UNIVERSIDAD NACIONAL DEL LITORAL

Hacia un Nuevo Paradigma en la Adquisición y Gestión del Dato Agro-Hidrológico

Emiliano Pedro López

FICH
FACULTAD DE INGENIERÍA Y CIENCIAS HÍDRICAS

INTEC
INSTITUTO DE DESARROLLO TECNOLÓGICO PARA LA INDUSTRIA QUÍMICA

CIMEC
CENTRO DE INVESTIGACIÓN DE MÉTODOS COMPUTACIONALES

sinc(*i*)
INSTITUTO DE INVESTIGACIÓN EN SEÑALES, SISTEMAS E INTELIGENCIA COMPUTACIONAL

UNIVERSIDAD NACIONAL DEL LITORAL
Facultad de Ingeniería y Ciencias Hídricas

Hacia un Nuevo Paradigma en la Adquisición y Gestión del Dato Agro-Hidrológico

Emiliano Pedro López

Tesis remitida al Comité Académico del Doctorado
como parte de los requisitos para la obtención
del grado de
DOCTOR EN INGENIERÍA
Mención Recursos Hídricos
de la
UNIVERSIDAD NACIONAL DEL LITORAL

2025

Comisión de Posgrado, Facultad de Ingeniería y Ciencias Hídricas, Ciudad Universitaria, Paraje El Pozo, S3000, Santa Fe, Argentina.

UNIVERSIDAD NACIONAL DEL LITORAL
Facultad de Ingeniería y Ciencias Hídricas
Instituto de Desarrollo Tecnológico para la Industria Química
Centro de Investigación de Métodos Computacionales
Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional

Hacia un Nuevo Paradigma en la Adquisición y Gestión del Dato Agro-Hidrológico

Emiliano Pedro López

Lugar de Trabajo
CEFHAL
Centro de Estudios Fluviales e Hidro-Ambientales del Litoral
Facultad de Ingeniería y Ciencias Hídricas (FICH)
Universidad Nacional del Litoral (UNL)

Director:
Carlos Alberto Vionnet CEFHAL. FICH - UNL / CONICET

Co-director:
Ricardo Germán Dunger Facultad de Ciencias Agrarias - UNL / CONICET

Jurado Evaluador:
Esteban Jobbágy Universidad Nacional de San Luis / CONICET, Argentina.
Raúl Giménez South Australian Research & Development Institute, Australia.
Alejo Bonifacio Universidad Nacional de Córdoba / CONICET, Argentina.

2025

DECLARACIÓN LEGAL

Declaro ser el autor de la presente tesis.

AGRADECIMIENTOS

Esta tesis es el resultado de la generosidad y colaboración de muchas personas. En primer lugar quiero agradecer a Carlos V. por haberme incentivado desde el primer momento a trabajar en investigación. Sin su confianza y predisposición no hubiese transitado este camino.

Dos personas fundamentales fueron Guillermo C. y Jorge P. La gran mayoría del trabajo fue a su lado, los viajes a campo, la mejora permanente de los dispositivos, la instalación de sensores y un largo etc. Gracias por haber trabajado a la par, por la gran dedicación y sobre todo por su amistad.

Quisiera también agradecer a Germán D. y a José M. de la FCA - UNL por su predisposición para asesorarme desinteresadamente y evacuar mis dudas referidas a los cultivos de alfalfa.

Un agradecimiento especial al grupo del SANS - UPC, en especial a Jorge G. V., José María B. O. por su cálido recibimiento y generosidad, cada trabajo conjunto fue un gran aprendizaje y satisfacción.

A mis colegas y amigos del CEFHAL - FICH - UNL por todos estos años de trabajo conjunto en un espacio ameno y colaborativo, en especial a Francisco L. y Lucas D.

Una dedicatoria especial a Diana Alberto, por haberme acompañado con amor tanto en las alegrías como en las frustraciones, en muchos viajes a campo, secando de la alfalfa, con la granulometría y un largo etcétera. Esta tesis es para vos.

Además, quisiera enaltecer el aporte de los jurados Esteban J., Raúl G. y Alejo B. por sus recomendaciones para mejorar el trabajo realizado.

Agradezco también a los estudiantes que de un modo indirecto aportaron a esta tesis realizando sus proyectos finales de carrera en temas afines, en especial a Mariana P.E., Dario V. y Renato M.

Por último, quisiera agradecer a la Universidad Pública Argentina y a todos los docentes e investigadores de excelencia que la forman, sobre todo a quienes luchan por ella.

Índice general

Índice de figuras

Índice de tablas

RESUMEN

Esta tesis pretende mitigar, en parte, la creciente necesidad que existe por mejorar la eficiencia en la gestión de recursos hídricos y agrícolas en un contexto de cambio climático y escasez de agua. El tratamiento de datos es el proceso que ocurre entre la recolección y su posterior traducción en información utilizable. La problemática central radica en la carencia de sistemas de monitoreo precisos y asequibles que puedan proporcionar datos confiables sobre las condiciones hidrológicas y su vinculación a procesos íntimamente relacionados como el crecimiento y rendimiento de cultivos. Para abordar esta problemática, se desarrollaron dispositivos IoT (del inglés, Internet of Things) de bajo costo para la medición automática de variables hidrológicas, ambientales y biológicas. Los dispositivos, basados en plataformas de hardware libre / open-source, permitieron la recopilación de información precisa sobre la dinámica hidrológica y la medición de la evolución del crecimiento de cultivos en la región centro de la Provincia de Santa Fe. Por tanto, la tesis propone una transformación en la forma en que se adquiere y se agrega valor a los datos agro-hidrológicos. El estudio incluye el diseño y desarrollo de hardware y software, la calibración y validación de sensores utilizando técnicas de aprendizaje automático, y la implementación de sistemas de monitoreo agro-hidrológicos. Además, se emplearon modelos de crecimiento y rendimiento de cultivos basados en datos recolectados por estos dispositivos, demostrando su eficacia en la mejora de la gestión de recursos hídricos y su impacto en la producción agrícola. El desarrollo ha mostrado resultados alentadores para su uso como herramienta de soporte en la toma de decisiones. La investigación concluye con recomendaciones para la implementación y mejora continua de estos sistemas, subrayando su potencial impacto positivo en la agricultura y la gestión del agua.

ABSTRACT

This thesis seeks to partially mitigate the growing need to enhance efficiency in managing water and agricultural resources amid climate change and water scarcity. Data processing is a key activity that bridges the gap between data collection and its transformation into actionable information. The central challenge stems from the absence of precise, affordable monitoring systems capable of delivering reliable data on hydrological conditions and their connections to closely related processes, such as crop growth and yield. To address this issue, low-cost IoT (Internet of Things) devices were developed for the automated measurement of hydrological, environmental, and biological variables. Based on open-source hardware platforms, these devices facilitated the accurate collection of data on hydrological dynamics and tracked crop growth trends in the central region of Santa Fe Province. Consequently, this thesis proposes a paradigm shift in how agro-hydrological data is gathered and adds value. The study encompasses the design and development of hardware and software, sensor calibration and validation using machine learning techniques, and the instrumentation of agro-hydrological monitoring systems. Furthermore, crop growth and yield models, driven by data from these devices, were utilized, proving their effectiveness in optimizing water resource management and boosting agricultural productivity. The development has yielded encouraging results, establishing it as a valuable decision-support tool. The research concludes with recommendations for the implementation and continuous improvement of these systems, emphasizing their potential to positively impact agriculture and water management.

Sobre la modalidad de la tesis

Esta tesis doctoral fue presentada bajo la modalidad denominada *tesis por compilación*, conformada por artículos científicos que constituyen el núcleo de la investigación y que fueron enviados y publicados en revistas internacionales. A continuación se presenta una descripción técnica general donde se resume el trabajo realizado a fin de proporcionar al lector un hilo conductor que vincule la problemática general con las investigaciones presentadas en cada artículo.

1 Introducción

1.1 La importancia de los datos hidro-ambientales en la gestión hídrica

En el campo de las ciencias ambientales y en particular en el estudio de los procesos hidrológicos y sus aplicaciones en la agronomía, contar con mediciones precisas resulta fundamental. Las observaciones de largo plazo son la base de la hidrología como ciencia; permiten establecer referencias, observar tendencias y ciclos naturales; son insumos básicos para evaluar modelos predictivos y dan soporte en la toma de decisiones para la gestión del agua y el medio ambiente (*Tetzlaff et al.*, 2017).

La disponibilidad de datos continuos con una adecuada resolución espacial es esencial para comprender y modelar de manera precisa la dinámica hídrica de una región. Por tanto, para gestionar adecuadamente los recursos hídricos es un requisito indispensable conocer qué y dónde medir, la cantidad y calidad de la información, con qué frecuencia y por cuánto tiempo. El verdadero valor de los datos se reconoce cuando se destinan a producir servicios y productos que puedan generar beneficios socioeconómicos y ambientales. Más aún, el acceso sin restricciones a la información facilita la innovación y el descubrimiento de nuevos usos (*Stewart*, 2015) (Figura 1.1).

Algunas de las series de datos más antiguas a nivel mundial se encuentran en el Central England Temperature Data Series[1], sus mediciones de temperatura ambiente se remontan a 1659 convirtiéndolo en uno de los registros meteorológicos más extensos del mundo (*Manley*, 1953). Registros de este tipo son utilizados, por ejemplo, para validar modelos que reconstruyen la temperatura del planeta en los últimos milenios mediante indicadores indirectos (*Mann & Jones*, 2003).

Otro caso excepcional es el vinculado a la fecha de floración de los cerezos, un fenómeno de gran importancia en la cultura japonesa que ha sido documentado durante más de 1200 años. En los últimos años se produjo la fecha de floración más temprana desde que existen registros. *Christidis et al.* (2022) atribuyen este adelanto al avance del cambio climático inducido por el ser humano. Estos ejemplos son una muestra sobre la importancia que tiene mantener series a lo largo del tiempo y su utilidad en otros campos de la ciencia para el que originalmente no fueron contemplados.

En las últimas décadas, la gestión de los recursos hídricos se ha convertido en un desafío para la humanidad. A medida que la población mundial continúa su crecimiento, el suelo cultivable disponible por habitante disminuye constantemente. Al mismo

[1]HadCET: www.metoffice.gov.uk/hadobs/hadcet, *accedida 28-08-2024*

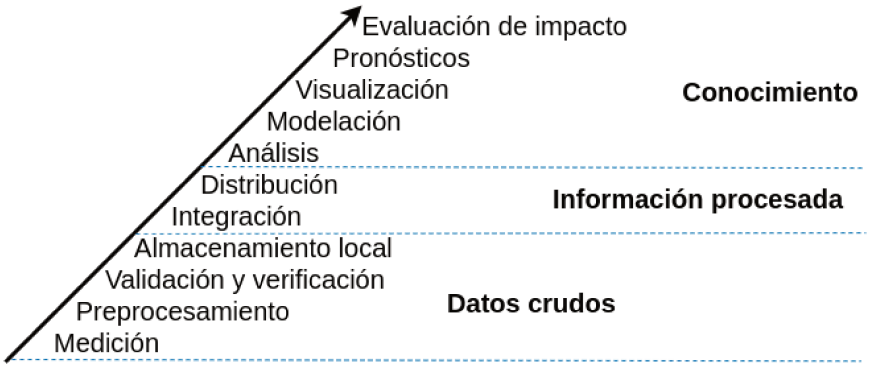


Figura 1.1: Agregado de valor a los datos desde su medición hasta su uso en predicción (basado en *Stewart*, 2015)

tiempo, las actividades antrópicas han alterado los flujos naturales de agua a una escala sin precedentes. La superficie terrestre ha experimentado cambios drásticos debido a la transformación impulsada por el ser humano en el uso del suelo, así como en la gestión de los sistemas de aguas superficiales y subterráneas (*Wada et al.*, 2017). Esta alteración significativa de los procesos hidrológicos naturales ha generado una creciente necesidad de mediciones precisas y continuas que permitan comprender y predecir el impacto de estos cambios, y por tanto, desarrollar estrategias efectivas de gestión y adaptación (*Yang et al.*, 2021).

Conocer la evolución de los recursos hídricos de una región permite evaluar su dinámica histórica y el estado de vulnerabilidad hídrica que presenta. Esto, a su vez, favorece el estudio de la interrelación o impacto en otras áreas de vital importancia e íntimamente vinculadas con el ciclo del agua, como por ejemplo la producción de alimentos. Registrar un alto contenido de agua en la zona no saturada del suelo y elevados niveles del agua subterránea (cercana a la superficie) en una determinada región, implica una dualidad eventual desde la perspectiva de la gestión del recurso hídrico. Por un lado, existe un mayor riesgo de inundaciones (*Zhang et al.*, 2002), y por otro, un rendimiento óptimo de cultivos si la napa freática se ubica en una franja cercana a la superficie (*Nosetto et al.*, 2009; *Kuppel et al.*, 2015; *Houspanossian et al.*, 2023)

La primera red pluviométrica Argentina fue conformada en 1887 en las estaciones de las líneas ferroviarias de los Ferrocarriles del Sur, Central Argentino y Central Norte (*SMN*, 2017). Posteriormente fue desmantelada con la privatización en la década de los 90, a fines del siglo pasado, perdiendo y discontinuando gran parte de esta serie de datos. Series de datos históricas con información agro-meteorológica relevante que nunca ha sido utilizada aún se encuentra almacenada en formato papel con un riesgo elevado de perderse (Figura 1.2).

De acuerdo con *Antico et al.* (2018) esto es un problema común en Sudamérica y son escasos los esfuerzos realizados para preservar datos históricos. *Antico et al.* (2018) realizaron el relevamiento y procesamiento para recuperar registros de los niveles hidrométricos diarios del río Paraná desde 1875 (Figura 1.3). Estos datos de observaciones del nivel del agua son de suma importancia para comprender la variabilidad hidro-climática pasada y presente, predecir cambios futuros y evaluar los resultados de modelos hidrológicos.

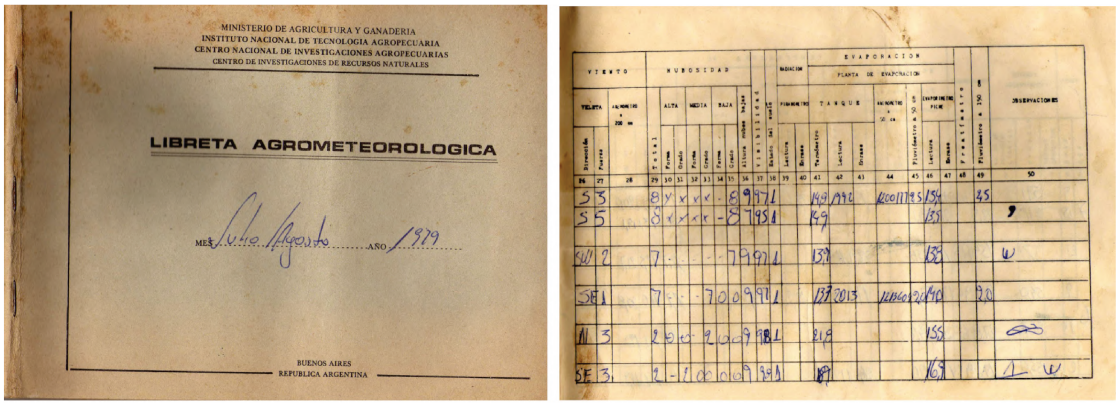


Figura 1.2: Libreta agro-meteorológica de la localidad de Tacuarendí (Prov. Santa Fe) de 1929 del Ministerio de Agricultura y Ganadería. Facilitada por el Centro de Informaciones Meteorológicas (FICH/UNL).

Variables como la precipitación y los niveles freáticos, entre otras, son registradas desde hace años por INTA. En consonancia con esto, la estación de INTA Rafaela (Prov. de Santa Fe) cuenta con más de 50 años de registros continuos de variables de interés para el agro. Series de datos continuas en el tiempo permiten estudiar y vincular el cambio en los patrones históricos, en el uso del suelo, en las prácticas agronómicas o en el régimen de precipitaciones y por tanto inferir sus causas y evaluar alternativas para minimizar su impacto (Figura 1.4).

De lo expuesto, surge entonces la relevancia que tiene *medir* a lo largo del tiempo y en múltiples sitios de observación. Sin embargo, asumir la inversión requerida para la instalación de redes de monitoreo y afrontar el costo asociado a su mantenimiento es un problema que dificulta la instrumentación a gran escala, sobre todo para países en vías de desarrollo. Un caso ilustrativo es el del sistema de alerta hidrológico del río Salado de la Prov. de Santa Fe. Luego de la catastrófica inundación de la ciudad de Santa Fe en el año 2003, el Gobierno Provincial implementó un costoso sistema de monitoreo conformado por 38 estaciones hidrométricas y pluviométricas en la cuenca inferior del río Salado. Con el tiempo se discontinuó el mantenimiento y en la actualidad apenas unas pocas estaciones se encuentran en funcionamiento.

Una alternativa al monitoreo puntual es a través del uso de sensores remotos. Entre las ventajas que presenta se destacan la capacidad de cubrir grandes superficies, la abundancia de datos generalmente de acceso gratuito, la disponibilidad de bibliotecas de software que facilitan su procesamiento, análisis y visualización, y la ausencia de requisitos de mantenimiento constante. Desde hace años existen sensores remotos capaces de inferir la humedad del suelo independientemente de la cobertura vegetal del terreno (*Babaeian et al.*, 2019).

La capacidad de estimar la humedad del suelo a partir de satélites o sensores aéreos es muy atractiva, especialmente en las últimas décadas donde el desarrollo de estas tecnologías ha tenido un aumento significativo (*Notarnicola & Solorza*, 2014). Estos productos utilizan como principal insumo las microondas (*Huang et al.*, 2019; *Arellana et al.*, 2023), por ejemplo las misiones SMAP o SAOCOM[2], y a partir de datos semi-

[2]SAOCOM: www.argentina.gob.ar/ciencia/conae/misiones-espaciales/saocom, *accedida 28-08-2024*

empiricos o modelos físicos calculan la constante dieléctrica del suelo o directamente la humedad del suelo. Por otro lado, la ausencia de observaciones de profundidad de napa freática es un desafío a escala global (*Ma et al.*, 2024), sin embargo, misiones como GRACE y GRACE-FO (del inglés Gravity Recovery and Climate Experiment Follow-On) estiman la variación temporal del campo gravitacional para luego inferir cambios en las masas de aguas subterráneas (*Adams et al.*, 2022; *Milewski et al.*, 2019).

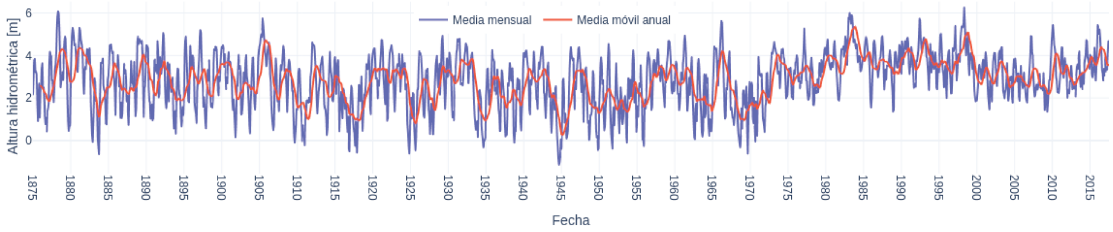


Figura 1.3: Niveles del río Paraná (serie de datos de *Antico et al.* (2018))

La teledetección es una herramienta de importancia insoslayable e independientemente de los beneficios por sobre la tediosa tarea y periplos asociados a medir *in situ* es una tecnología que requiere y se complementa de los datos obtenidos por instrumentos en campo. Cada uno de los casos citados previamente ha requerido datos de campo para calibrar o validar las mediciones remotas. Basta con descargar los datos más recientes de precipitación obtenidos mediante imágenes satelitales para corroborar las discrepancias sustanciales con los valores de campo. Las mediciones *in situ* siguen siendo esenciales y **gran parte de los esfuerzos de esta tesis se han centrado en mejorar la adquisición y calidad de los datos obtenidos en el terreno.**

1.2 Producción alimentaria: Argentina en el contexto mundial

Según proyecciones de la ONU, se espera que la población mundial crezca en aproximadamente 2.000 millones de personas en las próximas tres décadas, elevando el número de habitantes de los actuales 7.700 millones a 9.700 millones para el año 2050. Esta tendencia plantea un desafío significativo, especialmente en el ámbito agrícola. La FAO advierte que la demanda de productos agrícolas podría aumentar entre un 60 % y un 70 % lo que significaría un 35 % más de requerimientos de agua para producirlos, mientras que la capacidad de suministro se ve amenazada por la disminución de la disponibilidad de tierras y los efectos del cambio climático.

En este contexto, surge una imperiosa necesidad de incrementar la eficiencia y la productividad agrícola, logrando producir más con menos recursos ambientales. La agricultura de precisión impulsada por dispositivos IoT ha surgido como una herramienta innovadora para abordar los desafíos actuales que condicionan la sustentabilidad de las actividades agropecuarias.

En este aspecto, la República Argentina es un actor importante en la producción mundial de alimentos (Figura 1.5). Pasó de producir 34 millones de toneladas de granos en 1990 a más de 140 millones en 2019 (*De Abelleyra et al.*, 2019) a partir de combinar la siembra directa (*Peiretti & Dumanski*, 2014), el uso de agroquímicos y semillas genéticamente modificadas (*Qaim & Traxler*, 2005) y la expansión de la frontera agrícola (*Gasparri & Grau*, 2009).

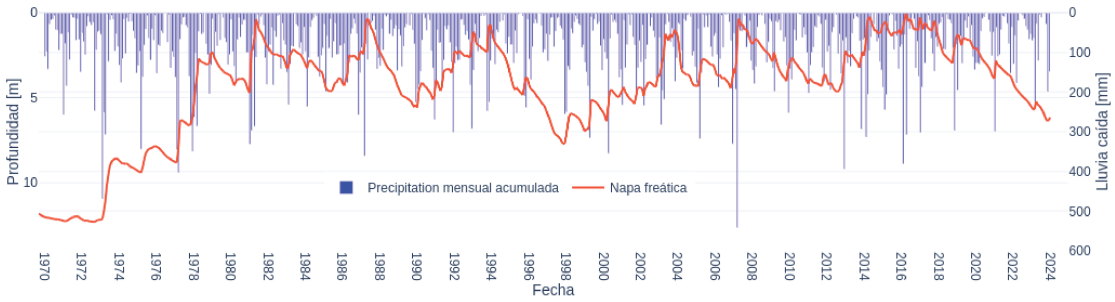


Figura 1.4: Registros de precipitación y niveles freáticos (datos INTA Rafaela)

En lo referido a la producción ganadera, la alfalfa (*Medicago sativa* L.) es la forrajera de mayor relevancia a nivel mundial (*Yang et al.*, 2008; *Noland et al.*, 2018; *Dhakal et al.*, 2020; *Feng et al.*, 2022; *Tucak et al.*, 2023). Esta leguminosa perenne se cultiva en aproximadamente 30-32 millones de hectáreas en todo el mundo (*Acharya et al.*, 2020; *Jáuregui et al.*, 2022), con Estados Unidos representando el 21 % del área total (*Russelle*, 2001; *Fink et al.*, 2022).

Los suelos más ricos de la región pampeana de Argentina históricamente han sido utilizados exclusivamente para la producción agrícola (*Urcola et al.*, 2015), con un cambio en el uso de la tierra desde 2003 (*Gasparri & Grau*, 2009; *Viglizzo et al.*, 2011). A pesar de estos inconvenientes, Argentina sigue siendo uno de los mayores productores de alfalfa del mundo y el primero en Sudamérica con 1.5 millones de hectáreas cultivadas (*Jáuregui et al.*, 2022), superando ampliamente a otros países de la región. El cultivo se concentra en la región Pampeana, donde las condiciones favorecen altos rendimientos en secano, y se expande a zonas como el Noroeste, Cuyo, Patagonia y Chaco (*Basigalup*, 2023).

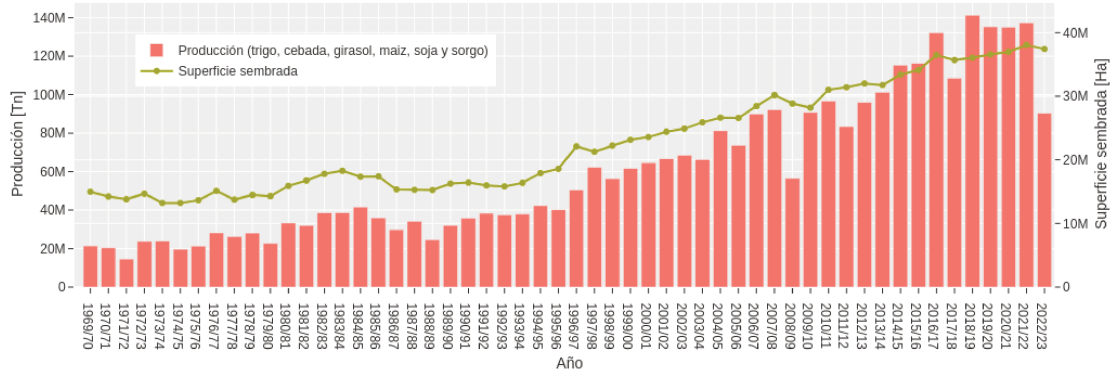


Figura 1.5: Producción de granos en Argentina (datos de la Subsecretaría de Agricultura de la RA.)

1.3 Tecnologías de monitoreo aplicadas al crecimiento de cultivos

En la zona no saturada tienen lugar los mecanismos de infiltración, evaporación, erosión, recarga subterránea, atenuación y una gran diversidad de procesos de transporte y transformación que dan soporte al crecimiento vegetal (*Holden & Fierer*, 2005).

Estos mecanismos, junto a otros, vinculados a las características del suelo, clima, etc., son utilizados como insumo en modelos de crecimiento de cultivos para simular el desarrollo, crecimiento y rendimiento vegetal. Estos modelos, generalmente mecanicistas, contribuyen a un mejor entendimiento sobre diferencias en los rendimientos y posibilitan incrementar la eficiencia en el manejo de los cultivos.

En este sentido, la medición y el análisis de las características fenotípicas de las plantas desempeñan un papel fundamental en el entendimiento de la respuesta a los cambios ambientales. Por ejemplo, la altura de las plantas es un indicador comúnmente reconocido del estado de crecimiento de los cultivos y está directamente relacionada con su rendimiento (*Fricke et al.*, 2011).

En consecuencia, se han desarrollado diversas técnicas basadas en tecnologías LiDAR (del inglés, Laser Image Detection and Ranging) para calcular la morfología vegetal y a partir de ella su biomasa (*Pittman et al.*, 2015; *Wang et al.*, 2017), aunque esto requiera de un elevado costo computacional (Figura 1.6). Otros autores han medido la altura mediante el uso de drones o sensores ultrasónicos. Estos sensores acústicos aún están infrautilizados en la agricultura y pueden ofrecer, a bajo costo, información muy útil a los investigadores y agricultores (*Bitella et al.*, 2024). La integración de sensores ultrasónicos para monitorear la evolución temporal de los cultivos de alfalfa mediante plataformas de hardware libre y su vinculación con modelos de rendimiento fue una de las principales herramientas utilizadas en esta investigación.

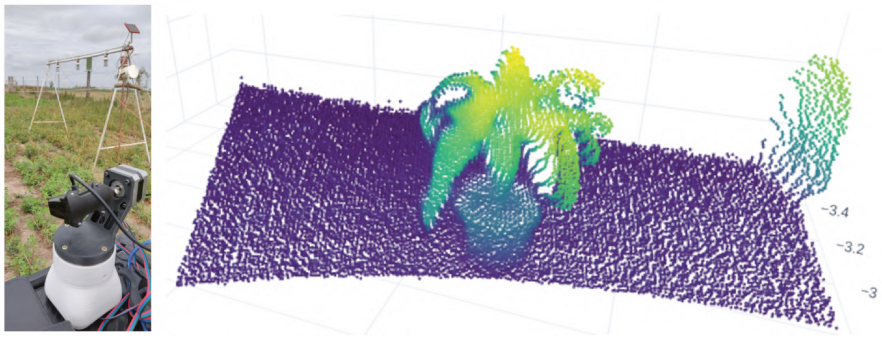


Figura 1.6: Ensayo en campo con un dispositivo LiDAR prototipo (izq.). Nube densa de 51200 puntos como resultado del escaneo en una planta *Dracaena fragans* (nombre común: Palo de agua) con el mismo dispositivo (der.). Experimentos preliminares realizados por quien suscribe la presente tesis.

La sofisticación y complejidad de la simulación del crecimiento y rendimiento de los cultivos han aumentado a un ritmo notable en las últimas décadas (*Paola et al.*, 2016). *Mattera et al.* (2013), *Berhongaray et al.* (2019) y *Otero & Castro* (2019), entre otros, encontraron que la temperatura del aire, la radiación solar, el espaciamiento entre hileras, la profundidad del nivel freático y la disponibilidad de agua en el suelo probablemente estén entre los parámetros más importantes que influyen en el crecimiento de la alfalfa. Por lo tanto, el crecimiento de las plantas es un proceso bioquímico, físico e hidrológico complejo. Registrar este crecimiento junto a las variables hidro-ambientales que inciden en su desarrollo, permitirá mejorar el conocimiento sobre este proceso y el rendimiento de los cultivos.

1.4 El papel del hardware libre y los dispositivos IoT en la medición de variables agro-hidrológicas

El registro automático *in situ* de variables hidro-ambientales o agro-hidrológicas suele realizarse mediante tecnología propietaria, en general, poco flexible y costosa, lo que dificulta contar con series continuas y a una adecuada resolución espacial.

La irrupción de las plataformas de hardware libre, también conocidas como hardware open-source (FOSH, del inglés Free Open Source Hardware), surgen como una alternativa atractiva para instrumentación científica e investigación, debido a su flexibilidad, a la gran cantidad de documentación técnica disponible y a la filosofía subyacente de sus licencias (GPL para el software, Creative Commons para la documentación, CERN Open Hardware License para diseños de hardware, entre otras) que permiten la libre distribución de los aportes y avances realizados por terceros (*Fisher & Gould*, 2012a; *Pearce*, 2012).

De acuerdo con *Pearce* (2020) las tecnologías FOSH proporcionan ahorros económicos en torno al 87 % en comparación con herramientas propietarias equivalentes, y llega al 94 % para aquellas plataformas que utilizaron microcontroladores de tecnología Arduino e impresiones 3D en sus desarrollos.

Como explica *Chagas* (2018) los esfuerzos para que la ciencia sea más abierta y accesible se concentran en aspectos que surgen antes y después de realizar experimentos (artículos, bases de datos y herramientas). Sin embargo, el acceso al instrumental científico necesario se encuentra fuertemente limitado por las restricciones económicas, lo que ha generado una disponibilidad desigual en todo el mundo, afectando predominantemente a países e instituciones de recursos limitados. El bajo costo de las tecnologías FOSH, combinado con la capacidad de modificarse las convierten en una valiosa herramienta para su uso en ciencias.

En este contexto, cabe destacar que el desarrollo de dispositivos IoT (del inglés, Internet of Things) basados en plataformas FOSH para medir parámetros agro-hidrológicos **ha sido una parte esencial y preponderante de la tesis actual**. El uso de este tipo de tecnologías se encuentra en ascenso continuo desde hace años. Por ejemplo, se han creado sistemas diversos para monitorear calidad de aire, crecimiento de cultivos, calidad de agua, entre muchos otros (*Mesas-Carrascosa et al.*, 2015; *Ripoll et al.*, 2019; *Rabault et al.*, 2020; *Gore et al.*, 2020). El uso de plataformas FOSH para el monitoreo hidro-ambiental aplicado al sector del agro implica un paradigma novedoso para un sector pujante de la economía de la República Argentina (RA) como es la producción de alimentos.

2 Contribución de la tesis

En el contexto descrito, donde se ha subrayado la crítica importancia de la medición de variables hidro-ambientales **esta tesis doctoral propone un aporte novedoso** en el avance de este campo multidisciplinario. El presente trabajo se ha enfocado en el desarrollo de dispositivos de medición innovadores, flexibles, de bajo costo y su aplicación para cuantificar y caracterizar variables hidrológicas, ambientales y biológicas (agro-hidrológicas) relevantes con el fin de comprender su interrelación (Figura 1.7).

A través de un enfoque integral, **esta investigación** ha buscado contribuir al cono-

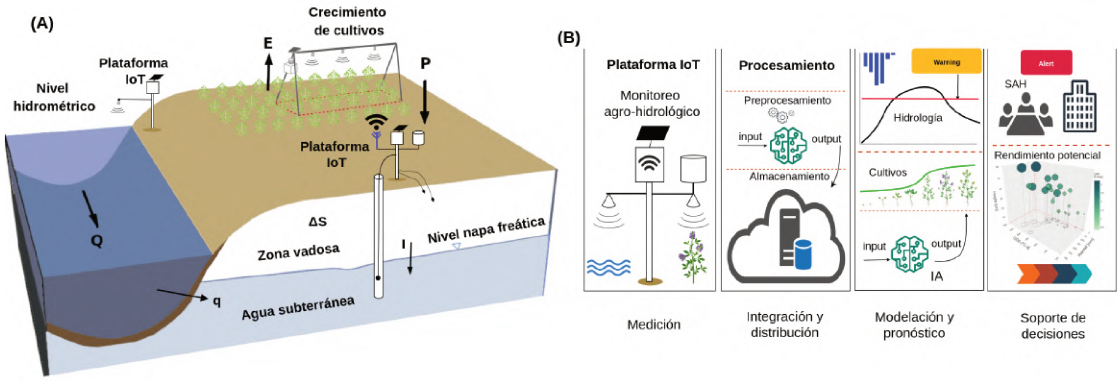


Figura 1.7: **(A)**: Registro de variables agro-hidrológicas mediante dispositivos automáticos. **(B)**: Resumen de las etapas de un sistema de monitoreo hasta el uso de los datos como soporte en la toma de decisiones (basado en *Plate*, 2007 y *Henonin et al.*, 2013).

cimiento y la aplicación práctica de herramientas de monitoreo que permitan evaluar el estado y la dinámica hídrica, y a su vez, vincular su impacto en el crecimiento y rendimiento de cultivos, específicamente de alfalfa, a escala de lote.

El aporte de esta tesis se ha cristalizado mediante cuatro artículos, dos de ellos orientados a datos hidrológicos, aplicados a una cuenca hidrográfica regional (cuenca inferior del río Salado de la Prov. de Santa Fe) y a una cuenca urbana de la ciudad de Santa Fe. Los otros dos artículos están vinculados a parámetros de importancia en el sector agrícola y su interrelación en el crecimiento y rendimiento del cultivo de alfalfa. En la Sección 4 se presentan cada uno de los artículos.

Esta tesis se centró en generar un aporte a la temática sin desprenderse ni de la aplicación práctica ni de la relevancia socio-económica de la problemática enunciada. Es así que gran parte del esfuerzo se concentró incialmente en facilitar el acceso a la información relevante sobre la vulnerabilidad hídrica de Santa Fe (cuenca inferior del río Salado y cuenca urbana de la ciudad de Santa Fe). Por último, y aprovechando la experiencia ganada en la primera parte del desarrollo de la tesis, el trabajo se orientó a producir una herramienta para que, en un futuro no muy lejano, el productor pueda tomar mejores decisiones a la hora de potenciar el rendimiento de su cultivo.

3 Objetivos

3.1 Objetivo general

El objetivo principal de la presente tesis fue, por un lado, el desarrollo y la instrumentación de tecnologías innovadoras para la adquisición de datos agro-hidrológicos mediante dispositivos IoT capaces de adaptarse al monitoreo de diferentes fenómenos hidrológicos. Y por otro, la utilización del dato recolectado para analizar y modelar el crecimiento y rendimiento de cultivos de alfalfa. Por tanto, la tesis consta de tres etapas bien diferenciadas: *i*) desarrollo tecnológico de instrumental de bajo costo y monitoreo de variables agro-hidrológicas, *ii*) validación de los datos recopilados y estimación de incertidumbres y *iii*) modelación y simulación de crecimiento vegetal.

3.2 Objetivos específicos

- Diseñar dispositivos de medición flexibles y de bajo costo basados en plataformas de hardware libre / open source, destinados tanto a la medición de parámetros agro-hidrológicos (para su eventual uso en la gestión del recurso hídrico), como para su aplicación al sector agropecuario.
- Investigar y aplicar algoritmos de aprendizaje automático para la calibración de sensores de bajo costo, con el propósito de mejorar la precisión de las mediciones y su adaptabilidad a diversas condiciones ambientales, así como su integración efectiva en los dispositivos de medición desarrollados.
- Monitorear el crecimiento de los cultivos integrando sensores y tecnologías adecuadas para la medición continua y automática del proceso biológico de crecimiento.
- Desarrollar modelos de crecimiento y rendimiento de cultivos basados en la información recopilada, permitiendo la creación de curvas predictivas que faciliten, eventualmente, la implementación de mejores prácticas agronómicas.
- Validar la efectividad del paradigma propuesto comparando los resultados obtenidos con métodos tradicionales.

4 Contenido de la tesis

El Apéndice A ofrece una visión integral de la investigación realizada en esta tesis a través del artículo de revisión "Artificial Intelligence of Things for Smart Water Systems" (*Ferrer-Cid et al.*, 2024a), actualmente **aceptado** aunque en proceso de corrección para su publicación en *IAHR Water Monograph Series*. Este trabajo de revisión establece las bases teóricas y el estado del arte de los sistemas IoT aplicados a fenómenos hídricos y la integración de técnicas de inteligencia artificial, ejemplificando con casos de Santa Fe (hidrológicos y agro-hidrológicos) el potencial de las tecnologías AIoT para la monitorización y validación de datos.

El Apéndice B presenta la investigación aplicada "A cost-effective redundant communication system for improving the reliability of a flood early warning system" (*López et al.*, 2020), **publicada** en *Journal of Hydroinformatics*. El estudio aborda las deficiencias del Sistema de Alerta Hidrológico del río Salado en Santa Fe, documentando sus fallas recurrentes e implementando un sistema de transmisión redundante que mejora significativamente la confiabilidad de las estaciones propietarias, proporcionandole mayor flexibilidad operativa y robustez.

El Apéndice C incluye una contribución **publicada** en la revista *Sensors* bajo el título "A low-power IoT device for measuring water table levels and soil moisture to ease increased crop yields" (*López et al.*, 2022). El trabajo desarrolla dispositivos IoT de bajo costo basados en hardware libre para medir parámetros hidro-ambientales agrícolas, optimizando su consumo energético e implementando algoritmos de machine learning para la calibración de sensores de nivel freático y humedad del suelo.

Finalmente, en el Apéndice D, se presenta el estudio sobre la medición de parámetros agro-hidrológicos y su impacto en el crecimiento y rendimiento de alfalfa. Este

trabajo titulado “Estimating alfalfa growth and yield by measuring plant height with a low-cost, close-range scanning device” (*López et al.*, 2024), se encuentra actualmente en proceso de **evaluación** la revista *Field Crops Research*. En el artículo se aborda el desarrollo de una estación de monitoreo de altura de cultivos con hardware libre mediante sensores ultrasónicos. Se implementaron modelos de crecimiento y rendimiento de cultivos basados en la información recolectada.

5 Metodología

5.1 Sistemas de medición hidrológica

Sistema de alerta en una cuenca regional

El uso de dispositivos de tecnología propietaria para la medición de parámetros hidroambientales dificulta, debido a su naturaleza cerrada, su adaptación a diferentes escenarios. El Sistema de Alerta Hidrológica (SAH) del río Salado se conformó, en su momento, con 38 estaciones de este tipo a un coste elvado. Su misión era registrar parámetros como precipitación, nivel freático y altura hidrométrica, entre otros; para que los tomadores de decisión contaran con información valiosa con la suficiente antelación. Uno de los mayores inconvenientes que padeció la red desde sus inicios fue la pérdida frecuente de datos por fallas en la transmisión satelital ORBCOMM (Figura 1.8).

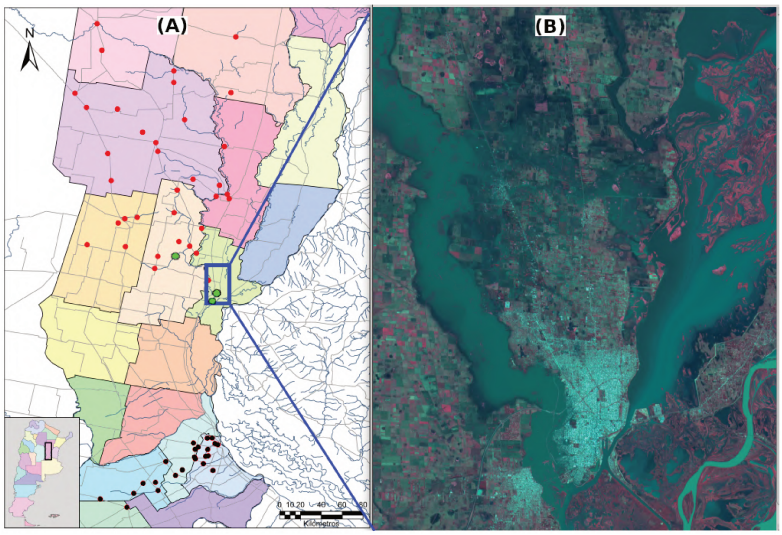


Figura 1.8: **(A)** Sistema de alerta hidrológico de la Prov. de Santa Fe (círculos rojos) y estaciones instrumentadas con el sistema redundante (círculos verdes). **(B)** Inundación de Santa Fe, imagen satelital SPOT-4 del 03-05-2003.

En esta investigación, se desarrolló una interfaz de comunicación integrada por componentes de software y hardware, diseñada específicamente para interactuar con un dispositivo cerrado mediante el software propietario proporcionado por el fabricante. El enfoque adoptado permitió emular la operación manual del dispositivo, como si fuera manejado por un usuario, permitiendo la automatización del proceso de extracción de

datos. Posteriormente, se implementó un sistema que facilitara la descarga de la información registrada y su transmisión a través de un enlace inalámbrico alternativo.

El sistema se instrumentó en las estaciones de la ciudad de Santa Fe, Santo Tomé (sobre el río Salado) y sobre el Arroyo de cañada Flesia (círculos verdes en Figura 1.8A). Este enfoque fue utilizado para "manipular" diferentes estaciones de monitoreo (meteorológicas, eddy covariance, entre otras) dado que es frecuente contar con dispositivos heterogéneos en diversas aplicaciones hidrológicas en un mismo sitio de observación. Los detalles sobre el SAH y la implementación electrónica y de software están presentados en el artículo del Apéndice B.

Entre los equipos de tecnología propietaria se encuentran aquellos que es posible gestionarlos prescindiendo del software del fabricante debido a que proveen comandos que pueden ser ejecutados desde una terminal, generalmente a través de comunicación serial (RS232 o USB). Estos equipos (denominadas *open stations* en el artículo del Apéndice B) permitieron reducir a su mínima expresión el hardware necesario para manipularlos. Para estos casos se modificó un router hogareño para que ejecute un sistema operativo Linux embebido (OpenWRT) y así posibilitar el acceso a los datos almacenados y gestionar el dispositivo según se requiera.

Monitoreo hidrológico en una cuenca urbana

La ciudad de Santa Fe se encuentra circundada por los ríos Paraná (sistema A° Leyes/Paraná) y Salado, ambos con valles de inundación que incluyen parte del entramado urbano (Apéndice A). Para la protección de las inundaciones fluviales cuenta con terraplenes que conforman el anillo de defensa y hacia su interior por reservorios que acumulan los excedentes pluviales del área urbana. Posteriormente, son evacuados por medio de estaciones de bombeo al río Salado. Dos de estos reservorios forman parte de las 142 ha de la Reserva Natural Urbana del Oeste (RNUO) (Figura 1.9) creada para amortiguar los excesos pluviales y proteger la biodiversidad regional (*López et al.*, 2023).

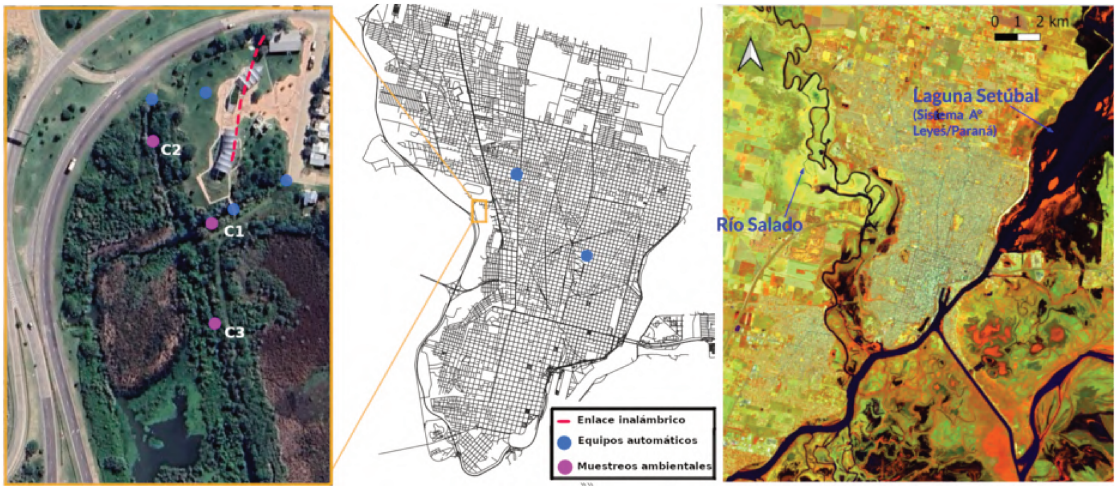


Figura 1.9: Red de monitoreo hidrológico implementada en la ciudad de Santa Fe. Recuadro: Reserva Natural Urbana del Oeste donde descargan los desagües pluviales.

Estos reservorios reciben el escurrimiento de dos subcuencas urbanas de una su-

perficie cuasi-impermeable de 1800 ha. Pese a su importancia, la cantidad de agua que ingresa y permanece en estos reservorios es aún desconocida con exactitud, representando un riesgo tanto para la población que vive sobre las márgenes y utiliza sus recursos, como para la integridad ecológica de los ríos que reciben su aporte (*Vaschetto et al.*, 2021). Para dar solución a este inconveniente se implementó un sistema de monitoreo hidrológico utilizando dispositivos desarrollados en base a plataformas de hardware libre / open-source (detalles en la Sección 5.2). Esta red mantuvo su funcionamiento desde diciembre de 2022 a julio de 2024, y una parte continúa operativa en la actualidad (Figura 1.9). Las variables de interés registradas son precipitación, nivel freático y altura hidrométrica en los ductos de desagüe troncal de la salida de la cuenca y en el reservorio principal[3] (Figura 1.10).

A diferencia de los equipos de tecnología propietaria, la flexibilidad de los dispositivos desarrollados facilitaron su adaptación a cada caso. De esta manera, la totalidad de las variables fueron registradas por el datalogger desarrollado que fue adaptado y programado en función del sensor utilizado (Apéndice C). En cuencas urbanas cuyas dimensiones suelen ser de unos pocos kilómetros es de gran interés registrar los procesos de transformación *lluvia - escorrentía*, más aún en ciudades de gran vulnerabilidad hídrica como Santa Fe. Debido a que estos procesos son de rápida respuesta se modificó la frecuencia de muestreo para una elevada resolución temporal. Los pluviómetros registran la precipitación acumulada cada 15 minutos, el nivel hidrométrico en la salida de la cuenca (alcantarilla de dos vanos de 1.3 m de ancho por 1.5 m de alto cada uno) se registra cada 1 minuto, la altura del reservorio y la profundidad de la napa freática se registran cada 30 minutos.

Este nivel de resolución temporal permitió explorar el proceso de transformación *lluvia-escorrentía* mediante modelos para series temporales basados en datos. En el Apéndice A.4.1 se muestran resultados preliminares del método de aprendizaje automático Random Forest. Es importante destacar que para entrenar este modelo se requiere una cantidad considerable de eventos que permitan capturar la variabilidad en su intensidad, duración, etc. para obtener un mejor desempeño[4].

