



Tecnológico Nacional de México

Centro Nacional de Investigación y Desarrollo Tecnológico

Tesis de Maestría

**Desarrollo de un algoritmo de limpieza de datos para
dispositivos del Internet de las Cosas Subterráneas**

presentada por
Ing. Julio Víctor Sánchez Hernández

como requisito para la obtención del grado de
Maestro en Ciencias de la Computación

Director de tesis
Dr. Hugo Estrada Esquivel

Codirector de tesis
Dr. Fernando Pech May

Cuernavaca, Morelos, México. Agosto de 2025



Educación
Secretaría de Educación Pública



TECNOLÓGICO
NACIONAL DE MÉXICO



Centro Nacional de Investigación y Desarrollo tecnológico
Departamento de Ciencias Computacionales

Cuernavaca, Mor., 16/Junio/2025

OFICIO No. DCC/190/2025

Asunto: Aceptación de documento de tesis
CENIDET-AC-004-M14-OFICIO

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial de **Julio Víctor Sánchez Hernández** con número de control **M23CE084**, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis de grado titulado **"Desarrollo de un algoritmo de limpieza de datos para dispositivos del Internet de las Cosas Subterráneas"** y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

ATENTAMENTE

Excelencia en Educación Tecnológica®
"Conocimiento y Tecnología al Servicio de México"

Dr. Hugo Estrada Esquivel
Director de tesis

Dra. Alicia Martínez Rebollar
Revisor 1

Dr. Fernando Pech May
Codirector de tesis

Dr. Jarniel García Morales
Revisor 2

C.c.p. Depto. Servicios Escolares.
Expediente / Estudiante



2025
Año de
La Mujer
Indígena

Interior Internado Palmira S/N, Col. Palmira,
C. P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 3201,
e-mail: dcc_cenidet@tecnm.mx tecnm.mx | cenidet.tecnm.mx

cenidet
Centro Nacional de Investigación
y Desarrollo Tecnológico





Centro Nacional de Investigación y Desarrollo tecnológico
Subdirección Académica

Cuernavaca Mor, 23/junio/2025

Oficio No. SAC/147/2025

Asunto: Autorización de impresión de tesis

JULIO VÍCTOR SÁNCHEZ HERNÁNDEZ
CANDIDATO AL GRADO DE MAESTRO
EN CIENCIAS DE LA COMPUTACIÓN
P R E S E N T E

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Desarrollo de un algoritmo de limpieza de datos para dispositivos del Internet de las Cosas Subterráneas"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

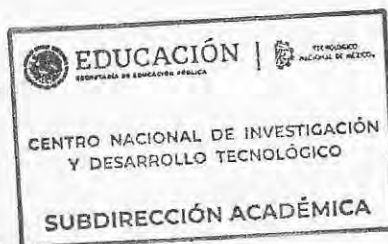
Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica®

"Conocimiento y Tecnología al Servicio de México"

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO



c.c.p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz



2025
Año de
La Mujer
Indígena

Interior Internado Palmira S/N. Col. Palmira.

C.P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 4104,
e-mail: ac2d.cenidet@tecnm.mx | tecnm.mx | cenidet.tecnm.mx

cenidet
Centro Nacional de Investigación y Desarrollo Tecnológico



Dedicatoria

*Dedico esta tesis a Dios, por ser mi guía, mi fuerza y mi esperanza en cada momento del camino.
A mis padres, por su amor incondicional, su ejemplo de esfuerzo y por estar siempre a mi lado
con palabras de aliento y confianza. Y a quienes me inspiraron a seguir adelante con fe, pasión
y convicción, incluso en los momentos más difíciles.*

Agradecimientos

Al concluir esta nueva etapa de superación y preparación profesional, quiero expresar mi más profundo agradecimiento a quienes hicieron posible este logro y me acompañaron a lo largo del camino.

- A **DIOS**, por su guía constante, por darme fuerza en los momentos difíciles y por iluminar cada paso de este proceso.
- A la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), por el respaldo económico brindado durante mi formación, el cual fue vital para continuar mis estudios.
- Al Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), por ofrecerme un espacio académico de excelencia y crecimiento profesional.
- A mis padres, Magdalena Hernández y Julio Sánchez, por su amor, valores y apoyo incondicional, y por brindarme las bases para alcanzar una formación profesional.
- A mi director de tesis, el Dr. Hugo Estrada Esquivel, por su guía académica, su compromiso inquebrantable y la confianza depositada en mí desde el inicio.
- A mi codirector, el Dr. Fernando Pech May, por su amistad, consejos e impulso constante para seguir adelante. Su cercanía y apoyo fueron fundamentales en este proceso.
- A mis revisores, la Dra. Alicia Martínez Rebollar y el Dr. Jarniel García Morales, por su tiempo, disposición y valiosas observaciones. Su mirada crítica y sus comentarios constructivos fortalecieron significativamente este trabajo y contribuyeron al logro de una versión más sólida y rigurosa.
- A mis queridas Jessica Sánchez e Iris Díaz, por su amor, apoyo incondicional y compañía en cada momento de este camino. Su presencia fue un impulso invaluable para seguir adelante.
- A mis amigos y a todas las personas significativas en mi vida que han contribuido de manera positiva, gracias por su compañía y aliento en los momentos clave de esta etapa.

A todos ustedes, mi gratitud sincera. Este logro también es suyo.

Tabla de contenido

Resumen	iii
Abstract	iv
1. Introducción	1
1.1. Introducción	1
1.2. Planteamiento del problema	3
1.3. Objetivos.....	4
1.3.1. Objetivo general	4
1.3.2. Objetivos específicos.....	4
1.4. Metodología de solución	5
1.4.1. Construcción del nodo sensor subterráneo.....	5
1.4.2. Desarrollo del algoritmo de limpieza de datos en flujo	6
1.4.3. Evaluación del sistema integrado de limpieza de datos	7
1.5. Estructura del documento	7
2. Fundamentación teórica y estado del arte	8
2.1. Fundamentación teórica	8
2.1.1. Internet de las Cosas	8
2.1.2. Internet de las Cosas Subterráneas	10
2.1.3. Limpieza de datos en flujo.....	12
2.2. Estado del arte	13
2.2.1. Clasificación y criterios de análisis de los trabajos de investigación	13
2.2.2. Sistemas del Internet de las Cosas en la Agricultura de Precisión.....	14
2.2.3. Sistemas del Internet de las Cosas Subterráneas en la Agricultura de Precisión	16
3. Construcción del nodo sensor subterráneo	24
3.1. Selección técnica de componentes de hardware.....	24
3.2. Diseño esquemático del nodo sensor subterráneo.....	28
3.3. Ensamble del nodo sensor subterráneo	30

4 Desarrollo del algoritmo de limpieza de datos en flujo	34
4.1. Definición de requerimientos del algoritmo	34
4.2. Codificación del algoritmo de limpieza de datos.....	44
4.3. Integración del algoritmo al nodo sensor subterráneo.....	46
 5 Evaluación del sistema integrado de limpieza de datos.....	 49
5.1. Evaluación experimental.....	49
5.1.1. Diseño y ejecución de escenarios experimentales.....	53
5.1.2. Evaluación funcional del sistema integrado	55
5.2. Resultados obtenidos	56
5.2.1. Subconjunto de datos con 10% de anomalías	56
5.2.2. Subconjunto de datos con 30% de anomalías	60
5.2.3. Subconjunto de datos con 50% de anomalías	64
 6 Conclusiones y trabajos futuros.....	 69
6.1. Productos derivados de la investigación	70
 Referencias	 72

Resumen

El Internet de las Cosas Subterráneas representa una extensión innovadora del paradigma del Internet de las Cosas, orientada al monitoreo en tiempo real de variables subterráneas relevantes para la Agricultura de Precisión. Sin embargo, la operación de dispositivos en entornos bajo tierra impone restricciones severas en términos de comunicación, energía y procesamiento, lo que dificulta garantizar la calidad de los datos recolectados. En este contexto, la presente tesis propone el diseño e implementación de un algoritmo para la limpieza autónoma de datos en flujo, implementado sobre un nodo sensor subterráneo de bajo consumo energético y orientado a operar directamente en campo.

El sistema embebido integra un nodo sensor subterráneo con sensores de temperatura, humedad y pH del suelo, un microcontrolador de arquitectura abierta y un algoritmo de limpieza que detecta, imputa y depura datos anómalos en tiempo real. La solución fue evaluada experimentalmente mediante la inyección controlada de anomalías (datos atípicos, duplicados y faltantes) en tres subconjuntos con 10 %, 30 % y 50 % de alteraciones. Los resultados demuestran que el algoritmo detectó entre 80 % y 92 % de los valores atípicos, imputó hasta el 96 % de los datos inconsistentes y eliminó prácticamente todos los duplicados, incluso en escenarios con alta perturbación.

Estos hallazgos confirman la viabilidad de implementar técnicas de preprocesamiento en plataformas con recursos limitados, lo que representa un aporte significativo para el desarrollo de soluciones robustas, autónomas y eficientes en aplicaciones agrícolas que demandan alta confiabilidad de datos.

Abstract

The Internet of Underground Things represents an innovative extension of the Internet of Things paradigm, aimed at the real-time monitoring of subsurface variables relevant to Precision Agriculture. However, operating devices in underground environments imposes severe constraints on communication, energy availability, and computational capacity, which significantly complicates the assurance of data quality. In this context, the present thesis proposes the design and implementation of an autonomous data stream cleaning algorithm, deployed on a low-power underground sensor node designed for direct in-field operation.

The embedded system integrates a subterranean sensor node equipped with temperature, humidity, and soil pH sensors, an open-architecture microcontroller, and a cleaning algorithm capable of detecting, imputing, and filtering anomalous data in real time. The solution was experimentally validated through the controlled injection of anomalies (namely outliers, duplicates, and missing values) into three subsets containing 10%, 30%, and 50% of altered data. Results show that the algorithm detected between 80% and 92% of outliers, imputed up to 96% of inconsistent values, and effectively eliminated nearly all duplicates, even under high-noise conditions.

These findings confirm the feasibility of implementing preprocessing techniques on resource-constrained platforms, representing a significant contribution to the development of robust, autonomous, and efficient solutions for agricultural applications requiring high data reliability.

1

Introducción

En este capítulo se contextualiza el área de estudio, se plantea el problema, se definen el objetivo general y los objetivos específicos de la investigación, y se expone la metodología de solución propuesta. En conjunto, estos componentes constituyen el marco inicial para el desarrollo de la tesis y fundamentan su pertinencia en el contexto del Internet de las Cosas Subterráneas aplicado a la Agricultura de Precisión.

1.1. Introducción

El avance de las tecnologías digitales ha transformado profundamente el sector agrícola, destacando la incorporación de sistemas inteligentes conectados que permiten la recolección, transmisión y análisis en tiempo real de variables agroambientales. Estos avances han dado lugar a una agricultura más eficiente, resiliente y sostenible (Alahmad y otros, 2023).

El Internet de las Cosas describe redes de dispositivos físicos interconectados capaces de operar de forma autónoma, recolectando y compartiendo datos. En el ámbito agrícola, esta tecnología ha impulsado el surgimiento de la agricultura 4.0, basada en el análisis intensivo de datos para optimizar prácticas como el riego, la fertilización y el monitoreo del suelo y del clima (Gubbi y otros, 2013) (Wolfert y otros, 2017).

Dentro de esta evolución surge el Internet de las Cosas Subterráneas, una tecnología enfocada en recolectar datos desde entornos bajo superficie mediante sensores inalámbricos enterrados, integrados con algoritmos de procesamiento local. Esta

tecnología permite monitorear variables como humedad, temperatura, salinidad y pH en profundidad, aportando información crítica para la Agricultura de Precisión (Cariou y otros, 2023)

La implementación de sistemas del Internet de las Cosas Subterráneas busca optimizar la productividad y el uso racional de recursos, alineándose con los Objetivos de Desarrollo Sostenible, al mejorar la seguridad alimentaria, fomentar infraestructuras digitales resilientes y promover una gestión eficiente de insumos (Alahmad y otros, 2023). No obstante, el despliegue de sensores subterráneos enfrenta retos técnicos significativos: heterogeneidad del suelo, atenuación de señales, interferencias, restricciones energéticas y alta variabilidad en los datos recolectados. Estos desafíos exigen soluciones que permitan realizar procesamiento de datos directamente en los nodos del Internet de las Cosas, sin que se requiera del envío de datos a un servidor externo, un ejemplo de ello es poder permitir que los dispositivos validen y limpien los datos obtenidos de los sensores directamente en los nodos, sin depender de infraestructura externa.

El diseño de algoritmos especializados para la validación y limpieza de datos en tiempo real es esencial para lograr la independencia de los dispositivos del Internet de las Cosas Subterráneas. Dichos algoritmos deben ser eficientes, robustos y adecuados para operar bajo limitaciones de energía, memoria y procesamiento, con el fin de asegurar la integridad de la información utilizada en la toma de decisiones agrícolas.

En este contexto, la presente tesis se enfoca en el desarrollo de un algoritmo de limpieza de datos en flujo para dispositivos del Internet de las Cosas Subterráneas basados en microcontroladores de arquitectura abierta. El trabajo de investigación incluyó el diseño del nodo sensor subterráneo, el desarrollo del algoritmo de limpieza de datos y la validación del sistema integrado en el entorno subterráneo. Con esto se busca garantizar la confiabilidad de la información recolectada para su aplicación en la Agricultura de Precisión.

1.2. Planteamiento del problema

La incorporación de tecnologías emergentes ha impulsado mejoras significativas en la supervisión y optimización de procesos agrícolas. Sin embargo, en entornos subterráneos persisten barreras técnicas que dificultan la obtención de datos confiables. En respuesta a esta problemática, el Internet de las Cosas Subterráneas ha surgido como una evolución del paradigma convencional, orientada al monitoreo en tiempo real de variables subterráneas como humedad, temperatura, salinidad y pH, con aplicaciones directas en la Agricultura de Precisión.

En el área de las Ciencias Computacionales, el desarrollo de sistemas embebidos de bajo consumo ha facilitado la adquisición y transmisión de datos en redes distribuidas, mediante la integración de sensores, microcontroladores y módulos de comunicación. No obstante, los entornos subterráneos presentan condiciones hostiles que afectan la calidad de los datos recolectados, generando anomalías como valores atípicos, duplicados o ausentes, producto de fallas de sensores, interferencias o pérdida de conectividad.

A pesar de los avances en técnicas de validación de datos, la mayoría de las soluciones existentes dependen de infraestructuras con amplios recursos computacionales y energéticos. Esto limita su aplicabilidad en campo, donde es necesario contar con algoritmos que se ejecuten localmente en dispositivos con recursos restringidos, sin comprometer autonomía ni eficiencia. Actualmente, son escasos los enfoques que abordan esta necesidad desde una perspectiva de ejecución embebida en tiempo real. Frente a este contexto, la investigación plantea la siguiente pregunta central: ¿Cómo desarrollar e implementar un algoritmo de limpieza de datos en flujo eficiente, autónomo y adaptable que pueda ejecutarse directamente en dispositivos subterráneos con recursos limitados, garantizando la calidad de los datos capturados para Agricultura de Precisión? Para responder esta pregunta, se definen tres preguntas de investigación:

- ¿Qué técnicas existentes en la literatura pueden adaptarse o rediseñarse para ejecutarse eficientemente en redes subterráneas?

- ¿Cómo asegurar la continuidad y confiabilidad de las series temporales en entornos agrícolas bajo superficie?
- ¿Qué impacto tiene la implementación local del sistema de limpieza en flujo sobre la autonomía de los nodos sensores subterráneos?

Dar respuestas a estas interrogantes permitirá diseñar, implementar y validar un algoritmo integral de limpieza de datos en flujo que opere sobre un nodo sensor subterráneo con condiciones de cómputo restringido. El enfoque propuesto en estas tesis contribuirá al avance de los sistemas embebidos aplicados al Internet de las Cosas Subterráneas y ofrecerá una solución concreta para mejorar la calidad de los datos en Agricultura de Precisión, beneficiando a productores, desarrolladores y responsables de políticas públicas orientadas al uso sostenible del suelo.

1.3. Objetivos

1.3.1. Objetivo general

Desarrollar un sistema de limpieza de datos para un nodo sensor subterráneo que mejore la precisión, consistencia y utilidad de los datos recolectados en aplicaciones del Internet de las Cosas Subterráneas, utilizando un microcontrolador de código abierto y técnicas estadísticas de limpieza de datos en tiempo real.

1.3.2. Objetivos específicos

- Diseñar y construir un prototipo funcional de nodo sensor subterráneo para la captura de variables físicas y químicas del suelo, empleando un microcontrolador de código abierto y componentes electrónicos compatibles, garantizando su operación en condiciones ambientales propias del subsuelo agrícola.
- Desarrollar un algoritmo de limpieza de datos específico para ser ejecutado en dispositivos subterráneos, capaz de realizar detección y corrección de datos atípicos, imputación de datos faltantes y eliminación de duplicados, utilizando técnicas estadísticas optimizadas para las restricciones de capacidad computacional del nodo sensor.

- Evaluar y validar experimentalmente la eficacia del sistema para determinar el impacto del algoritmo de limpieza de datos en la mejora de la calidad de la información recolectada mediante la ejecución de pruebas en entornos subterráneos y utilizando métricas de calidad de datos como precisión, completitud, consistencia y reducción de errores, de acuerdo con estándares establecidos para sistemas del Internet de las Cosas Subterráneas.

1.4. Metodología de solución

La metodología empleada para el desarrollo de esta tesis se dividió en tres fases: construcción del nodo, desarrollo del algoritmo de limpieza de datos y evaluación en condiciones reales (ver Figura 1).

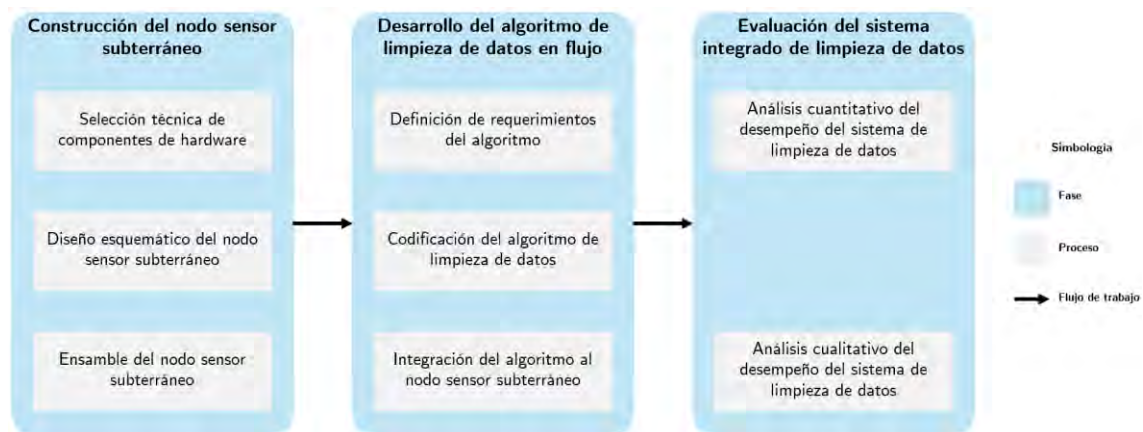


Figura 1. Metodología de solución implementada para el desarrollo del algoritmo de limpieza de datos en flujo. *Fuente: Elaboración propia.*

1.4.1. Construcción del nodo sensor subterráneo

La primera fase de la metodología comprendió el diseño y construcción del nodo sensor subterráneo, concebido como una unidad autónoma para la adquisición y procesamiento de datos ambientales del subsuelo. Esta fase se dividió en tres etapas: selección de componentes, diseño esquemático y ensamblaje con validación preliminar. Esta fase inició con una revisión técnica del estado del arte sobre sensores para variables fisicoquímicas del suelo y plataformas de procesamiento embebido. Con base en ello, se aplicó una evaluación multicriterio, inspirada en el enfoque de Salam y Raza (2020), que

permitió comparar opciones considerando precisión, robustez, eficiencia energética, compatibilidad eléctrica y facilidad de integración.

Una vez definidos los componentes, se elaboró el diseño esquemático en Fritzing del nodo sensor subterráneo, modelando las conexiones entre sensores, unidad de procesamiento, almacenamiento, fuente de alimentación y periféricos. Este diagrama sirvió de guía para el ensamblaje del prototipo físico, el cual fue encapsulado en una carcasa IP65 para asegurar su operación en condiciones subterráneas. Finalmente, se realizaron pruebas funcionales en un entorno controlado, verificando la adquisición y almacenamiento estructurado de datos en archivos CSV.

1.4.2. Desarrollo del algoritmo de limpieza de datos en flujo

La segunda fase se enfocó en el diseño e integración del algoritmo de limpieza de datos capaz de operar en tiempo real y de forma autónoma dentro del nodo sensor subterráneo. El objetivo fue asegurar la calidad de los datos desde su origen, sin requerir validaciones externas, incluso en entornos adversos. En esta fase se analizaron técnicas estadísticas para flujos continuos, priorizando precisión, eficiencia energética y baja complejidad computacional. Con base en ello, se definieron los requerimientos del sistema siguiendo el estándar IEEE/ISO/IEC 29148-2018, estableciendo criterios de confiabilidad, robustez y compatibilidad con el hardware embebido. El algoritmo se implementó considerando las limitaciones del microcontrolador, utilizando un entorno optimizado para bajo consumo.

Una vez integrado el algoritmo desarrollado al nodo sensor subterráneo, se verificó la correcta secuencia de adquisición, limpieza y almacenamiento de datos, así como la gestión energética del sistema. Finalmente se realizaron pruebas en un entorno subterráneo controlado, evaluando detección de atípicos, imputación de faltantes y eliminación de duplicados. Los resultados permitieron afinar el rendimiento del algoritmo y confirmar su estabilidad operativa.

1.4.3. Evaluación del sistema integrado de limpieza de datos

La tercera fase tuvo como objetivo evaluar el desempeño del sistema de limpieza de datos en condiciones subterráneas reales, validando su eficacia y autonomía. Se analizó tanto el comportamiento del algoritmo como la estabilidad operativa del nodo sensor subterráneo. En una etapa inicial, se evaluó la capacidad del sistema para mejorar la calidad de los datos mediante el análisis de registros operativos, enfocándose en la detección de valores atípicos, la imputación de datos faltantes y la eliminación de duplicados. Para ello, se emplearon métricas como la tasa de detección de anomalías, falsos positivos y negativos, porcentaje de registros imputados y efectividad en la deduplicación. Estas métricas fueron analizadas de forma integrada para estimar la precisión, sensibilidad y eficiencia general del sistema. Los resultados obtenidos se sintetizaron en tablas resumen, facilitando su interpretación y permitiendo contrastar el desempeño del algoritmo en escenarios operativos reales.

