UNIVERSITÀ DEGLI STUDI DI PALERMO
Information And Communication Technologies
ING-INF/03-Telecomunicazioni
Dipartimento di Ingegneria

Internet of Things and Artificial Intelligence for Smart Sustainable Agriculture

Ph.D. Candidate
Ing. Antonino Pagano

Coordinator
Prof. Ilenia Tinnirello

Tutor
Prof. Ilenia Tinnirello

Co-Tutors
Prof. Daniele Croce
Prof. Gianpaolo Vitale
Ph.D. Roberto La Rosa

XXXVI CYCLE - ACADEMIC YEAR 2023-2024

Abstract

The main objective of this thesis work involves the improvement and dissemination of digitization in agriculture. In fact, in the near future, the agricultural sector is called upon to face a significant challenge due to increasingly scarce resources, extreme weather conditions, population growth, and the reduction of cultivable land. One solution is to transition from the old concept of agriculture to Smart Agriculture, with the adoption of the Internet of Things (IoT) and Artificial Intelligence (AI). Despite the investments made in recent years to connect rural areas to the internet, a barrier to the digitization of agriculture is represented by limited radio connectivity in rural areas. Radio networks are essential for easily connecting sensors/actuators over large areas and enabling new modes of interaction between the real world and the digital world. The purpose of this research is to bring the Cloud closer to the agricultural environment, by developing IoT systems for data collection and processing directly in the field, at the edge of the network, in order to rationalize the natural resources employed, reduce pollution, and optimize agricultural production, also with the assistance of artificial intelligence. The basic idea of this research is to develop decision support systems that minimize traffic generated towards the Internet and both the costs and energy consumption of the devices used for monitoring and controlling farms.

This present work has been divided into two parts. The first part discusses sensors, hardware, and wireless technologies for smart agriculture, while the second part reports on some specific applications for sustainable agriculture using artificial intelligence algorithms and graph signal processing.

In order to develop IoT solutions for sustainable agriculture 4.0, the initial part of this work analyzes the state of the art in precision agriculture and network-level technological aspects, in order to understand current solutions and potential future challenges. Indeed, from the study conducted in Chapter 1, it is deduced that LPWAN

technologies meet the requirements for developing IoT solutions in smart agriculture, such as wide coverage, low consumption, and low cost. In particular, several studies show that LoRa technology is suitable for agricultural applications such as soil and air monitoring, irrigation, livestock farming, etc.

In Chapter 2, an in-depth study is presented on the coexistence between LPWAN technologies such as LoRaWAN and Sigfox in realistic urban and rural scenarios with different duty cycles and traffic conditions. The choice of such scenarios and simulation parameters is supported by a detailed literature review and real-world experiments conducted in rural areas. Additionally, a mathematical model for interference evaluation is proposed, and the results are quantified using both the developed model and the SEAMCAT simulator. These results offer new perspectives on the coexistence of LoRaWAN and Sigfox for emerging IoT applications, such as smart agriculture in rural areas and vertical urban farming.

Chapter 3 presents an innovative wireless piezoresistive sensor that uses an energy harvesting system to convert ambient light into electrical energy. The Wireless Sensor Node (WSN) achieves high sensitivity and accuracy in strain measurements due to the carefully designed piezoresistive element and time-domain-to-digital conversion (TDDC) technology that eliminates the need for expensive analog-to-digital converters and reduces power consumption. Its energy autonomy feature and wireless communication capabilities make it an ideal solution for applications such as a wireless and battery-free resistive sensor, enabling real-time monitoring of various agricultural scenarios.

In Chapter 4, adaptive transmission algorithms will be studied to enhance the performance of battery-less IoT sensors based on the LoRa protocol. We design appropriate capacity-based storage, considering a renewable energy source (e.g., a photovoltaic panel), and analyze the probability of energy failures by exploiting both theoretical models and real solar energy traces. We explore various rural agricultural scenarios and dynamically adjust key network parameters, including inter-packet transmission time, data redundancy, and packet size, to optimize device performance. Additionally, we aim to optimize data sampling and transmission even when the natural energy source is absent, for example, during the night.

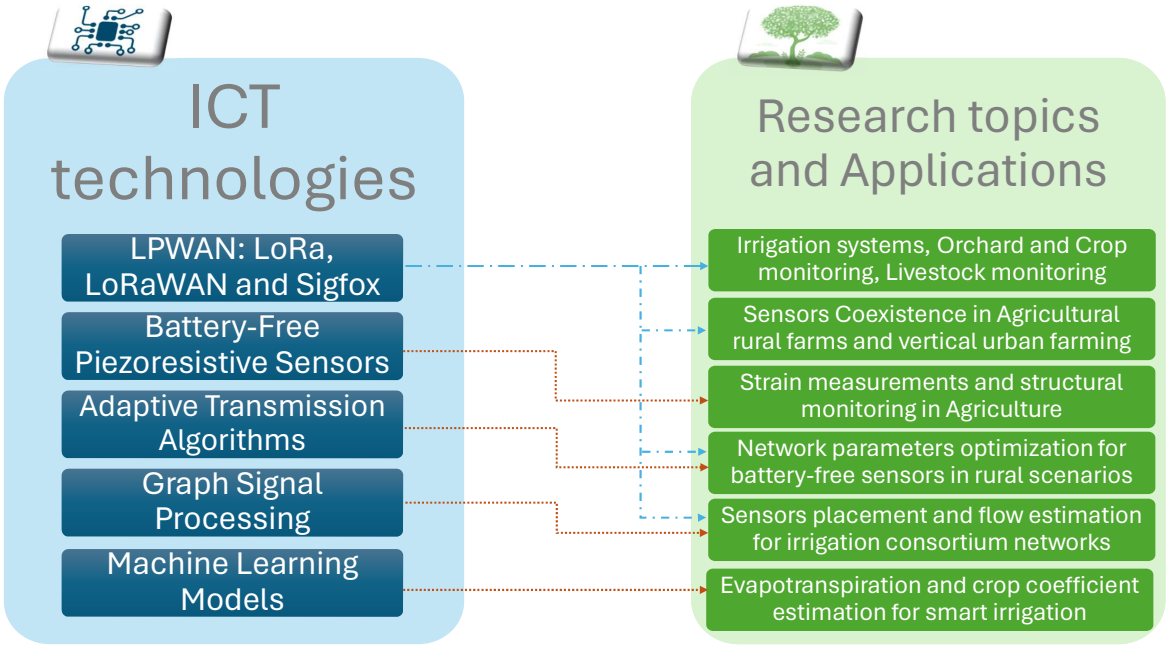
Chapter 5 introduces the second part of the thesis, which focuses on specific applications in agriculture. In particular, it presents an original approach for Smart Water Grids through the Internet of Things (IoT). This approach involves the integration of multiple sensors placed across irrigation consortium networks to accurately measure water flow, facilitated by a graph-based model.

Chapter 6 provides a successful application of Internet of Things (IoT) and Artificial Intelligence (AI) technologies for developing Smart and Sustainable Agriculture. In particular, it presents an example of an IoT system designed to monitor and predict soil water content, actual evapotranspiration, and other environmental variables, with the objective of using AI for precise irrigation scheduling in Mediterranean tree crops. The data collected during the monitoring period are used for training Machine Learning (ML) models and predicting daily actual evapotranspiration (ET_a) in a citrus orchard with a regulated deficit irrigation (RDI) strategy, using different feature combinations.

Finally, the chapter 7 presents a framework that leverages the potential of machine learning (ML) algorithms in combination with the seasonal decomposition (STD) algorithm to effectively estimate the time series dynamics of the crop coefficient (K_c). The thesis work concludes with Chapter 8.

Graphical Abstract

The following figure graphically represents the work done in this doctoral dissertation, illustrating the connections between the various chapters and the links between the software or hardware technologies used in the research and their applications in the context of smart and sustainable agriculture.



Published Content

This thesis is based on the following papers:

1. **A. Pagano**, D. Croce, I. Tinnirello and G. Vitale, "*A Survey on LoRa for Smart Agriculture: Current Trends and Future Perspectives*," in IEEE Internet of Things Journal, vol. 10, no. 4, pp. 3664-3679, 15 Feb.15, 2023, doi: 10.1109/JIOT.2022.3230505.

2. D. Garlisi, **A. Pagano**, F. Giuliano, D. Croce and I. Tinnirello, "*A Coexistence Study of Low-Power Wide-Area Networks based on LoRaWAN and Sigfox*," 2023 IEEE Wireless Communications and Networking Conference (WCNC), Glasgow, United Kingdom, 2023, pp. 1-7, doi: 10.1109/WCNC55385.2023.10118692.

3. **A. Pagano**, F. Amato, M. Ippolito, D. De Caro, D. Croce, A. Motisi, G. Provenzano, I. Tinnirello, "*Machine learning models to predict daily actual evapotranspiration of citrus orchards under regulated deficit irrigation*", Ecological Informatics, Volume 76, 2023, https://doi.org/10.1016/j.ecoinf.2023.102133.

4. F. Giuliano, **A. Pagano**, D. Croce, G. Vitale, and I. Tinnirello, 2023 "*Adaptive algorithms for batteryless LoRa-based sensors,*" in Sensors 23, no. 14: 6568, https://doi.org/10.3390/s23146568.

5. **A. Pagano**, F. Amato, M. Ippolito, D. De Caro, D. Croce, A. Motisi, G. Provenzano, and I. Tinnirello. "*Internet of Things and Artificial Intelligence for Sustainable Agriculture: A Use Case in Citrus Orchards,*" Presented at IEEE 9th World Forum on Internet of Things (IEEE WFIoT2023), Aveiro, Portugal.

6. M. Costanza, **A. Pagano**, S. Margueron, I. Tinnirello and R. La Rosa. *"An Energy-Autonomous and Battery-Free Resistive Sensor using a Time-Domain*

to Digital Conversion with Bluetooth Low Energy connectivity," submitted to IEEE International Symposium on Circuits and Systems (IEEE ISCAS2024), Singapore.

7. M. Costanza, A. Laaraibi, **A. Pagano**, G. Jodin, F. Razan, I. Tinnirello, S. Margueron and R. La Rosa *"Innovative Battery-Free Wireless Piezoresistive Sensor for Green IoT Applications*", Submitted to IEEE Transactions on Green Communications and Networking.

8. **A. Pagano**, F. Amato, M. Ippolito, D. De Caro, D. Croce. *"A Machine Learning Framework to Estimate Crop Coefficient Dynamics of Citrus Orchards*," submitted to Computers and Electronics in Agriculture.

9. T. Cattai, S. Colonnese, D. Garlisi, **A. Pagano**, F. Cuomo, *"Graph-based Energy efficient IoT sensing for Water Distribution Network monitoring,*" Submitted to ACM Transactions on Sensor Networks.

Table of Contents

List of Figures

List of Tables

Part 1: Sensors, Hardware, and Wireless Technologies for Smart Agriculture

Chapter 1

LoRa Technology for Smart Agriculture: Current Trends and Future Perspectives

1.1 Overview

The chapter provides a comprehensive survey on the integration of LoRa technology in the agricultural sector. It critically evaluates cutting-edge solutions for Smart Agriculture, assessing the potential of LoRa in various on-field applications. Specifically, the study explores four key scenarios: irrigation systems, plantation and crop monitoring, tree monitoring, and livestock monitoring. These scenarios present diverse requirements concerning network bandwidth, density, sensor complexity, energy consumption, and decision-making latency. The analysis examines the viability of LoRa-based solutions within these contexts, including their scalability, interoperability, network architecture, and energy efficiency. Additionally, potential avenues for future research are outlined, and emerging challenges that could shape research trends in the coming years are highlighted.

1.2 Introduction

In the near future, the agricultural sector is called to face a significant challenge due to increasingly scarce resources, extreme weather conditions, a growing population, and a reduction in arable land [5]. Indeed, according to the FAO, the Food and Agriculture Organization of the United Nations, by 2050 the world's population will reach about 10 billion and, to be able to feed everyone, we will have to produce 70% more food [6, 7]. A practical and feasible solution is to move from the old farming concept to

Smart Agriculture, with the adoption of Information and Communications Technologies (ICT) that help farmers to monitor, manage and optimize their operations more effectively [8]. In particular, the introduction of Internet of Things (IoT) applications, every single step of agricultural production can be improved: from soil management to minimizing water consumption, from plant protection to animal health and farm automation [9, 10]. Smart devices located infield are able to collect information and control the evolution of the different processes at various production stages. Besides, the miniaturization of electronic components allows to implement IoT sensors with reduced form factor and energy consumption, monitoring many variables of interest such as temperature, humidity, wind speed and direction, soil conditions, chemical concentrations, crop growth, and solar exposure, as well as possible damages caused by drought, hail, or flooding. IoT systems integrate all this and other data and turn it into useful statistics: for example, predictive analysis allows farmers to use the strictly necessary resources (water, pesticides, fertilizers, etc.) and only where there is a real need, e.g. for poorly irrigated areas of the field, weak or sick plants, etc. [11, 12]. Despite such potential benefits, the deployment of smart agriculture systems is still in its infancy. Indeed, an obstacle to the digitization of agriculture is the lack or limitations of Internet connectivity in many areas. In the literature, several communication protocols have been proposed, with different characteristics related to cost, coverage, power consumption, and reliability [13]. Among the available technologies (summarized in figure 1.1 in terms of power consumption and coverage range), Low Power Wide Area Networks (LPWAN, enlightened in a dashed box in figure 1.1) represent the best solution for supporting Smart Agriculture requirements. One of the most adopted LPWAN technology is LoRaWAN, which offers wide network coverage, built-in security, low cost, and limited power consumption during operation [14].

Indeed, LoRaWAN is an open system based on a very robust modulation (called LoRa), which provides several interesting features for covering rural areas with simple devices [13]. For such reasons, LoRa has been widely employed and tested in the agricultural field, connecting environmental sensors measuring temperature, air/soil moisture, etc., or to control different kinds of actuators (e.g. irrigation valves), and in applications such as tractor communications, livestock monitoring and location tracing [15–17].

In this chapter, we analyze the adoption of LoRa systems from a holistic perspective, contrasting it with existing works that either group LoRa with all other IoT solutions or concentrate solely on specific technological aspects. For example, LoRa is cited in [9] in the general framework of IoT; [11] cites LoRa among the most promising

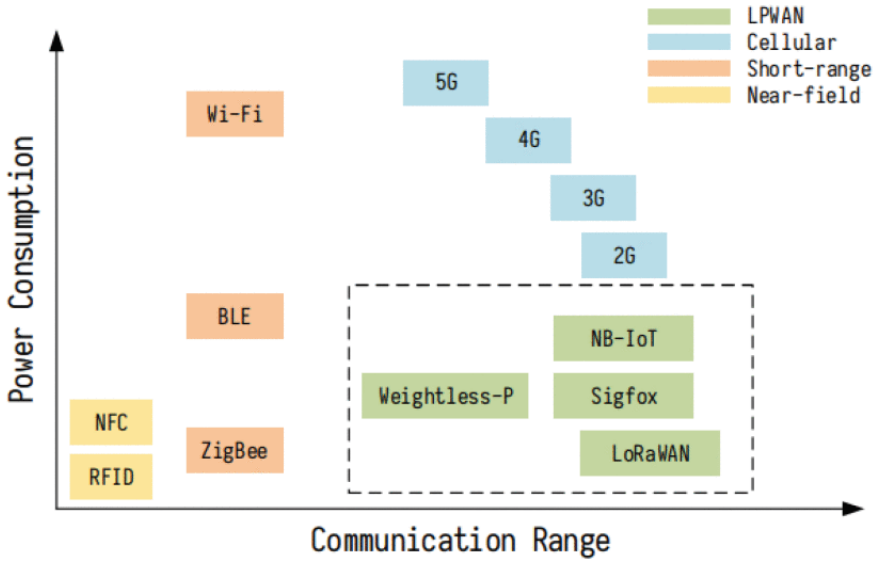


Figure 1.1: Classification of wireless technologies: power consumption versus communication range [1].

technologies for agricultural IoT, and the same plan is followed by [12] which dedicated a subsection to LoRa in Enabling Communication Technologies; similarly [18] includes LoRa in IoT communication protocols suitable for smart agriculture. On the other hand, [19] is specific to LoRa but is focused only on the protocol performance, [16] takes into consideration only the energy consumption of LoRa, [20] discusses the application of the technologies of industry 4.0 in the context of smart agriculture. Papers [21] and [22] deal with specific issues meaning the decision support system and robotics in agriculture respectively. Instead, in this chapter we strive to provide a thorough and focused analysis on LoRa/LoRaWAN application in smart agriculture, offering a comprehensive view of the advancements and in-field applications of this IoT technology.

LoRaWAN relies on LoRa modulation, a robust chirp-based modulation scheme, patented by Semtech [23]. It supports wireless connectivity with limited data rates over large areas and without the need of an operator. LoRaWAN is widely used in smart industry, smart home, smart city and, increasingly, in the smart agriculture environment. Unfortunately, there is still a gap between the digitization of industries and cities when compared to the digitization of agriculture. In the view of the author of this work, LoRaWAN possesses five main strengths (low-cost, long-range, low-power, no-operator and unlicensed spectrum) that can bridge the gap between smart agriculture and smart cities or industries. Moreover, several recent improvements on the resource allocation, channel access protocol and network planning, can

enhance the efficiency of LoRaWAN networks, reducing capital and operational costs [24]. In rural areas, several experiments have demonstrated good coverage of LoRa [25, 26, 15]. Coverage ranges of up to 5 km and 47 km have been obtained in the non-line-of-sight (NLOS) and line-of-sight (LOS) propagation conditions, respectively. In addition, in the case of NLOS propagation, the coverage range can be increased by using an unmanned aerial vehicle (UAV) [27, 28]. In terms of power consumption, LoRa offers up to 15 years of battery life to its devices. The low power consumption is a key feature of LoRa that makes it an ideal choice for Smart Agriculture applications. It was demonstrated experimentally that the estimated battery lifetime of a LoRa device may be six times that of a Wi-Fi device and two times that of a ZigBee device [29].

In this chapter, a comprehensive examination is conducted on LoRa-based Smart Agriculture systems, analyzing the state-of-the-art and highlighting for each solution the potential adoption of Machine Learning, control automation techniques, and energy autonomy features. These works are further classified into four main categories: (i) irrigation systems, (ii) plantation and crop monitoring, (iii) tree monitoring, and (iv) livestock monitoring. Moreover, these LoRa systems are analyzed in terms of scalability, interoperability, network architecture, energy-efficiency, and point out some open issues which traverse most of the current Smart Agriculture systems. References have been selected based on recent papers dealing with smart agriculture; however, some contributions in different fields useful as a benchmark are cited as well. Finally, lessons learned are presented, and future research directions are drawn, which are deemed crucial for the success and widespread adoption of such technologies.

The rest of the chapter is structured as follows: Section 1.3 provides an overview of LoRa technology and its perspectives in agriculture, while section 1.4 offers a brief summary on Smart Agriculture applications and their main challenges. Section 1.5 reports on general purpose LoRa-based IoT platforms applied to Smart Agriculture. Section 1.6, instead, discusses specific vertical solutions for Smart Agriculture, according to the four above-mentioned categories. Lessons learned and future research directions are proposed in Section 1.7. Finally, conclusions are drawn in Section 1.8.

1.3 LoRa and its perspectives in agriculture

LoRa technology is a proprietary physical layer telecommunications technology patented by Semtech, which is revealing as a promising solution for large-scale low power IoT deployments, including smart agriculture applications. Indeed, by operating in the

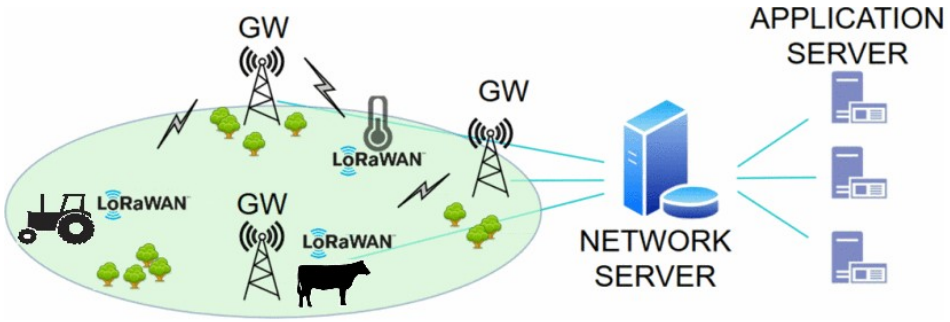


Figure 1.2: A LoRaWAN typical architecture.

unlicensed Industrial, Scientific and Medical (ISM) radio bands and with a robust chirp-based modulation scheme, LoRa provides a cheap solution for supporting wireless connectivity with limited data rates (from 0.3 kbps to 27 kbps) in large areas and without the need of an operator. Moreover, LoRa transmissions are regulated by having a maximum transmission power of 25 mW (14 dBm) in the uplink, a configurable bandwidth of 125kHz, 250kHz or 500kHz, and a duty cycle of 0.1%, 1.0% and 10%, which permit low energy consumption. In some scenarios, the battery of LoRa devices can last up to 15 years.

Although LoRa technology is limited to the physical layer, different network solutions can be built on top of it, by exploiting its transmission interfaces. Among these, the most consolidated one is the open-source solution promoted by the LoRa Alliance, which is called LoRaWAN [30]. LoRaWAN networks are based on a simple star of star topology (figure 1.2): end-devices (EDs), such as sensors or actuators deployed infield, transmit packets on the wireless medium to fixed nodes called gateways (GWs), which in turn forward the collected packets to a central Network Server (NS) interacting with several Application Servers (ASs) [31]. The network infrastructure between GWs, NS and ASs is typically based on a wired Internet technology, while EDs are not associated to a specific GW, which greatly simplifies implementation (e.g. in case of mobility [32]): in case a duplicate packet is simultaneously received by multiple GWs, the NS is responsible of filtering these packets and performs other simple decisions on network configuration.

To minimize the protocol complexity and the energy consumption, LoRaWAN employs a simple Aloha MAC protocol and defines three classes of devices (figure 1.3). Device classes represent different ways of managing the reception operations performed by the EDs. Class A devices, corresponding to the lowest energy profile, can receive downlink packets only in two time windows following the transmission of their own packet to the GW. In other words, devices can sleep all the time and downlink trans-

missions are triggered only after an uplink one. Class B devices add to this possibility a periodic scheduling of reception windows, by keeping a time synchronization with the GW. Lastly, class C devices are constantly listening to the channel for downlink packets. Any time that a new packet is ready for transmission, devices attempt to transmit by randomly selecting one of the available channels in the ISM bands (e.g. in the 868MHz there are 16 channels in Europe), together with a modulation parameter called Spreading Factor (SF).

More into details, six different SFs are used in LoRa (from SF7 to SF12), which result in different symbol times and in almost orthogonal transmissions: when two signals modulated at different SFs overlap, the GW is able to decode both transmissions in a wide range of power ratios among the signals [33]. Unlike many other IoT technologies, the LoRaWAN specification offers dedicated end-to-end encryption to application providers, together with network-level security primitives, which allow sharing the same network among multi-tenant applications [34].

Summarizing, the ease of deployment with excellent coverage, the availability of devices with very low energy demand and intrinsic security mechanisms, make these systems very suitable for innovative agriculture applications. Indeed, several state-of-the-art IoT applications in Smart Agriculture are based on LoRa/LoRaWAN networks. For example, LoRa is used to connect sensor nodes measuring environmental parameters or to control different kinds of actuators (e.g. solenoid valve for irrigation purposes), and in applications such as livestock monitoring and location tracing [16, 14, 17]. These applications are not critical for data rates and latency, but often require to work in large rural areas, with limited access to energy grids and the Internet, and with decision mechanisms which benefit from data-driven learning.

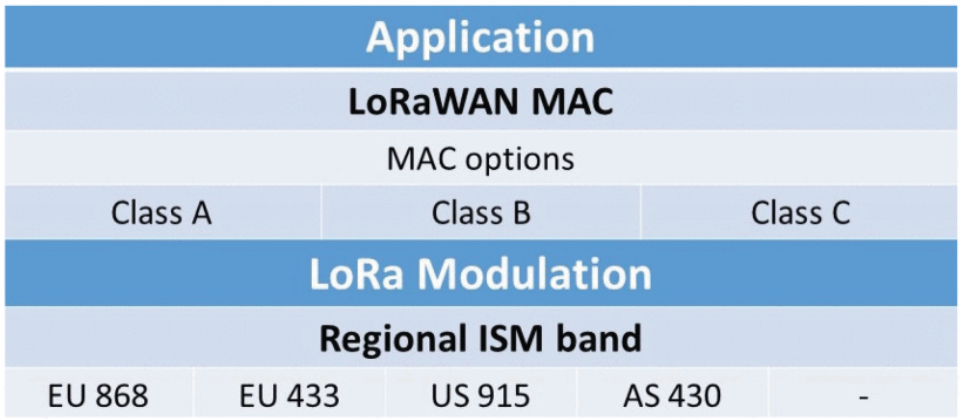


Figure 1.3: LoRaWAN technology stack.

1.4 Smart agriculture applications and challenges

While industrial production processes have already become smarter and autonomous thanks to the implementation of the so called Industry 4.0 concept, the integration of technologies such as IoT, AI, robotics, and big data is more recent in agriculture. The availability of IoT technologies for supporting wireless connectivity in rural areas and controlling infields smart objects shows a great potential for improving the agricultural sector, towards the so called Smart Agriculture [20–22]. Indeed, farm monitoring and automation can make production more efficient and sustainable [11], by promptly detecting and reacting to water or moisture stress, wastes of raw materials, crops' diseases, pests, and nutrient deficiencies, as well as problems related to the well-being of farm animals. The interest on the development of Smart Agriculture applications has been demonstrated by the recent commercialization of agricultural sensors and robots (called Agribots), specifically designed for reducing the intense physical labor traditionally required in agriculture [18].

Apart from the availability of smart devices for interacting with the farm in the physical world, smart agriculture applications require to build a digital representation of the farm status and a decision logic based on the collected data. Different protocols can be envisioned both for providing the wireless connectivity to heterogeneous devices (from simple low-cost temperature sensors, to complex remote-controlled robots) and exporting data for analysis and decisions [35, 36].

Since a large amount of data can be produced by agricultural sensors, big data analysis can provide efficient monitoring and processing methods [37]. Data processing, may involve various features such as data loading, validation, aggregation, prediction, classification, image or video processing, and data mining. Thus, based on the acquired data, decision support systems can optimize the productivity and reduce the ecological footprint of the farm.