Figura 1.10: Equipos instalados en una cuenca urbana de la ciudad de Santa Fe.

[3]Proyecto ganador del concurso Aguas Claras de la Fundación Bunge y Born (2022).
[4]Monitor de datos: `https://monitorReservaSfe.streamlit.app`, *accedida 27-06-2024*.

5.2 Desarrollo de dispositivos basados en hardware libre

El acceso a las estaciones del SAH del río Salado permitió una evaluación en forma simultánea de dispositivos basados en plataformas FOSH en su etapa embrionaria para registrar niveles hidrométricos (ver Apéndice A). Las dificultades e inconvenientes asociados a la escasa flexibilidad de las estaciones del SAH resultaron una motivación que llevó al desarrollo de dispositivos de medición automáticos utilizando como base plataformas FOSH. La descripción detallada sobre el diseño del datalogger, sus componentes y los sensores utilizados se encuentra en el artículo del Apéndice C. A continuación se presenta un resumen de las características de los dispositivos.

Diseño del equipo

Un aspecto clave en el desarrollo de dispositivos FOSH es la elección de una plataforma de uso masivo, que cuente con presencia en el mercado local, con documentación profusa y con un consumo energético reducido considerando que estará funcionando en forma ininterrumpida en campo. El dispositivo fue diseñado para funcionar bajo cuatro estados:

- Medición: obtiene las lecturas de los sensores.
- Almacenamiento: registra los datos en memorias persistentes.
- Transmisión: envía los datos en forma inalámbrica (en caso que corresponda).
- Suspensión: duerme, ingresa en un modo de bajo consumo hasta que nuevamente sea el momento de medir.

Bajo este esquema de trabajo se programó el firmware de un datalogger con funcionalidades básicas, teniendo en cuenta que la transmisión es una opción de la que se puede prescindir ya que en la mayoría de los sitios de campo la conectividad es deficiente o nula. En base a esto el *datalogger base* fue conformado por un módulo de almacenamiento de datos persistentes (memoria micro SD), un reloj de tiempo real (módulo RTC), el microcontrolador principal (Arduino Pro Mini), un sistema de alimentación con baterías recargables mediante energía solar y una serie de componentes electrónicos necesarios para la adaptación del voltaje y la optimización del consumo energético. En la Figura 1.11 se muestra la evolución de las placas prototipo que conforman el datalogger base.

El análisis de consumo energético fue realizado en laboratorio individualizando la carga de cada componente para el sistema completo. Se estableció que con un consumo en torno a 1 mA se alcanzaba la suficiente autonomía para 3 meses de funcionamiento con dos baterías en serie de 18650 Li-ion de 3.7 V, 2600 mAh. Este período es lo suficientemente amplio teniendo en cuenta que las baterías se cargan diariamente mediante paneles solares. Para alcanzar estos niveles de consumo se utilizó un circuito electrónico basado en transistores que desconectan los módulos RTC y SD cuando no están siendo utilizados y los conecta para el almacenamiento y la consulta de la fecha y hora. A su vez, se incorporaron conexiones extras para la alimentación de sensores con este mismo esquema de corte, en caso que presenten un consumo elevado. Se agregaron líneas

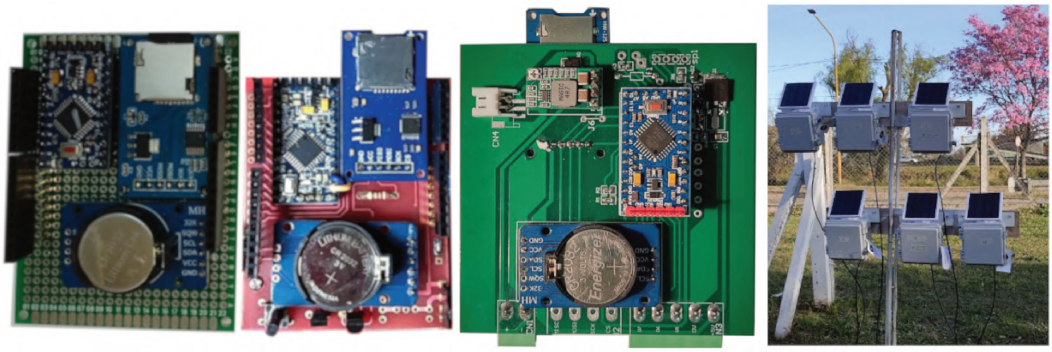


Figura 1.11: Evolución en el diseño de la placa principal del datalogger y varios equipos en fase de prueba antes de su instalación en campo.

con alimentación permanente para sensores que requieren un tiempo considerable de alimentación, por ejemplo, sensores químicos utilizados para medir calidad de aire.

Contar con transmisión inalámbrica es un desafío de múltiples aristas. Un aspecto importante es el elevado consumo energético, otro, desde una perspectiva económica, es el abono mensual que implica el uso de telefonía celular. A estos inconvenientes se le suma la deficiente o nula cobertura en la mayoría de los sitios de campo (en Argentina). Los mayores avances en este sentido se realizaron en laboratorio (*Manso et al.*, 2018) y su uso en campo fue parcial. La topología desplegada fue a través de enlaces con nodos intermedios hasta un sitio con acceso a Internet (casa de familia o escuela cercana). Debido a su bajo costo y consumo energético reducido, la tecnología seleccionada fueron las placas Nordic nRF24L01+ de 2.4GHz. Los datos recibidos en el nodo con acceso a Internet se envían mediante protocolo *http* a un sistema en la nube desarrollado para tal fin (*Boutet*, 2020). Más detalles en el Apéndice C.2.6.

Parámetros agro-hidrológicos

El componente principal e ineludible de todo equipo de medición es el datalogger. Sin embargo, la calidad de los datos estará condicionada por la bondad del sensor utilizado ya que es, en última instancia, quien provee el vínculo con el mundo físico. Durante el proceso de investigación, análisis y validación de sensores, algunos fueron descartados por no cumplir con los requerimientos de precisión, estabilidad, consumo energético o, simplemente, la estrategia para medir la variable de interés fue desestimada.

Entre los parámetros monitoreados, algunos fueron de validación directa y otros requirieron un proceso más elaborado de calibración. Para los primeros, el trabajo principal consistió en el encapsulado para su protección del medio ambiente y la programación, de entre estos se pueden mencionar sensores de humedad y temperatura ambiente (HTa), anemómetros (An), pluviómetros (P) y temperatura de suelo (Ts). En el otro extremo se encuentran hidrómetros (Hd), sensores de humedad del suelo (Hs), sondas de nivel freático (Nf) y altura de cultivos (Ac).

El caso del sensor de presión diferencial utilizado para el freatímetro o limnímetro requirió un armado artesanal de una dedicación especial. Las dimensiones del integrado y sus puertos de conexión al aire y agua son de un diámetro cercano al milímetro por tanto se utilizaron sondas nasogástricas pediátricas y soportes diseñados con impresoras

3D para finalmente encapsularlo con resina epoxi. De este modo se conformó una sonda de 12 metros de longitud que permite medir hasta 10 metros de columna de agua. Para establecer el diálogo con este sensor se programó el firmware específico[5], no así para el resto de los sensores que contaban con sus propias bibliotecas. Estos sensores pueden utilizarse para registrar altura de ríos y arroyos aunque requieren una instalación de mayor complejidad, por tal razón se optó por distanciómetros ultrasónicos para estos casos.

Para el registro de precipitación se utilizó el método convencional por cangilones. Inicialmente se programó el firmware del datalogger utilizando el mecanizado de un pluviómetro comercial. En etapas posteriores se realizaron diseños propios mediante impresiones 3D. La validación se realizó utilizando en simultáneo equipos de referencia. El ciclo de funcionamiento del datalogger (*medición - almacenamiento - transmisión - suspensión*) cambia levemente para el caso del pluviómetro por cangilones ya que cada volcado del cangilón se debe contabilizar aunque el equipo se encuentra en modo bajo consumo (*suspensión*). En este caso se utilizó el esquema de llamadas a interrupciones (IRQ, del inglés interrupt request) que provee el microcontrolador Atmel368p (presente en la placa Arduino Pro Mini), de modo que ante cada vuelco del cangilón se emite un pulso que "despierta" al microcontrolador, se contabiliza y luego retorna a su estado previo. De este modo, cuando el programa retoma su ciclo de trabajo, almacena las variables actualizadas.

Para el caso de los sensores de humedad del suelo de bajo costo, se utilizó resina epoxi en el encapsulado para proteger el circuito electrónico y como elemento de tracción. Esta resina es comúnmente utilizada en sensores comerciales (por ej. Hydraprobe, Decagon 10HS, etc.), aun así se verificó que el material no afecta sus lecturas mediante pruebas de laboratorio asegurando así la inocuidad en las mediciones. En la Tabla 1.1 se listan los sensores utilizados, su tecnología de medición y protocolo de comunicación con el datalogger para los diferentes sistemas de medición desarrollados.

Tabla 1.1: Variables registradas por el dispositivo desarrollado

Parámetro	**Modelo/Marca**	**Tecnología**	**Protocolo**
Nf	HSC / Honeywell	Presión diferencial	SPI
Hs	HydraProbe II / Stevens	TDR	SDI-12
	SKU: SEN0193 / DFRobot	Capacitivo	Analógico
Hd y Ac	MB7092 / Maxbotix	Ultrasónico	Serie/Analógico
HTa	AM2302 / Aosong Electronics	Capacitor / termistor	DHT
Ts	DS18b20 / Maxim Integrated	Semiconductor	1-Wire
An	Adafruit	Efecto Hall	Analógico
Pr	Propio	Cangilón	IRQ

Calibración y validación de sensores

La calibración se refiere al proceso de corrección de errores sistemáticos en las lecturas de los sensores, generalmente comparando un primer instrumento de referencia con un segundo sensor no calibrado. Con este procedimiento se busca ajustar sus parámetros

[5] Aporte a github.com/vwls/Honeywell_pressure_sensors, *accedida 28-08-2024*

y proporcionar una estimación precisa. Más específicamente, definimos como y_i la medición de referencia, y como $\mathbf{x}_i$=[x_{i1}, ..., x_{iM}] el vector que inclye las mediciones por calibrar y las mediciones externas, el proceso de calibración consiste en hallar la función f:$\mathbb{R}^M \rightarrow \mathbb{R}$ que mejor aproxima estas mediciones a la de referencia:

$$y_i = f(\theta, \mathbf{x}_i) + \epsilon_i \quad , \tag{1.1}$$

con $i = 1 \cdots N$ mediciones, $f(\cdot)$ es la función utilizada para calibrar el sensor y ϵ_i es ruido aleatorio distribuido siguiendo la distribución normal de media cero y varianza σ^2, es decir, N(0, σ^2) y θ son los parámetros del modelo de calibración para ser optimizados. Entre los diferentes algoritmos para estimar la función $f(\cdot)$ se utilizó Regresión Lineal Múltiple (MLR) para aquellos casos donde se considera que las mediciones tienen un comportamiento lineal y K-Nearest Neighbors (KNN), Support Vector Regression (SVR) y Random Forest (RFO) para los casos donde el comportamiento considerado es no lineal (ver Apéndice C).

Los sensores pasaron tres etapas de ensayos para su validación: una primera en laboratorio, donde se contrasta con mediciones manuales o pruebas rudimentarias, luego una a la intemperie junto a otros instrumentos considerados confiables y finalmente su instalación en campo. En general, las dos últimas instancias son utilizadas para obtener datos para la posterior calibración. A continuación se describen en forma resumida aquellas variables que han requerido un proceso de calibración más elaborado.

Freatímetro / limnímetro
El sensor de presión diferencial, utilizado para diseñar el freatímetro / limnímetro, fue validado en laboratorio mediante el registro de descargas por gravedad de columnas de agua de cuatro metros y comparando al inicio con mediciones manuales y modelos teóricos como balance de cantidad de movimiento y balance de energía simplificado (Bernoulli). En una segunda instancia, la sonda fue evaluada en situaciones reales, en conjunto de otras dos sondas comerciales. En este ensayo se deprimió la napa mediante bombas de extracción y luego se recargó el pozo con seis metros de columna de agua, registrando hasta su descarga natural. Se utilizaron regresiones lineales para evaluar la relación existente entre dos instrumentos considerados de referencia y el desarrollado. En la Figura 1.12 se muestra la evolución de la sonda desde el primer prototipo a su versión final[6].
Humedad del suelo
Los sensores de humedad del suelo fueron primeramente validados en laboratorio realizando análisis gravimétrico con muestras de diferentes tipos de suelo. Estos primeros ensayos dan una idea del rango de medición de los sensores, sensibilidad y estabilidad a fin de constatar su viabilidad para instalar en campo. Para el análisis gravimétrico fue integrado un sistema de cámara fotográfica automática para tomar las lecturas de las balanzas y en simultáneo las mediciones de los sensores. Se desarrolló un sistema OCR (del inglés, optical character recognition) para identificar los valores de las balanzas, vinculando las lecturas de los sensores. Con este ensayo se verificaron las curvas de secado obtenidas por los sensores de bajo costo (Figura 1.13).

Para la calibración se utilizó como instrumento de referencia el sensor HydraProbe II de Stevens y para la temperatura del suelo, otra de las variables predictoras, el sensor

[6]La empresa CCG Ingeniería SA escaló la fabricación del equipo para su comercialización mediante un acuerdo de vinculación tecnológica con la Universidad Nacional del Litoral

Figura 1.12: Evolución de sonda freatimétrica hasta su versión final.

DS18b20 (Tabla 1.1), ambos sensores de uso extendido. Estas variables fueron registradas en simultáneo en diferentes sitios de campo. Para la determinación de los tipos de suelo se realizó un análisis granulométrico de cada muestra. Con estos datos se realizaron calibraciones evaluando los diferentes algoritmos de aprendizaje automático para resolver la Ec. 1.1, esto es, MLR, KNN, SVR y RFO (ver más en Apéndice C.2.9).

Para las regresiones no lineales se requiere una instancia previa de cálculo de hiper-parámetros de cada modelo. Para esto, los datos se dividieron en un conjunto de entrenamiento y de prueba (proporción 70-30 % respectivamente). El conjunto de datos de entrenamiento se utilizó para obtener los hiper-parámetros con la técnica de validación cruzada en $k = 10$ particiones, mientras que con el conjunto de datos de prueba se verificó el desempeño de los modelos. Las predicciones se evaluaron utilizando la raíz del error cuadrático medio (RMSE) y el coeficiente de determinación (R^2). En la etapa previa al ajuste, para la selección de los hiper-parámetros, se utilizó el error medio sesgado (MBE).

Hidrómetros

Los niveles de altura hidrométrica se registraron con sensores distanciómetros ultrasónicos de resolución centimétrica y un rango de medición de hasta 750 cm. El sensor utiliza el principio TOF (del inglés Time of Flight), esto es, computa el tiempo entre la emisión y recepción del pulso ultrasónico luego de interactuar con el objeto destino y finalmente se calcula la distancia utilizando la velocidad del sonido en el aire. Su funcionamiento se validó en laboratorio con mediciones manuales utilizando cintas métricas contra superficies sólidas. Una segunda instancia fue en campo sobre agua en superficie libre en la Laguna Setúbal, contrastado con los registros manuales tomados diariamente por Prefectura Naval Argentina (Figura A.5). Esta prueba tuvo un doble propósito, validar las lecturas del sensor acústico y evaluar la robustez del dispositivo datalogger para funcionar en campo durante más de un año. Posteriormente, y luego de las sucesivas mejoras al equipo, fue utilizado para medir niveles hidrométricos en la salida de una cuenca urbana de la ciudad de Santa Fe, esto es, en un ducto troncal de descarga y en uno de los reservorios pertenecientes a la Reserva Natural Urbana del Oeste (ver más en Apéndice A.4.1).

Crecimiento de cultivos

Para medir el crecimiento de alfalfa en altura se utilizó el mismo distanciómetro ultrasónico que para la altura hidrométrica. Los sensores se ubican a una distancia fija en

Figura 1.13: Ensayos varios de validación y calibración de sensores de humedad del suelo

dirección al cultivo, perpendicular al suelo. Dependiendo de la densidad de las hojas de alfalfa y de cómo el haz ultrasónico rebote en ellas, la cobertura de la alfalfa puede comportarse como una combinación de múltiples capas de material blando y poroso, con la superficie del suelo como un límite rígido final, o por el contrario, como una superficie dura y rugosa con buenas propiedades reflectantes.

Las estructuras de soporte se modificaron durante el transcurso de la investigación. El primer prototipo fue un brazo de un sensor único, luego una estructura con cinco sensores tipo estrella y finalmente una estructura lineal de tres sensores. Si bien este sensor fue validado contra superficies "duras", es decir, no porosas, su evaluación con cultivos de alfalfa fue *in situ* mediante un equipo portátil. Posteriormente, se instaló en campo de manera permanente desde noviembre de 2020 a la actualidad, con algunos períodos sin mediciones por fallas en el equipo (ingreso de agua, por ejemplo) que requirieron su reparación. A partir de estos inconvenientes se realizaron mejoras sucesivas sobre el datalogger y la estructura de la estación (Figura 1.14). Las lecturas de los sensores ultrasónicos se contrastaron con mediciones manuales, evaluando la altura máxima, mínima y promedio de las plantas.

5.3 Modelación de crecimiento y rendimiento de alfalfa

Sitio de estudio

Los muestreos y la instalación de los dispositivos fue realizada en la estancia Don Silvano SRL (latitud 31° 19′ 13,359″ S, longitud 61° 8′ 23,982″ O), cercana a la localidad de Humboldt (Prov. de Santa Fe, Argentina), a 80 km de la ciudad de Santa Fe (Figura 1.15A). La principal actividad agrícola de la estancia es el cultivo de pasturas (alfalfa y maíz) para la producción de leche y carne.

En Humboldt, los veranos son calurosos, húmedos y mayormente despejados, y los inviernos son cortos, frescos y parcialmente nublados. Durante el año, las temperaturas generalmente varían entre 7 °C y 32 °C, raramente bajan de 0 °C o suben por encima de 36 °C. El régimen de precipitaciones de la estación de INTA Rafaela marca un promedio anual de 1004 mm en los últimos 54 años. El estudio fue realizado entre enero de 2021 y abril de 2024, con un período de lluvias por debajo del promedio (Figura 1.15B) atribuido al inusual fenómeno de tres años consecutivos de condiciones de La Niña (*Li et al.*, 2023).

Datos y muestreos

Para la evaluación del rendimiento de alfalfa se realizaron muestreos destructivos y luego el cálculo de su biomasa. El procedimiento consistió en cortes a 12 ± 2 cm del suelo, pesando en balanzas de precisión antes y luego del secado, obteniendo así la materia fresca y seca de cada muestra (MF y MS, respectivamente). El secado se realizó en hornos a 65 °C durante 48 hs. El rendimiento del cultivo se calculó dividiendo la materia seca por la superficie de corte (DMY, del inglés Dry Matter Yield). En detalle:

- Se realizaron 46 cortes de alfalfa, entre enero de 2021 y abril de 2024. En cada corte se registró la altura manual de cada parcela además de otros parámetros registrados automáticamente por el equipo. Con los pares de datos *altura-rendimiento* (R, DMY) se realizó una primera evaluación mediante regresión lineal simple.
- Entre los datos previos, un subconjunto de 25 cortes, cuentan además con el registro del resto de las variables hidro-ambientales como temperatura ambiente, precipitación, profundidad de la napa freática, entre otras. Con estos datos se evaluó el DMY en función de las otras variables mediante regresión lineal múltiple.
- Por último, se utilizaron 12 ciclos de crecimiento de alfalfa completos, esto es, el proceso evolutivo que va desde su rebrote hasta alcanzar su altura máxima, cuando se realizó el nuevo corte. Con estos datos y el cálculo de unidades térmicas se modeló el proceso evolutivo mediante tres modelos de crecimiento logístico (Apéndice D).

La discrepancia entre la cantidad de datos utilizados fue consecuencia del mantenimiento de los equipos, con más fallas en los primeros prototipos aunque dieron lugar

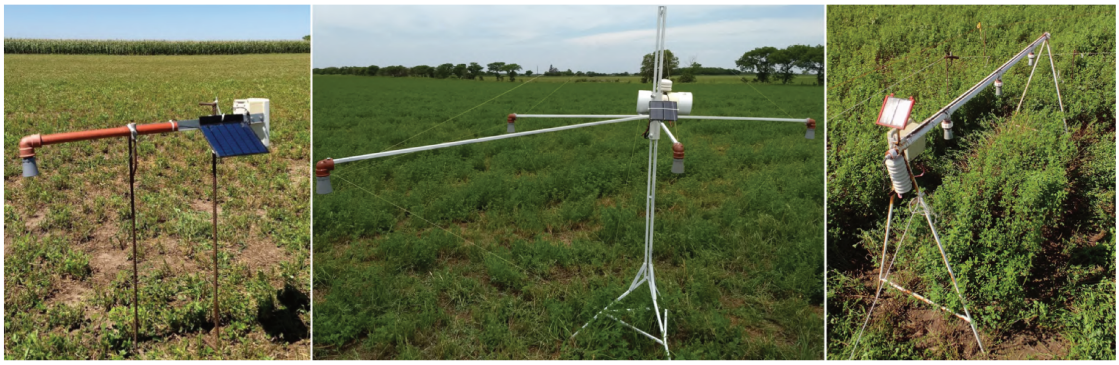


Figura 1.14: Evolución del diseño de la estación de cultivos.

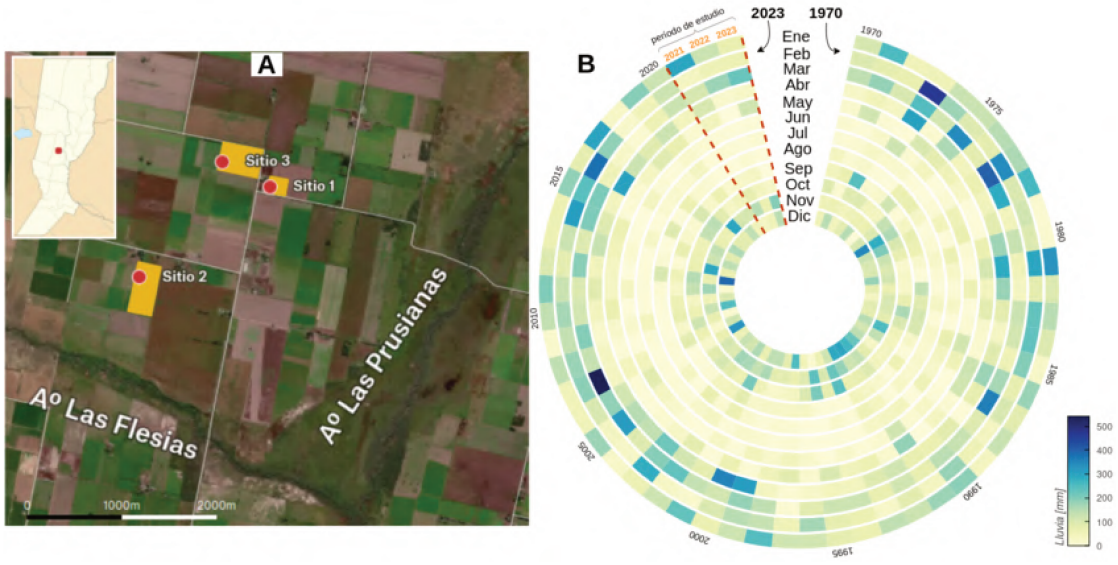


Figura 1.15: **(A)** Sitios de campo donde se realizó la instalación de la estación de cultivos. **(B)** Régimen mensual de precipitaciones entre 1970 a 2023 de la estación INTA Rafaela, en naranja la ventana en la que se realizó el estudio.

a mejoras sucesivas a medida que se ganaba experiencia. Si bien estas tareas de mantenimiento se tradujeron en ausencia de datos (debido al reemplazo de sensores, retiro de equipo para reparación en laboratorio, etc.) dieron como resultado una mejora en la calidad de los datos obtenidos posteriormente y equipos mas robustos.

La frecuencia de muestreo de los equipos fue programada cada 30 minutos. En el caso de la estación que registra la altura de la alfalfa, en cada instancia de medición se realizan 10 lecturas en ráfaga. Durante el período de esta investigación se realizaron más de 50 viajes a campo para llevar a cabo tareas de mantenimiento de equipos, limpieza de sensores, cortes de alfalfa, desmalezamiento, traslado de la estación, descarga de datos, entre otras actividades.

Modelación

Se modelaron dos aspectos del cultivo de alfalfa, en primer lugar el proceso evolutivo de crecimiento en altura de la planta, y posteriormente, los aspectos vinculados al rendimiento de materia seca (ton/ha). Para el primer caso, se utilizaron modelos teóricos sigmoideos o curvas en forma de ***S*** capaces de describir la dinámica completa del crecimiento del cultivo. Inicialmente el cultivo exhibe un crecimiento exponencial, luego lineal y finalmente de agotamiento o saturación. Los modelos teóricos utilizados fueron el de Fisher, Landau y Gompertz. Para el segundo aspecto, se modeló la producción total de materia seca (DMY) totalizando las variables consideradas más relevantes durante el ciclo de desarrollo del cultivo hasta el momento del corte.

Crecimiento de alfalfa

El crecimiento de la alfalfa fue modelado en función de unidades térmicas o acumulación de calor (G_{DD}, del inglés growing degree days) del cultivo, calculadas a partir de los datos registrados de temperatura ambiente. El concepto de unidades de calor tiene

un efecto significativo en el ciclo de crecimiento de la planta, desde la siembra hasta la cosecha, y se calcula en términos de los G_{DD} (*Elnesr & Alazba*, 2016).

La forma más sencilla de calcular los G_{DD} es tomar la temperatura máxima más la temperatura mínima diaria, dividirla por dos (o la temperatura media) y restar la temperatura umbral o base T_b que debe superarse para que ocurra el crecimiento. En resumen, el desarrollo de la alfalfa es insignificante a temperaturas por debajo de 5 °C. A partir de la temperatura registrada por el dispositivo (cada media hora) se calculó el G_{DD} por cada ciclo de crecimiento mediante el método trapezoidal de integración numérica (ver Ec. D.1) (Figura 1.16).

Los datos de altura correspondientes a cada ciclo de crecimiento considerados como función del G_{DD} acumulado fueron ajustados por los modelos teóricos mediante el método de mínimos cuadrados no lineal. Los ciclos de crecimiento de alfalfa (en altura) fueron de aproximadamente 30 días de duración en diferentes períodos entre los años 2021 al 2023, para un régimen de precipitaciones anuales (entre 600 y 900 mm) por debajo de la media histórica (1004 mm) y con dos ciclos para el año 2024 (Figura 1.15B).

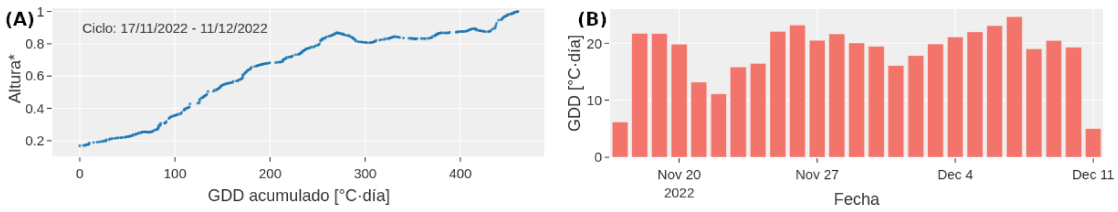


Figura 1.16: **(A)** Altura de alfalfa normalizada, presentada como función de los G_{DD} acumulados. **(B)** G_{DD} diarios para el mismo ciclo de crecimiento

Para el ajuste se utilizaron tres modelos teóricos, considerando una ecuación evolutiva en el tiempo $dR/dt = f(t, \mathbf{c})$, donde R es la altura de la planta y el vector de parámetros $\mathbf{c}$ contiene las interrelaciones desconocidas, convirtiendo luego la dependencia temporal en la variación de G_{DD}.

El primer modelo utilizado es conocido como de crecimiento logístico, también llamado ecuación de Verhulst o de Fisher. El segundo modelo es la ecuación de Landau. *Landau* (1944) propuso una ecuación para explicar la evolución de una perturbación en la transición a la turbulencia de flujos newtonianos, basada en consideraciones fenomenológicas. Por último, el modelo de Gompertz, ampliamente utilizado en muchos aspectos de la biología para describir el crecimiento de animales y plantas (*Tjørve & Tjørve*, 2017), y que puede considerarse como una generalización de la función logística.

La normalización de R y G_{DD} es requerida por razones de convergencia y estabilidad numérica, pero además, porque de este modo se establece un marco de referencia universal que permite comparar los ciclos de crecimiento durante diferentes épocas del año, con períodos de duración variable y diferente altura final alcanzada (G_{DD} y R, respectivamente).

El modelo de Fisher fue ajustado para cada uno de los ciclos de crecimiento. Luego, haciendo uso del marco de referencia común provisto por la normalización, se obtuvo una *curva universal* de crecimiento promediando la totalidad de los ciclos. Finalmente, se ajutó cada modelos a la curva promedio, en este caso empleando el concepto de

máxima probabilidad (maximun likelihood), es decir, minimizando el valor de la variable chi-square χ^2. La formulación matemática completa, así como también los detalles del procedimiento para calcular el G_{DD} se encuentra descrito en el Apéndice D.

La dispersión de la señal debido a la penetración dispar del eco acústico en el cultivo es producto de su "porosidad" y densidad de cobertura por lo que se requiere un procesamiento previo con el fin de obtener una curva suave que describa el crecimiento de la planta a lo largo del tiempo y que será en última instancia utilizada para ajustar los modelos teóricos. El procedimiento para obtener datos suavizados consistió en la limpieza de valores anómalos y seguidamente:

- **A)** Obtención de las envolventes superiores (e_{sup}) para los datos de cada sensor. Se calculó la media (μ_w) y desviación estándar móvil (σ_w) con una ventana (w) de 30 datos, con un intervalo de confianza (z) del 95 %: $e_{sup} = \mu_w + z \cdot \sigma_w$

- **B)** Suavizado de las envolventes mediante filtrado. La envolvente de cada sensor fue suavizada mediante el procedimiento estándar de filtrado LOWESS (del inglés locally weighted scatterplot smoothing). En forma simplificada, el algoritmo estima el valor suavizado con los N puntos más cercanos usando una regresión lineal ponderada (*Cleveland*, 1979). Otros procedimientos (filtro de Wiener por ejemplo) fueron evaluados y descartados por un desempeño inferior al seleccionado.

- **C)** Promediado. Se obtuvo el promedio entre las curvas suavizadas del paso previo. Esta curva fue utilizada como el crecimiento evolutivo del cultivo para los análisis posteriores.

En la Figura 1.17 se muestra el tratamiento de datos para un ciclo de crecimiento de alfalfa correspondiente a marzo de 2023, en el que se pueden observar las salidas de cada una de estas instancias; que luego son utilizadas como insumo para la siguiente fase. Estas tareas previas a la modelación son fundamentales para preparar o mejorar los datos antes de aplicarlos a modelos (ya sea analíticos o de IA). En el Apéndice A.3.1 se detallan esta serie de tareas comunes previas a su uso en el que también suelen asistir herramientas de IA. A su vez, en el Apéndice D se detalla el enfoque utilizado para los datos de cultivos que aquí se presentan en forma resumida.

Rendimiento de alfalfa

Al finalizar cada ciclo de crecimiento se realizaron cortes del cultivo a 12 ± 2 cm para calcular el rendimiento de su biomasa (DMY) (Figura 1.18). Los muestreos se realizaron en tres sitios, ya que la estación de cultivos fue desplazada entre diferentes potreros dependiendo del estado de la alfalfa (Figura 1.15). El rendimiento de la alfalfa es un fenómeno complejo que involucra la contribución de parámetros climáticos, del cultivo y del suelo. Una de los parámetros relevantes es la altura del cultivo (R), por lo que es razonable indagar sobre su correlación con la biomasa producida. Para los 46 pares de datos (R, DMY) se realizó una regresión lineal simple y, para un subconjunto de 25 cortes, donde además se registraron los otros parámetros agro-hidrológicos, se realizó una regresión lineal múltiple seleccionando los predictores (dependientes del tiempo e independientes entre sí) de acuerdo a la calidad, cantidad y disponibilidad de datos, entre ellos, la profundidad de la capa freática, la lluvia caída, la altura de la planta,

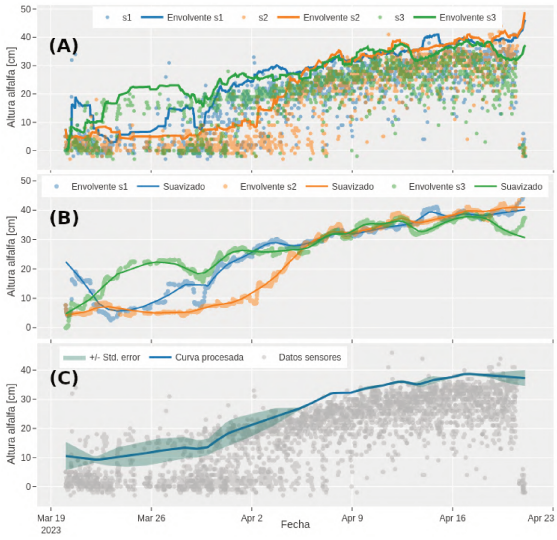


Figura 1.17: Procesamiento de datos obtenidos por la estación de crecimiento de cultivos. **(A)** Envolvente de datos de cada sensor. **(B)** Suavizado de envolventes mediante filtrado LOWESS. **(C)** Promediado de curvas suavizadas, salida utilizada para el ajuste de los modelos propuestos.

G_{DD} acumulados, etc. Todos los datos seleccionados correspondieron al valor final o acumulado al final del ciclo del cultivo.

La altura final de las plantas se determinó promediando las lecturas de los sensores. En situaciones donde la lluvia aplastaba el cultivo se reemplazaron los valores de la altura de la planta con las lecturas anteriores a la afectación. Este fenómeno se hizo evidente frente a lluvias intensas gracias al registro simultáneo de las variables agrohidrológicas. En contraposición, el valor utilizado para el G_{DD} acumulado fue el de la fecha del muestreo destructivo, en parte porque se consideró que la biomasa continuó aumentando hasta ese momento. Una vez seleccionados, los datos fueron estandarizados restando la media y dividiendo todo el conjunto por su desviación estándar ($z = (x - \mu)/\sigma$).

Para evaluar los modelos de crecimiento de cultivos se utilizó el coeficiente de determinación (R^2) y la raíz del error cuadrático medio (RMSE, del inglés root mean square error, también conocido como *rmsd*) normalizado por el rango entre los valores máximo y mínimo de los datos de la medición. A su vez, como evaluación de la bondad de los ajustes y del posible sesgo en el tratamiento de los datos se analizó la distribución de los residuos entre las curvas de cada ciclo respecto a la curva promedio y a los modelos teóricos ajustados.

6 Resultados y Discusión

La presente sección expone los resultados organizados en cuatro ejes: primero, las mejoras en sistemas de alerta propietario y el diseño y desarrollo del sistema de monitoreo hidrológico conformado integramente con plataformas FOSH, mostrando su viabilidad y ventajas; segundo, el análisis del desempeño de dispositivos y sensores, destacando

Figura 1.18: Muestreo destructivo para determinar el rendimiento de materia seca.

precisión, eficiencia energética y robustez; tercero, los hallazgos de la estación de monitoreo de alfalfa, demostrando la capacidad de generar modelos precisos de crecimiento y rendimiento; y cuarto, los desafíos y perspectivas de implementación, donde se analizan los aspectos necesarios para la adopción de estas tecnologías a gran escala. Los resultados confirman que las tecnologías de hardware libre han sido efectivas para obtener datos valiosos sobre procesos hidrológicos y biológicos.

6.1 Mejoras en el sistema de alerta y desarrollo del sistema de monitoreo hidrológico

El aporte inicial de esta tesis radica en el diseño de una interfaz común de comunicación que permitió dialogar con sistemas originalmente cerrados y poco flexibles, brindando la posibilidad de gestionar la estación, manipular y enviar los datos medidos por el Sistema de Alerta Hidrológico (SAH) del río Salado a través de enlaces inalámbricos alternativos. En particular, los datos de altura hidrométrica para tres estaciones del SAH (instrumentada por el Gob. de la Provincia de Santa Fe) fueron enviados por el sistema redundante diseñado con una pérdida de datos del 0.7 % comparado con el sistema satelital ORBCOMM con pérdidas de un 20.3 %. Este desarrollo permitió dotar de cierta flexibilidad a equipos variados de tecnología propietaria como estaciones meteorológicas, eddy covariance, entre otros, de diferentes fabricantes (Pegasus, Spectrum, Campbell, Stevens, Genica, etc.).

Durante el período que transcurrió la investigación en el SAH del río Salado se evaluaron los primeros prototipos, principalmente el datalogger y los sensores ultrasónicos para el registro de altura hidrométrica. Los resultados demostraron completa fiabilidad ($R^2 = 0.99$) en las lecturas en comparación con datos oficiales (ver Figura A.5). En esta primera instancia se observó un elevado consumo energético identificando así el punto débil de estas plataformas, sin embargo demostraron ser una alternativa viable y en muchos casos superadora respecto a las estaciones comerciales.

El sistema de monitoreo hidrológico desplegado en una cuenca de la ciudad de Santa Fe se realizó utilizando los dispositivos FOSH desarrollados para tal fin. Los datos registrados desde diciembre de 2022 a la actualidad demostraron una excelente calidad y permitieron observar la dinámica de la cuenca con un gran nivel de detalle, obteniendo información hidrológica desconocida por la gestión municipal. La red de monitoreo permite analizar el impacto que tiene la precipitación en la fluctuación de las variables relevantes como el nivel hidrométrico en la salida de la subcuenca (alcantarilla troncal), el nivel del agua del reservorio de almacenamiento y las napas freáticas.

A modo de ejemplo, en la Figura 1.19A se observa la alcantarilla troncal inundada, alcanzando un nivel pico de 73 cm una hora posterior al inicio de la lluvia. Para ese mismo evento la napa freática subió 85 cm, llegando a aproximadamente 1 m de la superficie y retornando a su nivel previo al evento recién 10 días después. Esto sucedió para una lluvia caída en dos días consecutivos, 56 mm y 16.5 mm respectivamente, con una intensidad máxima de 60 mm/h. En la Figura 1.19B se muestra el comportamiento del sistema ante una lluvia caída de 100 mm que produjo un incremento en el nivel del reservorio de 70 cm, un pico máximo en la alcantarilla troncal de 60 cm luego de una hora posterior al inicio de la lluvia. En 12 horas la napa ascendió 1 m, llegando a 1.23 m de profundidad. Para este caso la intensidad de la lluvia fue de 37 mm/h.

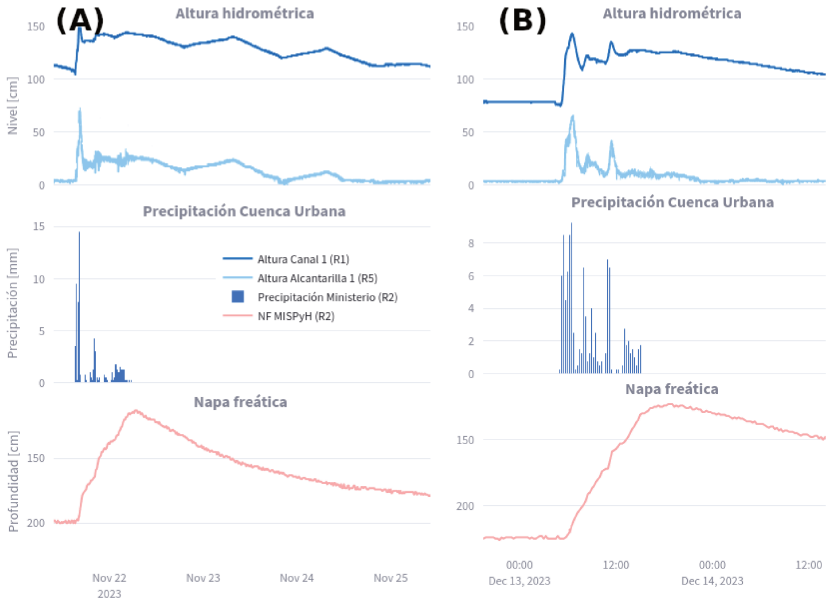


Figura 1.19: Respuesta del sistema ante dos eventos de lluvia. **(A)** Precipitación de 60 mm/h de intensidad en la que la alcantarilla troncal de descarga al reservorio se inunda (celeste), se observa 150 cm de nivel del reservorio y su descenso escalonado producido por bombas de desagote. **(B)** Precipitación de 37 mm/h de intensidad, alcantarilla troncal sin inundarse, reservorio a 150 cm de nivel.

6.2 Dispositivos y sensores desarrollados para el monitoreo de variables agro-hidro-ambientales

Los dispositivos desarrolados han permitido la obtencion de información valiosa para la comprensión de procesos hidrologicos y su dinámica como también para el estudio del crecimiento de cultivos. El equipo fue instalado en campo para registrar variables hidrológicas, ambientales y biológicas. Las pruebas en campo fueron de 2 a 4 años, y dependiendo del equipo continúan en la actualidad. Las técnicas de optimización de consumo energético aplicadas dieron como resultado mejoras significativas, con un consumo de aproximadamente 0.35 mA cuando se encuentra en modo inactivo (“suspensión”) y un promedio de 53 mA en el “ciclo de trabajo”, esto es, midiendo, almacenando y transmitiendo.

El consumo energético en el momento de la medición y almacenamiento es considerable aunque esta acción se realiza en una fracción de segundo, de modo que el valor gravitante es el que se produce fuera del ciclo de trabajo. El nivel de consumo alcanzado lo hicieron apto para su funcionamiento en campo con un sistema de alimentación solar de requisitos mínimos y por tanto sin que los costos se incrementen de manera significativa.

El *datalogger base* fue replicado y desplegado en campo para registrar parámetros hidro-ambientales como precipitación, humedad del suelo, profundidad de napa freática, altura hidrométrica, humedad y temperatura ambiente, velocidad del viento y altura de cultivos. Estos equipos fueron utilizados para instrumentar cuencas urbanas de la ciudad de Santa Fe y en potreros con cultivos de alfalfa. El desempeño de los equipos en campo fue estable, dando cuenta que es una tecnología robusta capaz de tolerar las inclemencias climáticas con el debido mantenimiento para evitar lecturas erróneas de los sensores por suciedad o insectos (Figura 1.20).

En lo que respecta a las técnicas de aprendizaje automático lineales y no lineales (MLR y KNN, SVR, RFO respectivamente) para la calibración de los sensores de humedad se obtuvieron errores relativos promedio de 2.1-4.6 % y 8.9-13.2 % para los dos tipos de suelo analizados, ambos instalados en sitios de campo. En todos los casos, con el modelo Random Forest se obtuvo el menor RMSE relativo (2.1 % y 8.9 % para suelos del tipo 2 y 3 respectivamente). El sensor de presión diferencial utilizado para el diseño de la sonda freatimétrica demostró un comportamiento lineal respecto a la columna de agua medida, de una paridad 1 en 1. Los resultados obtenidos en contraste con sondas comerciales dieron cuenta de este desempeño con un $R^2 = 0.99$. Una descripción detallada sobre el desempeño alcanzado puede encontrarse en la sección de resultados del Apéndice C.

Figura 1.20: Afectación en el funcionamiento de los equipos debido a insectos

6.3 Estación de monitoreo de la altura y rendimiento de alfalfa

La estructura soporte de la estación de cultivos fue modificada en el tiempo obteniendo los mejores resultados con la disposición lineal conformada por tres sensores. Este diseño facilitó la alineación de cada sensor sobre el surco donde crecen las plantas de alfalfa a diferencia de la estructura de brazos concéntricos.

La correlación entre las mediciones automáticas y las realizadas manualmente mostraron un comportamiento lineal con un coeficiente de determinación $R^2 = 0.85$. Un resultado considerado más que aceptable teniendo en cuenta que la altura del cultivo no es un valor absoluto o único y que la medición manual presenta ciertas incertidumbres y subjetividades frente a las lecturas objetivas e integradas que brinda el sensor.

El modelo de Fisher fue ajustado para los 12 ciclos de crecimiento tomando como variable explicativa los G_{DD} acumulados (ver Tabla D.1). Estos ajustes arrojaron un RMSE relativo entre 3.2 % y 11.4 % y un R^2 de 0.83 a 0.99. Como ejemplo, en la Figura 1.21 se muestra el ajuste para un único ciclo. La curva universal, obtenida promediando todos los ciclos normalizados, fue ajustada con los modelos de Fisher, Landau y Gompertz con un RMSE relativo de 2.5 %, 2.7 % y 4.9 % respectivamente. El coeficiente de determinación para Fisher y Landau fue de $R^2 = 0.99$ y para Gompertz un $R^2 = 0.98$. En la Figura 1.22 se muestra la dispersión de los tres modelos respecto a la curva promedio universal.

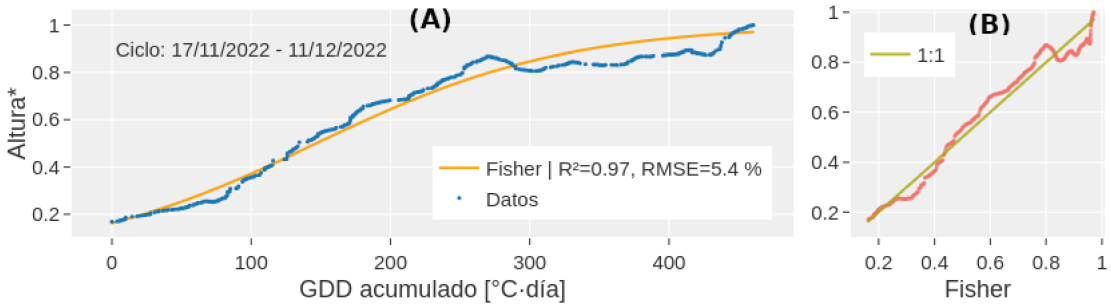


Figura 1.21: **(A)** Ajuste del modelo Fisher para un ciclo de crecimiento (Altura* es la altura del cultivo normalizada en [0, 1]). **(B)** Comparación entre los datos observados y el modelo de Fisher.

El análisis de incertidumbre muestra que la distribución de los residuos entre todos los ciclos de crecimiento normalizados respecto a la curva promedio y a los modelos teóricos presentan un comportamiento Gaussiano cayendo la mayoría dentro del rango de ±2 desvíos, entorno al 95 % de los residuos. Además, los errores se distribuyen normalmente en torno a cero, indicando que son aleatorios y no hay sesgos significativos. Estos resultados permiten analizar la bondad de los ajustes, indicando que el modelo propuesto de Landau tuvo un desempeño superior a los modelos conocidos de Fisher y Gompertz (más detalles en el Apéndice D).

El rendimiento de materia seca (DMY) utilizando la altura como única variable explicativa mediante una regresión lineal arrojó un $R^2 = 0.74$ para un conjunto de 48 muestreos destructivos (Figura 1.23). Para el análisis multivariado se descartó el uso de la profundidad de la napa freática como variable predictora del DMY. Esto se debe a que la alfalfa nunca superó los 3 años de antigüedad desde su fecha de siembra, por lo que es poco probable que la extensión de las raíces hayan alcanzado el agua subterránea en niveles inferiores a los 7.5 m de profundidad. Del mismo modo para los sensores de humedad del suelo, por fallas sistemáticas y dificultad de mantenimiento e instalación.