De manera complementaria, se realizó un análisis cualitativo mediante la inspección visual de los registros generados, utilizando gráficas temporales que permitieron identificar momentos clave de intervención del algoritmo.

1.5. Estructura del documento

La presente tesis se organiza en seis capítulos. En el Capítulo 1 se introduce el problema de investigación, se establecen los objetivos y se describe la metodología empleada. El Capítulo 2 aborda la fundamentación teórica sobre Internet de las Cosas, su variante subterránea y la limpieza de datos en flujo, complementado con un análisis del estado del arte. El Capítulo 3 describe el proceso de diseño, selección de componentes y ensamblaje del nodo sensor subterráneo. En el Capítulo 4 se detalla el desarrollo del algoritmo de limpieza de datos, desde su codificación hasta su integración en el sistema embebido. El Capítulo 5 presenta la evaluación del algoritmo integrado a partir de escenarios experimentales. Finalmente, el Capítulo 6 expone las conclusiones del trabajo de investigación y los productos derivados.

Fundamentación teórica y estado del arte

Este capítulo presenta los conceptos fundamentales sobre el Internet de las Cosas, Internet de las Cosas Subterráneas y limpieza de datos en flujo, así como una revisión de trabajos de investigación recientes en estas áreas. La revisión permite identificar avances, limitaciones y vacíos tecnológicos que justifican la propuesta de este estudio.

2.1. Fundamentación teórica

2.1.1. Internet de las Cosas

El Internet de las Cosas constituye una innovación tecnológica fundamental que permite la conexión de dispositivos físicos capaces de capturar, intercambiar y procesar datos en redes digitales. Esta infraestructura ha transformado sectores clave como la salud, la industria y la agricultura, al integrar tecnologías de información y comunicación interoperable (Atzori y otros, 2010) (Madakam y otros, 2015).

Su arquitectura general se organiza en cuatro capas funcionales: percepción, red, procesamiento y aplicación. La Figura 2 ilustra este modelo, en el que la capa de percepción comprende sensores y actuadores que obtienen datos físicos; la capa de red se encarga de su transmisión a través de tecnologías como LoRaWAN, Wi-Fi o 5G; la capa de procesamiento transforma los datos mediante análisis locales o en la nube; y la capa de aplicación permite la interacción con el usuario a través de servicios como monitoreo remoto, automatización y análisis predictivo (Ray, 2018).



Figura 2. Arquitectura en capas jerárquicas ascendentes en un sistema del Internet de las Cosas.
Fuente: Elaboración propia.

En el ámbito agrícola, el Internet de las Cosas ha dado lugar al desarrollo de la Agricultura de Precisión, un enfoque que integra sensores distribuidos, plataformas de análisis y sistemas de actuación automática para optimizar la producción. Entre sus aplicaciones destacan los sensores de humedad del suelo, estaciones meteorológicas conectadas, drones para monitoreo aéreo y sistemas de riego automatizado (Wolfert y otros, 2017). Estos dispositivos permiten una captura continua de datos que, al ser analizados, facilitan decisiones informadas sobre el uso eficiente de agua, fertilizantes y pesticidas, incrementando el rendimiento de los cultivos y reduciendo el impacto ambiental.

Sin embargo, la implementación del Internet de las Cosas en contextos rurales enfrenta obstáculos relevantes, como la conectividad limitada, la heterogeneidad de dispositivos y las preocupaciones en torno a la seguridad y privacidad de los datos (Sfar y otros, 2018). Además, el volumen de información generado requiere soluciones de procesamiento y almacenamiento eficientes, lo que representa un reto particular para pequeños y medianos productores. Para atender estas limitaciones, han surgido alternativas como el *Edge Computing*, que permite realizar el procesamiento de datos directamente en los nodos o puertas de enlace, disminuyendo la latencia, el consumo de

ancho de banda y la dependencia de servicios en la nube (Shi y otros, 2016). A su vez, la estandarización de protocolos y la implementación de mecanismos robustos de autenticación y cifrado resultan esenciales para garantizar la seguridad en estas redes agrícolas inteligentes.

2.1.2. Internet de las Cosas Subterráneas

El Internet de las Cosas Subterráneas representa una extensión especializada del paradigma del Internet de las Cosas, enfocada en la recolección y transmisión de información desde entornos bajo la superficie terrestre. Su característica distintiva es la ubicación total o parcial de sensores y dispositivos bajo tierra (Salam & Raza, 2020), lo que introduce desafíos adicionales relacionados con la transmisión de datos, la alimentación energética y el mantenimiento de los dispositivos.

A diferencia de las tecnologías convencionales diseñadas para operar en la superficie, esta variante enfrenta limitaciones inherentes al medio subterráneo, tales como la atenuación de señales, interferencias electromagnéticas y la heterogeneidad de las propiedades físicas del suelo (Silva & Vuran, 2010). Para mitigar estos problemas, se han desarrollado Redes de Sensores Inalámbricos Subterráneos, las cuales integran nodos autónomos capaces de recolectar y enviar datos de manera eficiente, optimizando el uso de la energía y minimizando la pérdida de información (Akyildiz & Stuntebeck, 2006).

La principal aplicación del Internet de las Cosas Subterráneas se encuentra en la Agricultura de Precisión, donde permite la medición continua de variables críticas como humedad, temperatura, salinidad, pH y contenido volumétrico de agua, además de ofrecer información sobre condiciones climáticas locales (Anisi y otros, 2015). Esta capacidad incrementa la resolución espacial del monitoreo edáfico, optimizando prácticas agrícolas fundamentales como el riego y la fertilización. Desde el punto de vista arquitectónico, estos sistemas incorporan nodos heterogéneos diseñados para operar en entornos hostiles y aislado. Las principales funcionalidades incluyen:

- **Detección in situ.** Sensores enterrados que capturan datos precisos y localizados sobre las propiedades del suelo.

- **Comunicación inalámbrica en entornos adversos.** Transmisión de datos a través del suelo, adaptada a las variaciones dinámicas del entorno físico.
- **Interconexión con maquinaria agrícola y plataformas en la nube.** Capacidad de vinculación con estaciones base, puertas de enlace y plataformas en la nube para habilitar el monitoreo y control remoto en tiempo real.

La Figura 3 ilustra esta arquitectura, mostrando los componentes clave del sistema. Esta configuración permite implementar un flujo de datos bidireccional, que favorece una gestión agrícola más eficiente y adaptativa.

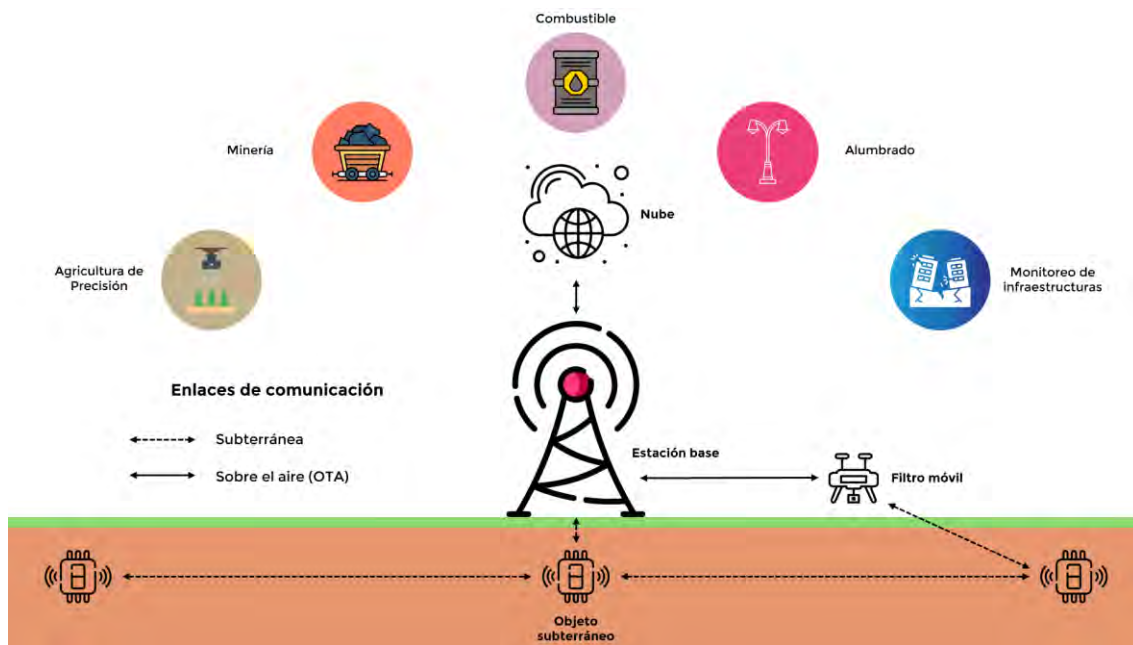


Figura 3. Arquitectura del Internet de las Cosas Subterráneas, se muestran sus componentes esenciales y aplicaciones. *Fuente: Elaboración propia.*

Si bien el Internet de las Cosas Subterráneas aporta beneficios sustanciales para la toma de decisiones basadas en datos, aún enfrenta desafíos técnicos importantes. Entre los más relevantes se encuentran la necesidad de adaptar los sistemas a la heterogeneidad edáfica, la correcta determinación de la topología de red y de la ubicación estratégica de los nodos sensores, así como la reducción del consumo energético para prolongar su autonomía sin requerir mantenimientos frecuentes (Akyildiz & Stuntebeck, 2006) (Anisi y otros, 2015) (Silva & Vuran, 2010).

2.1.3. Limpieza de datos en flujo

La limpieza de datos, también llamada depuración, consiste en identificar, corregir o eliminar registros incorrectos, corruptos, mal formateados, duplicados o incompletos dentro de un conjunto de datos (Erl y otros, 2016). Este proceso es fundamental para asegurar la precisión y consistencia de los datos, condiciones necesarias para generar resultados analíticos confiables. La presencia de datos deficientes puede conducir a interpretaciones erróneas, decisiones inadecuadas o modelos inexactos (Alotaibi y otros, 2023).

Entre los principales beneficios de la limpieza destacan la reducción de errores, la mejora en la confianza de los análisis y la optimización de recursos computacionales al disminuir la carga de procesamiento posterior (Han y otros, 2012). Este proceso forma parte del ciclo de calidad de datos, que incluye además integración, transformación y validación. Según su momento de ejecución, la limpieza puede realizarse en modalidad de procesamiento en lote o en flujo (ver Figura 4). El primero se aplica una vez que los datos han sido recolectados en su totalidad, siendo útil en escenarios sin urgencia temporal. Por su parte, el procesamiento en flujo permite depurar datos en tiempo real, siendo indispensable en contextos como monitoreo con sensores, transacciones en línea o vigilancia de sistemas.

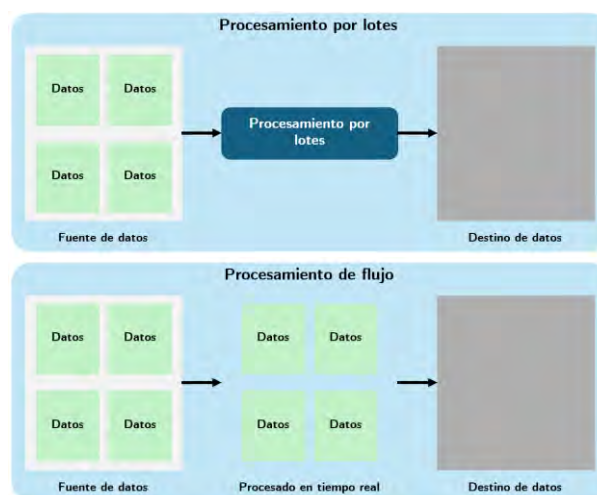


Figura 4. Diferencia entre procesamiento por lotes y en flujo: el primero analiza datos agrupados; el segundo, en tiempo real conforme se generan. *Fuente: Elaboración propia.*

La limpieza de datos en flujo constituye un desafío central en sistemas que requieren procesamiento en tiempo real, como el Internet de las Cosas. A diferencia del procesamiento en lote, en los flujos de datos no es viable almacenar ni revisar exhaustivamente la información, por lo que se requieren algoritmos eficientes, incrementales y adaptables (Abedjan y otros, 2016). Este proceso enfrenta problemas recurrentes como valores faltantes, duplicados, atípicos y registros irrelevantes. Los datos ausentes pueden imputarse mediante promedios, medianas o métodos predictivos (Alotaibi y otros, 2023), mientras que los duplicados se eliminan aplicando técnicas de deduplicación que preservan solamente una instancia del registro. Los valores atípicos pueden ser descartados, corregidos o marcados según el contexto (Han y otros, 2012). Asimismo, los datos no pertinentes se eliminan para no afectar los modelos analíticos.

La limpieza en flujo emplea detección de anomalías mediante estadísticas en línea, reglas heurísticas y algoritmos de aprendizaje automático, lo que permite asegurar la calidad de los datos desde su origen antes del almacenamiento (Abedjan y otros, 2016) (Shi y otros, 2016).

2.2. Estado del arte

2.2.1. Clasificación y criterios de análisis de los trabajos de investigación

Con base en los objetivos de esta tesis, se revisaron trabajos científicos que aplican tecnologías del Internet de las Cosas y del Internet de las Cosas Subterráneas en el contexto de la Agricultura de Precisión. Estos estudios se clasifican en dos categorías: por un lado, sistemas del Internet de las Cosas en la Agricultura de Precisión; por otro, sistemas del Internet de las Cosas Subterráneas en la Agricultura de Precisión.

Para el análisis comparativo, se definieron criterios que permiten valorar cada estudio desde distintos enfoques: el objetivo principal del trabajo, las tecnologías de comunicación empleadas, el hardware y software utilizados, la metodología de validación, los resultados alcanzados, así como las conclusiones, limitaciones y futuras líneas de investigación. Esta estructura facilita la identificación de avances relevantes y vacíos

tecnológicos, especialmente en lo que respecta a la necesidad de integrar algoritmos de limpieza de datos directamente en dispositivos embebidos operando en condiciones subterráneas.

2.2.2. Sistemas del Internet de las Cosas en la Agricultura de Precisión

En el trabajo de investigación de (Meriç, 2025), el autor diseñó un sistema inteligente de riego por goteo que integra una red de sensores inalámbricos, basada en módulos de radiofrecuencia a 868 MHz, microcontroladores STM32 y sensores de humedad, presión y caudal. El autor estructuró el sistema con una arquitectura modular y control remoto de válvulas, y validó su funcionamiento en un campo de cultivo de olivos, donde logró una reducción del 15 % en el consumo de agua y una alta eficiencia operativa. Además, el autor propuso como mejoras futuras la incorporación de tecnologías como LoRa o NB-IoT y el uso de algoritmos inteligentes.

En el trabajo de investigación de (Abidin y otros, 2025), los autores desarrollaron una red de sensores inalámbricos sin conexión a Internet, empleando módulos nRF24L01+ con topología en malla. Los autores implementaron nodos sensores utilizando Arduino Pro Mini y sensores ambientales como DHT11 y sensores de humedad del suelo. En sus pruebas, los autores lograron una comunicación confiable con envíos cada minuto y bajo consumo energético. No obstante, los autores identificaron limitaciones en el alcance de la señal, lo cual motivó la recomendación de adoptar tecnologías con mayor cobertura.

En el trabajo de investigación de (Raza y otros, 2024), los autores propusieron un sistema combinado de censado y comunicación, compuesto por un sensor ultra compacto de humedad del suelo, basado en resonadores espirales complementarios que operan entre 86 y 180 MHz, junto con una antena reconfigurable en la banda de 2.45 GHz. Los autores diseñaron dicha antena con la capacidad de generar seis patrones de radiación distintos, mediante el uso de diodos varactores. Según los resultados obtenidos, los autores reportaron una alta sensibilidad (2.05 %) y una buena capacidad de penetración

en volúmenes amplios de suelo. Sin embargo, los autores reconocieron que aún no validaron el sistema en condiciones reales de campo.

En el trabajo de investigación de (Singh y otros, 2025), los autores presentaron un modelo de agricultura inteligente basado en la integración del Internet de las Cosas y la Computación en Niebla. Los autores utilizaron sensores ambientales, módulos ESP8266 y SIM808, y desarrollaron un enfoque de procesamiento distribuido para reducir la latencia y mejorar la eficiencia en la toma de decisiones agronómicas. En sus pruebas de campo, los autores evidenciaron mejoras en precisión y tiempos de respuesta. No obstante, los autores destacaron la necesidad de fortalecer la escalabilidad y seguridad del sistema para su implementación a gran escala.

En el trabajo de investigación de (Catota y otros, 2024), los autores desarrollaron un invernadero autónomo para cultivos de frijol, con el objetivo de optimizar el uso de agua y fertilizantes mediante automatización y monitoreo remoto. Los autores utilizaron comunicación Wi-Fi con protocolo MQTT a través de Arduino Cloud, y un sistema basado en ESP32, sensores ambientales y motores paso a paso. La validación se llevó a cabo durante tres meses, registrando variables eléctricas y del suelo. Los resultados mostraron un control eficiente de humedad y temperatura, mejorando el desarrollo del cultivo. Los autores concluyen que su prototipo es funcional y sostenible a pequeña escala, aunque limitado en escalabilidad y cobertura de variables. Proponen integrar visión artificial y escalar el sistema en futuras investigaciones.

Finalmente, en el trabajo de investigación de (Melesse, 2025), el autor realizó una revisión sistemática sobre las aplicaciones de gemelos digitales en el monitoreo de cultivos. A partir de la literatura analizada, el autor identificó cinco enfoques de modelado predominantes: físico, híbrido, basado en datos, espacial y basado en agentes. El autor señaló que estos enfoques permiten predicción de rendimiento, monitoreo continuo del desarrollo vegetal y generación de recomendaciones personalizadas. Sin embargo, el autor advirtió que esta tecnología aún se encuentra en una etapa temprana y requiere mayor validación empírica y estandarización antes de su adopción generalizada.

En la Tabla 1 se muestra un resumen de otros trabajos analizados, cuyos autores desarrollaron propuestas correspondientes a esta primera clasificación.

2.2.3. Sistemas del Internet de las Cosas Subterráneas en la Agricultura de Precisión

En el trabajo de investigación de (Herring y otros, 2022), los autores diseñaron un nodo sensor subterráneo basado en tecnología LoRaWAN, orientado al monitoreo de incendios forestales. Para validar su funcionalidad, los autores evaluaron la resistencia térmica del dispositivo bajo condiciones de fuego controlado, utilizando un termopar tipo K conectado a una radio enterrada a 10 cm de profundidad. A través de mediciones experimentales, los autores demostraron que el nodo puede sobrevivir a incendios breves, dependiendo de variables como la profundidad de enterramiento y el tipo de suelo. Como línea futura de trabajo, los autores propusieron integrar modelos que consideren la humedad del suelo para mejorar la caracterización térmica.

En el trabajo de investigación de (Boebel y otros, 2023), los autores desarrollaron un sistema autónomo sin baterías para el monitoreo de tuberías de agua urbana, utilizando generadores termoelectricos que aprovechan el gradiente térmico entre el suelo y las tuberías. Los autores validaron el sistema durante un periodo de un año en cuatro ubicaciones diferentes, obteniendo una tasa de operación del 94 % sin intervención externa y con comunicación mediante LoRaWAN. No obstante, los autores identificaron como principal limitación la dependencia de las condiciones térmicas estacionales. Como mejora futura, los autores sugirieron incorporar más sensores y optimizar la eficiencia energética del sistema.

En el trabajo de investigación de (Cariou y otros, 2023), los autores propusieron una solución para la recolección de datos provenientes de sensores enterrados mediante el uso de vehículos aéreos no tripulados, con el objetivo de superar las limitaciones de comunicación típicas de las redes subterráneas. Para ello, los autores integraron sensores equipados con radio LoRa y drones actuando como recolectores móviles. En sus pruebas experimentales, los autores confirmaron la viabilidad del sistema tanto en vuelo

estacionario como en movimiento. No obstante, los autores destacaron la necesidad de optimizar la trayectoria y velocidad del UAV para aplicaciones a mayor escala.

En el trabajo de investigación de (Zhang y otros, 2017), los autores presentaron el sistema Thoreau, una red subterránea conformada por sensores alimentados con baterías de larga duración y conectados mediante la red Sigfox. Los autores desplegaron este sistema en un campus universitario para el monitoreo continuo de humedad, temperatura y conductividad eléctrica del suelo, logrando una operación estable durante varios meses. Sin embargo, los autores señalaron que la dependencia de la cobertura de Sigfox representa una limitación para su adopción en otras regiones. En consecuencia, los autores plantearon como trabajo futuro la expansión de la red y la evaluación de tecnologías alternativas de transmisión.

En el trabajo de investigación de (Said y otros, 2024) los autores desarrollaron un prototipo autónomo para estimar la permitividad dieléctrica del suelo vegetal, clave en sistemas subterráneos aplicados a la agricultura de precisión. Utilizaron sensores enterrados comunicados por LoRa a 433 MHz y una sonda capacitiva basada en reflectometría en frecuencia. El sistema fue validado mediante simulaciones y pruebas empíricas, mostrando estimaciones precisas de permitividad entre 6 y 15. Los autores concluyen que el diseño es viable para monitoreo subterráneo de bajo costo, aunque con limitaciones para valores bajos de permitividad. Proponen como mejoras futuras optimizar la sensibilidad y validar el prototipo en campo real.

Finalmente, en el trabajo de investigación de (Levintal y otros, 2022), los autores diseñaron un sistema inalámbrico, abierto y de bajo costo para el monitoreo de humedad, oxígeno y temperatura del suelo, utilizando tecnología LoRa y microcontroladores de bajo consumo energético. Los autores validaron el sistema durante cinco meses en un campo experimental, obteniendo resultados favorables en términos de autonomía y confiabilidad. Asimismo, los autores destacaron la modularidad de su diseño como una ventaja para adaptaciones futuras. No obstante, los autores advirtieron que el desempeño del sistema puede variar en función de las características específicas del suelo. Como línea de mejora,

los autores plantearon escalar el sistema y optimizar la eficiencia en la comunicación de datos.

En la Tabla 2 se presenta un resumen de otros trabajos analizados, cuyos autores desarrollaron propuestas correspondientes a esta segunda clasificación.