Researchers recognize that digitization of farming processes and activities is an important challenge for the adoption of smart agriculture technologies [35, 38]. In particular, the major challenges to digitization in agriculture can be categorized in:

- **Communication issues:** as we will detail later, large-scale implementations of IoT solutions require robust and secure network architectures. The reliability of communicating information still represents a challenge to be addressed in the agricultural context and justifies the adoption of LoRa/LoRaWAN technologies.

- **Energy Management:** power supply in devices for Smart Agriculture is a significant challenge and energy harvesting systems are a relevant area of research. The main issue concerns the sensor's power supply and how to optimize efficiently the power consumption. Moreover, distributed nodes can execute some computations (Edge computing) which consumes more energy, while sensor batteries have a limited capacity. Consequently, smart devices require efficient energy storage and supply.

- **Data/device Heterogeneity:** in general, the agricultural data is produced by heterogeneous sensors (soil sensors, weather sensors, trunk sensors, leaf sensors, etc.). In addition, IoT devices generally use different network protocols and platforms. Thus, in addition to sensors heterogeneity, network and protocol heterogeneity should be considered as well. Getting these technologies to work together is often an issue, especially for unskilled farmers.

- **Physical Deployments:** spatial deployment of devices on farms proves to be a significant challenge, especially when the entire farm needs to be monitored across a large area and with different application scenarios (soil, plants, trees, animals, etc.).

- **Data management:** the difficulty of interpreting the data can be a huge barrier: indeed, numerous sensors are necessary and big data analysis could be required to better understand and forecast the unpredictability of agricultural ecosystems.

- **Generic platform:** to promote the adoption of Smart Agriculture technologies is often required to develop user friendly software platforms. The challenge here is to build a universal platform that can be easily modified to support different types of monitoring ranging from specific crop to livestock

These challenges, together with cost of infrastructure investment, complexity of technologies, lack of farmers' education and training, data ownership, privacy and security concerns, has motivated the research and development of innovative platforms, specific network technologies and new architectures for Smart Agriculture [35, 38].

1.5 Generic LoRa-based platforms

Since agricultural applications are widely different, varying from soil and air monitoring, to irrigation automation and livestock breeding, several general purpose IoT

platforms have been adapted for farmers to accommodate all these applications together under a unified, easy to understand and simple to use interface. Therefore, in this section, we will discuss some of these LoRa-based platforms horizontally designed for Smart Agriculture, while in the next one we will dig into more vertical and application-specific systems, focusing on the four reference scenarios depicted in figure 1.4.

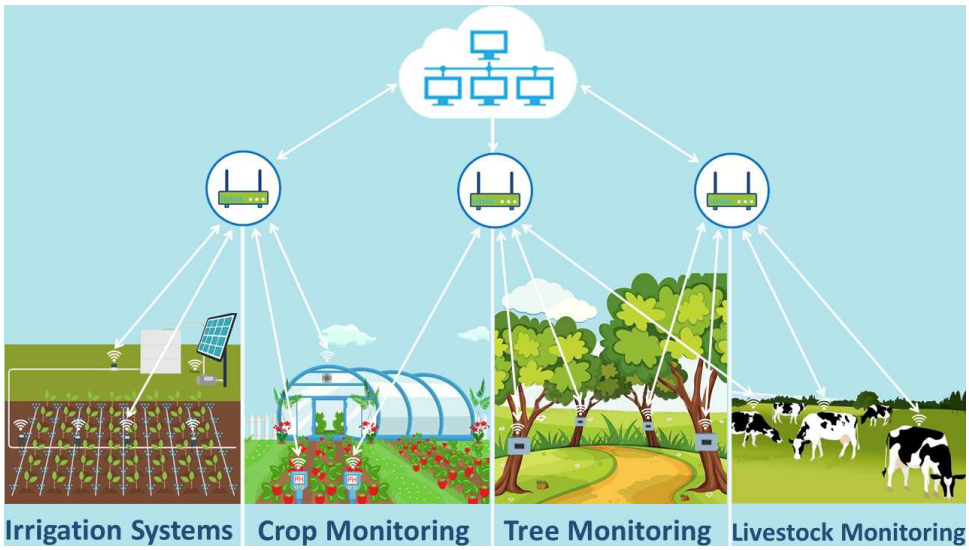


Figure 1.4: Four reference applications in Smart Agriculture.

Generic and open IoT platforms can indeed help to digitize farms by integrating numerous agriculture applications, harmonizing specific sensing devices, actuators, and decision logics, which exhibit heterogeneous requirements in terms of network bandwidth, latency, sensors' complexity, and energy requirements. A clear example is constituted by FIWARE [39], a powerful open source platform, sponsored by the European Commission, that provides standardized interfaces for many different IoT sectors including agriculture. The FIWARE platform includes several parts called Generic Enablers (GEs), which provide components and reference implementations that support specific APIs, and can integrate data collected from heterogeneous sensors using different communication technologies, to create custom applications [40, 41]. Several GEs are available making it easier to interface with IoT systems, and the IoT Agent for LoRaWAN offers a bridge between LoRaWAN and the FIWARE Context Broker (the core component of the "Powered by FIWARE" platforms). Moreover, FIWARE can be combined with other third-party platforms to provide accessible tools to worldwide farmers and consumers too [42]. Another example is the work in [43] where low-cost, LoRa-based devices are used for soil temperature and humidity

monitoring, and the data is processed and sent to the Cayenne IoT Platform for storage and visualization [44].

This platform is a drag and drop project builder for developers and engineers that can be used in different IoT applications. It encompasses cloud-based web applications as well as mobile apps for Android and IOS devices. Cayenne can integrate any tool into the library with a wide variety of IoT ready-to-use devices and connectivity options. Other LoRaWAN-based IoT platforms are more specific to the agricultural world, aiming at improving the management of generic farms in a highly-customizable way. For example, the LoRaFarM platform [2] has a generally applicable "core" infrastructure, which can be completed with specialized ad-hoc modules depending on the farm's characteristics and requirements. The LoRaFarM platform derives its topological structure from the LoRaWAN architecture, since low-level communication patterns are built around the LoRaWAN technology (see figure 1.5). Hence, expansion modules can be added at farm-level (or low-level), if they include physical hardware to be installed in the deployment (sensors or actuators), as well as a high level, in case data processing is needed. The middleware, in the LoRaFarM domain, refers to the set of entities and technologies by which data coming from farm-level modules are collected, stored, and exposed to high-level modules. This middleware can be defined as a sort of "connecting layer" between the farm and the back-end domain.

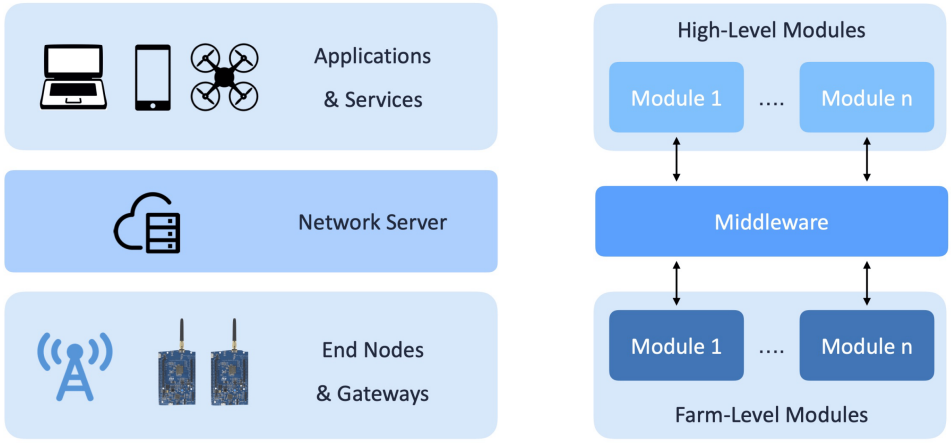


Figure 1.5: LoRaFarM platform: levels and parallelism with LoRaWAN [2].

LoRaFarM has been deployed and validated in an organic farm in Italy that covers both open-field and greenhouse cultivation together with the production of several different agricultural products. The platform thus integrates two farm modules: a vineyard module, useful to monitor soil parameters (soil moisture and temperature) of the farm vineyards, and a greenhouse module, which collects the environmental

conditions of the greenhouse. One of the main advantages of the LoRaFarM platform is that heterogeneous sub-networks, in terms of capabilities (transmission range, data throughput, energy consumption), can be incorporated without altering the platform structure and, thus, making it highly scalable, flexible, and suitable for a wide range of scenarios. Indeed, this gives the freedom to choose the most suitable communication protocols and traffic policy to monitor and control the farm different areas, such as greenhouses and fields. Messages between nodes employing different protocols are translated by a multi-protocol GateWay (mpGW), enabling communications between non-LoRaWAN-enabled nodes and the LoRaFarM middleware, in a seamless way. Its protocol translation functionality, the mpGW can be enriched with edge computing features, to process and aggregate sensor data. Moreover, LoRaFarM can be extended with new functionalities like data analysis and prediction of the evolution of environmental parameters to prevent plant diseases, relying on Artificial Intelligence and Machine Learning techniques.

Finally, the mySense environment proposed by Silva et al. [45] is a sensor data integration framework aimed to systematize data acquisition procedures to address common Smart Agriculture issues. It facilitates the use of low cost platforms such as Arduino and Raspberry Pi, making available a set of free tools based on the DIY (Do It Yourself) concept. The mySense platform builds over a 4-layer technological structure (sensor nodes, crop field and sensor networks, cloud services and front-end applications) and is accordingly divided into four levels of operation: Level 1, for data collection using common data transfer technologies (ZigBee, GSM/GPRS, LoRa, etc.); Level 2, for GWs (possibly) running local tasks according to the fog or edge computing paradigms; Level 3 for storing data in the cloud; and Level 4 for high-level applications. Data can arrive from any device provided that complies with the data formats allowed by the platform.

Section summary and insights. This section discussed LoRa-based platforms which can be exploited to unify different applications into one simple and easy-to-use platform. Platforms such as Fiware, Cayenne, LoRaFarM, and mySense provide standardized interfaces to integrate different agricultural applications with each other. These platforms provide ready-to-use solutions and connectivity between heterogeneous networks. With these platforms, LoRa can integrate and complement existing systems based on other network technologies (ZigBee, Bluetooth, etc.), making them highly scalable.

1.6 Application-specific LoRa platforms

In this section, an in-depth review is provided, focusing on four reference scenarios: irrigation systems, plantation and crop monitoring, tree monitoring, and livestock monitoring, which broadly encompass most Smart Agriculture applications.

1.6.1 Irrigation systems

Accurate monitoring of the soil water status allows to achieve seasonal water savings of up to 90% compared to traditional management, increasing productivity and introducing significant savings in energy costs for the water pumps management [46]. To improve water management in agriculture, it is necessary to analyze and monitor the complex water interactions that occur in field, following the concept of soil-plant-atmosphere (SPA) continuum systems [47]. Indeed, the knowledge of the water status of the SPA system plays a significant role for understanding the crop water stress and implement water saving mechanisms with a minimal effect on the production [46]. Measuring the evapotranspiration (ET), which refers to the amount of water that passes from the soil into the air due to the combined effect of plant transpiration and evaporation, is another complex task. Examples of these sensors are the lysimeters or sophisticated micro-meteorological sensors (e.g. Eddy covariance), whose cost and complexity limit their application to research studies [48]. Cheaper systems are the Time Domain Reflectometry (TDR) technique or gravimetric methods [49], whose main limit is the difficulty in calibration and automation. Some LoRa based irrigation systems are implemented using development boards such as Arduino, ESP32, Pycom or STM32, e.g. [50–54, 3]. Few of them also include energy harvesting modules, such as an hydroelectric generator, allowing them to operate for decades. For example, the LoRaWAN-based irrigation system in [3] comprises an energy-neutral irrigation node (figure 1.6) with the following modules: controller module, power module, irrigation module, and transmitter module.

Exploiting AI and data coming from different sensors, such as air temperature and humidity, soil temperature and humidity, light intensity, etc. makes possible to develop and train specific irrigation models to calculate the exact amount of water to be distributed. For example, the works [55–57] provide machine learning-based smart irrigation systems, all employing LoRa technology. In particular, in [56] a Random Forest Classifier predicts the soil moisture and thus irrigation is planned accordingly. In [55], instead, multiple linear regression algorithm is employed to train the model using two highest correlation coefficient features: light intensity and soil humidity.

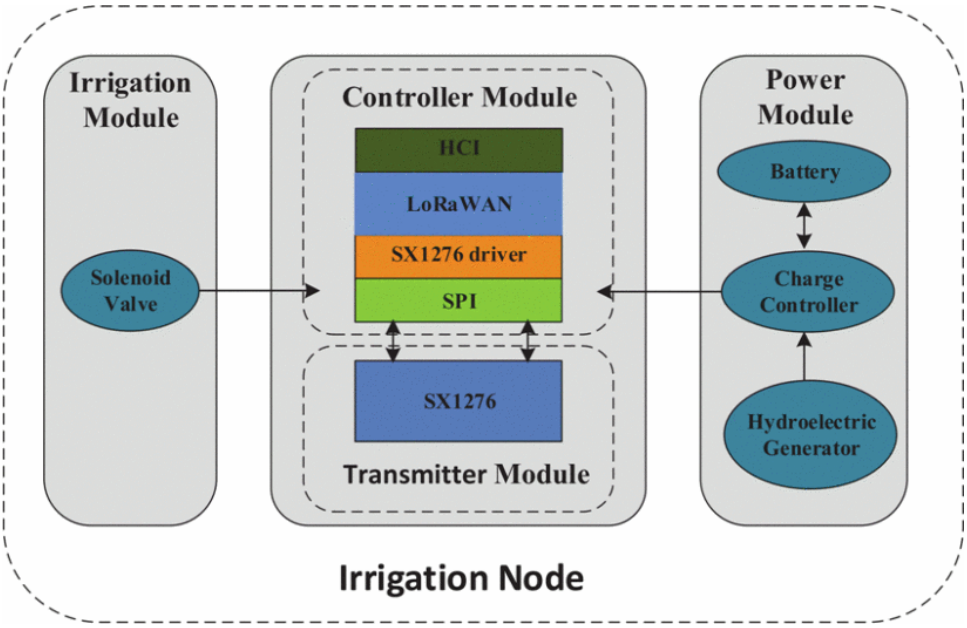


Figure 1.6: The energy-neutral irrigation node described in [3].

Data is collected with a LoRa P2P network, which uses a master-slave and TDMA-based MAC protocol. Each slave node has a unique address and can transmit a packet in each of the reserved TDMA time slots. Alternatively, a Penman-Monteith [58] based irrigation model allows for an optimal irrigation strategy for different crop growth periods and uses the ET parameter to estimate the amount of water. This solution requires an integration of actuators, sensors and a meteorological station in a LoRa network [59] [60]. In addition, third-party services such as weather information or fog computing may be needed to decide on irrigation schedules [61]. Since in LoRaWAN the latency of downlink communication from GW to Class-A nodes (sensors or actuators) is relatively long (must first wait for an uplink transmission), few systems employ alternative Master/Slave protocols [62, 63, 55]. These protocols increase the stability of the LoRa irrigation system, avoiding packet collisions, and thus can save water during the close command of the solenoid valve.

Finally, AREThOU5A [50] is an example of a water management system that combines data collected from wireless sensor networks in the field and satellite data provided by international weather forecast services, to achieve efficient water usage strategies for farmers. It employs a wireless sensor network with two different sensors for measuring the temperature and the soil moisture in field. A routing subsystem controls and routes the data and information through LoRaWAN and TCP/IP with SSL network interfaces. The LoRa network is used to collect data from the EDs and

Table 1.1: LoRa-based irrigation systems comparison

Reference	Features			Sensors						
	Machine Learning	Energy harvesting	Automatic Control	Temperature	Humidity	Soil Moisture	Soil Water Potential	Solar radiation	Water flow	Others
Boursianis et al. [50]	✓	✓	x	✓	x	✓	x	x	x	x
Kodali et al. [51]	x	x	✓	✓	x	✓	x	x	✓	x
Nisa et al.[64]	x	x	✓	✓	✓	✓	x	x	x	x
Ali et al. [54]	x	x	x	x	x	x	x	x	x	Pump current
Jiang et al. [52]	x	✓	x	✓	✓	✓	✓	x	x	x
Zhang et al. [53]	x	✓	✓	x	x	✓	✓	x	x	Pressure
Emharraf et al. [59]	x	✓	✓	✓	✓	✓	x	✓	x	Battery level
Yuan et al. [60]	x	✓	✓	✓	✓	✓	x	x	✓	x
Zhao et al. [3]	x	✓	x	x	x	x	x	x	x	x
Usmonov et al [62]	x	x	x	x	x	x	x	x	x	x
Vu et al. [63]	✓	x	✓	✓	✓	✓	x	x	x	x
Chang et al. [55]	✓	✓	✓	✓	✓	✓	x	✓	x	x
Henna et al. [56]	✓	x	✓	✓	✓	✓	x	x	x	Pressure
Loganathan et al. [57]	✓	x	✓	✓	x	✓	x	x	x	x
Froiz-Míguez et al. [61]	x	x	✓	✓	✓	✓	x	x	x	x

perform administration processes, while the TCP/IP SSL works as a bridge to the rest of the network architecture.

Subsection insights: comparing the characteristics of different irrigation systems, summarized in Table 1.1, it is relevant to note that most of these LoRa-based irrigation systems adopt temperature, humidity and soil moisture sensors. However, albeit all cited papers are recently published, ML is used only in 1/3 of the applications.

Furthermore, only 13% of these irrigation systems used an evapotranspiration-based methodology. This strategy, which is often expensive, may be accomplished by combining inexpensive sensors and artificial intelligence (with a more comprehensive approach integrating meteorological variables measured by a weather station with variables measured by soil sensors into the system), significantly lowering the cost of direct evapotraspiration measurements. Such improvements could lead to more effective water management, with the simultaneous impact of decreasing water usage and increase crop output. Finally, LoRaWAN communications can be tuned to adapt the duty cycle and manage the system optimally: for example, when the irrigation system is not in use, sensor data could be collected every hour or even less, while when irrigation is taking place the measurements could be increased to every 5-10 minutes. This way, the use of water and energy could be further reduced [65].

1.6.2 Plantation and crop monitoring

Plantation and crop monitoring requires a large number of sensors to obtain an effective control and thus increase productivity, especially when agricultural fields are very heterogeneous. For example, in order to optimize the production while minimizing the ecological footprint, it is necessary to control the injection of pesticides and fertilizers [66, 67], increasing yields up to 10% and saving fertilizers up to 37% [68]. Such control can be performed by varying the pesticides and fertilizer application rate over time and space. Indeed, today fertilizers are often applied uniformly without considering the variability in soil/crop characteristics within fields. Crops do not always need a uniform application, as some areas have different requirements due to their location (sunlight, soil features, etc). Over-fertilization can deteriorate water quality, favor weed growth, and reduce profit. Vice-versa under-fertilization restricts yield or reduces crop quality [69]. The application rate can be modified based on weather impacts, nutrient availability, and seasonal cycles [67, 69]. Some optical or ultrasonic sensors indirectly assess the nutrient request (nitrogen, phosphorus, potassium, etc.) of the crop at the time of application [66]. In addition, to reduce the loss of productivity in crops, surveillance systems can be adopted [70]. Providing visual monitoring to growers can prevent crops from getting damaged by intruders, ensure the field conditions or enable the detection of pests attacks remotely. Although there are plenty of devices which can be exploited for building a real-time visual monitoring system, deploying them in a wide area and over wireless channels can be challenging [71–73].

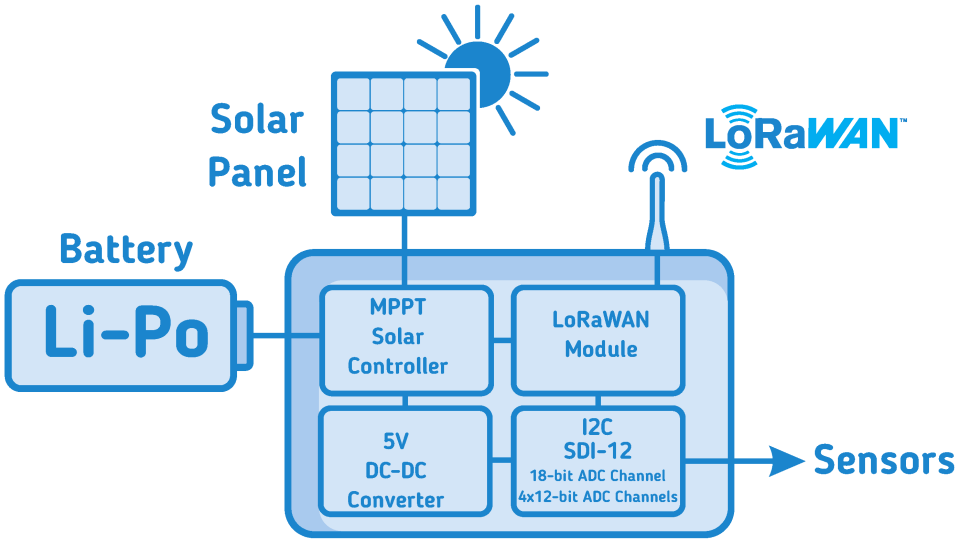


Figure 1.7: Block diagram of the sensor nodes taken from [4].

Table 1.2 summarizes the main characteristics of several plantation and crop monitoring systems, based on LoRa technology. The nodes used in these systems should be of small dimensions, self-sufficient in terms of energy, relatively cheap, and often able to acquire a large variety of parameters. For example, three different sensor nodes have been developed by Valente et al. [4] and tested in a vineyard field: node 1, with an ultrasonic anemometer (that measures the direction and speed of the wind) and a sensor that monitors bulk electrical conductivity, in addition to volumetric water content (by measuring soil permittivity) and soil and air temperature; node 2, an irrometer watermark soil water tension sensor; and node 3, an all-in-one weather station with 12 sensors to measure air temperature, relative humidity, vapor pressure, barometric pressure, wind speed, gust and direction, solar radiation, precipitation, lightning strike counter, and distance. It should be remarked that each node contains sensors which differ for the sampling rate, accuracy, and supplied energy. The nodes send data using LoRaWAN to a GW that is connected to a TTN (The Things Network) server. In the TTN server, data is decoded and sent to the ThingSpeak [74] platform for visualization and possible analysis and aggregation. Figure 1.7 summarizes the different blocks composing the nodes: a) a Maximum Power Point Tracker (MPPT) applied to a photovoltaic source and connected with a storage system, b) a DC/DC switching converter to interface the source with the storage system and loads; c) the LoRaWAN module for communication; d) the analog-to-digital converter (ADC) module to convert the signals available from sensors.

Table 1.2: LoRa-based plantation and crop monitoring systems comparison

	Features			Sensor					
Reference	Plantation	Energy harvesting	Automatic Control	Temperature	Humidity	Soil moisture	Carbon dioxide	Solar Radiation	Others
Valente et al. [4]	Vineyard	✓	x	✓	✓	✓	x	x	12 weather sensors
Shamshiri et al. [75]	Berry orchard	✓	x	✓	✓	✓	x	✓	Leaf sensor
Rachmani et al. [76]	Starfruit	x	x	x	x	✓	x	x	pH sensor
Ibrahim et al. [77]	Mushroom	x	✓	✓	✓	x	✓	x	x
Singh et al.[78]	Tomato	x	x	✓	✓	x	✓	✓	Electrical conductivity
Sacaleanu et al. [79]	Walnut	x	x	✓	✓	✓	x	✓	Pressure
Silva et al. [45]	Vineyard	✓	✓	✓	✓	x	x	✓	Pluviometers
Codeluppi et al. [2]	Vineyard	✓	✓	✓	✓	x	x	x	x
Brunelli et al. [80]	Apples	✓	x	x	x	x	x	x	Camera

In [76] a LoRa-based IoT monitoring system for starfruit plantation is presented. The LoRa network implemented includes three nodes and one master, and it can cover a range of 700 meters. For optimal growth, starfruit plants need soil pH conditions between 5,5-7,5. Thus, thanks to the proposed LoRa system, the farmers can make

important and precise decisions about how to grow the crop. Similarly, works [75, 77–79, 45] present solutions to increase production and fruit quality, with optimal use of resources through LoRa-based networks.

The Smart Mushroom Cultivation is a system used to automatize the production of expensive mushrooms [77]. The smart system includes devices to monitor and control humidity and CO_2 levels through sensors and actuators all connected using LoRaWAN. The sensor nodes measure the ambient condition inside Mushroom House (humidity, temperature, and CO_2), and data is sent to the remote server for monitoring and analysis. An automatic control maintains the ambient conditions between the required levels.

Lastly, there are cases where anomaly detection near the sensor is required to allow decisions and actions as soon as possible. In this direction, Brunelli et al. [80] propose a new paradigm of monitoring and pest detection to improve the performance of an apple orchard. They add intelligence to the LoRa nodes, shifting the detection of anomalies near the sensor. The system takes a picture of a pest trap, pre-processes it, classifies each insect, and eventually sends a notification to the farmer if any codling moth is detected. The application is developed on a low-energy platform powered by a solar panel, realizing an energy-autonomous system capable of operating unattended continuously over LoRa networks.

Subsection insights: plantation and crop monitoring requires the control of numerous parameters, captured by different heterogeneous sensors deployed in the agricultural fields. Some of the sensors used in the cited papers are specific to the type of crop, while others (e.g. temperature, humidity, etc.) are deployed in almost all of the literature works. In addition, the use of artificial intelligence is not yet widely adopted, and only 3 out of 9 papers adopt automatic control for the implementation of decision support systems (DSS). An innovative approach in this context would be to add intelligence to the LoRa nodes, while moving the DSS closer to the sensor. Finally, note that the maximum size of the LoRa payload is 250 bytes; this allows a wide variety of parameters to be monitored and transferred in a single packet. For example, in [79] the authors send 8 agri-meteorological measurements in a single LoRa packet of only 16 bytes.