Las variables utilizadas en la regresión lineal múltiple (MLR) se seleccionaron en base a su disponibilidad y calidad, entre ellas, la precipitación total, los G_{DD} acumulados y la altura final del cultivo. El resultado arrojó un $R^2 = 0.80$ y un $rmsd = 12.3$ %. En este

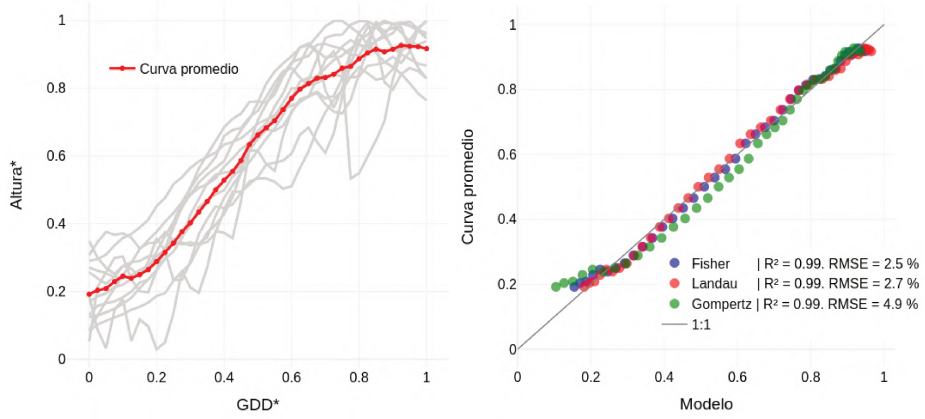


Figura 1.22: Comparación entre modelos de crecimiento de Landau, Fisher y Gompertz y la curva universal de altura promedio.

caso, la influencia de las variables predictoras estandarizadas mostraron que, por cada incremento de una desviación estándar en la altura de la alfalfa, en los G_{DD} acumulados y en la precipitación acumulada se aumenta el DMY en 0.621, 0.327 y 0.01 (desvíos) respectivamente. Estos resultados muestran la importancia relativa que tiene la altura del cultivo como factor preponderante en el rendimiento, en segundo lugar los G_{DD} y luego la lluvia acumulada. Aunque es importante destacar que se requiere una mayor cantidad de observaciones, ya que la mayoría de los cortes fueron realizados durante un período de sequía extraordinaria causado por tres años consecutivos de *La niña* (*Li et al.*, 2023).

Un análisis individualizado de los predictores de la producción de materia seca y rendimiento se observa en la Figura 1.23. Si bien se infiere que la precipitación incide en el rendimiento, hay una serie de valores en torno al rendimiento de 2 tn/ha para un rango amplio de lluvia acumulada entre 0 y 100 mm.

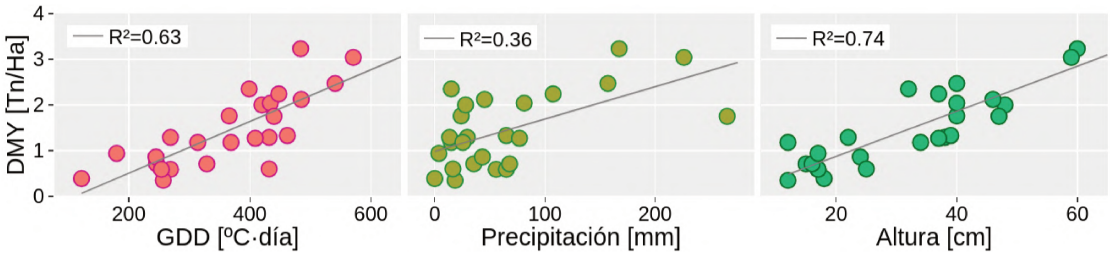


Figura 1.23: Relación individual entre el rendimiento (DMY) y cada una de las variables predictoras seleccionadas, el tamaño de las circunferencias indica el rendimiento.

Este fenómeno se presenta al totalizar la lluvia de cada ciclo, cuando se observó que el comportamiento difiere si esto se produjo en las etapas iniciales de crecimiento o al final, cercano al corte. Por ejemplo, un sesgo presente en este tipo de datos tiene que ver con una lluvia importante caída en días cercanos al corte del cultivo, cuya influencia impactará más en el ciclo subsiguiente que en el actual. Una subdivisión de la lluvia caída en diferentes momentos del ciclo bastaría para corregir este efecto indeseado. Aún así, los resultados obtenidos muestran que la selección de estas variables son

relevantes y que con pocos predictores medidos correctamente se puede realizar una buena aproximación al DMY (ver Apéndice D.3.2).

Un resultado destacable, que surge del uso de los sensores ultrasónicos en la medición de la altura del cultivo, está vinculado a la información extra que se puede obtener sobre su grado de cobertura, siendo uno de los factores preponderantes en el rendimiento. Al procesar los datos acústicos, se puede inferir la densidad vegetal según la penetración del eco. Un análisis preliminar muestra que en coberturas con diferentes niveles de precipitación (escasa y abundante), la distribución de los puntos de medición varía: en coberturas ralas es bimodal, mientras que en cultivos con follaje denso los puntos se concentran cerca de la altura máxima (Figura 1.24).

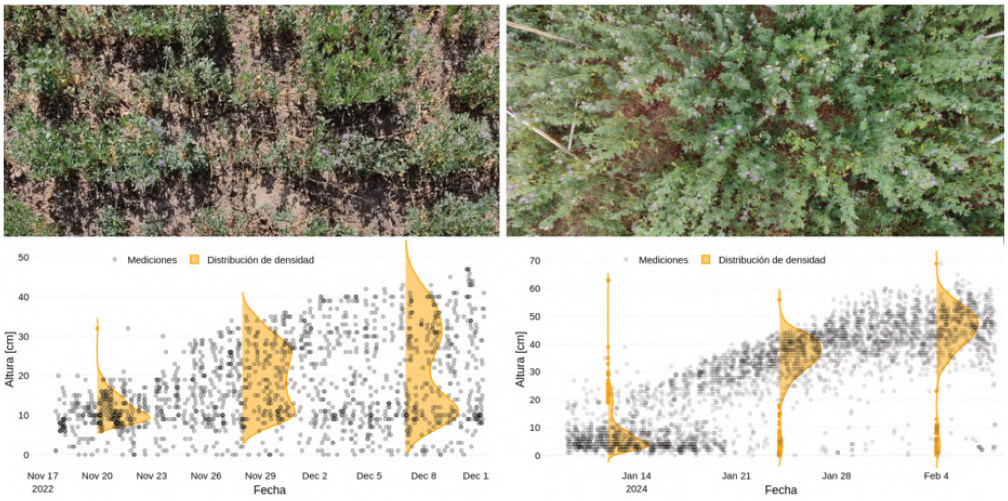


Figura 1.24: Penetración del eco acústico y distribución de densidad de mediciones para alfalfa de escasa y abundante cobertura vegetal.

6.4 Desafíos y perspectivas de implementación

Los desarrollos presentados han demostrado su viabilidad en condiciones de campo, aunque su adopción a mayor escala presenta varios desafíos. El sensor de nivel de napa freática demostró ser robusto y de fácil uso, ya que no requiere calibraciones sofisticadas y funcionó sin inconvenientes por más de tres años. Los sensores de humedad de suelo, si bien inicialmente presentaron un desempeño adecuado, con el transcurso del tiempo evidenciaron limitaciones significativas en cuanto a su fiabilidad y estabilidad operativa. En suelos de textura más gruesa (predominantemente arenosos) su funcionamiento fue comparativamente más consistente, lo que sugiere su posible uso para investigación con una calibración adecuada.

La estación de monitoreo de cultivos alcanzó un nivel de desarrollo y robustez que la hace viable para su adopción por productores. Sin embargo, para su implementación a mayor escala, sería necesario diseñar una interfaz que integre los componentes existentes y que facilite el uso por personal sin experiencia en condiciones de campo. Es importante señalar que estas tecnologías requieren un mantenimiento periódico debido a las condiciones ambientales adversas, característica que no es exclusiva de los equipos basados en hardware libre. Si bien en desarrollos sucesivos se implementaron mejoras

significativas en los sistemas de protección, factores como la presencia de insectos y la acumulación de tierra pueden interferir considerablemente en la precisión de las mediciones, comprometiendo la integridad de los datos obtenidos.

En cuanto a la aceptación institucional, tanto en el Sistema de Alerta Hidrológico como el la red de monitoreo urbana, aunque inicialmente hubo interés y compromisos para continuar el trabajo, los cambios de gestión dificultaron su continuidad. La falta de financiamiento y la discontinuidad de políticas impidieron consolidar las redes de alerta y monitoreo propuestas, evidenciando la dificultad de sostener estos desarrollos en el ámbito estatal.

7 Conclusiones y recomendaciones

El objetivo principal de esta tesis fue investigar fenómenos naturales vinculados con el ciclo hidrológico mediante el uso de sistemas de medición IoT desarrollados sobre plataformas de hardware libre / open source (FOSH). A lo largo de este estudio se desarrollaron sistemas de monitoreo flexibles que permitieron registrar diferentes variables hidro-ambientales tanto para estudiar su dinámica como para evaluar el impacto e interrelación en el crecimiento y rendimiento de cultivos de alfalfa.

Los resultados obtenidos pueden ser juzgados bajo dos aspectos. Primeramente, referido al desempeño demostrado por los dispositivos de medición y sus sensores asociados. Luego de exhaustivos ensayos, validaciones, calibraciones y más de 3 años de funcionamiento en campo se puede considerar que alcanzaron un elevado nivel de confiabilidad. Sumado a esto, probaron ser lo suficientemente flexibles para adaptarse a diferentes escenarios para la medición de procesos hidrológicos y problemas vinculados. El segundo aspecto está relacionado a la calidad de las mediciones obtenidas y el agregado de valor a los datos como insumo principal en modelaciones.

Los hallazgos de este trabajo dan cuenta del gran potencial en la integración de variables hidro-ambientales como la precipitación, profundidad de la napa freática y temperatura, entre otras, que junto a datos de crecimiento de cultivos permitieron mejorar estimaciones del rendimiento. Por tanto, el producto combinado de esta tesis, junto a desarrollos ulteriores (como la estimación de la morfología de la planta en base a la penetración del eco acústico) puede consituir una adecuada herramienta de soporte para la toma de decisiones del productor agropecuario.

Se ha presentado un enfoque innovador y de bajo costo para estimar el rendimiento de cultivos de alfalfa utilizando un dispositivo de escaneo ultrasónico. A través de este dispositivo, se obtuvieron mediciones precisas de la altura de las plantas de alfalfa a lo largo del tiempo. Los resultados mostraron la factibilidad de crear sistemas de medición flexibles que permiten obtener información novedosa y relevante tanto para el monitoreo hidrológico como para el sector agrícola.

A pesar de los aportes de este estudio, es importante reconocer algunas limitaciones. Desde el punto de vista del instrumental, corresponde evaluar en forma permanente el surgimiento de nuevas tecnologías, ya sea para incorporar funcionalidades o bien para facilitar el procesamiento de los datos *in situ* ya que el margen de recursos computacionales disponibles es mínimo. En lo referido a los cultivos, tanto en los modelos de crecimiento como en los de rendimiento, se debe tener en cuenta que la mayoría de los datos fueron recolectados en una situación anormal de sequía. Por tanto, una limitante

es que los resultados pueden estar segados a este régimen de precipitaciones.

Como trabajo a futuro se plantean tres líneas para continuar la investigación. En primer lugar, la integración de los componentes hasta aquí desarrollados en un único sistema predictor. Esto es, en base a los datos registrados inferir las curvas de crecimiento más probables y vincular la medición de altura del cultivo, las unidades térmicas y la lluvia acumulada con su potencial rendimiento. Otro aspecto importante a incorporar es la densidad de cobertura del cultivo como variable predictora del rendimiento. Un cultivo con una determinada altura pero con mayor desarrollo foliar tendrá un rendimiento superior a uno de igual altura pero menor follaje. Fue posible inferir el grado de densidad de la planta a partir de evaluar la penetración del eco acústico. En la Figura 1.25 se observa la diferencia entre un cultivo más poroso (cobertura rala), cuyo eco acústico alcanza hasta el suelo, con otro de una cobertura densa, donde se forma una franja superior que la señal acústica no logra atravesar. Este comportamiento bien diferenciado de dos ciclos de similar desarrollo en altura se explica por la ocurrencia de lluvias en la etapa temprana del desarrollo de la alfalfa.

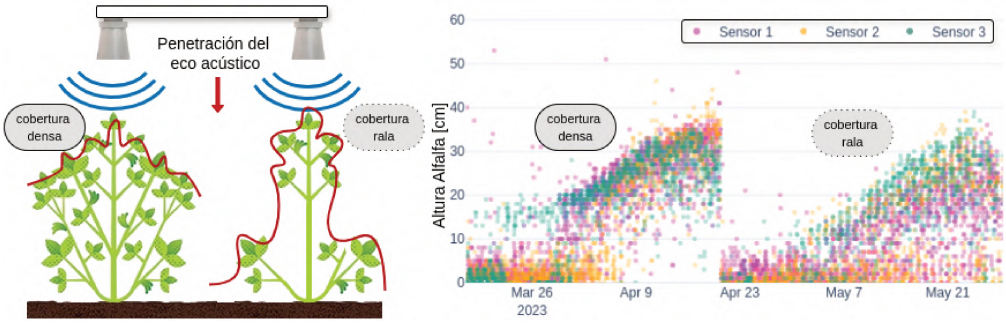


Figura 1.25: Penetración del eco acústico dependiendo del nivel de cobertura vegetal

Entre la gran variedad de sensores están los que realizan el cálculo de distancia al primer rebote recibido (en las hojas de datos se lo denomina envelope output) o aquellos que de la ráfaga que emiten promedian los ecos recibidos que poseen con mayor intensidad. Para el caso de superficies porosas como los cultivos, esta variante en la tecnología del sensor resultó ser la adecuada, ya que permite que la señal acústica penetre el cultivo y compute la distancia donde la planta presenta mayor densidad vegetal. Para incorporar este parámetro primeramente debe ser validado con un segundo instrumento que permita asociar el nivel de cobertura, por ejemplo, cámaras fotográficas. A su vez, el uso de fotografías junto al resto de variables agro-hidrológicas ayudaría a pronosticar el momento de floración, siendo un parámetro relevante para decidir el corte de alfalfa.

Por último, incrementar la cantidad de observaciones en otros escenarios hidrológicos tanto de ciclos de crecimiento como de rendimientos para aplicar regresiones de aprendizaje automático que permitan captar de mejor manera la no linealidad del fenómeno.

En conclusión, esta tesis ha demostrado la importancia de considerar factores hidrológicos, ambientales y biológicos para comprender el estado y la dinámica de diferentes fenómenos mediante el uso de dispositivos lo suficientemente flexibles para adaptarse a cada escenario. Los hallazgos presentados no solo enriquecen el conocimiento existente, sino que también tienen el potencial de influir en prácticas y políticas que promuevan un futuro más sostenible.

Apéndices

Apéndice A

Artificial Intelligence of Things for Smart Water Systems

Resumen Tradicionalmente las cuencas se caracterizaban estadísticamente basadas en registros a largo plazo, aunque esta metodología ya no es válida debido al cambio climático y otros efectos. La aparición de dispositivos IoT (del inglés, Internet of Things) y la Inteligencia Artificial (IA) han permitido el desarrollo de sistemas inteligentes de gestión de los recursos hídricos capaces de medir, monitorear, controlar y tomar decisiones automatizadas o semi-automatizadas en tiempo real. Estos sistemas inteligentes se conforman con la combinación de múltiples sensores, big data y tecnologías de IA para proporcionar una inestimable ayuda a tomadores de decisiones. Los componentes de las plataformas basadas en IoT incluyen sensores, dataloggers, módulos de comunicación, diferentes paradigmas de computación, estrategias de privacidad y seguridad a un bajo costo. Para la medición se utilizan diversos sensores como por ejemplo de temperatura, precipitación y niveles freáticos, mientras que el diseño de nodos y los métodos de comunicación permiten la transmisión de datos a largas distancias. Las tareas de mejora de la calidad de los datos registrados por las plataformas IoT mediante técnicas de IA incluyen la eliminación de ruido, la calibración de sensores, imputación de datos faltantes, fusión de datos y detección de anomalías, entre otros. Por otro lado, el uso de los datos abarca la modelación y predicción de fenómenos, detección de eventos y control automático de sistemas inteligentes. Dos casos de uso ejemplifican estas aplicaciones: en primer lugar, sistemas de monitoreo y alerta hidrológico en una cuenca urbana y en otra regional, y el segundo sobre aplicaciones agro-hidrológicas en campos de alfalfa utilizando sensores ultrasónicos. Se mencionan técnicas de aprendizaje automático para calibrar sensores, frecuentemente utilizadas para mejorar su desempeño. En conclusión, la combinación de tecnologías IoT e IA permite crear plataformas confiables y reducir los costos de implementación de redes de monitoreo, creando nuevas aplicaciones e innovaciones en la gestión inteligente del agua[1].

[1]Este capítulo fue enviado a *IAHR Water Monographs, Artificial intelligence in hydroinformatics*

1 Introduction

"Hydrological theories have arisen from knowledge gained through data analysis or have been confirmed by data supporting the theory" (*Wood*, 1998). In recent decades, however, the hydrologist's task has been complicated by two major changes. The first is the human alteration of natural systems and global warming, which has led to changes in hydrology (*Kundzewicz*, 2008). Previously, catchments were characterised statistically based on long-term records. Now the assumption of 'stationarity' is no longer valid due to climate change and other effects (*Han et al.*, 2020; *Houspanossian et al.*, 2023). The second challenge is the need to measure more and better a wider range of hydro-climatic parameters to track changes in hydrological conditions. In the past, the high cost of proprietary software and hardware and the need to replace spare parts has limited the long-term sustainability of data collection programmes.

With the emergence of the enabling technology known as the Internet of Things (IoT), the quest for an intelligent water management system has gained momentum. The accelerated development of electronic technologies led to a paradigm shift at the beginning of the second decade of this century. Open source hardware made it possible to assemble programmable microcontrollers with expandability through add-on boards and a development environment for creating custom software (*Fisher & Gould*, 2012b). Moreover, the combination of multiple sensors, big data and Artificial Intelligence (AI) technologies underpins the concept of intelligent water management and control systems. These can provide agricultural and livestock producers, manufacturers, water utility operators, and decision makers with the ability to measure, monitor, control and make automated or semi-automated decisions in real time (Figure A.1). Flood prevention, precision agriculture, pressure control in pipelines, real-time monitoring of marine leaks and water quality control in facilities such as fish farms or industries are some examples.

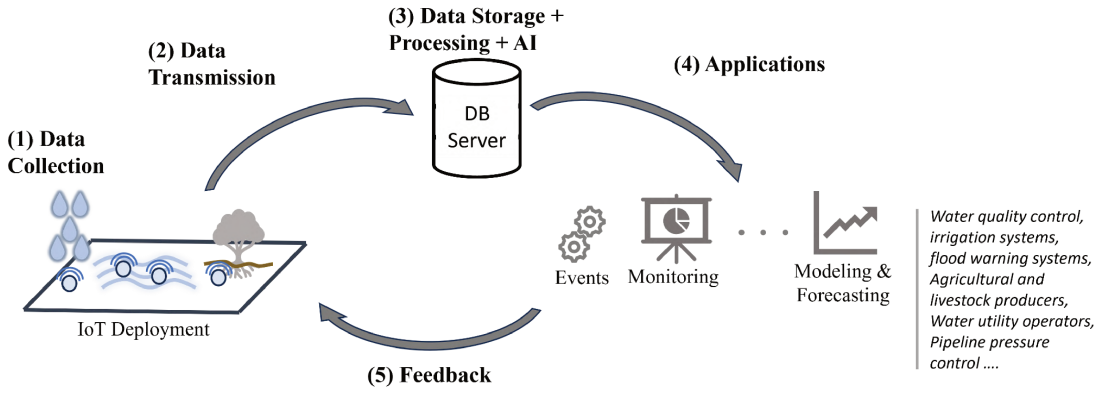


Figure A.1: IoT smart water system lifecycle: first, data collection using IoT devices; second, data transfer, storage and processing to a server, where it is fed into AI algorithms; and third, AI-triggered feedback to the IoT actuator platform.

Increasing urbanisation and industrialisation are exacerbating environmental problems. IoT systems provide a cost-effective alternative to ensure environmental sustainability, leading to the concept of smart cities to optimise water-related management

issues. Recently, IoT-based smart water systems have become an important component in the development of digital twins, which provide measures to particularise the modelling of a particular phenomenon (*Manocha et al.*, 2024). The incorporation of AI has improved predictive capabilities for the early detection of natural disasters. As a result, the decision-making processes and AI models responsible for obtaining information and deciding on actions are fed by the data collected by IoT platforms. The merging of AI and IoT has given rise to the so-called Artificial Intelligence of Things (AIoT), where AI is embedded in IoT platforms, processing the data collected and even guiding the design of smart devices.

This chapter delves into the development of AIoT for smart water systems. Firstly, it describes the IoT components of a smart water system, along with current trends and challenges. Then, it introduces the utilisation of AIoT, highlighting the role of AI in enhancing the quality of IoT data in what is referred to as the upstream pipeline. Additionally, it introduces some AI-powered downstream applications. The chapter concludes by examining two use cases: case A) the installation of IoT sensors at monitoring stations to enhance the capabilities of a flood warning system (FWS) in two distinct subwatersheds - one in a rural area and the other in an urban setting. These sensors contribute to an overarching FWS; case B) the monitoring of agro-hydrological parameters, including rainfall, soil moisture, water tables, temperature, and crop growth, in an alfalfa field using the same IoT system.

2 IoT enabling technologies

Next, the IoT-based platform components are explored: i) sensing, ii) node design, iii) communication, iv) computing paradigms, v) privacy and security, and vi) deployment and operational costs (Figure A.2).

2.1 Sensing

Miller et al. (2023) identified temperature, hydrogen potential (pH), dissolved oxygen, total dissolved solids (TDS), turbidity, electrical conductivity (EC), oxidation-reduction potential (ORP), chlorophylla, chemical oxygen demand (COD), and ammonia nitrogen, among others, as water quality parameters that can be measured using IoT technologies. According to *de Camargo et al.* (2023) the most commonly used water quality monitoring sensors are those from DFRobot (48%), followed by Atlas Scientific (8%) and Vernier (2%). Among the sensors mentioned in the literature, the following DFRobot models are the most commonly used: DO SEN0237, EC DFR0300, ORP SEN0165, pH SEN0161 and SEN0169, TDS SEN0244, temperature sensor DS18B20, turbidity sensor SEN0189 and level/depth sensor SEN0205. On the other hand, Atlas Scientific manufactures environmental and electrochemical sensors for environmental monitoring.

The most commonly used models are the env-20, env-40 and env-50 for measuring DO, pH, temperature, EC and ORP. Examples of Vernier sensor models are PH-BTA, ORP-BTA, SALT-BTA, DO-BTA, FLO-BTA and LJ-A, which are used to measure parameters such as pH, ORP, DO, salinity and flow velocity. Other sensors commonly found in the literature include temperature sensors from the DHT11, DHT12, DHT22,

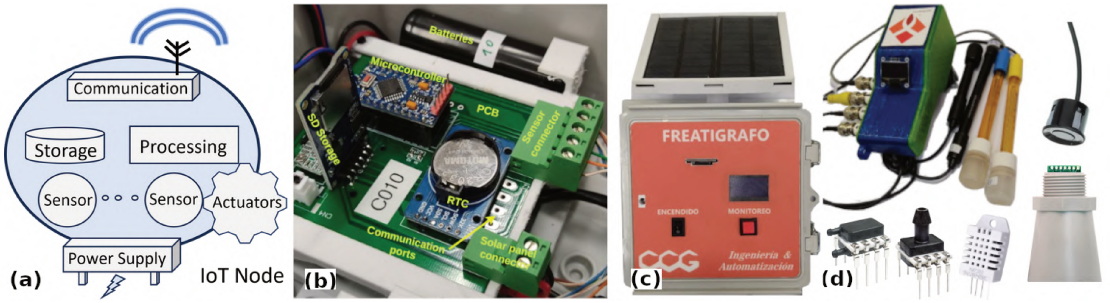


Figure A.2: (a) Diagram of a general IoT node design for measuring different phenomena. (b) Components of an IoT node mounted on a printed circuit board. (c) Front of an IoT node mounted in a waterproof enclosure. (d) pH, ORP, OD, and pH environmental sensors in a 3D self-built multi-parameter probe; pressure, temperature, relative humidity and ToF distance sensors.

PT100 and LM35 families, pH sensors E201-C, PH-4502C, DO D-6800, AR8210 and DOS-600, and turbidity sensors TSW-20M, BL5419 and ST100.

The time-of-flight (ToF) principle is used in rangefinder measurements to determine distances (water level, crop height, etc.). It measures distance by calculating the time taken for a signal to travel back and forth from its point of emission to its point of reflection from an object. Another method of measuring groundwater levels is to use submerged pressure sensors at predetermined distances, which record the pressure exerted by the water column above them. Difficulties in installation sometimes limit the use of these pressure sensors for river level measurement. For rainfall measurement, bucket rain gauges remain popular due to their durability and ease of connection to IoT devices.

2.2 Node design

The use of open source hardware platforms has grown rapidly, thanks in part to open/free licences such as Creative Commons and GPL, which are embedded in both the documentation and the distribution of software libraries. The core function of the node is to collect sensor readings, store information, possibly correct or process data, and transmit it to the cloud. These nodes are usually powered by batteries charged by solar panels and include a microcontroller, a real-time clock, actuators, a storage module and a transmission module. Sensors and modules with high power requirements can follow a power saving or duty cycle strategy to optimise power consumption. The most common protocols for module interconnection or sensor integration are I2C, SPI and Serial. Analogue sensors, typically 10 or 12 bit resolution, use the A/D converter on the microcontroller. For situations requiring higher resolution, external 16-bit converter modules are seamlessly integrated.

IoT devices can be classified as Low-End, Middle-End, or High-End (*Ojo et al.*, 2018). **Low-end** IoT devices are resource-constrained, with RAM and flash memory typically in the tens or hundreds of kilobytes, and a processor based on an 8-bit or 16-bit architecture. These devices are primarily made for simple sensing and actuation applications, and are typically programmed using low-level firmware. Many Arduino boards fall into the low-end category. **Middle-End** IoT devices offer more features

with greater processing capabilities. They typically have clock speeds and RAM in the range of hundreds of megahertz and kilobytes, respectively, and can support multiple communication technologies. Examples of middle-end IoT devices include the ESP8266, ESP32, and Arduino Yun boards. **High-end** IoT devices are typically single-board computers (SBCs) with a powerful processing unit, plenty of RAM, significant storage capacity, and possibly a graphical processing unit (GPU) to run traditional operating systems such as Linux or Windows 10 IoT Core. These devices can handle the execution of heavy machine learning algorithms. A notable example of a high-end IoT device is the Raspberry Pi.

Currently, the most widely used platforms are Arduino and ESP32, which can be seamlessly integrated with others with greater computing resources, such as Raspberry Pi, as required. However, for prototyping and experimental designs, other more specialised microcontroller units (MCUs) are used, such as the PIC16F877A, STM32L476RG or LPC2148 or CC2430 MCUs (*Silva et al.*, 2022).

2.3 Communication

Communication is one of the main tasks of an IoT node, where the collected data is transmitted to a central server or cloud for different purposes. This communication must not cause any losses, or delays in the case of a real-time application. In addition, the location of the nodes and restrictions in terms of power consumption can hinder or define the transmission mechanism and the use of communication protocols. Although there is a wide variety of technologies and communication protocols, the characteristics of IoT deployments for smart water systems (*López-Ramírez & Aragón-Zavala*, 2023), which tend to be highly energy-constrained, make those that offer reduced overhead and low power consumption the most relevant. These include long-range technologies with a coverage of several kilometres, such as cellular IoT technologies (M2M, NB-IoT, LTE-M), and low power WANs (LoRaWAN, Sigfox), and short-range technologies, with a coverage up to 100 metres, such as Bluetooth Low Energy (BLE), Wi-Fi HaLow (IEEE 802.11.ah), ZigBee, and WirelessHART (IEEE 802.15.4); (*Singh & Ahmed*, 2021). Moreover, these communication technologies impose hard restrictions in terms of data packet length, protocol overhead, etc, which implies the use of adaptation layers (e.g., 6LowPAN, 6TiSCH and 6Lo), header compression protocols (e.g., SCHC), and modifications of the standard data transfer protocols used in the Internet (e.g., CoAP and MQTT) (*Gomez et al.*, 2017; *Minaburo et al.*, 2021; *Shelby et al.*, 2014)

2.4 Cloud/fog/edge computing

The increase in the use of AI applications powered by data measured by IoT platforms and the need to process large amounts of data is a challenge. Computing paradigms need to adapt to IoT requirements and data volumes. The main approaches are the centralised and distributed paradigms (*Kalyani & Collier*, 2021). In the case of centralised paradigms, **cloud computing** is the most relevant, where all nodes send measurements to the cloud or a central server where data is stored and processed. Among the most used commercial technologies, there is Amazon Web Service (AWS) and Azure, which allow data processing through AI techniques. As for the distributed paradigms, there is

the edge and fog computing. **Edge computing** corresponds to the case where most of the data processing is performed at the data collection terminal so that the data can be sent to the cloud already processed. This approach requires the nodes to have increased computational and storage capacities, requiring more powerful microcontrollers, as is the case of Raspberry PI-based nodes. The **fog computing** case corresponds to an intermediate case where part of the processing is performed neither at the edge nor in the cloud, but at other devices that act as aggregators or intermediaries and that have a higher computing capacity than the IoT devices at the edge. Figure A.3 shows an illustrative example of the three paradigms described.

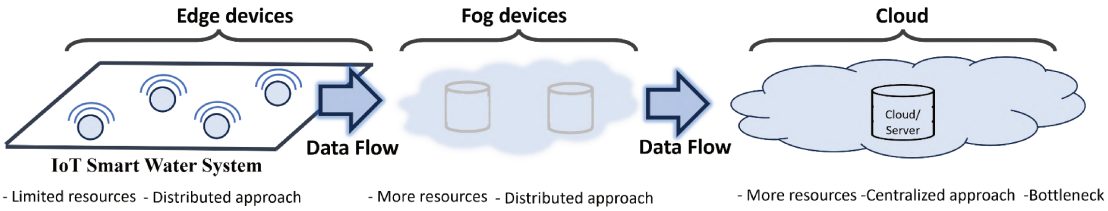


Figure A.3: Chart of cloud, fog and edge computing paradigms with their advantages and disadvantages.

2.5 Privacy and security

The most important aspects of data security include confidentiality, integrity, availability, and network security (*Al-Turjman et al.*, 2022). Therefore, it is essential to ensure data privacy and security when deploying IoT platforms, especially for industrial applications and critical smart water systems. IoT solutions include encryption to provide secure communication between nodes and servers while maintaining data confidentiality, with a particular focus on low-power, lightweight encryption. Another recently adopted alternative is the use of blockchain, which ensures data integrity using a decentralised paradigm (*Zeng et al.*, 2023). Recently, data security has been improved by detecting attacks and intrusions through the use of AI techniques. In particular, traffic logs can be used to feed intrusion detection models. Data poisoning attacks have been reported by introducing small perturbations to the AI model to produce undesired actions and results. Therefore, AI models capable of detecting data poisoning attacks through outlier and anomaly detection schemes have been proposed. Finally, federated learning (*Khan et al.*, 2021) is another trend aimed at preserving data privacy and ownership, mainly in heterogeneous networks with providers and third parties. It is a distributed approach to global training of AI models, where servers create independent models by accessing their own datasets. These models are then shared and unified by a central server, which limits the amount of data exchanged between servers. This approach is interesting when data from some smart meters cannot be shared directly due to data ownership issues.

2.6 Acquisition, deployment and operational costs

CAPEX (purchase) and OPEX (operational) dominate the costs associated with an IoT-based smart water system. The price of IoT node components, such as sensors, power

supplies, dust and waterproof enclosures, among others, make up the initial cost. Low-cost sensors and technologies are now a good alternative, although the trade-off for their low price is potential data quality issues. Then, there are the assembly costs, as well as the costs of setting up the cloud infrastructure, servers or fog, and the software required at IoT nodes for data collection, storage, and transmission. Secondly, there are the costs associated with the deployment phase, where nodes need to be placed in locations of interest that may be difficult to access. Finally, there are the operational or maintenance costs, which can be the most important part of a long-lasting smart system. These include replacing components when they fail, and relocating and maintaining nodes. Recently, AI techniques have been used to reduce operational costs by enhancing the data quality and performing predictive maintenance, e.g., identifying which nodes and sensors may have reliability problems and their cause (see section 3.1 for data quality-related AI tasks).

For example, high transmission rates in battery-powered nodes can require more power, more recharging, more wear and tear, or more frequent replacement, increasing operating costs. These costs are typically high because maintenance is critical to providing quality data. Another example is smart FWSs, which have evolved from fully integrated short-term hydrological forecasting to stand-alone in-situ measurement systems that warn their clients via SMS or other means (*Federico Vilaseca & Gorgoglione*, 2023; *Zakaria et al.*, 2023). The importance of an intelligent FWS becomes clear when the potential flood damage avoided is calculated, with a benefit of around €400 for every €1 invested. *López et al.* (2020) document a detailed five-year study that explains the main factor that caused the failure of the FWS of the lower Salado River basin (Santa Fe, Argentina). Failures were mainly due to sensor and data logger malfunctions, as well as power outages and transmission losses, or environmental factors such as lightning strikes or the presence of insects, and acts of vandalism. Finally, they estimated that a cost-effective open hardware alternative was on average eight times cheaper to maintain than the proprietary hardware-software system.

3 AIoT for smart water systems

AI in conjunction with IoT technologies have provided the necessary tools for the development of smart water systems (section 1). In particular, data obtained from IoT platforms can be used to perform different applications through AI techniques, e.g., event forecasting, early warning, or anomaly detection. AIoT is devoted to the development and use of AI techniques for processing data from IoT platforms, e.g., a smart meter network. Specifically, the AIoT tasks can be classified according to their purpose (*Ferrer-Cid et al.*, 2024b); a) the use of AI for improving the quality of the data collected by the IoT platform, in what is called the **upstream pipeline**, and b) the use of AI to carry out applications with the processed data in what is called the **downstream applications**.

3.1 Upstream tasks

The data collected must be of good quality if it is to be used for regulatory purposes and to provide reliable information. Data quality in this context refers to data accuracy

and completeness. However, the heterogeneous nature of IoT platforms implies that data quality issues can arise, either due to the faulty sensors and meters used or in transmitting the data. It is therefore important to introduce a series of AI-driven tasks to improve data quality as much as possible, known as upstream pipeline tasks.

Figure A.4 shows how data quality can be improved in two steps by using the upstream pipeline; i) AI1, improving the quality of each node's data streams to obtain enriched data, and ii) AI2, using data streams from different nodes to improve data quality, taking into account the consistency of the data with respect to the rest of the data measured by the platform. Therefore, data quality can be improved either by using data from each node individually or at the network level by using data from different nodes. The use of these blocks (AI1 and AI2) is complementary, i.e., both or only one of them can be used, depending on the application and the tasks to be performed at different levels. Regarding the improvement of raw data, phase AI1, the tasks to be performed are focused on the improvement of the data using the data streams available in each of the nodes. For instance, a node can measure temperature, relative humidity, and rainfall at the same time, where the information from each of the phenomena can help to detect or correct the data from the others. Among the different tasks for improving the quality of the raw data some examples include (*González-Vidal et al.*, 2019; *López et al.*, 2022):

- *Denoising:* signal denoising can be introduced to reduce signal-to-noise ratios and to smooth out the data to feed the tasks that follow with more accurate data. Examples include signal processing filters, e.g., low-pass filters, and multivariate methods such as principal components analysis (PCA).

- *Sensor calibration:* might be a necessary task when IoT nodes include low-cost sensors. In this situation, it is common practice to collocate the low-cost sensors with a high-precision instrument that provides ground-truth measurements to perform an *in-situ* calibration using supervised machine learning techniques.

- *Missing value imputation:* may be required when data losses occur, e.g., due to communication or computational failures. At this stage, univariate time series methods can be used to interpolate the data gaps as well as multivariate methods that make use of all streams available in an IoT node, e.g., matrix completion, multiple imputation by chained equations (MICE).

- *Data fusion and proxy creation:* can be used to combine data streams measuring a specific phenomenon to produce a more accurate estimate. For instance, one can combine the measurements of different low-cost sensors to obtain a more accurate estimation. Techniques include the use of supervised machine learning models and ensemble and consensus algorithms among others. In the case of proxy creation, an unmonitored phenomenon can be modeled from a set of indirect measurements, e.g., using a supervised machine learning model.

- *Outlier detection:* the detection of outlying and anomalous data can be critical to avoid feeding applications with erroneous data and to avoid retrieving misleading information. To this end, outlier methods such as z-score or empirical-cumulative-distribution-based outlier detection (ECOD) can be used.

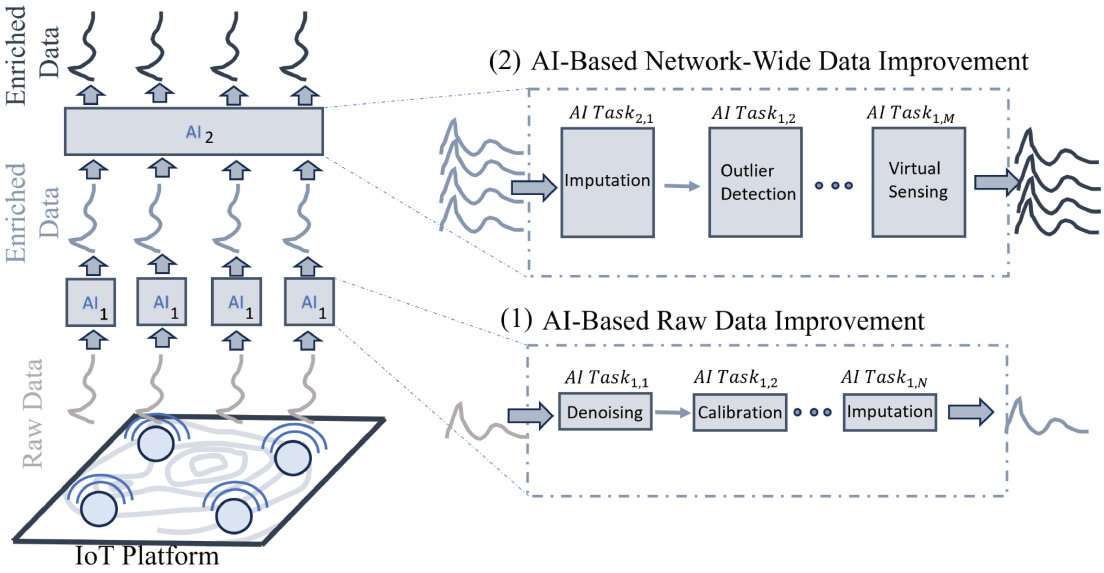


Figure A.4: The quality of data collected by an IoT platform can be improved by going through a series of AI-driven tasks, called upstream pipelining. Specifically, a distinction can be made between AI-driven enhancement of raw data and AI-driven enhancement of network-wide data.

In the case of the network-level data enhancement phase, AI2, the tasks use data streams provided by different nodes in the network. For instance, in the case of a soil moisture monitoring network, data from nodes in nearby locations or data from correlated nodes can be used to improve the quality of these data. Some tasks that can be used to provide data at the sensor network level with improved quality include (*Khampuengson & Wang*, 2023; *Shahra & Wu*, 2023):

- *Missing value imputation*: this task can be used to perform the data imputation using available data from the monitoring network, i.e., benefiting from the correlations present in the network. Techniques such as matrix completion, MICE, missForest, and others can be used.
- *Outlier detection:* can be used to detect incoherent or anomalous data using the measurements of the network. The methods that can be used at this level are similar to the ones already seen but are limited to multivariate such as ECOD, autoencoders, etc.
- *Virtual sensing:* sometimes it is necessary to obtain measurements at a location where a sensor is not available. In this situation, one can create a virtual sensor at the target location using the measurements from other sensors. For this purpose, supervised machine learning techniques as well as spatial interpolation techniques can be used.
- *Sensor placement:* to choose the optimal location of the different sensors and meters to monitor a phenomenon. Common techniques used for sensor selection and placement include the use of convex and nonconvex optimization problems.

The tasks and techniques to be implemented are not limited to the listed ones. The implementation and design of an upstream pipeline for an IoT platform is completely modular and dependent on the downstream applications to be performed with the data and their requirements. This means that the two different levels can be implemented with different tasks, a single level, or even replicate the same task at two levels, e.g., detecting anomalous data first using only node data and then network data. Moreover, since most techniques are data-driven, i.e., they may require training prior to their application, there are different computational paradigms to train and apply the different techniques (see section 2.4). Finally, it is important to mention that in applications where reliability is critical, apart from enriching the data, it is also possible to annotate it, i.e., to add metadata indicating which actions have been performed, with which methods and the confidence in the actions. In this way, properties such as data traceability and interpretability are maintained.

3.2 Downstream applications

AI has revolutionised many sectors, providing tools for obtaining relevant information and for the automatic assessment and control of infrastructures. Modeling and prediction of phenomena, event detection, and automatic control are among the many other applications of AI in the field of smart water systems (*Doorn*, 2021). In the case of modeling, there is the monitoring of phenomena. This can involve direct or indirect measurements using proxies and virtual sensors. The development of proxies is part of the upstream pipeline. An example of modeling is the monitoring of water quality from other parameters such as salinity and temperature. Forecasting can follow monitoring phenomena, where time series techniques and deep learning technologies are used to predict different phenomena over time. For example, recurrent neural networks can be used to predict water demand (*Kavya et al.*, 2023).

Event detection applications include automatic irrigation or early warning systems to mitigate the effects of natural disasters (floods, earthquakes, volcanic eruptions, forest fires, extreme pollution, etc.). Supervised classification techniques as well as supervised and unsupervised anomaly detection algorithms can be applied. Finally, in the case of the automatic control of intelligent systems, all the above applications and prescriptive tools can be used to support an automatic decision-making process (*Jenny et al.*, 2020). For example, computer vision techniques can be used to monitor crop yields in precision farming applications. Another application that is emerging and will be critical in the coming years is the creation of digital twins, where an IoT network provides measurements of a physical phenomenon that are fed to its virtual counterpart for characterisation. Many applications can be realised from digital twins, such as smart irrigation systems (*Manocha et al.*, 2024). In short, although a comprehensive review of IoT-derived applications is beyond the scope of this chapter, the goal of applications and the use of AI is to obtain valuable information, optimise decision-making processes, reduce operating costs, increase energy savings, and provide an optimised service to users.

4 AIoT use cases and examples

The upper part of the unsaturated zone is subject to fluctuations in water content and solute concentrations due to the mechanisms of evaporation from the soil surface, infiltration and deep percolation, and water uptake by plant roots (*Hopmans*, 2011). During periods when precipitation is greater than evapotranspiration, some of the excess water contributes to an increase in soil moisture storage, resulting in a rise in the water table. This increase can lead to waterlogging and flooding in both rural and urban areas. A lumped water balance summarises this: $\Delta S/\Delta t = P - E - I - (Q \pm q)$ (Figure A.5(a)), where ΔS is the change in soil moisture, t is time, P is precipitation, E is evapotranspiration, I is deep percolation, q is groundwater discharge/recharge due to the river-aquifer interaction, and Q is runoff, (*Rodriguez et al.*, 2008). For floods large enough to pose a threat to public safety, the soil is typically saturated, resulting in negligible values of E and I with minimal changes in ΔS. This highlights the need for observational data to estimate the conversion of precipitation to runoff and ultimately to river flow and water levels at critical points. However, not all high water table scenarios are negative. In arid regions with low groundwater levels, water storage in both unsaturated and saturated zones becomes critical to meet crop needs (*Nosetto et al.*, 2009). In water management, high soil moisture and groundwater levels are associated with an increased risk of flooding or higher crop yields (*Kuppel et al.*, 2015; *Houspanossian et al.*, 2023).

Regardless of the specific approach (hydrological or sensor calibration problem), regression models such as Multiple Linear Regression (MLR) are used when data are linear, or K-Nearest Neighbours (KNN), Support Vector Regression (SVR) and Random Forest (RFO) when data are non-linear (*Ferrer-Cid et al.*, 2019; *López et al.*, 2022).

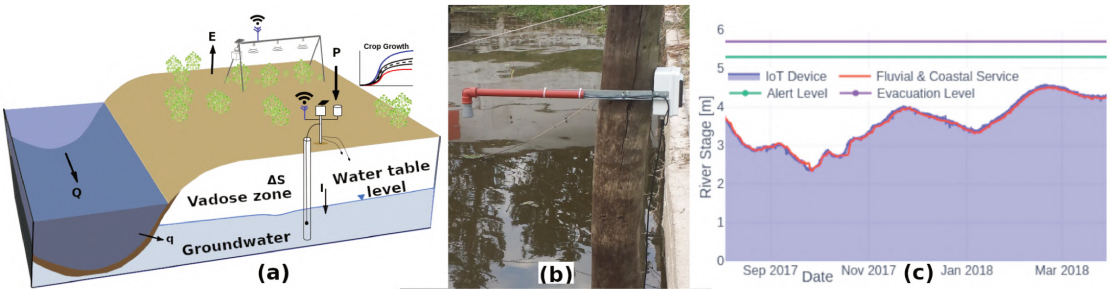


Figure A.5: (a): Integrated IoT system for monitoring agro-hydrological parameters for modelling crop growth curves or other applications. (b): Maxbotix MB7092 acoustic sensor in the Setúbal Lagoon (see Figure A.6(a)); (c): Acoustic sensor measurements vs. FCS readings (R^2=0.9949).

4.1 Case A: flood early warning systems applications

Almost all FWSs use gauges and communication tools to collect and distribute hydrometeorological measurements. The information is used in real time to predict whether a flood will occur, when it will reach critical points in the catchment and how severe it will be. Early flood warning involves several detailed sequential steps: detection, transmission, retrieval, forecasting, warning and transformation (see *López et al.* (2020)). The warnings are then transformed into mitigation actions.

Example 1: flood early warning in the lower Salado River catchment (Santa Fe, Argentina)

Following a major flood of the Paraná River in 1992, the Santa Fe Province invested more than €65 million in flood control works throughout the state. The city of Santa Fe is wedged between the confluence of the Salado River to the west and the alluvial system of the Paraná River to the east (see Figure A.6(a)) and is located at 31 degrees south latitude. The city's defences covered not only the eastern part along the Paraná River, but also the Salado River on the western side. The main defence against flooding from the Salado River was the western dike, which stretched along the city boundary for about 7 km and was built 5.2 meters above the floodplain (see Figure A.6(b)). Although the western levee was built in two stages, the third stage was never completed, leaving a small unfinished gap formed by piles of bare earth reinforced with sandbags. In the first quarter of 2003, unusually high rainfall in the province of Santa Fe saturated the lower Salado River basin. Between 21 and 29 April, more than 400 mm of rain fell in some areas, causing the river to burst its narrow and meandering channel, destroying bridges and roads, isolating small towns on the way to its outlet, and flooding a third of the city of Santa Fe (*Vionnet et al.*, 2006a).

The lower Salado River FWS was established after the 2003 flood in the city of Santa Fe (Figure A.6(c)). Frequent loss of information due to communication problems with the storage site increased the operational vulnerability of the network. To address this problem, a low-cost alternative solution using affordable equipment was proposed (*López et al.*, 2020). The upgrade included an interface that simultaneously transmitted network data via an electronic bypass without interfering with the FWS satellite communication system. Tests carried out at three stations showed that the low-cost interface was efficient in transmitting data via WiFi or cellular in real time. The results established a one-to-one data correspondence between the official system and the redundant system. As a result of this learning process, a low-cost data logger was designed and manufactured to measure hydrometric levels using a low-cost acoustic sensor. Laboratory experiments were carried out and the device was then installed at a yacht club in the city of Santa Fe, where the periodic fluctuations of the Paraná River are observed daily.