Tabla 1. Síntesis de trabajos de investigación enfocados en el desarrollo de sistemas del Internet de las Cosas para la Agricultura de Precisión

Título	Objetivo principal	Tecnologías de comunicación	Hardware y software	Metodología de investigación	Resultados	Conclusiones y trabajos futuros
Hybrid Sensing Platform for IoT-Based Precision Agriculture - (Bagha y otros, 2022)	Proponer una plataforma híbrida que combine sensores aéreos y terrestres para mejorar la precisión de datos en agricultura de precisión.	Red del Internet de las Cosas híbrida (UAV + sensores de suelo), comunicación inalámbrica con validación cruzada.	Sensores terrestres, UAVs con sensores infrarrojos y espectrómetros; uso de modelos estadísticos y de aprendizaje automático.	Experimentos en campo; comparación de datos aéreos calibrados con medidas de referencia del suelo.	Mejora en precisión de datos: 34% en humedad, 51% en temperatura, 66% en infrarrojo y 90% en espectro.	Alta precisión con bajo costo operativo; se requiere optimización del despliegue y algoritmos de fusión de datos más robustos.
Integrating IoT for Soil Monitoring and Hybrid ML in Predicting Tomato Crop Disease in a Typical South India Station - (Babu y otros, 2024)	Desarrollar un sistema del Internet de las cosas con Aprendizaje Maquina híbrido para predecir enfermedades en tomates mediante análisis de parámetros del suelo.	Red del Internet de las Cosas con sensores de suelo; transmisión de datos en tiempo real.	Sensores de humedad, temperatura, pH, NPK; algoritmos KNN, SVM, RF, LR, optimización bayesiana.	Pruebas en campo durante etapa vegetativa; análisis estadístico y evaluación de métricas de Aprendizaje Maquina.	El algoritmo híbrido alcanzó 95% de precisión, recall y F1; los modelos individuales entre 77-82% de precisión.	Sistema eficaz para detección temprana de enfermedades; se sugiere adaptar modelos a condiciones locales y ampliar base de datos.
Applications of AI in Precision Agriculture -	Revisar aplicaciones actuales y futuras	Sistemas del Internet de las Cosas, imágenes	Sensores, drones, cámaras, plataformas de	Revisión de casos reales y estudios	La Inteligencia Artificial mejora eficiencia de	Se requieren políticas para privacidad de

(Gupta & Pal, 2025)	de la inteligencia artificial en agricultura de precisión.	satelitales, sensores, redes inalámbricas.	visión por computadora; uso de IA, ML, CV, NLP.	implementados en campo.	riego, predicción de enfermedades, manejo de maleza y control de plagas.	datos y reducción de costos; tendencia a automatización total del campo.
Smart Sensors and Smart Data for Precision Agriculture: A Review - (Soussi y otros, 2024)	Revisar el papel de sensores inteligentes y datos en Agricultura de Precisión.	WSN, Internet de las Cosas, protocolos ZigBee, LoRa, APTEEN, Fuzzy-based Energy Protocol.	Plataformas Arduino, MDA300, EZ430, sensores múltiples, algoritmos de ML (PCR, PLSR).	Revisión de literatura científica (2018-2023) con enfoque en sistemas de monitoreo.	Uso extensivo de sensores mejora toma de decisiones y reduce insumos agrícolas.	Integrar Cloud, Edge Computing y estandarizar plataformas; falta mayor despliegue en campo real.
IoT-Driven Smart Agricultural Technology for Real-Time Soil and Crop Optimization - (Shahab y otros, 2025)	Proponer un sistema del Internet de las Cosas con sensores y aplicación móvil para optimización de cultivos en tiempo real.	Internet de las Cosas, transmisión en tiempo real, nube, aplicación móvil conectada.	Sensores de temperatura, humedad, salinidad, EC, pH, NPK; aplicación con Inteligencia Artificial para recomendaciones.	Pruebas de campo con cultivo de arroz en Pakistán; análisis comparativo de parámetros.	Mediciones precisas de 8 parámetros; la aplicación mejora fertilización e irrigación con datos en tiempo real.	Solución escalable con beneficios en sostenibilidad; se recomienda extenderla a otros cultivos y regiones.

Tabla 2. Síntesis de trabajos de investigación enfocados en el desarrollo de sistemas del Internet de las Cosas Subterráneas para la Agricultura de Precisión

Título	Objetivo principal	Tecnologías de comunicación	Hardware y software	Metodología de investigación	Resultados	Conclusiones y trabajos futuros
LWS: A LoRaWAN Wireless Underground Sensor Network Simulator for Agriculture Applications - (Wang y otros, 2019)	Desarrollar un simulador basado en Eventos Discretos (LWS) para modelar redes WUSN con LoRaWAN.	LoRaWAN en entorno subterráneo, con modelo de pérdida por trayectoria en el suelo.	Simulador basado en LoRaSim; incorpora modelos dieléctricos y parámetros de red LoRa.	Simulaciones con distintas profundidades, densidades de nodos y condiciones de suelo.	El simulador permite evaluar cobertura, escalabilidad y rendimiento antes del despliegue real.	Herramienta útil para planificación de redes WUSN; se propone ampliar con otros protocolos y escenarios reales.
WUSN Path Loss Model for Precision Agriculture (WUSN-PLM) - (Sambo y otros, 2020)	Proponer un nuevo Modelo de Pérdida de Trayectoria (WUSN-PLM) para comunicaciones subterráneas.	Ondas electromagnéticas a 433 MHz para UG2UG, UG2AG y AG2UG.	Sensores enterrados con transceptores de baja frecuencia; análisis por simulación y prueba real.	Pruebas intensivas con dos pares de transceptores en jardín botánico; comparación con modelos existentes.	El modelo propuesto alcanzó 87.13% de precisión y 85% de exactitud balanceada.	Modelo robusto para entornos agrícolas reales; futuras versiones integrarán más variables del suelo y sensores.
A WUSN Field Pilot for Agriculture and Ecology: Soil Moisture Mapping Using Signal Attenuation -	Evaluar el uso de WUSN para monitoreo continuo de humedad del suelo y mapearla mediante aprendizaje	Red Sigfox 901.2 MHz con nodos enterrados conectados a sensores TEROS12.	23 nodos WUSN con sensores de humedad, temperatura y conductividad eléctrica; algoritmos de	Despliegue de 9 meses en un campo agrícola; recolección continua y modelado con ANN-MLP, SVR y ELM.	R ² de 0.82 y RMSE de ± 0.012 usando RSSI combinado con variables climáticas.	Demuestra viabilidad para mapear humedad sin sensores costosos; se propone escalar a otras regiones y estaciones.

(Balivada y otros, 2022)	automático a partir de RSSI.			regresión y redes neuronales.				
LoRa Multi-Hop Networks for Monitoring Underground Mining Environments - (Scalambrin y otros, 2023)	Desarrollar y evaluar una red multi-hop basada en LoRa para entornos mineros subterráneos con enfoque en eficiencia energética y sincronización.	LoRa multi-hop en topología tipo árbol; sincronización TDMA con sub-40mpps de precisión.		Nodos con LoRa y protocolo MAC personalizado; modelo teórico de duty cycle y pruebas reales.	Simulaciones y pruebas en dispositivos reales; análisis de rendimiento según topología y ubicación.	Buena sincronización y bajo consumo para nodos remotos; útil para escenarios sin conectividad directa.	Aplicable a minería y agricultura bajo tierra; se propone ampliar a despliegues dinámicos y escenarios heterogéneos.	
Battery-Free Sensor Array for Wireless Multi-Depth In-Situ Sensing - (Guo & Kamrath, 2024)	Diseñar una matriz de sensores inalámbricos sin batería para monitoreo subterráneo a múltiples profundidades usando NFC.	Comunicación por inducción magnética (NFC HF, 13.56 MHz) con guía de onda virtual.		Sensores NFC ISO/IEC 15693; modelo analítico y prototipo experimental con lector móvil.	Modelado teórico y pruebas empíricas en laboratorio con sensor array.	Ampliación de rango de comunicación mediante diseño optimizado; sin necesidad de baterías.	Prometedor para entornos subterráneos y submarinos; se propone escalar red y validar en campo.	

La revisión del estado del arte muestra que, si bien existen importantes avances en el desarrollo de sistemas del Internet de las Cosas y del Internet de las Cosas Subterráneas aplicados a la Agricultura de Precisión, persisten vacíos críticos en relación con el tratamiento autónomo de la calidad de los datos. La mayoría de los estudios analizados se han centrado en aspectos como la conectividad, el consumo energético, la comunicación en medios hostiles o la incorporación de sensores especializados. Sin embargo, se ha identificado una ausencia generalizada de mecanismos que permitan garantizar la calidad de los datos directamente desde el nodo, particularmente en entornos subterráneos donde el acceso físico es limitado.

Las investigaciones revisadas no integran algoritmos dedicados a la detección de datos atípicos, imputación de valores faltantes o eliminación de duplicados dentro del propio dispositivo embebido. Además, la validación de muchos prototipos carece de experimentación, y los procedimientos de evaluación suelen limitarse a pruebas funcionales básicas, sin un análisis cuantitativo del impacto sobre la integridad de los datos.

En este contexto, esta tesis contribuye con una propuesta novedosa que integra un algoritmo de limpieza de datos en tiempo real dentro de un nodo subterráneo autónomo, capaz de operar de manera continua y sin supervisión en condiciones reales de campo. Esta propuesta incluye un algoritmo ligero y eficaz para detección de atípicos, imputación y deduplicación, así como mecanismos de autodiagnóstico y persistencia de registros. A través de una validación experimental, en un entorno subterráneo real, se demuestra su efectividad frente a distintas proporciones de anomalías.

Así, esta investigación aporta una respuesta integral a un vacío identificado en el estado del arte, al ofrecer un enfoque funcional, validado y reproducible para el tratamiento autónomo de datos en sistemas del Internet de las Cosas Subterráneas, fortaleciendo la confiabilidad de las aplicaciones agrícolas basadas en sensores enterrados.

Construcción del nodo sensor subterráneo

Este capítulo detalla las actividades de diseño, integración, ensamblaje y validación del nodo sensor subterráneo, componente físico esencial para la ejecución del algoritmo de limpieza de datos en flujo. El nodo fue concebido para operar de forma autónoma en entornos subterráneos, asegurando una recolección continua y confiable de variables ambientales.

3.1. Selección técnica de componentes de hardware

La primera actividad consistió en una revisión del estado del arte para identificar sensores y plataformas adecuadas al Internet de las Cosas Subterráneas en Agricultura de Precisión. Se consideraron las condiciones del entorno subterráneo y las restricciones energéticas y computacionales de los sistemas embebidos. El análisis incluyó trabajos de Salam y Raza (2020), Silva y Vuran (2010), Zhang y otros. (2024), entre otros, enfocados en arquitecturas subterráneas y sensores eficientes. A partir de esta revisión, se seleccionaron plataformas como M0 LoRa, Arduino MKR WAN 1310, Heltec WiFi LoRa 32 V2 y TTGO LoRa32, junto con sensores como el Decagon 5TE, FS200-SHT10, SHT20, SEN0193 y SEN0161. La elección final se realizó mediante una evaluación multicriterio que ponderó precisión, eficiencia, robustez e integración, siguiendo la metodología propuesta por Salam y Raza (ver Figura 5).

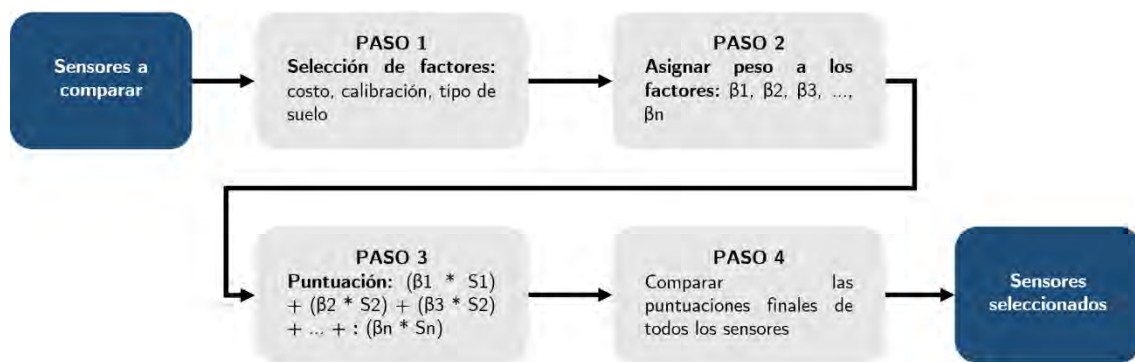


Figura 5. Marco de trabajo propuesto por Salam & Raza (2020) para la selección técnica de sensores y plataformas en entornos subterráneos. *Fuente: Elaboración propia.*

Con base en los requerimientos del sistema y la revisión técnica, se definieron seis criterios para evaluar sensores y plataformas en entornos subterráneos: precisión (25 %), consumo energético (20 %), resistencia estructural (20 %), rango de operación (15 %), tiempo de respuesta (10 %) y facilidad de instalación (10 %). Estos factores, ponderados según su relevancia en condiciones edáficas adversas, guiaron la comparación de los sensores Decagon 5TE, FS200-SHT10, SHT20 y SEN0193. Las especificaciones, obtenidas de fuentes oficiales y literatura científica, se resumen en la Tabla 3 para calcular su puntuación global.

Tabla 3. Especificaciones técnicas comparativas de sensores de temperatura y humedad del suelo para el nodo sensor del Internet de las Cosas Subterráneas

Sensor	Precisión	Consumo	Resistencia	Rango	Respuesta	Instalación
Decagon 5 TE	±1% VWC / ±1°C	Bajo (3.6–15 V)	Alta	-40 a 60°C / 0–100% RH	Moderado	Moderada
FS200-SHT10	±3% RH / ±0.5°C	1 mW activo / 30 µW reposo	Muy alta	-20 a 100°C / 0–100% RH	Rápido	Media
SHT20	±4% RH / ±1°C	Bajo	Media	-40 a 125°C / 0–100% RH	Moderado	Alta

SEN0193	±5% RH / ±1°C	Medio (alimentación de 3.3 a 5 V)	Baja	0 a 80°C / 0–100% RH	Lento	Alta
----------------	------------------	---	------	----------------------------	-------	------

Para seleccionar el sensor de temperatura y humedad más adecuado, se aplicó el modelo de evaluación multicriterio con las ponderaciones establecidas. Cada dispositivo fue calificado del 1 al 10 según su desempeño documentado, con el fin de identificar el que ofreciera el mejor equilibrio entre precisión, eficiencia, robustez e integración en entornos subterráneos. La Tabla 4 muestra los resultados de esta comparación.

Tabla 4. Resultados de la evaluación multicriterio de sensores de temperatura y humedad del suelo (ponderaciones y puntuaciones finales)

Sensor	Puntuación total
FS200 – SHT10	8.72
Decagon 5TE	8.25
SHT20	6.90
SEN0193	6.10

El sensor FS200–SHT10 obtuvo la mayor puntuación en la evaluación multicriterio, destacando por su robustez, amplio rango de operación y alta precisión, lo que lo posiciona como la opción más adecuada para entornos subterráneos en agricultura de precisión. Aunque el Decagon 5TE mostró buen desempeño, fue superado en resistencia estructural por el sensor FS200-SHT10. Para la medición de pH se eligió el SEN0161, reconocido por su resistencia y compatibilidad con sistemas embebidos. Respecto a la unidad de procesamiento, se evaluaron las placas Arduino MKR WAN 1310, M0 LoRa, Heltec WiFi LoRa 32 V2 y TTGO LoRa32, considerando criterios como consumo energético, autonomía, integración, almacenamiento local y comunicación inalámbrica. La Tabla 5 resume las especificaciones comparadas.

Tabla 5. Especificaciones técnicas comparativas de las placas de desarrollo para el nodo sensor del Internet de las Cosas Subterráneas

Placa	Precisión IO	Consumo	Resistencia HW	Almacenamiento	Comunicación	Integración
MKR WAN 1310	Alta	Muy bajo	Alta	Flash interno	LoRa	Alta
M0 LoRa	Alta	Bajo	Alta	Flash interno	LoRa	Alta
Heltec WiFi LoRa V2	Media	Medio	Media	Flash interno	WiFi + LoRa	Media
TTGO LoRa32	Media	Medio / alto	Media	Flash interno	WiFi + LoRa	Alta

Con base en las especificaciones técnicas, se aplicó un modelo de evaluación multicriterio a las placas de desarrollo, permitiendo identificar objetivamente su idoneidad para operar de forma autónoma en entornos subterráneos. Los resultados se presentan en la Tabla 6.

Tabla 6. Resultados de la evaluación multicriterio de las placas de desarrollo (ponderaciones y puntuaciones finales)

Sensor	Puntuación total
MKR WAN 1310	8.80
M0 LoRa	8.20
Heltec WiFi LoRa V2	6.90
TTGO LoRa32	7.10

Como resultado, se seleccionó la placa Arduino MKR WAN 1310 por sus cualidades clave para el entorno subterráneo: bajo consumo, almacenamiento integrado y conectividad LoRa. Esto permitió definir la configuración óptima del nodo sensor, integrando componentes compatibles y aptos para operar de forma autónoma en entornos subterráneos. La Figura 6 presenta los elementos seleccionados.



Figura 6. Sensores y placa de desarrollo seleccionados para el nodo sensor del Internet de las Cosas Subterráneas tras la evaluación multicriterio. *Fuente: Elaboración propia.*

3.2. Diseño esquemático del nodo sensor subterráneo

Una vez definida la configuración de hardware, se diseñó el esquema estructural del nodo sensor subterráneo, integrando los componentes seleccionados y representando sus conexiones para asegurar su correcto funcionamiento. El nodo incorpora los sensores FS200-SHT10 y SEN0161 para medir temperatura, humedad y pH, junto con la placa Arduino MKR WAN 1310, un reloj DS3231SN para sincronización precisa y una batería LiPo de 3.7 V para operación autónoma. La Figura 7 muestra los elementos del nodo diseñado para ejecutar el algoritmo de limpieza de datos en flujo.



Figura 7. Componentes seleccionados para la construcción del nodo sensor subterráneo, considerando criterios técnicos. *Fuente: Elaboración propia.*

El diseño del circuito se realizó en Fritzing, generando un esquema detallado de las conexiones del nodo. Para el montaje experimental se empleó un protoboard de 400 puntos. La Tabla 7 resume las conexiones entre módulos, y la Figura 8 presenta el esquema electrónico resultante.

Tabla 7. Detalle de las conexiones eléctricas entre los componentes del nodo sensor subterráneo asegurando una comunicación estable y un funcionamiento eficiente en condiciones subterráneas

Componente	Pin / Terminal	Conexión
Sensor FS200 – SHT10	VDD	Línea de 5V en el protoboard
	GND	Línea GND en el protoboard
	DATA	Pin D3 del Arduino MKR WAN 1310
	SCK	Pin D4 del Arduino MKR WAN 1310
Módulo RTC DS3231SN	VCC	Línea de 5V en el protoboard
	GND	Línea GND en el protoboard
	SDA	Pin SDA (I ² C) del Arduino
	SCL	Pin SCL (I ² C) del Arduino
Sensor SEN0161	VCC	Línea de 5V en el protoboard
	GND	Línea GND en el protoboard
	Analógico	Pin A0 del Arduino MKR WAN 1310
Batería LiPo 3.7V 3000mAh	JST PH	Puerto de alimentación del Arduino MKR WAN 1310

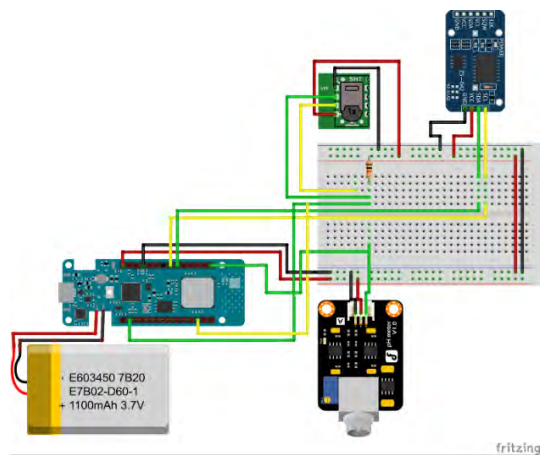


Figura 8. Esquemático de conexiones del nodo sensor subterráneo en protoboard utilizando el software Fritzing. *Fuente: Elaboración propia.*

El nodo sensor subterráneo fue diseñado para operar de forma cíclica, ejecutando cuatro etapas por intervalo: adquisición, procesamiento local, almacenamiento y transmisión. Los sensores FS200-SHT10 y SEN0161 registran temperatura, humedad y pH, sincronizados por el RTC DS3231SN. El microcontrolador ejecuta un algoritmo embebido que detecta atípicos, imputa valores faltantes y elimina duplicados, garantizando operación autónoma sin infraestructura externa. Los datos validados se almacenan en archivos .CSV en una tarjeta microSD. Aunque la transmisión LoRa aún no se implementa, se contempla como mejora futura para integrar una red de monitoreo agrícola subterráneo.

3.3. Ensamble del nodo sensor subterráneo

Antes del ensamblaje final, se calibró el sensor de pH SEN0161 para hacerlo compatible con el ADC del Arduino MKR WAN 1310, ya que por defecto entrega 0 V para pH 7, generando valores negativos en medios ácidos. Se ajustó el potenciómetro para que una señal neutra produjera 2.5 V, representando así el rango de pH 0–14 dentro de 0–5 V. La calibración se realizó simulando una señal neutra mediante un cortocircuito en el conector BNC y regulando el potenciómetro mientras se monitoreaba la salida desde el Arduino. Las Figuras 9 el procedimiento seguido para la calibración.



(a) Calibración mediante corto circuito inducido

```
calibrarsensorino
1 void setup() {
2   // initialize serial communication at 9600 bits per second:
3   Serial.begin(9600);
4 }
5 // the loop routine runs over and over showing the voltage on A0
6 void loop() {
7   // read the input on analog pin 0:
8   int sensorValue = analogRead(A0);
9   // Convert the analog reading (which goes from 0 - 1023) to a voltage (0 - 5V):
10  float voltage = sensorValue * (5.0 / 1023.0);
11  // print out the value you read:
12  Serial.println(voltage);
13  delay(1000);
14 }
```

(b) Sketch para la captura de los valores de voltaje

Serial Monitor × Output

Message (Enter to send message to 'Arduino MKR WAN 1310' on 'COM4')

```
2.48
2.48
2.48
2.49
2.50
2.50
2.50
2.50
2.50
2.50
2.50
```

(c) Valores de voltaje posterior a la calibración

Figura 9. Proceso de calibración del sensor. (a) Corto circuito inducido, (b) código para captura de lecturas, y (c) valores registrados tras la calibración. *Fuente: Elaboración propia.*

Una vez calibrado el sensor de pH, se ensambló el nodo siguiendo el diseño esquemático definido. Los componentes se interconectaron en un protoboard de 400 puntos, asegurando compatibilidad eléctrica entre sensores, módulo RTC, microcontrolador y fuente de energía. Para proteger el sistema en el entorno subterráneo, el circuito se montó en una caja de policarbonato con certificación IP65, garantizando aislamiento frente a humedad, suciedad y agentes corrosivos durante las pruebas de campo (ver Figura 10).