1.6.3 Tree monitoring

Trees are essential in modern society and are widely applied in a great number of scenarios including soil erosion prevention, air purification, wood or fruit production. For supporting the managers of urban/rural green infrastructures and forests, it is im-

portant to constantly monitor the tree conditions, in terms of growth rate and failure risk, as well as micro-climate parameters in the tree surrounding areas. The analysis of this data allows the characterization of the trees functional responses to their environment and a prompt action in case of problems. Tree monitoring also requires reliable long-range communications in the presence of foliage, large sensor densification (i.e. one sensor per tree), and measurements of various physiological/biological parameters from specific locations (at the root, the trunk, or the branch) as a function of vegetation type to obtain accurate readings [81]. In these systems, it is also important to measure changes in position over time or instantaneous trunk accelerations. Table 1.3 summarizes the main characteristics of relevant Monitoring systems.

Table 1.3: LoRa-based sensors applications in tree farms

Reference	Temperature	Humidity	Soil moisture	Solar Radiation	Others
Valentini et al. [82]	✓	✓	x	✓	Sap flow
Amaro et al. [83]	x	x	x	x	Impedance spectroscopy
He et al. [81]	✓	✓	x	✓	Carbon dioxide
Klaina et al. [84]	✓	✓	✓	✓	x
Yim et al. [85]	✓	✓	✓	✓	x
Park et al. [86]	✓	✓	x	✓	x

One of the main properties to be measured is the water transport in the xylem of the trunk (called the sap flow). A possible measurement method is the Heat Balance Method, developed by Granier [87, 88], which is based on analyzing the temperature difference among two probes inserted into the stem wood at a 10 cm distance along the vertical trunk axis. The probe in the higher position is heated, while the lower one provides the stem wood reference temperature. The temperature difference generated between the probes represents an index of the transpiration activity of the plant, expressed as a variation of the flux density. This method can be used for accurate measurements of sap flow in plants, providing a reliable calibration procedure to relate the temperature difference to the actual sap flow [89].

For example, the TreeTalker (TT) [82] is a device that measures sapflow (water transport in the trunk), wood temperature and humidity, multispectral signature of light transmitted through the canopy, tree trunk radial growth, accelerations along a 3D coordinate system used to detect tree movements, air temperature, and relative humidity, which can be additionally complemented by soil temperature and volumetric water content. A TT node is connected via LoRa wireless connection to a GW, that manages up to 48 devices in one cluster. The GW is in turn connected to the internet

via GPRS and sends data to a computer server. This technology can be applied to monitor the root plate tilt, as well as the flexion and the accelerations that tree trunks receive under the force of the wind for the evaluation of tree failure risk.

Another solution to analyze the health condition of a tree consists on the Electrical Impedance Spectroscopy (EIS), it is a well-known technique with a wide range of applications. EIS has been applied to characterize solids, liquids, both in the laboratory and industrial environments. Moreover, assessment of physiological states of some trees (pinus, chestnut, etc.) has also been studied. The method based on bioimpedance indexes allows determining three distinct physiological states: healthy and watered plants, plants with a high level of hydric stress, and plants with disease [90–92]. For example, Amaro et al. [83] integrate an EIS system in a sensor node to analyze the health condition of the tree and transfer the results through the LoRaWAN protocol.

Finally, tree monitoring systems are often influenced by the presence of foliage which can severely impact wireless communication systems performance. Indeed, discrete scatters such as randomly distributed leaves, twigs, branches, and tree trunks can cause attenuation, scattering, diffraction, and absorption of the radiated waves. This generally leads to node densification to increase coverage levels, especially in large areas, resulting in additional costs and constraining the design of LoRa systems in non-homogeneous vegetation environments [84, 85, 93]. For such reasons, a small drone with a GW is sometimes required for collecting data from nodes and solving the Fresnel zone radio propagation issues encountered in tree farms [94, 86].

Subsection insights: in this sub-section, some methodologies for monitoring tree health have been discussed. It is important to highlight how LoRa can be easily integrated into these systems, e.g. to measure the lymph flow or bio-impedance of trees. In case the parameter to be monitored involves roots, it has been shown that LoRa can be used for under-ground or near-ground communications too [15, 95]. Finally, the use of drones for data collection has been exploited to solve the problem of foliage scattering.

1.6.4 Livestock monitoring system

The implementation of these practices requires to monitor the general health conditions of the animals, by tracking some biological signals to be associated to symptoms of disease, estrus and calving [96]. Wearable sensor technologies provide the possibility of remotely managing individual animals facilitating urgent interventions, responding to time and labor-intensive concerns in a more efficient way [97]. In extensive livestock production systems, the absence of access to networking and animal contact

Table 1.4: LoRa-based sensors applications in Livestock Monitoring

Reference	Gyroscope	Accelerometer	Temperature	Humidity	GPS	Carbon dioxide	Illuminance	Others
Dos Reis et al. [97]	✓	✓	✓	x	✓	x	x	Magnetometer
Li et al. [98]	x	✓	x	x	✓	x	x	x
Germani et al. [99]	x	x	✓	✓	x	✓	✓	Ammonia
Ikhsan et al. [100]	x	x	✓	✓	x	x	x	Heartbeat

presents a barrier to the effective use of these technologies. Wearable sensors, to be more practical for extensive management settings, must:

- Network over longer distances
- Have reliable power supplies (preferably renewable)
- Be low-cost so that damaged and lost sensors are less economically impactful
- Have data being transmitted in real-time

For these reasons, LoRaWAN technology is indicated for above described applications, some of which are summarized in table 1.4. Primarily, these systems are used to monitor the animal health, but by integrating LoRa technology with a GPS, remote grazing systems can be implemented [97, 98].

Animal monitoring can involve completely different scenarios; as a consequence, the LoRa network architecture could require a more specific design effort to work either in indoor or outdoor settings. For example, the work in [99] proposes two different versions of GWs: an indoor GW, designed for installation in sheltered areas such as barns and cowsheds and oriented towards dairy cattle livestock scenarios, and an outdoor version, more specific for open areas such as paddocks and pasture lands, and designed for beef cattle livestock scenarios. The indoor GW is conceived for monitoring several important physical parameters typical of the shed environment, such as temperature, relative humidity, illuminance, carbon dioxide (CO_2), and ammonia (NH_3) concentration, while the main purpose of the outdoor GW is to manage nodes in remote areas, far from the shed, directly on the pasture land. In the open field scenario, weather parameters (temperature and humidity) are collected, for purposes of correlation with the animal health status.

Moreover, in presence of large herds, the high node density could cause an increase in collisions between sent packets. In such scenarios, a MAC layer that includes a listen-before-talk (LBT) mechanism could prevent as much as possible packet collisions among nodes. Indeed, LBT-based carrier-sense multiple access with collision avoidance (CSMA/CA) can be incorporated with the physical layer of LoRa [99]. The CA

mechanism is based on a random retransmission time that randomizes the access of the nodes to the wireless medium.

The size of the pasture area is another factor to consider in deploying the LoRaWAN network. In particular, it has been shown that in large areas of pasture, the use of a mobile GW that moves along the track is a better solution than the use of one or more static GWs [100]. Contrarily, when the livestock area is not too large, using only one static GW is preferable because the data extraction rate value is high enough and the energy consumption is lower compared to multiple static GWs or one mobile GW. In both scenarios, small and large pasture areas, 1500 nodes were considered. Each node is a collar equipped with heart rate, temperature, respiration, and humidity sensors. The instance of cattle monitoring in New Mexico, as described by Actility [101], is one of the successful illustrations of a large-scale LoRaWAN-enabled deployment. Due to the large size of these desert ranches (10,000 to 20,000 hectares) and the large number of cows to track (up to 7000), monitoring and obtaining information regarding cattle well-being can be time-consuming and expensive. Indeed, while the cattle were previously followed using traditional GPS devices, the absence of reliable cellular connection throughout the whole grazing region made this method ineffective. These issues were solved with an off-the-shelf LoRaWAN solution because of its extensive range and good coverage. Finally, LoRa technology can be used for sharing the short text messages and voice messages in absence of cellular coverage. For example, COWShED [102] is used for supporting livestock transhumance in Senegal.

Subsection insights: in this sub-section we showed how LoRa is used to collect information about the movements and health of livestock, as well as on the conditions of grassland. LoRa can also aid herders in achieving remote grazing by combining data with electronic fences, to identify whether animals have crossed it. In addition, LoRa has been used to monitor environmental parameters of barns, demonstrating how this technology can be adopted in both outdoor and indoor scenarios. Additionally, innovative MAC schemes, such as LBT method could be implemented to minimize packet collisions when big herds present, and to mitigate the limits on the effective duty cycle of channel occupation. Finally, in absence of cellular coverage, LoRa/LoRaWAN solutions have been used for large-scale cattle monitoring or even supporting livestock transhumance for text/voice messages.

1.7 Lessons learned and open issues

This section discusses the lessons learned and the open research challenges for using LoRa technology in smart agriculture. According to the aim of this chapter, it was learned that, given the wide variety of sensors used in smart farming systems, different communication protocols need to be integrated, particularly when different platforms/vendors coexist and data must be collected from the various subsystems. In addition, since power supplies are frequently unavailable in a large agricultural area, nodes should be as energy self-sufficient as possible. Using local or edge data processing could mitigate this problem, optimizing the energy consumption. Moreover, keeping the logic on the Edge of the network could alleviate the hurdle on LoRa's centralized communications (especially on the downlink). The development of interoperability in smart agriculture systems can also be accelerated by platforms such as FIWARE and Cayenne, while machine learning can be used to model and analyze technical problems, improving scalability of LoRa networks and predicting network congestion.

The experience gained in Industry 4.0 can be transferred to agriculture, considering some peculiarities, including the need to cover large spaces that cannot be manned. In addition, there is the need to provide device power supply and data security (partially solved by leveraging on LoRaWAN built-in security schemes). Another significant factor is the initial cost of the system, which must be as low as possible since the pay-back time also depends on elements that cannot be predicted during the year, such as weather. Finally, it has been recognized that although ICT has long-term sustainability issues to be solved, they show great potential for improving the usage of natural resources, especially when Cyber-physical systems (CPS) are combined with IoT, artificial intelligence, machine learning, and neuromorphic computing techniques [103].

Through the study carried out in this chapter, it is also possible to understand in which area LoRa has been applied and is emerging in recent years. In particular, among the application areas discussed in section 1.6, figure 1.8 shows in a pie chart that more than 40% of the analyzed studies focus on water management, while almost 25% are dedicated on crop monitoring, followed by tree monitoring. This result is in line with recent market surveys on LPWANs (e.g. [104]), and other general studies on communications protocols for smart agriculture [13].

All this confirms the great potential of implementing Smart Agriculture solutions using IoT, and LoRa technology in particular. However, there are still some open

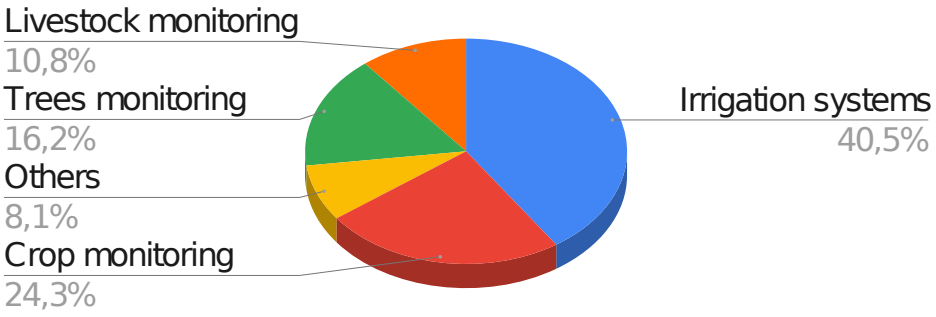


Figure 1.8: Distribution of the LoRa papers according to smart agriculture application areas.

issues that need to be faced: for example, LoRaWAN works quite well in uplink when it needs to collect data from sensors, while downlink connections might suffer high latency. In what follows, we briefly discuss future research directions related to downlink latency, energy management, device heterogeneity and interoperability, data management, and scalability. These open issues must be solved for LoRa systems to be widely adopted in Smart Agriculture. We conclude the section with an eye on other wireless technologies, different from LoRa/LoRaWAN.

1.7.1 LoRa downlink performance

The downlink performance of LPWAN systems still represents a challenge since it is related to the energy consumption. In particular, LoRaWAN allows different trade-offs between communication latency on the downlink channel and energy consumption. Nodes are classified by classes: they can receive only after an uplink transmission (Class A), or at regular time intervals (Class B), or at any time (Class C). The modern trend is to optimize energy efficiency, hence data are transmitted only when necessary or periodically. According to the authors' opinion, a further optimization could be retrieved by local data processing.

As a matter of fact, even if nodes remain asleep most of the time, as in [61], or with scheduling intervals of reception windows of 10-20 minutes as in [53, 60] or a few hours [45], local processing always lowers the data to be transmitted decreasing the transmission time; it has been successfully tested in [2] where edge processing on the GW allows a more effective control of the actuator nodes. This last approach improves also reliability since it allows farms to work even if the Internet connection of the LoRa GW is absent for a few hours. Reliability can also be improved by Master/Slave access control method for the LoRa network [62, 63, 55].

Alternatively, for short-range communication a Wake-up Radio (WuR) can be adopted. WuR technology is an ultra low power receiver that is continuously listening to the channel while spending a few nanowatts or microwatts depending on the circuit's design. WuRs work in parallel to the main LoRa transceiver and allow asynchronous wake-up of the nodes with low latency. With the LoRa-WuR scheme, the downlink latency can be reduced by almost 90% compared to the traditional LoRa protocol for a 10 nodes cluster [105, 106].

1.7.2 Energy efficiency considerations

In addition to downlink communication performance just explained, the energy consumption in an agricultural ED can include turning on booster pumps or solenoid valves, activating sensors over a long period, use of GPS and data transmission, etc. Nodes should be autonomous as much as possible since usually power supplies are not available in a wide agricultural area. Besides, the use of batteries needs to minimize disposal costs and pollution. Providing solutions to avoid the use of batteries by harvesting energy from the environment would encourage the deployment of wireless devices in Smart Agriculture. The use of different energy sources, such as solar energy, piezoelectricity, thermal, wind, water, and radiofrequency is consolidated [107]. However, making a device completely energy-neutral requires a thorough analysis of power consumption in different working states [108]. One facilitation is the availability of a renewable energy source as in [3]; on the other hand, a high energy consumption due to the heating of one probe as in [87] requires a different design or the remote monitoring of the energy available or harvested as in [109]. It is evident that there are many factors that influence the analysis of offering-demanding energy, it varies on a case-by-case basis and does not lend itself to systematic analysis; on the other hand in this context, machine learning algorithms can give a significant contribution. Infact, the ML approach has been already successfully applied in different contexts allowing to implement an efficient renewable energy selection based on the geographic location [110], or to retrieve a good energy prediction [111]. An application example is given by the energy-neutral system for pest detection [80] which takes advantage of ML algorithms.

1.7.3 Heterogeneity and interoperability

Smart Agriculture systems are quite heterogeneous in terms of sensors and, in some cases, it is also required to integrate different communication technologies, e.g. when

multiple platforms coexist and data arrives from different sub-systems. LoRa platforms are used with Zigbee to implement hybrid communications managing different sensors clusters or wuth IEEE 802.11s-based system to build a mesh networking architecture. The path for the integration of different technologies like cloud, IoT and Software-defined networking, with artificial intelligence (AI) is proposed in [112] with the related challenges and opportunities. Assuring communications in heterogeneous Smart Agriculture systems is a critical issue that has been studied for example in [113], where LoRa and Zigbee hybrid communications are implemented. Precisely, two LoRa sensor clusters and two Zigbee sensor clusters are used and combined with two Zigbee-to-LoRa converters to communicate in a network managed by a LoRa GW. The token ring protocol in the Zigbee network and polling mechanism in the LoRa network is used. The system can work with a packet loss rate of less than 0.5% when the communication distance is 630 m for the Zigbee network and 3.7 km for the LoRa network.

An hybrid LoRa/IEEE 802.11s-based mesh networking architecture is proposed in [114], where an effective network protocol selection mechanism is developed to choose the right interface. Protocol selection is based on multiple parameters, including network communication interface type, GNSS position of the APs, RSSI of nearby nodes, type and amount of data to be transmitted. Large data to be transferred in a short time can rely on the IEEE 802.11s-based network while small data can be transmitted through a LoRa-based mesh network.

Platforms such as FIWARE [115], Cayenne [44] and mySense [45], discussed in section 1.5, can also give a push to achieve interoperability in the Smart Agriculture systems. The above described solutions can benefit of an "industry 4.0"-based approach where the integration of different protocols cooperate to address the needs of automating computing and technology processes [112, 21].

1.7.4 Machine learning and big data management

The integration of big data analysis with Machine Learning can provide predictions about future outcomes, such as fruit quality or detect crops' diseases using historical data, analytical techniques, and statistical modeling [116]. The benefits of ML in the agriculture are relevant [117]. However, the deployment of models is the most challenging step to bring the ML algorithms in the production fields, and thanks to its advantages LoRa technology could make a big contribution to taking this step. Collected data can be used to implement an intelligent system capable of supporting the identification of varieties and predicting the quality of the final product [4]. In

fact, exploiting ML, the data can be used by the biologists to develop crop models and perform disease prediction [78].

The agricultural industry produces a large amount of data collected by heterogeneous sensors, so best practices should include the mechanisms to reduce the memory and time for data analysis. Thus, to pursue such objectives, Edge Computing models are also applied [118].

Distributed data process, such as MapReduce [119], may avoid bottlenecks when transferring all data to a single server, as in [120] where the proposed method adopts smart sensors to measure the soil quality indicators, while the pre-elaborated data is transmitted using the LoRaWAN protocol. The Apache Spark environment is then used to implement a parallel algorithm for statistical models based on the soil indicator data obtained from the experimental field.

1.7.5 LoRa scalability and network improvement

As concerns scalability, some open points, shared with general applications, are recognized. For this reason, most of the reference literature does not directly address issues related to agriculture. Indeed, scalability is a key feature in LoRa networks due to its long-range and large number of devices can concurrently reach a given GW. The network scales quite well if dynamic transmission parameters are used, in combination with multiple sinks. However, the correct behavior of the network server is not easy to be evaluated [121]. In fact, the network server presents some challenges from the point of view of its optimization, such as processing duplicate packets or packets from other networks, or bringing down the entire network in case of internet connection loss.

LoRa networks are bound by strict legal requirements, particularly where no listen-before-talk schemes are utilized. Transmission Duty Cycle (TDC) regulates the ISM bands to determine the maximum time that the band can be occupied, typically bounded to 1%. This implies that devices may not occupy the ISM band for more than 36 seconds per hour, forbidding the transmission of new packets when this limit is attained [122]. Machine learning can be applied to model and analyze technical problems, improving the scalability of LoRa networks and predicting network congestion [123]. Further developments could include enhanced ADR mechanisms, optimization of GW locations, and interference cancellation techniques [124].

Finally, some challenges remain such as the widespread adoption of multihop communications in LoRaWAN. Literature has shown that multihop or mesh topologies can extend the coverage of LoRaWAN networks and improve energy efficiency in certain

scenarios [125]. These solutions propose intermediate nodes to forward messages to other EDs to extend the coverage. Other open points include the use of GWs as intermediate nodes, GW-to-GW communications, and practical large-scale deployment of LoRaWAN mesh networks.

1.7.6 Other Communication Technologies

The choice of a specific communication technology is central to the performance of IoT-based agricultural applications. Other than LoRa, many standards for wireless communications can be employed, including Bluetooth, ZigBee, Z-Wave, RFID, Sigfox, NB-IoT. Some of them work well in the short-range (within 100 m), while others are more useful to cover long distances (up to tens of kilometers). Examples of the former are Bluetooth, ZigBee, Z-Wave, passive and active RFID systems, while in the latter standards are Sigfox and NB-IoT (and LoRa of course). As discussed previously, the deployment of a massive number of IoT devices might cause interference problems especially for technologies using the unlicensed spectrum, such as ZigBee, Wi-Fi, Sigfox, and LoRa. On the other hand, IoT devices operating with licensed spectrum eliminate interference problems but might increase costs significantly.

Several papers have analyzed different aspects of wireless communication protocols for smart agriculture, studying possible applications and comparing their performance. For example, ZigBee-based smart agriculture systems are described in [126–130]. The biggest challenges for ZigBee networks are the limited range and increased power consumption (compared to LPWANs) and relatively low data rate (e.g. compared to BLE or WiFi). Therefore, ZigBee is better suited for small-scale scenarios [129], while the use of this protocol is not suitable when the agricultural area is vast and the distance between sensor nodes is large. On the other hand, the works [131–134] represent successful examples of NB-IoT applications in smart agriculture. Indeed, extensive coverage, adaptable power consumption (depending on the mode of operation), and low interference among nodes, are features that make NB-IoT an interesting protocol for various agricultural systems [135]. However, NB-IoT employs licensed frequency channels, which results in higher subscription prices for the associated system even if it offers a higher data throughput than LoRa. Moreover, when there is an existing LTE infrastructure already in place, the need for hardware update may be another source of expense for such a system. This might be a drawback in the context of smart agriculture if the projected return on investment is not high enough to cover these costs [136].

Overall, the choice of the communication technology in smart agriculture needs to consider many factors and requirements, such as support for roaming, suitability of technology to small-scale, medium-scale and large-scale deployments, geographical location, costs, etc. For example, it has been shown that Sigfox and LoRaWAN excel on network capacity, battery lifetime and cost, whereas NB-IoT achieves higher quality of service and lower latency [30]. Finally, while LoRaWAN has been considered the most suitable communication network for IoT in smart agriculture [19], it is still difficult to tell which technology will dominate the market, or if several technologies will coexist, perhaps specializing on different application domains.

1.8 Summary

Although the expected transition to Smart Agriculture has already begun, researchers around the world are still looking for new solutions to improve agricultural productivity through IoT architectures. Indeed, albeit applications in agriculture can benefit from the experience gained in Industry 4.0, they require specific knowledge regarding sensor management, energy optimization, and data processing. LoRa technology is widely adopted, as it allows building an autonomous network that meets some of the requirements of the Smart Agriculture, such as low power and long-range communication. The adoption of LoRa-based systems in agriculture results in an effective way to improve the connectivity of farms, encourage the deployment of decision support systems and consequently improve their management, leading the agricultural sector towards Smart Agriculture. In order to provide a more focused and comprehensive view of the applications in the field, in this chapter we restricted our focus to LoRa/LoRaWAN technology and its uses in the context of Smart Agriculture. We presented many LoRa applications in the field, and we discussed some open issues and research areas for future improvements. The main challenges analyzed using LoRa Technology in Smart Agriculture, are: latency on the downlink channel, energy management, heterogeneity and interoperability of the devices, data management and scalability. All of these can benefit from the use of machine learning algorithms. Indeed, Artificial Intelligence and Edge Computing are still scarcely used but related algorithms and technologies are now mature and may be successfully applied in this field. Finally, optimization of multiple GW locations and multihop topologies to extend the coverage of LoRa networks have been recently tested to further improve performance and coverage.

Chapter 2

Comparative Analysis of LoRaWAN and Sigfox Coexistence in Rural and Urban IoT Networks

2.1 Overview

According to IoT Analytics, NB-IoT, LoRaWAN, and Sigfox are today the most popular technologies for low-power wide-area networks (86% of the market), both in terms of end-user adoption as well as ecosystem support. While NB-IoT utilizes licensed bands, LoRaWAN and Sigfox both employ the sub-GHz ISM bands, potentially interfering with each other. In this chapter, an in-depth study is presented on the coexistence between LoRaWAN and Sigfox under realistic urban and rural scenarios with different duty cycles and traffic conditions. The choice of such scenarios and simulation parameters is supported by a thorough literature review. Additionally, a mathematical model for interference assessment is proposed, and the results are quantified using both the developed model and the SEAMCAT simulator. The results offer new insights on the coexistence of LoRaWAN and Sigfox for emerging IoT applications like smart agriculture and urban vertical farming. Finally, as interference mitigation strategy, we analyze the performance obtained applying protection distance mechanisms.

2.2 Introduction

The Internet of Things (IoT) is a global network of devices gathering information from systems and environments, and interacting with each other. In this field, Low-Power Wide-Area Network (LPWAN) technologies have emerged as a viable alternative to traditional wireless technologies to provide power-efficient and cost-effective wide area connectivity for the IoT. Farther, from the application side, more focus is covered on smart agriculture applications. The advantages of the LPWAN architectures include a wide coverage in the order of kilometers, and a low power consumption, with batteries lasting up to 10 years. According to IoT Analytics, the number of connected IoT devices is forecasted to grow from 7 billions in 2018 to 22 billions in 2025 [137]. Moreover, according to the same study, three LPWAN technologies, namely LoRaWAN [23], Sigfox [138], NB-IoT [139], are used to connect about 86% of the devices. While NB-IoT works in licensed bands, LoRaWAN and Sigfox share the same ISM bands. Since both technologies employ an Aloha access protocol, the proliferation of such devices can produce coexistence issues.

It is interesting to note that LoRaWAN and Sigfox employ very different modulation schemes: while Sigfox utilizes an Ultra-NarrowBand (UNB) 100 Hz signal and Binary Phase Shift Keying (BPSK) modulation, LoRaWAN exploits a 125-500 kHz Chirp Spread Spectrum (CSS) modulation. In the rest of the chapter, we refer to LoRa/LoRaWAN as CSS and Sigfox as UNB and vice-versa. Given the different modulation employed, it is thus important to evaluate the coexistence of these technologies in realistic application scenarios, e.g. digital agriculture, indoor urban vertical farming, smart metering, etc. Although some previous works have analyzed the interference among these technologies, e.g. [140, 141], these works either employ arbitrary simulation settings or fail to provide insights on the achievable performance in large-scale scenarios. In contrast, the present work takes into account realistic node deployments, such as the ones in [142], considering variable node densities and traffic conditions, different deployments (e.g. indoor/outdoor, elevation, etc.), intra- and inter-technology interference, as well as possible interference mitigation strategies. In particular, in this work we study the interference between the CSS and UNB systems in urban and rural scenarios, focusing especially on the uplink which is the most critical. Accurate path-loss models are employed, and realistic traffic scenarios (derived from a thorough literature review) are analyzed, considering node density distributions, Duty Cycle (DC), and modulation parameters typical of the considered technologies. We measure the interference probability, resulting from collisions of the

same technology (self-interference), from the other technology (cross-interference), or both (mixed-interference) for CSS and UNB systems. This chapter also introduces a theoretical model for evaluating interference in LPWANs, focusing on a circular cell with a victim receiver in the center. The importance of accurate calculation of offered traffic is discussed, and an attempt is made to define a critical distance beyond which interference becomes negligible. In addition, the model examines the distribution of distances of interfering nodes and analyzes the probability of interference based on the number of active nodes and the minimum distance between interferer and victim.

