en varias iteraciones, en búsqueda de reducir su consumo energético, aligerar su carga de procesamiento y mantener los requisitos de tamaño.

despliegan las series de tiempo de las variables que son medidas en los pozos monitoreados. Por último, en (3) se presenta un panel de información de los pozos que son visualizados en el mapa interactivo. La interfaz se encuentra disponible para su uso con datos de prueba en <http://agua.niclabs.cl:3001/>.

Siguientes pasos

El diseño, implementación y pruebas del prototipo de la Plataforma Open Water tuvieron grandes aciertos, pero también muchos desafíos. El desarrollo durante la pandemia de COVID-19 obligó al trabajo colaborativo de forma remota durante el total del tiempo del proyecto. Los laboratorios de pruebas tuvieron que trasladarse a casas y en contadas ocasiones se logró acceso a pruebas en terreno, debido a las múltiples restricciones durante las cuarentenas y desplazamientos controlados fuera de la Región Metropolitana. Sin embargo, la motivación del equipo fue grande y así se logró tener un prototipo 100%

funcional con un costo inicial unitario, sin producción en volumen, de aproximadamente USD \$500. El prototipo muestra la operación de la plataforma desde el punto de medición en la fuente hasta el procesamiento y visualización de los datos en una plataforma interactiva y de alta usabilidad para los usuarios gestores del recurso hídrico. En los próximos pasos, el equipo de investigación y desarrollo planea continuar con pruebas de los sensores que mostraron un menor desempeño (i.e., conductividad eléctrica y turbidez), así como ajustes al diseño de la carcasa sumergible. La placa electrónica se adaptará para cumplir funciones acotadas que nos permitan jugar con dispositivos sensores que contengan sensores para medir solo un subconjunto de los parámetros, tratando de atender necesidades de proyectos que requieren mediciones puntuales. En cuanto al sistema de información y visualización, esperamos abordar los retos de escalabilidad para soportar la incorporación de un número alto de dispositivos, pero también la conexión de estaciones diferentes a la nuestra, para aquellas iniciativas que

busquen poner a disposición sus datos en la plataforma abierta de Open Water.

Este proyecto contó con el arduo trabajo y compromiso de un equipo multidisciplinario de investigación y desarrollo conformado por Matías Taucare, Francisco Jaramillo, Benoit Viguier y Miguel Solís (postdocs), Rodrigo Muñoz, Matías Macaya y Felipe Lalane (ingenieros de investigación y desarrollo), Gabriela Mendoza, Gabriel Flores, Sebastián Cifuentes, Maximiliano Jones y Carlos González (memoristas y tesisistas), María Jesús Ugarte, Maite González, Michael Alan Acevedo, Pablo Martín, Amilcar Aravena y Gonzalo Maureira (asistentes de investigación y practicantes). También hicieron parte Felipe Astaburuaga (propiedad intelectual) y Patricia Martínez (asistente administrativa), y el equipo de +D en el diseño de la carcasa. Las empresas asociadas tuvieron una comprometida participación a través de Juan Salamanca (University of Illinois at Urbana, Champaign), Ernesto Ramírez y Nicolás Ramírez (Arcadis SpA), Mariela Vicencio, Bárbara Zurita y Daniel Velásquez (Esval), y Andrea Osses y Mauricio Lorca (DGA). Partes de este artículo están basados en los reportes de hitos elaborados por el equipo y por la/os investigadora/es para presentación ante FONDEF. Este proyecto fue financiado por ANID FONDEF Concurso IDeA I+D 2019 Folio ID19110363 y recibió apoyo del Centro Avanzado en Ingeniería Eléctrica y Electrónica, AC3E, Proyecto Basal FB0008, ANID. ■

REFERENCIAS

- [1] Contraloría General de la República de Chile. "Informe final 265-2021 Auditoría a las acciones desarrolladas por la DGA del MOP", 2022. <https://www.contraloria.cl/pdfbuscador/auditoria/57a62247294b58c216d729f456e2987c/html>.
- [2] P. Barría, M. Rojas, P. Moraga, A. Muñoz, D. Bozkurt, and C. Álvarez, "Anthropocene and streamflow: Long-term perspective of streamflow variability and water rights," *Elem Sci Anth*, vol. 7, no. 1, p. 2, Jan. 2019.
- [3] G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, pp. 3146-3154. 2017.
- [4] A. Shafi, "TabNet: The End of Gradient Boosting?," in *Towards Data Science*. 2021. <https://towardsdatascience.com/tabnet-e1b979907694>.