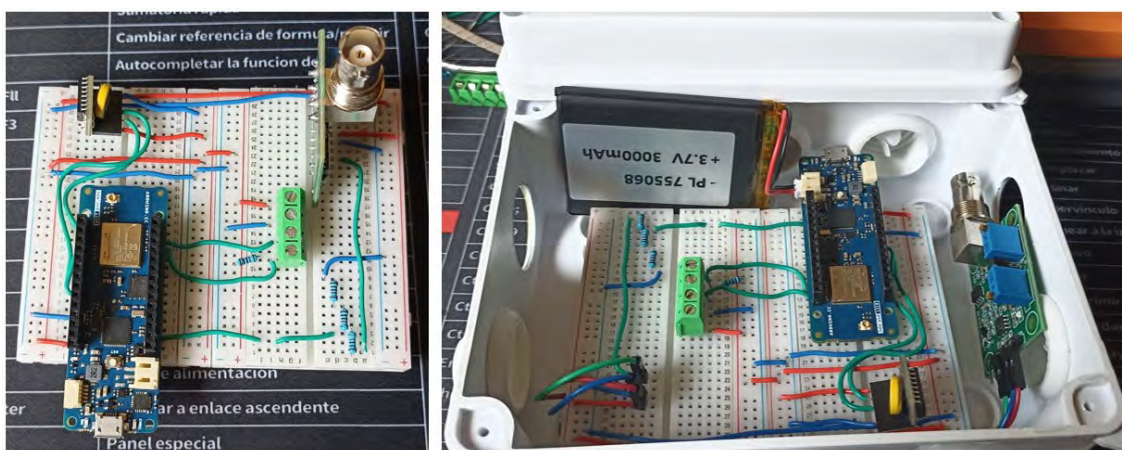


Figura 10. Ensamblaje físico del nodo sensor subterráneo. De lado izquierdo se muestra la disposición de los componentes y de lado izquierdo la integración de los componentes en una caja con protección IP65. *Fuente: Elaboración propia.*

El sistema ensamblado fue sometido a pruebas funcionales individuales para verificar el correcto desempeño de cada componente. El sensor FS200-SHT10 fue evaluado mediante un sketch que mostraba lecturas en el monitor serie, comprobando su respuesta ante variaciones térmicas al ser sumergido en agua a temperatura ambiente y luego en agua fría (ver Figura 11). El sensor SEN0161 fue probado en soluciones buffer estándar (pH 4.0, 6.86 y 9.18), mostrando lecturas estables y coherentes, lo que validó su calibración (ver Figura 12). Finalmente, el módulo RTC DS3231SN fue validado al confirmar la precisión y estabilidad de las lecturas de fecha y hora en tiempo real.



Figura 11. (a) Lectura de temperatura y humedad en agua a temperatura ambiente. (b) Lectura con agua fría para observar variación térmica y de humedad. *Fuente: Elaboración propia.*

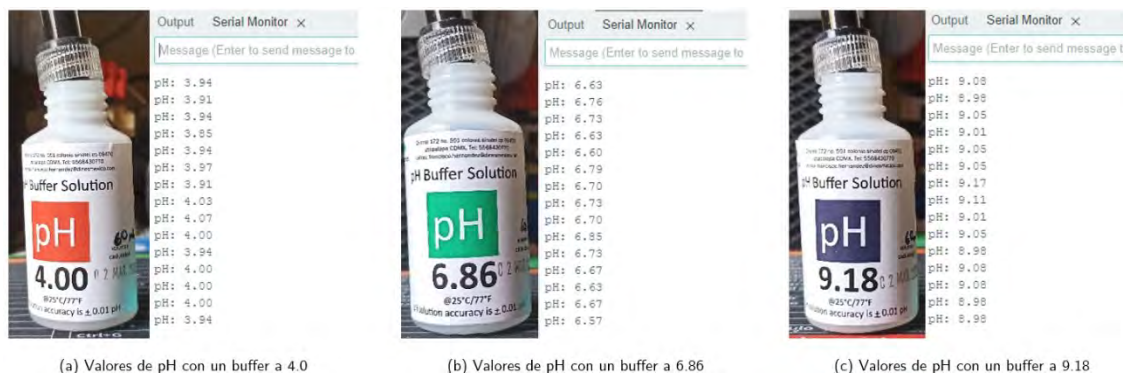


Figura 12. (a) Lecturas con buffer pH 4.00, (b) con buffer pH 6.86 y (c) con buffer pH 9.18. *Fuente: Elaboración propia.*

Finalmente, se evaluó la estabilidad operativa del nodo mediante una prueba continua de seis horas, enterrándolo a una profundidad controlada en condiciones simuladas de campo. Durante este periodo, los sensores registraron datos periódicos con marca de tiempo, almacenados en archivos .CSV. Se monitoreó la estabilidad del sistema

y la integridad de los datos. Los resultados, presentados en la Figura 13, confirman que el nodo operó de forma autónoma, continua y sin errores de captura.



Figura 13. Registro de marca de tiempo y valores de sensores (pH, temperatura y humedad) capturados en una lectura conjunta. *Fuente: Elaboración propia.*

4

Desarrollo del algoritmo de limpieza de datos en flujo

Este capítulo expone el desarrollo del algoritmo de limpieza de datos implementado para mejorar la calidad de la información recolectada por el nodo sensor. Se describen las técnicas analizadas, los criterios para definir los requerimientos del sistema y las decisiones de diseño adoptadas para asegurar su ejecución eficiente en tiempo real. También se detalla su codificación e integración con el prototipo, garantizando compatibilidad operativa y capacidad para gestionar datos en entornos subterráneos.

4.1. Definición de requerimientos del algoritmo

La calidad de los datos recolectados en sistemas del Internet de las Cosas Subterráneas es crucial para un análisis confiable, pero su operación en entornos hostiles, con conectividad limitada y restricciones de hardware, incrementa la presencia de anomalías como valores atípicos, datos faltantes y duplicados.

Esta actividad tuvo como objetivo definir los requerimientos operativos y estadísticos del algoritmo de limpieza a integrar en el nodo sensor subterráneo, mediante una revisión de literatura especializada y la evaluación de técnicas adaptadas a condiciones de campo, priorizando simplicidad, eficiencia y confiabilidad. La limpieza de datos en flujo ha sido ampliamente estudiada, especialmente en sistemas de adquisición continua como los del Internet de las Cosas. La Tabla 8 resume las técnicas encontradas en la literatura para la limpieza de datos según el tipo de problema, con una breve descripción de su funcionamiento.

Tabla 8. Técnicas identificadas en la literatura científica para limpieza de datos en flujo en sistemas del Internet de las Cosas

Proceso	Técnica	Descripción
Detección de atípicos	K – Sigma	Detecta valores fuera de $\pm k$ desviaciones estándar respecto a la media.
	Welford	Calcula media y varianza en línea, sin almacenar todo el histórico de datos.
	Rango Intercuartílico	Detecta valores fuera del rango intercuartílico; adecuado para datos no normales.
	Promedio móvil con ventana deslizante	Suaviza datos y detecta anomalías según variaciones en una ventana temporal.
	Concept Drift Detection	Detecta cambios estadísticos en el comportamiento de la variable; costoso computacionalmente.
	Reservoir Sampling	Mantiene una muestra aleatoria de los datos; útil para análisis posteriores, no inmediato.
Imputación de datos	Última Observación Llevada Adelante (LOCF)	Sustituye datos faltantes con el último valor válido registrado.
	Promedio móvil	Imputa usando el promedio de lecturas recientes; requiere ventana histórica.
	Interpolación lineal	Estima valores intermedios a partir de lecturas adyacentes.
	Media global	Rellena con el valor promedio de toda la serie histórica.

			Regresión lineal	Modela y predice valores a partir de la relación entre variables.
			K-Nearest Neighbors (KNN)	Imputa con base en patrones de instancias similares; alto costo computacional.
Detección de duplicados	Comparación de marcas de tiempo			Compara marcas de tiempo para detectar registros repetidos; determinista y eficiente.
	Filtros de Bloom			Estructura probabilística que indica presencia de elementos; puede generar falsos positivos.
	Hashing simple			Genera claves hash para comparar registros; requiere tablas adicionales.
	UUIDs (Identificadores únicos)			Asigna un identificador único a cada registro; útil en sistemas distribuidos.

Para optimizar el desempeño en entornos subterráneos, se definieron cuatro criterios: simplicidad, bajo uso de memoria, operación en línea y precisión determinista. Con base en ellos, se seleccionaron tres técnicas compatibles con sistemas embebidos de baja potencia. La Tabla 9 muestra su correspondencia con dichos criterios.

Tabla 9. Correspondencia entre criterios de selección y técnicas adoptadas para el desarrollo del algoritmo de limpieza de datos en flujo

Criterio	Justificación técnica	Técnica seleccionada
Simplicidad algorítmica	Uso de operaciones matemáticas básicas (sumas, medias, comparaciones)	Welford + K-Sigma, LOCF, marcas de tiempo
Bajo uso de memoria	No requiere estructuras auxiliares ni almacenamiento histórico	Welford, LOCF, marcas de tiempo
Operación en línea	Cálculo y evaluación de cada dato en tiempo real	Welford, K-Sigma, LOCF
Determinismo y precisión	Clasificación clara, sin componentes probabilísticos ni aproximaciones	K-Sigma, LOCF, marcas de tiempo

La combinación de Welford y K-Sigma permite detectar anomalías en tiempo real de forma eficiente y con bajo uso de memoria. El algoritmo de Welford actualiza media y varianza sin almacenar todas las observaciones, lo cual es ideal para dispositivos con recursos limitados.

Las estadísticas se actualizan solo con datos válidos, y K-Sigma evalúa si la nueva lectura x_n está dentro del rango $\mu \pm 3\sigma$. Si cae fuera, se considera atípica y no se usa para actualizar los parámetros. Esta combinación ofrece un equilibrio entre sensibilidad, eficiencia y estabilidad. La Figura 14 ilustra cómo se identifica una lectura anómala (L4) fuera del rango, mientras que una lectura válida (L5) contribuye al ajuste estadístico.

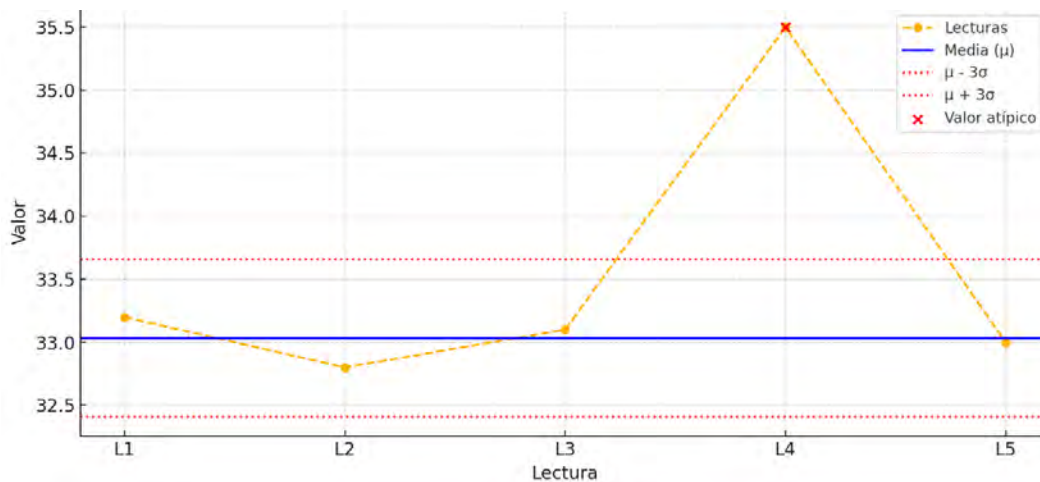


Figura 14. Ejemplo de detección de valor atípico mediante Welford + K-Sigma en flujo de dato, esta combinación permite la detección de anomalías en tiempo real. *Fuente: Elaboración propia.*

La técnica de la Última Observación Llevada Adelante imputa valores faltantes sustituyéndolos por el último valor válido registrado, lo que permite mantener la continuidad de la serie sin cálculos complejos ni uso intensivo de memoria. Esta técnica es especialmente eficaz en variables de variación lenta, como temperatura y humedad del suelo. En el algoritmo desarrollado, los valores imputados mediante la Última Observación Llevada Adelante no se utilizan para actualizar las estadísticas de Welford, evitando sesgos. Gracias a su complejidad constante $O(1)$, es ideal para microcontroladores. La Figura 15 muestra cómo esta técnica reemplaza valores faltantes, manteniendo la secuencia temporal con mínima carga computacional.

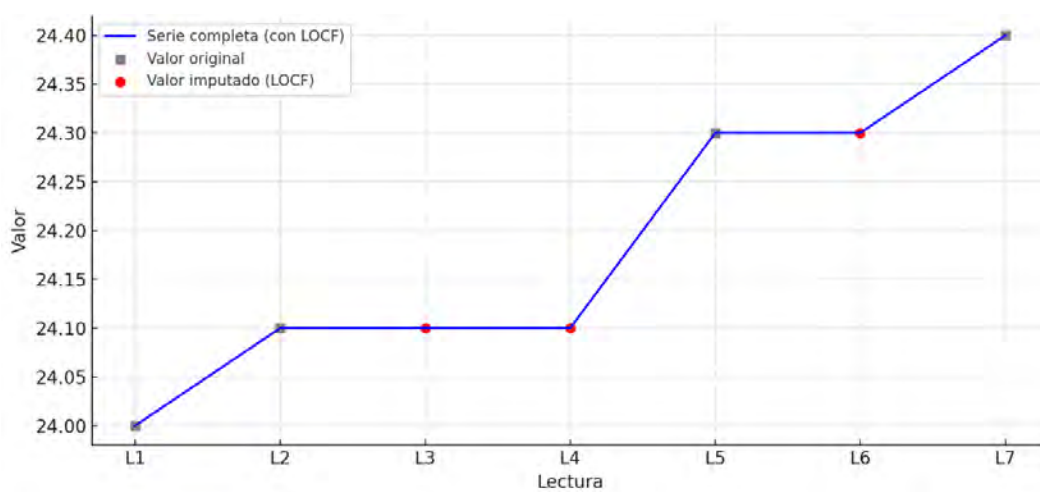


Figura 15. Ejemplo de imputación de datos faltantes mediante la técnica de la Última Observación Realizada. *Fuente: Elaboración propia.*

La detección de duplicados se basa en comparar la marca de tiempo de cada nueva lectura con la última registrada. Si coinciden, el dato se descarta por considerarse retransmitido. Este método, eficiente y determinista, no requiere estructuras adicionales y es especialmente efectivo con un RTC como el DS3231, que garantiza precisión temporal.

Definidas las técnicas estadísticas, se realizó el levantamiento formal de requerimientos del algoritmo según el estándar IEEE/ISO/IEC 29148-2018, el cual establece una metodología estructurada para especificar funcionalidades, restricciones y criterios técnicos del sistema. El objetivo fue desarrollar un algoritmo confiable y eficiente para la plataforma Arduino MKR WAN 1310, capaz de procesar lecturas ambientales, corregir inconsistencias y asegurar la integridad de los datos antes del almacenamiento o transmisión.

El algoritmo opera con lecturas cada 15 minutos e incluye: validación de la primera lectura, detección de atípicos con K-Sigma y Welford, imputación con la Última Observación Llevada Adelante, eliminación de duplicados por marca de tiempo, almacenamiento en tarjeta SD y activación del modo de descanso profundo. El algoritmo debe ejecutarse dentro del intervalo sin estructuras complejas, tolerar fallos menores y mantener un historial de al menos 30 días. La Tabla 10 presenta los requerimientos funcionales, enfocados en adquisición, validación y manejo energético. La Tabla 11 resume los requerimientos no funcionales, orientados a eficiencia y estabilidad. La Tabla 12 detalla las interfaces externas, incluyendo sensores, almacenamiento y monitoreo del sistema.

Tabla 10. Requisitos funcionales del algoritmo, funciones esenciales que debe cumplir el sistema para procesar, validar e imputar datos ambientales en flujo

Identificador	Requisito funcional	Descripción
RF1	Lectura periódica de sensores ambientales.	El sistema debe realizar una lectura de temperatura, humedad y pH cada 15 minutos, utilizando sensores conectados vía I2C.
RF2	Validación de la primera lectura tras el arranque del nodo.	Se deben tomar hasta 10 lecturas consecutivas al iniciar el sistema. Si ninguna está dentro del umbral predefinido, se imputará un valor promedio.
RF3	Detección de datos atípicos en tiempo real.	Se aplicará el método K-Sigma sobre estadísticas calculadas con el algoritmo Welford para detectar valores fuera de ± 3 desviaciones estándar.
RF4	Imputación de datos faltantes o anómalos.	Para todo dato atípico o ausente se empleará la técnica Última Observación Llevada Adelante para completar el flujo.
RF5	Eliminación de datos duplicados.	Se verificará si la marca de tiempo de una lectura coincide con el inmediatamente anterior. Si es así, se descarta como duplicado.
RF6	Registro estructurado de los datos procesados.	Cada lectura debe incluir: marca de tiempo, valor original, valor imputado (si aplica), y una bandera que indique el tipo de procesamiento aplicado.

RF7

Activación del modo de ahorro energético.

El sistema debe entrar en modo de descanso profundo entre cada lectura para preservar la batería del nodo.

Tabla 11. Requisitos no funcionales del sistema, condiciones técnicas complementarias que garantizan eficiencia, tolerancia a fallos y bajo consumo energético

Identificador	Requisito no funcional	Descripción
RNF1	Bajo costo computacional.	Todos los procesos del algoritmo deben completarse en menos de un minuto y utilizar únicamente operaciones aritméticas simples para garantizar la ejecución.
RNF2	Uso eficiente de memoria.	El algoritmo no debe almacenar historiales completos; debe operar en línea utilizando estructuras optimizadas como los cálculos incrementales de Welford.
RNF3	Equilibrio entre precisión y eficiencia.	La selección de métodos debe priorizar la confiabilidad del resultado sin comprometer el tiempo de respuesta ni la capacidad de almacenamiento o cómputo.
RNF4	Tolerancia a fallos en sensores.	El sistema debe detectar fallas de lectura y continuar su operación imputando o descartando valores sin interrumpir el servicio.
RNF5	Maximización de la autonomía energética.	El código debe minimizar las operaciones fuera del modo de sueño profundo y evitar el uso de periféricos no esenciales para extender la duración de la batería.

Tabla 12. Interfaces externas del sistema, componentes físicos y lógicos que permiten la interacción del algoritmo con sensores, almacenamiento y usuario

Tipo	Elemento involucrado		Descripción
Usuario	Puerto serial		Proveen retroalimentación básica sobre el estado del nodo (encendido, lectura, error) y permiten depuración durante pruebas.
Hardware	Sensores	FS200-SHT10 y SEN0161	Conectados vía I2C, estos sensores miden temperatura, humedad y pH en el entorno subterráneo.
Almacenamiento	Módulo de tarjeta SD		Registra los datos procesados en archivos .CSV, asegurando la persistencia de información sin conexión a red.
Tiempo real	Módulo RTC DS3231SN		Proporciona marcas de tiempo precisos y autónomos, independientes del ciclo de encendido del microcontrolador.
Comunicación	Módulo LoRa		Permite enviar lecturas a una pasarela remota; su uso es opcional y depende del despliegue.
Software	Bibliotecas de Arduino (Wire, SD, RTC, etc.)		Se emplean librerías oficiales y optimizadas para interactuar con los sensores, módulo SD, y temporizador de bajo consumo.

4.2. Codificación del algoritmo de limpieza de datos

En esta etapa se presenta la implementación del algoritmo de limpieza de datos en flujo desarrollado para el nodo sensor del Internet de las Cosas Subterráneas. El objetivo es traducir los requerimientos previamente establecidos en una solución funcional y eficiente, adecuada para ejecutarse en el microcontrolador Arduino MKR WAN 1310. La codificación se llevó a cabo respetando las limitaciones de memoria, tiempo y consumo energético del dispositivo.

El algoritmo se compone de varias fases funcionales. En primer lugar, al inicio del sistema, se ejecuta una etapa de calibración inicial, donde cada sensor realiza 10 lecturas consecutivas. Durante esta fase, los datos son validados contra los límites físicos de operación; en caso de detectarse valores fuera de rango, se imputan utilizando un promedio móvil basado en las últimas dos lecturas válidas. Esta etapa permite establecer un conjunto inicial de parámetros estadísticos calculados mediante el algoritmo de Welford, excluyendo valores imputados.

Posteriormente, durante la operación regular, el sistema realiza una lectura cada 15 minutos, validando en primer lugar que la marca de tiempo sea nueva. Cada lectura es verificada contra los límites físicos y, posteriormente, contra los umbrales estadísticos dinámicos definidos por la regla de K-Sigma ($\mu \pm 3\sigma$). Las lecturas clasificadas como atípicas o fuera del rango físico se imputan utilizando Última Observación Llevada Adelante, y si se presentan más de tres imputaciones consecutivas, se sustituye esta técnica por un promedio móvil de las tres últimas lecturas válidas.

Adicionalmente, el algoritmo implementa una regla de excepción por error en cadena: si más del 80% de las lecturas han sido imputadas, el sensor se marca como posiblemente degradado y se registra esta condición en el archivo de diagnóstico. Todas las lecturas, su estado y posibles modificaciones se registran en la tarjeta SD. Tras completar el ciclo, el nodo entra en suspensión profunda hasta el siguiente intervalo. La Figura 16 presenta el pseudocódigo que resume el funcionamiento completo del algoritmo.