The interference results were quantified using both our mathematical model and the SEAMCAT simulator [143], which represents a novel approach for IoT technology. Finally, the performance obtained by applying protection distance to the victim gateways is analyzed to mitigate the impact of interference.

The main contributions of this chapter are:

1. the path-loss models that best fit the considered technologies (including both indoor and outdoor settings) for urban and rural agriculture scenarios are studied.

2. we introduce a theoretical model for evaluating interference in LPWAN technologies;

3. we configure SEAMCAT for realistic UNB and CSS deployments and offer all simulation files to the scientific community;

4. we analyze the interference probability in several traffic scenarios, showing the impact of self, cross and mixed interference for both configured systems;

5. we discuss possible interference mitigation strategies, focusing on the performance obtained with protection distance mechanisms.

The remainder of the chapter is organized as follows. A detailed analysis of the state of the art can be found in Sec. 2.3. Sec. 2.4, briefly presents the selected technologies and their interference. Sec. 2.6 recalls the SEAMCAT architecture and details the path-loss model employed. Sec. 2.6.1 introduce a theoretical model to examine the interference of LPWAN networks. In Sec. 2.7 we describe the experiment scenarios and present the results of our experiments. Sec. 2.8 discusses possible interference mitigation schemes and analyzes the performance obtained with protection distance. Finally, Sec. 2.9 concludes the chapter.

2.3 Related Work and Motivation

In this section, the motivation behind this study is presented, with a focus on mass-scale IoT for agriculture, monitoring in indoor vertical farming, and monitoring in rural scenarios. Additionally, relevant literature studies regarding the coexistence of LoRa and Sigfox, as well as other related works employing SEAMCAT for interference studies, are examined.

2.3.1 Motivation

The LoRaWAN and Sigfox technologies are widely used in agricultural contexts and can coexist without mutually influencing each other, when the number of sensors deployed in the field is relatively low, as observed in reference [15]. However, the deployment of massive IoT sensors for agricultural and other purposes will cause interference problems, especially with IoT devices that use unlicensed spectrum, such as ZigBee, Wi-Fi, Sigfox, and LoRa [9]. The interference can lead to data loss and reduce the reliability of the IoT agro-ecosystem. Imagining agricultural scenarios where sensors are deployed on a large scale, such as one sensor per plant, or even in tree crop contexts where it could be beneficial to have at least three sensors per plant, one in the soil to monitor moisture in the root zone, one in the trunk to assess sap flow, and another in the leaves to measure turgor, might highlight the need for carefully planned wireless network deployment. In a hypothetical future of precision agriculture, examining the coexistence and potential interference among transmission technologies sharing the same frequency bands could prove useful. Thorough analysis and accurate research can help mitigate any interference-related issues, providing tailored solutions and technical requirements of each agricultural context in both rural and urban environments. In fact, considering the growth forecasts of the world population, which will reach 9 billion by 2050, of which 70% will live in urban centers, the need arises for vertical agriculture. Vertical agriculture allows the urban cultivation of fruits, vegetables, and cereals within the buildings of a city or urban centers [144].

2.3.2 Coexistence studies on Sigfox and LoRa

One of the first works evaluating the differences between CSS and UNB networks is [140], where both physical level and packet level performance are considered. The authors show that an UNB network has a larger coverage, while CSS-based networks are less sensitive to interference. Moreover, they analyze the case in which both technolo-

gies coexist and show that active rate and frequency management are fundamental for improving performance. Similarly, in [141] the impact of different types of ISM interference on CSS modulation is evaluated and is demonstrated that narrowband interference suppression can significantly improve the resilience of LoRa. However, these works either employ unrealistic scenarios (i.e. with an arbitrary number of nodes and without considering indoor/outdoor settings), or fail to provide insights on the achievable performance in large-scale scenarios.

Authors in [145] experimentally evaluate the coexistence of LPWAN technologies in the ISM band. The results show a significant impact on the performance of LoRa and Sigfox from already installed IoT devices in smart homes, business parks, smart agriculture, etc. The work in [146] exploits a testbed setup to evaluate the impact of different sub-GHz technologies (Sigfox, Z-wave, and IO Home Control) on LoRa. Results show that there is a significant loss under Sigfox interference, especially when the interferer starts during the preamble and header time while loss is reduced if the interferer starts during the payload time. However, LoRa is only seen as the victim and the opposite is not considered. In [147, 148] performance limits of LoRa have been discussed, and in [149] the impact of interference among LoRa nodes is analysed experimentally. However, these works do not consider the interference that LoRa and Sigfox might suffer when deployed in the same environments. Finally, [150] surveys the challenges of LPWANs and points out their design objectives and limits, analyzing coverage, energy efficiency, scalability, interference management, and quality of service. However, all these existing works do not consider interference in mixed high-density scenarios where multiple IoT technologies coexist in the same area and use the same ISM bands. Moreover, in this chapter, we analyze the performance obtained with protection distance, as a possible interference mitigation mechanism.

2.3.3 SEAMCAT-based interference studies

The SEAMCAT simulator has been used in several previous works. For example, authors in [151] analyzed the interference between Wi-Fi HaLow and Long Term Evolution (LTE) User Equipment (UE). The study employs both a theoretical method based on Minimum Coupling Loss and statistical MC simulations exploiting SEAMCAT. In [152] the authors studied interference of three industrial IoT technologies (WirelessHART, WiFi and Bluetooth) with a 5G uplink Macro-cell. All simulation scenarios are composed of one victim technology, and one or more interfering systems in turn, with and without the presence of 5G uplink traffic. A coexistence work between LoRaWAN and SRDs is proposed in [153], where SEAMCAT is employed

to measure the interference probability. In [154, 155] authors adopt the SEAMCAT Extended-Hata propagation model to analyze LoRa performance in urban scenarios and for coexistence between IEEE802.11ah and 802.15.4g, respectively. The above literature confirms that SEAMCAT is a powerful tool to conduct this study, and to the best of our knowledge, no previous work studied interference between Sigfox and LoRa technologies using SEAMCAT.

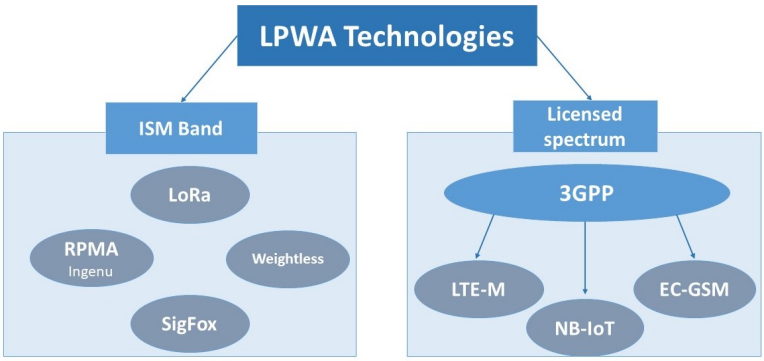


Figure 2.1: Classification of LPWAN technologies.

2.4 LPWAN Technologies and their interference

Figure 2.1 shows the different LP-WAN technologies separated based on the frequency spectrum used, licensed or unlicensed (ISM). Among these, LoRa and Sigfox are the main emerging technologies [30]. The most prominent requirements for serving remote IoT nodes are long-range and low-power connections. Consequently, this chapter focuses on LPWANs which are tailored to provide wide-area communication to powerconstrained devices. In this chapter we will focus of the latter cluster and in particular to the most prominent technologies in that cluster LoRaWAN and Sigfox. These technologies can easily support the transmission of long-range data packets, minimizing power consumption, design complexity, and therefore cost. Furthermore, new cellular communication modes and terminal categories are defined, e.g., NB-IoT, for Machine-Type Communication and IoT applications. NB-IoT can operate both in subGHz and conventional cellular frequency bands. Notably, other technologies exist but are less adopted by the LPWAN vendors or are focused on higher throughput communication. For instance, two other cellular technologies are designed to support machine-type communication. EC-GSM extends the coverage of legacy Global System for Mobile Communications (GSM) but is less or not adopted by network vendors. In addition to NB-IoT, another technology based on Long Term Evolution (LTE) is

specified, i.e., LTE-M. This standard, while less adopted than NB-IoT, also targets a different market, i.e., Machineto-Machine (M2M). Furthermore, other technologies tailored for LPWAN communication such as RPMA, NB-Fi, Weightless, Wi-Fi Halow, DASH7 have seen less of an adoption by the LPWAN market and are therefore not further considered in this work.

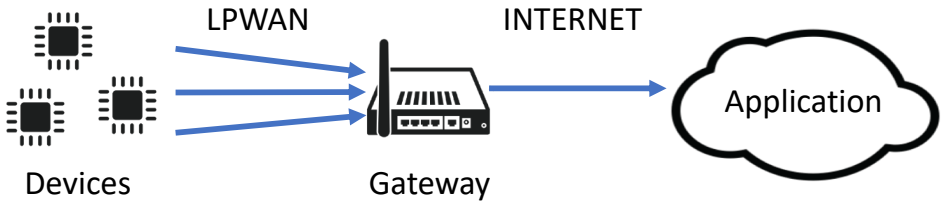


Figure 2.2: LPWAN technology architecture.

As shown in Fig. 2.2, a typical LPWAN architecture comprises: i) IoT devices providing sensing, measurement, and/or control functionalities in a large geographic area; ii) wireless connection to a gateway (GW) or base station; and iii) a server or cloud platform, which processes, stores, and keeps available data for visualization and usage. To extend the communication range, often LPWAN technologies i) employ low frequency bands with favorable propagation characteristics, ii) apply modulation schemes which are robust to low SNR (Signal to Noise Ratio), iii) adopt simple MAC schemes to reduce power consumption.

In LPWANs, communication is mostly device-initiated, meaning that an Up-Link (UL) message is transmitted when the device has some data to send. For the remainder of the time, the device is kept in sleep mode, while Down-Link (DL) traffic is generally permitted only after a UL transmission for a tunable window time. Among the different LPWANs, LoRaWAN is an open-source solution promoted by the LoRa Alliance. LoRaWAN uses Long Range (LoRa) modulation, patented by Semtech, based on CSS [23]. This technique encodes a set of bits into a single chirp. A chirp is a sinusoidal signal where the frequency is linearly increased or decreased in time over a fixed bandwidth. The Spreading Factor (SF) determines the duration of the chirps, and defines two modulation features: i) the time duration of each symbol; and ii) the number of raw bits encoded by that symbol, equal to SF. In LoRaWAN, the available Data Rates (DRs) are determined by the SF, bandwidth, and number of redundancy bits (Coding Rate). The bandwidth can be 125 kHz or 250 kHz (typically 125 kHz is used in the 868 MHz ISM band). Finally, LoRaWAN defines six different SFs (from SF7 to SF12), which result in different symbol times and almost orthogonal transmissions. The range can be extended in LoRaWAN by increasing the SF, thereby lowering the demodulation threshold. For example, with SF12 a signal with an SNR of -21.9 dB

can be demodulated, while for SF7 a minimum SNR of -8 dB is required in order to receive the packet. Any time that a new packet is ready for transmission, devices attempt to transmit randomly selecting one of the available channels, as detailed later.

Sigfox is also a proprietary technology, where GWs and network servers are managed directly by Sigfox or other national operators authorized by Sigfox. Devices are certified by Sigfox before joining the network. In contrast to LoRaWAN, Sigfox uses a UNB modulation scheme: data is modulated by Differential Binary Phase Shift Keying (DBPSK) at 600 bps, generating a 100 Hz signal. The advantage of such a method is that the noise level present in a narrow-band communication channel is minimal.

For redundancy, data is transmitted 3 times on different channels (to ensure frequency diversity) at different consecutive time intervals. There are 1920 channels available for the UL, and the same number for the DL. Moreover, together with time and frequency diversity, determined by the frequency hopping mechanism, Sigfox also uses spatial diversity since the network is designed so that every node is in the range of at least 3 different GWs [156]. As discussed previously, LoRaWAN and Sigfox both exploit the same ISM bands. In Europe, in particular, the band between 863 MHz and 870 MHz is allocated for license-free communications [157]. As depicted in Fig. 2.3, five LoRaWAN UL and DL channels are accommodated between 865.0 MHz and 868.0 MHz, while in the second sub-band (from 868.0 MHz to 868.6 MHz) are allocated 1920 Sigfox UL channels and 3 UL/DL LoRaWAN channels. In this band, about 200 kHz are used by Sigfox ($1920 \times 100\,\text{Hz} = 192\,\text{kHz}$), completely overlapping the 125 kHz LoRaWAN channel centered on 868.1 MHz. Both sub-bands 1 and 2 (see Fig. 2.3) are limited to 25 mW (14 dBm) transmission power and 1% DC (basically, this means 36 s of Time on Air (ToA) per hour).

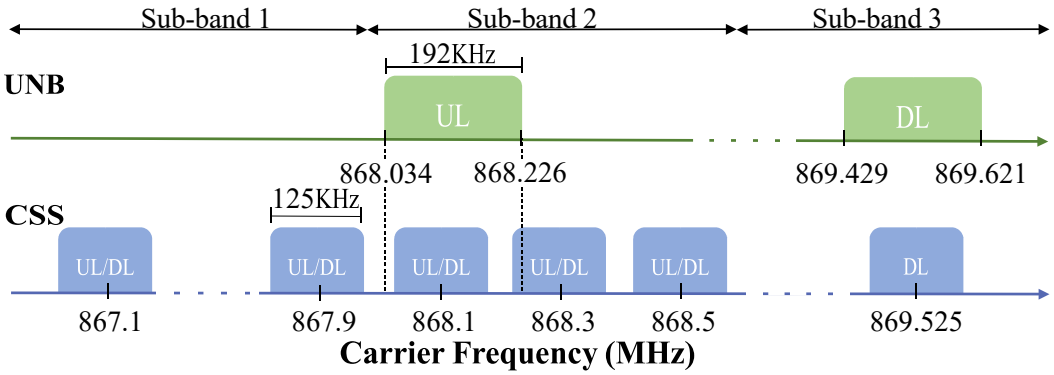


Figure 2.3: LoRaWAN and Sigfox channels overlapping in the 868 MHz bands.

Sub-band 3 is interesting because the maximum transmission power is to 500 mW with a 10% DC. While an end device would not be able to transmit such power running

on batteries, a GW attached to the grid could use is it for long range DL communications. Moreover, the larger DC allows the GW to communicate with many nodes. Nevertheless, LPWAN technologies are mainly employed for UL communications and the density of the GWs is usually less critical, so we omit this DL scenario in our analysis.

2.5 A large-scale model for interference evaluation

In this section, we introduce a theoretical model to examine the interference of LPWAN networks characterized by a circular cell with a victim receiver positioned at the center and constant reception power. Identifying and analyzing the circular area from which significant interference emanates is the primary objective of this study. Initially, this region may appear indefinite, but we aim to establish the critical distance beyond which the influence of interference becomes negligible from a practical standpoint. A maximum radius (R_{max}) can be considered within which the average interfering power is exceedingly low and practically negligible. Consequently, interference generated by nodes deployed at distances greater than R_{max} can safely be disregarded. Therefore, to accurately calculate the offered traffic (A), it is imperative to have knowledge the value of the maximum radius, as is shown in formula 2.1.

$$A = \delta \cdot R_{\max}^2 \cdot \pi \cdot \mathrm{DC}_{\mathrm{eq}} \cdot \frac{0.01}{n_{ch}} \tag{2.1}$$

where δ is the density of the nodes, DC_{eq} is the equivalent duty cycle in terms of fraction of 0.01 legacy limits, and n_{ch} is the number of channels, which is equal to 8 for CSS and 3 for UNB. Generally, increased R_{max} value result in increased traffic offered. However, once a certain point is reached, this increase is limited by the minimum distance at which the last interfering device is placed. Additionally, it could happen that multiple nodes are active simultaneously, so in the case of overlap, it is necessary to consider the worst-case interference. This means that the probability of having a certain number of active nodes simultaneously needs to be evaluated. When the number of nodes is high, the probability of having a certain number of active nodes (K) can be approximated using the following Poisson equation.

$$\Pr[k] = \frac{e^{-A} \cdot A^k}{k!} \tag{2.2}$$

According to the equation 2.2 it can be affirmed that the bigger the cell, the higher the load, the higher the most probable number of active nodes at the same time.

2.5.1 Distance of the interfering nodes

We will show that in the case of several active nodes, the contribution to interference is only made by the node closest to the victim and that the distance between node and victim does not depend on the size of the cell. In case of single interferer, assuming it is uniformly spread into the cell, the distance of the nodes is linearly distributed between 0 and R_{max}. By normalizing the distance d in as a fraction of R_{max}, i.e. $x = d/R_{max}$, the probability density function $f(x)$ is:

$$f(x) = 2x \quad \text{with} \quad x \in [0, 1] \tag{2.3}$$

In this case, it is necessary to know the distance of the main interferer, because when two or more nodes are simultaneously active, we will be interested in the distribution of the nearest node, neglecting interference contributions from greater distances. To do this we need to know the cumulative function (cdf) of $f(x)$, i.e.

$$F(x) = x^2 \quad \text{with} \quad x \in [0, 1] \tag{2.4}$$

Whatever the form of $F(x)$, the minimum of k random variables that are linearly distributed in [0,1] can be written as follows

$$1 - (1 - F(x))^k = 1 - (1 - x^2)^k \tag{2.5}$$

It follows that the density function is $2kx(1-x^2)^{k-1}$. We obtain a density as a function of k, from which we can derive the most probable value (expected value), and as k increases, the distance approaches the minimum distance of the nearest interferer. We remember that k is the number of devices transmitting simultaneously, and it can range from one to infinity according to the Poisson distribution. Consequently, if we have a large number of simultaneously active devices, the value obtained becomes more and more probable. The same thing happens when the we have bigger R_{max}, the higher the load and consequently the closest to the victim is the value of the most probable normalized distance.

2.5.2 From interference distance to the interference probability

Assuming we know the minimum distance between the interferer and the victim, the number k of simultaneously active nodes, and we include log-normal fading, we can

calculate the probability of interference as follows.

$$P[\Pr(x_{TN}) > I_{th}|x] = \begin{cases} 0.5 \cdot \operatorname{erfc}\left(\frac{|P_{av}(x)-I_{th}|}{\sqrt{2}\sigma}\right), P_{av}(x) < I_{th} \\ 1 - 0.5 \cdot \operatorname{erfc}\left(\frac{|P_{av}(x)-I_{th}|}{\sqrt{2}\sigma}\right), P_{av}(x) > I_{th} \end{cases} \tag{2.6}$$

If we know the minimum distance, then we know the average power ($P_{av}(x)$) at that distance, so we can calculate the probability of going below or above the threshold (I_{th}). Lastly, it is necessary to integrate over all possible distances x, as shown in equation 2.7.

$$\Pr\left[\Pr(x_{TN}) > I_{th} \mid k\right] = \int_0^1 \Pr\left[Prx_{TN}(x) > I_{th} \mid x\right] \cdot 2k \cdot x \cdot (1-x^2)^{k-1} dx \tag{2.7}$$

And integrating on k becomes:

$$[\Pr x_{TN} > I_{th}] = \sum_{k=1}^{\infty} \frac{e^{-A} A^k}{k!} \Pr\left[\Pr x_{TN} > I_{th} \mid k\right] \tag{2.8}$$

It can be shown that as R_{max} increases, this probability converges to a limit value. Because the load will increase and with it the probability of finding more active nodes, but the average probability given by k will decrease, the cell being larger. Finally, we tested the model for evaluating interference in urban environments, considering scenarios of self-interference and cross-interference, and compared them with SEAMCAT simulations, as reported in the following section.

2.6 The SEAMCAT simulator and path-loss tuning

In this analysis, SEAMCAT (version 5.4.2) is employed (version 5.4.2), a complex statistical simulator based on the MC method, devised to assess the interference between different radio communication technologies. SEAMCAT is developed by CEPT/ECC Working Group Spectrum Engineering (WGSE) within its sub-entity SEAMCAT Technical Group (STG). The simulator is based on the definition of a victim link, characterized by a transmitter and a receiver of a given technology, as well as one or more interfering links (including different technologies). For each technology, it is possible to specify several physical parameters of the node, including the propagation model, location (e.g. indoor/outdoor, height), antenna radiation diagram, transmission power (including emission mask), receiver blocking mask, etc., and any of these parameters can have a statistical distribution among nodes.

The evaluation of the interference probability is performed by averaging the results of multiple simulated events. Indeed, following the statistical distributions of the physical parameters, for each event, the impact of interference is computed by comparing the signal strength of the victim link with the sum of the interfering signals, filtered by the transceiver power masks (including also adjacent channels). Thus, for each event the power received by the victim is computed taking into account both the transmissions power and the relative path-loss (PL), evaluated by considering the propagation model, and the environment parameters (e.g. position, height, etc.).

An event is flagged as interfered (or good) when the Carrier to Interference ratio (C/I) value is lower (or higher) than a target threshold. One of the main characteristics is thus related to the accurate computation of the PL.

In the rest of this section, two objectives will be pursued: first, path loss (PL) in LoRaWAN and Sigfox technologies in the 868 MHz bandwidth will be investigated, and second, the model used for the simulation will be corrected. Based on a thorough study of the literature and real-life experimental evaluation, a correction to the path loss model available in SEAMCAT will be proposed.

2.6.1 The propagation model

Generally, propagation models include empirical models. In empirical models, the parameter values are derived by fitting measurement data to an appropriate function for a particular environment. This gives a more generic model that can be used by systems operating in similar areas. The PL is the different value between the radiated power (Pt) and the received power (Pr).

$$PL(dB) = Pt - Pr \tag{2.9}$$

The Log-distance PL Model, also referred to as the one slope model, is a general PL model that has been used in a large number of indoor and outdoor environments. It assumes that PL varies exponentially with distance according to the following equation:

$$PL(d)[dB] = 10n \log_{10}(\frac{d}{d_0}) + PL_0 + X_\sigma \tag{2.10}$$

where n is the PL exponent, d is the distance between the transmitter and the receiver, and PL_0 is the PL at a reference distance d_0. Shadow fading is represented by a zero-mean Gaussian random variable X_σ with standard deviation σ (in dB). In order to characterize the indoor PL, the most diffuse approach is to consider additional attenuation incurred by walls and floors. Thus, PL is modeled as:

$$PL(d)[dB] = 10n\log_{10}(\frac{d}{d_0}) + PL_0 + X_\sigma + W_{AF} + F_{AF} \tag{2.11}$$

where W_{AF} and F_{AF} are wall and floor attenuation factors based on the number of traversed walls and floors between the transmitter and the receiver. Examples of models that take into account attenuation of wall and floor are Cost 231 [158] and the ITU-R model [159]. For the our simulation scope, we select EXTENDED HATA, that according with presented literature, best fit the experimental results. The Hata Model for Urban Areas (also known as the Okumura-Hata model), is a widely used propagation model for predicting path loss in urban areas. This model takes into account the effects of diffraction, reflection and scattering caused by city structures. The model also has formulations for predicting path loss in Suburban and Open Areas.

Table 2.1: Path-loss comparison for Sigfox and LoRa technologies in urban scenarios

Ref.	Technology	Location	Scenario	d_0 (m)	PL(d_0) (dB)	η	PL(100m) (dB)
[160]	LoRa	Sydney, Australia	indoor	1	36.5	4	140.4
[121]	LoRa	Lancaster, United Kingdom	indoor	40	127.41	2.08	135.7
[143]	LoRa/SigFox	SEAMCAT, Extended Hata	indoor	-	-	-	114.5
[161]	LoRa	Beirut, Lebanon	outdoor	1000	140.7	3.12	109.5
[162]	LoRa	Oulu, Finland	outdoor	1000	128.95	2.32	105.8
[163]	LoRa	Dortmund, Germany	outdoor	1000	132.25	2.65	105.8
[164]	LoRa	Delft, The Netherlands	outdoor	1	23.9	3.89	101.7
[165]	LoRa	Dakar peninsula, Senegal	outdoor	1000	133.6	2.8	105.6
[166]	LoRa	Bologna, Italy	outdoor	1	31	3.84	107.8
[167]	Sigfox	Brno, Czech Republic	outdoor	100	118.04	3.76	118.0
[168]	Sigfox	Prague, Czech Republic	outdoor	1000	121	1.68	104.2
[169]	Sigfox	Belfast, United Kingdom	outdoor	1000	139.8	3.2	107.8
[143]	LoRa/Sigfox	SEAMCAT, Extended Hata	outdoor	-	-	-	87.1

2.6.2 Path-Loss tuning

In this subsection, a comprehensive literature review is presented, focusing on papers featuring real-world experiment campaigns for the estimation of path loss (PL) in LPWAN (Low-Power Wide-Area Network) technologies. Table 2.1 reports a broad list of such studies in urban scenarios, detailing the experiment location, the attenuation exponent η, and the PL_0 received at the reference distance d_0 specified in the paper. The wall and floor attenuation are also described in indoor studies [160, 121], where internal walls attenuate about 4 dB and concrete walls up to 10-20 dB. To better

compare the values, we also provide the PL at a common reference distance of 100 m (last column in the table). We compared these values to the propagation model provided by SEAMCAT Extended Hata model (last row in the table). The difference between PL values extracted from SEAMCAT and the experimental ones is quite large: 23.56 dB for the indoor case (114.49 dB against 138.05 dB, in average) and 20.64 dB for the urban case (87.1ḋB versus 107.74 dB).

Therefore, to align SEAMCAT simulations with real-world measurements, in our experiments we reduce the transmitted power accordingly. Finally, we provide the configuration files (including node density, duty cycle, and height distribution of the nodes) to the scientific community for repeatability and future integration of the research [174]. A similar procedure was carried out for the rural model, considering the literature and experimental measures reported in Table 2.2. Specifically, two experimental campaigns were carried out. The first one was performed in an area adjacent to the campus of the University of Palermo, which includes an urban park. Figure 2.4 shows the map of the urban park with the relative locations of the GateWay (GW) and the two locations used for the experimental campaign (ED1 and ED2). The second measurement campaign was carried out in the rural area of Roccamena (Pa), Italy, mainly characterized by wheat, grape, olive, and almond crops on the Pomilla farm (2.5b).