Example 2: hydro-environmental monitoring of the west urban natural reserve (Salado River, Santa Fe, Argentina)

The city of Santa Fe faced challenges not only from river flooding on the Salado and Paraná Rivers sides, but also from waterlogging during high-intensity storms. The city began building an extensive network of drainage channels and reservoirs to collect rainwater. Pumping stations regulate water levels in these peripheral reservoirs and release excess water into surrounding rivers. Two of these reservoirs together form the West Urban Nature Reserve (WUNR), which covers 142 hectares (see Figure A.6(b)). An interconnected network of channels drains the urban catchment area, of approximately 1,700 hectares in size, that discharges into the WUNR.

The quantity and quality of water entering and remaining in the reservoirs posed a risk to the local population and the integrity of the receiving water bodies. To address this challenge, a system for continuous, open-access monitoring of hydrological and

Figure A.6: (a) Santa Fe City (Argentina) map with the location of several references used in this work; (b) aerial view of the west urban natural reserve; (c) the FWS network in the lower basin of the Salado River (SFe Prov, Argentina).

water quality parameters was designed and implemented using IoT and low-cost sensors. The new network is part of a hierarchical flood warning and water management system for the city. The final operational network covers the inner canal tributaries, the water bodies within the WUNR and the catchment sectors (see Figure A.7(a)). Groundwater levels are recorded in the upper, middle and lower parts of the catchment using the data logger and low-cost sensor used previously. Water levels at the entrances to two main sewers and the primary reservoir are also monitored, along with precipitation in the contributing catchment. Preliminary analysis shows satisfactory results, with a relative Root Mean Square Error (RMSE) = 7.9% and R^2=0.84, when using the Random Forest model to reproduce the surface elevation at the outlet of the urban catchment, using cumulative precipitation as a predictor variable with different time lags (see inset Figure A.7(b)).

4.2 Case B: agro-hydrological applications

The importance of alfalfa (*Medicago sativa* L.) as a forage crop for livestock production is widely recognised. This perennial legume is grown on approximately 30-32 million hectares worldwide, with the USA accounting for 21% of the total area. Argentina remains one of the largest alfalfa producers in the world and the first in South America. Crop yield is closely linked to the amount of rainfall and soil moisture for plant uptake, groundwater depth and cumulative temperature. It follows two examples with the same IoT system plus AI for the continuous monitoring of parameters in an alfalfa field (Figure A.8).

The presence of water in the soil causes the electrical permittivity to vary, thus changing its capacity. The SKU: SEN0193 (DFRobot) is a very low-cost capacitive soil moisture sensor. Its output is an analogue value that is inversely proportional to soil moisture. The HydraProbe II sensors (Stevens Water Monitoring Systems, Inc., Portland, OR, USA) measure electrical permittivity using a variant of the Time Domain Reflectometry (TDR) technique. Unlike the capacitive sensor, the output is digital (SDI-12

Figure A.7: (a): Instrumentation of the WUNR (see aerial view in Figure A.6(b)) at various points along the urban catchment. The outlet of the main drainage channel is also shown. (b): from top to bottom, the temporal evolution of the water level at the reservoir, at the drainage outlet, the precipitation and the groundwater level in the urban catchment and at the entrance to the reservoir. The incept shows the water level predicted by the random forest model.

protocol) and directly represents the volumetric water content (VWC). The HydraProbe II sensor was therefore used for calibration as a reliable measure of soil moisture (see the deployment of both sensors in Figure A.8(a)).

Calibration (see Section 3.1) refers to the process of correcting for systematic errors in sensor readings, often by comparing a reference measurement from a first instrument with an uncalibrated measurement from a second instrument in order to adjust the parameters of the second instrument to provide an accurate estimate. Machine learning algorithms were used to calibrate low-cost soil moisture sensors for soils of different textures. At the same time, soil and ambient temperature and relative humidity were measured to assess their influence. Using the HydraProbe II sensor as a reference, these variables were used as predictors. Figure A.8(b) shows the results obtained with the calibration technique compared to the VWC of the reference sensor. The implemented linear and non-linear machine learning techniques gave a relative RMSE ranging from 2.1% to 13.2%.

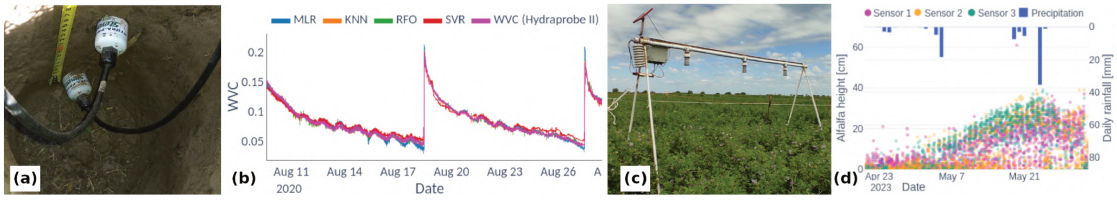


Figure A.8: (a): colocation of sensors in the field; (b) Predictions from machine-learning models superimposed on reference soil-moisture sensor records; (c) agro-IoT station recording alfalfa height; (d): raw data on alfalfa height.

The same IoT device was used to collect precipitation, temperature, groundwater depth and crop height data over time in the alfalfa plot shown in Figure A.8(c). The ul-

trasonic distance meter was used to record alfalfa growth. By applying anomaly detection and smoothing algorithms, a logistic type experimental growth curve was obtained, which is consistent with the accumulated experience in this regard (Figure A.8(d)). To assess crop yield, the forage is periodically cut to estimate biomass production (dry weight/area).

5 Concluding remarks

The main elements that make up an IoT system for measuring water-related phenomena have been summarised. Then, the use of AI in the data coming from IoT-based smart water systems has been described, emphasising the use of AI techniques to improve data quality in the so-called upstream pipeline. This is crucial to provide reliable data to the applications that feed on this data, called downstream applications. From this, the main tasks and AI techniques that can be used have been reviewed, as well as possible downstream applications that have an increasing impact today. Finally, in order to put all the elements described into context, two examples of use cases in which AIoT plays a key role have been shown. Overall, IoT and AI technologies can be combined to create a reliable platform with optimised maintenance, enabling new applications and providing a cost-effective alternative for smart water systems.

Apéndice B

A cost-effective redundant communication system for improving the reliability of a flood early warning system

Resumen: Los sistemas de alerta temprana de inundaciones funcionan con estaciones de medición y comunicación para registrar y transmitir los datos hidrometeorológicas colectados en tiempo real. El objetivo es permitir que los tomadores de decisión y autoridades lleven a cabo acciones antes que una crecida llegue a lugares críticos. El sistema de alerta hidrológico (SAH) de la cuenca inferior del río Salado, Argentina, se conformó después de la inundación de la ciudad de Santa Fe en 2003. La pérdida frecuente de información debido a problemas de comunicación satelital aumenta la vulnerabilidad operativa de la red. Primero, se analiza la frecuencia de inundaciones de la región utilizando distribuciones estadísticas conocidas. Luego, se ofrece una solución de bajo costo al problema de la pérdida de información basada en una interfaz de software que transmite en paralelo los datos de la red a través de un *bypass* electrónico sin interferir con el sistema de comunicación satelital del SAH. Las pruebas realizadas en tres estaciones muestran que la interfaz resultó eficiente en la transmisión de datos vía WiFi o comunicaciones móviles en tiempo real. Los resultados establecen una correspondencia uno a uno de los datos cuando el sistema satelital y el sistema redundante están activos. La interfaz está basada en computadoras de placa única y routers de bajo costo. La variedad de sensores comerciales probados hace que la interfaz desarrollada sea portable a situaciones similares, y los ahorros logrados en costos de mantenimiento son significativos[1].

[1]Este artículo fue publicado en el Journal of Hydroinformatics (2020) 22 (4): 856–875.

Abstract: All early flood warning systems (FEWS) work with gauges and communications systems to collect and distribute hydro-meteorological measurements in real-time. The objective is to allow the authorities to take action before the flood reaches critical places. The FEWS of the lower Salado River basin, Argentina, were established after the flood of Santa Fe city in 2003. Frequent loss of information due to communication problems with the repository site increases the operational vulnerability of the network. First, the document analyses the flood frequency of the region using known statistical distributions. Then, it offers a low-cost solution to the loss of information problem based on an interface that transmits in parallel the network data through an electronic bypass without interfering with the FEWS satellite communication system. The tests carried out along three stations show the interface proved efficient in transmitting data via WiFi or mobile communications in real-time. The results establish a one-to-one data correspondence when the satellite system and the redundant system are active. The interface is based on low-cost single-board computers and routers. The variety of commercial sensors tested makes the developed interface portable to similar situations, and the savings achieved – maintenance cost – are significant.

1 Introduction

Nowadays it is common to find a wide variety of early warning systems, ranging from assessing the financial risk of bank conglomerates (*Faranda et al.*, 2015) and famine spreading by drought (*Senay et al.*, 2015) to volcanoes eruptions (*Gregg et al.*, 2015) and the mitigation of natural disasters effects (*Balis et al.*, 2013; *Calvello et al.*, 2015). Nevertheless, it is the management of flood hazards one of the most frequent and critical components of public safety due to its usually disruptive impacts on public health and infrastructure. (*Plate*, 2007; *Chen et al.*, 2014). Therefore, the establishment of Flood Early Warning Systems (FEWS) has been very active in the last decades to help decision makers deal with hazardous floods (*Garrote & Bras*, 1995; *Asante et al.*, 2007; *Henonin et al.*, 2013; *Verdin et al.*, 2016).

As defined by *Plate* (2007), a FEWS is a set of actions oriented to protect lives and properties whenever the river free surface elevation attains a pre-established threshold value. Almost all flood warning networks work with gauges and communications system to collect and distribute hydro-meteorological measurements. The information is used in real time to predict whether a flood is about to happen, when it is going to arrive at some critical points within the watershed, and how severe it is going to be.

Floods triggered by heavy rainfalls are of considerable interest in urban regions bounded or crossed by large or medium size rivers (*Plate*, 2007; *Fakhruddin et al.*, 2015). As such, a FEWS is irreplaceable for local authorities since it allows them to take proper actions before the flood arrives (*Asante et al.*, 2007; *Henonin et al.*, 2013).

Consequently, an early flood warning embraces few detailed sequential steps (*Plate*, 2007): capturing, transmitting, retrieving, forecasting, warning, and transforming. The last two stages comprise warnings to local decision makers and the population in general and then transforming the warnings into remedial actions. Figure B.1 depicts the subsequent phases that make up a FEWS.

Flood warning systems that integrate rainfall monitoring with flood forecasting are now operational in many countries all over the world (*Henonin et al.*, 2013). Indeed,

any FEWS has evolved from fully integrated flood management tools involving short-term hydrological forecasting (*Velázquez et al.*, 2009; *Merkuryeva et al.*, 2015) to a stand-alone system that takes in-situ measurements to warn their clients via SMS (short message service) (*Kuantama et al.*, 2013). The relevance of an operational FEWS becomes evident if the potentially avoided flood damages are computed, yielding an approximated benefit of 400€ for every invested € (*Pappenberger et al.*, 2015).

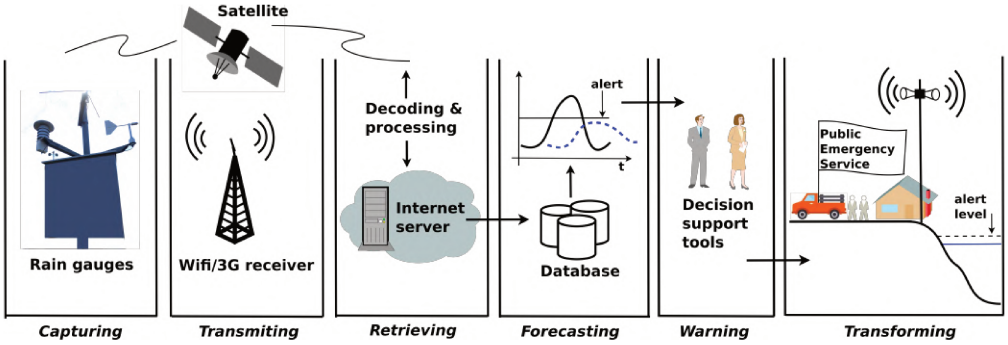


Figure B.1: Sequential steps that make up a FEWS (see, e.g. *Plate* (2007); *Henonin et al.* (2013)).

However, the purpose of this work is not to review the state of the art underneath the operation of any FEWS, nor going through the latest developments in technologies to ease the communications between the data capture and the required short-term flood forecasting (*Guo et al.*, 2012; *Pumo et al.*, 2016). Instead, this paper deals with the operation of the FEWS located in the lower watershed of the Salado River in Midwest Santa Fe Province, Argentina, and a cost-efficient solution to circumvent part of its transmission losses.

The Santa Fe City implemented its early warning network after the catastrophic flood suffered in 2003 (*Vionnet et al.*, 2006b). Nevertheless, and among other persistent problems, the network has been hampered by communication problems ever since. As a possible solution to this shortcoming, this work proposes an alternative way to transmit the data captured by the monitoring network without interfering with the proprietary hardware of the functioning FEWS, given the frequent missing of data. Although the implemented solution relies on affordable open source components, it is not similar to the alternative followed by *Sadler et al.* (2016a), who resorted to Arduino platforms to design a low-cost data logging and transmission system for environmental sensors connected to a network. The solution consists of installing an interface that runs on a minicomputer coupled with a router to transmit the data through WiFi or mobile communications. The interface then accesses the data utilizing an electronic switch that avoids interfering with the satellite communication system embedded in the FEWS.

Thus, and according to Figure B.1, the authors strictly restrict the discussion to the communication phase of a working FEWS; the low-cost solution found to circumvent the problem, and the test performed to check if the data packages transmitted by the alternative channel were indeed unaltered.

The aspects that contain the occasional or potential failure of a flood warning system are often minimized. An early and accurate warning is what defines an adequate warning network. In this work, a detailed field survey of more than seven years is presented,

involving the direct action of some of its co-authors, who visited the Salado River Flood Warning Network several times, documenting the occasional and specific failures of each monitoring station. The survey is summarized in this paper, providing a clear picture about which item of the network fails most frequently. To the best of the authors knowledge, there are few detailed statistics spanning seven years that document the failures of a flood warning system.

The survey triggered the need to find a technical solution, once it was detected that communication blockage was the most frequent failure. Therefore, the way to solve the problem of an operational Flood Warning Network, using pieces of hardware of common use combined with free software, is also a novel aspect if one takes into account that it was addressed from the academic, and not from the professional field. However, it is well-known that equipment developed by private industry for environmental applications is usually based on patented technology that manufacturers lock away from the public reach. As a result, the user is soon in a somewhat uncomfortable situation due to the impossibility of introducing changes in the hardware or software that controls the equipment.

In tune with the problems mentioned, one of the main contributions of the present work was the developed software to automatically emulate the manual process that operators generally perform in front of the station when they download data, without interfering with the normal operation of the network. The operator interacts with a Graphical User Interface (GUI) using images rather than commands executed line by line. The selected library (PyAutoGUI) allowed controlling, through a Python script, the mouse and keyboard to automate the interaction with the proprietary GUI software and thus emulating any desired manual operation. Such methodology was previously tested against several commercial devices such as a Campbell® station, Steven® and Spectrum® soil moisture sensors, and a Genica® phreatic level sensor and a Pegasus® station (the earlier three made in the USA, and the former two made in Argentina). The variety of commercial sensors and the scenarios tested make the proposed development easily transferable to other situations with restrictions of similar characteristics (closed software, lack of backup communication system). In this aspect, the solution proposed here is in tune with the work of *Sadler et al.* (2016a).

In summary, the novelty of this work relies on three well-separated aspects; first, a detailed field survey of more than seven years detected that communication blockage of the flood warning system was the most frequent failure of the network. Second, the automatic manipulation of a GUI is a technical contribution that emerged due to the impossibility of accessing the manufacturer's application of each data unit of the network, with an end product that sends the data with incredible fidelity. Last but not least, the testing of the methodology with different devices and environments ensures a trouble-free and transferable product, making it attractive to similar systems or situations.

The remainder of the paper describes the main floods of the Salado River throughout the 20th century to present days, including a brief description of the damage suffered by the two bridges that cross the river since their construction in 1972. Then, the document continues with the method developed to solve the problem mentioned above. It includes a description of the FEWS of the Salado River, and the architecture of the proposed alternative to transmitting the data of the network without using the satellite communication system. The following section presents the results, which are mainly based on the comparison between several data packets sent simultaneously through the

satellite and redundant communication systems. The paper ends by discussing the results and the far-reaching implication of the proposed solution. Finally, and to keep text to a minimum, the previous research conducted on commercial devices is summarized in Appendix 7. There, it is possible to find out how the developed interface interacts automatically with the proprietary GUI software of some devices. All developments are freely available to the reader in a Web repository.

2 Material and methods

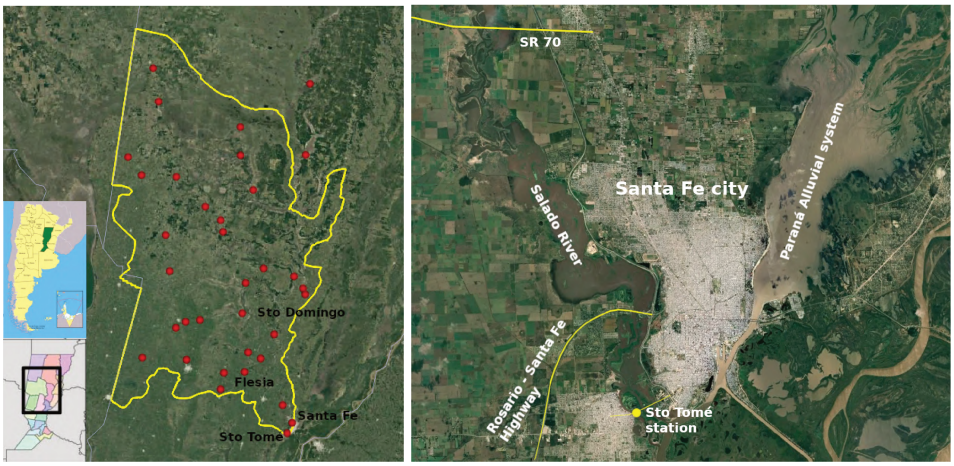
2.1 The major floods of the Salado River, Santa Fe, Argentina

The design of a FEWS requires the estimation of maximum water levels attained for floods corresponding to given return periods. This information is crucial for the city authorities and people located on the floodplain, along the riverbanks. Nonetheless, it is not the purpose of this work to carry out an exhaustive statistical analysis of the extreme floods of the Salado River. This section offers a brief account of the historical floods of the river, with an overview of the frequency distribution of the water heights. The flood of 2003 caused, for the first time, loss of human life besides severe damage to the public and private infrastructure. The official death toll was 23 (*Vionnet et al.*, 2006b). The serious consequences of the flood were the reasons behind the decision that pushed forward the local authorities to establish an early warning system.

The Salado River crosses several provinces of Argentina. It flows for about 1150 km from the northwest of the country to its central part, near Santa Fe City (Figure B.2a. The city borders on the west with that river, and to the east with the Paraná alluvial system (Figure B.2b). Table 1 shows the great floods of the Salado River that affected the city in the last 100 years. The listed floods were generated in the lower watershed of the Salado River (Figure B.2a. The region comprises a relatively flat area of approximately 30,000 km^2. This region is located in the mid-west of the Province. In the last fifty years, the basin has undergone modifications in land use, as well as the development of a wide network of drainage channels. Areas with frequent flood risk have been occupied by the population over the years, increasing the vulnerability of the region. Despite these periodic setbacks, the lower river watershed supports a healthy economy that is vital for the region and the country.

The Santa Fe-Rosario Highway (Province of Santa Fe) was inaugurated at the end of 1972. It was necessary to construct two bridges to cross the Salado river, stretching 157 m each in a place where the alluvial valley was 2000 m wide (Figure B.2b). At that time, the wetted cross-section was about 500 m^2 under the bridges. Both structures were destroyed during the 1973 flood. Surveys carried out after the flood revealed that the cross-section had tripled its size, with erosions deeper than 5 m from the initial bed level. The bridges were rebuilt with a pile foundation placed at a much greater depth but still with the same span, 157 m. The flood of 2003 again affected the bridges, collapsing one of the abutments (Figure B.3a). The bridges were rebuilt once again, this time stretching a total of 578 m (Figure B.3b).

It is possible to infer the statistics of extremes floods of the Salado River by looking at the 1953–2019 daily water level data. The records correspond to the Santo Tomé station by the Salado River outlet at the Paraná alluvial system, and the cross-section



(a) (b)

Figure B.2: The Salado River, Santa Fe, Argentina. (a) The network and limits of the lower watershed; (b) the Santa Fe-Rosario Highway (satellite imagery source: Google Earth).

SR70, located 25 km upstream (Figure B.2b). The data set comprises of 22,327 readings of the river stage, which were converted to water surface elevation above sea level (the official datum of the Argentine Republic (IGN: www.ign.gob.ar/)). The 67-year long record has missing readings between the years 1988–1992.

Figure B.4 shows the histogram of the water level readings at both stations. The histogram describes not only the spread of the data but also the disparate behavior of the water surface elevation at both sites. The hydrology regimes of the Salado and the Paraná rivers are independent. The histogram of the water height at the Santa Tomé station, located nearby the Salado River mouth, is affected by the backwater effects induced by the water surface elevation of the Paraná River. Consequently, it is unable to spot the two peaks of the bi-modal distribution of the SR70 station, which is 25 km upstream. These two peaks correspond to periods of low water, or dry seasons, and medium to high water, or periods of wet seasons. However, the relevance of the Santo Tomé scale to the warning system is crucial. The failure of the protection levee, built to protect people living in lowlands from the periodic flooding of the Salado River, was so dramatic during the 2003 flood that at one point the water was 2.48 m higher within the city than on the riverside, near the location of the Santo Tomé scale (*Vionnet et al.*, 2006b).

Finally, the available data set allowed the estimation of the return period of the 2003 flood using traditional two- and three-parameter distributions (Appendix 7).

Table B.1: Major floods recorded on the Salado River, Santa Fe, Argentina

Year	1914	1973	1977	1998	2003	2014	2016
Q(m^3/s)	2750	2451	1701	2513	3980	1747	2080

(a) (b)

Figure B.3: Bridges of the Santa Fe-Rosario Highway over the Salado River during and after the 2003 flood. (a) Collapse of the westward abutment; (b) Reconstruction after the flood.

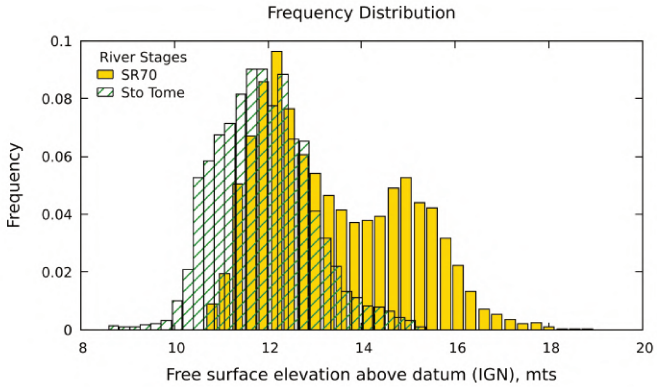


Figure B.4: Histogram of river stages at the Santo Tomé station and the State Road 70 (SR70) station. The zero level for both scales is 8.07 m for the Santo Tomé station, and 11.09 m for the SR70 station.

2.2 The FEWS at the lower watershed of the Salado River, Argentina

After the catastrophic flood suffered by the area and the Santa Fe City in 2003 (*Vionnet et al.*, 2006b), the Government of the Santa Fe Province called for the implementation of a FEWS, a bid awarded to a joint venture between a Buenos Aires-based company and a firm in charge of the hydro-meteorological measurements nationwide (*Tecmes & Evarsa*, 2020).

The network initially consisted of 41 stations, with only 36 fully operational until now (Figure 2(a)). Among them, 24 measure river stages, nine rainfall and soil moisture, and five rainfall, soil moisture, relative humidity, atmospheric pressure, phreatic levels, solar radiation, evaporation, and wind speed (*Contini G. F. et al.*, 2014).

Each station has an outer enclosure which protects an inner cabinet that complies with the IP65 standard (water-dust resistant) and hosts the datalogger with the Remote Transmission Unit (RTU), the wiring of the terminal blocks, the satellite module transmitter and the power system (battery, charger, and voltage regulator). In turn, each unit also has a service port allowing in-situ manual management, so providing access to data already stored in the datalogger with the aid of the proprietary software.

All stations, either fed with solar energy or from the power grid, measure the state-of-charge of their batteries, log the information internally, and whenever there is storage capacity available, upload and transmit the collected data to the final government repository site through the Orbcomm satellites (LEO – low Earth orbit satellites operated by the company Orbcomm, www.orbcomm.com).

Failures in any FEWS are mostly due to flaws in the functioning of sensors and dataloggers, as well as a power outage and transmission losses or environmental factors (e.g. lightning strikes). Vandalism is another source of partial or total failure. However, the incidence of any factor does not always end in complete data loss.

In other to analyze the occurrence of those factors, the FICH at the UNL and the Ministry of Public Works of the Government of Santa Fe Province signed an agreement for surveying the network performance for the period 2008–2013 (*Macor et al.*, 2016). It was possible to totalize 2573 visits, with an average of 11.9 per station per year (Table 2, *Contini G. F. et al.* (2014)). The lower number of visits in 2011–2012 was due to insufficient funds to afford the field trips.

Table B.2: Number of visits to the Salado River FEWS network (2008–2013)

Year	**08**	**09**	**10**	**11**	**12**	**13**
Visits	538	502	449	309	303	472

During the field campaigns, the FEWS' companies supplier (*Tecmes & Evarsa*, 2020) and the labor inspectors (*Macor et al.*, 2016) agreed to fill out several surveys throughout the elapsed time of the consulting contract, and so they obtained a database organized according to the detected fault of each station:

- sensors: when observing unstable readings or losses of the parametrization settings,
- dataloggers: due to flaws in the proprietary software or connection problems which nullify the data reception,

- power outage: related to missing or broken panels, or to a faulty voltage controller of a battery close to its useful life,
- transmission losses: triggered by missing or broken antennas, a coaxial cable deterioration, or with the Orbcomm system,
- bugs: due to the presence of bees, wasps, ants and nests that could damage the electronic components,
- weather: produced as a result of some extreme weather events (hail, lightning, or high winds), and
- vandalism: damages provoked by antisocial behavior (e.g. gunfire, robbery, or fire).

Table B.3 shows the number of failures observed in the years 2008–2013. Figure B.5 depicts a stacked frequency histogram, which makes it easier for the quantitative comparison among the contributing factors to failure. The most frequent failure was the transmission losses, with an annual average of 22.6% occurrences. This flaw manifests itself as a delayed reception of the information or the eventual absence of the data package in the final repository site. The frequent repetition of this failure determines the operational vulnerability of the FEWS network, a critical aspect when the watershed is under extreme weather conditions or heavy rainfall (when certain thresholds are exceeded, and the network must activate warning alerts to local authorities).

The second factor affecting the FEWS was the presence of bugs, with an average frequency of 20% followed by the failure of sensors with a mean of 16%. The losses attributed to vandalism were steady and around 10% yearly.

Despite the significant drop in transmission losses observed in 2013 (only 6% of the total, Figure B.5), an extension of the consulting study with the Ministry of Public Works allowed the testing of this issue once again for the whole network during 2015. Figure B.6 depicts, in solid lines, the continuous transmission of the referenced station, whereas the blanks indicate periods of silence with no link between the station and the Orbcomm satellites. The Orbcomm system is in charge of uploading and forwarding the information to the final repository site. The loss of information due to transmission problems was still there.

Table B.3: Factors affecting the Salado's River FEWS (2008–2013)

Failure	**08**	**09**	**10**	**11**	**12**	**13**	**Total**
Sensors	13	16	7	4	12	18	70
Datalogger	11	7	5	3	7	4	37
Power	5	7	14	16	10	22	74
Bugs	10	16	11	14	8	29	88
Atmosphere	2	8	2	0	0	2	14
Vandalism	8	8	7	8	6	9	46
Transmission	36	17	14	11	13	5	96
Total	85	79	60	56	56	89	425

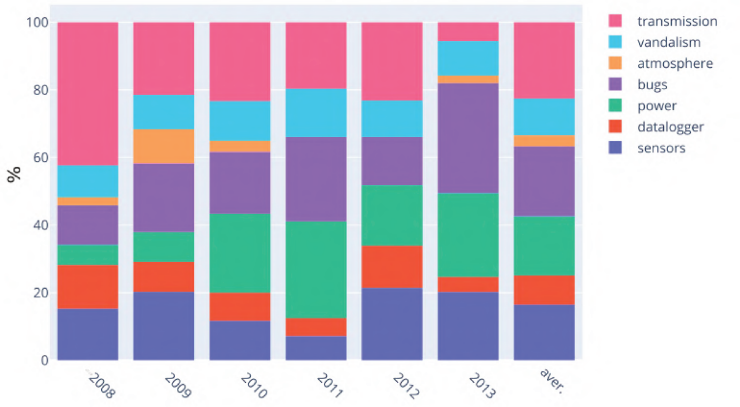


Figure B.5: Frequency of factors affecting the FEWS performance.

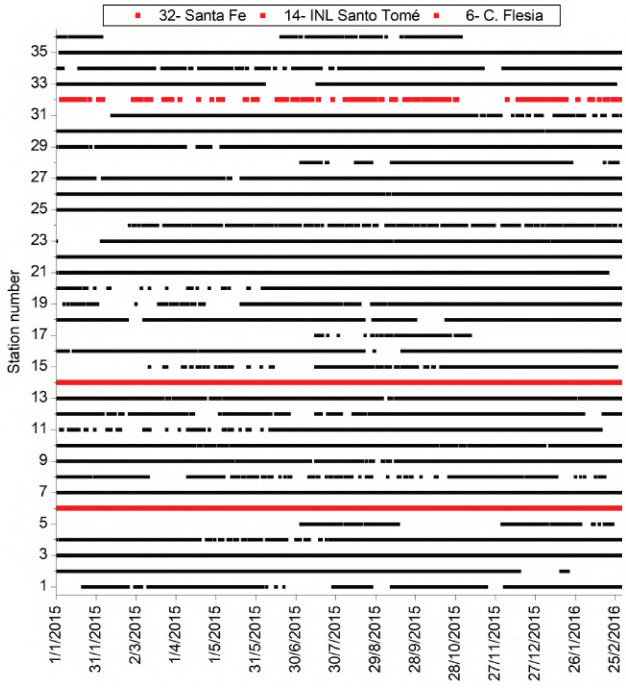


Figure B.6: Signaling of 36 stations of the Salado River's FEWS network during 2015.

Functioning of the FEWS'network

Each station of the network acquires the hydro-meteorological data through a current loop working in the range 4–20 mA or 0–4000 mV of voltage. The readings go in through a channel datalogger and convert to digital within the RTU according to a previously stipulated agenda. Then, a fixed and removable internal memory stores the data. If the measured value is above a preset risk threshold value; it broadcasts an immediate alarm. Otherwise, it is put on hold until the time for its transmission is attained.

The feeding of the Tecmes® acquisition and recording unit (RTU) depends on the installation site. The two options are: 1) a 12 Vdc battery to 220 Vac with automatic charger; or 2) a 12 Vdc battery with charge controller and solar panel. This unit is responsible for managing the data acquired by sensors, and transferring them to a communication module satellite of the Orbcomm network.The module continues to receive data until it has communication with the satellite.

The data uploaded to the satellite transponder is stored until there is communication with the Earth station in charge of managing the Orbcomm network. At that moment, the information is sent to the servers, and from there, to the data recipients through a directory with the clients' email addresses. Then, an email enters the central station of the FEWS through a specific software that decodes the information, which deploys the value of each remote station on a watershed map.

A low-cost interface to solve the transmission lost shortcomings

An interface software which runs on some added hardware was installed on three selected stations of the network to upload the measured data to an Internet server without affecting the pre-established link between the nodes and the repository site.

The expected operations to be executed by the interface were: (i) to set up or modify the working parameters for each installed sensor device, and (ii) to upload the recorded data to an Internet server which stores the measurements of the monitoring station.

Towards that aim, the work was broken down into successive stages: 1) conditioning the added hardware, 2) interacting with the proprietary software, 3) automating the download of historical or real-time data and reconfiguring sensors, 4) submitting data to the web through a VPN (virtual private network), and 5) viewing the data on a website.

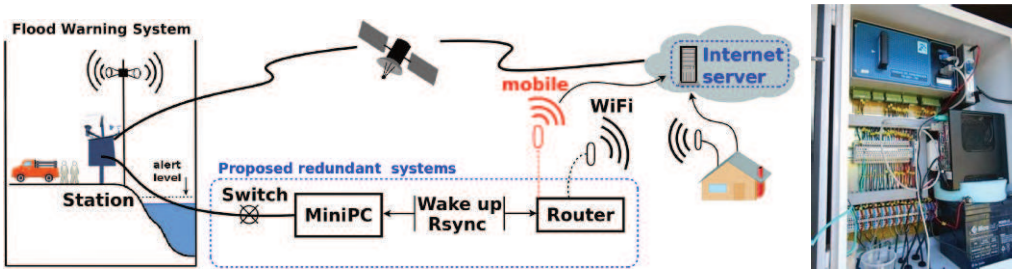
This interface was used before to drive remotely commercial equipment purchased without the wireless communication module supplied by the manufacturer (*López*, 2012).

Conditioning the added hardware

The interface works as a multi-platform Common Datalogger Interface (CDI) with minimal software dependencies. Figure B.7a depicts the interface architecture layout.

The system has two critical constituents: 1) a host program used for uploading and transferring the data from the station dataloggers; and 2) a server, which receives, stores and finally re-transmits the data originated at each site belonging to the FEWS' network.

The interface requires two small devices to run, a SOHO (Small Office/Home Office) router and a general-purpose minicomputer (called MiniPC hereafter), the latter connected by USB or serial wire to the datalogger (Figure B.7a. Consequently, a bundle of devices was ensemble first with the following features (Figure B.7b): 1) one MiniPC



(a) (b)

Figure B.7: (a) System architecture (in red): isolated stations (in black), stations within reach of an access point of the existing network. (b) The RTU with the MiniPC.

equipped with a processor INTEL 2.41GHZ J1800 DC, a motherboard AsRock, Memory RAM 4 GB DDR3 1600 MHz, a solid state disk (SSD) with a capacity of 128 GB, and a Mini-ITX cabinet with a power source of 60 W; 2) one SOHO router with an average power consumption of 0.6 W, equipped with a processor Atheros 400 MHz, Memory RAM 32 MB, 100 Mbps WAN/LAN Port, Wireless standard IEEE 802.11 ngb, and USB 2.0 Port.

It was essential to count with motherboards equipped with: i) Real time clock alarm which allows the MiniPC to turn on automatically at any preset date and time, and ii) Wake on LAN (WoL), which allows the MiniPC to be turned on remotely. The configuration of the former functionality is carried out in the power management section of each motherboard BIOS. For that reason, the network cards of the motherboards must support this protocol (this feature must be enabled from the BIOS). Also, it was vital to link the MiniPC to the Internet; a critical step since the proprietary software only works under Microsoft Operating System (OS), previously installed in each MiniPC. The task scheduler was duly programmed to run the developed interface automatically.

Depending on the FEWS' station locations, two options were considered as a function of the available power supply: one for those that have access to the grid power and are close to an Internet connection, and the other for the remote ones which are fed by a photovoltaic system. For the first alternative, an 802.11 wireless network was implemented that connects to the Internet access point existing at the location (Figure B.7a, in black). For the second option, the Internet access was achieved through a mobile phone (3G/GPRS/GSM, Figure B.7a, in red).

The MiniPC was installed at each station, always off albeit with a self-start according to a preset schedule. The router, of low consumption, has dual functionality. On the one hand it provided access to the Internet (WiFi or 3G/GPRS/GSM) while at the same time it offered an environment for a general purpose embedded GNU/Linux for remote access.

As previously indicated, some stations are powered by solar panels. For this reason, a battery of 12 V was attached to connect the MiniPC and the router, and a timer, with a voltage regulator to cut the battery power for the added router, with a window of 30 min for downloading and transmitting the data as a preset scheduled task.

Interaction with the proprietary software

Due to the impossibility of establishing direct communication with the RTU of each station without using the proprietary application, a software was developed to simulate the manual process performed by the network operators. It was crucial to avoid any interference with the proper functioning of the FEWS. The procedure yields an alternative dual route for wireless transmission, either for sending the records to the Internet or for receiving instructions remotely to modify the sensors parameters.

The software acts as an interface between the proprietary application provided by the manufacturer and the actions executed from a configuration file that is updated and modified from the Internet. The different functions that drive the interface resemble a layered architecture where each establishes some degree of communication with the underlying layer, known as a Common Datalogger Interface (CDI) (Figure B.8).

The organization of the CDI relied on a single configuration file consisting of subsections, each containing the necessary specifications for each equipment to which the station is connected. For the present situation, the configuration file was adapted to work with a FEWS station. There are several subsections for managing commercial devices of different brands when more than one sensor is connected at the same time. In the current implementation, the interface software runs in Python. Consequently, and depending on the ASCII configuration file, the adopted design invokes one by one the Python module corresponding to each sensor.

Several languages allow manipulating GUI applications, such as PyAutoGUI[2] which was finally adopted, and whose purpose is to provide cross-platform Python modules for their automatization. PyAutoGUI offered the best possible interaction with the RTU software. One of the main advantages is that it provides a Python library with an automation environment which, in turn, allows the use of the language's benefits, such as object orientation, easy reading, and writing of configuration files. It simulates movements and mouse clicks, keystrokes combinations and also a search of images on the screen (Figure B.8).

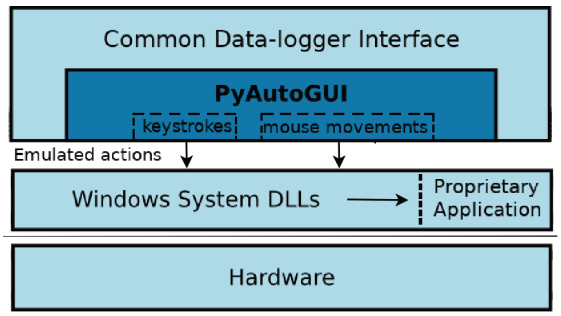


Figure B.8: Common data-logger interface (CDI).

Automating the data download

The design of the monitoring station posed a difficulty from the start. Whenever an operator connects a laptop to the service port of any station of the network (e.g. in

[2]http://pyautogui.readthedocs.io

a maintenance visit, Figure B.7b), it automatically overrides the satellite transmission module. Since the service port of the RTU was the natural choice to plug in the MiniPC to upload the data through the interface, it was necessary to build an electronic switch (on/off) to avoid the permanent overriding of the Orbcomm module (Figure B.9).

The switch is energized with a 12 V Molex connector from a motherboard and connected to pin 3 of the parallel port of the MiniPC. Through the operating system, the switch is set in the high state to close the relay and to enable the communication between the MiniPC and the RTU, and in the low state when all tasks end. In this way, the in-situ manual connection and disconnection for downloading data is emulated.

The operation of the implemented system consists of a series of steps, some of which are configurable, ranging from setting the power-on to the automatic shutdown of the MiniPC (Figure B.10). Briefly speaking, they amount to:

- Self-starting: by using the MiniPC real time clock alarm Protocol (configurable from the BIOS).
- Connection: by using the electronic switch (the software sets in a high state a pin of the parallel port so closing the circuit and thus emulating a manual connection).
- Extraction: by downloading the data stored in the RTU and then by re-configuring the station.
- Disconnection: by setting in a low state the pin of the parallel port so leaving the circuit open and thus emulating a manual disconnection.
- Transmission: by using the Rsync protocol to transmit to an Internet server.

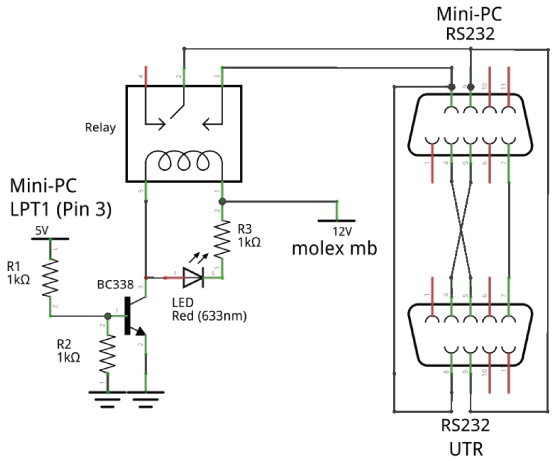


Figure B.9: Electronic switch circuit diagram.

Uploading data to the Web via a VPN

The use of a VPN enables remote access to computers that have a private IP from the Internet. For the current development, the VPN was set up in the MiniPC, which together

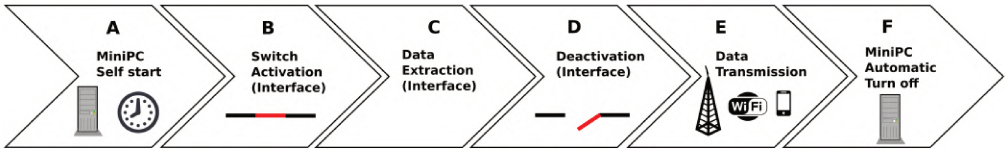


Figure B.10: Operation workflow.

with the TeamViewer package[3], allowed the remote access to the stations to perform unscheduled tasks (Figure B.7a).

Then, and whenever scheduled, the acquired data are uploaded to an Internet server using the remote synchronization protocol Rsync to avoid loss or duplication of data. It uses an algorithm to reduce the amount of data sent over the network, transmitting only the differences between the source files and the destination files (*Zhou et al.*, 1994).

Once a station node awakes, it downloads a configuration file if it has been changed remotely since the last login, and then uploads the associated data files to the cloud following a preset schedule. Listing B.1 shows the associated bash commands for Station 32 (see Figure B.6).

Listing B.1: Rsync commands

```
rsync.exe -avrPO /local/st32 ser@serverIP::sfe/st32
rsync.exe -avrPO user@serverIP::sfe/st32 /local/st32
```

The information then integrates a database for further processing before being transmitted to the final official repository site.

Viewing the data on a website

Most web services report according to the concepts of two layers; the front-end layer, and the back-end layer which includes all the logic processes in charge of generating the results. The front-end layer is an outlet to portray the notifications produced by the back-end. The back-end layer uses a GNU/Linux server[4] that hosts the data processing system consisting of a MySQL database, a Rsync daemon and Apache web server with PHP module and Python.

For the front-end layer, a dynamic website allows end users to obtain the information using customized queries. Furthermore, the raw files are available in a public repository using the original data format of each commercial device.

Once the interface synchronizes each field data, a set of Python scripts parses and stores them in a database. The system relies on the utility pynotify[5] that automatically executes when it detects any change in any directory. A Python script looks for new data files, and if they are present, the processing starts. It reads line by line getting the timestamp and the hydro-environmental recorded variable to put them all into the designed database.

[3] www.teamviewer.com
[4] www.tektonic.net
[5] https://github.com/seb-m/pyinotify

Preliminary testing

The proposed alternative communication system sought to link a station of the FEWS network to a nearby stakeholder, or to a rural school having Internet access, otherwise through mobile technology. Consequently, some performance and stability checks were first run at laboratory scale to continue then at the field before moving the equipment to the selected monitoring sites.

An internal test brought by Ubiquiti equipment reaches speeds of about 180.39 Mbps in mixed transfer mode and 148.76 Mbps in one direction, so in the transmission to the terminal devices, the 100 Mbps RJ-45 interface reaches the limit before the maximum WiFi speed. The result showed a state of a permanent link, with no fall or reconnection and with a constant speed over time. Therefore the connection results were satisfactory against a certified WiFi access point 820.11n. Also, since for WiFi dish antennas the communication path must be a line of sight, to prove its duplex capacity a test was made in the laboratory by connecting a satellite dish directly to the rear connector of the router.

The second test consisted of placing the antenna on the first floor of the FICH Department, connected in turn to its Internet network, and putting the TP-Link MR3420 router with an antenna at two distant points (Figure B.11). This device was configured to connect automatically at a different point to the signal emitted by the antenna. To verify the connection; the TP-Link MR3420 router was connected to a laptop computer, to which a few sample files were downloaded from a PC linked to the department's local network. Once the communication between the antennas was verified, both were taken to the field to establish the distance of reach between them. With that objective, one of them was held fixed to one side of the road while the other was carried in a vehicle until the signal faded (Figure B.11).

In the field, the group also tested other types of WiFi antennas depending on the type of power supply available. Thirteen visits were made to the selected stations along the project, seven to Santa Fe, two to Santo Tomé and four to Flesia (see Figure B.2a for locations).

The Santa Fe station (Figure B.12) feeds from the power grid with AC. On the contrary, the Flesia station, which is on an isolated field site by the Flesia Creek, and the Santo Tomé station located by the Salado River (Figure B.13), are fed by batteries powered, in turn, by solar panels without access to the Internet. In these cases, a 3G modem was used instead for the remote connection.

The Santa Fe station is located in the northwest part of the city, on premises owned by the Ministry of Public Works, and has free access to the Internet. There, an 802.11 wireless link between two Ubiquiti antennas was installed, one of them adjacent to the administrative office and connected via Ethernet to the available network.

The TP-Link MR3020 was configured with OpenWRT (*Holt et al.*, 2014), which allowed remote access and to power on the MiniPC. Then, the remaining antenna was installed together with the MiniPC in the FEWS station. The Santa Fe station was frequently visited, given its convenient location, to verify the operation of the redundant communication system during the tests.

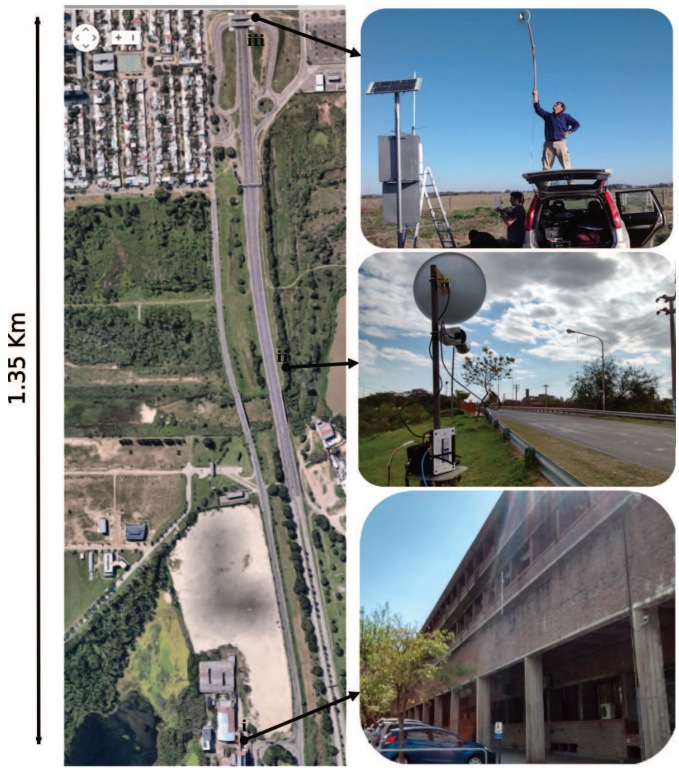


Figure B.11: Preliminary tests in lab and field.

Figure B.12: Setting up the redundant transmission system on the Santa Fe station.

Figure B.13: The Santo Tomé station.

3 Results

Both raw and verified data are uploaded and archived within the public server system as comma-separated, ASCII text files. Superusers or administrators can set the configuration parameters for each datalogger, to enable/disable sensors, and to increase/decrease the sampling and the data uploading frequency. The website allows users to access and send their customized queries. Then, users can choose different observation points and types of hydro-environmental variables between dates to download the information, or plot the temporal evolution of the records.

Although the system can preview filtered comma separated values from any formatted file, the relevance of the proposed solution is better appreciated whenever the transmitted data through the Orbcomm and the redundant channel are contrasted one against the other.

All sensors at the Santa Fe station were malfunctioning, except for the temperature sensor (Figure B.14). The parity between both registers is visible, with five different and lasting gaps throughout the official record filled correctly by the backup system.