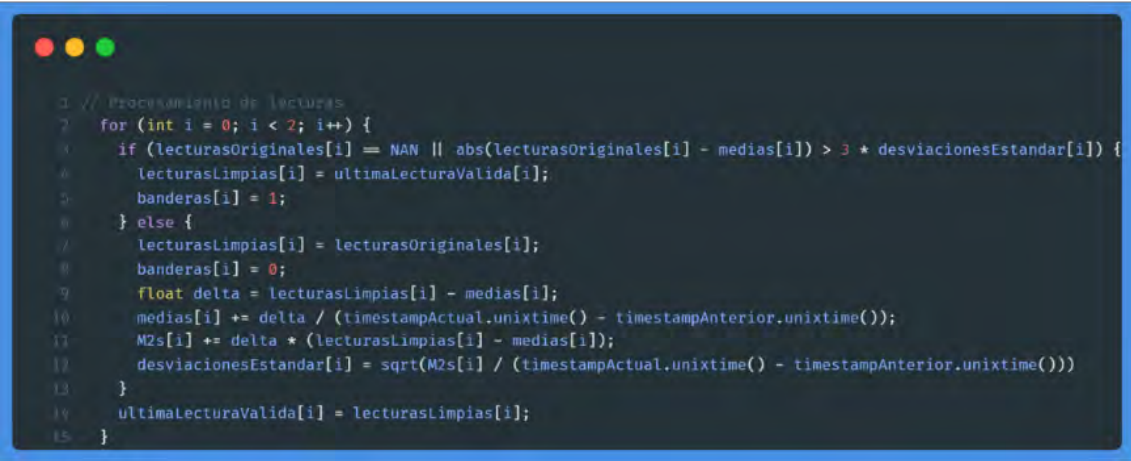
Algorithm 1 Algoritmo de limpieza de datos en flujo para dispositivos del IoUT

```
Inicialización
1: for all sensor en SENSORES do
2:   Inicializar variables: media, varianza, sigma, contador_validos
3:   Inicializar buffers: buffer_validos, buffer_ventana[12], indice_ventana  $\leftarrow$  0
4:   imputaciones_consecutivas  $\leftarrow$  0, sensor_degradado  $\leftarrow$  FALSO
5: end for
Fase 0 – Calibración inicial
6: for all sensor en SENSORES do
7:   while no se completan 10 lecturas válidas o imputadas do
8:     lectura  $\leftarrow$  Leer sensor
9:     if lectura está dentro del rango físico then
10:       Actualizar estadísticas iniciales
11:       Agregar a buffer_validos
12:     else if hay al menos 2 válidas previas then
13:       Imputar con promedio móvil de 2 valores previos
14:       Agregar a buffer_validos
15:     end if
16:   end while
17: end for
18: while el sistema esté activo do
19:   Esperar 15 minutos
20:   timestamp_actual  $\leftarrow$  leer RTC
21:   if timestamp_actual = timestamp_anterior then
22:     continuar
23:   end if
24:   for all sensor en SENSORES do
25:     lectura  $\leftarrow$  Leer sensor
26:     bandera  $\leftarrow$  "válida"
27:     if lectura fuera de rango físico then
28:       Imputar con LOCF o promedio móvil (si  $\geq 3$  imputaciones)
29:       bandera  $\leftarrow$  "fuera_rango_fisico"
30:     end if
31:     if lectura fuera de límites dinámicos ( $\mu \pm 3\sigma$ ) then
32:       Registrar límites
33:       Imputar como en la condición anterior
34:       bandera  $\leftarrow$  ".atípica"
35:       Incrementar imputaciones_consecutivas
36:       Marcar 1 en buffer_ventana
37:     else
38:       Actualizar estadísticos con Welford
39:       Reiniciar imputaciones_consecutivas
40:       Marcar 0 en buffer_ventana
41:     end if
42:     Avanzar indice_ventana
43:     if 10 o más imputaciones en las últimas 12 lecturas then
44:       sensor_degradado  $\leftarrow$  VERDADERO
45:       Registrar evento en diagnostico.csv
46:     end if
47:     Registrar en datos.csv los valores procesados
48:   end for
49:   if cambio de día detectado then
50:     Generar resumen diagnóstico diario
51:   end if
52:   timestamp_anterior  $\leftarrow$  timestamp_actual
53:   Activar modo deep sleep por 15 minutos
54: end while
```

Figura 16. Pseudocódigo generalizado del algoritmo de limpieza de datos en flujo desarrollado para el sistema del Internet de las Cosas Subterráneas. *Fuente: Elaboración propia.*

4.3. Integración del algoritmo al nodo sensor subterráneo

El algoritmo de limpieza en flujo fue implementado en C++, adaptado para plataformas embebidas de bajo consumo, y desplegado sobre el nodo sensor descrito en el Capítulo 5, asegurando su funcionamiento en condiciones reales. Se integró con los módulos del sistema para realizar lecturas periódicas, validar datos, imputar errores, almacenar registros en la tarjeta SD y gestionar el consumo energético, operando en ciclos de 15 minutos. La Figura 17 muestra un fragmento representativo del código, donde se evidencian las decisiones estructuradas del algoritmo en el microcontrolador.

A screenshot of a code editor showing a C++ function for processing sensor readings. The code is written in a dark-themed editor with a blue border. It includes comments and logic for handling NaN values, calculating standard deviations, and updating data arrays. The code is as follows:

```
3 // Procesamiento de lecturas
7 for (int i = 0; i < 2; i++) {
8     if (lecturasOriginales[i] == NAN || abs(lecturasOriginales[i] - medias[i]) > 3 * desviacionesEstandar[i]) {
9         lecturasLimpias[i] = ultimaLecturaValida[i];
10        banderas[i] = 1;
11    } else {
12        lecturasLimpias[i] = lecturasOriginales[i];
13        banderas[i] = 0;
14        float delta = lecturasLimpias[i] - medias[i];
15        medias[i] += delta / (timestampActual.unixtime() - timestampAnterior.unixtime());
16        M2s[i] += delta * (lecturasLimpias[i] - medias[i]);
17        desviacionesEstandar[i] = sqrt(M2s[i] / (timestampActual.unixtime() - timestampAnterior.unixtime()))
18    }
19    ultimaLecturaValida[i] = lecturasLimpias[i];
20 }
```

Figura 17. Fragmento del código implementado en C++ para la ejecución del algoritmo de limpieza de datos en el nodo sensor subterráneo. *Fuente: Elaboración propia.*

El código fuente completo del algoritmo, junto con los archivos de configuración y los scripts de prueba, puede ser consultado en el siguiente repositorio: <https://github.com/Julio-Schez/StreamDataCleaning.git>. Se realizaron pruebas de implementación en un entorno subterráneo controlado, para ello, el nodo sensor fue enterrado a una profundidad de 25 cm (ver Figura 18), replicando las condiciones físicas de un entorno subterráneo. Durante varios días se ejecutaron ciclos de lectura automática, con almacenamiento local y transmisión de datos al nodo receptor ubicado a 10 metros de distancia. Estas pruebas permitieron validar la ejecución continua del algoritmo, la respuesta del sistema ante datos anómalos y la estabilidad de la comunicación.



Figura 18. Proceso de instalación del nodo sensor subterráneo embebido con el algoritmo de limpieza de datos en flujo para realizar las pruebas de integración. *Fuente: Elaboración propia.*

Se evaluó la calidad de la señal inalámbrica entre el nodo subterráneo y el receptor mediante el análisis del Indicador de Intensidad de Señal Recibida (RSSI). Como se muestra en la Figura 19, el 5 de diciembre se obtuvo una media de -95.17 dBm, con valores entre -109.9 y -75.6 dBm. Estos resultados confirman una comunicación estable y confiable, incluso en condiciones subterráneas, con recepción completa de los registros.

Timestamp	Sensor	Valor Original	Valor Limpio	Procesado	RSSI
05/12/2024 00:00	Temperatura	13.9	13.9	0	-105.8
05/12/2024 00:15	Temperatura	13.8	13.9	0	-109.2
05/12/2024 00:30	Temperatura	13.7	13.7	0	-80.6
05/12/2024 00:45	Temperatura	13.5	13.5	0	-93.7
05/12/2024 01:00	Temperatura	13.4	13.4	0	-101.6
05/12/2024 01:15	Temperatura	13.3	13.3	0	-91.9
05/12/2024 01:30	Temperatura	13.2	13.2	0	-99.5
05/12/2024 01:45	Temperatura	13.1	13.1	0	-103.9
05/12/2024 02:00	Temperatura	13	13	0	-100.3
05/12/2024 02:15	Temperatura	12.9	12.9	0	-101.2
05/12/2024 02:30	Temperatura	12.8	12.9	0	-81.1
05/12/2024 02:45	Temperatura	12.8	12.8	0	-76.1
05/12/2024 03:00	Temperatura	12.7	12.7	0	-97.6
05/12/2024 03:15	Temperatura	12.6	12.6	0	-100.1
05/12/2024 03:30	Temperatura	12.5	12.5	0	-93.9
05/12/2024 03:45	Temperatura	12.4	12.4	0	-76.6
05/12/2024 04:00	Temperatura	12.3	12.3	0	-107.2
05/12/2024 04:15	Temperatura	12.2	12.2	0	-99.2

Figura 19. Fragmento de las lecturas de temperatura registradas, incluyendo los valores de RSSI para cada transmisión entre el nodo sensor subterráneo y el nodo receptor. *Fuente: Elaboración propia.*

Se realizó una prueba controlada introduciendo un valor atípico de -300 en la lectura de temperatura. El algoritmo lo detectó como inválido, lo imputó con el último valor válido (14.2) y marcó correctamente su estado, asegurando la continuidad de la serie temporal (ver Figura 20).

Timestamp	Sensor	Valor Original	Valor Limpio	Procesado	RSSI
05/12/2024 08:00	Temperatura	13.7	13.7	0	-80.6
05/12/2024 08:15	Temperatura	14	14	0	-98.6
05/12/2024 08:30	Temperatura	14.2	14.2	0	-102.5
05/12/2024 08:45	Temperatura	-300	14.2	1	-84.4
05/12/2024 09:00	Temperatura	-300	14.2	1	-77.6
05/12/2024 09:15	Temperatura	15.4	15.4	0	-100.5
05/12/2024 09:30	Temperatura	16	16	0	-80.4
05/12/2024 09:45	Temperatura	16.6	16.6	0	-87.3
05/12/2024 10:00	Temperatura	17.2	17.2	0	-105.6
05/12/2024 10:15	Temperatura	17.7	17.7	0	-82.2
05/12/2024 10:30	Temperatura	18.2	18.2	0	-105.4

Figura 20. Simulación de error en el sensor de temperatura (-300) y su imputación automática con el último valor válido (14.2), validando el funcionamiento del algoritmo. *Fuente: Elaboración propia.*

5

Evaluación del sistema integrado de limpieza de datos

Este capítulo evalúa el desempeño del algoritmo integrado de limpieza de datos en condiciones subterráneas, operando de forma autónoma. Se analizan métricas clave como detección de anomalías, imputación de datos faltantes, eliminación de duplicados y errores tipo I y II. Además, se realiza un análisis cualitativo de los registros, observando la evolución de umbrales y la lógica del algoritmo. Esta evaluación permite validar la eficacia, robustez y adaptabilidad del sistema frente a variaciones ambientales.

5.1. Evaluación experimental

Durante la validación, el sistema operó de forma autónoma del 25 de febrero al 4 de marzo de 2025, recolectando 1959 lecturas de humedad, temperatura y pH cada 15 minutos. Cada registro fue procesado en tiempo real mediante el algoritmo de limpieza y almacenado en una tarjeta SD. Se confirmó la ausencia de eventos anómalos, lo cual, valida la estabilidad del sistema, aunque limita la evaluación en escenarios críticos. La Figura 21 muestra una humedad entre 24.90 % y 44.55 %, la Figura 22 una temperatura entre 12.3 °C y 29.94 °C, y la Figura 23 un pH entre 5.92 y 7.70, con mayor dispersión atribuible a la variabilidad natural del suelo.

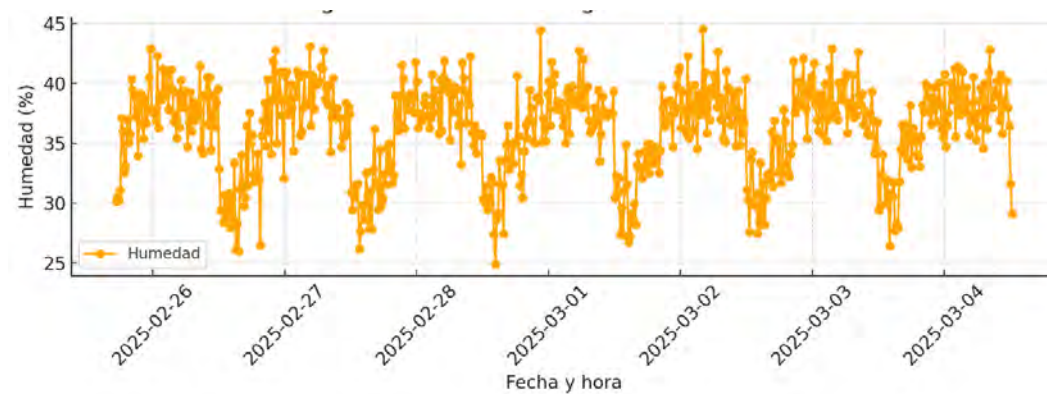


Figura 21. Evolución temporal de las lecturas de humedad durante el periodo de validación. Se observa una señal estable y continua. *Fuente: Elaboración propia.*

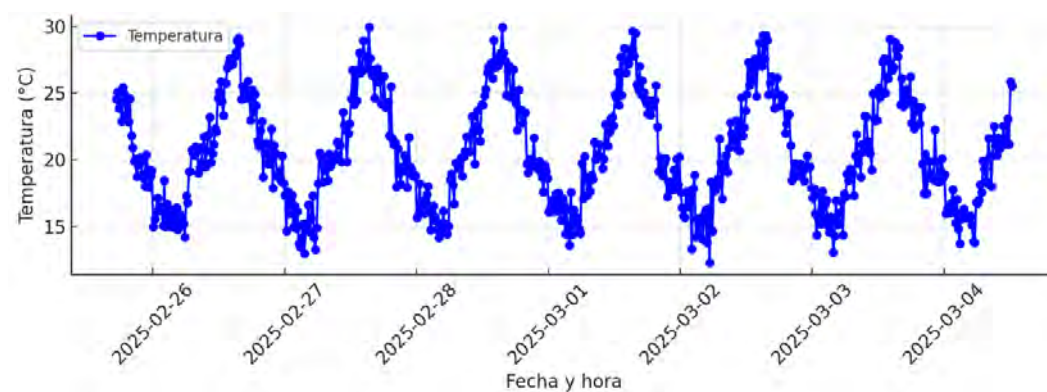


Figura 22. Registro de lecturas de temperatura del suelo obtenidas por el sistema embebido. La variación cíclica corresponde a cambios ambientales naturales. *Fuente: Elaboración propia.*

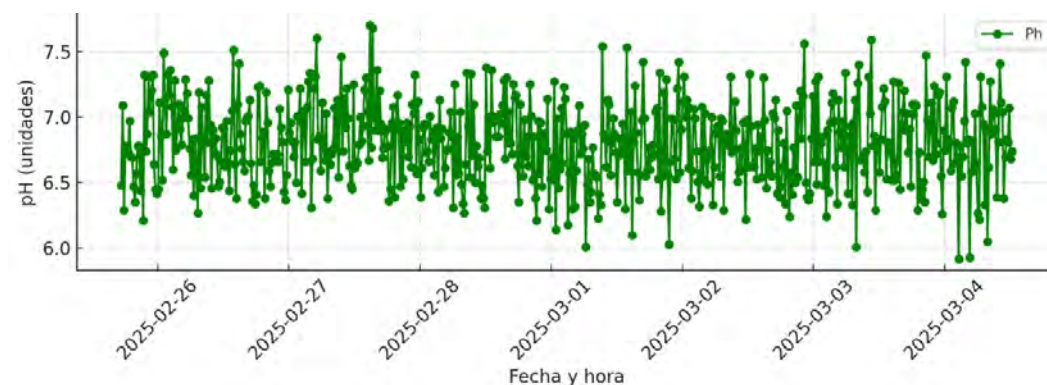


Figura 23. Lecturas de pH del suelo durante el periodo experimental. Aunque se aprecia una mayor dispersión en los datos, estas variaciones corresponden al comportamiento natural del sensor en condiciones subterráneas, sin representar eventos anómalos. *Fuente: Elaboración propia.*

El análisis inicial mostró un funcionamiento estable del sistema, pero evidenció la necesidad de evaluarlo bajo condiciones críticas. Por ello, se diseñó una metodología experimental complementaria basada en la generación de subconjuntos con 10 %, 30 % y 50 % de anomalías, incorporando atípicos, omisiones y duplicados de forma controlada. Luego, se evaluó cuantitativa y cualitativamente el desempeño del algoritmo ante dichas perturbaciones. La Figura 24 muestra la metodología empleada para la experimentación.

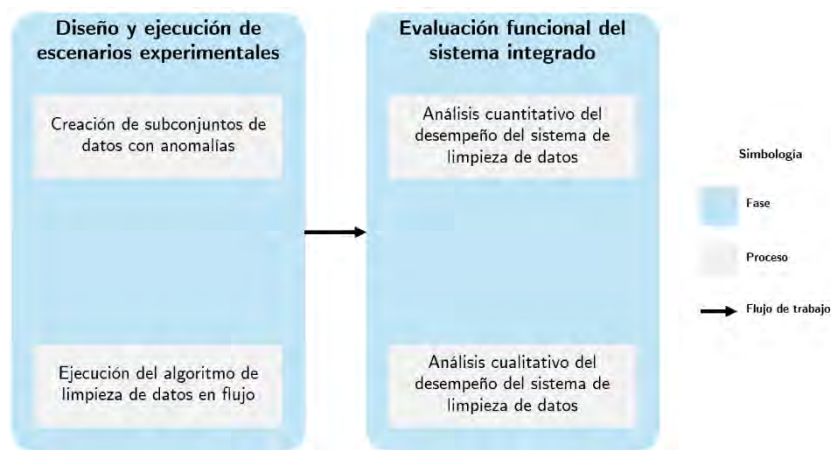


Figura 24. Metodología experimental para la validación del sistema de limpieza de datos, estructurada en dos fases. *Fuente: Elaboración propia.*

La experimentación se apoyó en recursos tanto físicos como no físicos, seleccionados bajo criterios de confiabilidad, trazabilidad y repetibilidad, con el fin de asegurar condiciones controladas durante la validación del sistema.

Recursos físicos

Se utilizó el nodo sensor subterráneo desarrollado en el Capítulo 5, compuesto por una unidad de procesamiento (Arduino MKR WAN 1310), sensores de humedad, temperatura y pH del suelo, reloj de tiempo real, almacenamiento en tarjeta microSD y sistema de alimentación autónomo. La Tabla 13 resume los componentes clave y su función general en el sistema.

Tabla 13. Componentes físicos utilizados en la evaluación del sistema. Incluye el nodo sensor subterráneo desarrollado y la estación de trabajo para programación y respaldo

Recurso	Descripción técnica	Función
Nodo sensor subterráneo	Unidad embebida con sensores ambientales, microcontrolador, RTC y almacenamiento local.	Recolección, procesamiento y almacenamiento autónomo de lecturas.
Computadora portátil	Estación de trabajo con herramientas de programación y monitoreo.	Carga del algoritmo, respaldo y análisis de datos

Recursos no físicos

Los recursos intangibles incluyeron herramientas de desarrollo, bibliotecas y plataformas de análisis utilizadas para implementar, ejecutar y validar el sistema. La Tabla 14 presenta un resumen de estos recursos.

Tabla 14. Recursos no físicos empleados en la implementación y validación del sistema. Comprenden herramientas de desarrollo, bibliotecas embebidas y entornos de análisis de datos

Recurso	Descripción técnica	Función
Algoritmo de limpieza de datos	Código embebido en C++ con detección de atípicos, corrección de fatos faltantes y duplicados.	Núcleo lógico del sistema, ejecutado en tiempo real.
Arduino IDE	Entorno de desarrollo oficial para placas Arduino.	Programación y carga del algoritmo al microcontrolador.
Bibliotecas de Arduino	Wire.h, SD.h, RTCLib.h, SPI.h, entre otras	Soporte para sensores, reloj, sistema de archivos y comunicación
Python 3 con Jupyter Notebooks	Plataforma de desarrollo para análisis estadístico y generación de gráficas	Procesamiento de datos, simulación de escenarios y validación
Librerías Python	pandas, numpy, matplotlib, seaborn	Cálculo de métricas y visualización de resultados

Google Sheets	Hoja de cálculo en línea	Verificación cruzada de registros detectados o imputados manualmente
----------------------	--------------------------	--

Con base en estos recursos, se desarrolló la metodología experimental descrita anteriormente, cuyas fases y actividades clave se detallan a continuación.

5.1.1. Diseño y ejecución de escenarios experimentales

Esta fase se centró en la generación de subconjuntos modificados con distintos niveles de anomalías, así como en la ejecución del algoritmo sobre dichos conjuntos, replicando condiciones críticas representativas de fallas reales.

Creación de subconjuntos de datos con anomalías

Se generaron tres subconjuntos artificiales a partir del conjunto base, con niveles de anomalías del 10 %, 30 % y 50 %, representando distintos grados de afectación operativa. Las anomalías se introdujeron de forma controlada, garantizando trazabilidad y realismo. En cada conjunto, los registros se seleccionaron aleatoriamente y se distribuyeron equitativamente entre temperatura, humedad y pH. A cada segmento se le aplicó un tipo de error: valores atípicos, duplicados o datos faltantes, respetando la secuencia temporal. La Tabla 15 detalla la cantidad y distribución de anomalías inducidas.

Tabla 15. Distribución de anomalías introducidas en los subconjuntos con niveles de alteración del 10 %, 30 % y 50 %. Cada conjunto incluye valores atípicos, duplicados y faltantes en proporciones iguales para simular escenarios con diferentes grados de alteración

Subconjunto	Nivel de anomalías	Registros modificados	Atípicos	Duplicados	Faltantes
V1	10%	196	65	65	66
V2	30%	588	196	196	196
V3	50%	979	326	327	326

Ejecución del algoritmo de limpieza de datos

Para evaluar los subconjuntos con anomalías, se adaptó el sistema embebido para simular un flujo de datos desde archivos CSV en lugar de sensores en tiempo real. Esto permitió ejecutar el algoritmo bajo condiciones controladas, manteniendo el ciclo lógico de validación, imputación y almacenamiento. La Figura 25 compara el funcionamiento original del nodo con la versión modificada, donde los registros provienen de archivos con anomalías simuladas, lo que facilitó la evaluación sin alterar la arquitectura ni el flujo operativo.