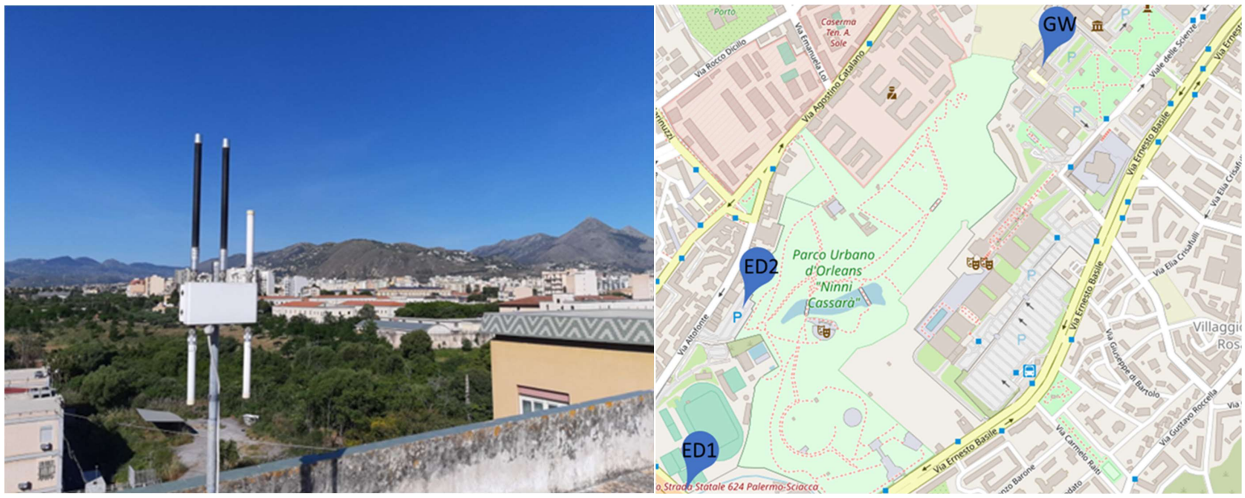
Based on the analysis reported in Table 2.2, a corrective factor of 12.03 dB was applied to the rural path loss model. This factor corresponds to the difference between the PL(100m) value extracted from SEAMCAT, which is 85.5 dB (as indicated in the last row of Table 2.2), and the average of the other values, which is 97.53 dB.

Table 2.2: Path-loss comparison for Sigfox and LoRa technologies in rural and suburban scenarios

Ref.	Technology	Location	Scenario	d_0 (m)	PL(d_0) (dB)	η	PL(100m) (dB)
[170]	LoRa	Perugia, Italy	suburban-hilly	1	33.39	3.73	107.99
This work	LoRa	Palermo (ED1), Italy	suburban	1160	119.81	2.5	93.20
This work	LoRa	Palermo (ED2), Italy	suburban	811	106.72	2.5	83.99
[171]	LoRa	West Lafayette, United States	rural	1107	128.1	2.5	102.00
[94]	LoRa/Sigfox	Simulation	rural	5000	107	2	93.02
[169]	Sigfox	Belfast, United Kingdom	rural	1000	121.4	2.78	79.2
[172]	LoRa	San Rossore Park, Pisa, Italy	rural	767	138.05	2.73	113.89
[173]	LoRa	Cuenca, Ecuador	rural	1000	133	3.2	101
This work	LoRa	Roccamena (Pa), Italy	rural	731	130	3.07	103.47
[143]	LoRa/Sigfox	SEAMCAT, Extended Hata	rural	-	-	-	85.5

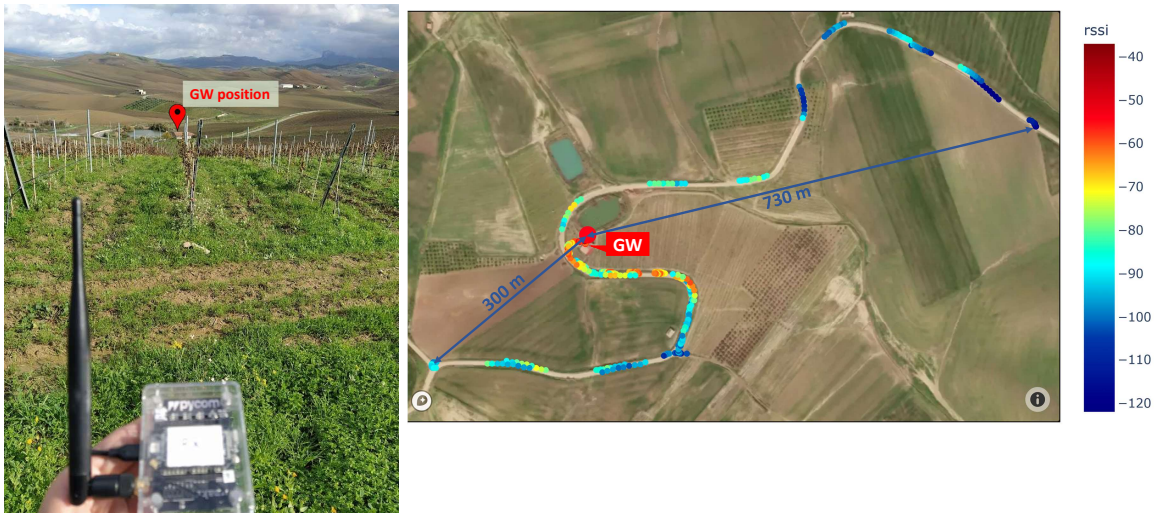
2.7 Simulation scenarios and results

In designing accurate simulation scenarios, important factors such as propagation environments, node density, transmission DC, elevation of the antennas, etc., were taken into account. In particular, we exploited the realistic distributions described in the *ETSI Technical Report 103 526* [142], with a statistical distribution of the device elevation of 1.5, 9, and 30 meters. Furthermore, we consider the urban scenario with 90% of the devices placed indoor and the remaining 10% deployed in an outdoor environment. Regarding the network density, we define a base scenario with 1000 devices/km^2, following [175], and we vary the receiver signal strength (RSS) of the victim technology from -130 dBm to -80 dBm. For the CSS network, we considered a setup with a distribution of different SFs, following *ETSI TR 103 526* (Fig. 6 in the report) [142] and *CEPT-SE24* [175], which describe the percentage of adopted SFs per number of gateways. Moreover, we exploited the results obtained in [153] to define a scenario with an average number of gateways equal to 0.5/Km^2. Since the LoRaWAN standard also includes the Adaptive Data Rate (ADR) algorithm which dynamically changes the SFs of the nodes, for CSS simulations we divided nodes in predefined groups each using a specific SF. In particular, unless specified otherwise, we use 65% of the devices working at SF7 and 35% at SF10. Indeed, according to the above studies, SF7 is the most significantly used, while SF10 has a ToA consumption



(a) (b)

Figure 2.4: Details of the experimental campaign conducted in a suburban setting near the Palermo University campus: (a) gateway position and (b) LoRa end-device placed at different positions (ED1, ED2).



(a) (b)

Figure 2.5: Details of the experimental campaign conducted in a rural area of Roccamena, Palermo: (a) test coverage in the Pomilla farm vineyard, (b) heat map of the Received Signal Strength Indicator (RSSI).

per day that is close to the weighted average of the ToA for all SFs from 8 to 12. Finally, regarding the UNB network, no particular tuning is required apart from the node distribution, assumed uniform.

Two different load conditions are considered: in the first (referred to as *Equal DC*), both UNB and CSS are configured to transmit at maximum duty cycle (1%). In this case, the data transmitted by CSS is greater than UNB: assuming a payload size of 12 bytes, 13,714 and 2,107 packets/day are sent using SF7 and SF10 respectively, compared to 144 UNB. In the second case (called *Equal data rate*) both networks transmit 144 packets/day, corresponding to a DC of 0.0105%, 0.068% and 1% for CSS SF7, CSS SF10 and UNB respectively. Since UNB re-transmits the same packet in three random channels for redundancy, considering a packet ToA of ≈ 2 seconds (for 12 bytes), the total channel occupancy becomes 6 seconds/packet. On the opposite side, we considered CSS with 13 bytes of overhead in addition to the 12 bytes payload.

These configurations are summarized in Table 2.3, while other parameters specific to UNB and CSS devices are reported in Table 2.4. In our simulation campaigns we consider the following scenarios: (i) *Urban self-interference*, when only one system is deployed; (ii) *Urban cross-interference*, evaluating the interference on a victim technology by a different one; (iii) *Urban mixed-interference*, when both self- and

Table 2.3: Configuration DC of the experiment scenarios

		Equal DC (1%)			**Equal data rate**		
	ToA (12 Bytes) (sec)	**packets/day**	**ToA/day (sec)**	**DC (%)**	**packets/day**	**ToA/day (sec)**	**DC (%)**
SF7	0.063	**13714**	863.94	**1.0**	**144**	9.07	**0.0105**
SF10	0.410	**2107**	863.91	**1.0**	**144**	59.04	**0.068**
UNB	6.0	**144**	864.00	**1.0**	**144**	864.00	**1.0**

cross-interference are present; and (iv) *Rural mixed-interference*, the same scenario as third one but in rural environments. Finally, we compare the mathematical model and simulations in the first two scenarios.

In the next subsections, we present the results for the defined scenarios, the simulation statistics are derived by configuring SEAMCAT to consider 50 000 simulation events. The simulation radius was set in relation to that typical of IoT systems, i.e., 1 km for urban scenarios and 5 km for rural scenarios.

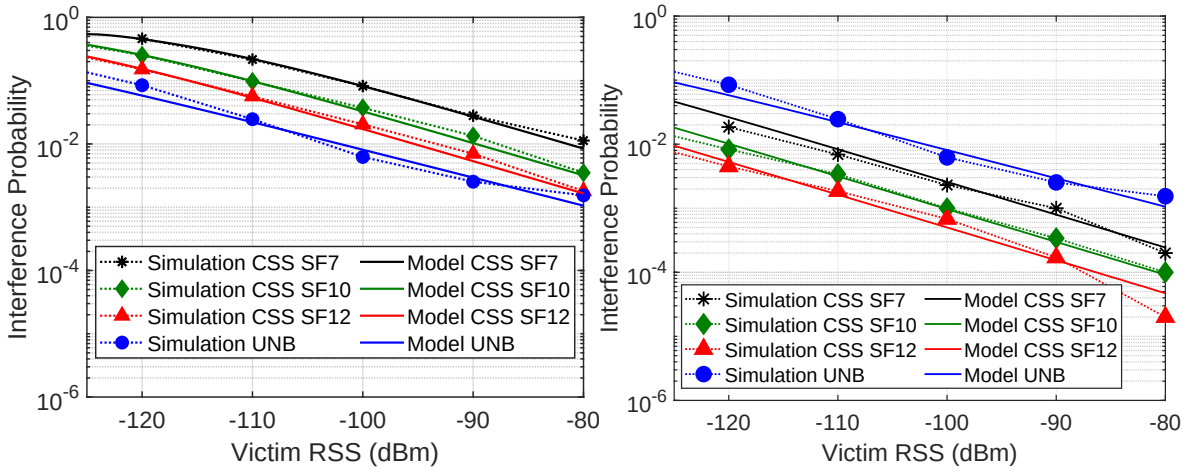
Table 2.4: Sigfox and LoRaWAN SEAMCAT systems parameters

Parameters	**UNB device**	**UNB GW**	**CSS device**		**CSS GW**		
TX PW(dBm EIRP)	16.25	29.25	16.25		29.25		
Ant. Gain(dBi omni)	0	5.15	0		5.15		
TX BW(Hz)	250	1K	125K		125K		
Adaptive Data Rate	N/A	N/A	Enabled		Enabled		
Power Control	N/A	N/A	Enabled		Enabled		
SF	-	-	**7**	**10**	**7**	**10**	**12**
RX Noise Figure (dB)	8	7	7	7	7	7	7
Req. SINR (dB)	7	7	-8	-16.3	-8	-16.3	-21.9
Sensitivity (dBm)	-124	-136	-124	-132	-124	-132	-137

2.7.1 Urban self-interference scenario

Fig. 2.6, shows the simulation results for the urban self-interference scenario, in the two load conditions discussed previously. In particular, Fig. 2.6a shows the results when the networks are loaded with equal DC, i.e. when nodes transmit with the maximum DC of 1%. The blue curve (circular marker) represents the urban self-interference probability for UNB technology, which is 0.21 in the worst case, when the received power of the victim is lowest. The other curves describe the CSS self-interference for three different SFs, namely SF7 (asterisks), SF10 (diamonds) and SF12 (triangles). Since, according to [142] (Table 10), the sensitivity of SF7 is -124 dBm, we do not evaluate the interference probability for RSS of -130 dBm, as no packet can be received with RSS lower than -124 dBm.

From the figure, the urban self-interference of CSS shows an interference probability about 20% higher compared to UNB. This difference depends on the modulation scheme and bandwidth efficiency (data rate divided by the bandwidth per hertz), whereas UNB is more robust than CSS modulation but also has lower bandwidth efficiency. Indeed, Fig. 2.6b shows the results in case of the equal data rate scenario, with a traffic load of 144 packets/day, and in this case, CSS experiences a significantly reduced urban self-interference probability compared to UNB. In particular, the results show that the urban self-interference of the CSS systems is around one order of magnitude lower than UNB, due to the much lower ToA of the CSS packets.

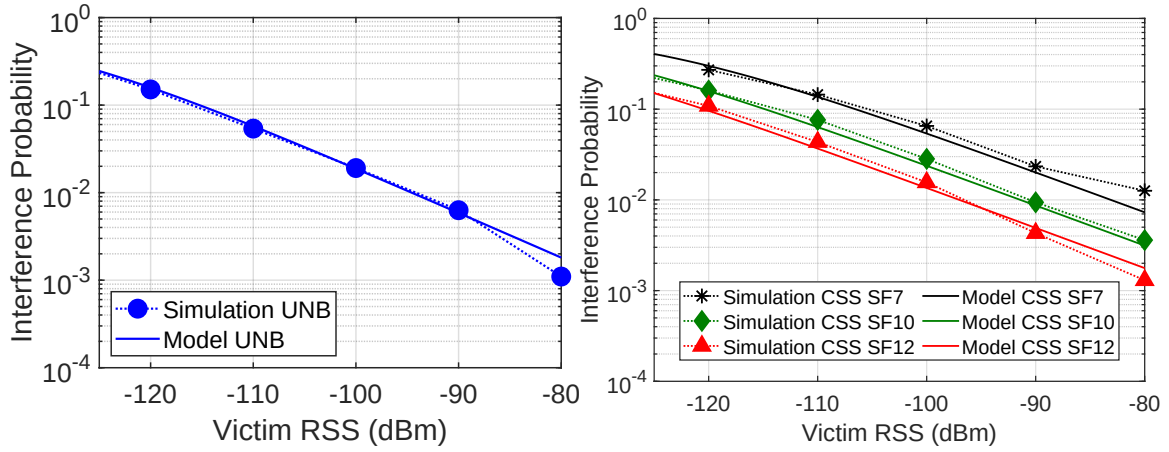


(a) Equal DC (b) Equal data rate

Figure 2.6: Urban self-interference probability results for UNB and CSS networks in the equal DC (a) or equal data rate (b) scenarios.

2.7.2 Urban cross-interference scenario

Fig. 2.7 describes the probability of urban cross-interference between CSS and UNB, as a function of the victim's RSS, in the equal DC condition which is the most challenging in terms of coexistence. In particular, Fig. 2.7a shows the interference probability for UNB as the victim when interfered by CSS, and in the worst scenario, i.e. with -130 dBm RSS for the victim, the interference probability is 0.36. Vice-versa, Fig. 2.7b shows the urban cross-interference results when CSS represents the victim. In this case, we tested three different SFs independently (namely, SF7, SF10, SF12), and both SF10 and SF12 are more robust than UNB (in the worst-case scenario of -130 dBm RSS), with the interference probability equal to 0.30 and 0.21 respectively.

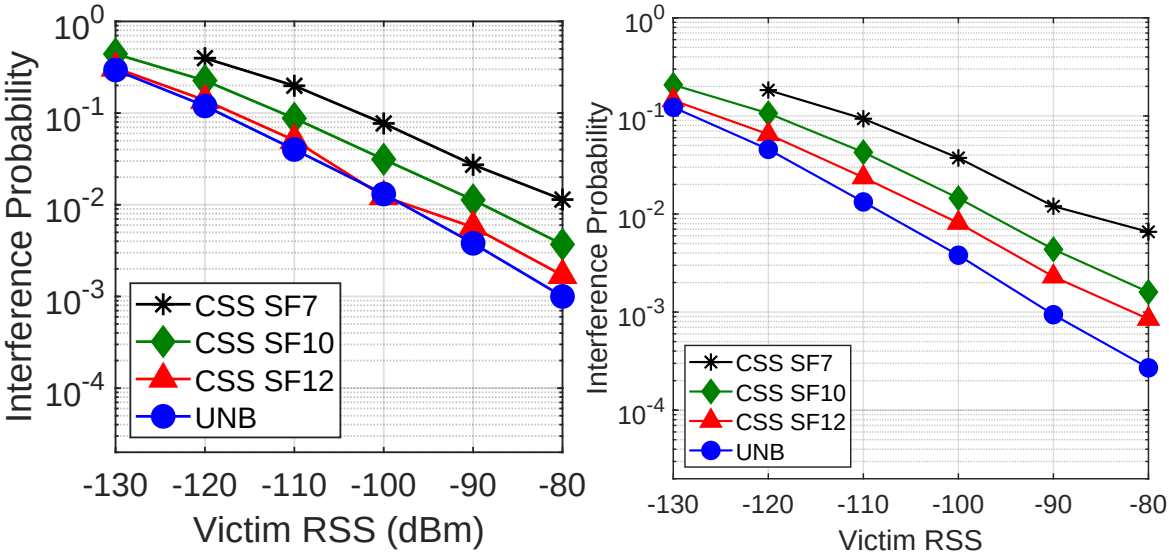


(a) UNB Victim, DC=1% (b) CSS Victim, DC=1%

Figure 2.7: Urban cross-interference probability for the equal DC scenario for the UNB system (a), and CSS systems (b).

Instead, SF7 is less robust to interference, with 0.27 of interference probability against 0.15 of UNB (measured at -120 dBm of RSS).

2.7.3 Urban mixed-interference scenario



(a) Equal DC (b) Equal data rate

Figure 2.8: Urban mixed-interference for the equal DC (a) or equal data rate (b) scenarios.

Fig. 2.8 shows the simulation results in the urban mixed scenario. In this case, the interfering nodes are equally distributed among UNB and CSS (i.e. 50% of the nodes using UNB and 50% using CSS). For the equal DC experiment, Fig. 2.8a shows an interference probability for UNB very close to CSS at SF12, while the other SFs are more sensitive to interference. In case of equal data rate, Fig. 2.8b shows a lower interference probability for all systems, especially for UNB (due to the lower ToA consumed by LoRa in this experiment). Moreover, compared to urban cross-interference, the urban mixed scenario has less impact on UNB (e.g for RSS=-130 dBm the interference probability is 6.5% less), while CSS is affected negatively (e.g. SF12 loses about 10% at -130 dBm). However, if we consider performance in terms of average throughput, in the same experiment scenario of equal DC, Fig. 2.9 shows that the result is opposite: in the worst-case scenario, UNB delivers only 101 packets/day, against 403 pkts/day, 1177 pkts/day and 8270 pkts/day for CSS with SF12, SF10, and SF7 respectively. Moreover, Fig. 2.9 shows that for RSS values close to -80 dBm, the daily throughput tends towards the values reported in the table 2.3.

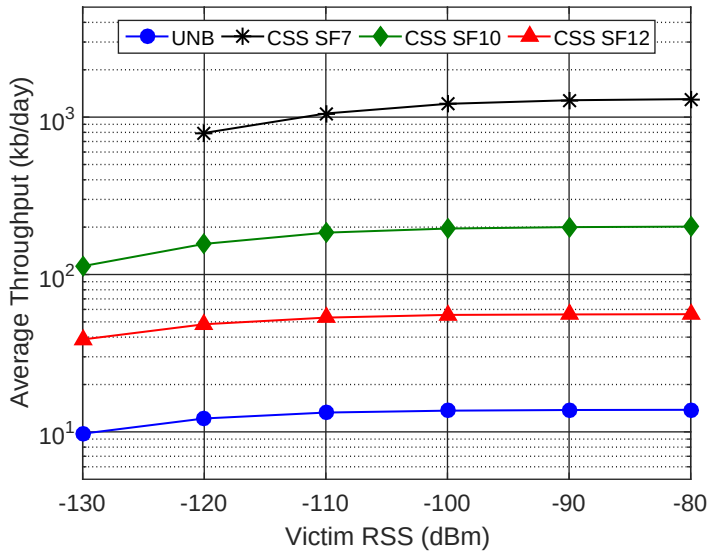
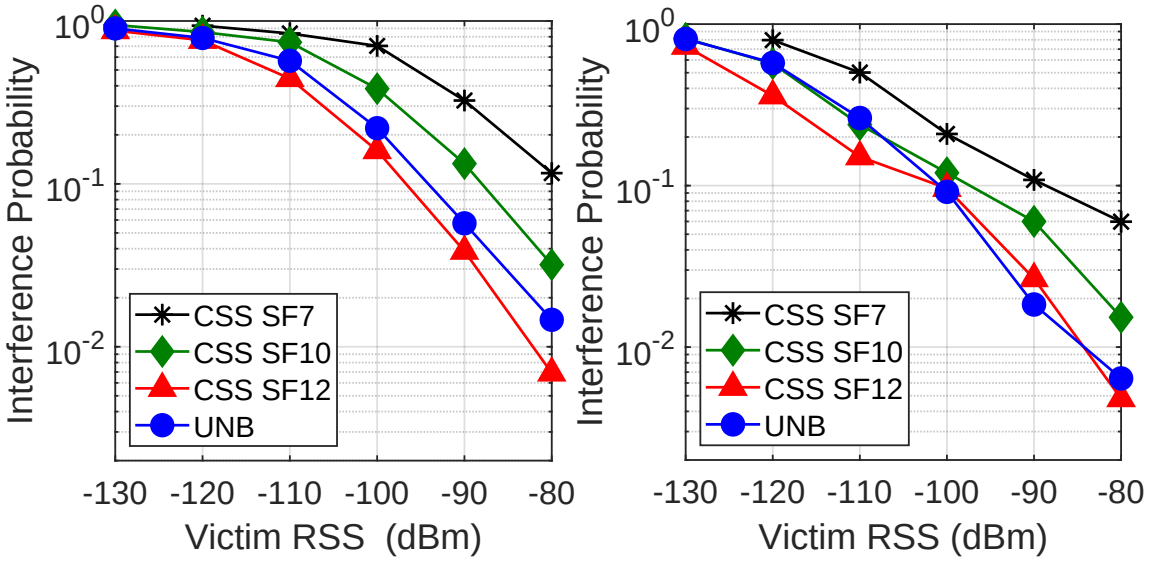


Figure 2.9: Average throughput/day with DC=1% and density of 1000 nodes/Km^2.

2.7.4 Rural mixed-interference scenario

The rural scenario also involves a device density of 1000/km^2, with simulations extended to a larger radius compared to the urban case, namely from 1km to 5km. Consequently, to maintain a constant density, the number of devices during simulation is increased from 3140 (of the urban case) to 78539 devices.



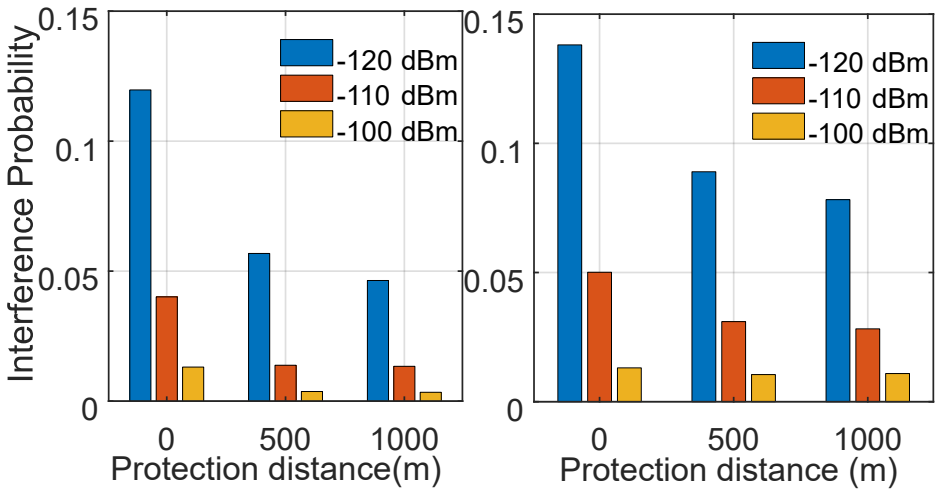
(a) Equal DC (b) Equal data rate

Figure 2.10: Rural mixed-interference for the equal DC (a) or equal data rate (b) scenarios.

The results of the simulations, in terms of interference probability in the mixed rural scenario, are shown in the figure 2.10. In case of equal DC, Fig. 2.10a shows that for RSS equal to or less than -120 dBm all victims technologies have an interference probability greater than 0.7 With an almost partial overlap of the curves. For power levels greater than 120 dBm, the curves begin to differentiate. Under these conditions, the most robust technology appears to be SF12, which, similar to the urban case, deviates only slightly from UNB. In fact, for an RSS of -90 dBm, the interference probability was 0.04 and 0.056 for SF12 and UNB, respectively. Compared with the case of equal DC, the interference probability is reduced for each technology in the case of equal data rate, as shown in Figure 2.10b. Specifically, in this rural scenario UNB, SF12 and SF10 technologies had an interference probability of less than 0.07 for an RSS of -90 dBm for the three victims considered.

2.8 Interference problem: Discussion and Possible Solutions

The presented results highlight the importance of mitigating the interference caused by heterogeneous technologies sharing the same frequency bands. Possible solutions,



(a) UNB victim (b) CSS SF12 victim

Figure 2.11: Protection distance effects in mixed-interference scenario with DC=1% and density of 1000 nodes/Km^2 for UNB (a) and CSS (b) victims.

most of these discussed in [150], are: (i) cognitive radios, to exploit opportunistic channel access techniques; (ii) protection distance, arranging devices at a distance that significantly reduces the probability of interference; (iii) smart antennas, to direct power only in the areas where devices are most concentrated; (iv) data rate adaptation to reduce the ToA as much as possible; and, finally, (v) multiple gateways and power control, to increase diversity, reduce the transmit power and, consequently, lower the interference. Regarding this last option, indeed different gateways experience different power ratios between the strongest received packet and the interfering signals. When the power ratios are higher than a minimum threshold, the capture effect assures that in case of collision multiple gateways can correctly decode different packets simultaneously, up to the number of deployed gateways. This can significantly increase the overall capacity of the network [176]. Finally, advanced receiver schemes can be used to improve the decoding performance in case of interference, for example by implementing Successive Interference Cancellation (SIC) mechanisms [177].