Figure B.15 depicts the river stage evolution on the Flesia station over approximately 7 weeks. The data transmitted by the redundant system was one by one with the official record when both channels were on and correctly filled in the gaps left by the Orbcomm system in the three field experiments conducted.

Similarly, Figure B.16 displays the outcome of one month of recording of the river stage at the Santo Tomé station; in coincidence with the peak of an ENSO (El Niño-

Southern Oscillation) anomaly in the region. In this opportunity, in addition to failures in the Orbcomm system for periods that vary between 1 and 6 days, the RTU failed for about 3 days. The redundant system interrupted its transmission only once during a short period, less than a day on January 4, 2016. The developed interface could recover the data; but just the previous recorded day. Once again, a one-to-one correspondence is appreciated when both communication systems are active, as well as the achievement of the interface to fill in the missing information gaps.

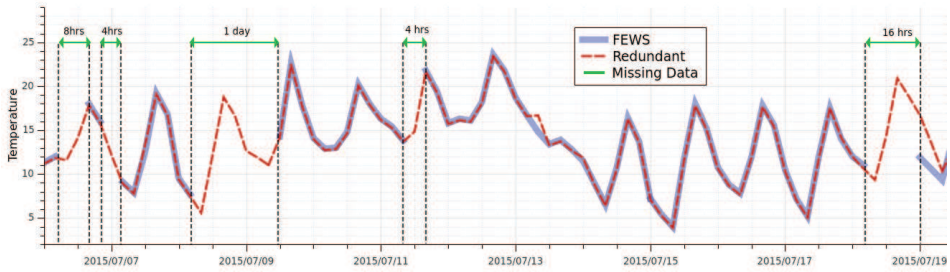


Figure B.14: Orbcomm vs redundant transmission system on the Santa Fe station.

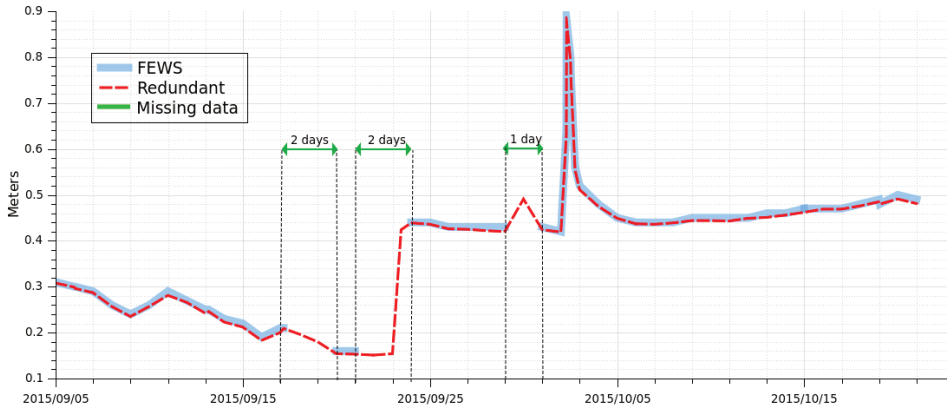


Figure B.15: Orbcomm vs redundant transmission system on the Flesia station.

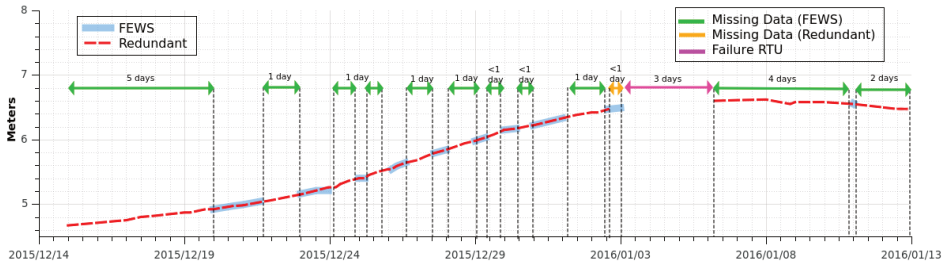


Figure B.16: Orbcomm vs redundant transmission system on the Santo Tomé station.

4 Discussion

Table B.4 summarizes the total number of test days when both systems were working simultaneously, and their failure rates. Surprisingly, the official system exhibits a similar average failure rate (here weighted by observation days) to that detected in the consulting work carried out by FICH (Table B.3 and Figure B.5), while for the developed interface, the failure rate was lower than 1%.

Table B.4: Failure rate of both systems

Station	Number of days	Official FEWS %	Redundant System %
Santa Fe	13	18.0	0.0
Santo Tomé	29	38.0	1.7
Flesia	47	10.0	0.0
Weighted average	89	20.3	0.7

Without discussing the fact that, in the end, what matters most is the cost of maintenance, it is relevant to emphasize here that the proposed interface is based on affordable hardware parts in some cases and open source components in others. This is in sharp contrast to the costly monthly service imposed by the Orbcomm satellite system, necessary to make data available in real-time.

Each station would require one MiniPC (cabinet, motherboard, and processor), with one solid state disk of 120 Gb, 4 Gb of RAM DDR3, an electronic switch, a waterproof resistance cabinet, a Tp-Link MR3020 router, and a voltage regulator, all totaling €295 approximately (Table B.5). The WiFi option will require two Ubiquiti NanoStation M5 antennas, two Ubiquiti Loco M5 antennas, and several meters of outdoor UTP cable CatE (about 15 m), all adding up to approximately €186 (Table B.6). On the contrary, the 3G alternative will demand a 3G Modem Huawei, one timer, USB cable (around 5 m or so), and a 12 V battery, totaling €53 more or less. The latter will add up to €168 of the annual mobile or cellular service charge, totaling €220 approximately (Table B.7).

In summary, in round figures, the alternatives successfully tested in the current configuration correspond to an initial investment that could add, per station, up to €480 for the WiFi alternative, or up to €516 for the 3G option. In the latter case, the annual cost of the cellular telephone service provider was included. These figures should be compared with the annual expenditure of €13,000 for having the Orbcomm system in operation, a system that fails, on average, 20% of the time.

The redundant communication system could be located in four or five key stations to have a reliable backup system whenever necessary.

4.1 Maintenance of the redundant system

The test period of the three stations together does not span more than three months. It is a very short period to obtain conclusions or reliable data on inherent system failures that would require the replacement of some or all of its components. However, based on previous and ongoing work carried out by the research group, which was the basis for the software development later transferred to the FEWS of the Salado River, it was

Table B.5: Required hardware components for each station

Common components	**Cost (€)**
MiniPC (Cabinet + Motherboard + Processor)	143
Solid state disk SSD 120 GB	67
Memory RAM 4 GB DDR3	28
Key	14
Waterproof cabinet	19
Router Tp-Link MR3020	17
Voltage regulator	7
Subtotal	295

Table B.6: WiFi components

WiFi	**Cost (€)**
Antenna Ubiquiti NanoStation M5	111
Antenna Ubiquiti Loco M5	70
Outdoor cable UTP CatE S7 (15 meters used)	5
Total WiFi alternative	186

Table B.7: Mobile components

3G Communications	**Cost (€)**
Modem release 3G Huawei	13
Annual Mobile provider service	168
Timer	12
USB cable (5 m)	5
12 V Battery	23
Total 3G Alternative	221

possible to extract the following failure frequency of low-cost devices used in outdoor environments:

i. Ten months of data collection in the agro-technical school of Santo Domingo (Figure 2(a)), operating phreatic level, soil moisture and temperature sensors[6];

ii. Seventeen months of observation with heat, water vapor, and carbon-dioxide flux sensors (Campbell eddy covariance[7]);

iii. Three years of water level collection (acoustic sensor device[8]),

iv. Five years of observation of hydro-meteorological data with a Pegasus station[9].

Table B.8: Observed failures during GUI-testing, sensors, and other alternatives

Sensor	**Reason of failure**	**Component lost**	**Cost (€)**
i. Santo Domingo	Power outage	MiniPC	143
ii. Eddy covariance	Lightning	Antenna Ubiquiti	70
iii. Water level	Power outage	Voltage regulator	7
iv. Pegasus station	Power outage	MiniPC	143

Overall, the operation of the experiments detailed in (i)–(iv) just served to provide a glimpse of the occasional failure of the low-cost components used in different outdoor settings, either by natural causes or by inherent faults (Table B.8). The weighted average of the values shown in Table B.8 yields a yearly maintenance cost for the low-cost system on the order of €104 (Table B.9). The observed failures were always attributed to an electrical fault, which constitutes the chilles' heel of low-cost systems.

Before applying the interface to the FEWS problem, all field experiments detailed in Table B.8 served to test the development of the software with different devices and environments to ensure a portable and trouble-free product. In Appendix 7, the reader will find a block diagram explaining how the interface works, and a repository address with the Python scripts where all the examples are readily available.

With the weighted average cost to replace damaged components for the low-cost system, plus the cost of communication (Table B.7), it is possible to set up the following comparison between both systems on an annual basis (Table B.9).

These numbers should be taken with caution since there are always hidden costs, expenses not contemplated, or changes in technology that can push the figures 20–25% above or below the initial estimate. Nevertheless, the low-cost system is, on average, eight times cheaper to maintain than the equipment of the official network.

Finally, the backup system was transferred to the local authorities of the Province of Santa Fe at the end of 2015. A technical talk was given to the technicians and the authorities in charge of the alternative network[10]. The system was easy to operate since,

[6] www.specmeters.com, www.genica.com.ar
[7] www.campbellsci.com
[8] www.maxbotix.com
[9] www.tecmes.com, fich.unl.edu.ar/cim/
[10] https://www.santafe.gov.ar/noticias/noticia/220436/

Table B.9: Maintenance costs between the official and the alternative network, on a unit basis

Item	Official FEWS	Redundant system
Communication	€1405 (Orbcomm)	€168 (3G)
Replace of components	€712	€104
Total annual cost	€2117	€272

among other things, it relied on a dynamic website to visualize the data downloaded and transmitted through the redundant system. Unfortunately, immediately after transferring the product, there was election time in the Province, and new authorities took over the State, freezing the collaborative effort for a while. Nowadays, there is a renewed interest by the state authorities after the devaluation of the Argentinian currency, given the high cost to replace damaged components and the obsolescence of the official network.

5 Conclusions

The Flood Early Warning System (FEWS) of the lower-catchment of the Salado River, Argentina, was set after the catastrophic flood suffered by the Santa Fe City, in 2003. A quick flood frequency analysis estimated the return period of the flood as around 130 years. However, the often loss of information due to communication problems in the final repository site of the FEWS has affected the network operational vulnerability ever since.

This paper proposes a solution based on a developed interface to transmit the data in parallel, previously captured by the monitoring network, through a dual channel that uses an electronic bypass without interfering with the original satellite communication system employed by the FEWS.

Several laboratory-scale trials were performed first to test the feasibility of the interface to transmit the data stored by the proposed redundant communication system. Three field experiments later complemented these tests carried out at lab scale. During the testing phase, the official communication system exhibited an average failure rate of the order of 20% compared to the developed interface, which was less than 1%. In all cases, the results had shown not only the consistency of the information transmitted by satellite and the proposed alternative, when both were active, but also the achievement of the latter to fill in the information gaps when the official system experienced a "blackout" for some reason.

The automatic manipulation of a GUI was a technical contribution that emerged due to the impossibility of accessing the manufacturer's application of each data unit of the network, with an end product that sends the data with incredible fidelity. Moreover, the testing of the methodology with different commercial devices and environments ensures a trouble-free, portable product, making it attractive to any FEWS with similar characteristics. The software was designed to add support to other commercial devices. All software of the developed interface was uploaded to a public site, where the readers can study the source code, modify it, and use it the way they want.

The interface relies on low-cost, single open source board computers and routers that eventually reduce the maintenance cost of the satellite-based FEWS system signif-

icantly. This aspect of the proposed solution, in turn, can increase the reliability of any FEWS with similar characteristics.

6 Acknowledgements

Support from the National Council for Scientific and Technical Research (CONICET), research grant no. 112-201100100384, and from the Universidad Nacional del Litoral, research grant CAID 279/013-I+D 1-11, are greatly acknowledged. Thanks are also given to Senior Eng. Ricardo Giacosa and other authorities of the Ministry of Public Works of the Province of Santa Fe for their financial assistance and encouragement during the execution of this project. The authors also thank three anonymous reviewers for their useful criticisms and comments that helped to improve the first version of the manuscript.

7 Supplementary material

The Supplementary Material for this paper is available online at
https://dx.doi.org/10.2166/hydro.2020.216

Appendix B.1

Python software to automate open or closed monitoring environmental stations

The Common Data-logger Interface (CDI, Figure 8) is part of a Python Software to Automate Monitoring Stations (namely PyAMS). It was published in a public software repository[11], where the source codes can be freely downloaded, used, studied, modified, or simply to suggest new features or to report or fix bugs. The software was designed to easily add support for other devices. When the GUI application is the only option to deal with the proprietary (closed) device, the CDI can interact with it adding the instructions to execute the mouse movements and keystrokes to automate the GUI. On the other hand, if the devices are not tied to any particular driving software (open), they are readily accessible with customized software using the hardware manufacturer information to communicate with the equipment (Figure B.17).

The methodology was previously tested with several commercial monitoring stations, or data-loggers, such as Campbell Eddy Covariance, Pegasus and Stevens (Figure B.18). The examples to drive these devices, including the software to drive the FEWS of the Salado River were published in a public repository. There, the reader can find simple instructions on how to give support to its own device.

[11] http://gitlab.com/emilopez/PyAMS

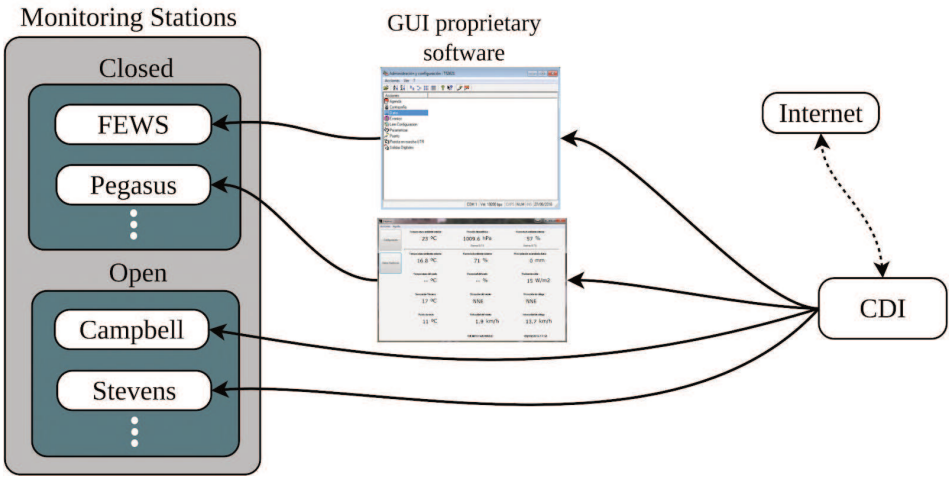


Figure B.17: PyAMS structure to interact with closed or open devices.

Figure B.18: Eddy covariance station, Pegasus station and Stevens data-logger.

Appendix B.2

Fitting Maximum Daily Annual River Discharges

The Gumbel and the Weibull models are two limiting cases of the generalized extreme value (GEV) distribution of maxima used to describe extreme phenomena in hydrology and risk assessment. The exponential and log-normal distributions have also been used to analyze the annual maximum values of daily river discharges.

For the Gumbel, Weibull and the Exponential distributions, a series of logarithmic transformations simplify the calculation of their parameters. For the lognormal distribution, it is necessary to use asymptotic expansions (or Tables). The analysis of frequency in maximum annual daily flows is carried out using probabilities of exceedance, which are related to the return period, T_r, through the expression,

$$P(q \geq Q) = 1 - F(q) = 1/T_r,$$

where $F(q) = P(Q \leqslant q)$ is the non-exceedance probability and q is the annual maximum of daily river discharges. The fitting of the theoretical distributions to data is most conveniently done by least squares. The data are first sorted in ascending order, and "ranked" by using the Weibull plotting position (*Makkonen*, 2006)

$$F_i = i/(n + 1),$$

where i is the ascending rank of the river discharge and n is the total number of observed values in the data set. The bimodal distribution of the river height observed on the SR70 station (Figure 4) suggests the existence of two well-separated hydrological responses in the lower Salado River basin. Field observations suggest $q_{bf} \simeq 320 m^3/s$ for the bank-full discharge value. Consequently, only those annual maxima that satisfy $q \geq q_{bf}$ are kept for the statistical analysis, so the original 62 years-long series shortens to 43 values.

For the Gumbel and Exponential distributions, the unknown parameters are easily estimated by a direct application of the least-squares method. However, a Newton-Raphson iterative scheme is required to solve for the unknown parameters for the Weibull distribution. The linear Taylor expansion in the unknown coefficients is approximated with ordinary least squares fitting to obtain new values at each step until convergence is achieved. For the lognormal distribution, it is required to compute the inverse of the cumulative density function, which leads to a straightforward application of the least-squares fit.

All models overestimate T_r, that is, underestimate the risk when compared with the return period of the 2003 flood, with a peak of 3980 m^3/s. Gumbel is mostly off the observed range of q except for medium values. The Weibull fitting is remarkable for the lowest values of the river discharge and then runs almost parallel to the lognormal distribution for large values of q (Figure B.19). The predicted values for several T_r-years flood events are summarized in Table B.10. The average of the values predicted by the Exponential and Weibull distributions for T_r=130 years is very close to the peak value recorded in the 2003 flood. This estimate is in line with public records (Table 1) since an event like the 2003 flood was seen only once in more than 105 years.

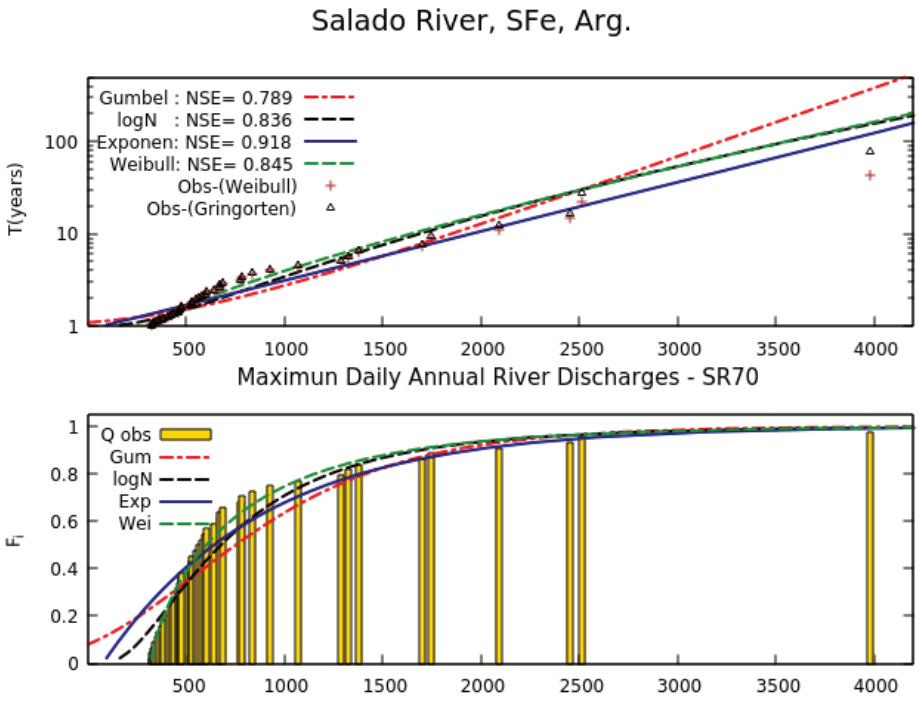


Figure B.19: Observed and theoretical distributions of maximum daily annual river discharges (Salado River, Santa Fe, Argentina.

Table B.10: Estimation of the T_r-year flood event, in m^3/s

T_r (years)	Log normal	Exponential	Weibull
100	3556	3826	3550
130	3811	4040	3789
200	4248	4390	4189

Apéndice C

A low-power IoT device for measuring water table levels and soil moisture to ease increased crop yields

Resumen: La medición simultánea del contenido de agua en el suelo y los niveles de la napa freática es de gran interés agronómico e hidrológico. No solo la humedad del suelo representa el agua disponible para el crecimiento de las plantas, sino que también la profundidad a la que se encuentra la napa freática pueden afectar la productividad de los cultivos. Además, monitorear estos parámetros hidrológicos es esencial para una alerta temprana de situaciones de lluvias extremas. Sin embargo, la medición utilizando instrumentos propietarios tiene ciertas desventajas, como el alto costo de adquisición y mantenimiento. Sumado a esto, el manejo de dispositivos comerciales dificulta su adaptación a los requisitos específicos de los agricultores o tomadores de decisiones. Las plataformas IoT (Internet de las cosas) basadas en hardware open-source están surgiendo como una alternativa atractiva para desarrollar dispositivos flexibles y de bajo costo. Este artículo describe el diseño de un dispositivo de medición en plataformas de hardware de open-source para registrar niveles de la napa freática y de humedad del suelo para aplicaciones agronómicas. Se detallan las técnicas de ahorro de energía implementadas, siendo uno de los puntos débiles de estas plataformas y la transmisión inalámbrica. Se realizó la calibración y validación del sensor de freátimetro con datos de laboratorio y de campo. Finalmente, se muestra como las técnicas no lineales de aprendizaje automático mejoran las calibraciones para sensores de humedad de suelo de bajo costo por sobre las herramientas clásicas[1].

[1]Este artículo fue publicado en Sensors 2022, 22(18), 6840; https://doi.org/10.3390/s22186840

Abstract: The simultaneous measurement of soil water content and water table levels is of great agronomic and hydrological interest. Not only does soil moisture represent the water available for plant growth but also water table levels can affect crop productivity. Furthermore, monitoring soil saturation and water table levels is essential for an early warning of extreme rainfall situations. However, the measurement of these parameters employing commercial instruments has certain disadvantages, with a high cost of purchase and maintenance. In addition, the handling of commercial devices makes it difficult to adapt them to the specific requirements of farmers or decision-makers. Open-source IoT hardware platforms are emerging as an attractive alternative to developing flexible and low-cost devices. This paper describes the design of a data-logger device based on open-source hardware platforms to register water table levels and soil moisture data for agronomic applications. The paper begins by describing energy-saving and wireless transmission techniques. Then, it summarizes the linear calibration of the phreatimeter sensor obtained with laboratory and field data. Finally, it shows how non-linear machine-learning techniques improve predictions over classical tools for the moisture sensor (SKU: SEN0193).

1 Introduction

The upper soil layers constitute the unsaturated or vadose zone that undergoes periodic fluctuations in water content and solute concentrations. Those changing conditions are consequences of the acting evaporation, infiltration, deep percolation and water uptake by plant roots (*Hopmans*, 2011). Consequently, during periods with precipitation above evapotranspiration, some of the remaining water increases the soil moisture storage and raises the water table. Monitoring soil saturation and water table levels are essential for an early warning of extreme rainfall occurrences (*López et al.*, 2020). Such a situation eventually leads to episodes of waterlogging and flooding in rural and urban areas (*Vionnet et al.*, 2006a; *Kuppel et al.*, 2015). However, not all scenarios of high water table levels appear to have negative impacts. Water storage in unsaturated and saturated zones is critical to meet crop water requirements in arid areas with shallow water table depth (*Zhao et al.*, 2020). Indeed, groundwater could exert both a positive and a negative influence on crop production. Depending on its depth, shallow groundwater can represent a valuable source of water supply to crops in drought periods or a stress agent causing waterlogging in rainfed crops (*Mercau et al.*, 2016).

Figure C.1 sketches the variation in crop yields depending on the groundwater depth. There appears to be a range of water table depths in which crop productivity is optimal (*Nosetto et al.*, 2009). When groundwater is close enough to the surface, crop yields decline sharply, suggesting the effects of waterlogging, root anoxia or salinity. Finally, if the groundwater depth falls beyond the optimum band, crop yields seem to decrease at an approximately exponential rate (*Nosetto et al.*, 2009). Therefore, soil moisture and water table levels directly affect crop productivity and food security (*Kopittke et al.*, 2019).

Monitoring these hydrological cycle components in any agricultural activity is the first step toward efficient water resource management. For example, if the producer knows in advance the water content of the soil profile, the timing and height of the irrigation lamina required by crops can be optimized (*Rodriguez et al.*, 2006; *Babaeian*

et al., 2019; *Vereecken et al.*, 2014). In rainfed (non-irrigated) agriculture situations, the knowledge of both parameters is equally relevant to quantifying water availability during the crop cycle (or at some key moment in the crop cycle) and to supporting decision making. By knowing the water reserves stored in the soil, farmers can decide whether or not to intensify their crop rotations, delay the sowing date to favour soil profile moisture recharge and decide which summer and winter crop sequences are best. The information may allow the producer to bring forward sowing to avoid waterlogging or loss of support in their plots (if the water table is very high). The producer can also define, based on field data, the dose and timing of fertilizer to be applied (*Mercau & Otegui*, 2014). Unlike rainfall, whose precipitable volume cannot be predicted at the beginning of the season or during the crop cycle (*Barros et al.*, 2000), the height of the water table and soil moisture are measurable components that reduce the uncertainty of water availability.

There are a variety of simple procedures and devices to survey water table levels. Manual measurement of the groundwater depth includes electronic indicators made of a conductor wire, a probe attached to the end and an indicator that emits a characteristic sound when it touches the water surface (*EPA*, 2020). For example, the HOBO U20L[2] is a popular logger used for monitoring changing water levels in a wide range of applications. The logger consists of a ceramic pressure sensor inside a durable housing for deployment in existing wells.

The determination of soil moisture is much more difficult and subtle. Existing approaches to soil moisture estimation differ in their accuracy, the spatial area covered, the depth of soil measured, the frequency of measurements and the cost of acquisition (*Dorigo et al.*, 2011; *Babaeian et al.*, 2019). At one extreme, there are estimates derived from satellite missions specifically designed to measure soil moisture (*Kerr et al.*, 2001; *Entekhabi et al.*, 2010; *Lozza*, 2019). These remote-sensing methods provide soil moisture estimates over large areas on a global scale, with a coarse spatial and temporal resolution. Another shortcoming is the limited soil depth measured, from a few centimetres for optical products to $\simeq$ 50 cm for longer wavelength radars in dry and bare soils. At the other extreme, proximal electromagnetic sensors such as time domain reflectometry (TDR) or capacitive sensors produce accurate *in situ* estimates of soil moisture at user-defined depths, locations and times. Their major limitation is the point value of the measurement since it represents the moisture of a small volume of soil at a fixed position (*Babaeian et al.*, 2019).

A feasible alternative to recording data with a high spatio-temporal resolution of soil moisture and water table levels at plot scale is by several proximal sensors integrated into a wireless monitoring network connected to a central station (*Bogena et al.*, 2007; *Babaeian et al.*, 2019). This technical solution requires affordable, robust and reliable sensors that are easy to install and have low power consumption (*Barcelo-Ordinas et al.*, 2013). Several commercial sensors meet most of these conditions (*Bogena et al.*, 2007; *Vereecken et al.*, 2014). Unfortunately, most commercial devices are beyond the reach of users in countries with emerging economies.

In recent years, Arduino, an open-source programmable microcontroller, proved to be a competitive player in driving the development of sensors and electronic components based on free software and hardware aimed at recording, storing and trans-

[2]https://www.onsetcomp.com/HOBO

mitting environmental parameters at low cost (*Sadler et al.*, 2016b; *Beddows & Mallon*, 2018). At the same time, its customized design has contributed to solving availability and energy efficiency problems (*Manso et al.*, 2018). In particular, the capacitive sensor SKU: SEN0193[3] and the differential pressure sensor Honeywell HSCDAND015PDSA3, both Arduino-compatible (see Figure C.1b), are good candidates for implementing an energy-efficient, low-cost monitoring network of agro-hydrological parameters (*Placidi et al.*, 2020; *González-Teruel et al.*, 2019). Recently, there have been numerous works on machine learning applied to sensor calibration to contrast surface moisture measurements with satellite data (*Coopersmith et al.*, 2016) and for estimating the impact of temperature on soil moisture detection in the root zone (*Bogena et al.*, 2017; *Chen et al.*, 2018; *Kapilaratne & Lu*, 2017).

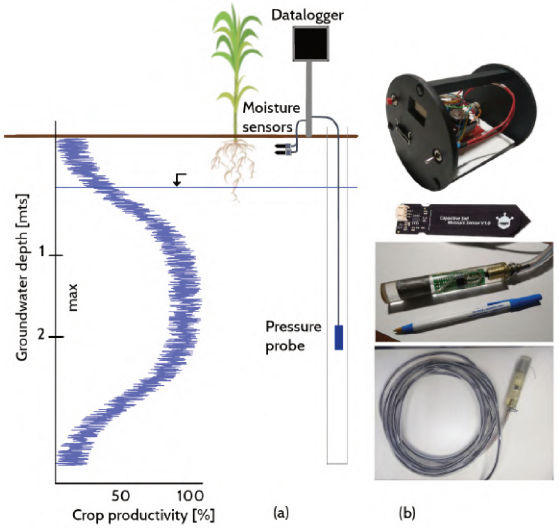


Figure C.1: **(a)** Schematic of crop yields, which may be optimal within certain water table elevation ranges (the fitted curve has a wide scatter band and varies for corn, soybeans and wheat, according to Nosetto et al. *Nosetto et al.* (2009)); **(b)** datalogger housing, sensor SKU: SEN0193 and pressure probe with Honeywell sensor.

In light of the above issues, the objective of this work is twofold:

- First, to present the implementation of an energy-efficient IoT device intended for real-time monitoring of two environmental parameters of great relevance for cereal producers. The hardware used for the datalogger utilizes the open-source, mass-produced electronics platform Arduino. A switch diminishes the power consumption by saving energy when the system is in sleep mode. Subsequently, the IoT device integrates the core of a non-optimized wireless unit to transmit and store environmental data.

- Second, to discuss the advantage of using nonlinear machine-learning techniques over classical linear regression techniques to calibrate the sensors used to collect the information.

[3]https://wiki.dfrobot.com/Capacitive_Soil_Moisture_Sensor_SKU_SEN0193

The first objective does not exhaust the comprehensive treatment that numerous researchers have given to the open-source programmable microcontroller platforms (see, e.g., *Fisher & Gould* (2012a); *Sadler et al.* (2016b); *Beddows & Mallon* (2018); *Thompson et al.* (2021)). While *Sadler et al.* (2016b) employ Arduino UNO, which is energy-inefficient, *Beddows & Mallon* (2018) use power-optimization techniques based on small form-factor 3.3 V Arduino to achieve long-term data storability in submerged environments. *Thompson et al.* (2021) also employ a low-cost system based on an Arduino microcontroller. This is an approach that, despite its similarities with ours, has its disparities. The difference lies in the electronic components used and in the interoperability of the IoT with low-cost sensors, whose firmware was developed here from scratch. For example, *Thompson et al.* (2021) not only do not measure the water table, which according to *Nosetto et al.* (2009) can lead to optimal crop production, but they also use commercial soil moisture sensors to measure volumetric water content. Here, part of the focus is on the inexpensive SKU: SEN0193 sensor. This aspect is in line with the second objective. The difficulty in obtaining a reliable calibration curve or a unique response is well known *Nagahage et al.* (2019); *Placidi et al.* (2020). There will be as many readings as there are SKU: SEN0193 sensors deployed in the field. We resort to machine-learning techniques, in particular the Random Forest (RFO) algorithm, to circumvent this shortcoming, obtaining remarkable results when compared to the reference data (provided by a known commercial sensor). An interesting application of the RFO algorithm can be seen in the work of *Carranza et al.* (2021), although in a completely different context. They applied it to interpolate/extrapolate spatial–temporal soil moisture data from information collected at 15 stations distributed in an experimental watershed.

2 Materials and Methods

2.1 Hardware and Firmware in Context

Figure C.2 shows the architecture of the designed system. The remaining sections describe the methods employed to implement the low-cost IoT logging and data transmission device configured to automatically input data collected with two probes, one for soil moisture and one for the water table. The test methods implemented to validate the inter-connectivity between the open hardware/software and the readings from the probes are also described—first, against soil moisture determinations obtained by gravimetry. Then, data were produced with commercial sensors for soil water content[4] and water table levels[5]. Finally, the reader can also find supplementary material in the Appendices.

2.2 Low-Power IoT Device

The IoT device was developed at Universidad Nacional del Litoral, Argentina, by researchers from the Engineering and Water Sciences Department. Figure C.3a shows the

[4]https://stevenswater.com/products/hydraprobe/
[5]http://www.genica.com.ar

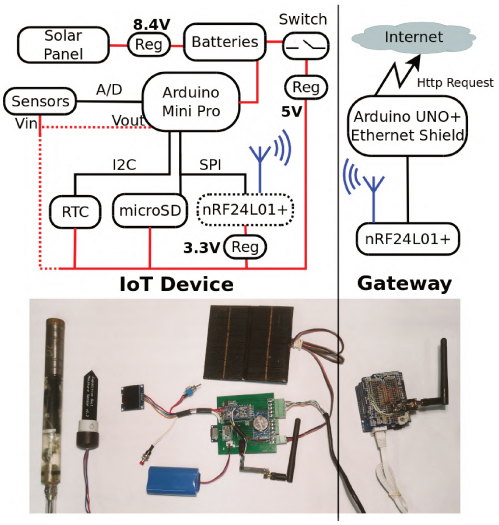


Figure C.2: Schematic of the system architecture (Figure C.15 of Appendix A contains a diagram of the electronic circuit showing the interconnection between the components of the IoT device).

layout of the printed circuit board (PCB) supporting the datalogger. Its principal function is to store the values read by the sensors and occasionally transmit them at regular intervals to a cloud repository. It consists of a microcontroller (Arduino Pro Mini 5 V 16 MHz), a storage module (catalex microSD) and a real-time clock (RTC DS3231). Depending on the requirements, a wireless transmission module is added (in this case the nRF24L01+). Figure C.1a shows a schematic with a node supposedly equipped with a transmission module.

The average power consumption was analysed in an idle state by successively connecting the different modules. Then in the working state, i.e., when it measures, stores and transmits data. The analysis was performed in the laboratory using an Owon VDS1022 oscilloscope. The voltage drops at the ends of a resistor of known value at the input of the device power supply were analysed. Thus, it was possible to know, through Ohm's law, the current consumed and the exact time of each task. These results have been summarized in Figure C.3b.

Once the measurements are stored, a transistor circuit cuts the power to the RTC and microSD modules and the wireless module as well if it is in use and enters into sleep mode until the next measurement. In detail, a 5 V Arduino Pro Mini with micro SD and RTC modules uses about 33 mA in idle state. In addition, to reduce its power consumption was necessary to cut the two tiny power connectors to the LEDs that come with the Pro Mini board and add a sleeping mode via firmware. However, despite this encouraging result, the device still demanded an energy consumption of 33 mA once the modules were connected. The introduction of a power gating with two transistors to cut off the energy of the SD and RTC modules reached less than 1 mA of consumption (Figure C.3b). The ultimate design of the datalogger has a 1.3″ Oled display that shows the measurements, date and battery voltage in real-time by pressing the monitoring button.

Once the sensor readings are ended, the program disengages the RTC and the SD module and enters into sleep mode. The internal watchdog wakes up the microcontroller every 8 seconds to check if it is time to measure or continue in sleep mode. If it is time

to measure, it reads the values of each sensor, energizes the modules and stores the information in the SD memory. The modular design of the device allows it to operate in stand-alone mode or to send data to a cloud repository via a wireless transmission module.

A complete list of components used to assemble the IoT device, the gateway node, as well as the water table probe and the soil moisture sensor employed can be found in Appendix B. Moreover, an exploded view of the housing for the IoT device is shown in Figure C.16.

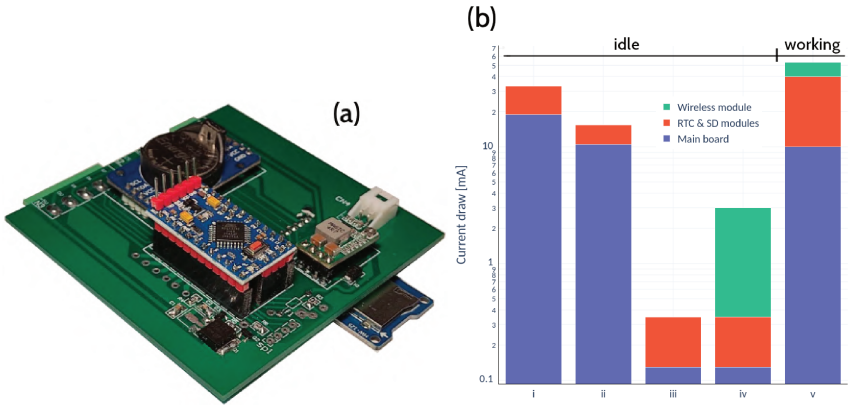


Figure C.3: **(a)** The PCB with the datalogger components mounted (the microcontroller, the storage unit and the real-time clock); **(b)** successive trials leading to the final energy-saving scheme: (i) Arduino Pro Mini with SD and RTC modules (≈ 33 mA), (ii) case (i) in sleep mode (≈ 11 mA), (iii) power gating technique with no LED after (ii) (≈ 0.35 mA), (iv) with wireless module added to (iii) (≈ 3 mA), (v) average power consumption of about 53 mA in full working mode, i.e., with sensor reading, storing data and transmitting.

2.3 Power Supply Systems

The power system consists of rechargeable 18650 Li-ion batteries of 3.7 V, 2600 mAh and 6 V, 1 W solar cells. Two or three batteries are connected in series like solar cells depending on the energy needs of the equipment. The voltage at the panel output passes through a regulator and a charge controller to ensure that each battery receives the correct voltage. In the soil moisture calibration stage, 12 V reference sensors were used. The flexibility of the design allowed it to be adapted to this voltage using 3 solar cells and 3 18650 batteries. The power supply to the Arduino Pro Mini was reduced by employing 4 diodes in series, entering a voltage of around 9.2 V. Then, for all other cases, two solar cells charge the 2 batteries and the Arduino Pro Mini is powered with 8.4 V. Both voltage values are recommended in the datasheet so as not to overload the internal voltage regulator of the Arduino Pro Mini and avoid undesired behaviour. Figure C.15 displays further details about the electronic circuit of the IoT device.

2.4 Sensors

The value of the dielectric constant of air is 1, while that of water is 80 (*Domínguez-Niño et al.*, 2019). The presence of water in the soil generates a variation in the electrical permittivity, thus modifying its capacitance. The SKU: SEN0193 (DFRobot) is a very low-cost capacitive soil moisture sensor that operates in the range of 3.3 to 5 volts as supply voltage. Its output is an analogue value inversely proportional to soil moisture. The coplanar sensor is made of corrosion-resistant material, which increases its durability. A pulling element has been added to the sensors SKU: SEN0193 for easy handling. In addition, they were protected with an electrically insulating film plus a crystal epoxy resin to prevent soil moisture from damaging the electronic circuits, protecting the sensor readings from any influence of soil moisture. This type of resin is commonly used in commercial soil moisture sensors ([6]) and has been used for the same purpose in similar works *González-Teruel et al.* (2019).

For the SKU: SEN0193 calibration, the Stevens HydraProbe II (Stevens Water Monitoring Systems, Inc., Portland, OR) sensor was used as a reliable reading for the soil moisture value. The HydraProbe II sensors determine electrical permittivity using a variant of the time domain reflectometry (TDR) technique, i.e., based on the time delay between the emission and reception of an electromagnetic pulse (*Domínguez-Niño et al.*, 2019). Unlike the capacitive sensor, the output is digital (SDI-12 protocol) and directly represents the volumetric water content (VWC). It is an impedance sensor with three tines surrounding one centre tine that measures the real and the imaginary dielectric permittivity separately from the response of a reflected standing EM wave at a radio frequency of 50 MHz. The device works with a 12 V power supply. The temperature has a great influence on soil moisture *Bogena et al.* (2017); *Chen et al.* (2019). Therefore, its value was recorded with the DS18B20 digital sensor manufactured by Maxim Integrated Products, Inc., San Jose, CA. Simultaneously, the ambient temperature and relative humidity were measured with the DHT22 sensor (Aosong Electronics Co Ltd., Guangzhou, China) to evaluate their impact on soil moisture determination. Both sensors are digital and operate with a 5 V power supply.

The firmware for the Honeywell digital differential sensor (HSCDAND015PDSA3) was designed from scratch to record changes in water table levels. The sensor operates on 3.3 V with a 12 bit resolution that records up to 15 PSI of the water column. This sensor has two ports, one to measure water pressure and the other to compensate for atmospheric pressure. A pediatric nasogastric tube was used for the first port and a glass tube for the second. Crystal epoxy resin was used to protect the electronics of the sensor. The exploded view of the housing for the water table probe and the SKU sensor is presented in Figures C.17 and C.18.

2.5 Communication with the Sensors

The firmware communicates with the sensors through specifically developed libraries[7]. For the HydraProbe II soil moisture sensors and the capacitive sensors, the SDI-12 digital protocol was programmed and an analog measurement communication system was

[6] https://www.stevenswater.com/resources/documentation/hydraprobe/HydraProbe_Manual_Jan_2018.pdf, https://www.campbellsci.com/cs655

[7] https://gitlab.com/emilopez/monito

established, respectively. For the water table reading with the differential pressure sensor, the SPI digital protocol was used. The digital protocols require specific parameters from both sensors. In the case of SDI-12, characters are sent to establish communication with the sensor (see sensor datasheet) to request the measured values. In the case of the pressure sensor, two bytes are read from a specific memory location 0x00; with this data a transfer function is applied that converts it to PSI,

$$Pressure_{PSI} = Pressure_{min} + (Output - Output_{min}) \cdot \Delta P / \Delta O \tag{C.1}$$

where $Pressure_{min}$ is the minimum value of pressure range (PSI), $Output$ is the digital pressure reading (counts), $Output_{min}$ is the output at minimum pressure (counts), ΔP is the difference between maximum and minimum pressure range (PSI) and ΔO is the difference between the output at maximum and minimum pressure (counts). Finally, the $Pressure_{PSI}$ obtained is converted to centimetres of water column.

2.6 Data Wireless Transmission

A wireless transmission module was added to the IoT device for data communication. The same power gating circuit was used in the RTC and SD modules. Nordic's 2.4 GHz technology (nRF24L01 + PA + LNA SMA) was used to establish the links. This is a half-duplex wireless communication module with a maximum peak receive/transmit transmission rate of up to 2 Mbps, average power consumption of 14 mA and a maximum range of 1100 meters.

The nRF24L01+ operates in the worldwide ISM frequency band of 2.400-2.4835 GHz. It is a single-chip 2.4 GHz transceiver with an integrated baseband protocol engine (Enhanced Shock Burst™), suitable for very low-power wireless applications. It is possible to operate and configure the nRF24L01+ through a serial peripheral interface (SPI). From its datasheet, the user can find the following additional information; the built-in baseband protocol engine (Enhanced ShockBurst™) is based on packet communication and supports various modes from manual operation to the advanced autonomous protocol operation. The radio front end uses GFSK modulation. It has user-configurable parameters such as frequency channel, output power, and on-air data rate. The nRF24L01+ supports 250 kbps, 1 Mbps and 2 Mbps over-the-air data rates. The high data rate over the air, combined with two power-saving modes, make the nRF24L01+ well suited for very low-power designs.

2.7 Three Standard Tests for the Pressure Sensor

Three deployments of the equipment were conducted at the Water Resources and Engineering premises at the Universidad Nacional del Litoral (FICH-UNL, Santa Fe, Argentina) to demonstrate the proof-of-concept of the low-cost datalogger connected to the Honeywell pressure sensor in real-time situations. In the first deployment, data were recorded at a rate of one sample per second. In the second and third trials, data were recorded at a rate of 10 samples per second and one every 7 s, respectively.

Although it was not necessary to calibrate the Honeywell pressure sensor, the firmware required development from scratch. The first test aimed to validate the readings of the sensor and its coupling with the low-power IoT device. Figure C.4a shows the best fit to

the dispersion data (a 45-degree line, zero intercepts) obtained by filling and emptying a phreatimeter not shown here for the sake of brevity in the presentation, albeit with characteristics quite similar to the observation well detailed in the third test. The depth of the phreatimeter is 6 m. The commercial sensor GENICA only registers up to 4 m of the water column. The HOBO sensor was placed at the bottom of the phreatimeter. There is a weak, positive and systematic bias when the data are compared to the GENICA readings and a negative bias when the comparison is against the HOBO.

The second test was a transient fluid pressure experiment during the draining of a cylindrical tube. The tube has a constant cross-sectional area, A, and volume $V = Ah(t)$, where $h(t)$ is the vertical distance from the end of the tube to the top of the water surface (see the inset in Figure C.4b). The tube has a small circular hole near the bottom with a much smaller area a . The water that flows down, and leaves the tube at flow velocity v, obeys the mass conservation statement $va + Adh/dt = 0$. After applying Bernoulli's principle between the two extreme points, separating variables and integrating from the initial time when the water height is h_0 to the time t leads to the well known solution for the water height as a function of time,

$$h(t) = \left(\sqrt{h_0} - \frac{kt}{2} \right)^2 , \qquad k = c\sqrt{2g}\frac{a}{A} \tag{C.2}$$

where $c \leq 1$ is the discharge coefficient of the orifice, and g the acceleration of gravity.

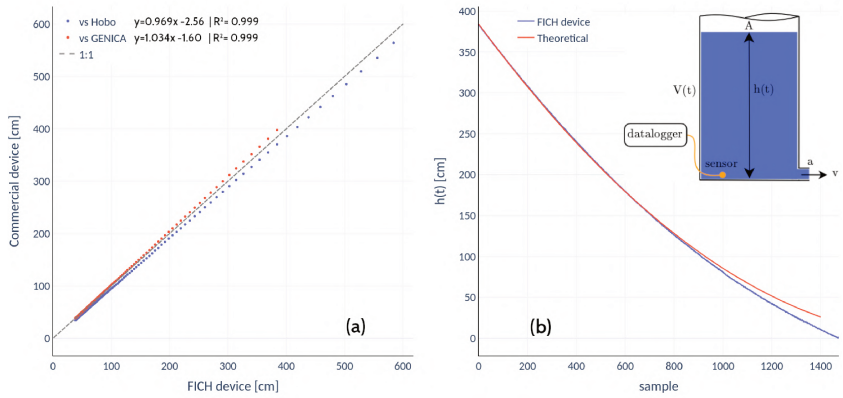


Figure C.4: **(a)** The best fit line obtained between the Honeywell sensor vs two commercial devices available in the market; **(b)** draining-tube test: theoretical vs experimental results ($a/A = 0.04$, $h_0 = 3.9$ m).

When there is no friction ($c = 1$), the flow follows Torricelli's law and the tube drains at time $t = t_1$. When the discharge coefficient is less than one, even after neglecting the tube wall friction, the above expression predicts that the tube drains at time $t = t_2$, where $t_2 > t_1$. The actual situation is likely to be between these two extreme cases. Consequently, instead of fitting the value of c using the least square method, it is much better to displace the experimental data to match the upper portion of the theoretical curve (Figure C.4b).

Moreover, as the height of the water decreases towards zero inside the tube, a mathematical model solely based on the mass conservation constraint and a simplified Bernoulli principle is likely to be not entirely accurate. Nevertheless, these two

tests show that the low-power datalogger coupled to the Honeywell pressure sensor can capture fast time-varying pressure phenomena with high accuracy.

Finally, a pumping test (Figure C.5a) was carried out with a nearby observation well (Figure C.5b), separated by 10 meters. The data shows the initial draw-down and the fast recovery once the pumping stopped (Figure C.5c).