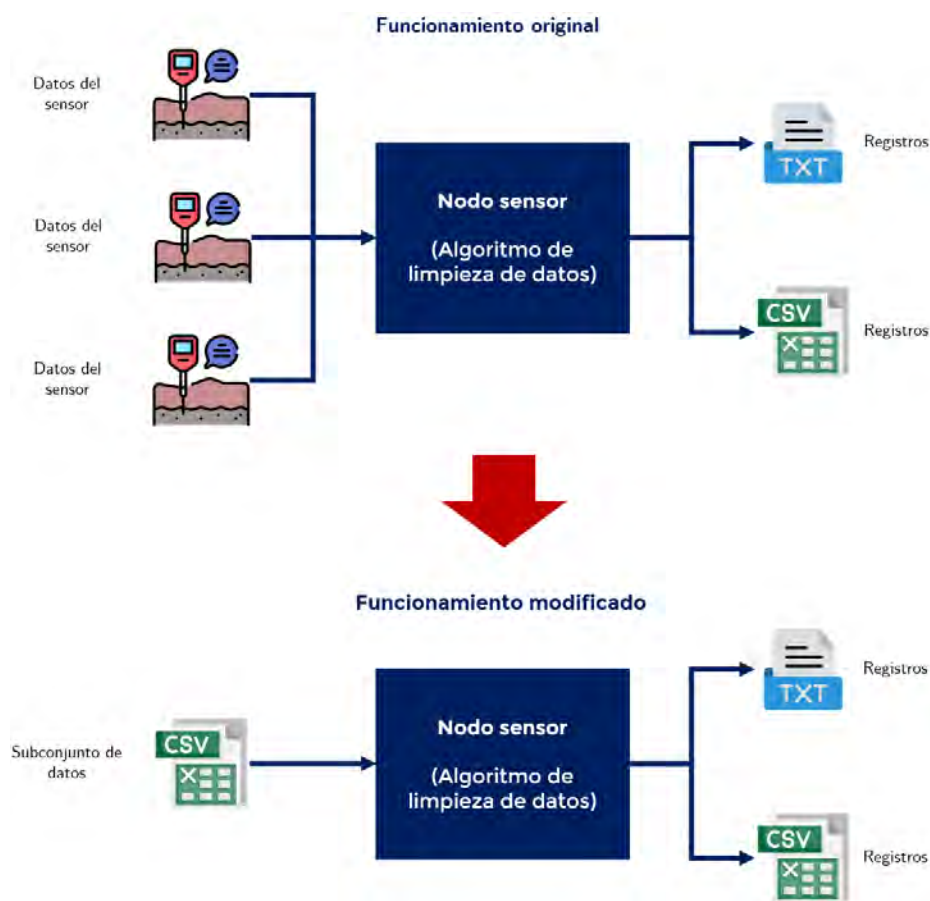


Figura 25. Esquema comparativo del funcionamiento original y modificado del nodo sensor. En el modo original, el algoritmo procesa datos de sensores físicos en tiempo real; en el modo modificado, los datos provienen de archivos CSV con errores inducidos. *Fuente: Elaboración propia.*

Durante la simulación, el sistema operó de forma autónoma, registrando todas las decisiones tomadas. Las estadísticas internas se actualizaron solo con datos válidos,

asegurando umbrales dinámicos confiables. La Tabla 16 resume el comportamiento del algoritmo ante cada tipo de dato procesado.

Tabla 16. Comportamiento del algoritmo ante los distintos tipos de datos procesados. Se describen los criterios de identificación, las acciones ejecutadas por el sistema y los resultados esperados durante el procesamiento en flujo

Tipo de dato	Criterio de identificación	Acción aplicada por el algoritmo	Resultado esperado
Dato válido	Dentro de los límites (media $\pm 3\sigma$)	Actualización de estadísticas internas	Registro aceptado
Dato atípico	Fuera de los límites definidos por K-Sigma + Welford	Excluido de estadísticas, imputado con la Última Observación Válida o Promedio Móvil	Registro sustituido
Dato faltante	Valor ausente en el registro	Imputado con la Última Observación Válida o Promedio Móvil	Continuidad de la serie mantenida
Dato duplicado	Marca de tiempo igual al registro anterior	Registro descartado	Evitar redundancia y sobrecarga de información

Este procedimiento permitió validar el funcionamiento autónomo del sistema ante situaciones anómalas, reflejando condiciones críticas similares a las que podrían encontrarse en entornos subterráneos reales.

5.1.2. Evaluación funcional del sistema integrado

Esta fase analizó los resultados del algoritmo aplicado a subconjuntos con distintos niveles de anomalías, evaluando su capacidad para mantener la calidad de los datos. Se combinó análisis cuantitativo y cualitativo para obtener una visión integral del desempeño del sistema.

Análisis cuantitativo del desempeño del algoritmo de limpieza de datos

Se calcularon indicadores clave para cada subconjunto con anomalías: tasa de detección de anomalías, falsos positivos, falsos negativos, porcentaje de imputaciones y tasa de eliminación de duplicados. Estos indicadores permitieron estimar la precisión, sensibilidad y eficacia del sistema, y se organizaron en una tabla resumen para facilitar la comparación entre escenarios.

Análisis cualitativo del desempeño del algoritmo de limpieza de datos

De forma complementaria, se realizó un análisis cualitativo mediante visualizaciones temporales de los registros procesados, enfocándose en fragmentos con imputaciones por datos atípicos o faltantes. Se verificó la evolución de los umbrales dinámicos, confirmando que solo se actualizaron con datos válidos. Este análisis permitió identificar patrones, validar la coherencia del sistema y detectar casos puntuales de falsos positivos o negativos, fortaleciendo la interpretación de los resultados.

5.2. Resultados obtenidos

Este apartado presenta los resultados de la evaluación experimental del sistema tras ejecutar el algoritmo sobre subconjuntos con distintos niveles de anomalías. Se analizan indicadores cuantitativos por escenario y observaciones cualitativas basadas en fragmentos representativos de las series procesadas.

5.2.1. Subconjunto de datos con 10% de anomalías

En el conjunto con 10 % de anomalías se inyectaron 65 valores atípicos, 65 duplicados y 66 datos faltantes (196 anomalías en total). El algoritmo detectó 60 de los 65 atípicos (92.31 %), con 1 falso positivo (1.64 %) y 5 falsos negativos (7.69 %). Imputó 126 registros, equivalentes al 96.18 % de los 131 datos imputables (faltantes y atípicos), y eliminó el 100 % de los duplicados.

La Figura 26 muestra las primeras 50 lecturas de humedad, destacando la calibración inicial y la posterior estabilización de los umbrales. La única imputación

registrada confirma la estabilidad del sistema y la correcta actualización estadística del algoritmo.

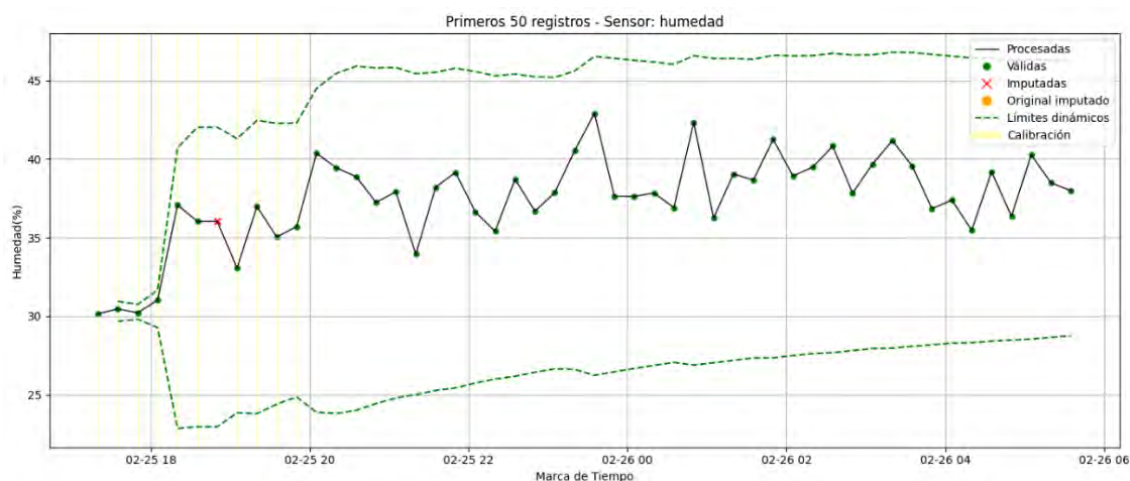


Figura 26. Evolución de las primeras 50 lecturas del sensor de humedad. La gráfica evidencia una operación estable posterior a la calibración, con límites adaptativos y sin imputaciones erráticas.
Fuente: *Elaboración propia.*

La Figura 27 muestra la evolución de la humedad durante el monitoreo, con una calibración inicial seguida de una operación estable. Las imputaciones, concentradas entre el 27 de febrero y el 3 de marzo, responden a posibles perturbaciones externas. Aun así, los límites se mantienen constantes, lo que confirma que las estadísticas solo se actualizan con datos válidos, según lo previsto en el diseño.

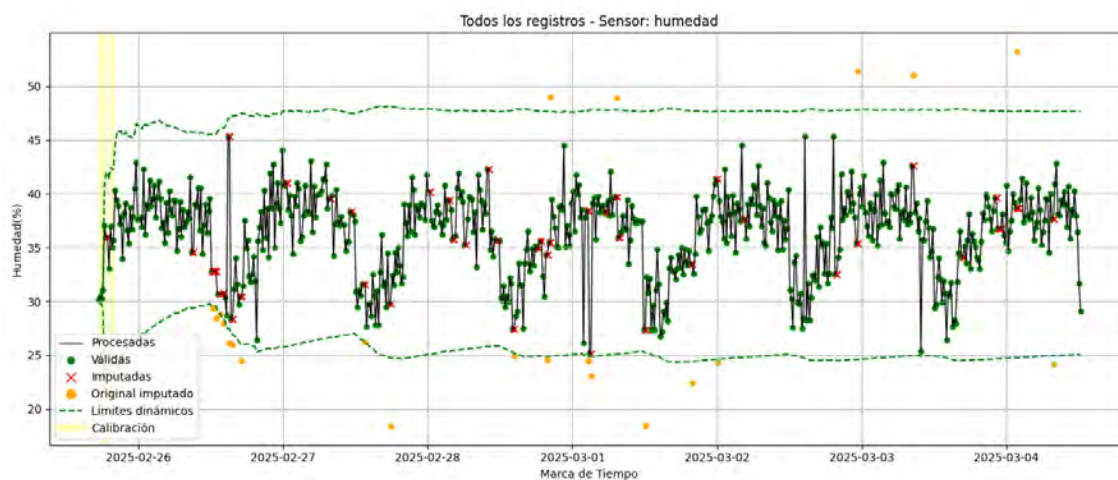


Figura 27. Serie completa de lecturas del sensor de humedad. El sistema muestra una operación estable y una detección eficaz de valores atípicos. Fuente: *Elaboración propia.*

La Figura 28 muestra las primeras 50 lecturas de temperatura tras el arranque del sistema. Durante la fase de calibración, los límites dinámicos se ajustan con base en datos válidos, permitiendo una estimación confiable de umbrales. Las lecturas iniciales son estables, lo que facilita una calibración eficaz. Posteriormente, se observa una disminución natural de temperatura y, al final, un ascenso súbito por encima de 28 °C, seguido de un descenso inmediato. Aunque inusual, esta variación se mantuvo dentro de los umbrales y fue clasificada como válida. En este fragmento no se registraron imputaciones, lo que sugiere un entorno estable o una sensibilidad conservadora del sistema ante transiciones rápidas.

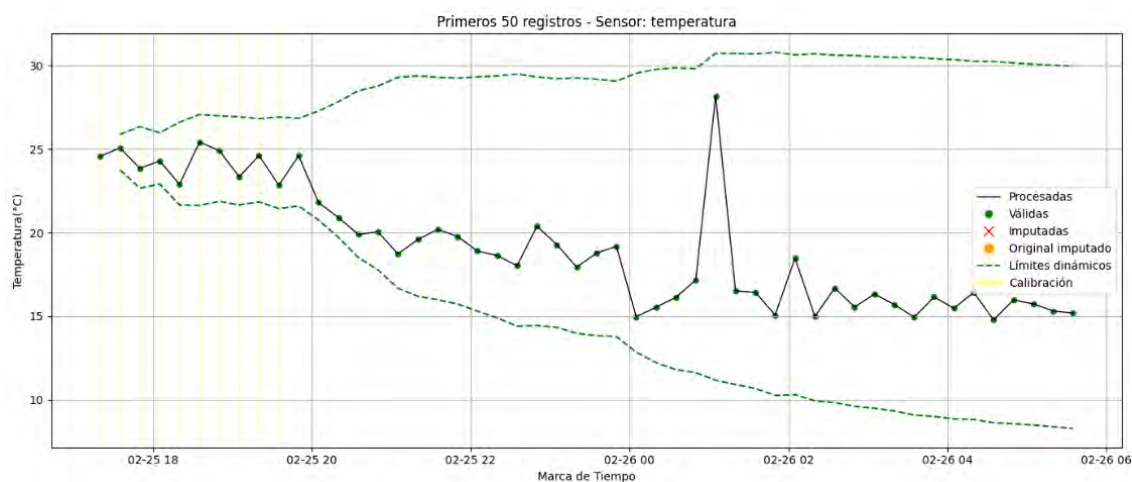


Figura 28. Comportamiento de las primeras 50 lecturas del sensor de temperatura. La señal presenta un descenso progresivo de temperatura y un pico aislado, que es aceptado como válido, no se registraron imputaciones en este intervalo. *Fuente: Elaboración propia.*

La Figura 29 muestra la serie completa de lecturas de temperatura durante la evaluación en campo. Al inicio, la fase de calibración permite ajustar los límites dinámicos con lecturas válidas, estableciendo umbrales estadísticos representativos. Tras esta etapa, el sistema opera de forma continua, registrando variaciones cíclicas coherentes con los cambios día-noche. Se identifican varias imputaciones, concentradas en picos que superan el límite superior, lo que indica una detección efectiva de perturbaciones puntuales. Los límites se mantienen estables y adaptados a la señal, sin saltos abruptos, lo que confirma que las estadísticas se actualizaron correctamente con datos válidos.

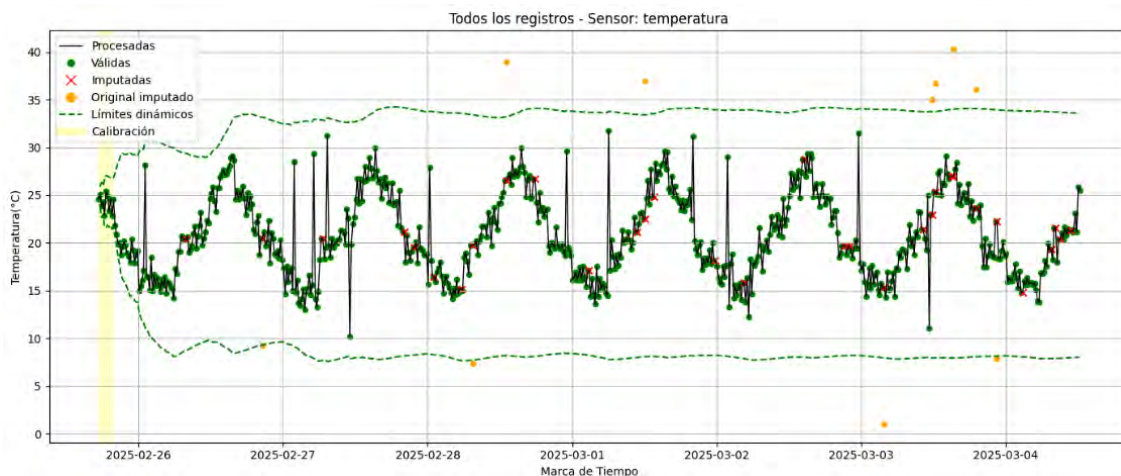


Figura 29. Serie completa del sensor de temperatura. La señal presenta oscilaciones cíclicas consistentes con el entorno, y el sistema actúa oportunamente corrigiendo lecturas atípicas sin afectar la estabilidad general del proceso de limpieza. *Fuente: Elaboración propia.*

La Figura 30 muestra las primeras 50 lecturas del sensor de pH tras el arranque. En la fase inicial, el sistema calibra los límites dinámicos a partir de lecturas válidas, estableciendo umbrales estadísticos para la detección de valores anómalos. Finalizada esta etapa, el sistema opera de manera estable, con la mayoría de las lecturas dentro del rango esperado. Hacia la segunda mitad del gráfico, se registran varias imputaciones realizadas mediante la técnica de Última Observación Llevada Adelante. Las observaciones originales descartadas se marcan como puntos naranjas, destacando.

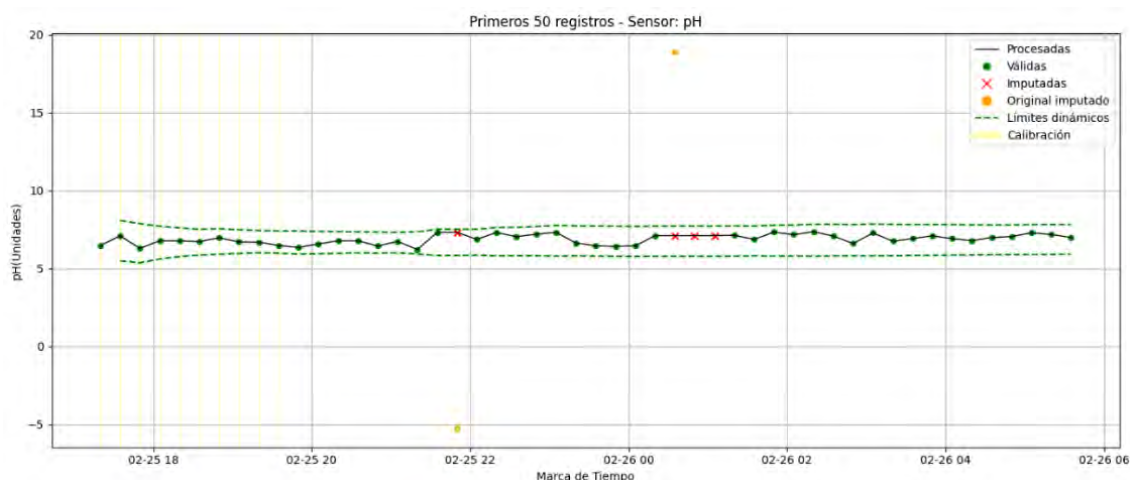


Figura 30. Comportamiento de las primeras 50 lecturas del sensor de pH. El sistema detecta oportunamente valores atípicos extremos y mantiene estabilidad estadística durante el proceso de imputación. *Fuente: Elaboración propia.*

La Figura 31 muestra la serie completa de lecturas del sensor de pH durante el periodo de operación del nodo subterráneo. Al inicio se observa la fase de calibración, donde el sistema ajusta los límites dinámicos a partir de datos válidos para establecer una base estadística confiable. Posteriormente, el sistema opera de forma continua, con la señal de pH mantenida de manera estable entre aproximadamente 6.5 y 7.5, lo que indica condiciones constantes en el entorno. Se registran varias imputaciones, correspondientes a lecturas descartadas por estar fuera de los umbrales definidos.

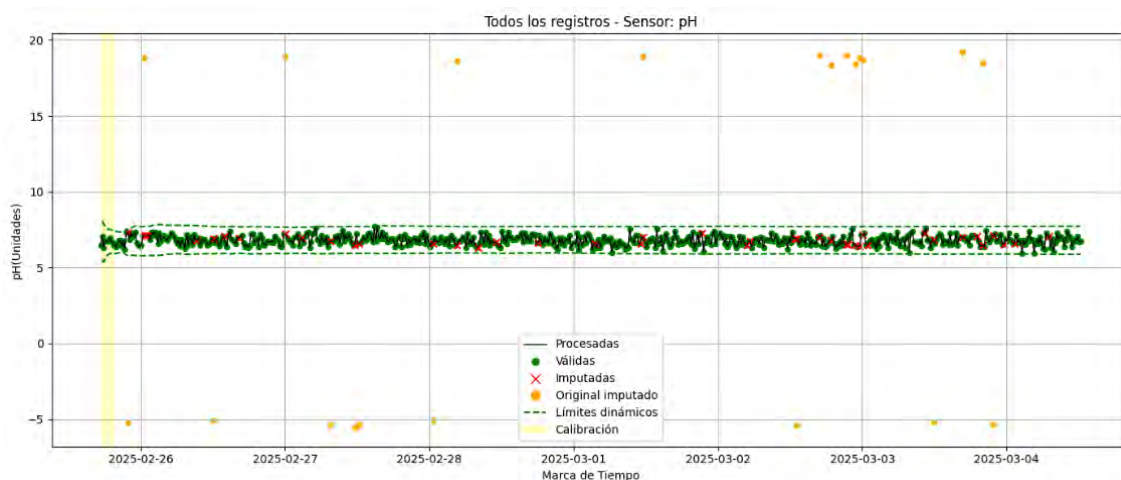


Figura 31. Serie completa del sensor de pH. El sistema detecta y corrige valores atípicos extremos sin comprometer la estabilidad estadística de la señal a lo largo del periodo. *Fuente: Elaboración propia.*

5.2.2. Subconjunto de datos con 30% de anomalías

En el conjunto con 30 % de anomalías se incorporaron 196 valores atípicos, 196 duplicados y 196 datos faltantes, sumando 588 anomalías. El algoritmo mostró un desempeño favorable, especialmente en la eliminación de duplicados (100 %) y la imputación de datos faltantes y atípicos (356 registros, 90.82 % de 392 casos imputables).

El sistema detectó el 81.63 % de los valores atípicos, con 3.03 % de falsos positivos y 18.37 % de falsos negativos, principalmente en temperatura y humedad. La Figura 32 muestra múltiples imputaciones en las primeras 50 lecturas de humedad tras la calibración, asociadas a secuencias anómalas descartadas por estar fuera del rango estadístico.