Since multiple gateway deployment and SIC schemes have been widely discussed in the literature, here we briefly present the results obtained exploiting protection distance to mitigate cross-interference. Focusing on the mixed scenario with Equal DC (the most challenging), and considering a protection distance of 0, 500 and 1000 meters, we run two different experiments in SEAMCAT with UNB or CSS as victim technology respectively. In each scenario, the protection distance was applied so that

the devices placed close the victim are only the ones of the same technology. Moreover, we defined three different levels of RSS received by the victim: -120 dBm (worst case), -110 dBm (typical) and -100 dBm (best case). In Fig. 2.11a, we observe that for UNB a protection distance of 500 m reduces the probability of interference, in the worst case of -120 dBm, from 12% to 6% (thus a 50% reduction) compared to no protection distance. The improvement obtained using a protection distance of 1 Km is an additional 33% reduction in interference probability. Fig. 2.11b shows the results of the experiments when the victim is CSS. In this case the probability of interference decreases from 14% to 7% with protection distance of 500m and victim RSS equal to -120 dBm, again obtaining a decrease of 50%. For higher RSS, the interference is below 5% and 2% respectively suggesting that protection distances higher than 500m do not provide much benefit. Indeed, self-interference due to the same technology of the victim node is still present and is not concerned by the protection distance mechanism.

2.9 Summary

In this work, we presented a coexistence study of CSS and UNB technologies, in realistic deployment scenarios and traffic conditions. The environment and system parameters are supported by an in-depth literature review in order to build accurate LPWAN simulation scenarios in terms of node density, DC, indoor/outdoor locations, antenna elevation, etc. The experimental evaluations are executed using the SEAMCAT simulator and the configuration files are made available to the scientific community to enable future integration of the research. Results have demonstrated that, if the RSS is high enough (e.g. above -90 dBm), UNB and CSS systems may coexist safely. However, in more extreme conditions (e.g. at the cell edge), UNB is more robust to interference, although CSS generally offers higher throughputs. Finally, we consider possible interference mitigation strategies, showing the performance obtained using protection distance. Overall, this study represents an important analysis and risk assessment for future IoT scenarios where both LPWANs might coexist.

Chapter 3

Energy-Autonomous and Battery-Free Wireless Piezoresistive Sensors for Smart Agriculture Applications

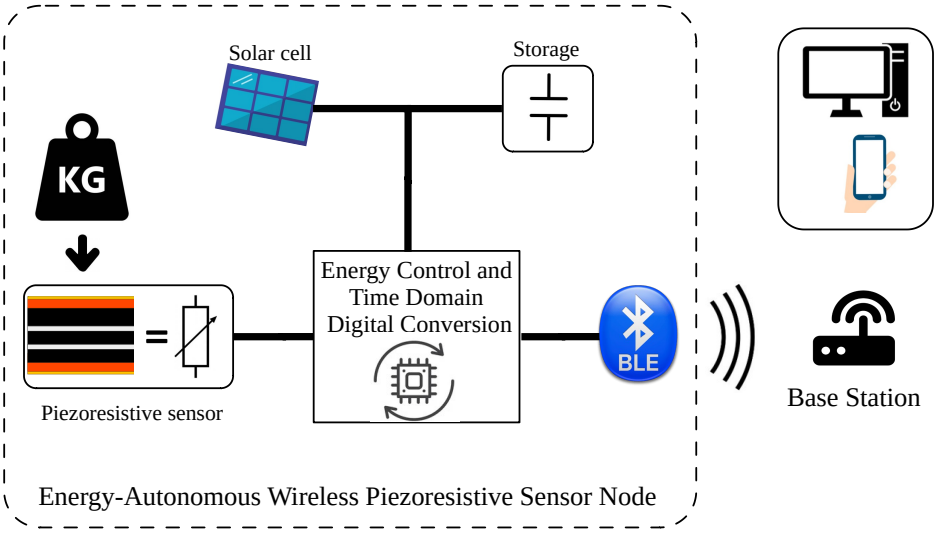


Figure 3.1: Graphical chapter overview

3.1 Overview

This chapter presents an innovative wireless piezoresistive sensor that utilizes an energy-harvesting system to convert ambient light into electrical energy, making it energy-autonomous and ideal for mass-scale IoT applications in agriculture. The

Wireless Sensor Node (WSN) achieves high sensitivity and accuracy in strain measurements through its carefully engineered piezoresistive element and Time Domain to Digital Conversion (TDDC) technology that eliminates the need for power consuming and costly analog-to-digital converters and reduces power consumption. Its Bluetooth Low Energy (BLE) wireless communication enables data transmission to a remote central Base Station (BS) without requiring an external power source. The energy-autonomous feature provides long-term, maintenance-free operation, making it suitable for mass-scale agricultural applications in remote and inaccessible environments. This novel wireless piezoresistive sensor has the potential to revolutionize the field of mass-scale agricultural IoT sensing and enable new applications that were previously impossible due to power constraints. Its energy-autonomous feature and wireless communication capabilities make it an ideal solution for large-scale IoT deployments, enabling real-time monitoring and control of infrastructure, automation, and other applications in agriculture.

3.2 Introduction

Wireless Sensor Nodes (WSNs) are utilized in smart agriculture to gather data such as soil moisture, temperature variations, humidity levels, and more [178]. The direct monitoring of plant parameters in a rural context, with many plants to monitor, requires sensor technology that is energy-autonomous and maintenance-free. In fact, the maintenance issue would bring farmers the costly problem of sensor maintenance, compromising the effectiveness of the detection technology from an economic point of view. In such IoT applications, energy efficiency and maintenance are critical issues for agricultural wireless network reliability, as the installed wireless sensor nodes require constant and reliable power to perform IoT monitoring, interact with the rest of the network, and transmit information to the server. A mass-scale IoT sensor deployment requires synergy between environmental health, economic prosperity, and social progress of current and future communication networks to promote sustainability [179]. As a result, energy harvesting is becoming an increasingly popular choice for powering sustainable wireless nodes, as the maintenance required for periodic battery replacement can be a discouraging factor for mass-scale IoT deployments in agriculture. State-of-the-art wireless sensor nodes use batteries, which require periodical physical maintenance that can be tedious, environmentally harmful, and costly [180].

Indeed, battery replacement costs can exceed the value of the IoT device itself. Even with a 10-year battery lifespan, there would still be several million battery re-

placements per day for IoT devices [181]. Therefore, achieving energy autonomy for wireless sensor nodes is a primary goal in deploying mass-scale IoT in agriculture, which drives engineers to adopt an energy-harvesting approach. Depending on the environment, several types of harvesters are available, and making an optimal choice among the available options is crucial. For example, solar panels can be a convenient option for home automation or smart agriculture [182], and piezoelectric harvesters can be a good choice near vibrating elements, such as machines and engines [183]. However, using harvesters also presents other challenges in energy management, as the available energy is intrinsically variable and typically very low. A previously published work [184, 185] shows a solution that combines ultra-low-power devices, which are capable of implementing proper power management [186], with the Time Domain to Digital Conversion (TDDC) techniques [184, 185]. TDDC is well-known in the literature as a technique for measuring ambient light in agriculture [182, 187]; nevertheless, this is also proposed for various IoT applications: for measuring Time-of-Flight [188], in LiDAR applications [189, 190], as a temperature sensor [191, 192], in time-domain ADCs [193], in RFID (radio frequency identification) tag sensors [194], in ADPLL (all digital phase-locked loop) [195], and as flow meters [196]. Recently, the authors in [197] introduced innovative methods for Resistance-to-Time-to-Digital conversion. These techniques can be implemented for resistive sensors and offer rapid, accurate, and low-energy consumption readings of the sensors. The main advantage is that the sensor reading implements direct interface circuits [198, 199] and the TDDC techniques. In fact, instead of using analog-to-digital converters (ADC), these solutions adopt a different approach based on converting time to digital and then to resistance. In this process, time measurements are obtained through charging-discharging cycles of a capacitor and subsequently processed by a microprocessor that expresses the results in terms of the number of internal clock cycles. These solutions can be applied to resistive, inductive, and capacitive sensors [200, 201]. Furthermore, [197] has demonstrated that energy consumption can be reduced by up to 75 %, making the TDDC technique ideal for developing ultra-low power energy-autonomous wireless sensor nodes in Smart Agriculture. In this context, the objective of the present study is to develop an Energy-Autonomous Wireless Piezoresistive Sensor Node (EAWPSN) integrated with time-to-digital conversion. The chapter presents a system designed for sensing mechanical stress or strain with an energy-autonomous and battery-free system. It comprises a piezoresistive sensor, a solar cell and a storage capacitor for harvesting and storing ambient energy, a Bluetooth Low Energy (BLE) wireless com-

munication interface for data transmission, and an ultra-low-power microcontroller that implements the power management and the time-to-digital conversion.

Furthermore, piezoresistive sensors are used in a broad range of agricultural applications, including soil moisture monitoring [202], treatment of agricultural waste using flexible pressure biosensors [203], detection of various soil macro-nutrients [204], agricultural pathogen detection [205], and to sense mechanical stress and deformations [206, 207]. This chapter is motivated by the awareness that a combined implementation of EAWPSN and TDDC techniques could significantly improve the energy performance, reliability, and sustainability of sensing in Smart Agriculture. Indeed, the use of batteries today poses a bottleneck for the widespread adoption of IoT in agriculture. Moving towards a future with trillions of IoT devices, battery replacement will be prohibitively expensive and environmentally irresponsible. Energy-autonomous and battery-free sensors could reshape Smart Farming into a wider and more sustainable agricultural infrastructure [208]. In this direction, the proposed system represents an innovative contribution to achieving massive and sustainable monitoring in various mechanical and strain applications in agriculture.

The chapter introduces several innovative aspects, including:

- The first ever demonstration of a TDDC technique for force and resistive measurements, in the context of wireless, battery-free sensor nodes;
- An approach that allows for low-complexity and low-cost designs, making it compatible for truly mass-scale, battery-less IoT applications in agriculture;
- The presentation of a mathematical model for the resistance to time domain digital conversion technique;
- The development of two calibration methods of the TDDC technique for either force and resistance measurements.

These innovations have the potential to significantly advance the field of wireless, battery-free sensor nodes, and enable new agricultural applications that were previously impossible due to power constraints.

The remainder of this article is organized as follows: section 3.3 describes the system architecture and discusses the design procedure followed to achieve the optimum energy and cost efficiency. Section 3.4 describes the time-to-digital conversion technique used to measure the sensor's resistance. Section 3.5 shows the experimental results. Section 3.7 reports the conclusions.

3.3 System Description and Design

3.3.1 System Description

The EAWPSN works in a distributed sensing IoT architecture, where direct and easy access to the sensors' environment could be hard to reach. Under this objective, each EAWPSN performs time measurements locally using the TDDC technique and reports the acquired values to a remote Base Station (BS) through a wireless communication protocol such as BLE, LoRa, NB-IoT, and the like. In this work, the EAWPSN implements BLE connectivity.

The BS is a device equipped with a Bluetooth receiver that is powered continuously either through a main power source or a battery. As a result, the device does not encounter any power-related issues. It collects data sent by one or more nodes and can perform energy-intensive operations.

Being energy-autonomous and battery-free, the proposed EAWPSN device is a WSN that is inherently maintenance-free, sustainable, environmentally friendly [209], safe [210], low-cost, low-volume, and easy to deploy and expand.

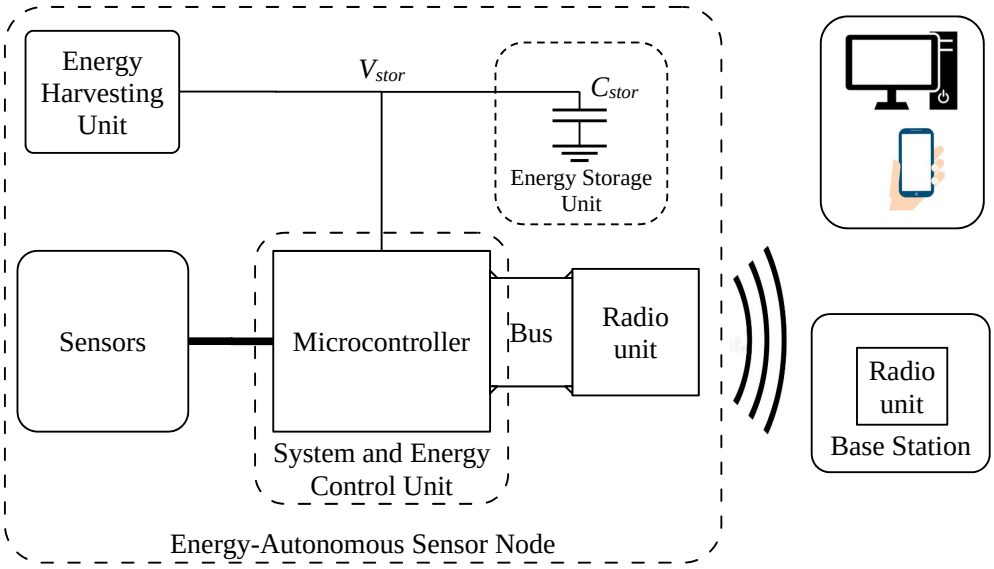


Figure 3.2: Block diagram of the IoT system.

Fig. 3.2 illustrates the architecture of the EAWPSN. Each WSN is composed of a System and Energy Control Unit (SECU), an Energy Harvester Unit (EHU), one Energy Storage Unit (ESU), and Sensor Unit (SU). The EHU can include one or more energy harvesters and is the only system power source. The SECU executes the finite

state machine, monitors and manages the power generated by the EHU, the power flowing into the ESU, and biases the sensors accordingly.

The EAWPSN works by alternating an *harvesting phase* to an *active phase*. During the harvesting phase, the system is configured to work under the lowest possible power consumption to minimize the quiescent power consumption and allow the harvested energy to flow efficiently into the ESU. In the active phase, the system uses the harvested energy to perform an energy-intensive action, like sensing or communicating with external devices through the radio.
This approach possesses the secondary advantage of reducing the minimum requirements of the EHU, allowing for lower costs and complexity for the node.

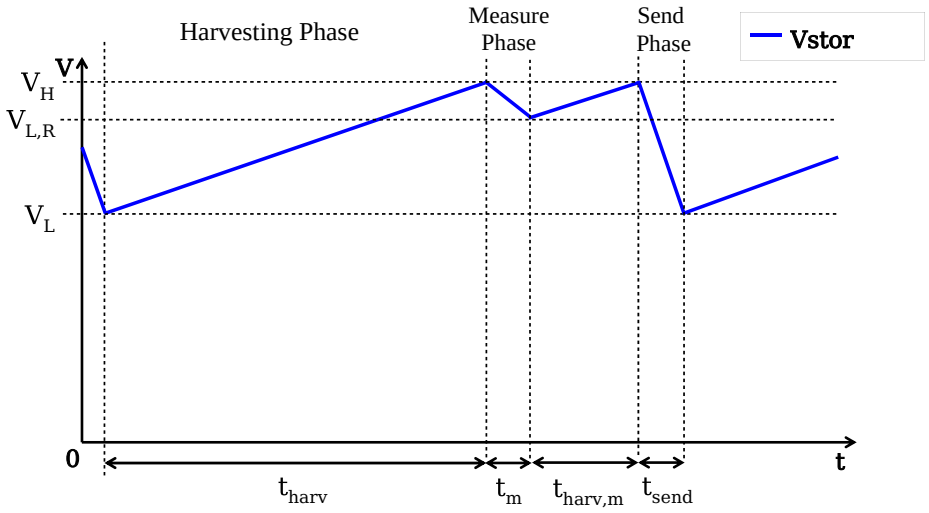


Figure 3.3: Typical work cycle of the EAWPSN.

The ESU is a low-cost and low-volume ceramic capacitor C_{stor}. Fig. 3.3 shows the typical work cycle of the EAWPSN. Each active phase follows cyclically after the harvesting phase. During the harvesting phase, the voltage V_{stor} reaches a defined voltage threshold value V_H, which corresponds to the highest level of energy harvested in the capacitor C_{stor}. In the active phase, as the power consumption of the EAWPSN (in the order of tens of milliwatt) is much greater than the power provided by the EHU (in the order of tens of microwatts), the voltage V_{stor} drops steeply down to the lowest voltage threshold V_L. The value of the capacitor C_{stor}, calculated as in (3.1), can store enough energy so that the voltage V_{stor} never drops below the lowest voltage V_L, while the system accomplishes the assigned tasks during the active phase. The voltage V_L is set to minimize the value of C_{stor} at the lowest possible voltage that keeps the Microcontroller out of its reset state.

$$C_{stor} = \frac{2 \cdot E_{harv}}{V_H^2 - V_L^2} \tag{3.1}$$

In steady-state operation, the time t_{harv} it takes the EAWPSN to harvest the energy depends on the energy E_{harv}, the power generated by the harvester P_{geh}, and the quiescent power of the system P_q as in (3.2):

$$t_{harv} = \frac{E_{harv}}{P_{geh} - P_q} \tag{3.2}$$

t_{harv} is digitally converted to the number of clock pulses N_{harv} counted by the timer LPTIM that runs at the frequency f_{clk} as in (3.3).

$$t_{harv} = \frac{N_{harv}}{f_{clk}} \tag{3.3}$$

It is worth noting that, as already described in *La Rosa et al.* [187], once the parameters V_H, V_L and C_{stor} are fixed, the energy value E_{harv} during the harvesting phase is constant. Therefore, the knowledge of N_{harv} provides the value of the power P_{geh} (3.4).

$$P_{geh} = \frac{E_{harv} \cdot f_{clk}}{N_{harv}} + P_q \tag{3.4}$$

In practical applications, for the system to be capable of harvesting energy, the minimum power generated by the harvester P_{geh_min} must be greater than the quiescent power P_q so that the condition in (3.5) must be satisfied.

$$P_{geh\_min} \gg P_q \tag{3.5}$$

3.3.2 System Design

The WSN proposed in this work exploits BLE for the low-energy system requirement and the widespread compatibility of the protocol with commercially available IoT devices such as computers, smartphones, and tablets. Specifically, the device used for the radio system is the commercially available BlueNRG-2 provided by STMicroelectronics, configured to send beacons at a transmission power of +8 dBm with a peak current of ≈ 14 mA. It is worth noting that the Bluetooth standard requires beacon emissions to be distant in time no less than 100 ms [211]. For this purpose, the EAWPSN uses a specific ultra-low-power timer LPTIM to measure the advertising time,

implement the TDDC technique and assure a Bluetooth standard-compliant elapsed timing between consecutive broadcast events.

The SECU is implemented with the off-the-shelf ultra-low-power Cortex-M0+ microcontroller from STMicroelectronics (STM32L031F4) [186]. During the harvesting phase, the microcontroller is configured in stop mode with a quiescent current I_q of 1 µA that results in a quiescent power P_q of 2.5 µW. The device can supply external devices, i.e., the BLE radio, the sensors, and the like, through a General Purpose Input/Output (GPIO) capable of driving a maximum current of 10 mA. Depending on the target Smart Agriculture application of the device, several types of harvesters can be used. In this chapter, the harvester used was the commercially available AM-1606C amorphous silicon photovoltaic cell provided by Panasonic, with an open-circuit voltage of 3.7 V and capable of providing $\approx$ 10 µW at the defined optimal operating point [212]. This photovoltaic cell offers several benefits, including the biasing of the system with a minimum light intensity of 200 lux, as it provides a current of 3 µA. Moreover, its small size (1.5 cm$\times$1.5 cm) allows for easy integration into the developed system requiring miniaturization.

The value of the capacitor C_{stor} must ensure that during the energy-intensive active phase, the voltage V_{stor} doesn't drop below the minimum bias voltage $V_{\text{dd\_min}}$ of the used electronic devices, which for the off-the-shelf devices used, corresponds to a minimum voltage of 1.8 V. For the proposed device, a capacitor C_{stor} of 440 µF provides storage for 1.1 mJ of energy when $V_{\text{H}} = 3.0$ V and $V_{\text{L}} = 2.0$ V.

Several transducers and technologies are available that can be categorized by the physical phenomenon for sensing pressure or applied force:

- Piezoelectric transducers use the generation of an electric potential whose value is proportional to the applied mechanical stress on piezoelectric crystals [213],
- Electret-based transducers use dielectric materials possessing a permanent electric dipole exploiting the voltage variation from the deformation caused by applied stress [214],
- Piezoresistive transducers rely on the piezoresistive effect, which causes a change in the electrical resistance of a material under mechanical stress [215].

Piezoelectric and Electret-based transducers require an amplifier, adding to the cost and power consumption of the EAWPSN. Piezoresistive sensors are cheap, widely used, and their value can be sensed via an ADC and a test resistor [216]. The EAWPSN embeds the piezoresistive transducer described in Section 3.3.3.

3.3.3 Description of the piezoresistive sensor

Piezoresistive polymers are a type of material that exhibit changes in electrical resistance when under mechanical stress. These polymers are flexible, inexpensive, and reliable, which makes them an excellent choice for monitoring pressure distribution [217]. These polymers are used in a wide range of applications, including the design of flexible IoT sensors [218, 219] and in mechatronic and biomedical devices [220].

The piezoresistive effect is responsible for the changes in electrical resistivity observed in these materials when subjected to mechanical stress. This effect is illustrated in Fig. 3.4, which shows that, by applying a compressive force, the distance between charged particles inside the piezoresistive film decreases, thereby increasing the number of conductive paths and resulting in lower resistivity.

Piezoresistive sensors can be fabricated utilizing different materials, such as Velostat® [221], a polymer composite of polyethylene impregnated with carbon black. Table 3.1 provides a summary of the physical properties of the material used in the manufacture of these sensors.

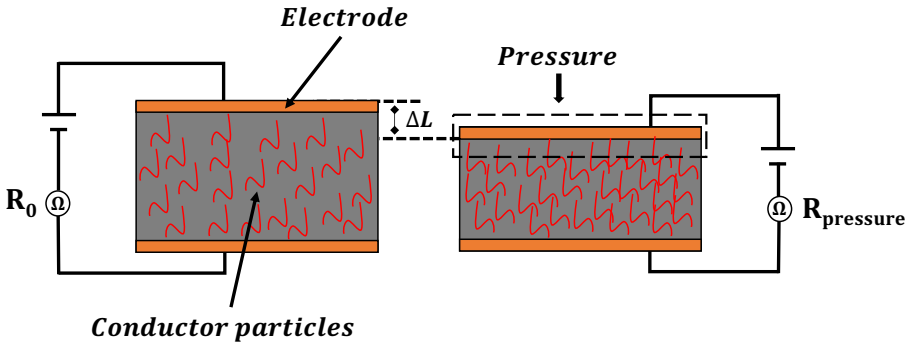


Figure 3.4: Schematic illustration of piezoresistivity, and the effect of the applied pressure tends to change the particles distances. With $R_{pressure} < R_0$

Table 3.1: Selected mechanical and electrical properties.

Technical specifications	
Thickness	0.1 mm
Size	4.2 cm × 2.6 cm
Temperature Limits	-45°C to 65°C (-50°F to 150°F)
Volume fraction of carbon particles	0.2873
Diameter of carbon particles	500 nm
Volume Resistivity	≤500 Ω·cm

The model proposes the use of multilayer piezoresistive sensor, as shown in Fig 3.5. These sensors consist of six layers of Velostat, each with a thickness of 0.1 mm. The layers are arranged in series using an Anisotropic Conductive Film (ACF) [222] adhesive to bond them together. Copper tapes, 0.065 mm thick, are positioned above and below the Velostat without any adhesive, serving as electrodes. In order to optimize the compatibility of the piezoresistive sensor with conditioning electronics, two layers of adhesive insulators (polyimide), each 0.076 mm thick, are applied above and below the electrodes for environmental protection and sensor assembly consolidation.

This enhances the resistance value of the sensor, according to [216], research suggests that multilayer sensors utilizing ACF adhesive exhibit superior sensitivity compared to those lacking ACF.

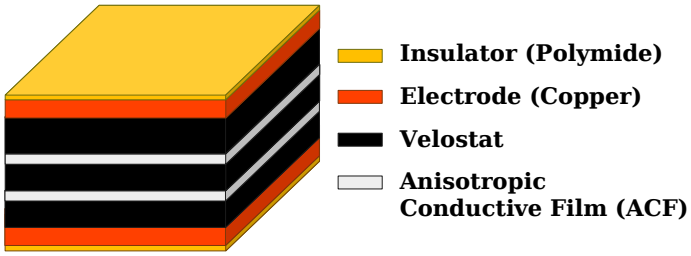


Figure 3.5: Multilayer sandwich structure with the ACF adhesive between the piezoresistive layers.

ACF is an electrically anisotropic, conductive pressure-sensitive adhesive tape infused with silver particles that enables interconnection through the adhesive thickness (z-axis) between substrates. Contact between the adhesive and Velostat is established by applying finger pressure at 15 psi (0.10 MPa).

In this context, the use of ACF ensures continuous electrical contact, even without pressure, facilitating higher resistance, which is advantageous for electronic integration.

3.4 Resistive sensing technique

3.4.1 System Setup

The sensing of the resistance R is performed through several methods and techniques. Most of these techniques typically require an (Analog to Digital Converter) ADC, a conditioning circuit, or a separate discrete device, which implies an extra, typically significant amount of energy to account for the sensing activity. In this work, the

EAWPSN computes the resistance R, using the TDDC technique that measures the discharge time t_m of the voltage V_{stor} across the capacitor C_{stor} through the resistor R between two voltage thresholds V_H and $V_{L,R}$ with $V_H > V_{L,R}$.

Fig. 3.6 shows the block diagram of the EAWPSN. The resistor R to measure is connected to the microcontroller directly through the GPIO1. Similarly, two calibration resistors R_{cal1} and R_{cal2} are respectively connected through GPIO2 and GPIO3 to cyclically calibrate the system and ensure the reliability and accuracy of the measurements.