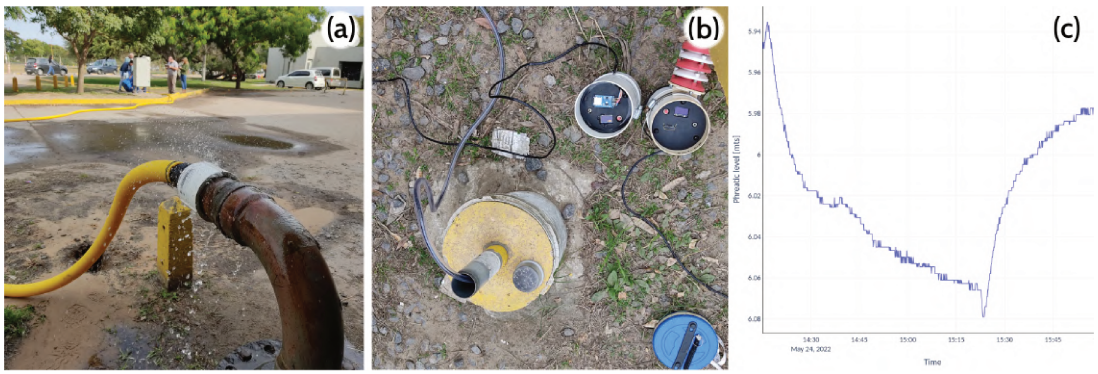


Figure C.5: **(a)** Running the pumping test at FICH; **(b)** the observation well where the pressure sensor was connected to the datalogger; **(c)** the collected data vs time.

2.8 Testing the SKU:SEN0193 Sensor

Previous research has shown that the SKU: SEN0193 is far from being a flawless sensor. It has been reported that porosity can severely influence capacitive soil moisture measurements. Its electronic design may also lead to parasitic capacitance that can misinterpret the soil water content. Consequently, they should operate at a high frequency; the higher the operating frequency, the lower the effect of losses related to the imaginary part of the permittivity (*Placidi et al.*, 2020). Others authors have suggested that the sensor accuracy depends on the soil mixture constituents (*Nagahage et al.*, 2019). Consequently, the study compared the sensor response in different soils and environmental conditions (laboratory and field).

Laboratory Tests

Initially, three tests were conducted for three types of soil, one in the laboratory with a clean sand (soil type 1) and the other two in the field with a sandy-loam soil (Table C.1). In all cases, the soil moisture was measured at 7 cm depth. In the laboratory, the gravimetric water content was determined by weight and then converted to volume. The samples were dried in an oven at 105^oC for 48 hours and then placed on two precision scales with water added until saturation. The readings from the two scales and the SKU sensors were stored simultaneously (Figure C.6a). Three SKU sensors were inserted into each container in such a way that, beyond the protector insulating film (see Section 2.4), there was no interference in their readings. The drying process was automatically recorded with a camera every 15 min.

In more detail, the data acquisition system consists of a webcam connected to a Raspberry Pi, with the SKU capacitive sensors connected to an Arduino Mega. Both systems were linked via serial connection. The Raspberry Pi takes photographs driven

by a Python script, which in turn receives readings from the Arduino platform. Both sets of data are stored with a timestamp for further processing. The readings from the scales were linked to the sensor measurements using a Python software that performs cropping, image enhancement and optical character recognition (OCR) on the photographs (PYSSO[8]), thus obtaining the weight of each scale. This procedure allowed to know the exact amount of water in the sample and the simultaneous reading of the sensors for a month.

The results show that the sensors reached saturation at approximately 80% of soil moisture, with a relatively slow response under drying conditions. In addition, the calibration changes from soil to soil and also from sensor to sensor, as shown in the boxplots from the lab and field tests (Figure C.6b).

Table C.1: Particle Size Analysis

Size (μm)	**Type 1 (lab)**	**Type 2 (field)**	**Type 3 (field)**
> 1000	0.0026	0.0126	0.1032
> 250	0.4859	0.1138	0.2620
> 63	0.4398	0.5281	0.3157
< 63	0.0717	0.3455	0.3191
Organic material (gr)	0.3978	4.2351	4.1549

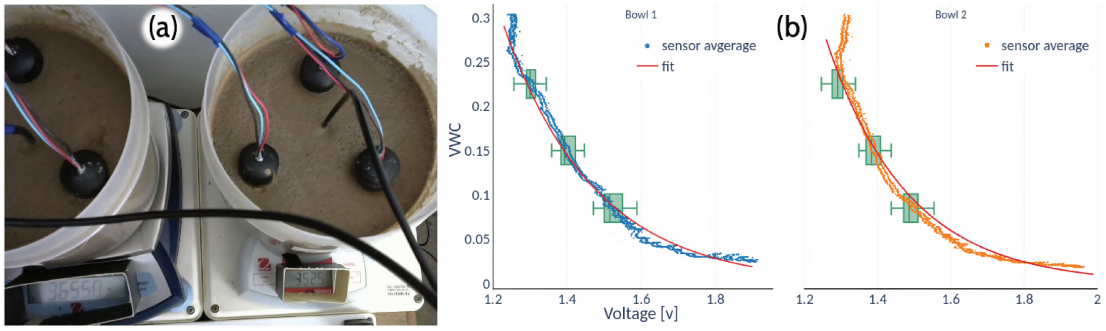


Figure C.6: **(a)** Acquisition of real-time data during a soil-drying test. **(b)** Collected data in the lab vs average voltage reading of the coplanar SKU:SEN0193 sensors and its fitted curves (bowl 1: $y = 43.23e^{-4.07v}$, bowl 2: $y = 77e^{-4.47v}$)

.

Field Test

The study site was located at Don Silvano SRL farm (latitude $-31°20'$, longitude $-61°8'$), close to the town of Humboldt (Santa Fe Province, Argentina, Figure C.7) whose main agricultural activity is the planting of pastures (alfalfa and corn) for the production of milk and meat. Two field trials were conducted for different soils, with the moisture sensors installed at 7 cm depth. The Hydraprobe II sensor, widely used in agronomy, provided the reference value of the soil moisture. The Hydraprobe II sensor estimates the

[8]https://gitlab.com/emilopez/pysso

volumetric water content (VWC) value directly. Simultaneously, the low-cost capacitive sensors' voltage output is inversely proportional to soil moisture. For the soil type 3 test, two SKU sensors kept reading for approximately three months (June–September 2022). For soil type 2 (Table C.1), the readings of six capacitive sensors were compared against the HydraProbe II sensor for approximately 1 month (August 2021, Figure C.8a).

Figure C.7: Location of the study site. The IoT device installed in the field above the soil surface counts, besides the groundwater level sensor (in green, 2) and the soil moisture sensors (indicated as types 2 and 3—see Table C.1), with several other devices. The groundwater monitoring well is located in a zone dominated by two regional topographic gradients, one oriented towards the little creek Las Flesias, and the other towards the creek Las Prusianas. Satellite imagery courtesy of Google Earth®. The two-month transmission test was carried out between the equipment installed at point 3 and the gateway node located 350 m away (orange point).

2.9 Calibration Algorithms and Dataset for the Soil Moisture Sensor

The term calibration refers to the process of correcting systematic errors in sensor readings, often by comparing a reference measurement from a first device with an uncalibrated measurement from a second device to adjust the parameters governing this second device in order to provide an accurate estimate *Barcelo-Ordinas et al.* (2019). In addition, the calibration process can be influenced by other measurements external to the second device due to correlations or cross-sensitivities present between different devices *Ferrer-Cid et al.* (2019, 2022). Specifically, we define as y_i the reference measurement, and as $\mathbf{x}_i$=[x_{i1}, . . . , x_{iM}] the vector that includes the measurement to calibrate and the external measurements, the calibration process consists of finding the function f:$\mathbb{R}^M \rightarrow \mathbb{R}$ that best approximates these measurements to the reference measurement:

$$y_i = f(\theta, \mathbf{x}_i) + \epsilon_i \quad , \tag{C.3}$$

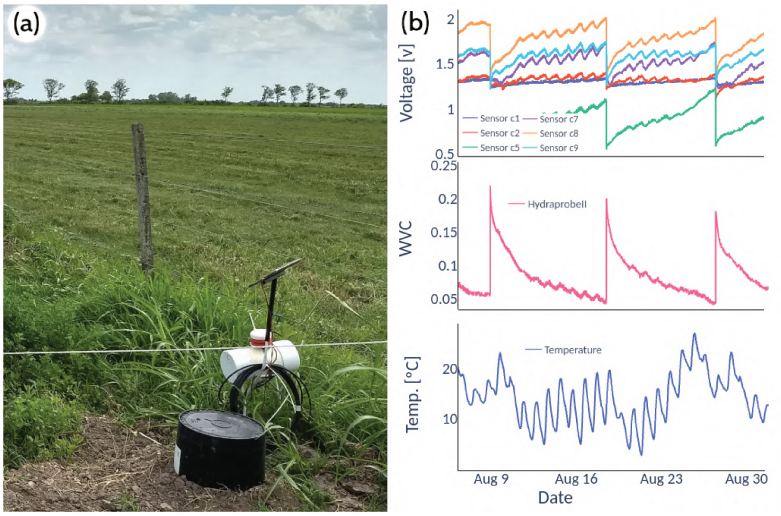


Figure C.8: **(a)** Acquisition of real-time data in field conditions; **(b)** field data of sensors SKU, Hydraprobe II and soil temperature.

with i=1,$\cdots$, N measurements, $f(\cdot)$ is the function used to calibrate the sensor and ϵ_i is random noise distributed following a normal distribution of mean zero and variance σ^2, i.e., N(0, σ^2). θ are the calibration model parameters to be optimized. There are different algorithms to estimate the function $f(\cdot)$, among which we have Multiple Linear Regression (MLR) if we consider the measured data to have a linear behaviour, K-Nearest Neighbors (KNN), Support Vector Regression (SVR) and Random Forest (RFO) if we consider the data to have a non-linear behaviour. The following is a brief description of the machine-learning methods used in the calibration process (more details on the methods can be found in *Hastie et al.* (2009)).

Multiple Linear Regression (MLR)

We consider that the dependent variable y has a linear dependence with the array of M sensors $\mathbf{x}$. The multi-matrix calibration model considers that the model for estimating the soil moisture for a new value $\mathbf{x}$ is:

$$\hat{y}(\mathbf{x}) = \beta_0 + \sum_{j=1}^{M} \beta_i x_j, \tag{C.4}$$

where the offset β_0 and the gains $\beta \in \mathbb{R}^M$ are the unknown parameters found by minimizing the sum of squares of the residuals.

K-Nearest Neighbor (KNN)

K-Nearest Neighbor is a memory-based model since the training data points are the model itself. Indeed, to obtain a new prediction for a point $\mathbf{x}$, we find the k closest points in the cloud and average their values.

$$\hat{y}(\mathbf{x}) = \frac{1}{k} \sum_{\mathbf{x}_i \in N(\mathbf{x})} y(\mathbf{x}_i), \tag{C.5}$$

where $N(\mathbf{x})$ is the set of points belonging to the neighbourhood of $\mathbf{x}_i$, and this neighbourhood is defined using the Minkowski distance as a metric. The hyper-parameters defining this model are the number of neighbours k and the power p of the Minkowski distance.

Support Vector Regression (SVR)

Support Vector Regression is a kernel, non-linear method that uses a convex optimization method to estimate the dependent variable. As the data in the data space need not be linear, SVR maps the data into a higher dimensional space called feature space, where a function can be found that is a linear combination of the mapped (feature) data. However, SVR uses the *kernel trick*, which consists of performing the calculations in the data space instead of the feature space by means of the kernel function k(x,x'). The estimation function is as follows:

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) k(\mathbf{x}, \mathbf{x}_i) + b \tag{C.6}$$

where $\mathbf{x}_i$ with i=1, $\cdots$, N are the training data points. The values for the parameters α_i, α_i^* are found by solving a quadratic optimization problem. The RBF (radial basis function) kernel was selected as the kernel function. The hyper-parameters of the model are the variance of the kernel, the margin in the loss function and a penalty term.

Random Forest (RFO)

Random Forest is a non-linear ensemble method, which constructs several uncorrelated decision trees from the training data and averages the response of all trees to produce the prediction; thus, the variance of the response is reduced. The prediction for a new observation is given by:

$$\hat{y}(\mathbf{x}) = \frac{1}{T} \sum_{i}^{T} tree_i(\mathbf{x}) \tag{C.7}$$

In this model, the required hyper-parameters are the number of trees T, the number of predictors F and the maximum depth D of the tree.

Dataset

For the field test (soil types 2 and 3, Table 1), which lasted approximately one and three months, respectively, the HydraProbe II volumetric soil moisture sensor was used as a reference value. Simultaneously, soil temperature, ambient temperature and relative humidity were measured to assess their impact. These variables were used as predictors taking into account that the reference reading or target variable is the one given by the HydraProbe II sensor.

Data for the type 2 soil (Table C.1) calibration were recorded between August 5 and August 31, 2021, totalling 3000 samples, using six low-cost soil moisture sensors in the test (c1, c2, c5, c7, c8 and c9). For the type 3 soil (Table C.1) test, the record lasts from June 17 to September 22, 2021, totalling 4200 samples, using two low-cost soil moisture sensors in this case (c10 and c11).

To calculate the hyper-parameters in each model, the dataset was divided into a training dataset and a test dataset. The training dataset was used with the 10-fold cross-validation technique to obtain the hyper-parameters, while the test dataset was used to check the performance for model comparison. A size of 70% was chosen for training data, i.e., with which the model is fitted, and the remaining 30% for testing, with which the goodness of fit of the calibration is analysed. This ratio was modified from 80%/20% to 60%/40% to assess the impact of the training and testing size. The predictions for the validation dataset were evaluated using the root mean-squared error (RMSE) and the coefficient of determination (R^2). In the pre-fitting stage, for the selection of the hyper-parameters, the mean bias error (MBE) was used.

3 Results

3.1 Low-Power IoT Datalogger

The IoT device was evaluated in the field for more than 12 months in stand-alone mode recording water table levels, soil moisture, soil temperature, relative humidity and ambient temperature (Figure C.8). To evaluate the real energy performance of each device, the battery voltage was also recorded through an analogue input of the microcontroller (with a voltage divider to ensure a range less than 5 V). A separate field test measured the power consumption of the IoT device with the wireless module for two months approximately, in conditions well above those required to evaluate its performance. The wireless transmission is a point-to-point link between the station and a gateway node with Internet access. The gateway node is connected to the power grid and listens permanently for connections from the measurement station. Once received, the gateway node transmits the data via an Ethernet connection to an API-REST system on the Internet[9].

Figure C.9a (in blue) represents the battery voltage of the IoT device in stand-alone mode, recording variables every 5 min with total solar exposure. Under this scenario, the amplitude of the charge–discharge cycle of the battery was 50 mV. Figure C.9b, in red, shows the voltage of the device measuring and transmitting wirelessly every 1 min and with partial solar exposure, 3 h per day ((i) and (iii)), and without harvesting solar energy (ii). Under situation (i), the range between the daily maximum and minimum was 80 mV.

3.2 Water Table Levels

Figure C.10 shows rainfall events with the IoT device and the Honeywell sensor recording the time variation of the water table levels. Since the rainfall was recorded manually by a farmer located 1 km away from the sensors, it is reasonable to expect some spatial and temporal discrepancy between the rainfall and groundwater records. Nevertheless, a rapid response of the water table to precipitation events is observed, followed by two slow time-scale responses. The slow groundwater response can be attributed to the rainfall water in excess, which, once infiltrated, is transported by the regional gradient. There is a slow albeit faster contribution due to groundwater discharge to the Flesias

[9] https://gitlab.com/ceneha/sentinel-iot

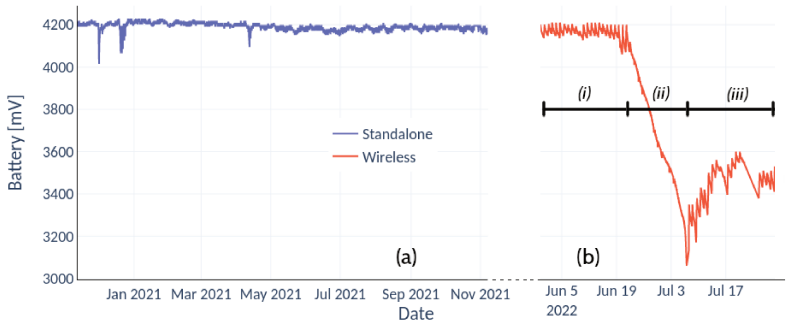


Figure C.9: **(a)** Battery voltage of the IoT device in stand-alone mode storing variables every 5 minutes with full solar exposure; **(b)** voltage variation when the IoT device transmits data every minute in conditions of reduced solar exposure, (i) and (iii), and when the solar panel is cut off (ii).

creek, followed by a slower temporal response given the larger surface area draining the water table aquifer, the groundwater discharge to the Prusianas creek (see Figure C.7).

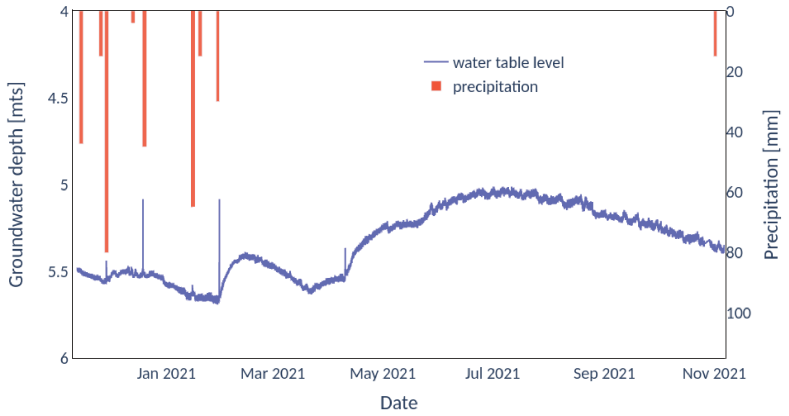


Figure C.10: Water table level variation at the study site (see Figure C.7, location 2).

3.3 Soil Moisture Sensors

On the one hand, the gravimetric test for clean sand uncovered the limited response of low-cost soil moisture sensors under controlled conditions. On the other hand, the field essays carried out in outdoor environmental conditions tested the endurance capabilities of the IoT devices based on open-source hardware.

Figure C.11 shows the calibrations obtained with the field test data for each sensor and soil type. In both cases, the nonlinear fits produced by the RFO algorithm achieved optimal results, with the lowest RMSE and highest R^2 (circles and crosses, respectively). Table C.2 shows the relative RMSE for each fit. For soil type 2, the average RMSE ranges between 2.1% and 4.6%, with a clear improvement of the nonlinear machine-learning techniques over Multiple Linear Regression. For soil type 3, the average RMSE represents 8.9% in the best case (RFO) and 13.2% for the worst fit (KNN). Figure C.12 shows the scatter plots between the RFO and the actual VWC value recorded with the

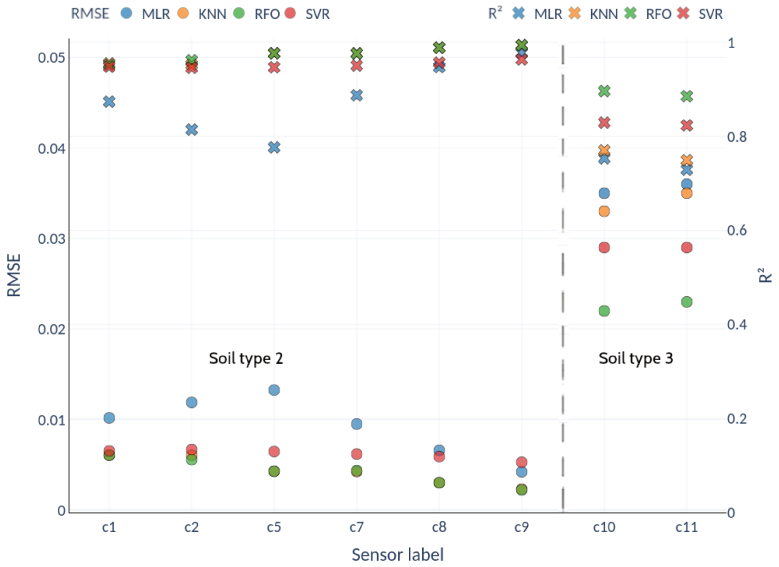


Figure C.11: R^2 (crosses) and RMSE (circles) values for each sensor with MLR, KNN, RFO and SVR machine-learning calibrations for soil type 2 (sensors c1 to c9) and soil type 3 (sensors c10 and c11).

Hydraprobe II sensor for the best-performing sensor (soil types 2 and 3, sensors c9 and c11, respectively).

Table C.2: Relative RMSE for the different calibration methods for soil types 2 and 3 (sensors c1-c9 and c10-c11 respectively)

Method	**c1**	**c2**	**c5**	**c7**	**c8**	**c9**	$\bar{x}$	**c10**	**c11**	$\bar{x}$
MLR	5.1	5.9	6.6	4.8	3.3	2.1	4.6	10.2	10.7	10.4
KNN	3.0	3.0	2.1	2.1	1.5	1.2	2.1	13.0	13.5	13.2
RFO	3.0	2.8	2.2	2.2	1.5	1.1	2.1	8.8	9.1	8.9
SVR	3.3	3.3	3.2	3.1	2.9	2.6	2.5	9.8	10.0	9.9

Figures C.13 and C.14 depict the predictions achieved with each of the aforementioned methodologies for two sensors of each soil type compared to the reference values measured with the HydraProbe II sensor.

4 Discussion

The IoT device was tested under adverse conditions in the field for more than 12 months with satisfactory performance. It is worth mentioning that *López et al.* (2020) details a field study of more than seven years collecting information on the factors leading to occasional failures of a flood warning system. The warning network, equipped with commercial devices, was exposed to outdoor environmental conditions similar to those mentioned here. The work documented the occasional and specific failures at each monitoring station, constituting one of the few known statistical works of this type.

In turn, the work analyses the associated costs of maintenance and replacement of the damaged equipment.

The energy consumption of the IoT device went below 1mA for stand-alone mode and 3mA when incorporating the wireless module, both in an idle state. Such a low threshold allowed recording variables at a frequency close to real time. In stand-alone mode and recording every 5 min, the equipment has an autonomy of approximately 20 d without solar harvesting. This estimate is arrived at by looking at the daily voltage drop of the batteries. Therefore, the minimum recommended values of the batteries are within reach in this period in case of not having a solar panel. It is possible to extend the period of 20 d by reducing the sampling frequency. Although its frequency of occurrence is low, periods of up to 14 consecutive cloudy days were observed, so the harvesting of solar energy guarantees continuous operation. The flexibility of the data logger allowed the incorporation of wireless communication modules. With transmission and measurements every minute, under a solar exposure of 3 h per day, the equipment achieved autonomy for 12 d.

In this case, the daily voltage drop was observed when the solar panel was disconnected from the equipment. This measurement is shown in Figure C.9b (ii). In that period, the voltages reached minimum values (below 3.2V), and although the equipment continued to operate, its behaviour could experience instability. The results show that a sampling frequency of 5 min with a few hours of sunshine per day is more than enough to guarantee continuous operation. Each device registers the voltage of its batteries, allowing it to anticipate a failure or to know if it is at the end of its battery life cycle.

The recording of water levels using the Honeywell differential pressure sensor showed that the measurements concerning commercial equipment are 1 to 1 (in both cases $R^2 > 0.999$), so calibration was not necessary. The housing made of crystal epoxy resin protects against salt water, where corrosion generally damages the probes.

The gravimetric test for the low-cost soil moisture sensors reported a considerable change in voltage to small increases in moisture for values below 0.01 VWC, while for values above 0.25 VWC there is no variation (Figure C.6b). Above 0.25 VWC, the sensor response is rigid, i.e., it reached saturation even though it reports readings up to 0.30 VWC.

For the field tests, the impact of the variables measured as predictors in all possible

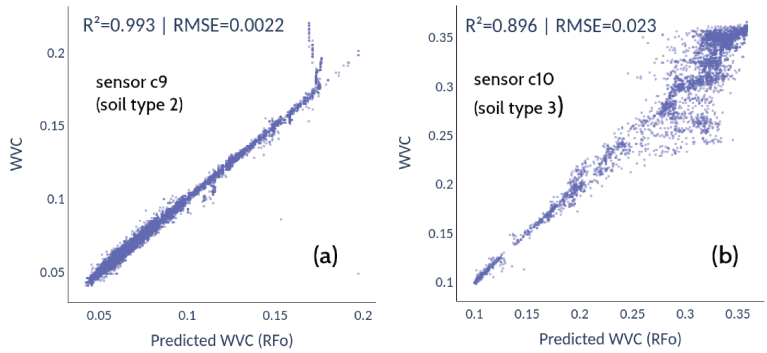


Figure C.12: The best fits for each soil type were obtained with Random Forest (RFO) for sensors c9 (left) and c10 (right).

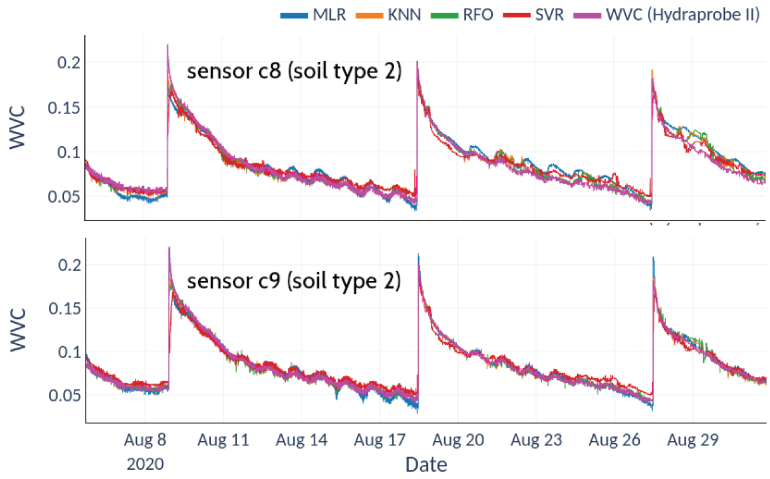


Figure C.13: Predictions from machine-learning models superimposed on HydraProbe II records for soil type 2.

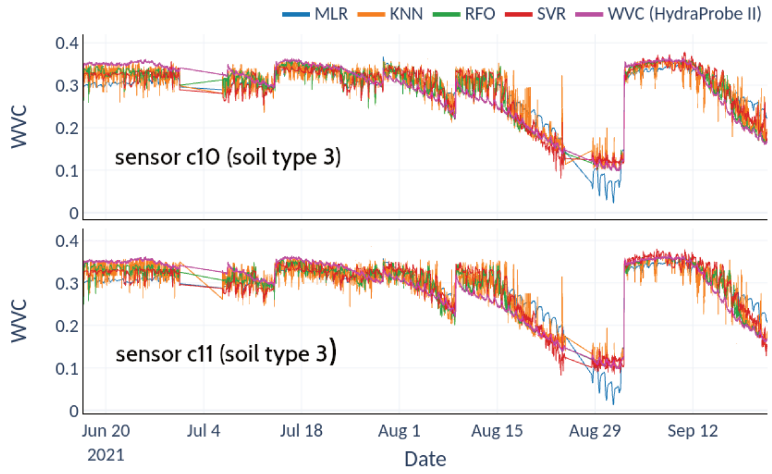


Figure C.14: Predictions from machine-learning models superimposed on HydraProbe II records for soil type 3.

combinations was analyzed, showing the preponderance of soil temperature. This observation led to discarding the rest of the variables (humidity and ambient temperature). Calibrations using nonlinear and linear machine-learning techniques gave lower errors for soil type 2 than for soil type 3 (maximum RMSE 4.6% and 13.2%, respectively).

The lower scatter exhibited by the RFO model in soil type 2 (Figure C.12a) translates into a fairly faithful reproduction of the reference observation (Figure C.13). In contrast, the higher scatter obtained for soil type 3 (Figure C.12b) manifests itself as a noisy variance reflected in the observed prediction (Figure C.14). Figures C.13 and C.14 show the predictions of the model, where the reconstructions for soil type 2 are more accurate than for soil type 3. The goodness of these regressions depends not only on the quantity and quality of the data but are also due to the training datasets not presenting enough variability to cover the full range of each variable. In soil type 2, there are many saturation and drying cycles, while soil type 3 exhibits only one drying cycle,

remaining saturated most of the time, which explains why there is not enough data in the training dataset, leading to overfitting and worse prediction at high values. Even so, the predictions for soil 3 follow the general trend of the soil moisture value, with a higher variance but without losing the average trend given by the HydraProbe II.

Last but not least, including preliminary tests and the field studies described above, the SKU sensors exhibited failures ranging from one in three to one in five (a failure rate between 20 and 33%).

5 Conclusions

In rainfed agriculture situations, knowledge of the water table level and the soil moisture is equally relevant to quantifying the water availability during the crop cycle and supporting decision making. Knowing the water reserves stored in the soil, farmers can decide whether or not to intensify their crop rotations, delay the sowing date to favour soil profile moisture recharge and decide which summer and winter crop sequences are best. The information may allow the producer to bring forward sowing to avoid waterlogging or loss of support in their plots (if the water table is very high).

This work demonstrated, first, how easy it is to install devices connected to the Internet, developing an "environmental monitoring" system of agricultural impact. Following *Beddows & Mallon* (2018), complex environmental systems are easy to implement with modest research budgets with the additional advantage of offering solutions of equivalent quality to high-cost commercial equipment, as demonstrated by *López et al.* (2020). Rephrasing *Fisher & Gould* (2012a), open source platforms allow people with short electronic experience to build instruments capable of collecting primary data. Here, the use of an IoT device based on an Arduino microcontroller with its wireless communication modules connected to two sensors capable of measuring variables of agro-hydrological interest was the starting point. The energy consumption was optimized to achieve sufficient autonomy to ensure continuous measurements equipped with a small solar power system. The designed probe to record water levels was the Honeywell differential pressure sensor. The lab test consisted of comparison against well known cases and with widely used commercial equipment. In all cases, the performance was satisfactory, recording the fast-time response close to real-time with characteristics superior to commercial alternatives with a significantly lower cost.

Another added advantage was the use of machine-learning techniques. The performances of the soil moisture sensor SKU: SEN0193 were analysed under laboratory (gravimetric test) and field conditions. To evaluate the field performance of the soil moisture sensor, the machine-learning calibration techniques of Multiple Linear Regression, K-Nearest Neighbor, Support Vector Regression and Random Forest were used. The models fitted the trend of soil moisture recorded by the reference sensor. Despite the very good predictions of the models, the relevance of having a trained dataset with a wide range of variability to improve calibrations was established. The RFO technique demonstrates that it is not necessary to have a calibration curve. Only a good set of reference data must be available, although this requirement and the complexity of the method limit the widespread use of the technique.

On the other hand, despite the sensor SKU: SEN0193 having an acceptable performance in sandy and silty soils, the recommendation is that a minimum of three sensors

should be deployed in the field and their readings averaged. In the group's experience, up to one sensor out of three stopped working in a brief period after deployment in the field. Further research includes testing the Decagon ECH20 10HS sensor[10], as it is an interesting option to pursue (it is 3 to 4 times cheaper than the HydraProbe II). In such conditions, the IoT would be able to integrate a comprehensive monitoring station of great utility for agronomic practices. Another issue is the current consumption since it rises to 3mA when the nRF24L01+ module is added. Despite this value being low, it is undesired behaviour since the wireless module is de-energizing like the SD and RTC modules. It was identified that this current consumption comes from the SPI port data lines (CS, MOSI, MISO, SCK). Subsequently, a way to avoid this should be analyzed to keep the current consumption as low as possible.

Author Contributions: Conceptualization, E.L, C.V., J.B.O. and J.G.V.; methodology, E.L, C.V., J.B.O., J.G.V., G.C. and J.P; software, P.F.C and E.L.; validation, E.L., G.C., J.P., J.M. and C.V; formal analysis, P.F.C. and J.B.O.; investigation, P.F.C., E.L and C.V.; data curation, E.L, J.P., G.C. and J.M; writing—original draft preparation, E.L, J.B.O. and C.V; writing—review and editing, C.V. and J.B.O.; visualization, E.L.; supervision, J.B.O. and C.V.; project administration, J.G.V., C.V and E.L; funding acquisition, C.V., E.L., J.B.O. and J.G.V. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by Universidad Nacional del Litoral, research grants CAID 5052019 0100249LI and CAAT 03012020, and the National Spanish funding PID 2019 107910RB-I00, by regional project 2017SGR-990, and with the support of Secretaria d'Universitats i Recerca de la Generalitat de Catalunya i del Fons Social Europeu.

Appendix C1

Figure C.15 shows a schematic diagram of the electronic circuit with interconnection between the components of the IoT device.

Appendix C2

The components and prices for the IoT device, the gateway node as well as the water table and the soil moisture probes with their commercial alternatives are listed in Table C.3.

Appendix C3

This Appendix includes a detailed view of the environmental housings for the IoT device, the water table probe and the soil moisture sensor.

[10]https://www.metergroup.com/en/meter-environment/products/ech20-10hs-soil-moisture-sensor

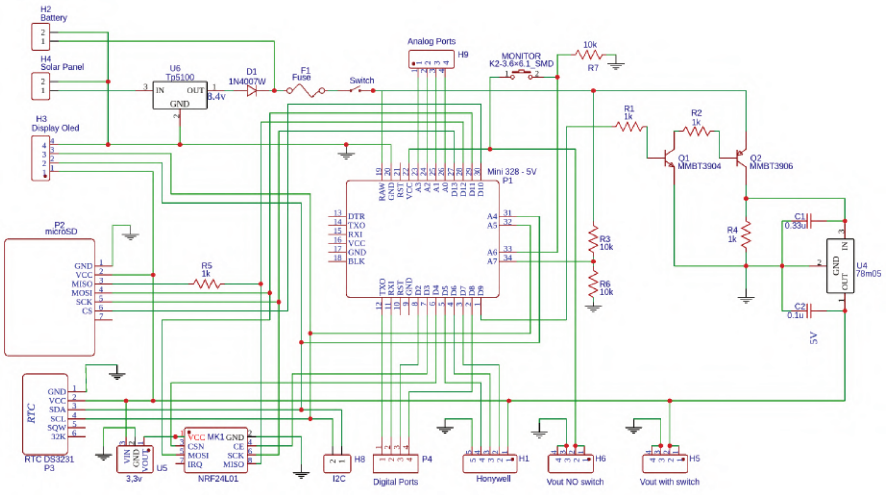


Figure C.15: Schematic diagram showing connectivity between different components of the IoT device.

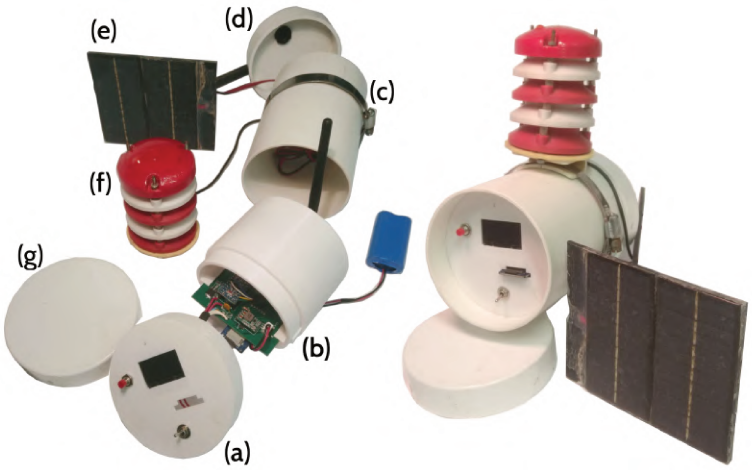


Figure C.16: Exploded view of the housing components of the IoT device: **(a)** Front of the device containing monitor button, ON/OFF switch, SD and Oled display (3D-printed part). **(b)** support for datalogger board, batteries and wireless module. **(c)** Four-inch PVC pipe. **(d)** and **(g)** Back and front cover. **(e)** Solar panel. **(f)** Relative humidity and air temperature sensor shield. On the right-hand side, all assembled housing parts.

Table C.3: Components and price list for IoT device, the water table probe and soil moisture sensor. Prices are based on delivery to Santa Fe, Argentina. Items without a suggested supplier are generally available at electronics retailers or hardware stores.

Component	Source	Cost/Unit
IoT Data-Logger		
PCB	Mayer S.A.	$8.60
microSD module	MercadoLibre	$1.65
RTC DS3231 module	MercadoLibre	$5.85
Arduino Pro Mini 5V 16 MHz	MercadoLibre	$8.15
nRF24L01+ PA + LNA SMA	MercadoLibre	$7.30
Battery 18650	MercadoLibre	$23.60
Solar panel	MercadoLibre	$12.85
CR2032 RTC battery	MercadoLibre	$0.80
Oled Display 1.3	MercadoLibre	$8.60
32 GB microSD (Sandisk)	MercadoLibre	$7.00
Housing misc parts: Poly Vinyl Chloride (PVC) pipe, PLA Filament, etc	various	$8.00
Electronic misc parts: wire, resistors, regulators, etc	various	$12.80
	Total in US Dollars:	$105.20
Gateway Node		
Arduino UNO	MercadoLibre	$18.00
nRF24L01+	MercadoLibre	$7.30
Ethernet shield	MercadoLibre	$30.00
Housing misc parts: PVC pipe, PLA Filament, cables, etc	MercadoLibre	$8.00
	Total in US Dollars:	$63.30
Water Table Probe		
Honeywell sensor	Mouser Electronics, Inc.	$50.80
UTP Cable C5 15m	MercadoLibre	$5.10
Crystal epoxy resin	MercadoLibre	$3.80
Glass tube 15m	various	$5.80
Steel ballast	various	$4.00
Housing and electronic misc parts:	various	$2.85
	Total in US Dollars:	$72.35
Soil Moisture Sensor		
SKU sensor	MercadoLibre	$3.90
Housing misc parts:	various	$1.50
	Total in US Dollars:	$5.40
Commercial Devices/Sensors		
GENICA (water table station, datalogger included)	Genica	$1000.00
HOBO U20L (water table station, datalogger included)	MeteoSur SRL.	$800.00
Aroya Solus Soil moisture sensor	Amazon	$580.15
Stevens HydraProbe Soil Moisture Sensor	Fondriest Inc.	$695.00
Decagon ECH20 10HS Soil Moisture Sensor	IdealSur	$182.00

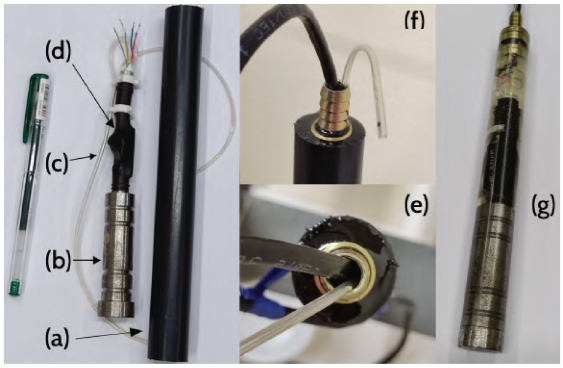


Figure C.17: Exploded view of the housing components of the water table probe: **(a)** One-inch plastic tube to contain the crystal epoxy resin during the filling process. **(b)** Steel ballast. **(c)** Nasogastric tube for water pressure. **(d)** Heat-shrink tubing to protect pressure sensor. **(e)** and **(f)** Crystal epoxy resin filling process. **(g)** Water table probe ready.

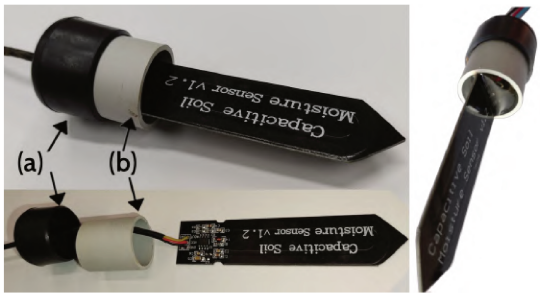


Figure C.18: Exploded view of the housing components of the low-cost soil moisture sensor: **(a)** Plastic cape tube to contain the crystal epoxy resin during the filling process. **(b)** Section of a plastic tube to avoid movement of the sensor during the filling process. On the right-hand side, soil moisture sensor ready.

Apéndice D

Estimating alfalfa growth and yield by measuring plant height with a low-cost, close-range scanning device

Resumen La alfalfa (*Medicago sativa* L.) es un cultivo forrajero que se puede cortar varias veces en el año, exhibiendo un alto rendimiento de materia seca (DMY, por sus siglas en inglés). Las estimaciones de DMY se basan en dos métodos: 1) correlación de datos y modelado sigmoidal del crecimiento, o 2) recopilación de datos mediante plataformas remotas o dispositivos de escaneo de corto alcance. Este trabajo demuestra que es posible utilizar un sensor acústico de corto alcance y bajo costo para monitorear el crecimiento de la alfalfa en diferentes etapas junto con otros sensores que registren parámetros agro-hidrológicos relevantes. Se utilizaron tres modelos analíticos independientes del crecimiento para evaluar la calidad de las mediciones de altura de la alfalfa. Los datos de cada ciclo de crecimiento se unificaron en un marco de referencia universal, lo que facilitó la comparación entre ciclos de diferente duración y temperatura acumulada. Una de los modelos surgió como una métrica novedosa para evaluar la precisión de las mediciones y los procedimientos de posprocesamiento realizados. Luego, se analizó la variabilidad del DMY utilizando regresión lineal múltiple. Se demostró que la altura de la planta es un factor significativo para determinar los valores finales de DMY. El sensor acústico con su hardware open-source asociado constituyen una herramienta de bajo costo para monitorear el crecimiento de la alfalfa que puede ayudar a los agricultores a tomar mejores decisiones sobre el momento adecuado para el corte y las expectativas de rendimiento[1].

[1]Este capítulo será enviado a *Field Crops Research*

Abstract: Alfalfa (*Medicago sativa* L.) is a forage crop with multiple annual harvests and high dry matter yield (DMY). DMY estimates are based on 1) correlation or modelling with sigmoidal growth curves, or 2) estimates using remote platforms or short range scanning devices. This work demonstrates the use of a low-cost, short-range acoustic sensor for continuously monitoring plant height and also other agro-hydrological parameters. Three analytical, growth-independent models were used to assess the quality of alfalfa height measurements. One of them is a novel metric (the Landau equation) for the evaluation of the accuracy of sensor measurements. The quality of the data collected for each growth cycle was improved through a series of AI-driven tasks and then reduced to a single reference frame to handle cycles with different cumulative temperatures. Forage yield was then analysed using multiple linear regression. As plant height was found to be a significant factor in determining final DMY values, the low-cost acoustic sensor and associated open-source hardware could provide a growth monitoring tool for alfalfa that could help farmers make more informed decisions about harvest and yield expectations.

1 Introduction

Plant height is used as a rapid indicator of plant response in phenotyping for the selection of high-yielding or stress-tolerant genotypes. It has been shown that plant height is one of the most important traits in plant ecology and is the result of many genetic and environmental factors (*Barmeier et al.*, 2016; *Bitella et al.*, 2024). It is related to yield and is a relevant parameter for rapid determination of crop response to variability in external factors (rainfall, soil moisture, temperature). However, the most common method for measuring plant height is still a ruler. Frequent visual measurements of plant height throughout the season are time-consuming, labour-intensive, costly and prone to human error (*Tucker*, 1980).

This work focuses on monitoring the growth of alfalfa (*Medicago sativa* L.) due to its widely recognized importance as a forage crop for livestock production (*Noland et al.*, 2018; *Dhakal et al.*, 2020; *Tucak et al.*, 2023). Continuous measurements are achieved using a low-cost, near-field scanning device.

This perennial legume is grown on approximately 30-32 million hectares[2] worldwide (*Acharya et al.*, 2020; *Jáuregui et al.*, 2022), with the USA accounting for 21% of the total area (*Fink et al.*, 2022). Argentina remains one of the largest alfalfa producers in the world and the first in South America with 1.5 million ha under cultivation (*Jáuregui et al.*, 2022; *Basigalup*, 2023).

Traditional field data collection or so-called direct methods for determining dry matter yields (DMY) in alfalfa are limited to destructive sampling and plant maturity assessment (*Tucker*, 1980; *Pittman et al.*, 2015). Indirect methods for DMY estimation include 1) methods based on data correlation (*Sanderson et al.*, 1994; *Noland et al.*, 2018; *Fink et al.*, 2022) or the use of sigmoidal growth models combined with the concept of growing degree days (G_{DD}) (*Mattera et al.*, 2013; *Zhou & Wang*, 2018; *Barriball et al.*, 2022), or 2) methods based on how data are collected, either using remote platforms (*Tucker*, 1980; *Erdle et al.*, 2011; *Kümmerer et al.*, 2023) or locally using short-range scanning

[2]One hectare -ha- is equal to an area of 10000 square meters.

devices (LiDAR-Light Detection and Ranging or acoustic sensors) (*Fricke et al.*, 2011; *Pittman et al.*, 2015; *Barker et al.*, 2016; *Banerjee et al.*, 2022).

Although correlation methods are useful, they require a large field dataset of specific agrohydrological parameters, which can be a burdensome task and locally dependent (*Fink et al.*, 2022; *López et al.*, 2022). It also requires the trimming of crop samples for further processing to determine fresh and dry matter, namely FM and DM, respectively (*Tucker*, 1980; *Mattera et al.*, 2013; *Noland et al.*, 2018). As a result, many crop management programmes only collect yield data at the end of the growing season without phenotypic information on yield and crop adaptation to stressful conditions (*Furbank & Tester*, 2011; *Barker et al.*, 2016).

The evolution of non-destructive techniques for the assessment of crops has resulted in a multitude of innovations over time. In particular, the advancement of ultrasonic and light detection technologies has been notable, with significant developments observed over the past two decades. The advent of close-range scanning sensor technologies has enabled the estimation of dry matter (DM) at the local level, utilising measurements derived from canopy or plant height. This has led to the establishment of a continuous record of parameters throughout the crop growth cycle.

Fricke et al. (2011) used an acoustic sensor mounted on a two-wheeled vehicle 1 m above ground to record grassland height measurements. To measure plant characteristics rapidly, *Barker et al.* (2016) built a mobile platform with two infrared thermometers, an ultrasonic sensor, and a multi-spectral canopy sensor. *Palleja & Landers* (2017) also demonstrate the ability of ultrasonic sensors to measure canopy density in apple orchards and vineyards. *Zhou et al.* (2021) propose a method to calculate canopy thickness using echo signals from ultrasonic sensors. These technologies were further validated by *Montazeaud et al.* (2021), who compared measurements from a low-cost ultrasonic sensor with manual determinations, although the autonomy of the Arduino® platform (*Mellis et al.*, 2007) only lasted for a few hours (see *López et al.* (2022)). More recently, *Bitella et al.* (2024) used a low-cost proximal platform with an ultrasonic sensor to flexibly measure plant height in either static or on-the-go mode, but also suffered from the low autonomy of the open-source hardware platform.

However, at the same time as acoustic sensor applications were being developed, another non-invasive technology was showing great promise. *Llorens et al.* (2011) demonstrated that a LiDAR sensor was an excellent choice for obtaining detailed insights into plant morphology, while ultrasonic sensors were suitable for determining average canopy characteristics in vineyards. *Pittman et al.* (2015) conclusively proved that the forage biomass of alfalfa, bermudagrass and wheat can be accurately estimated using a ground-based mobile LiDAR and an ultrasonic sensor. Their results were confirmed by *Banerjee et al.* (2022), who also found high correlations between manually measured plant height, FM and DM with measurements from a mobile LiDAR mounted 1.8 m above the ground. *Moreno & Andújar* (2023) and *Debnath et al.* (2023) provide a definitive review of several airborne and terrestrial LiDAR systems for use in agriculture. Despite the inherent differences between the two sensors, i.e. a LiDAR emits light waves that estimate biomass volumes (3D information) more accurately than 1D acoustic pulses, both acoustic and light detection must be consistent with accepted biomass estimation metrics, which implies correlation with destructive FM removal data (*Pittman et al.*, 2015).