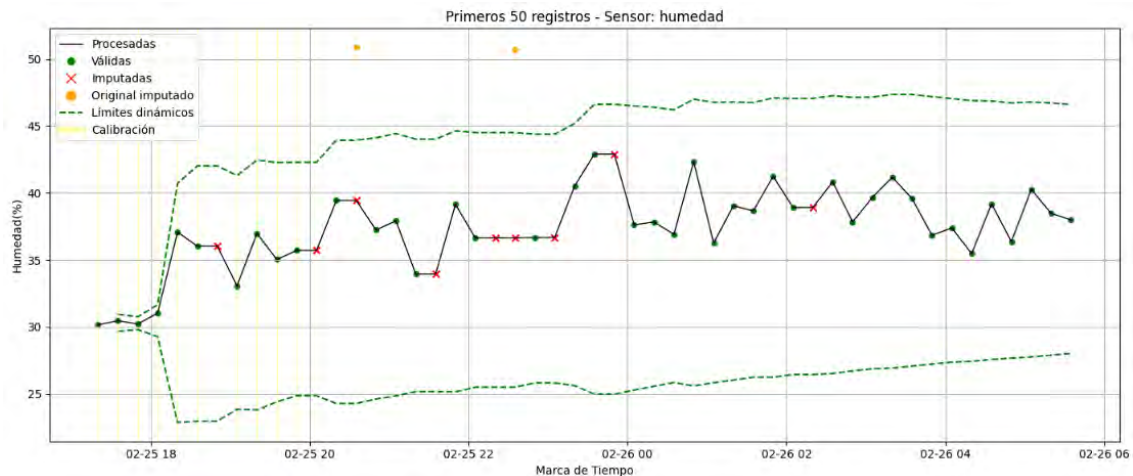


Figura 32. Primeras 50 lecturas del sensor de humedad con 30% de anomalías. El sistema detecta y corrige adecuadamente atípicos individuales y secuenciales. *Fuente: Elaboración propia.*

La Figura 33 muestra la serie completa de humedad con 30 % de anomalías. Tras la calibración, el sistema estabiliza los umbrales y opera de forma continua, imputando eficazmente valores fuera del rango esperado (por encima de 50 % o debajo de 20 %) mediante la técnica de Última Observación Llevada Adelante para conservar la continuidad de la señal.

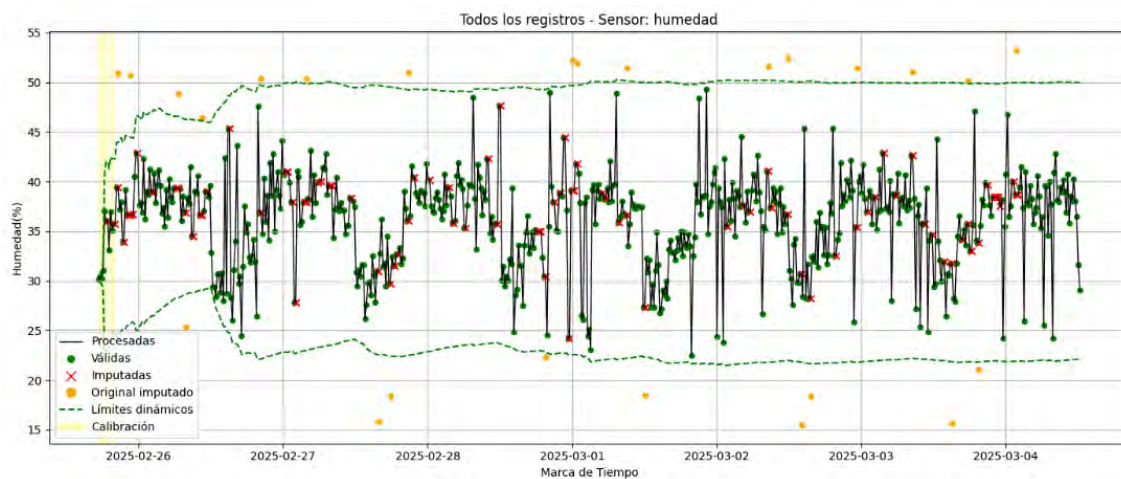


Figura 33. Serie completa del sensor de humedad con 30% de anomalías. El sistema mantiene estabilidad estadística y responde eficazmente ante perturbaciones frecuentes y severas. *Fuente: Elaboración propia.*

La Figura 34 muestra las primeras 50 lecturas de temperatura en el subconjunto con 30 % de anomalías. Durante la fase inicial de calibración, el sistema ajusta los límites

dinámicos a partir de datos válidos, siguiendo el algoritmo estadístico definido. Luego, la serie presenta una tendencia descendente con oscilaciones, incluyendo un pico cercano a 28 °C que se mantuvo dentro de los límites y fue clasificado como válido. Se observan también varias imputaciones, correspondientes a lecturas descartadas por exceder los umbrales, como una registrada cerca del minuto 25 con un valor superior a 31 °C, claramente fuera del comportamiento esperado.

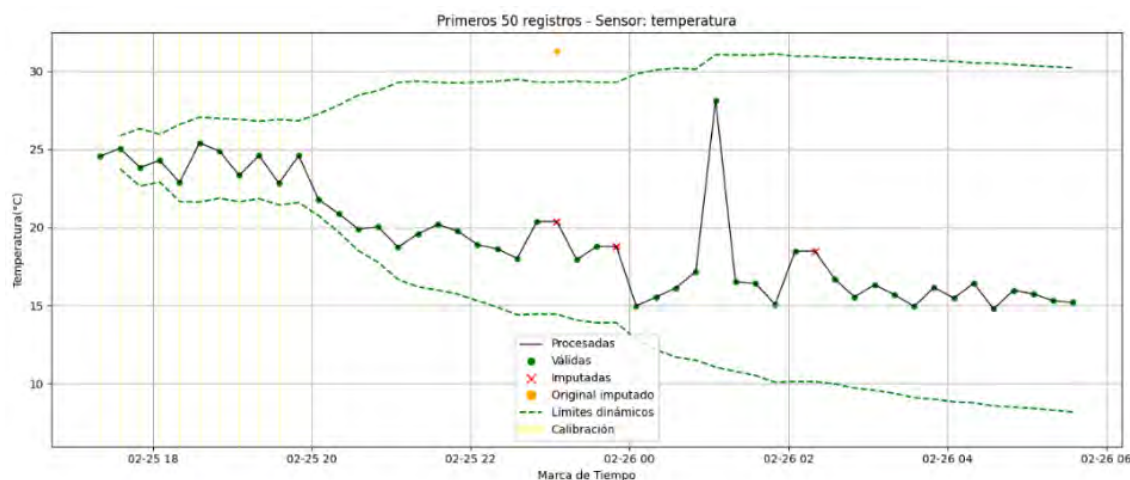


Figura 34. Primeras 50 lecturas del sensor de temperatura con 30% de anomalías. El sistema responde adecuadamente a la presencia de anomalías y conserva la estabilidad estadística de los límites. *Fuente: Elaboración propia.*

La Figura 35 muestra la serie completa de temperatura para el subconjunto con 30 % de anomalías. Al inicio se observa la fase de calibración, donde se establecen los límites dinámicos a partir de lecturas válidas, mediante un algoritmo estadístico robusto. Durante la operación continua, la señal presenta un patrón cíclico asociado a la variación térmica día-noche. Sin embargo, se registran numerosas imputaciones distribuidas a lo largo del periodo, resultado de la detección de valores atípicos. Los datos descartados incluyen picos por encima de 35 °C y caídas por debajo de 5 °C, claramente fuera del rango esperado.

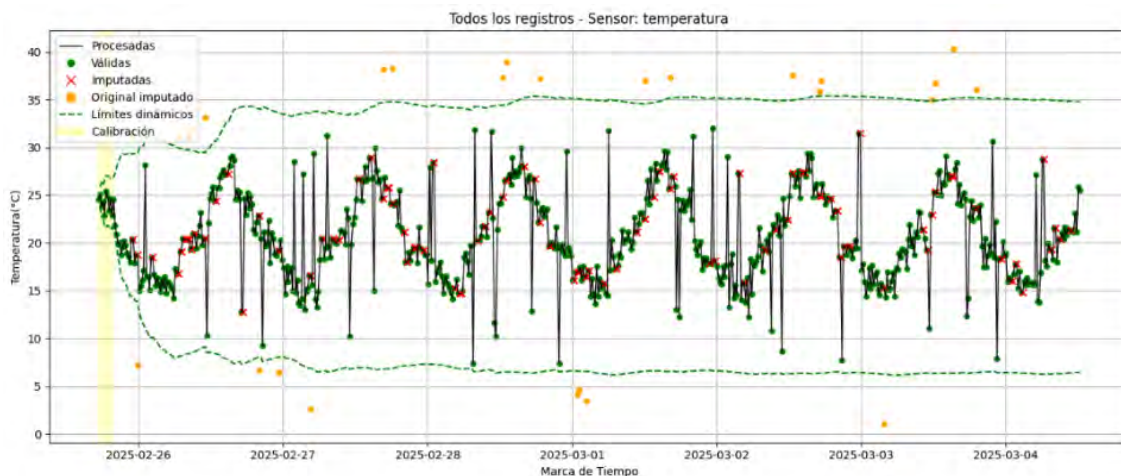


Figura 35. Serie completa del sensor de temperatura con 30% de anomalías. El sistema conserva la integridad de los ciclos térmicos y responde eficazmente ante múltiples perturbaciones sin desestabilizar las estadísticas internas

La Figura 36 muestra las primeras 50 lecturas del sensor de pH en el subconjunto con 30 % de anomalías. Durante la fase de calibración, el sistema ajusta los límites estadísticos a partir de datos válidos. Tras esta etapa, la señal se mantiene mayormente estable, con lecturas válidas entre 6.5 y 7.5 unidades. No obstante, debido a anomalías, se registran múltiples imputaciones, algunas en secuencia, resultado de lecturas descartadas por exceder los umbrales.

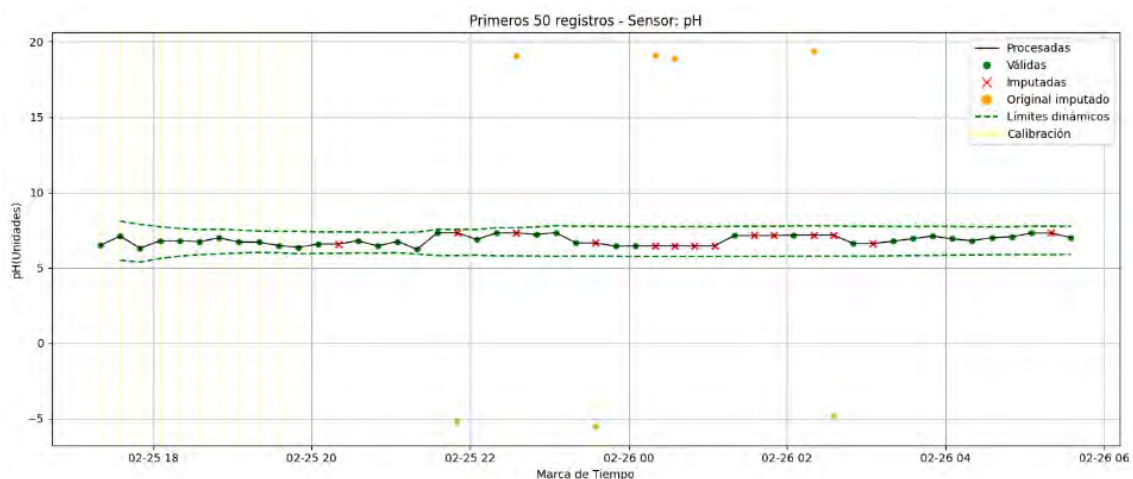


Figura 36. Primeras 50 lecturas del sensor de pH con 30% de anomalías. El sistema identifica y corrige valores extremos sin comprometer la integridad de la señal. *Fuente: Elaboración propia.*

La Figura 37 muestra la serie completa de pH durante el monitoreo del subconjunto con 30 % de anomalías. La fase de calibración permite ajustar los límites dinámicos a partir de datos válidos. Tras esta etapa, el sistema mantiene límites estables, ya que las estadísticas se actualizan solo con valores no anómalos. La señal se mantiene mayormente estable, con lecturas válidas entre 6.5 y 7.5 unidades. No obstante, se identifican múltiples valores atípicos, tanto positivos como negativos, que fueron descartados y corregidos mediante imputación por Última Observación Llevada Adelante.

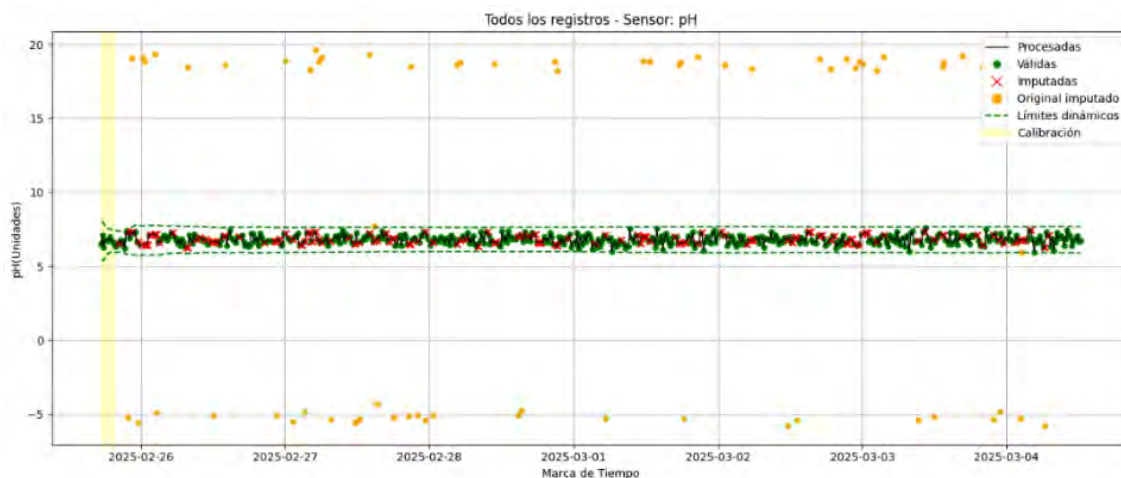


Figura 37. Serie completa del sensor de pH con 30% de anomalías. El sistema mantiene la estabilidad de la señal y actúa eficientemente ante múltiples perturbaciones dispersas. *Fuente: Elaboración propia.*

5.2.3. Subconjunto de datos con 50% de anomalías

En el conjunto con 50 % de anomalías se inyectaron 326 valores atípicos, 327 duplicados y 326 datos faltantes, sumando 979 anomalías. El algoritmo mostró buen desempeño en la eliminación de duplicados (99.69 %) y en la imputación de datos faltantes y atípicos (583 registros, 89.45 % de 652 casos imputables). En detección de valores atípicos, se identificaron correctamente 261 de los 326 (80.06 %), pero se clasificaron erróneamente 13 datos válidos (4.74 % de falsos positivos), y 65 atípicos no fueron detectados (19.94 % de falsos negativos), principalmente en temperatura y humedad. La Figura 38 muestra las primeras 50 lecturas de humedad, con límites estabilizados tras la calibración y una alta densidad de imputaciones asociadas a secuencias anómalas que superan el 50 % y exceden los umbrales del sistema.

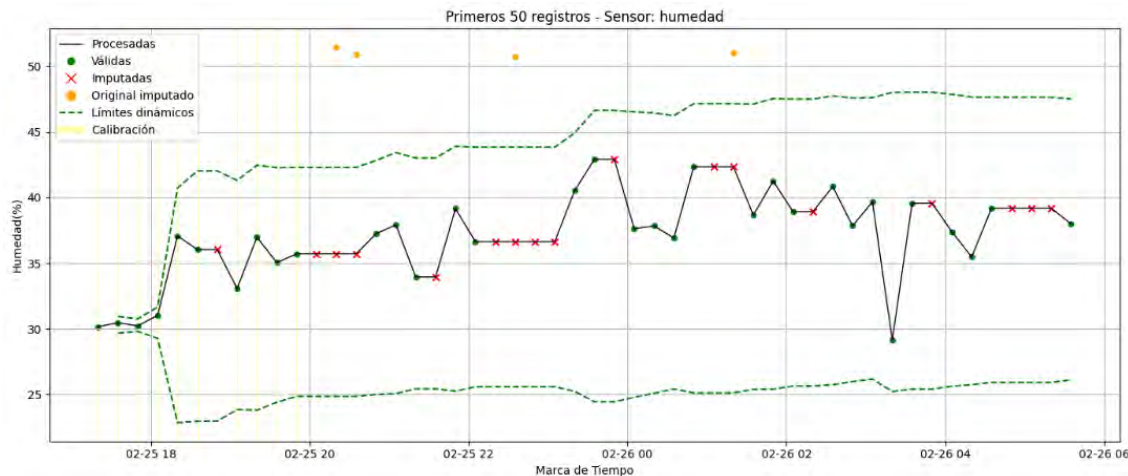


Figura 38. Primeras 50 lecturas del sensor de humedad con 50% de anomalías. El sistema actúa correctamente ante valores atípicos aislados y secuenciales, manteniendo estabilidad estadística.
Fuente: Elaboración propia.

La Figura 39 muestra la evolución completa de las lecturas de humedad en el subconjunto con 50 % de anomalías. Tras la fase de calibración, el sistema ajusta y estabiliza los límites dinámicos, que se mantienen consistentes durante todo el monitoreo. A partir de ese punto, se observa una alta frecuencia de valores atípicos, tanto aislados como en secuencia, lo que se refleja en numerosas imputaciones a lo largo de la serie. Las lecturas descartadas incluyen registros extremos, con valores por encima de 50 % o por debajo de 20 %, claramente fuera del rango esperado en condiciones normales de campo.

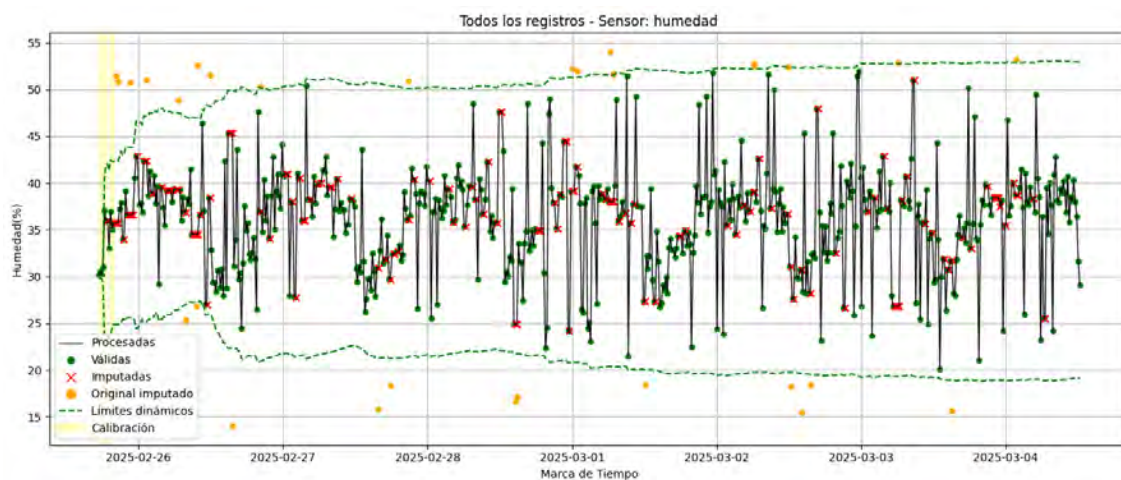


Figura 39. Serie completa del sensor de humedad con 50% de anomalías. A pesar del nivel elevado de perturbación, el sistema mantiene integridad estadística y continuidad en la señal.
Fuente: Elaboración propia.

La Figura 40 muestra las primeras 50 lecturas de temperatura en el subconjunto con 50 % de anomalías. La fase de calibración permite ajustar los límites dinámicos, que se estabilizan tras recibir lecturas válidas. Las lecturas descartadas superan los 28 °C o caen por debajo de 10 °C, fuera del comportamiento esperado. También se observan dos picos abruptos cercanos a 28–29 °C; uno fue clasificado como válido por mantenerse dentro de los límites, y el otro fue correctamente imputado al superar el umbral.

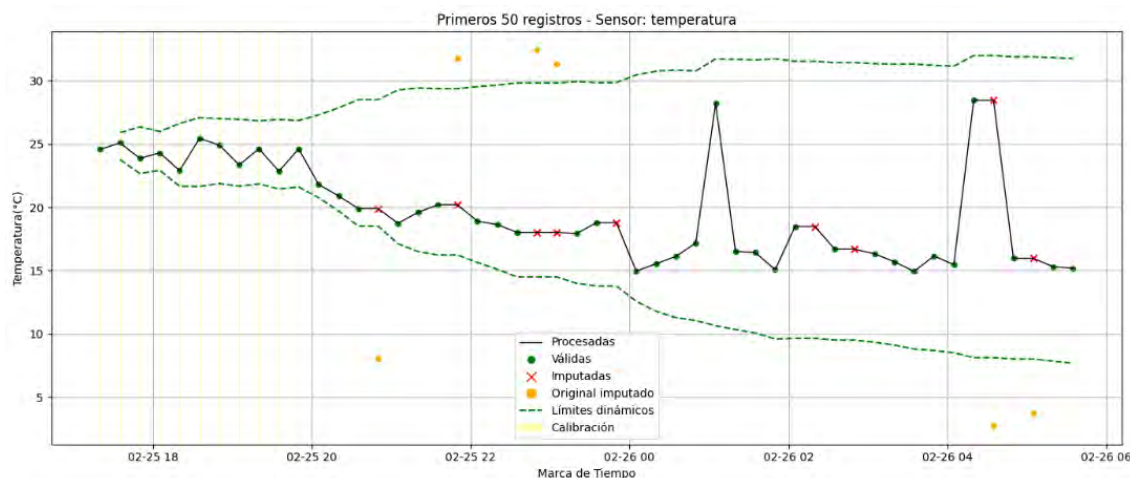


Figura 40. Primeras 50 lecturas del sensor de temperatura con 50% de anomalías. El algoritmo mantiene la estabilidad estadística pese a la inyección de múltiples perturbaciones. *Fuente: Elaboración propia.*

La Figura 41 muestra la serie completa de temperatura en el subconjunto con 50 % de anomalías. Tras la fase de calibración, el sistema estabiliza los límites dinámicos y entra en operación continua. La señal presenta ciclos térmicos regulares, coherentes con la variación día-noche en el entorno subterráneo. No obstante, se observan múltiples valores anómalos inyectados, manifestados como saltos extremos por encima de 35 °C o por debajo de 5 °C.