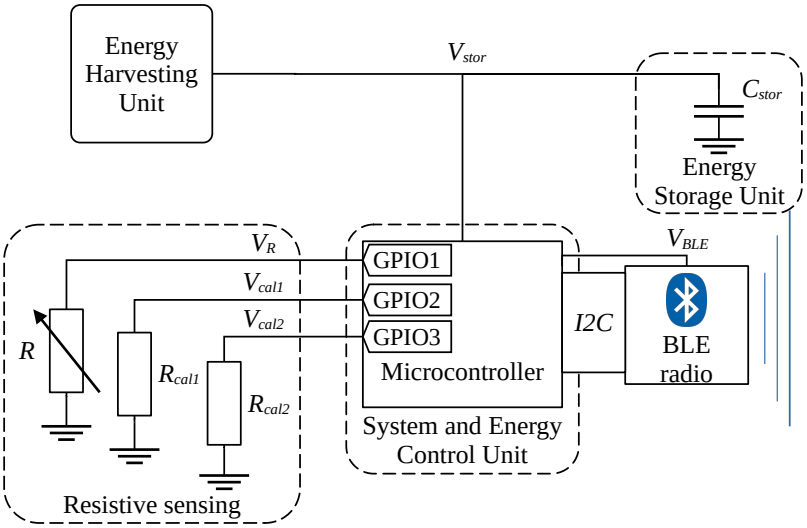


Figure 3.6: Block diagram of the resistance measurement system.

Referring to Fig. 3.7, during the measurement phase, the resistor R is connected to the voltage V_{stor} through GPIO1 so that the voltage V_R through the resistor R is equal to the voltage V_{stor}.

This cause the capacitor C_{stor} to discharge through the resistor R and the voltage V_{stor} drops exponentially with the time constant $\tau = R \cdot C_{stor}$. Therefore, the resistance R can be indirectly calculated from the time t_m needed to discharge the storage capacitor C_{stor} by exploiting the fact that its value is known and static. The voltage V_{stor} varies between two defined voltage thresholds, the high voltage threshold V_H and the low threshold voltage $V_{L,R}$.

In addition to being inherently low-power, this approach translates into minimal use of hardware resources and system simplicity as it takes advantage of the internal hardware peripherals of the microcontroller, i.e., PVD (Programmable Voltage Detector) and LPTIM (Low-Power Timer), which are already required to monitor the V_{stor} voltage during the energy harvesting stage and just one additional GPIO per each

measured resistor. This approach has the relevant advantage of leading to the use of simplified and low-cost microcontrollers.

Another advantage of this method is that the energy E_{meas} used per each measure, given by (3.6), is constant. The voltage thresholds V_H and $V_{L,R}$ are defined simply via the PVD with on-the-fly configurable voltage values in the range from 2.0 V to 3.2 V with programmable discrete steps of 200 mV [186].

E_{meas} represents the energy required to perform the measurement, and it is lower than E_{harv} since the latter is necessary to complete a BLE transmission. As a result, $V_{L,R}$ is usually higher than V_L.

It is worth noting the dual use of the storage capacitor C_{stor} that is simultaneously used as a storage device to harvest the energy and as a measurement device in the measurement process of the resistor R.

$$E_{meas} = \frac{1}{2} \cdot C_{stor} \cdot \left(V_H^2 - V_{L,R}^2\right) \tag{3.6}$$

The measurement of the time t_m is executed as follows:

1. The system wakes up from the harvesting phase, i.e., ($V_{stor} = V_H$),
2. The LPTIM clock is configured to run at the highest frequency to maximize the measurement resolution and accuracy,
3. The signal V_R is connected to V_{stor} through the GPIO1,
4. The Microcontroller monitors the value of the voltage V_{stor} through the PVD,
5. The value of t_m is digitally converted in the number of clock pulses N_m through the timer LPTIM,
6. The value of N_m is recorded when the threshold $V_{L,R}$ is reached,
7. The signal V_R is disconnected from V_{stor} through GPIO1,
8. The MCU enters in Stop Mode, i.e., in the energy harvesting phase.

Equation (3.7) evaluates the power balance of the system:

$$P_{geh} + P_{C_{stor}} = P_R + P_q \tag{3.7}$$

P_R and $P_{C_{stor}}$ are respectively the power dissipated through R and the power flowing through C_{stor}.

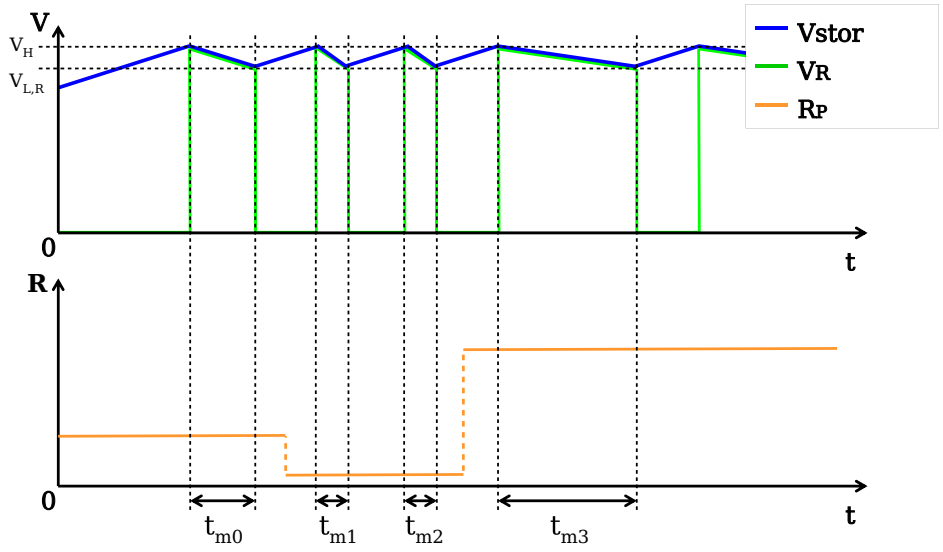


Figure 3.7: V_{stor} and V_R evolution during the resistor measurement phase.

As shown already in (3.5) the quiescent power P_q is negligible with respect to P_{geh}, so that (3.7) can be approximated as in (3.8).

$$P_{geh} + P_{C_{stor}} \approx P_R \tag{3.8}$$

In practical cases, the voltage V_{stor} is nearly constant as the EAWPSN is designed so that V_{stor} varies between the two values V_H and $V_{L,R}$ which differ for a couple of hundred millivolts. Therefore, P_{geh} can be considered invariant with V_{stor} leading to (3.9).

$$P_{geh} - C_{stor} \cdot V_{stor}(t) \cdot \frac{dV_{stor}(t)}{dt} = \frac{V_{stor}^2(t)}{R} \tag{3.9}$$

Solving (3.9), and imposing the initial condition $V_{stor}(0) = V_H$, the voltage $V_{stor}(t)$ is given by (3.10):

$$V_{stor}(t) = \sqrt{(V_H^2 - R \cdot P_{geh}) \cdot e^{-2t/RC_{stor}} + R \cdot P_{geh}} \tag{3.10}$$

The system is designed so that the voltage V_{stor} drops from the voltage V_H to the voltage $V_{L,R}$ within the time t_m. Therefore, by imposing the condition $V_{stor}(t_m) = V_{L,R}$ in (3.10), the time t_m is given by (3.11):

$$t_m = \frac{R \cdot C_{stor}}{2} \cdot \ln\left(\frac{V_H^2 - R \cdot P_{geh}}{V_{L,R}^2 - R \cdot P_{geh}}\right) \tag{3.11}$$

Equation (3.11) cannot be easily inverted to calculate the value of R and must be solved numerically. Therefore, this equation can not be used on the EAWPSN, and R can only be calculated remotely at the BS.

Equation (3.11) can be simplified to the linear (3.12) when the applicative condition of (3.13) is met.

$$t_m \approx R \cdot C_{stor} \cdot \ln\left(\frac{V_H}{V_{L,R}}\right) \tag{3.12}$$

$$V_{L,R}^2 \gg P_{geh} \cdot R \tag{3.13}$$

Given the currently available power from the harvester, the condition from (3.13) defines the maximum range of validity for (3.12), as in (3.14). Therefore, the condition in (3.5) de facto defines the maximum measurable value of R as in (3.15).

$$R \ll \frac{V_{L,R}^2}{P_{geh}} \tag{3.14}$$

$$R_{max} \ll \frac{V_{L,R}^2}{P_{geh\_min}} \tag{3.15}$$

Equation (3.15) reveals that to maximize $\mathrm{R_{max}}$, $\mathrm{V_{L,R}}$ must be maximized and set at the highest possible value, as in (3.16).

$$V_{L,R} \approx V_H \tag{3.16}$$

By defining $\mathrm{V_{avg}}$ and $\mathrm{V_{\Delta V}}$ as in (3.17) and (3.18) respectively

$$V_{avg} = \frac{V_H + V_{L,R}}{2} \tag{3.17}$$

$$V_{\Delta V} = V_H - V_{L,R} \tag{3.18}$$

$\mathrm{V_H}$ and $\mathrm{V_{L,R}}$ can be calculated as in (3.19) and (3.20).

$$V_H = \frac{V_H + V_{L,R}}{2} + \frac{V_H - V_{L,R}}{2} = V_{avg} + \frac{V_{\Delta V}}{2} \tag{3.19}$$

$$V_{L,R} = \frac{V_H + V_{L,R}}{2} - \frac{V_H - V_{L,R}}{2} = V_{avg} - \frac{V_{\Delta V}}{2} \tag{3.20}$$

Combining (3.11), (3.19) and (3.20) leads to (3.21).

$$t_m \approx R \cdot C_{stor} \cdot \ln\left(\frac{1+\frac{V_{\Delta V}}{2\cdot V_{avg}}}{1-\frac{V_{\Delta V}}{2\cdot V_{avg}}}\right), \tag{3.21}$$

The condition expressed in (3.16) implies that $V_{\Delta V}/V_{avg} < 1$ that allows to further simplify (3.21) as in (3.22).

$$t_m \approx 2 \cdot R \cdot C_{stor} \cdot \frac{V_H - V_{L,R}}{V_H + V_{L,R}} \tag{3.22}$$

Equation (3.22) can be inverted, and the value of R can be calculated also by the EAWPSN by simply applying (3.23).

$$R \approx \frac{1}{2 \cdot C_{stor}} \cdot \frac{V_H + V_{L,R}}{V_H - V_{L,R}} \cdot t_m \tag{3.23}$$

t_m is digitally converted to the number of clock pulses N_m counted by the timer LPTIM that runs at the frequency f_{clk} as in (3.24).

$$t_m = \frac{N_m}{f_{clk}} \tag{3.24}$$

From (3.23) and (3.24) derives the value of R in terms of the number of pulses counted by the timer LPTIM as in (3.25).

$$R \approx \frac{1}{2 \cdot C_{stor} \cdot f_{clk}} \cdot \frac{V_H + V_{L,R}}{V_H - V_{L,R}} \cdot N_m \tag{3.25}$$

Equation (3.25) shows that the minimum measurable resistance, i.e., the Limit of Detection (LoD) of the measurament system is given by (3.26)

$$R_{min} \gg \frac{1}{C_{stor} \cdot f_{clk}} \cdot \frac{V_H + V_{L,R}}{V_H - V_{L,R}} \tag{3.26}$$

Even if (3.25), in the measurement interval from R_{min} to R_{max}, can supply an accurate approximation in the considered measurement system, it cannot yet provide sufficiently precise results. Indeed, especially to evaluate low values of R, i.e., close to R_{min}, it is necessary to take into account the value of the resistor of GPIO1 that connects C_{stor} to R. Therefore, two main factors can influence the accuracy of R, i.e., the variation of the slope S_R, whose nominal value is given in (3.28), and the intercept R_0 of the function R vs N_m shown in Fig. 3.8 and represented by (3.27).

$$R = R_0 + S_R \cdot N_m \tag{3.27}$$

Where the nominal value of S_R is given by (3.28).

$$S_R = \frac{1}{2 \cdot C_{stor} \cdot f_{clk}} \cdot \frac{V_H + V_{L,R}}{V_H - V_{L,R}} \tag{3.28}$$

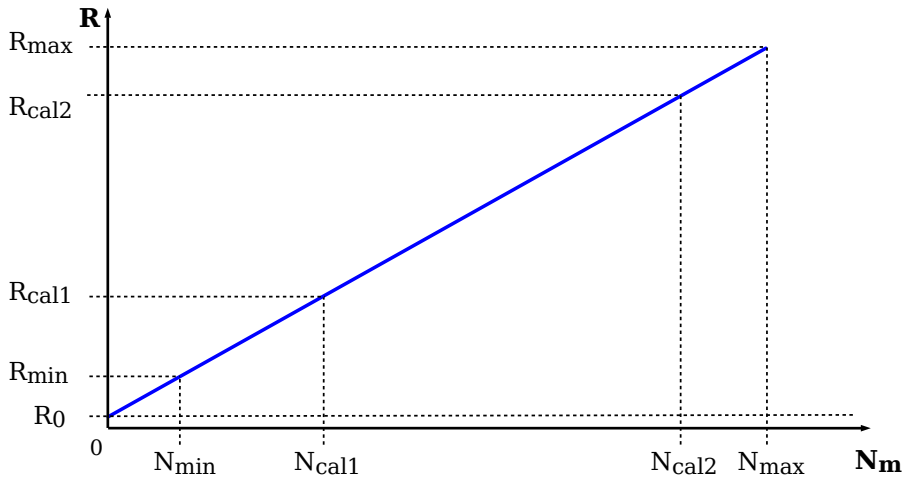


Figure 3.8: R vs N_m

R varies from its nominal value due to the manufacturing variations and tolerances from the nominal values of the parameters C_{stor}, V_H, $V_{L,R}$, and f_{clk} and (3.29) shows how the value of R deviates from its nominal value by dR given in (3.30).

$$R \pm dR = R_0 + S_R \cdot N_m \tag{3.29}$$

$$dR = dS_R \cdot N_m \tag{3.30}$$

Where dS_R is expressed in terms of the relative variations of C_{stor}, f_{clk}, V_H and $V_{L,R}$ by (3.31).

$$dS_R = \frac{\partial S_R}{\partial C_{stor}} \cdot dC_{stor} + \frac{\partial S_R}{\partial f_{clk}} \cdot df_{clk} + \frac{\partial S_R}{\partial V_H} \cdot dV_H + \frac{\partial S_R}{\partial V_{L,R}} \cdot dV_{L,R} \tag{3.31}$$

By combining (3.29) and (3.30), R can be expressed as in (3.32).

$$R = R_0 + S_R \cdot \left(1 \mp \frac{dS_R}{S_R}\right) \cdot N_m \tag{3.32}$$

By defining α as in (3.33), a more accurate equation for the measurement of R is given by (3.34).

$$\alpha = 1 \mp \frac{dS_R}{S_R} \tag{3.33}$$

$$R = R_0 + \alpha \cdot S_R \cdot N_m \tag{3.34}$$

Finally, (3.35) represents the sensitivity of S_R. Interestingly, it provides design insight for achieving the best accuracy of R, indicating the need to select a capacitor C_{stor} with sufficient accuracy, maximize f_{clk}, and suggest maximizing the voltage difference between V_H and $V_{L,R}$. This last condition contradicts the one previously expressed in (3.16) and highlights how the optimal choice of $V_{L,R}$ results from balancing the measurement range and the desired accuracy.

$$\frac{dS_R}{S_R} = -\frac{dC_{stor}}{C_{stor}} - \frac{df_{clk}}{f_{clk}} - \frac{2 \cdot V_{L,R}}{V_H^2 - V_{L,R}^2} \cdot dV_H + \frac{2 \cdot V_H}{V_H^2 - V_{L,R}^2} \cdot dV_{L,R} \tag{3.35}$$

3.4.2 System Calibration

The calibration procedure to derive the parameters R_0 and α of (3.34) is performed as described in the following steps:

1. Through the GPIO2 the MCU connects the capacitor C_{stor} to a known calibration resistor R_{cal1} ($R_{min} < R_{cal1} < R_{max}$), as shown in Fig. 3.6,
2. The MCU counts the numbers of pulses N_{cal1} related to the time t_{cal1} elapsed to discharge C_{stor} through R_{cal1},
3. Through the GPIO3 the MCU connects the capacitor C_{stor} to a known calibration resistor R_{cal2} ($R_{min} < R_{cal2} < R_{max}$ and $R_{cal1} < R_{cal2}$), as shown in Fig. 3.6,
4. The MCU counts the numbers of pulses N_{cal2} related to the time t_{cal2} elapsed to discharge C_{stor} through R_{cal2}.

Once recorded N_{cal1} and N_{cal2}, the parameters α and R_0 are provided in (3.37) and (3.38) respectively as the unique solution of the system of equations (3.36).

$$\begin{cases} R_{cal1} = R_0 + \alpha \cdot S_R \cdot N_{cal1} \\ R_{cal2} = R_0 + \alpha \cdot S_R \cdot N_{cal2} \end{cases} \tag{3.36}$$

$$\alpha = \frac{1}{S_R} \cdot \frac{R_{cal2} - R_{cal1}}{N_{cal2} - N_{cal1}} \tag{3.37}$$

$$R_0 = \frac{R_{cal1} \cdot N_{cal2} - R_{cal2} \cdot N_{cal1}}{N_{cal2} - N_{cal1}} \tag{3.38}$$

3.5 Measurement Setup and Experimental results

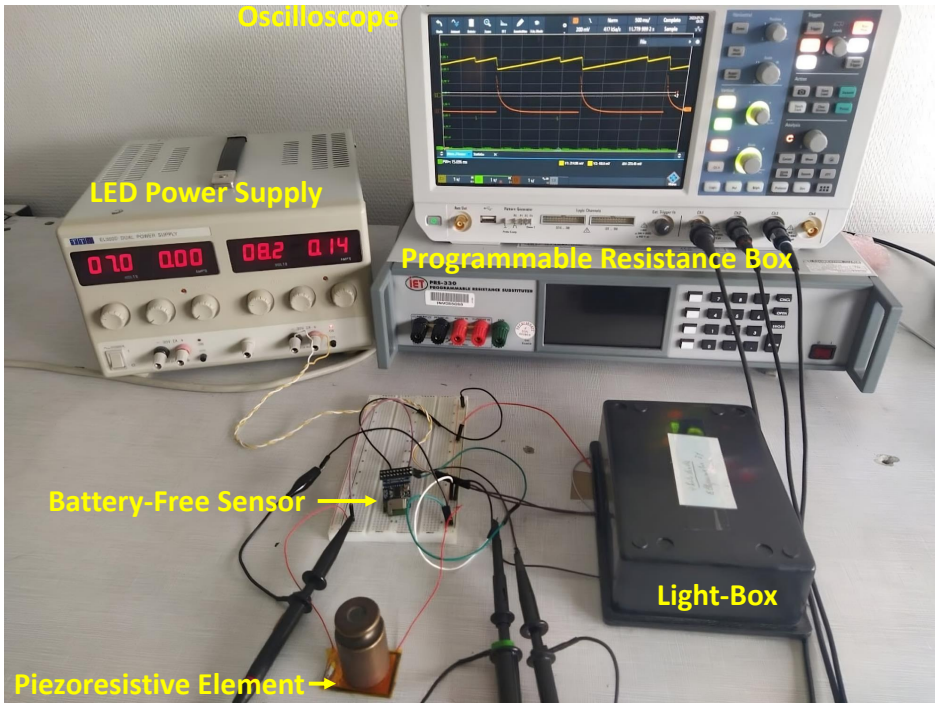
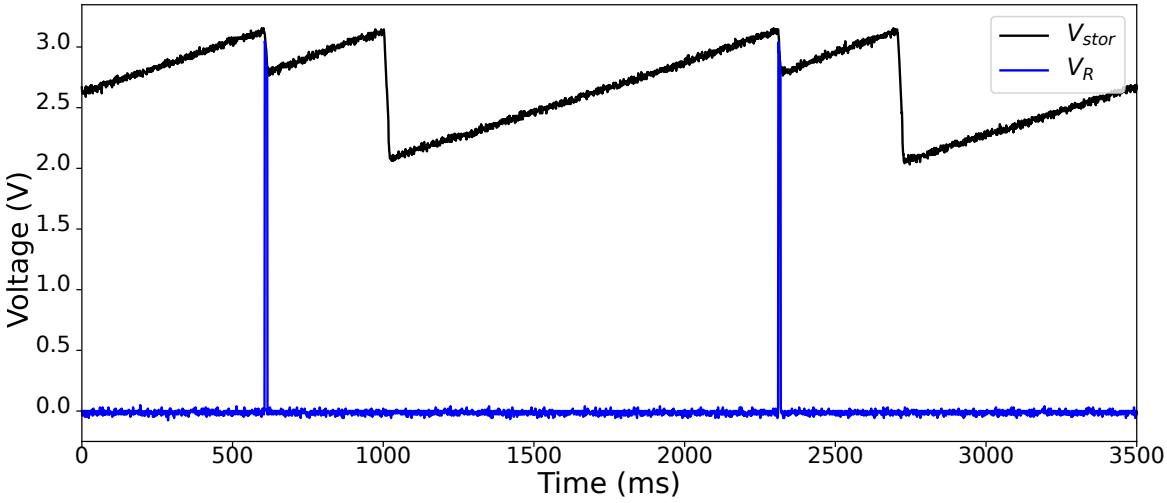


Figure 3.9: Experimental setup.

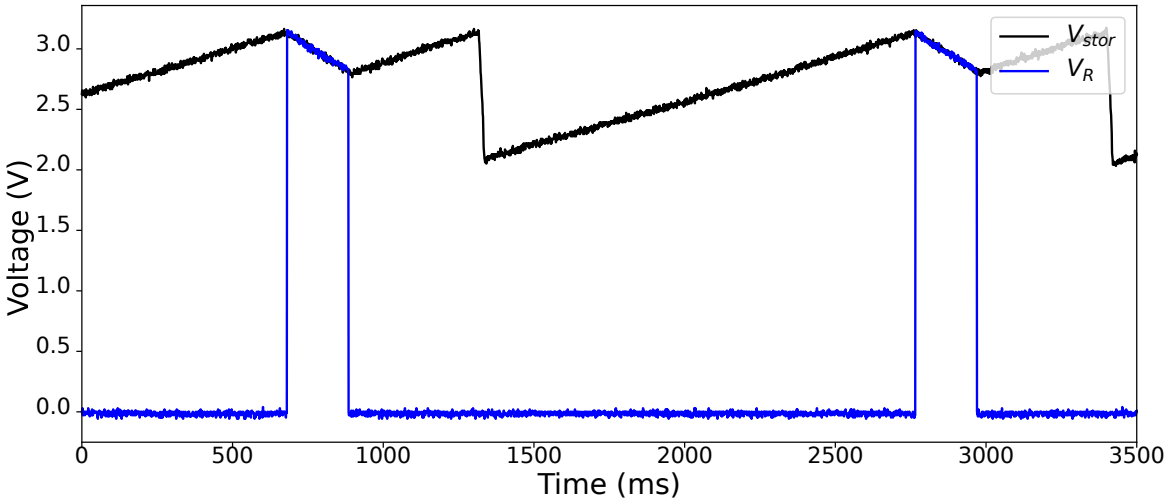
3.5.1 System validation

Several tests were conducted to measure the electrical resistance of the piezoresistive sensor in the EAWPSN, as explained in section 3.5.2. Figure 3.9 shows the experimental setup and highlights its most important parts. Additionally, the energy harvester and an LED, used as a light source, were contained within the light-box to achieve a controlled and repeatable light environment. The measurement data were simultaneously captured via Bluetooth using a computer connected to a remote BS and directly probing the signals V_R and V_{stor} on the EAWPSN using an oscilloscope. The TDDC resistance measurement technique was initially tested and calibrated with various resistance values using the PRS-330 Precision Programmable Resistance Box [223]. Later, the piezoresistive sensor was connected to the system and measured with different masses, as shown in subsection 3.5.2.

Fig. 3.10 shows the evolutions of V_{stor} and V_R during two reading cycles of time t_m for two sample resistance values: 0.4 kΩ, 6 kΩ. These two values were chosen because they belong to the range of resistance values obtained from piezoresistive sensor during



(a)



(b)

Figure 3.10: Evolution of the voltage V_{stor} and V_R vs. time, during the time readings t_m for two different sample resistance values: (a) 0.4 $k\Omega$, (b) 6 $k\Omega$.

static characterization. In this case, the measured times t_m were respectively 13.5 ms for 0.4 kΩ (Fig. 3.10a) and 200 ms for 6 kΩ (Fig. 3.10b). These results were obtained when the harvester was subjected to a controlled lighting condition with a luminous power of 1900 lux, along with the setup parameters summarized in Table 3.2. Under these conditions, the operating range of the EAWPSN is between 40 Ω and about 15 kΩ according to formulas 3.26 and 3.14. In particular, when substituting the parameters from Table 3.2 in 3.26, an approximate value of 1.2 Ω is obtained. As a result, the value $R_{min} = 40\,\Omega$ is selected to ensure strict compliance with the inequality (3.26). Furthermore, this value represents the Limit of Detection (LoD) of the EAWPSN.

Table 3.2:
Setup parameters

Parameter	Value
C_{stor}	440 μ F
V_H	3.15 V
V_L	2.00 V
$V_{L,R}$	2.85 V
f_{clk}	37 kHz
P_{geh}	1 mW
Light Intensity	1900 lux

Table 3.3:
Calibration parameters

Parameter	Value
R_{cal1}	400 Ω
R_{cal2}	10000 Ω
$t_{m,cal1}$	13.5 ms
$t_{m,cal2}$	351 ms
N_{cal1}	499
N_{cal2}	12980
α	1.25
R_0	16.36 Ω

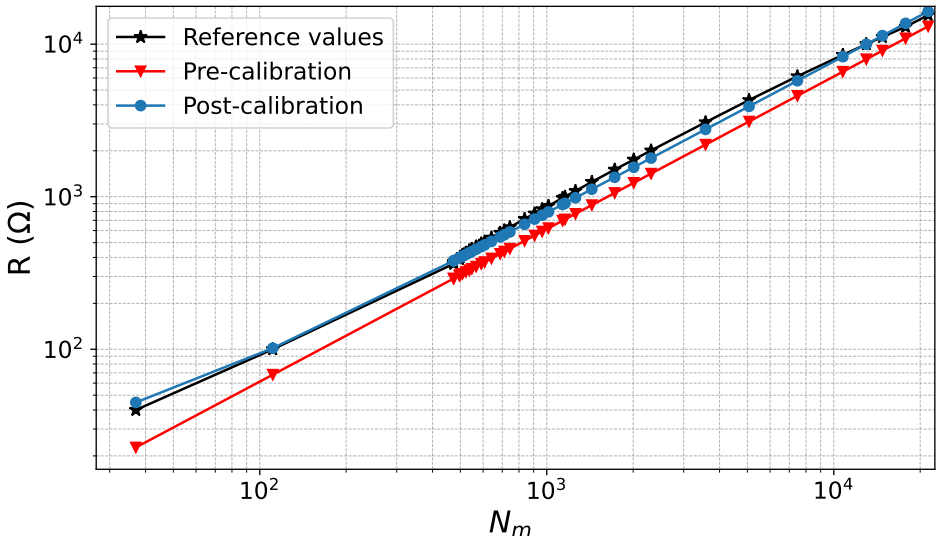


Figure 3.11: Resistance values as a function of the number of clock pulses.