The specific objectives of this work were therefore twofold. First, to demonstrate the

feasibility of using a low-cost, short-range acoustic sensor to monitor alfalfa growth at different stages. Three ultrasonic sensors were integrated with low power, open source hardware and firmware. The acoustic sensors were used to determine the effective height of the forage on a static basis. Other agro-hydrological parameters were measured simultaneously using low-cost sensors. Subsequently, three analytical growth-independent models were used as figures of merit to assess the quality of the alfalfa height measurements. Among the three models, a novel metric (the Landau equation) for evaluating the accuracy of sensor measurements is introduced. Considerable effort is devoted to the treatment of data measurement uncertainty. In addition, data collected from alfalfa growth cycles were reduced to a single reference frame, a technique that avoids the need to handle cycles with different cumulative G_{DD} values. Second, the goal was to determine the intrinsic ability of data collected by the instrument to estimate biomass production in comparison to DMY data obtained by destructive extraction. The variability of forage yield was analysed using multiple linear regression, which demonstrated the importance of plant height in determining the final values of DMY. The paper ultimately concludes that the acoustic sensor model may serve as a reliable means of inferring plant density through the vertical.

2 Material and methods

2.1 Data acquisition system

A low-cost, energy-efficient, multi-sensor data logger was used for in-situ collection of key agro-hydrological parameters (Fig. D.1**A**). This versatile design was adapted into different configurations to monitor: crop height using ultrasonic distance sensors, groundwater levels using differential pressure sensors (Honeywell HSCDAND015PDSA3) and meteorological variables using a rain gauge and anemometer. The basic system consists of a microcontroller (Arduino Pro Mini 5 V 16 MHz), a memory module (catalex microSD) and a real-time clock (RTC DS3231). The power system consists of two 18650 3.7V batteries in series, charged by a solar panel. Additional electronics condition the power to suit the device. The crop height monitoring platform contains three Maxbotix ultrasonic distance sensors and a DHT-22 ambient temperature and relative humidity sensor (Aosong Electronics Co Ltd, Guangzhou, China).

The low-power device, designed to measure agro-hydrological parameters in different environments, was developed at the Universidad Nacional del Litoral (Santa Fe, Argentina, *López et al.* (2020, 2022)). Fig. D.1**B** shows a snapshot of the low-cost system used at the Don Silvano SRL farm. Although historical rainfall records were not available at the study site, a weather station at INTA[3] Rafaela (see *Mattera et al.* (2013); *Berhongaray et al.* (2019) for more details), located 34 km to the west, provides a sufficiently long period of data that can be used, with some limitations, as an indicator of the prevailing hydrology in the region.

[3]National Institute of Agricultural Technology, Argentina

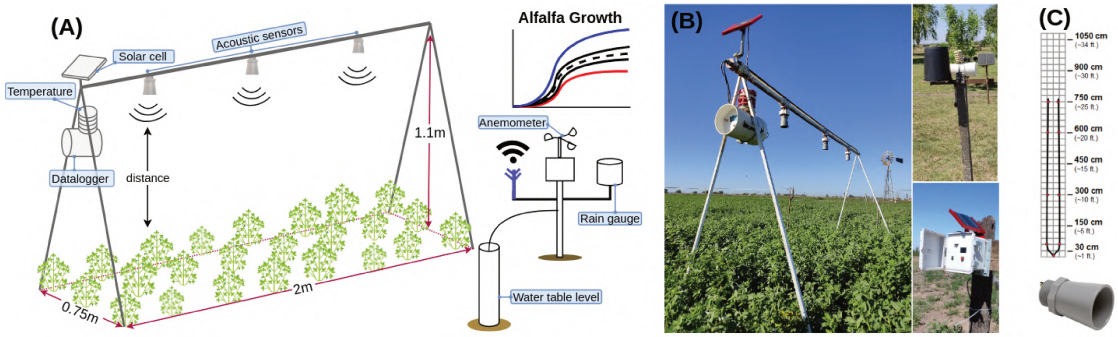


Figure D.1: **(A)** Schematic of the agro-hydrological station for measuring alfalfa height and other parameters, **(B)** low-cost station deployed in an alfalfa field, **(C)** MB7092 sensor and ultrasonic beam diagram (drawing taken from datasheet 2005-2022 ©MaxBotix Incorporated)

2.2 The ultrasonic distance sensor

The accuracy an ultrasonic sensor can achieve depends on the reflective properties of the target and other factors. The acoustic reflection from a hard, soft or rough surface will be different and may affect the achievable accuracy. The sensor model used for measuring alfalfa growth over time has certain characteristics that make it suitable for pinging on small plant leaves. Unlike other models, the Maxbotix MB7092 emits a coherent beam of sound, it receives the backscatter produced by all the surfaces reached within the 60 cm diameter of the beam and produces a single average value of all the reflections (*Palleja & Landers*, 2017). The sensor is housed in a weatherproof PVC case and the beam pattern is shown in Fig. D.1**C** (the manufacturer's drawing has been made at a scale of 1:95 for ease of comparison), with a detection range of up to 750 cm. The actual beam angle varies throughout the range.

The ultrasonic sensor operates at 42 KHz and provides a high-performance acoustic output along with real-time automatic sensitivity adjustment to changing conditions (supply voltage dips, acoustic or electrical noise). The Maxbotix MB7092 is equipped with noise rejection, enabling a resolution of 1 cm. It operates from a supply voltage of 3 V to 5 V and the output was configured for an analogue voltage signal.

Previous testing by the group has demonstrated the ability of the Maxbotix MB7092 ultrasonic sensor to record reliable information when the acoustic signal strikes a hard surface (see *Ferrer-Cid et al.* (2024b)). However, when the acoustic signal bounces off alfalfa leaves, the reflectivity of the surface can be considered both soft and rough. It was therefore necessary to implement signal filtering techniques to produce reliable echo data, as explained later in the text.

2.3 Study site and field experiments

The field trials were carried out between 2021 and 2024 at the Don Silvano SRL farm latitude 31° 19′ 13.359″ S, longitude 61° 8′ 23.982″ W, near the town of Humboldt (Santa Fe Province, Argentina). The main agricultural activity of the farm is the cultivation of pastures (alfalfa and maize) for milk and meat production. In Humboldt, summers are hot, humid and mostly clear, and winters are short, cool and partly cloudy. During the

year, temperatures generally vary between 7 °C and 32 °C, rarely falling below 0 °C or rising above 36 °C.

Many field trials have been carried out over 3 years (Nov 2020 to Feb 2024) on three different plots within the farm. The measurements and samples were randomly selected from time to time within these three locations. The first version of the equipment was installed in a field with a two-year-old alfalfa crop. A second version of the equipment was then installed in the same field. The improved and final version of the equipment was placed in a alfalfa field that was less than one year old. The farm soil is defined as PIL-03 - 2w(P)-67 silt loam and the crops are rainfed. The alfalfa variety used was Alphatec 921 (Forratec). Before sowing, the seedbed was prepared with a disc harrow. The seeds were sown in rows at a distance of 17.5 cm and at a depth of about 1 cm. Seed density was 13.5 kg/ha. No fertiliser was applied before sowing and pesticides were used to control weeds, insects and other diseases that could affect the crop.

Fig. D.2 shows, from top to bottom, a historical series of rainfall recorded at INTA's Rafaela station and other parameters measured at the study site by the low-cost device, including rainfall, ambient temperature, crop height (acoustic bounce before filtering), groundwater level, and wind speed. While INTA's Rafaela station recorded an average annual rainfall of 1004 mm over the last 54 years (Fig. 2, top), the last four years had annual rainfall of 682, 909, 607, and 900 mm. Immediately after a significant rainfall event, which occurred at the end of January 2021, a rapid response (peak) of the groundwater level is observed. The sensitivity of the pressure sensor used was extensively tested (*López et al.*, 2022). The groundwater depth was about 5.5 m. It is likely that a small portion of the plant's water intake was supplied by groundwater.

Records of all variables listed above spanned fifteen complete growing cycles of alfalfa. Destructive bulk sampling of the forage to determine FMY and DMY was performed manually by mowing the alfalfa approximately every 30 days. Before each mowing, the height of the crop was automatically measured by the acoustic sensor and recorded by the data logger. At the same time, the height of the crop was measured manually with a ruler. The sampling area was 2 m long, 0.56 m wide and covered 3 rows. One sample was taken at a time for each regrowth period, leaving a stem of approximately 12 ± 2 cm in height. Each sample was then preserved for weighing and dried in an oven at 65°C for 48 hours.

2.4 Data processing and filtering

Depending on the density of the alfalfa leaves and how the ultrasonic beam bounces off them, the canopy can behave like a multi-layered mixture of soft, porous material with the soil surface as the final rigid layer. Or, conversely, a hard, rough surface with good reflective properties. Fig. D.3 shows a number of situations where the ultrasonic beam pings on alfalfa with different textures. Most of the field experiments were carried out in years of pronounced drought, with deep groundwater in the last three years (more than 7.5 m below surface). The groundwater levels shown in Fig. D.2 were recorded in 2021 when the annual precipitation was about 10% below its mean, whereas in 2022 it was on the order of 40% below average. Fig. D.3 covers a data window of seven growing cycles and, for simplicity, six snapshots, showing situations of low, medium and high leaf density. The first two are the result of little or no rainfall. The next two are with moderate rainfall and the last three with significant rainfall during the first week after

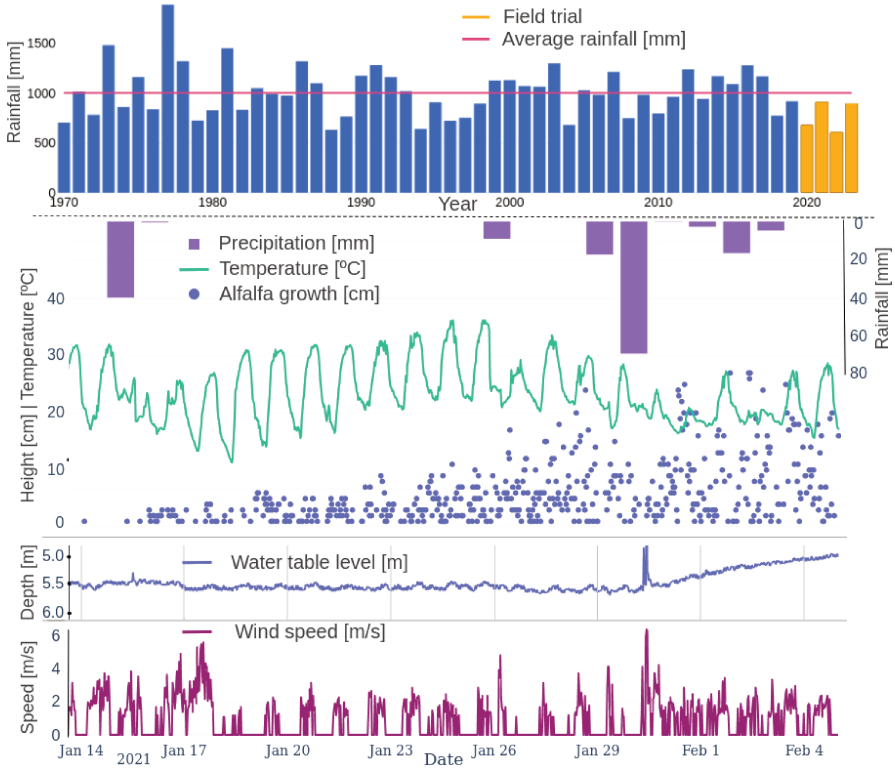


Figure D.2: Recorded variables for an alfalfa growth cycle. From top to bottom: Historic record of rainfall (INTA Rafaela), local measurement of daily rainfall, ambient temperature, plant height (ecosounder bounce), groundwater levels and wind speed.

each cut. At the top of Fig. D.3, the arrows pointing down indicate the dates on which the snapshots were taken. As can be seen, the problem arises when water stress deficit causes the plant to behave like a porous material as it grows. This causes significant scatter in the received signal.

Dry cycles with sparse canopy cover (snapshots on the left) show scattered data from the ground up to the maximum height of the crop (envelope). The case of medium canopy density (snapshots in the middle) shows a better-defined envelope. However, it still contains highly scattered data below the envelope. Finally, the wet cycles show a high concentration of points around the top of the canopy, with empty space between the canopy and the ground.

It was therefore necessary to filter the raw noisy signal to obtain a robust estimate of plant height as a function of time between cuts. Data pre-processing was carried out using the Python scientific ecosystem[4] (*Barba*, 2021). The quality of the data collected was improved through a series of AI-driven tasks (*Ferrer-Cid et al.*, 2024b). First, anomalous values for alfalfa height were identified and removed from each sensor's readings. Next, the upper envelope of the data set measured by each sensor was calculated and smoothed using a LOEWSS filter (locally weighted scatter plot smooth (*Cleveland*, 1979), see Fig.D.4).

[4] Python 3.11, Numpy 1.24.2, Pandas 1.5.3, Scipy 1.10, Sklearn 1.2.1, Statsmodels 0.13.5 and Plotly 5.13.0

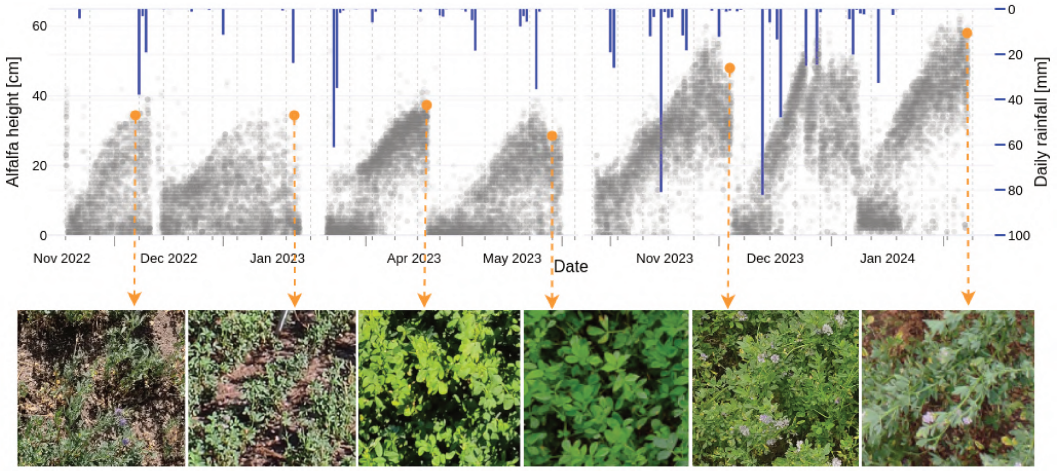


Figure D.3: Top: Raw data for the ultrasonic sensor for different growth cycles. Bottom: Snapshots of the canopy taken at the moment indicated by the orange arrow (no snapshot of the penultimate growth cycle was included to maximise the size of the other images).

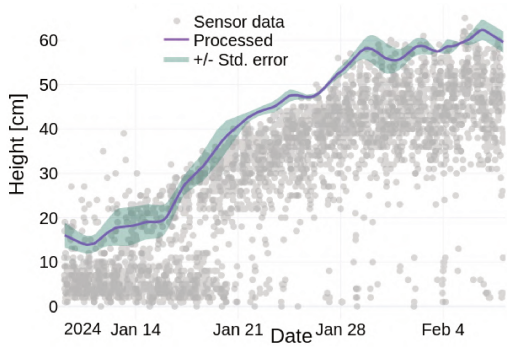


Figure D.4: Growth curve envelope after raw data processing.

The signal processing described above is in line with the capabilities of the sensor. The sensor averages the acoustic bounces it receives. The advantage of this is that a density function can be obtained from the penetration of the signal in the vertical profile of the crop. Fig. D.5**A** shows that during dry cycles with low vegetation cover, the signal has a bimodal distribution (green colour). It covers the whole range of heights from the ground to the top of the crop (envelope). In contrast, wet cycles (purple) show a clear concentration of points around the upper part of the canopy, with virtually no leaf or lower stem values.

The concept of heat units has a major impact on the growing cycle of the plant, from sowing to harvest, and is calculated in terms of G_{DD} (*Elnesr & Alazba*, 2016). The simplest way to calculate G_{DD} is to take the maximum plus the minimum daily temperature divided by two (or the mean temperature) minus the threshold or base temperature T_b, which must be exceeded for growth to occur. In short, alfalfa development is negligible at temperatures below 5 °C. The data logger was set to measure ambient temperatures, $T = T(t)$, every half hour ($= \Delta t$), where T is temperature and t is time (fraction of day, $= d$). So that using the Heaviside function $H(T)$, which is = 0 when $T_{avg} - T_b \leq 0$ and = 1 otherwise, where $T_{avg}(t) = (T(t)+T(t+\Delta t))/2$, the cumulative G_{DD} [°Cd] calculated

for a crop growth cycle using the trapezoidal rule is

$$G_{DD} = \sum_{i=1}^{m_r} H\left(T_{avg}(t_i) - T_b\right)\left(T_{avg}(t_i) - T_b\right)\Delta t \quad , \tag{D.1}$$

where i represents each measured temperature and m_r is the total number of readings up to the end of the regrowth period. Only 12 out of 15 alfalfa growth data cycles were reliable. Three data sets were discarded because the signal was either corrupted or too noisy. The relationship between G_{DD} and time can be constrained to an approximately wedge-shaped linear variation with a slope α ranging from 11.1 °C to 19.2 °C, with a mean about 15.2 °C (Fig. D.5**B**).

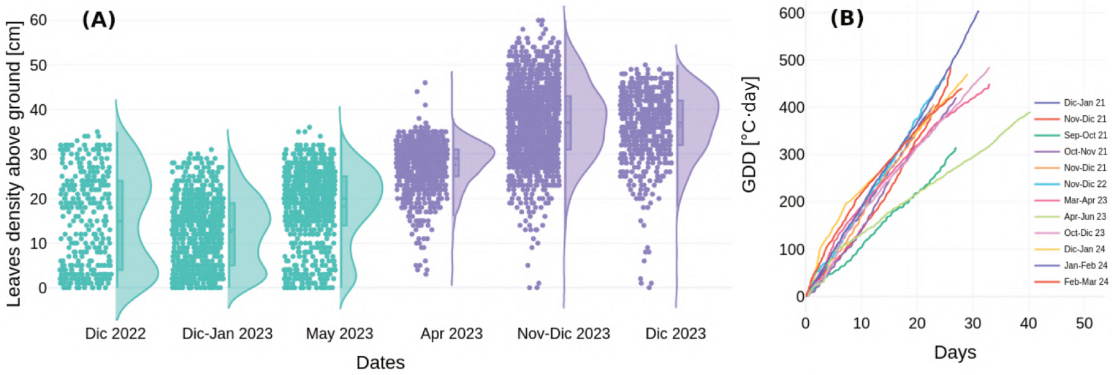


Figure D.5: **(A)** Acoustic signal distributions for low and high leaf density (green and violet respectively), **(B)** Curves of cumulative G_{DD} [°C·d] as a function of time [d :day], for all growing cycles.

2.5 Statistical analyses

The accuracy of the sensor has previously been validated by comparing acoustic bounces from a hard surface with manual readings (*Ferrer-Cid et al.*, 2024b), a test that is repeated here with a soft surface target (leaves) instead. However, random uncertainties are an unavoidable consequence of measurement, no matter how accurate the sensor. Fig. D.4 and D.5**A** illustrate some of the large variability in the observed values and the uncertainties associated with the plant growth process. Whenever more than one measurement is made, the values obtained will not be equal but will be scattered around a mean value considered the most reliable. This range of values is known as random uncertainty. Assume that $y_i = y(t_i)$ are observations of plant height at time t_i and that $\hat{y} = \hat{y}(t, \mathbf{c})$ is the expected growth model according to an unknown distribution or "true model" $\hat{y}(t)$. The expected growth model depends on the time t and a set of parameters $\mathbf{c} = (c_1, \ldots, c_p)$. If it is further assume that each data point has a completely random measurement error with standard deviation σ_i and is normally (Gaussian) distributed around the "expected model" $\hat{y}(t, \mathbf{c})$, then an appropriate merit function is

$$\chi^2(\mathbf{c}) = \sum_{i=1}^{n} \left(\frac{y(t_i) - \hat{y}(t_i, \mathbf{c})}{\sigma_i} \right)^2 \quad , \tag{D.2}$$

called the "chi-squared", where n is the number of observations. Finding the best set of $\mathbf{c}$-parameters that minimises the sum of $\chi^2(\mathbf{c})$ will validate the model (maximum likelihood estimator), while allowing some error bars to be assigned to the data points (*Press et al.*, 1989). If the residuals are $e_i = y_i - \hat{y}_i$, then the fractional, or normalised, root mean square deviation (*rmsd*), and the coefficient of determination (r^2), defined as

$$rmsd = \frac{1}{(y_{max} - y_{min})} \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^2} \quad , \tag{D.3}$$

$$r^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=n}^{n} (y_i - \bar{y})^2} \quad , \tag{D.4}$$

are also useful metrics for assessing the uncertainty associated with the fit of a model to experimental data. Above, $\bar{y}$ is the mean of the measured data $y_i = y(t_i)$ and $\hat{y}_i = \hat{y}(t_i, \mathbf{c})$ is the fitted model that is as close as possible to the true model.

2.6 Crop growth modelling

The growth of plants is subject to a number of environmental factors, making it impossible to determine the precise route traversed by the plant from the initial stages to the final stages of the growth cycle. It is therefore beneficial to utilise straightforward yet autonomous phenomenological growth models to evaluate the dependability of alfalfa height measurements obtained by the acoustic sensor. If the scatter of the data is minimal or the random errors are normally distributed, the models may not diverge too much from the true growth model. Three models are presented here for sigmoidal or S-shaped growth.

$$y(\tau, \mathbf{c}) = \begin{cases} \dfrac{1}{2} + \dfrac{1}{2} tanh(\mathrm{c_1}\tau - \mathrm{c_2}) & : \text{ Fisher,} \\ \sqrt{\dfrac{1}{2} + \dfrac{1}{2} tanh(\mathrm{c_1}\tau - \mathrm{c_2})} & : \text{ Landau,} \\ \dfrac{1 - tanh[exp(\mathrm{c_2} - \mathrm{c_1}\tau)/2]}{1 + tanh[exp(\mathrm{c_2} - \mathrm{c_1}\tau)/2]} & : \text{ Gompertz ,} \end{cases} \tag{D.5}$$

where τ and $y(\tau, \mathbf{c})$ are dimensionless $\mathrm{G_{DD}}$ and plant height respectively. The interpretation of $\mathrm{c_1}$ and $\mathrm{c_2}$ varies across models; however, statistical fitting provides greater precision. All models predict growth saturation at $\tau_* = \mathrm{c_2}/\mathrm{c_1}$. The inflection point is easy to identify for Fisher and Landau. The Gompertz solution can be cast in terms of chained exponential functions, but finding the onset of saturation requires going up to the second derivative (*Chu*, 2020). See 5.

2.7 Yield model

An effective approach to identifying patterns in the data is to use linear regression (SLR) and multiple linear regression (MLR). A linear regression consists of a standard procedure, whereas a DMY can be obtained as follows

$$\mathrm{DMY} = a_1\, \varphi_1(t) + a_2\, \varphi_2(t) + \ldots + a_m\, \varphi_m(t) \quad , \tag{D.6}$$

where the constant at the origin is absent if standardised values of the time-varying set of m-independent predictors, $\varphi_1(t), \ldots, \varphi_m(t)$ are used. These include water table depth, rainfall, plant height, G_{DD}, etc. They are selected according to data quality, quantity and availability. See 5 for more details.

3 Results

3.1 Data Acquisition and Processing

The acoustic sensors are weatherproof (IP67 certification), but the conical shape created an environment conducive to the accumulation of cobwebs, which interfered with the signal and the accuracy of the readings. The sensor performed well in the field, but the datalogger had an electrical failure in 2022 due to water ingress, requiring extensive repairs in the laboratory. This setback led to a redesign of the station structure to facilitate relocation. Sensors were positioned in a linear configuration for measurement across a furrow (see 5).

A scatter plot of data from alfalfa samples assessed the ultrasonic instrument's performance. Comparing the acoustic sensor readings with manual measurements before destructive sampling showed a strong correlation, with an $r^2 = 0.85$ (Fig. D.6**A**). The three curves for each sensor were averaged to produce a single growth curve for each period, then *i)* the growth models were fitted with 12 complete growth cycles (Fig. D.6**B**), *ii)* the yield models used data from 46 alfalfa cuts (linear regression) and a subset of 25 observations with summed values of the predictor variables (see 5 for more details on MLR).

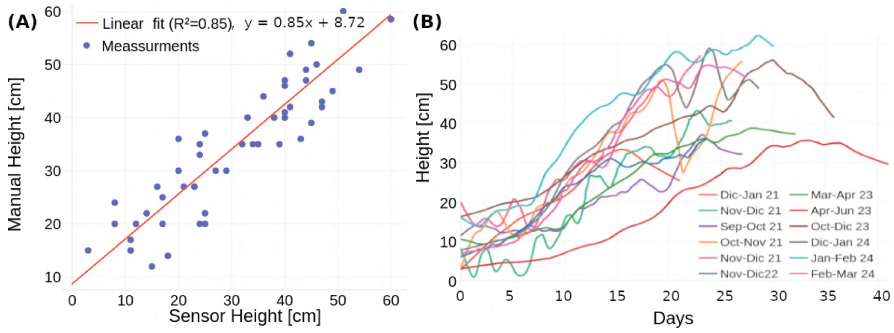


Figure D.6: **(A)** Scatter plot between manual and automatic measurements, **(B)** Growth curves as a function of time after filtering and smoothing.

3.2 Growth modeling

Table D.1 shows the parameters obtained after fitting the 12 alfalfa regrowth periods observed between November 2020 and March 2024. The tabulated data are: the duration of the regrowth cycle (in days), the time window of the field experiment, the number of points recorded in-situ by the automatic station, the total accumulated G_{DD}, the fitting constants $\mathbf{c_1}$ and $\mathbf{c_2}$, the onset of saturation in terms of G_{DD}, and finally the *rmsd* (in percentage) and the r^2 coefficients. The Fisher's fit (Fig. D.7**A**) has a relative

days	Date	pts	G_{DD_M}	$\mathbf{c_1}$	$\mathbf{c_2}$	G_{DD*}	*rmsd*	r^2
26	Sep-Oct 21	1088	314	1.575	0.626	125	11.4	0.87
27	Oct-Nov 21	1129	420	1.755	0.722	173	10.2	0.83
25	Nov-Dic 21	964	485	3.206	1.551	235	6.6	0.96
23	Nov-Dic 21	908	405	1.849	0.838	184	8.2	0.93
21	Dic-Jan 21	954	409	2.611	0.840	132	7.6	0.93
24	Nov-Dic 22	512	462	2.569	0.820	147	5.4	0.97
32	Mar-Apr 23	1168	448	2.648	1.028	174	6.1	0.98
41	Apr-Jun 23	999	409	3.735	1.549	169	4.2	0.98
32	Oct-Dic 23	911	484	1.875	0.490	126	5.3	0.97
28	Dic-Jan 24	648	571	2.616	0.826	189	9.0	0.88
30	Jan-Feb 24	1174	603	2.864	0.834	176	4.0	0.99
29	Feb-Mar 24	755	440	3.326	1.183	157	3.2	0.99
						$\tau_* = \mathbf{c_2}/\mathbf{c_1}$		
–	Fisher	41	$\sigma_i \neq 1$	2.313	0.849	0.367	2.5	0.99
–	Landau	41	$\sigma_i \neq 1$	2.987	1.688	0.565	2.7	0.99
–	Gompertz	41	$\sigma_i \neq 1$	3.500	0.794	0.227	4.4	0.98

Table D.1: Parameter values obtained after fitting the logistic or Fisher growth curve to different regrowth periods. G_{DD*}=G_{DDM} $\mathbf{c_2/c1}$ indicates the accumulated temperature at the onset of saturation. The last three rows show the fit obtained with the $\chi^2(\mathbf{c})$ (Eq.(D.2)) over the normalised range, and the dimensionless G_{DD} at the onset of saturation.

rmsd between 3.2% and 11.4%, with an r^2 between 0.83 and 0.99. Although the forage growth cycles were measured at different times of the year, a climatic factor that influences the time required to accumulate enough solar energy to reach maturity, the fit predicts that the onset of saturation is reached on average at G_{DD*} = 166 ± 30 °Cd. Fig. D.7**B** shows the scatter plot obtained by comparing all the Fisher's fits to the normalised height measurements.

The three theoretical models were fitted by minimising the χ^2 with respect to the mean values of all normalised growth cycles (see 5). The fit gave a relative *rmsd* of 2.5, 2.7 and 4.9% for the Fisher, Landau and Gompertz models respectively (Fig. D.8**A** and Fig. D.8**B**). Similarly, the fit gives an r^2 = 0.99 for Landau and Fisher and = 0.98 for Gompertz (see Table D.1). The poor ability of the r^2 coefficient as a metric for determining which model best fits the data can be circumvented by keeping in mind that as the number of data increases, the χ^2-distribution becomes approximately Gaussian. Thus, a good fit will have nearly 95% of the data within ±two normalised residuals of the theoretical function (see Table D.2), and its histogram will be approximately Gaussian (Fig. D.8**C**).

3.3 Crop yield

It is possible to obtain a fair and statistically significant SLR (r^2 = 0.74) between crop yield and plant height readings produced by the acoustic sensors. Fig. D.9 illustrates

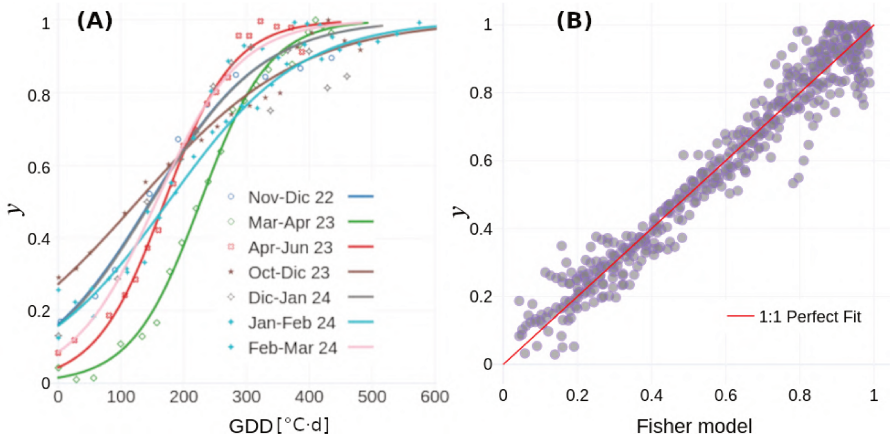


Figure D.7: **(A)** Fit curves for different growth cycles overlaid on the measured data with the accumulated G_{DD} as the independent variable (not all cycles are shown for brevity), **(B)** Scatter plot between measured data and Fisher model for the 12 growth cycles.

Std	Gauss [%]	$\bar{y}$ [%]	Fisher [%]	Landau [%]	Gompertz [%]
$\pm\sigma$	68.3	66.3	64.8	64.8	64.6
$\pm 2\sigma$	95.4	95.7	94.7	95.9	92.3
$\pm 3\sigma$	99.7	99.8	100.0	99.6	99.8

Table D.2: Frequency of the sample data for the 12 growth cycles showing the distribution of normalised residuals ($\sigma = 1$). Nearly 95% of the data lie within ± 2 error bars for the Fisher and Landau models.

the correlation, with the manual plant height measurements superimposed on the plot to demonstrate the consistency of the result.

An optimal trade-off between quantity and quality in forage crops represents the most relevant objective for the producer (*Testa et al.*, 2011). Nevertheless, the DMY/FMY ratio demonstrates a seemingly gradual oscilation over time, with values oscillating be-

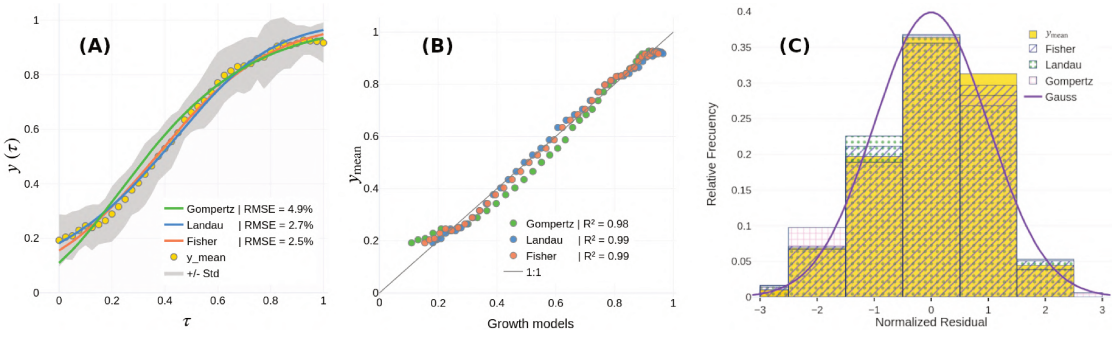


Figure D.8: **(A)** Averaged growth cycle after minimising χ^2 with the Fisher, Landau and Grompertz models (grey area shows data spread), **(B)** scatter plot between each growth model and the means, **(C)** distribution of normalised residuals.

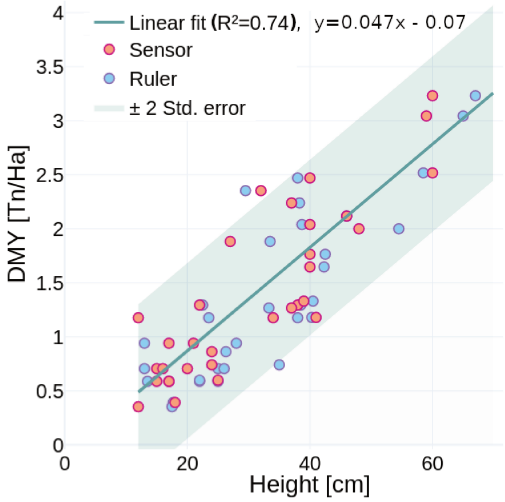


Figure D.9: DMY versus automatic readings of alfalfa height.

tween 20 and 35% (Fig. D.10**A**). However, this narrow range of variation is not observed when the role of other parameters is taken into account. A multiple linear regression (MLR) was derived with cumulative rainfall ($\varphi_1(t)$), final plant height ($\varphi_2(t)$), and cumulative growing degree days (G_{DD}) ($\varphi_3(t)$) as predictors during each regrowth period. In both instances, data from the automated station were employed. The calculation did not include groundwater depth, which was observed to be more than 7.5 m below the ground surface for the majority of the observation window. The size and colour of each bubble in Fig. D.10**B** illustrate the relationship between DMY and these three key variables. All values measured during the growing cycle are final or cumulative. The yield and therefore dietary protein content can be up to three times higher than under less optimal conditions, depending on the cumulative G_{DD} and total rainfall. The final MLR equation, with $r^2 = 0.80$ and relative $rmsd = 12.3\%$, is

$$\text{DMY} = 0.010\ \varphi_1(t) + 0.621\ \varphi_2(t) + 0.327\ \varphi_3(t) \tag{D.7}$$

The scatter plot between the predicted (MLR model) and measured values is shown in Fig. D.10**C**.

4 Discussion

Alfalfa height is a key indicator of healthy growth. This paper presents three parts: *i)* data collection using a low-cost device, *ii)* data processing and filtering, *iii)* modelling of plant growth and yield. The paper describes a close-range ultrasonic instrument used to record the growth of alfalfa. It works with an automated station to measure other hydro-environmental parameters. The system is open source hardware. Unlike other close-range devices, the acoustic sensor is low-cost and accessible. The acoustic sensor offers a low-cost, accessible alternative to other close-range devices like LiDAR. It can track plant height and morphology, and is reliable enough for use in field studies.

Fig. D.5**A** shows that the acoustic sensor can penetrate the plant from top to bottom. Despite this, this work was limited to obtaining plant height data to infer crop

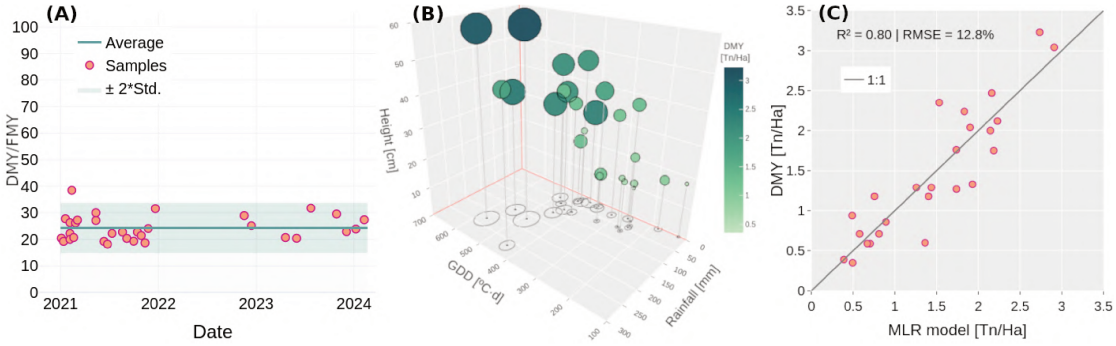


Figure D.10: **(A)** DMY/FMY ratio, **(B)** DMY versus rainfall ($\varphi_1(t)$), crop height ($\varphi_2(t)$), and accumulated G_{DD} ($\varphi_3(t)$), during the growth cycle, **(C)** Scatter plot between modelled and observed DMY.

growth and yield. Subsection 3.2 shows that sigmoidal growth curves fit the experimental data. This communication's novelty was using normalised variables to treat all alfalfa regrowth periods in a unified framework. This helped deal with the uncertainty of measured plant height and filter out noise.

The results show that all measurements are reliable and that the instrument and the post-processing procedures did not introduce any bias. Three different growth models fitted by maximum likelihood estimation confirmed that the data were normally distributed around the mean (Table D.2). Two sigmoidal growth models (Fisher and Landau) divide growth into two halves and show symmetrical behaviour. The Gompertz model is asymmetric. Table D.2 and Fig. D.8**C** show that the Fisher and Landau models best approximate the data's Gaussian distribution. The Landau model fits slightly better.

Incidentally, the use of the Landau equation is another novel contribution of the paper. Apart from the Gompertz model, Fisher and Landau predict linear growth after exponential growth (around the inflection point of the hyperbolic tangent). Landau predicts a much later time (57% of the total cumulative temperature in the cycle, see Table D.1), while Fisher predicts much earlier temperatures (37%, corresponding to a saturation growth point around $G_{DD*} = 166 \pm 30$ °Cd).

The difference in response is due to the variation in the Malthusian coefficient of exponential growth. Every living organism initially exhibits exponential growth: $y' = ry$, where r is the Malthusian coefficient and $'$ is time differentiation. But this behaviour is modified by the introduction of a braking mechanism, $y' = \hat{r}y$, where $\hat{r} = r(1 - y)$ for Fisher or logistic, and $\hat{r} = r(1 - y^2)$ for Landau, which is in fact a cubic equation. Using the definition of the dimensionless coefficient c_1 (Eq.(D.15)), its fitted values (Table D.1), and the mean value of G_{DD} and α ($\approx 15^oC$), a rough estimate of the Malthusian coefficient is between $0.08d^{-1} \leq r \leq 0.1d^{-1}$ for the Fisher and Landau models. These values agree with those reported by *Schiphorst et al.* (2023) on the effects of daily irradiance on *Arabidopsis thaliana*.

MLR uses different weights based on predictor variance to estimate the target variable (DMY). If a predictor has significant variance, it has less influence on the final fit. Goodness of fit is therefore measured in units of standard deviation. A value of $\varphi_2 = 0.62$ means that a one-standard-deviation increase, $\varphi_{2o} + 1$, in crop height leads to a 0.62-unit increase in crop yield, followed by 0.33 for G_{DD} (φ_3) and finally 0.01 for

precipitation (φ_1). An additional yield of 0.62 per additional unit of plant height is just a numerical result. The subscript o indicates a reference value of the predictor variable. In numerical jargon, the conditioning number means that if the plant height value is uncertain by one unit, DMY is only uncertain by 0.62. In other words, the model reduces the uncertainty in the data. Alfalfa cannot grow indefinitely; it either falls under its own weight or flowers and is no longer digestible.

Most of the alfalfa cuts were made during an exceptional period of drought caused by a rare "triple dip" phenomenon of La Niña (*Li et al.*, 2023). Rainfall per cycle can affect the results as it has a different effect at the beginning and end of the cycle (when the phreatic level is well below the root zone). Dividing the rainfall into periods would refine the analysis, but more observations are needed for more complex methods (*López et al.*, 2022)).

5 Conclusions

This study presents an innovative method for estimating alfalfa crop growth and yield using a short-range scanning device. It was established that open source hardware devices can be used to integrate data to improve the prediction of plant growth and yield for alfalfa. The acoustic device accurately measured alfalfa plant height for almost three years. The work included two novel aspects: a unified frame of reference and a new sigmoidal function. The Landau equation predicts a higher accumulated temperature for alfalfa to reach maturity saturation than the logistic model.

Similarly, a set of data collected by the automatic station was used to infer the final yield with an acceptable level of fit ($r^2 = 0.80$). It was established that plant height is a key predictor for DMY values. The low-cost acoustic sensor and open-source hardware could help growers monitor alfalfa growth and make informed harvest and yield decisions. This approach can also be applied to other crops.

As acoustic sensor data can infer canopy density, yield predictions can be improved by using an independent device to contrast plant morphology. Combining low-cost 3D imagery with satellite imagery will allow the use of machine learning algorithms with more predictors for non-linear regressions.

Acknowledgements

We would like to express our sincere gratitude to Matías Blumenthal for his invaluable help with the initial cutting of alfalfa, an essential step to start this research. We also thank Diana Alberto for her dedicated work in handling and drying the samples at INALI (CONICET), which was essential for the success of this study.

Declaration of Competing Interest

We declare that we have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

CRediT authorship contribution statement

EP López: Funding acquisition, Investigation, Methodology, Software, Formal analysis, Data curation, Validation, Visualization, Writing – original draft. CA Vionnet: Project administration, Conceptualization, Funding acquisition, Formal analysis, Software, Validation, Writing – original draft. G Dunger: Funding acquisition, Conceptualization, Supervision, Writing – review and editing. J Maiztegui: Conceptualization, Supervision, Validation. G Contini: Methodology, Software. G Contini: Investigation,Methodology, Software. J Prodolliet: Investigation, Methodology, Software, Visualization.

Funding

This work was supported by the research grant CAID 50520190100249LI of the Universidad Nacional el Litoral and the upscaling grant CAAT 03012020. Special recognition is also given to the National Council for Scientific and Technical Research (CONICET) for its financial support.

Appendix D1: Growth Models

The biochemical, physical, and hydrological interactions behind plant growth can be summarized in a single phenomenological equation: $dR/dt = f(t, \mathbf{c})$, where t is time, R is plant height, and $\mathbf{c}$ is the vector of parameters that lumps all these unknown interrelationships. Three models are used here. The first two predict symmetric sigmoidal curves, while the third predicts asymmetric growth curves. The time dependence is then converted into G_{DD} variation by the chain rule $\alpha dR/d\mathrm{G_{DD}} = f(\mathrm{G_{DD}}, \mathbf{c})$, where $\alpha \approx 15°\mathrm{C}$ (cf. Fig.D.5**B**)

$$\alpha \frac{dR}{d\mathrm{G_{DD}}} = r\frac{R}{R_M}(R_M - R) \qquad \text{: Fisher} \tag{D.8}$$

$$\alpha \frac{dR^2}{d\mathrm{G_{DD}}} = 2Q^2R^2 - 2P^2R^4 \qquad \text{: Landau} \tag{D.9}$$

$$\alpha \frac{dR}{d\mathrm{G_{DD}}} = r \ln\left(\frac{R_M}{R}\right)R \qquad \text{: Gompertz} \tag{D.10}$$

The first equation is the logistic equation, also known as the Fisher equation (Kudryashov and Zakharchenko, 2014), where $r : [T^{-1}]$ is the Malthusian parameter and $R_M : [L]$ is the maximum plant height. The second model is due to *Landau* (1944), who proposed it to explain the evolution of turbulence in Newtonian flows. The coefficients $Q : [T^{-1/2}]$, $P : [L^{-1}T^{-1/2}]$ are empirical constants required for dimensional reasons, where L and T have the dimensions of length and time respectively. Substituting $z = y^2$ reduces Landau's model to Fisher's equation. The third independent model is the generalized logistic function or Gompertz model (*Chu*, 2020). Introduction of normalized variables, necessary for theoretical and numerical reasons

$$\tau = \frac{\mathrm{G_{DD}}}{G_{\mathrm{DD_M}}} \quad , \quad y(\tau) = \frac{R(\tau)}{R_M} \quad , \tag{D.11}$$

where G_{DD_M} is the total, or maximum, cumulative G_{DD} for each growing cycle. Then, the above models become

$$\frac{dy}{d\tau} = c_1\, y(1-y) \qquad : \text{Fisher}, \tag{D.12}$$

$$\frac{1}{2}\frac{dy^2}{d\tau} = c_1\, y^2(1-y^2) \qquad : \text{Landau}, \tag{D.13}$$

$$\frac{dy}{d\tau} = c_1\, ln\left(\tfrac{1}{y}\right) y \qquad : \text{Gompertz}. \tag{D.14}$$

The exact solutions are included in Eq.(D.5), with c_1 and c_2 given by

$$\begin{aligned} c_1 &= \tfrac{r\, G_{DD_M}}{\alpha} \quad , \; c_2 = ln\,(1/y_0 - 1) && : \text{Fisher} \\ c_1 &= \tfrac{Q^2\, G_{DD_M}}{\alpha} \quad , \; c_2 = ln\left(\sqrt{1y_0^2 - 1}\right) && : \text{Landau} \\ c_1 &= \tfrac{r\, G_{DD_M}}{\alpha} \quad , \; c_2 = ln\,((ln(1/y_0)) && : \text{Gompertz} \end{aligned} \tag{D.15}$$

where $y(0) = y_0$ is some initial condition. All models show initial exponential Malthusian growth followed by linear growth. Note that fitting the Gompertz model can be done by direct linearisation (similar to the logistic equation, see *Chu* (2020))

$$\begin{aligned} y(\tau) &= \frac{1 - tanh[exp(c_2 - c_1\tau)/2]}{1 + tanh[exp(c_2 - c_1\tau)/2]} \\ &= exp\left[-exp(c_2 - c_1\tau)\right] \\ ln\left[ln\left(\frac{1}{y(\tau)}\right)\right] &= -c_1\tau + c_2 \end{aligned}$$

The models were fitted to the data in three steps. First, a nonlinear least squares method[5] was applied to Eq. (D.2) with $\sigma_i = 1$ using the logistic equation to check the suitability of sigmoidal growth models. Normalized variables provide a common frame of reference, avoiding the problem of dealing with cycles of different lengths or cumulative G_{DD} values. The interval $[0, 1]$ was divided into 40 equidistant segments. From each cycle, the G_{DD} value closest to the equidistant G_{DD_i} value (converted from τ_i) was selected along with its corresponding plant height value. This resulted in 12 data points for each τ_i, with a deviation σ_i for $i = 1, \ldots, 41$. Third, χ^2 (Eq.(D.2)) was minimized with $\sigma_i \neq 1$. Eqs.(D.15) gives the initial seeds for each method. For the linearized form, the Newton method performs one iteration. All computations were performed using Python (see footnote 4 for details) and the GNU fortran 2008 (Ubuntu 11.4.0-1ubuntu1 22.04) to compile some in-house developed routines. The results were identical.

Appendix D2: Crop yield models

All selected data used to fit the MLR model corresponded to the final or cumulative value of the trait variable at the end of the growing cycle. Plant height was determined

[5] Newton's scheme as a fixed point iteration.

by averaging the sensor readings, except in situations where rain flattened the crop. Cumulative G_{DD} corresponds to the date of destructive sampling, as biomass continued to increase until then. Data were standardised by subtracting from the mean and dividing by the standard deviation. Standardisation makes direct comparisons easier by overcoming the problem of variables measured in different units.