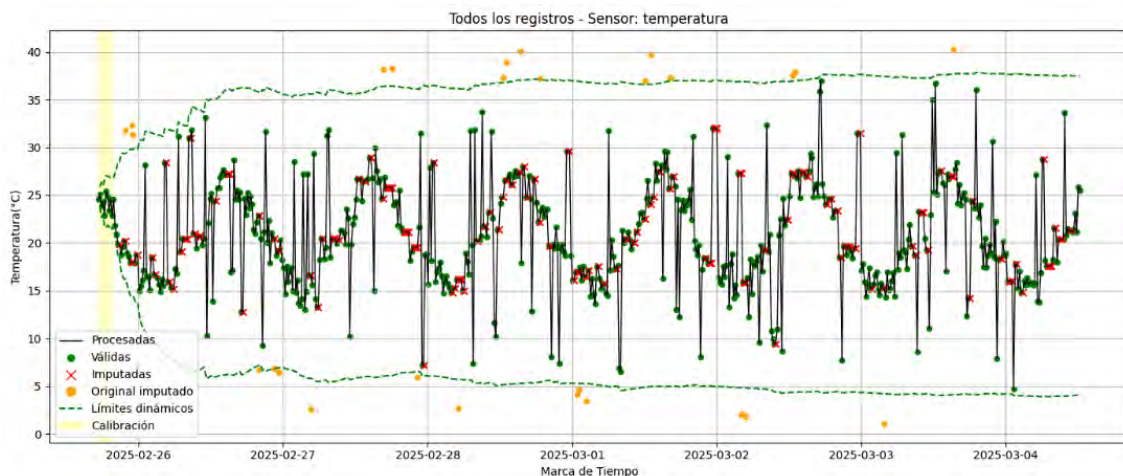


Figura 41. Serie completa del sensor de temperatura con 50% de anomalías. El sistema conserva la forma de la señal y detecta eficazmente anomalías severas distribuidas en todo el periodo.
Fuente: *Elaboración propia.*

La Figura 42 muestra las primeras 50 lecturas del sensor de pH en el subconjunto con 50 % de anomalías. Durante la fase de calibración, el sistema ajusta los límites dinámicos con datos válidos, estableciendo umbrales adaptativos. Finalizada esta etapa, la señal evidencia múltiples perturbaciones, con varias imputaciones consecutivas originadas por lecturas descartadas que superaron los umbrales, incluyendo valores cercanos a 19–20 o inferiores a –6, ambos fuera del rango plausible para pH en suelos.

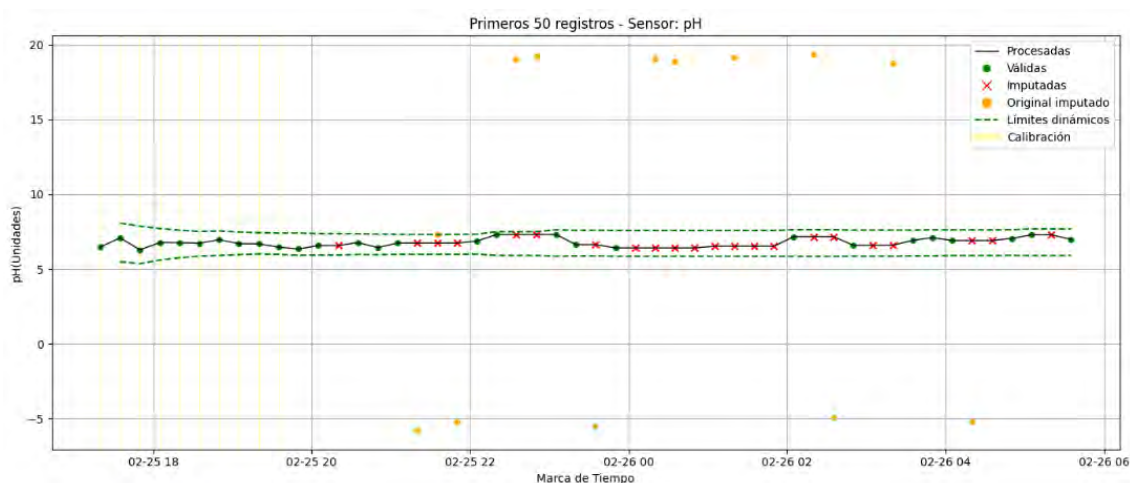


Figura 42. Primeras 50 lecturas del sensor de pH con 50% de anomalías. El sistema responde eficazmente ante la aparición de valores atípicos extremos, manteniendo la estabilidad estadística. Fuente: *Elaboración propia.*

La Figura 43 muestra la serie completa de pH en el subconjunto con 50 % de anomalías. Tras la fase de calibración, el sistema ajusta los límites dinámicos con datos válidos, logrando una estabilización sostenida. Durante la operación continua, se observa una alta densidad de valores atípicos distribuidos de forma homogénea. Estas lecturas, marcadas como imputaciones, fueron correctamente descartadas por superar los umbrales, con magnitudes físicamente inviables, representadas como puntos naranjas.

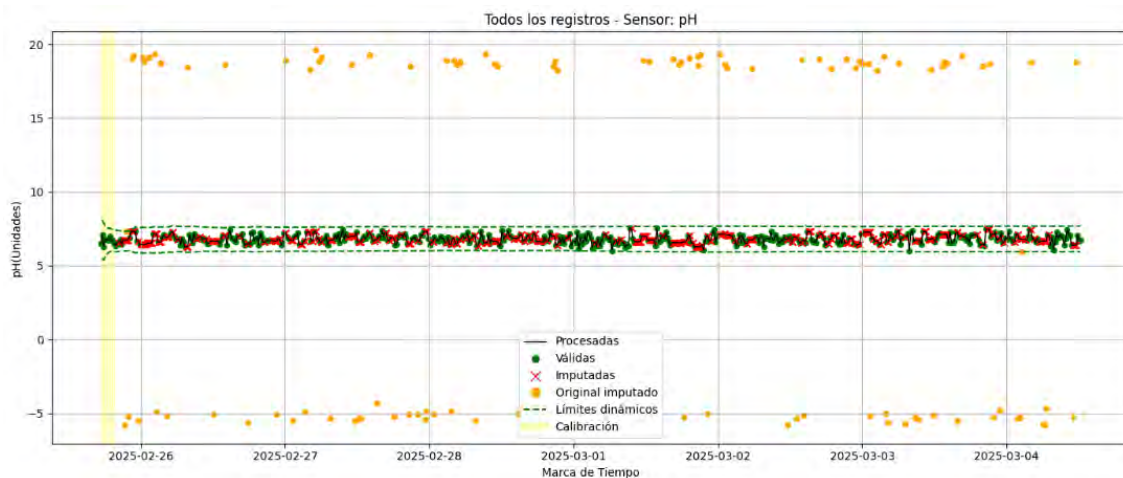


Figura 43. Serie completa del sensor de pH con 50% de anomalías. El sistema mantiene estabilidad estadística y continuidad en la serie, incluso bajo perturbaciones severas. *Fuente: Elaboración propia.*

6

Conclusiones y trabajos futuros

El sistema de limpieza de datos propuesto en esta investigación demostró ser una solución funcional y eficiente para entornos subterráneos con restricciones computacionales, al permitir la depuración autónoma de datos en flujo directamente en nodos embebidos. La validación experimental, realizada mediante la inyección controlada de anomalías en tres subconjuntos de datos con niveles crecientes de perturbación (10 %, 30 % y 50 %), permitió evaluar rigurosamente el comportamiento del algoritmo en condiciones adversas, confirmando su capacidad para operar con precisión y adaptabilidad.

En el subconjunto con 10 % de anomalías, el sistema logró detectar el 92.31 % de los valores atípicos, eliminó el 100 % de los registros duplicados e imputó el 96.18 % de los datos faltantes y atípicos. En condiciones más exigentes, como en el conjunto con 30 % de anomalías, el desempeño del sistema se mantuvo sólido, con una tasa de detección de outliers del 81.63 %, una imputación del 90.82 % de los valores posibles y una eliminación total de duplicados. Finalmente, en el escenario más complejo (50 % de anomalías), se conservaron niveles aceptables de eficacia, con una detección del 80.06 % de los datos atípicos, una imputación del 89.45 % y una tasa de eliminación de duplicados del 99.69 %. En todos los casos, los falsos positivos se mantuvieron por debajo del 5 %, mientras que la mayoría de los falsos negativos se concentraron en las variables de temperatura y humedad, evidenciando áreas específicas de mejora.

Estos resultados, junto con los análisis cualitativos de las gráficas por variable y por subconjunto, confirmaron que el algoritmo fue capaz de ajustar sus umbrales dinámicos

tras la fase de calibración, imputar adecuadamente datos mediante Última Observación Válida, y preservar la integridad de las series temporales a lo largo de todo el periodo evaluado. La arquitectura del sistema permitió un procesamiento completamente autónomo, sin necesidad de infraestructura externa ni supervisión continua, lo que representa una contribución significativa para el monitoreo confiable de condiciones edáficas en Agricultura de Precisión.

Si bien los resultados fueron satisfactorios, existen oportunidades claras para mejorar la precisión del sistema y ampliar su aplicabilidad. Un primer eje de trabajo consiste en fortalecer la detección de anomalías en variables como temperatura y humedad, donde se concentraron la mayoría de los falsos negativos. Para ello, se propone la incorporación de métodos híbridos que combinen estadísticas en línea con técnicas ligeras de aprendizaje automático, lo cual podría mejorar la sensibilidad del sistema frente a patrones no lineales o transiciones bruscas.

También se sugiere optimizar el consumo energético del nodo mediante esquemas de hibernación inteligente y priorización de eventos, así como explorar configuraciones multinodo que permitan validar lecturas de forma colaborativa y descentralizada. La integración de nuevas variables del suelo, como conductividad eléctrica o salinidad, ampliaría el alcance del sistema, permitiendo diagnósticos edáficos más completos.

Por último, se contempla el desarrollo de una interfaz remota para visualización en tiempo real del estado del nodo y los datos limpios generados, con el fin de facilitar su adopción en entornos agrícolas operativos. Estas mejoras permitirán avanzar hacia una plataforma robusta, escalable y aplicable a distintos escenarios agroambientales, consolidando así el papel del Internet de las Cosas Subterráneas en el futuro de la agricultura inteligente.

6.1. Productos derivados de la investigación

Como resultado del desarrollo de esta investigación, se generaron diversos productos académicos que reflejan la contribución científica, técnica y formativa del

trabajo realizado. Estos productos incluyen artículos científicos, participaciones en congresos, presentaciones en instituciones académicas y una estancia de investigación internacional:

Artículos científicos

- Sánchez-Hernández, J. V., Estrada-Esquivel, H., Martínez-Rebollar, A., Pech-May, F. (2025). *Marco de evaluación para la óptima selección de componentes de hardware para sistemas del Internet de las Cosas*. **Revista Computación y Sistemas**.
- Sánchez-Hernández, J. V., Estrada-Esquivel, H., Martínez-Rebollar, A., Pech-May, F. (2024). *Development of a systematic framework for the selection of development boards and sensors for Internet of Things systems*. **Revista Tecnología y Ciencia Aplicada**, 7(2), 689-694.
- Islas-Ávila, A. L., Sánchez-Hernández, J. V., González-Águila, M., Bueno, E., Colindres-Bahena, A., & Hernández, Y. (2024). *Evaluación del rendimiento académico estudiantil con aprendizaje automático: Un caso de estudio de Chipre*. **Encuentro Nacional de Computación, Eureka ENC 2024**.

Póster académico

- Sánchez-Hernández, J. V., Estrada-Esquivel, H., Martínez-Rebollar, A., Pech-May, F. (2023). *Desarrollo de un Sistema Autónomo de Preprocesamiento de Datos para Dispositivos del Internet de las Cosas Subterráneas* [Póster]. **1er Congreso Internacional de Tecnología y Ciencia Aplicada**, Cuernavaca, Morelos.

Estancia de investigación

- Universidad de Magallanes, Punta Arenas, Chile. Departamento de Computación. **Del 22 de marzo al 4 de mayo de 2025**.

Referencias

- Abedjan, Z., Chu, X., Deng, D., Castro-Fernández, R., Ilyas, I., Ouzzani, M., . . . Tang, N. (2016). Detecting Data Wrrors: Where are We and What Needs to be Done? *Proceedings of the VLDB Endowment*, 9(12), 993 - 1004. <https://doi.org/10.14778/2994509.2994518>
- Abidin, Z., Falah, R., Setyawan, R., & Wardana, F. (2025). Wireless Sensor Network Using nRF24L01+ for Precision Agriculture. *Bulletin of Electrical Engineering and Informatics*, 14(2), 1003 - 1013. <https://doi.org/10.11591/eei.v14i2.8481>
- Akyildiz, I., & Stuntebeck, E. (2006). Wireless Underground Sensor Networks: Research Challenges. *Ad Hoc Networks*, 4(6), 669 - 686. <https://doi.org/10.1016/j.adhoc.2006.04.003>
- Alahmad, T., Neményi, M., & Nyéki, A. (2023). Applying IoT Sensors and Big Data to Improve Precision Crop Production: A Review. *Agronomy*, 13(10), 2603 - 2633. <https://doi.org/10.3390/agronomy13102603>
- Alotaibi, O., Pardede, E., & Tomy, S. (2023). Cleaning Big Data Streams: A Systematic Literature Review. *Technologies*, 11(4), 101 - 121. <https://doi.org/10.3390/technologies11040101>
- Anisi, M., Abdul-Salaam, G., & Abdullah, A. (2015). A Survey of Wireless Sensor Network Approaches and their Energy Consumption for Monitoring Farm Fields in Precision Agriculture. *Precision Agriculture*, 2015(1), 216 - 238. <https://doi.org/10.1007/s11119-014-9371-8>
- Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15), 2787 - 2805. <https://doi.org/10.1016/j.comnet.2010.05.010>
- Babu, G., Gokuldhev, M., & Brahmanandam, P. (2024). Integrating IoT for Soil Monitoring and Hybrid Machine Learning in Predicting Tomato Crop Disease in a Typical South India Station. *Sensors*, 24(19), 6177 - 6189. <https://doi.org/10.3390/s24196177>

- Bagha, H., Yavari, A., & Georgakopoulos, D. (2022). Hybrid Sensing Platform for IoT-Based Precision Agriculture. *Future Internet*, 14(8), 233 - 244. <https://doi.org/10.3390/fi14080233>
- Balivada, S., Grant, G., Zhang, X., Ghosh, M., Guha, S., & Matamala, R. (2022). A Wireless Underground Sensor Network Field Pilot for Agriculture and Ecology: Soil Moisture Mapping Using Signal Attenuation. *Sensors*, 22(10), 3913 - 3925. <https://doi.org/10.3390/s22103913>
- Boebel, M., Frei, F., Blumensaat, F., Ebi, C., Meli, M., & Rust, A. (2023). Batteryless Sensor Devices for Underground Infrastructure—A Long-Term Experiment on Urban Water Pipes. *Journal of Low Power Electronic Applications*, 13(2), 31 - 45. <https://doi.org/10.3390/jlpea13020031>
- Cariou, C., Moiroux-Arvis, L., Pinet, F., & Chanet, J. (2023). Internet of Underground Things in Agriculture 4.0: Challenges, Applications and Perspectives. *Sensors*, 23(8), 4058 - 4078. <https://doi.org/10.3390/s23084058>
- Dutta, L., Bharali, S., Barman, P., & Singh, A. (2024). An IoT-Enabled Smart pH Monitoring and Dispensing System for Precision Agriculture Application. *Agricultural Research*, 1(2024), 309 - 318. <https://doi.org/10.1007/s40003-024-00702-0>
- Erl, T., Khattak, W., & Buhler, P. (2016). *Big Data Fundamentals: Concepts, Drivers & Techniques* (Vol. 1). Indiana: Pearson. Obtenido de <https://dl.icdst.org/pdfs/files3/053e44f660b0c2f405e42ac1f8f1a408.pdf>
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions. *Future Generation Computer Systems*, 29(7), 1645 - 1660. <https://doi.org/10.1016/j.future.2013.01.010>
- Guo, H., & Kamrath, A. (2024). Battery-Free Sensor Array for Wireless Multi-Depth In-Situ Sensing. *ICC 2024-IEEE International Conference on Communications*, 1(2024), 4173 - 4178. <https://doi.org/10.48550/arXiv.2401.11479>

- Gupta, G., & Pal, S. (2025). Applications of AI in Precision Agriculture. *Discover Agriculture*, 3(61), 1-15. <https://doi.org/10.1007/s44279-025-00220-9>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (1 ed.). Massachusetts: Morgan Kaufmann. Obtenido de <https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques>
- Herring, B., Sharp, T., Roberts, T., Fastier-Wooler, J., Kelly, G., Sahin, O., . . . Woodfield, P. (2022). Underground LoRa Sensor Node for Bushfire Monitoring. *Fire Technology*, 58(1), 1087 - 1095. <https://doi.org/10.1007/s10694-022-01224-3>
- Levintal, E., Ganot, Y., Taylor, G., Freer-Smith, P., Suvocarev, K., & Dahlke, H. (2022). An Underground, Wireless, Open-Source, Low-Cost System for Monitoring Oxygen, Temperature, and Soil Moisture. *Soil*, 8(1), 85 - 97. <https://doi.org/10.5194/soil-8-85-2022>
- Madakam, S., Ramaswamy, R., & Tripathi, S. (2015). Internet of Things (IoT): A Literature Review. *Journal of Computer and Communications*, 3(1), 164 - 173. <https://www.doi.org/10.4236/jcc.2015.35021>
- Melesse, T. (2025). Digital Twin-Based Applications in Crop Monitoring. *Heliyon*, 11(2), 1 - 10. <https://doi.org/10.1016/j.heliyon.2025.e42137>
- Meriç, M. (2025). Implementation of a Wireless Sensor Network for Irrigation Management in Drip Irrigation Systems. *Scientific Reports*, 2025(1), 14157 - 14168. <https://doi.org/10.1038/s41598-025-97303-w>
- Payero, J., Nafchi, A., David, R., & Khalilian, A. (2017). An Arduino-Based Wireless Sensor Network for Soil Moisture Monitoring Using Decagon EC-5 Sensors. *Open Journal of Soil Science*, 7(10), 10 - 22. <https://doi.org/10.4236/ojss.2017.710021>
- Ray, P. (2018). A Survey on Internet of Things Architectures. *Journal of King Saud University - Computer and Information Sciences*, 30(3), 291 - 319. <https://doi.org/10.1016/j.jksuci.2016.10.003>

- Raza, A., Keshavarz, R., & Shariati, N. (2024). Precision Agriculture: Ultra-Compact Sensor and Reconfigurable Antenna for Joint Sensing and Communication. *IEEE Transactions on Instrumentation and Measurement*, 73(1), 1 - 13. <http://doi.org/10.1109/TIM.2024.3350126>
- Salam, A., & Raza, U. (2020). A Guide for Sensor Selection. En A. Salam, & U. Raza, *Signals in the Soil: Developments in Internet of Underground Things* (1 ed., págs. 286 - 287). Switzerland: Springer. <https://doi.org/10.1007/978-3-030-50861-6>
- Salam, A., & Raza, U. (2020). Current Advances in Internet of Underground Things. En A. Salam, & U. Raza, *Signals in the Soil: Developments in Internet of Underground Things* (Vol. 1, págs. 321 - 356). Suiza: Springer. https://doi.org/10.1007/978-3-030-50861-6_10
- Salam, A., & Raza, U. (2020). IOUT Applications in Decision Agriculture, Real-Time Soil Monitoring, and Irrigation Automation. En A. Salam, & U. Raza, *Signals in the Soil: Developments in Internet of Underground Things* (Vol. 1, págs. 321 - 356). Switzerland: Springer Cham. https://www.doi.org/10.1007/978-3-030-50861-6_10
- Sambo, D., Forster, A., Yenke, B., Sarr, I., Gueye, B., & Dayang, P. (2020). Wireless Underground Sensor Networks Path Loss Model for Precision Agriculture (WUSN-PLM). *IEEE Sensors Journal*, 20(10), 5298 - 5313. <https://doi.org/10.1109/JSEN.2020.2968351>
- Scalambrin, L., Zanella, A., & Vilajosana, X. (2023). LoRa Multi-Hop Networks for Monitoring Underground Mining Environments. *2023 IEEE Globecom Workshops (GC Wkshps)*, 1(2023), 696 - 701. <https://doi.org/10.1109/GCWkshps58843.2023.10464954>
- Sfar, A., Natalizio, E., Challal, Y., & Chtourou, Z. (2018). A Roadmap for Security Challenges in the Internet of Things. *Digital Communications and Networks*, 4(2), 118 - 137. <https://doi.org/10.1016/j.dcan.2017.04.003>

- Shahab, H., Naeem, M., Iqbal, M., Aqeel, M., & Ullah, S. (2025). IoT-Driven Smart Agricultural Technology for Real-Time Soil and Crop Optimization. *Smart Agricultural Technology*, 10(2025), 1 - 9. <https://doi.org/10.1016/j.atech.2025.100847>
- Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge Computing: Vision and Challenges. *IEEE Internet of Things Journal*, 3(5), 637 - 646. <https://www.doi.org/10.1109/JIOT.2016.2579198>
- Silva, A., & Vuran, M. (2010). Development of a Testbed for Wireless Underground Sensor Networks. *EURASIP Journal on Wireless Communications and Networking*, 2010(1), 620307 - 620327. <https://doi.org/10.1155/2010/620307>
- Singh, G., Singh, J., Juneja, S., Gulzar, Y., Gupta, D., Ghafoor, K., & Kumar, M. (2025). Empowering Self-Sustained Agriculture: An IoT Field Tracking Model Driven by Fog Computing. *Discover Internet of Things*, 2025(1), 1 - 13. <https://doi.org/10.1007/s43926-025-00140-6>
- Singh, H., Halder, N., Singh, B., Singh, J., Sharma, S., & Shacham-Diamand, Y. (2023). Smart Farming Revolution: Portable and Real-Time Soil Nitrogen and Phosphorus Monitoring for Sustainable Agriculture. *Sensors*, 23(23), 5914 - 5928. <https://doi.org/10.3390/s23135914>
- Soussi, A., Zero, E., Sacile, R., Trincherio, D., & Fossa, M. (2024). Smart Sensors and Smart Data for Precision Agriculture: A Review. *Sensors*, 24(8), 2647 - 2660. <https://doi.org/10.3390/s24082647>
- Wang, K., Wu, S., Ivoghlian, A., Salcic, Z., Austin, A., & Zhou, X. (2019). LWS: A LoRaWAN Wireless Underground Sensor Network Simulator for Agriculture Applications. *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 1(2019), 475 - 482. <https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00123>

- Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. (2017). Big Data in Smart Farming – A review. *Agricultural Systems*, 153(1), 69 - 80. <https://doi.org/10.1016/j.agry.2017.01.023>
- Zhang, X., Andreyev, A., Zumpf, C., Negri, M., Guha, S., & Ghosh, M. (2017). Thoreau: A Subterranean Wireless Sensing Network for Agriculture and The Environment. *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2017(1), 74 - 84. <https://doi.org/10.1109/INFOCOMW.2017.8116356>
- Zhang, X., Feng, G., & Sun, X. (2024). Advanced Technologies of Soil Moisture Monitoring in Precision Agriculture: A Review. *Journal of Agriculture and Food Research*, 18(2024), 101473 - 101485. <https://doi.org/10.1016/j.jafr.2024.101473>