Appendix D3: Crop Monitoring Station Design

This appendix presents the design of the plant growth monitoring station, including technical diagrams and materials used. The station is supported by structural square tubing ($12 \times 12 \times 1.60$ mm, Fig. D.11). The ultrasonic sensors have a 3/4" male thread, screwed into a 3/4-inch polypropylene T-fitting. Wiring was routed through the interior with matching end caps. This info supports reproducibility and further crop growth monitoring research.

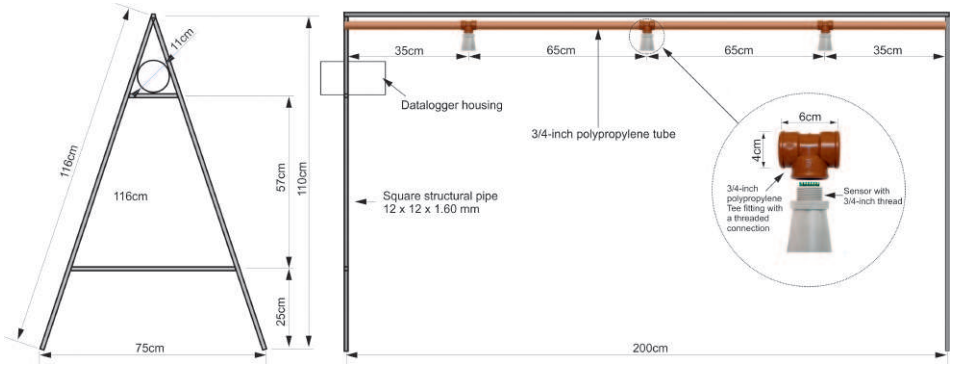


Figure D.11: Design of the crop growth monitoring station

Acrónimos

AI	Artificial Intelligence.
CDI	Common Datalogger Interface.
DM	Dry Matter.
DMY	Dry Matter Yield.
ENSO	El Niño Southern Oscillation.
FAO	Food and Agriculture Organization of the United Nations.
FEWS	Flood Early Warning System.
FOSH	Free Open Source Hardware.
FM	Fresh Matter.
G_{DD}	Growing Degree Days.
INTA	Instituto Nacional de Tecnología Agropecuaria.
IoT	Internet of Things.
KNN	K-Nearest Neighbor.
LiDAR	Light Detection and Ranging.
MBE	Mean Bias Error.
MLR	Multiple Linear Regression.
OCR	Optical Character Recognition.
PCB	Printed Circuit Board.
RA	República Argentina.
RFO	Random Forest.
RMSD	Root Mean Square Deviation.
RMSE	Root Mean Square Error.
RNUO	Reserva Natural Urbana del Oeste.
RTC	Real Time Clock.
RTU	Remote Transmission Unit
SAH	Sistema de Alerta Hidrológica.
SVR	Support Vector Regression.
TDR	Time Domain Reflectometry.
TOF	Time of Flight.
VWC	Volumetric Water Content.
WUNR	West Urban Natural Reserve.

Bibliografía

Acharya, J. P., Y. Lopez, B. T. Gouveia, I. de Bem Oliveira, M. F. R. Resende, P. R. Muñoz, & E. F. Rios (2020), Breeding Alfalfa (Medicago sativa L.) Adapted to Subtropical Agroecosystems, *Agronomy*, *10*(5), doi:10.3390/agronomy10050742.

Adams, K. H., et al. (2022), Remote sensing of groundwater: Current capabilities and future directions, *Water Resources Research*, *58*(10), e2022WR032,219, doi:https://doi.org/10.1029/2022WR032219, e2022WR032219 2022WR032219.

Al-Turjman, F., H. Zahmatkesh, & R. Shahroze (2022), An overview of security and privacy in smart cities' iot communications, *Transactions on Emerging Telecommunications Technologies*, *33*(3), e3677, doi:10.1002/ett.3677, e3677 ETT-19-0133.R1.

Antico, A., R. O. Aguiar, & M. L. Amsler (2018), Hydrometric data rescue in the paraná river basin, *Water Resources Research*, *54*(2), 1368–1381, doi:https://doi.org/10.1002/2017WR020897.

Arellana, J., M. Franco, & F. Grings (2023), Using saocom data and bayesian inference to estimate soil dielectric constant in agricultural soils, *IEEE Geoscience and Remote Sensing Letters*, *20*, 1–5, doi:10.1109/LGRS.2023.3296094.

Asante, K. O., R. D. Macuacua, G. A. Artan, R. W. Lietzow, & J. P. Verdin (2007), Developing a flood monitoring system from remotely sensed data for the Limpopo basin, *IEEE Transactions on Geoscience and Remote Sensing*, *45*(6), 1709–1714, publisher: IEEE.

Babaeian, E., M. Sadeghi, S. B. Jones, C. Montzka, H. Vereecken, & M. Tuller (2019), Ground, proximal, and satellite remote sensing of soil moisture, *Reviews of Geophysics*, *57*(2), 530–616, doi:10.1029/2018RG000618.

Balis, B., T. Bartynski, M. Bubak, G. Dyk, T. Gubala, & M. Kasztelnik (2013), A development and execution environment for early warning systems for natural disasters, in *2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing*, pp. 575–582, IEEE.

Banerjee, B. P., G. Spangenberg, & S. Kant (2022), Cbm: An IoT enabled LiDAR sensor for in-field crop height and biomass measurements, *Biosensors*, *12*(16), doi:/10.3390/bios12010016.

Barba, L. A. (2021), The python jupyter ecosystem: Today's problem-solving environment for computational science, *Computing in Science and Engineering*, *23*(3), 5–9, doi:10.1109/MCSE.2021.3074693.

Barcelo-Ordinas, J. M., J.-P. Chanet, K.-M. Hou, & J. García-Vidal (2013), A survey of wireless sensor technologies applied to precision agriculture, pp. 801–808.

Barcelo-Ordinas, J. M., M. Doudou, J. Garcia-Vidal, & N. Badache (2019), Self-calibration methods for uncontrolled environments in sensor networks: A reference survey, *Ad Hoc Networks*, *88*, 142–159, doi:10.1016/j.adhoc.2019.01.008.

Barker, J., N. Zhang, J. Sharon, R. Steeves, X. Wang, Y. Wei, & J. Poland (2016), Development of a field-based high-throughput mobile phenotyping platform, *Computers and Electronics in Agriculture*, *122*, 74–85, doi:10.1016/j.compag.2016.01.017.

Barmeier, G., B. Mistele, & U. Schmidhalter (2016), Referencing laser and ultrasonic height measurements of barley cultivars by using a herbometre as standard, *Crop and Pasture Science*, *67*(12), 1215 – 1222, doi:10.1071/CP16238.

Barriball, S., A. Han, & B. Schlautman (2022), Effect of growing degree days, day of the year, and cropping systems on reproductive development of Kernza in Kansas, *Agrosystems, Geosciences & Environment*, *5*(3), e20,286, doi:10.1002/agg2.20286.

Barros, V., E. Castañeda, & M. Doyle (2000), Precipitation trends in southern south america, east of the andes: An indication of climate variability. southern hemisphere paleo-and neoclimates: Key sites, methods, data and models.

Basigalup, D. H. E. (2023), *Research, production and industrialization of alfalfa in Argentina (in Spanish)*, INTA, National Institute of Agricultural and Livestock Technologies.

Beddows, P. A., & E. K. Mallon (2018), Cave pearl data logger: A flexible arduino-based logging platform for long-term monitoring in harsh environments, *Sensors*, *18*(2), doi:10.3390/s18020530.

Berhongaray, G., M. Basanta, & J. M. Jauregui (2019), Water table depth affects persistence and productivity of alfalfa in central argentina, *Field Crops Research*, *235*, 54–58, doi:https://doi.org/10.1016/j.fcr.2019.02.018.

Bitella, G., et al. (2024), Monitoring plant height and spatial distribution of biometrics with a low-cost proximal platform, *Plants*, *13*(8), doi:10.3390/plants13081085.

Bogena, H., J. Huisman, C. Oberdörster, & H. Vereecken (2007), Evaluation of a low-cost soil water content sensor for wireless network applications, *Journal of Hydrology*, *344*(1), 32–42, doi:10.1016/j.jhydrol.2007.06.032.

Bogena, H. R., J. A. Huisman, B. Schilling, A. Weuthen, & H. Vereecken (2017), Effective calibration of low-cost soil water content sensors, *Sensors (Switzerland)*, *17*(1), doi:10.3390/s17010208.

Boutet, N. (2020), Diseño y desarrollo de un sistema de monitorización web para dispositivos de medición automáticos, *Proyecto final de carrera*, Facultad de Ing. y Cs. Hídricas, UNL, director: López, E.

Calvello, M., R. N. d'Orsi, L. Piciullo, N. Paes, M. Magalhaes, & W. A. Lacerda (2015), The Rio de Janeiro early warning system for rainfall-induced landslides: analysis of performance for the years 2010–2013, *International journal of disaster risk reduction*, *12*, 3–15, publisher: Elsevier.

Carranza, C., C. Nolet, M. Pezij, & M. van der Ploeg (2021), Root zone soil moisture estimation with random forest, *Journal of Hydrology*, *593*, 125,840, doi:https://doi.org/10.1016/j.jhydrol.2020.125840.

Chagas, A. M. (2018), Haves and have nots must find a better way: The case for open scientific hardware., *PLOS Biology*, *16*, 1–8, doi:10.1371/JOURNAL.PBIO.3000014.

Chen, G., S. Zhou, & H. Huang (2018), An integrated moisture and temperature sensor with model based temperature-dependent nonlinearity compensation, *IEICE Electronics Express*, *15*(8), 20180,200–20180,200, doi:10.1587/elex.15.20180200.

Chen, L., L. Zhangzhong, W. Zheng, J. Yu, Z. Wang, L. Wang, & C. Huang (2019), Data-driven calibration of soil moisture sensor considering impacts of temperature: A case study on FDR sensors, *Sensors (Switzerland)*, *19*(20), doi:10.3390/s19204381.

Chen, N., K. Wang, C. Xiao, & J. Gong (2014), A heterogeneous sensor web node meta-model for the management of a flood monitoring system, *Environmental Modelling & Software*, *54*, 222–237, publisher: Elsevier.

Christidis, N., Y. Aono, & P. A. Stott (2022), Human influence increases the likelihood of extremely early cherry tree flowering in kyoto, *Environmental Research Letters*, *17*(5), 054,051, doi:10.1088/1748-9326/ac6bb4.

Chu, K. H. (2020), Fitting the gompertz equation to asymmetric breakthrough curves, *Journal of Environmental Chemical Engineering*, *8*(3), 103,713, doi:doi.org/10.1016/j.jece.2020.103713.

Cleveland, W. S. (1979), Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association*, *74*(368), 829–836, doi:10.1080/01621459.1979.10481038.

Contini G. F., Carrillo E., Ferreira G., Macor J., & Veizaga E. (2014), Types of faults detected in a telemetric network. Case study: Salado River Basin, Province of Santa Fe, Argentina (in Spanish), UNL Press.

Coopersmith, E. J., M. H. Cosh, J. E. Bell, & R. Boyles (2016), Using machine learning to produce near surface soil moisture estimates from deeper in situ records at U.S. Climate Reference Network (USCRN) locations: Analysis and applications to AMSR-E satellite validation, *Advances in Water Resources*, *98*, 122–131, doi:10.1016/j.advwatres.2016.10.007.

De Abelleyra, D., S. Banchero, S. Verón, J. Mosciaro, & J. Volante (2019), National map of crops campaign 2018/2019 (in spanish), *Tech. rep.*, National Agricultural Technology Institute (INTA).

de Camargo, E. T., et al. (2023), Low-cost water quality sensors for iot: A systematic review, *Sensors*, *23*(9), doi:10.3390/s23094424.

Debnath, S., M. Paul, & T. Debnath (2023), Applications of lidar in agriculture and future research directions, *Journal of Imaging*, *9*(3), doi:10.3390/jimaging9030057.

Dhakal, M., C. P. West, C. Villalobos, P. Brown, & P. E. Green (2020), Interseeding alfalfa into native grassland for enhanced yield and water use efficiency, *Agronomy Journal*, *112*(3), 1931–1942, doi:/10.1002/agj2.20147.

Domínguez-Niño, J. M., H. R. Bogena, J. A. Huisman, B. Schilling, & J. Casadesús (2019), On the accuracy of factory-calibrated low-cost soil water content sensors, *Sensors (Switzerland)*, *19*(14), doi:10.3390/s19143101.

Doorn, N. (2021), Artificial intelligence in the water domain: Opportunities for responsible use, *Science of The Total Environment*, *755*, 142,561, doi:10.1016/j.scitotenv.2020.142561.

Dorigo, W. A., et al. (2011), The international soil moisture network: a data hosting facility for global in situ soil moisture measurements, *Hydrology and Earth System Sciences*, *15*(5), 1675–1698, doi:10.5194/hess-15-1675-2011.

Elnesr, M., & A. Alazba (2016), An integral model to calculate the growing degree-days and heat units, a spreadsheet application, *Computers and Electronics in Agriculture*, *124*, 37–45, doi:10.1016/j.compag.2016.03.024.

Entekhabi, D., et al. (2010), The soil moisture active passive (smap) mission, *Proceedings of the IEEE*, *98*(5), 704–716, doi:10.1109/JPROC.2010.2043918.

EPA (2020), Groundwater level and well depth measurement, *Lsasdproc-105-r4*, U.S. Environmental Protection Agency.

Erdle, K., B. Mistele, & U. Schmidhalter (2011), Comparison of active and passive spectral sensors in discriminating biomass parameters and nitrogen status in wheat cultivars, *Field Crops Research*, *124*, 74–84, doi:10.1016/j.fcr.2011.06.007.

Fakhruddin, S., A. Kawasaki, & M. S. Babel (2015), Community responses to flood early warning system: Case study in kaijuri union, bangladesh, *International Journal of Disaster Risk Reduction*, *14*, 323–331, publisher: Elsevier.

Faranda, D., F. M. E. Pons, E. Giachino, S. Vaienti, & B. Dubrulle (2015), Early warnings indicators of financial crises via auto regressive moving average models, *Communications in Nonlinear Science and Numerical Simulation*, *29*(1-3), 233–239, publisher: Elsevier.

Federico Vilaseca, C. C., Alberto Castro, & A. Gorgoglione (2023), Assessing influential rainfall–runoff variables to simulate daily streamflow using random forest, *Hydrological Sciences Journal*, *68*(12), 1738–1753, doi:10.1080/02626667.2023.2232356.

Feng, Y., et al. (2022), Yield and quality properties of alfalfa (medicago sativa l.) and their influencing factors in China, *European Journal of Agronomy*, *141*, 126,637, doi:/10.1016/j.eja.2022.126637.

Ferrer-Cid, P., J. M. Barcelo-Ordinas, J. Garcia-Vidal, A. Ripoll, & M. Viana (2019), A comparative study of calibration methods for low-cost ozone sensors in iot platforms, *IEEE Internet of Things Journal*, *6*(6), 9563–9571, doi:10.1109/JIOT.2019.2929594.

Ferrer-Cid, P., J. Garcia-Calvete, A. Main-Nadal, Z. Ye, J. M. Barcelo-Ordinas, & J. Garcia-Vidal (2022), Sampling trade-offs in duty-cycled systems for air quality low-cost sensors, *Sensors*, *22*(10), 3964, doi:10.3390/s22103964.

Ferrer-Cid, P., E. López, C. Vionnet, J. Barcelo-Ordinas, & J. Garcia-Vidal (2024a), *Artificial intelligence in hydroinformatics. Chapter 1. Artificial Intelligence of Things for Smart Water Systems*, IAHR Water Monograph Series, IAHR, International Association for Hydro-Environment Engineering and Research, en prensa.

Ferrer-Cid, P., J. A. Paredes-Ahumada, X. Allka, M. Guerrero-Zapata, J. M. Barcelo-Ordinas, & J. Garcia-Vidal (2024b), A data-driven framework for air quality sensor networks, *IEEE Internet of Things Magazine*, *7*(1), 128–134, doi:10.1109/IOTM.001.2300112.

Fink, K. P., P. Grassini, A. Rocateli, L. M. Bastos, J. Kastens, L. P. Ryan, X. Lin, A. Patrignani, & R. P. Lollato (2022), Alfalfa water productivity and yield gaps in the U.S. central Great Plains, *Field Crops Research*, *289*, 108,728, doi:/10.1016/j.fcr.2022.108728.

Fisher, D. K., & P. J. Gould (2012a), Open-source hardware is a low-cost alternative for scientific instrumentation and research, *Modern Instrumentation*, *1*, 8–20, doi:http://dx.doi.org/10.4236/mi.2012.12002.

Fisher, D. K., & P. J. Gould (2012b), Open-Source Hardware Is a Low-Cost Alternative for Scientific Instrumentation and Research, *Modern Instrumentation*, *1*, 8–20, doi:http://dx.doi.org/10.4236/mi.2012.12002.

Fricke, T., F. Richter, & M. Wachendorf (2011), Assessment of forage mass from grassland swards by height measurement using an ultrasonic sensor, *Computers and Electronics in Agriculture*, *79*(2), 142–152, doi:10.1016/j.compag.2011.09.005.

Furbank, R. T., & M. Tester (2011), Phenomics – technologies to relieve the phenotyping bottleneck, *Trends in Plant Science*, *16*(12), 635–644, doi:10.1016/j.tplants.2011.09.005.

Garrote, L., & R. L. Bras (1995), A distributed model for real-time flood forecasting using digital elevation models, *Journal of Hydrology*, *167*(1-4), 279–306, publisher: Elsevier.

Gasparri, N. I., & H. R. Grau (2009), Deforestation and fragmentation of chaco dry forest in NW Argentina (1972-2007), *Forest Ecology and Management*, *258*(6), 913–921, doi:10.1016/j.foreco.2009.02.024.

Gomez, C., J. Paradells, C. Bormann, & J. Crowcroft (2017), From 6lowpan to 6lo: Expanding the universe of ipv6-supported technologies for the internet of things, *Comm. Mag.*, *55*(12), 148–155, doi:10.1109/MCOM.2017.1600534.

González-Teruel, J. D., R. Torres-Sánchez, P. J. Blaya-Ros, A. B. Toledo-Moreo, M. Jiménez-Buendía, & F. Soto-Valles (2019), Design and calibration of a low-cost SDI-12 soil moisture sensor, *Sensors (Switzerland)*, *19*(3), doi:10.3390/s19030491.

González-Vidal, A., J. Cuenca-Jara, & A. F. Skarmeta (2019), Iot for water management: Towards intelligent anomaly detection, in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pp. 858–863, doi:10.1109/WF-IoT.2019.8767190.

Gore, N. R., P. Sagar, & M. Gawande (2020), Water quality monitoring, doi:10.1201/9781003062110.

Gregg, C. E., B. Houghton, & J. W. Ewert (2015), Volcano warning systems, in *The Encyclopedia of Volcanoes*, pp. 1173–1185, Elsevier.

Guo, Z., P. Chen, H. Zhang, M. Jiang, & C. Li (2012), IMA: An integrated monitoring architecture with sensor networks, *IEEE Transactions on Instrumentation and Measurement*, *61*(5), 1287–1295, publisher: IEEE.

Han, J., Y. Yang, M. L. Roderick, T. R. McVicar, D. Yang, S. Zhang, & H. E. Beck (2020), Assessing the steady-state assumption in water balance calculation across global catchments, *Water Resources Research*, *56*(7), e2020WR027,392, doi:10.1029/2020WR027392.

Hastie, T., R. Tibshirani, & J. Friedman (2009), *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York Inc., New York, NY, USA.

Henonin, J., B. Russo, O. Mark, & P. Gourbesville (2013), Real-time urban flood forecasting and modelling–a state of the art, *Journal of Hydroinformatics*, *15*(3), 717–736, publisher: IWA Publishing.

Holden, P., & N. Fierer (2005), *VADOSE ZONE, Microbial Ecology*, Elsevier.

Holt, A., C.-Y. Huang, A. Holt, & C.-Y. Huang (2014), OpenWRT, *Embedded Operating Systems: A Practical Approach*, pp. 161–181, publisher: Springer.

Hopmans, J. (2011), Infiltration and unsaturated zone. in: Peter wilderer (ed.) treatise on water science, Oxford Academic Press, Oxford.

Houspanossian, J., R. Giménez, J. I. Whitworth-Hulse, M. D. Nosetto, W. Tych, P. M. Atkinson, M. C. Rufino, & E. G. Jobbágy (2023), Agricultural expansion raises groundwater and increases flooding in the south american plains, *Science*, *380*(6652), 1344–1348, doi:10.1126/science.add5462.

Huang, J., et al. (2019), Assimilation of remote sensing into crop growth models: Current status and perspectives, *Agricultural and Forest Meteorology*, *276-277*, 107,609, doi:https://doi.org/10.1016/j.agrformet.2019.06.008.

Jenny, H., E. G. Alonso, Y. Wang, & R. Minguez (2020), Using artificial intelligence for smart water management systems.

Jáuregui, J. M., J. J. Ojeda, G. D. Berone, F. A. Lattanzi, J. Baudracco, S. R. Fariña, & D. J. Moot (2022), Yield gaps of lucerne (medicago sativa l.) in livestock systems of Argentina, *Annals of Applied Biology*, *181*(1), 22–32, doi:10.1111/aab.12745.

Kalyani, Y., & R. Collier (2021), A systematic survey on the role of cloud, fog, and edge computing combination in smart agriculture, *Sensors*, *21*(17), doi:10.3390/s21175922.

Kapilaratne, R. J., & M. Lu (2017), Automated general temperature correction method for dielectric soil moisture sensors, *Journal of Hydrology*, *551*, 203–216, doi:10.1016/j.jhydrol.2017.05.050, investigation of Coastal Aquifers.

Kavya, M., A. Mathew, P. R. Shekar, & S. P (2023), Short term water demand forecast modelling using artificial intelligence for smart water management, *Sustainable Cities and Society*, *95*, 104,610, doi:10.1016/j.scs.2023.104610.

Kerr, Y., P. Waldteufel, J.-P. Wigneron, J. Martinuzzi, J. Font, & M. Berger (2001), Soil moisture retrieval from space: the soil moisture and ocean salinity (smos) mission, *IEEE Transactions on Geoscience and Remote Sensing*, *39*(8), 1729–1735, doi:10.1109/36.942551.

Khampuengson, T., & W. Wang (2023), Novel methods for imputing missing values in water level monitoring data, *Water Resources Management*, *37*(2), 851–878.

Khan, L. U., W. Saad, Z. Han, E. Hossain, & C. S. Hong (2021), Federated learning for internet of things: Recent advances, taxonomy, and open challenges, *IEEE Communications Surveys & Tutorials*, *23*(3), 1759–1799, doi:10.1109/COMST.2021.3090430.

Kopittke, P. M., N. W. Menzies, P. Wang, B. A. McKenna, & E. Lombi (2019), Soil and the intensification of agriculture for global food security, *Environment International*, *132*, 105,078, doi:https://doi.org/10.1016/j.envint.2019.105078.

Kuantama, E., P. Mardjoko, & M. A. Saraswati (2013), Design and Construction of Early flood warning system through SMS based on SIM300C GSM modem, in *2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME)*, pp. 115–119, IEEE.

Kundzewicz, Z. W. (2008), Climate change impacts on the hydrological cycle, *Ecohydrology & Hydrobiology*, *8*(2), 195–203, doi:10.2478/v10104-009-0015-y.

Kuppel, S., J. Houspanossian, M. D. Nosetto, & E. G. Jobbágy (2015), What does it take to flood the pampas?: Lessons from a decade of strong hydrological fluctuations, *Water Resources Research*, *51*(4), 2937–2950, doi:https://doi.org/10.1002/2015WR016966.

Kümmerer, R., P. O. Noack, & B. Bauer (2023), Using high-resolution uav imaging to measure canopy height of diverse cover crops and predict biomass, *Remote Sensing*, *15*(6), doi:10.3390/rs15061520.

Landau, L. (1944), On the problem of turbulence, *C. R. Acad. Sci. URSS*, *44*(8), 339–349.

Li, X., Z.-Z. Hu, M. J. McPhaden, C. Zhu, & Y. Liu (2023), Triple-dip la niñas in 1998–2001 and 2020–2023: Impact of mean state changes, *Journal of Geophysical Research: Atmospheres*, *128*(17), e2023JD038,843, doi:https://doi.org/10.1029/2023JD038843, e2023JD038843 2023JD038843.

Llorens, J., E. Gil, J. Llop, & A. Escolà (2011), Ultrasonic and lidar sensors for electronic canopy characterization in vineyards: Advances to improve pesticide application methods, *Sensors*, *11*(2), 2177–2194, doi:10.3390/s110202177.

López, E. P., C. A. Vionnet, D. Villarreal, G. Contini, E. Veizaga, & C. G. Ferreira (2020), A cost-effective redundant communication system for improving the reliability of a flood early warning system, *Journal of Hydroinformatics*, *22*(4), 856–875, doi:10.2166/hydro.2020.216.

Lozza, H. F. (2019), Sistema para la aplicación de los datos de la misión satelital saocom en la agricultura, *XI Congreso de AgroInformática (CAI)-JAIIO 48*, (September), 70–83.

López, E. (2012), Captura y transmisión automática de datos hidro-meteorológicos (Capture and Automatic Transmission of Hydro-Meteorological Data), Ph.D. thesis, Universidad Nacional del Litoral, Santa Fe, Argentina.

López, E., C. Vionnet, P. Ferrer-Cid, J. M. Barcelo-Ordinas, J. Garcia-Vidal, G. Contini, J. Prodolliet, & J. Maiztegui (2022), A low-power iot device for measuring water table levels and soil moisture to ease increased crop yields, *Sensors*, *22*(18), doi:10.3390/s22186840.

López, E., et al. (2023), Monitoreo hidro-ambiental de la reserva natural urbana del oeste de la ciudad de santa fe, in *VII Simposio sobre Métodos Experimentales en Hidráulica, Hidrología e Hidrometeorología*, Córdoba, Argentina.

López, E., C. A. Vionnet, G. Dunger, J. Maiztegui, J. Prodolliet, & G. Contini (2024), Estimating alfalfa yield by measuring plant height with a low-cost, close-range scanning device, *Field Crops Research*, submitted for publication.

López, E. P., C. A. Vionnet, D. Villarreal, G. Contini, E. Veizaga, & C. G. Ferreira (2020), A cost-effective redundant communication system for improving the reliability of a flood early warning system, *Journal of Hydroinformatics*, *22*(4), 856–875, doi:10.2166/hydro.2020.216.

López-Ramírez, G. A., & A. Aragón-Zavala (2023), Wireless sensor networks for water quality monitoring: A comprehensive review, *IEEE Access*, *11*, 95,120–95,142, doi:10.1109/ACCESS.2023.3308905.

Ma, Y., E. Leonarduzzi, A. Defnet, P. Melchior, L. E. Condon, & R. M. Maxwell (2024), Water table depth estimates over the contiguous united states using a random forest model, *Groundwater*, *62*(1), 34–43, doi:https://doi.org/10.1111/gwat.13362.

Macor, J., G. Contini, E. Veizaga, M. Morresi, E. Carrillo, & C. Ferreira (2016), Asesoramiento a la Inspección de Obra de las Redes de Monitoreo de Cuenca del Río

Salado en la Prov. de Santa Fe y su Ampliación a la Prov. de Sgo. del Estero, de los A Saladillo y Ludueña y en la Cuenca del Río Carcarañá (Advice to the Construction Inspection of the Monitoring Networks of the Salado River Basin in the Prov. of Santa Fe and its Extension to the Prov. of Sgo. del Estero, of the Saladillo and Ludueña Streams and in the Carcarañá River Basin) . Report for the Ministry of Public Works and Transportation., *Tech. rep.*, FICH-UNL, Santa Fe, Argentina.

Makkonen, L. (2006), Plotting positions in extreme value analysis, *Journal of Applied Meteorology and Climatology*, *45*(2), 334 – 340, doi:10.1175/JAM2349.1.

Manley, G. (1953), The mean temperature of central england, 1698–1952, *Quarterly Journal of the Royal Meteorological Society*, *79*(340), 242–261, doi:https://doi.org/10.1002/qj.49707934006.

Mann, M. E., & P. D. Jones (2003), Global surface temperatures over the past two millennia, *Geophysical Research Letters*, *30*(15), doi:https://doi.org/10.1029/2003GL017814.

Manocha, A., S. K. Sood, & M. Bhatia (2024), Iot-digital twin-inspired smart irrigation approach for optimal water utilization, *Sustainable Computing: Informatics and Systems*, *41*, 100,947, doi:10.1016/j.suscom.2023.100947.

Manso, R., E. López, G. Contini, , & C. Vionnet (2018), Monitoring system based on a wireless sensor network (WSN) implemented with open source software and hardware (in spanish), pp. 316–322, 47 JAIIO-Jornadas Argentinas de Informática CAI 2018 - Argentine Agro-Informatics Congress, issn 2525-0949.

Mattera, J., L. A. Romero, A. L. Cuatrín, P. S. Cornaglia, & A. A. Grimoldi (2013), Yield components, light interception and radiation use efficiency of lucerne (Medicago sativa L.) in response to row spacing, *European Journal of Agronomy*, *45*, 87–95, doi:/10.1016/j.eja.2012.10.008.

Mellis, D., M. Banzi, D. Cuartielles, & T. Igoe (2007), Arduino: An open electronic prototyping platform, in *Proc. Chi*, vol. 2007, pp. 1–11.

Mercau, J., & M. Otegui (2014), A modeling approach to explore water management strategies for late-sown maize and double-cropped wheat–maize in the rainfed pampas region of argentina, in *Practical Applications of Agricultural System Models to Optimize the Use of Limited Water*, p. 351–373, John Wiley Sons, Ltd, doi:10.2134/advagricsystmodel5.c13.

Mercau, J. L., M. D. Nosetto, F. Bert, R. Giménez, & E. G. Jobbágy (2016), Shallow groundwater dynamics in the pampas: Climate, landscape and crop choice effects, *Agricultural Water Management*, *163*, 159–168, doi:https://doi.org/10.1016/j.agwat.2015.09.013.

Merkuryeva, G., Y. Merkuryev, B. V. Sokolov, S. Potryasaev, V. A. Zelentsov, & A. Lektauers (2015), Advanced river flood monitoring, modelling and forecasting, *Journal of computational science*, *10*, 77–85, publisher: Elsevier.

Mesas-Carrascosa, F., D. Verdú Santano, J. Meroño, M. Sánchez de la Orden, & A. García-Ferrer (2015), Open source hardware to monitor environmental parameters in precision agriculture, *Biosystems Engineering*, *137*, 73–83, doi:https://doi.org/10.1016/j.biosystemseng.2015.07.005.

Milewski, A. M., M. B. Thomas, W. M. Seyoum, & T. C. Rasmussen (2019), Spatial downscaling of grace twsa data to identify spatiotemporal groundwater level trends in the upper floridan aquifer, georgia, usa, *Remote Sensing*, *11*(23), doi:10.3390/rs11232756.

Miller, M., A. Kisiel, D. Cembrowska-Lech, I. Durlik, & T. Miller (2023), Iot in water quality monitoring—are we really here?, *Sensors*, *23*(2), doi:10.3390/s23020960.

Minaburo, A., L. Toutain, & R. Andreasen (2021), Static Context Header Compression (SCHC) for the Constrained Application Protocol (CoAP), RFC 8824, doi:10.17487/RFC8824.

Montazeaud, G., C. Langrume, S. Moinard, C. Goby, A. Ducanchez, B. Tisseyre, & G. Brunel (2021), Development of a low cost open-source ultrasonic device for plant height measurements, *Smart Agricultural Technology*, *1*, 100,022, doi:/10.1016/j.atech.2021.100022.

Moreno, H., & D. Andújar (2023), Proximal sensing for geometric characterization of vines: A review of the latest advances, *Computers and Electronics in Agriculture*, *210*, 107,901, doi:10.1016/j.compag.2023.107901.

Nagahage, E. A. A. D., I. S. P. Nagahage, & T. Fujino (2019), Calibration and validation of a low-cost capacitive moisture sensor to integrate the automated soil moisture monitoring system, *Agriculture*, *9*(7), doi:10.3390/agriculture9070141.

Noland, R. L., M. S. Wells, J. A. Coulter, T. Tiede, J. M. Baker, K. L. Martinson, & C. C. Sheaffer (2018), Estimating alfalfa yield and nutritive value using remote sensing and air temperature, *Field Crops Research*, *222*, 189–196, doi:/10.1016/j.fcr.2018.01.017.

Nosetto, M., E. Jobbágy, R. Jackson, & G. Sznaider (2009), Reciprocal influence of crops and shallow ground water in sandy landscapes of the inland pampas, *Field Crops Research*, *113*(2), 138–148.

Notarnicola, C., & R. Solorza (2014), Integration of remotely sensed images and electromagnetic models into a bayesian approach for soil moisture content retrieval: Methodology and effect of prior information, in *Dynamic Programming and Bayesian Inference*, edited by M. S. F. Nezhad, chap. 2, IntechOpen, Rijeka, doi:10.5772/57562.

Ojo, M. O., S. Giordano, G. Procissi, & I. N. Seitanidis (2018), A review of low-end, middle-end, and high-end iot devices, *IEEE Access*, *6*, 70,528–70,554, doi:10.1109/ACCESS.2018.2879615.

Otero, A., & M. Castro (2019), Variability of alfalfa (medicago sativa l.) seasonal forage production in the southwest of uruguay., *Agrociencia (Uruguay)*, *23*(1), 65–75, doi:10.31285/agro.23.1.9.

Palleja, T., & A. J. Landers (2017), Real time canopy density validation using ultrasonic envelope signals and point quadrat analysis, *Computers and Electronics in Agriculture*, *134*, 43–50, doi:10.1016/j.compag.2017.01.012.

Paola, A. D., V. Riccardo, & S. Monia (2016), An overview of available crop growth and yield models for studies and assessments in agriculture, *Journal of the Science of Food and Agriculture*, *96*(3), 709–714, doi:/doi.org/10.1002/jsfa.7359.

Pappenberger, F., H. L. Cloke, D. J. Parker, F. Wetterhall, D. S. Richardson, & J. Thielen (2015), The monetary benefit of early flood warnings in Europe, *Environmental Science & Policy*, *51*, 278–291, publisher: Elsevier.

Pearce, J. M. (2012), Building research equipment with free, open-source hardware, *Science*, *337*(6100), 1303–1304, doi:10.1126/science.1228183.

Pearce, J. M. (2020), Economic savings for scientific free and open source technology: A review, *Hardwarex*, doi:10.1016/j.ohx.2020.e00139.

Peiretti, R., & J. Dumanski (2014), The transformation of agriculture in Argentina through soil conservation, *International Soil and Water Conservation Research*, *2*(1), 14–20.

Pittman, J. J., D. B. Arnall, S. M. Interrante, C. A. Moffet, & T. J. Butler (2015), Estimation of biomass and canopy height in bermudagrass, alfalfa, and wheat using ultrasonic, laser, and spectral sensors, *Sensors*, *15*(2), 2920–2943, doi:/10.3390/s150202920.

Placidi, P., L. Gasperini, A. Grassi, M. Cecconi, & A. Scorzoni (2020), Characterization of low-cost capacitive soil moisture sensors for iot networks, *Sensors*, *20*(12).

Plate, E. J. (2007), Early warning and flood forecasting for large rivers with the lower Mekong as example, *Journal of Hydro-environment Research*, *1*(2), 80–94, publisher: Elsevier.

Press, W., S. Teukolsky, W. Vetterling, & B. Flannery (1989), *Numerical Recipes in fortran: The Art of Scientific Computing*, 1st edition ed., Cambridge University Press.

Pumo, D., A. Francipane, F. Lo Conti, E. Arnone, P. Bitonto, F. Viola, G. La Loggia, & L. Noto (2016), The SESAMO early warning system for rainfall-triggered landslides, *Journal of Hydroinformatics*, *18*(2), 256–276, publisher: IWA Publishing.

Qaim, M., & G. Traxler (2005), Roundup ready soybeans in Argentina: farm level and aggregate welfare effects, *Agricultural Economics*, *32*(1), 73–86, doi:10.1111/j.0169-5150.2005.00006.x.

Rabault, J., G. Sutherland, O. Gundersen, A. Jensen, A. Marchenko, & Øyvind Breivik (2020), An open source, versatile, affordable waves in ice instrument for scientific measurements in the polar regions, *Cold Regions Science and Technology*, *170*, 102,955, doi:https://doi.org/10.1016/j.coldregions.2019.102955.

Ripoll, A., M. Viana, M. Padrosa, X. Querol, A. Minutolo, K. Hou, J. Barcelo-Ordinas, & J. Garcia-Vidal (2019), Testing the performance of sensors for ozone pollution monitoring in a citizen science approach, *Science of The Total Environment*, *651*, 1166–1179, doi:https://doi.org/10.1016/j.scitotenv.2018.09.257.

Rodriguez, L., P. Cello, & C. Vionnet (2006), Modeling stream-aquifer interactions in a shallow aquifer, choele choel island, patagonia, argentina, *Hydrogeol Journal*, *14*, 591–602, doi:10.1007/s10040-005-0472-3.

Rodriguez, L. B., P. A. Cello, C. A. Vionnet, & D. Goodrich (2008), Fully conservative coupling of hec-ras with modflow to simulate stream–aquifer interactions in a drainage basin, *Journal of Hydrology*, *353*(1), 129–142, doi:10.1016/j.jhydrol.2008.02.002.

Russelle, M. P. (2001), Alfalfa: After an 8,000-year journey, the queen of forages stands poised to enjoy renewed popularity, *American Scientist*, *89*(3), 253–261.

Sadler, J. M., D. P. Ames, & R. Khattar (2016a), A recipe for standards-based data sharing using open source software and low-cost electronics, *Journal of Hydroinformatics*, *18*(2), 185–197, doi:10.2166/hydro.2015.092.

Sadler, J. M., D. P. Ames, & R. Khattar (2016b), A recipe for standards-based data sharing using open source software and low-cost electronics, *Journal of Hydroinformatics*, *18*(2), 185–197, doi:10.2166/hydro.2015.092.

Sanderson, M. A., T. P. Karnezos, & A. G. Matches (1994), Morphological development of alfalfa as a function of growing degree days, *Journal of Production Agriculture*, *7*(2), 239–242, doi:10.2134/jpa1994.0239.

Schiphorst, C., C. Koeman, L. Caracciolo, K. Staring, T. Theeuwen, S. Driever, J. Harbinson, & E. Wientjes (2023), The effects of different daily irradiance profiles on arabidopsis growth, with special attention to the role of PsbS, *Front. Plant Sci.*, *14*, 1–16, doi:10.3389/fpls.2023.1070218.

Senay, G., N. M. Velpuri, S. Bohms, M. Budde, C. Young, J. Rowland, & J. Verdin (2015), Drought monitoring and assessment: remote sensing and modeling approaches for the famine early warning systems network, in *Hydro-meteorological hazards, risks and disasters*, pp. 233–262, Elsevier.

Shahra, E. Q., & W. Wu (2023), Water contaminants detection using sensor placement approach in smart water networks, *Journal of Ambient Intelligence and Humanized Computing*, *14*(5), 4971–4986.

Shelby, Z., K. Hartke, & C. Bormann (2014), The Constrained Application Protocol (CoAP), RFC 7252, doi:10.17487/RFC7252.

Silva, G. M. e., D. F. Campos, J. A. T. Brasil, M. Tremblay, E. M. Mendiondo, & F. Ghiglieno (2022), Advances in technological research for online and in situ water quality monitoring—a review, *Sustainability*, *14*(9), doi:10.3390/su14095059.

Singh, M., & S. Ahmed (2021), Iot based smart water management systems: A systematic review, *Materials Today: Proceedings*, *46*, 5211–5218, doi:10.1016/j.matpr.2020.08.588, international Conference on Innovations in Clean Energy Technologies (ICET2020).

SMN (2017), Una historia de mediciones para entender el cambio climático, https://www.smn.gob.ar/noticias/una-historia-de-mediciones-para-entender-el-cambio-clim%C3%A1tico, accedido: 2024-06-05.

Stewart, B. (2015), Measuring what we manage–the importance of hydrological data to water resources management, *Proceedings of the International Association of Hydrological Sciences*, *366*, 80–85.

Tecmes, S., & S. Evarsa (2020), Joint Venture to Operate the Flood Early Warning System in the Lower-Catchment of the Salado River, SFe Prov, Arg, *Tech. rep.*

Testa, G., F. Gresta, & S. Cosentino (2011), Dry matter and qualitative characteristics of alfalfa as affected by harvest times and soil water content, *European Journal of Agronomy*, *34*(3), 144–152, doi:10.1016/j.eja.2010.12.001.

Tetzlaff, D., S. Carey, J. McNamara, H. Laudon, & C. Soulsby (2017), The essential value of long-term experimental data for hydrology and water management, *Water Resources Research*, *53*, 2598 – 2604, doi:10.1002/2017WR020838.

Thompson, A. I., H. H. Schomberg, S. R. Evett, D. K. Fisher, S. B. Mirsky, & S. C. Reberg-Horton (2021), Gateway-node wireless data collection system for environmental sensing, *Agrosystems, Geosciences & Environment*, *4*(4), e20,219, doi:https://doi.org/10.1002/agg2.20219.

Tjørve, K., & E. Tjørve (2017), The use of gompertz models in growth analyses, and new gompertz-model approach: An addition to the unified-richards family, *PLoS ONE*, *12*, doi:10.1371/journal.pone.0178691.

Tucak, M., D. Horvat, T. Čupić, G. Krizmanić, & M. Ravlić (2023), Assessment of alfalfa populations for forage productivity and seed yield potential under a multi-year field trial, *Agronomy*, *13*(2), doi:10.3390/agronomy13020349.

Tucker, C. J. (1980), A critical review of remote sensing and other methods for non-destructive estimation of standing crop biomass, *Grass and Forage Science*, *35*(3), 177–182, doi:/10.1111/j.1365-2494.1980.tb01509.x.

Urcola, H. A., X. A. de Sartre, I. Veiga, J. Elverdin, & C. Albaladejo (2015), Land tenancy, soybean, actors and transformations in the pampas: A district balance, *Journal of Rural Studies*, *39*, 32–40, doi:/10.1016/j.jrurstud.2015.03.001.

Vaschetto, P., L. Regaldo, W. Polla, V. Andrade, S. Gervasio, & A. M. Gagneten (2021), Plankton community responses to anthropogenic pollution in an argentinian urban reserve, *Water, Air & Soil Pollution*, doi:10.1007/s11270-021-05380-0.

Velázquez, J., T. Petit, A. Lavoie, M.-A. Boucher, R. Turcotte, V. Fortin, & F. Anctil (2009), An evaluation of the Canadian global meteorological ensemble prediction system for short-term hydrological forecasting, *Hydrology and Earth System Sciences*, *13*(11), 2221–2231, publisher: Copernicus Publications Göttingen, Germany.

Verdin, A., C. Funk, B. Rajagopalan, & W. Kleiber (2016), Kriging and local polynomial methods for blending satellite-derived and gauge precipitation estimates to support hydrologic early warning systems, *IEEE Transactions on Geoscience and Remote Sensing*, *54*(5), 2552–2562, publisher: IEEE.

Vereecken, H., et al. (2014), On the spatiotemporal dynamics of soil moisture at the field scale, *Journal of Hydrology*, *516*, 76–96, doi:10.1016/j.jhydrol.2013.11.061.

Viglizzo, E., F. Frank, L. Carreño, E. Jobbágy, H. Pereyra, J. Clatt, D. Pincén, & M. Ricard (2011), Ecological and environmental footprint of 50 years of agricultural expansion in Argentina, *Global Change Biology*, *17*(2), 959–973, doi:10.1111/j.1365-2486.2010.02293.x.

Vionnet, C., P. Tassi, L. Rodriguez, & C. Ferreira (2006a), Numerical modelling of the catastrophic flooding of Santa Fe City, Argentina, *International Journal of River Basin Management*, *4*(4), 301–314, doi:10.1080/15715124.2006.9635299.

Vionnet, C., P. Tassi, L. Rodriguez, & C. Ferreira (2006b), Numerical modelling of the catastrophic flooding of Santa Fe City, Argentina, *International Journal of River Basin Management*, *4*(4), 301–314, doi:10.1080/15715124.2006.9635299, publisher: Taylor & Francis.

Wada, Y., et al. (2017), Human–water interface in hydrological modelling: current status and future directions, *Hydrology and Earth System Sciences*, *21*(8), 4169–4193, doi:10.5194/hess-21-4169-2017.

Wang, H., Y. Lin, Z. Wang, Y. Yao, Y. Zhang, & L. Wu (2017), Validation of a low-cost 2d laser scanner in development of a more-affordable mobile terrestrial proximal sensing system for 3d plant structure phenotyping in indoor environment, *Comput. Electron. Agric.*, *140*, 180–189, doi:10.1016/j.compag.2017.06.002.

Wood, E. F. (1998), *Chapter 3: Hydrologic Measurements and Observations: An Assessment of Needs. Hydrologic Sciences: Taking stock and looking ahead*, National Research Council. National Academies Press, Washington, DC, USA.

Yang, D., Y. Yang, & J. Xia (2021), Hydrological cycle and water resources in a changing world: A review, *Geography and Sustainability*, *2*(2), 115–122, doi:https://doi.org/10.1016/j.geosus.2021.05.003.

Yang, S., M. Gao, C. Xu, J. Gao, S. Deshpande, S. Lin, B. Roe, & H. Zhu (2008), Alfalfa benefits from medicago truncatula: the RCT1 gene from m. truncatula confers broad-spectrum resistance to anthracnose in alfalfa, in *Proc Natl Acad Sci USA*, vol. 105(34), p. 12164 –12169, doi:10.1073/pnas.0802518105.

Zakaria, M. I., W. A. Jabbar, & N. Sulaiman (2023), Development of a smart sensing unit for lorawan-based iot flood monitoring and warning system in catchment areas, *Internet of Things and Cyber-Physical Systems*, *3*, 249–261, doi:10.1016/j.iotcps.2023.04.005.

Zeng, H., G. Dhiman, A. Sharma, A. Sharma, & A. Tselykh (2023), An iot and blockchain-based approach for the smart water management system in agriculture, *Expert Systems*, *40*(4), e12,892, doi:10.1111/exsy.12892.

Zhang, Y., C. Li, X. Zhou, & B. M. III (2002), A simulation model linking crop growth and soil biogeochemistry for sustainable agriculture, *Ecological Modelling*, *151*(1), 75 – 108, doi:http://dx.doi.org/10.1016/S0304-3800(01)00527-0.

Zhao, T., Y. Zhu, J. Wu, M. Ye, W. Mao, & J. Yang (2020), Quantitative estimation of soil-ground water storage utilization during the crop growing season in arid regions with shallow water table depth, *Water*, *12*(12), doi:10.3390/w12123351.

Zhou, B. B., R. P. Brent, & A. Tridgell (1994), Efficient implementation of sorting algorithms on asynchronous distributed-memory machines, in *Proceedings of 1994 International Conference on Parallel and Distributed Systems*, pp. 102–106, IEEE.

Zhou, G., & Q. Wang (2018), A new nonlinear method for calculating growing degree days, *Scientific Reports*, *8*(10149), 1–14, doi:10.1038/s41598-018-28392-z.

Zhou, H., W. Jia, Y. Li, & M. Ou (2021), Method for estimating canopy thickness using ultrasonic sensor technology, *Agriculture*, *11*(10), doi:10.3390/agriculture11101011.

Doctorado en Ingeniería
mención recursos hídricos

Título de la obra:

Hacia un Nuevo Paradigma en la Adquisición y Gestión del Dato Agro-hidrológico

Autor: Emiliano Pedro López

Lugar: Santa Fe, Argentina

Palabras Claves:

hardware libre / open-source, calibración de sensores,
IoT, Internet de las Cosas, AIoT, Inteligencia Artificial de las Cosas,
sensores de humedad del suelo, sensores de nivel freático,
monitoreo hidrológico, monitoreo agro-hidrológico,
modelos basados en datos, algoritmos de aprendizaje automático
crecimiento de cultivos, rendimiento de alfalfa