The table 3.4 compares the reference resistance values and those calculated by EAWPSN using the linear approximation as a function of digital time t_m before and after calibration. The relative error (R_{error}) on the resistance estimation was calculated using (3.39).

$$R_{error}(\%) = 100 \cdot \frac{R - R_{ref}}{R} \tag{3.39}$$

Table 3.4 also displays the number of clock pulses N_m counted by the LPTIM timer to obtain the measured digital time t_m. According to the findings, WSN reading accuracy is significantly improved by calibration. Before calibration, percentage errors were considerable, with values as high as 43 %. After calibration, the percentage errors notably decrease, with 95 % of values falling below 10 %. This affirms that calibration enhances the accuracy of WSN. The results demonstrate the importance

Table 3.4: Comparison between the reference resistance values and those calculated by the WSN through linear approximation as a function of digital time t_m, before and after calibration.

			Pre-calibration		Post-calibration	
$R_{ref}\,(\Omega)$	$t_m\,(ms)$	N_m	$R\,(\Omega)$	$R_{error}\ (\%)$	$R\,(\Omega)$	$R_{error}\ (\%)$
40	1.0	37	22.7	43.2	101.6	-11.9
100	3.0	111	68.2	31.8	379.5	-1.6
363	12.8	473	290.5	19.96	399.5	-4.6
400	13.5	499	306.5	23.37	417.9	0.1
430	14.2	523	321.2	25.29	439.4	2.8
455	14.9	551	338.4	25.62	473.2	3.4
494	16.1	595	365.5	26.02	510.1	4.2
539	17.4	643	395.0	26.72	564.6	5.4
602	19.3	714	438.6	27.15	663.6	6.2
715	22.8	843	517.8	27.58	760.4	7.2
828	26.2	969	595.2	28.11	797.2	8.2
871	27.5	1017	624.7	28.28	910.9	8.5
1000	31.5	1165	715.6	28.44	1123.6	8.9
1250	39.0	1442	885.8	29.14	1345.4	10.1
1505	46.8	1731	1063.3	29.35	1558.9	10.6
1750	54.3	2009	1234.0	29.48	1791.5	10.9
2010	62.5	2312	1420.2	29.35	2765.9	10.9
3075	96.8	3581	2199.6	28.47	3913.8	10.1
4300	137.2	5076	3117.9	27.49	5755.0	8.98
6155	202.0	7474	4590.9	25.41	8271.9	6.5
8500	290.6	10752	6604.4	22.3	9988.0	2.7
10000	351.0	12987	7977.3	20.23	11351.6	0.12
11150	399.0	14763	9068.2	18.67	13681.1	-1.81
13000	481.0	17797	10931.8	15.91	16380.0	-5.24
15500	576.0	21312	13090.9	15.54	16380.0	-5.68

of this calibration process in achieving more reliable and accurate TDDC conversions. This phenomenon is evident in Fig. 3.11, which compares the resistance values measured using the TDDC technique before and after calibration with reference values (black line). In particular, it emerges that, in the absence of calibration (red line), the TDDC technique tends to underestimate the resistance values, while when calibration is applied (blue line), the R values are very close to the reference values. Furthermore, in Fig. 3.11, it is possible to observe and appreciate the linearity of the sensor through its correlation with the number of pulses (N_m) counted by the LPTIM timer. Fig. 3.12 displays the percentage error of the proposed TDDC technique, representing the error in the WSN's operating range between the estimated resistance value to the BS and the reference resistance values connected to the EAWPSN.

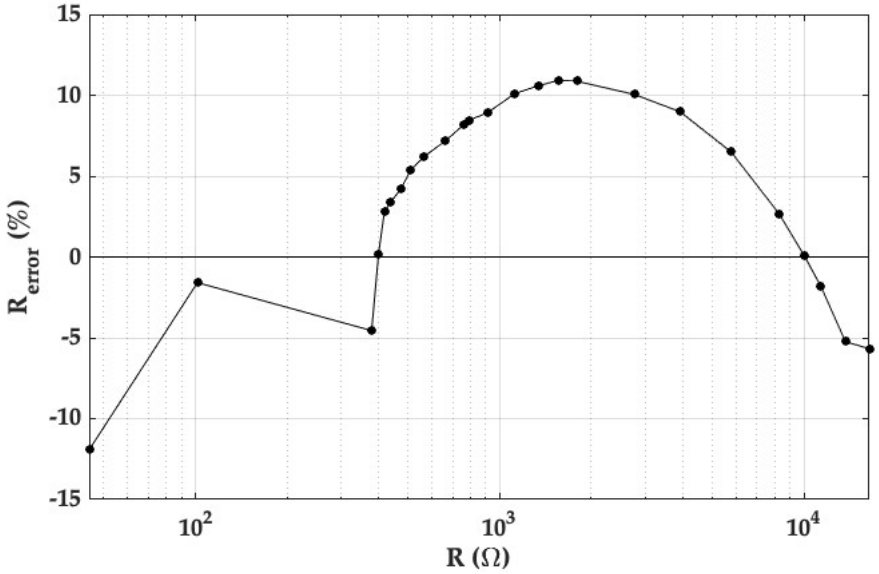


Figure 3.12: Percentage error in resistance estimation within the operational range of the EAWPSN.

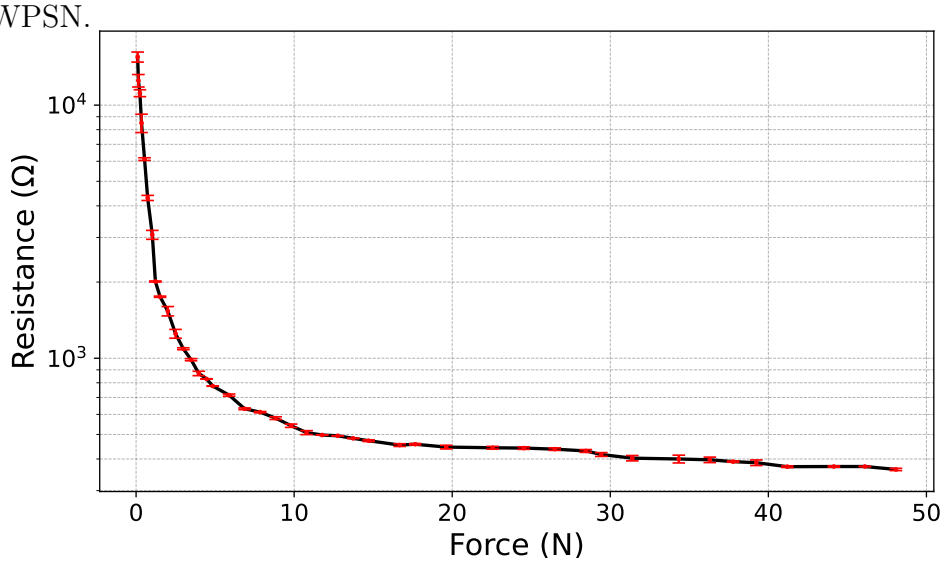


Figure 3.13: Static characterization of multilayer sensors with the ACF adhesive (error bars representing min and max).

3.5.2 Static characterization of the piezoresistive sensor

To obtain a static characterization of the relationship between resistance (R) and force (F) for the piezoresistive material, we performed the following procedure on the sensor:

- Connected the piezoresistive sensor to the Digital Multimeter M-3650, which has an accuracy of ±0.5 % of the reading + 3 digits.
- Mechanically loaded the piezoresistive sensor with a test mass.

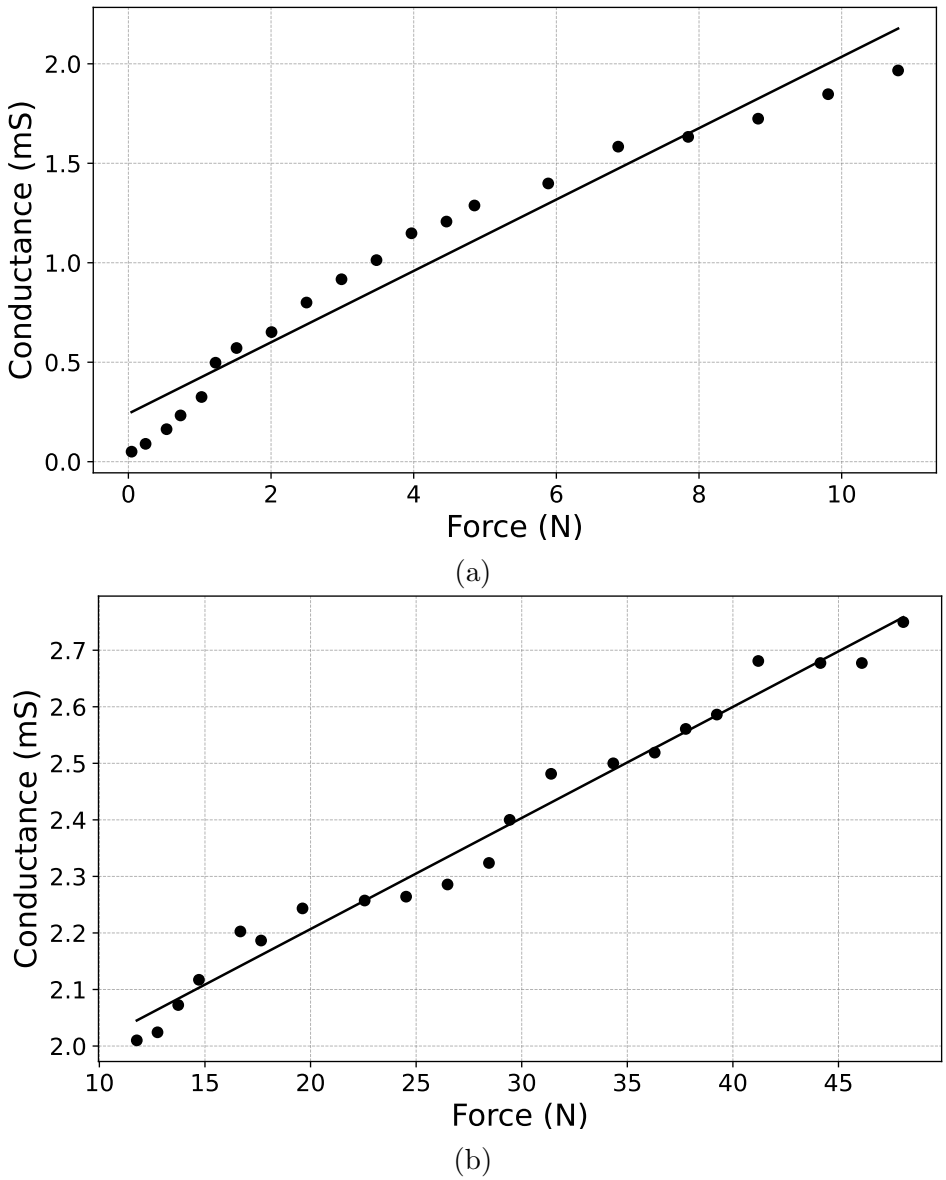


Figure 3.14: Conductance as a function of the force applied to the piezoresistive sensor. Black line is the best linear fit.

- Let the system rest for 4 min to allow any transient effect to dissipate.
- Measured the resistance value and plotted it versus the applied force.

We repeated the procedure for various test masses ranging from 15 grams to approximately 5 kilograms, and we calculated the dispersion and mean values, as shown in Fig. 3.13. The system showed lower accuracy when the weakest force was applied because, in this condition, the piezoresistive sensor is mechanically less stable, and specifically for this condition, the standard deviation was 0.70 kΩ, and the average

measured resistance value was 15.5 kΩ. However, the standard deviation decreased to 10 Ω for an applied force of 1.22 N and dropped further to a few ohms when the applied force exceeded 3 N. The minimum measured resistance value was 363 Ω with a standard deviation of only 3.78 Ω when a force of 48 N was applied to the sensor.

In addition to analyzing resistance values, the data were also explored in terms of conductance (G) to provide further insight into the static behavior of the piezoresistive sensor. In particular, this analysis revealed a limit of detection (LoD) in terms of conductance equal to 64 μS and a linear relationship between conductance values and the applied force (F). Furthermore, two regions with different sensitivities were identified. Specifically, for forces lower than 10 N, the sensitivity (dG/dF) was found to be 0.194 mS/N (Fig. 3.14a); while for forces greater than 10 N, dG/dF was equal to 0.02 mS/N (Fig. 3.14b).

3.5.3 Force-to-time-to-digital conversion results

This section presents the experimental results of the WSN readout based on force-to-time-to-digital conversion. Fig. 3.15 shows the relationship between the number of pulses counted (N_m) by the LPTIM using the TDDC technique and the different applied forces. Observations indicate that the curve has a monotonic trend, which allows for a linear approximation.

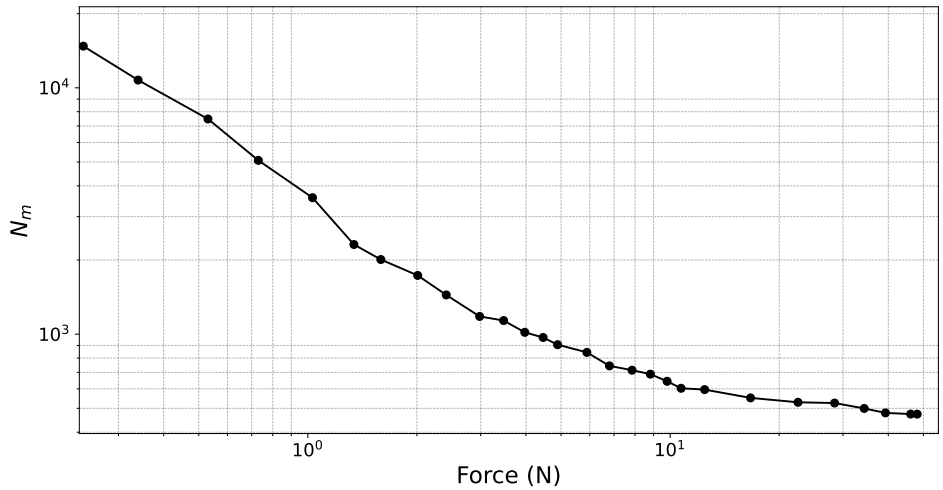


Figure 3.15: Experimental measurements of N_m vs applied force (log scale).

Since N_m and applied forces are inversely proportional, to further explore the linearity of the force-to-time-digital conversion method and understand the WSN's sensitivity, it is interesting to observe the relationships between $1/N_m$ and the applied forces, as depicted in Fig. 3.16. The analysis of the results revealed three distinct

sensitivities in the relationship between $1/N_m$ and F. As a result, the experimental measurement interval of applied forces, which spans from 0.25 N to 48.07 N, is divided into three sub-intervals characterized by the following linear relationships between force and $1/N_m$.

$$F = \frac{3.957 \cdot 10^3}{N_m} - 0.048 \quad \text{with } F \in [0.25\,\text{N} - 2.55\,\text{N}] \tag{3.40}$$

$$F = \frac{8.767 \cdot 10^3}{N_m} - 4.005 \quad \text{with } F \in [2.55\,\text{N} - 13.0\,\text{N}] \tag{3.41}$$

$$F = \frac{1.156 \cdot 10^5}{N_m} - 195.3 \quad \text{with } F \in [13.0\,\text{N} - 48.07\,\text{N}] \tag{3.42}$$

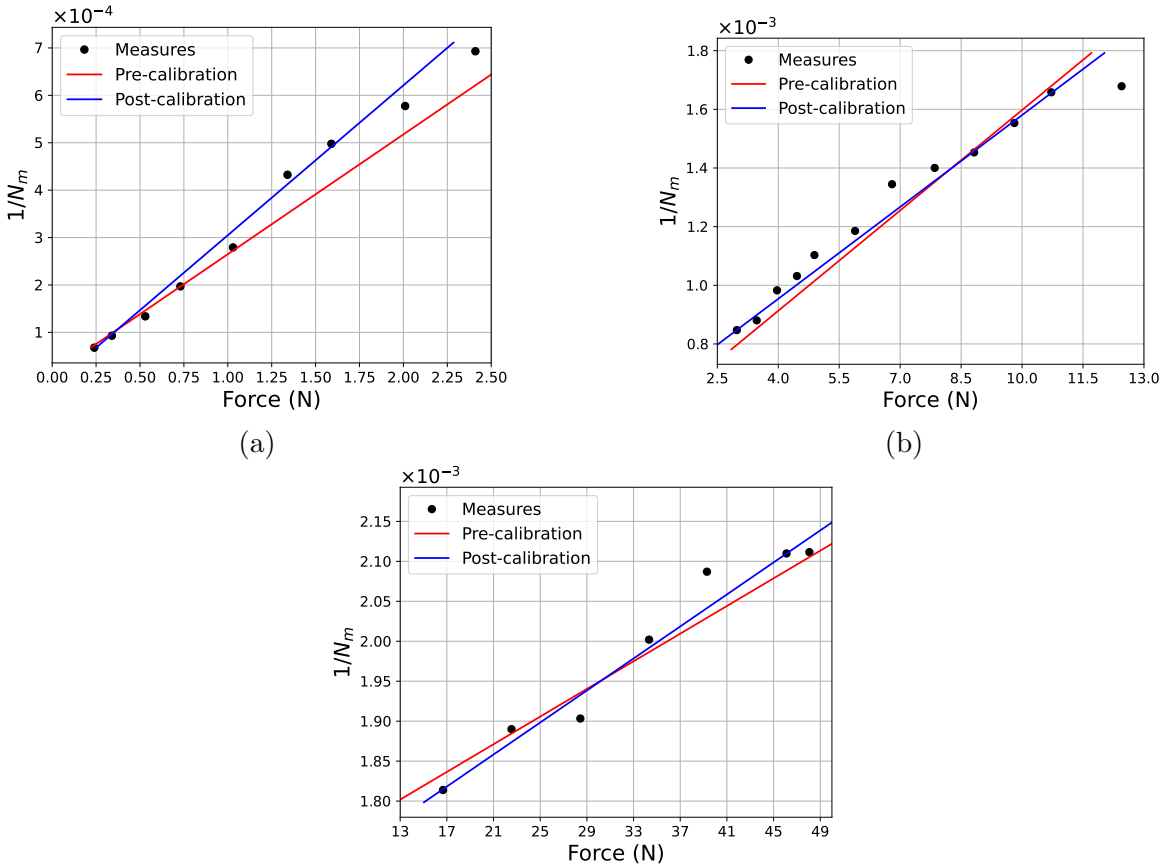


Figure 3.16: Experimental results of $1/N_m$ vs applied force and its linear fits before (red line) and after calibration (blue line).

Table 3.5: Comparison of reference forces and experimental results of the force-time-digital conversion method.

			Pre-calibration		Post-calibration	
F_{ref} (N)	t_m (ms)	N_m	F (N)	F_{error} (%)	F (N)	F_{error} (%)
0.25	**399.0**	**14763**	0.22	-13.6	0.25	0.0
0.30	351.0	12987	0.26	-15.4	0.28	-7.1
0.34	290.6	10752	0.32	-6.3	0.33	-3.0
0.72	131.2	4854	0.77	6.5	0.69	-4.3
1.03	90.8	3359	1.13	8.8	0.98	-5.1
1.37	63.5	2349	1.64	16.5	1.38	0.7
1.61	**54.3**	**2009**	1.92	16.1	1.61	0.0
2.01	46.8	1731	2.24	10.3	1.86	-8.1
2.45	38.0	1406	2.77	11.6	2.29	-7.0
2.98	**31.9**	**1180**	3.43	13.1	2.98	0.0
3.97	27.5	1017	4.62	14.1	4.28	7.2
4.46	26.2	969	5.04	11.5	4.75	6.1
5.89	22.8	843	6.40	8.0	6.22	5.3
7.85	19.3	714	8.28	5.2	8.28	5.2
9.81	17.4	643	9.63	-1.9	9.76	-0.5
10.72	**16.3**	**604**	10.51	-2.0	10.72	0.0
12.45	15.9	588	10.91	-14.1	11.15	-11.7
16.67	**14.9**	**551**	14.53	-14.7	16.67	0.0
22.54	14.4	534	21.21	-6.3	22.44	-0.4
28.45	14.2	525	24.93	-14.1	25.64	-10.0
34.33	13.5	499	36.40	5.7	35.55	3.4
39.29	12.9	479	46.08	14.7	43.91	10.5
46.11	**12.8**	**474**	48.62	5.2	46.11	0.0
48.10	12.7	469	51.22	6.1	48.35	0.5

Table 3.5 compares the reference forces (F_{ref}) with the results obtained by force-time-digital conversion before and after calibration. The data provided include the measured time (t_m), the number of pulses counted (N_m), the force calculated from the empirical relations (F) in Newtons and the percentage error (F_{error}) concerning the reference force. Some results before calibration, for example, those associated with forces around 1.37 N and 3.97 N, displayed low accuracy, with percentage errors exceeding 14%. The low accuracy observed for specific applied forces suggested the need for calibration. Consequently, the same calibration technique used for the resistance measurement was applied to the force estimation using the three linear regression equations. The empirically derived linear relationships in the three force sub-intervals can be expressed in the following form: $F = F_0 + \frac{\alpha_F \cdot S_F}{N_m}$; where S_F was equal to 3957, 8767, and $1.156 \cdot 10^5$ for 3.40, 3.41, and 3.42 respectively. To perform the calibration,

reference force pairs were selected in the three sub-intervals. Finally, for each value of reference force (highlighted in bold in Table 3.5), the associated N_m were measured to calculate α_F and F_0 as follows.

$$\alpha_F = \frac{N_{cal1} \cdot N_{cal2}}{S_F} \cdot \frac{F_{cal2} - F_{cal1}}{N_{cal1} - N_{cal2}} \tag{3.43}$$

$$F_0 = \frac{F_{cal1} N_{cal1} - F_{cal2} N_{cal2}}{N_{cal1} - N_{cal2}} \tag{3.44}$$

The results suggest that calibration is required to ensure high accuracy of the TDDC method over a wide range of forces. Indeed, after the calibration, the percentage errors decrease, with many values below 10%.

3.6 Piezoresistive sensors for Smart Agriculture Applications

This section briefly discusses some applications of piezoresistive sensors in agriculture found in the literature and also provides an idea for strain measurements based on a piezoresistive element for structural monitoring in agriculture scenarios.

The work [202] presented a wireless sensor system for piezoresistive-based soil moisture measurement for precise irrigation scheduling in tomato and chrysanthemum fields. The developed sensor was compared with values obtained from a soil tensiometer, an instrument used to measure soil matrix potential. The maximum coefficient of determination (R^2) observed was 0.84 in tomato and 0.83 in chrysanthemum. Liu et al. [203] proposed a method for easily and inexpensively modifying the interior of corn stover pith with aqueous solutions to convert it into a piezoresistive material. In this manner, the material exhibited a stable resistance change signal in the presence of external pressure, and in particular, the resistance decreased with increasing pressure. The study represents a green and sustainable solution to replace traditional sensors with biobased materials obtained from agricultural waste.

Article [204] demonstrated the use of piezoresistive microcantilevers for soil nutrient monitoring in precision agriculture. The sensor was capable of detecting nitrates, potassium, and tributyltin chloride for phosphate detection. The presence of the macronutrient caused the resistance of the microcantilever to change, and this change in resistance was measured by highly sensitive electronics.

Paper [205] proposes a piezoresistive microcantilever capable of detecting low concentrations of pathogens in the early stages of infection, enabling early detection and

thus saving crops. The developed sensor is highly sensitive for detecting the micromolar concentration of Ralstonia solanacearum antigen, a disease-causing bacterium in tomatoes, potatoes, tobacco, ginger, eggplants, and bananas. By exploiting the piezoresistive nature of the device, it is possible to indirectly determine the surface stress-induced deflection caused by the pathogen. This deflection causes a resistance change of 0.6% for the piezoresistor, which is measured using a 4-terminal configuration. However, the works reported in this subsection did not address solutions from both the energy consumption and wireless technology perspectives. Whereas the system described in this chapter is energy autonomous, battery-free, and the TDDC resistance measurement technique can be employed in the agricultural use cases presented in the literature. Additionally, the EAWPSN can be used for diagnosing the health status of agricultural structures and vehicles, such as barn sheds, greenhouses, silos, tractors, etc. Indeed, agricultural facilities and equipment are susceptible to environmental degradation due to frequent exposure to water, strong wind, agricultural chemicals and structural loading, so intelligent monitoring is necessary to mitigate subsidence in agricultural facilities. Due to their ability to sense mechanical stress and strain, piezoresistive sensors are used in a wide range of IoT applications for structural monitoring, including automotive, industrial, construction, sports, and in agricultural scenarios [215, 206]. However, compared with the state of the art, in this work we introduce a sustainable IoT solution with wireless and battery-free communication. In addition, experimental evidence (Figure 3.16) shows the strain measurements with the TDDC technique. An ultra-low power measurement technique that employs only 396 microjoules of energy (from formula 3.6), which is ideal for large-scale, sustainable, and maintenance-free monitoring in remote agricultural environments.

Moreover, the EAWPSN is capable of functioning in indoor environments, as it can operate with a minimum ambient light intensity of 200 lux. For this reason, the coverage distance of the sensor was evaluated in an indoor scenario, as shown in Fig. 3.17. In particular, the figure displays the values of the receiver signal strength (RSS) obtained in our experiment and the logarithmic trend (dashed line). In this experiment, a coverage of 64.2 meters was achieved, with the RSS at the last acquisition point being above -80 dBm. Regarding the coverage of the BLE module in outdoor environments, it maintains reliable communication with the BS within a maximum coverage range of 160 meters in an environment characterized by foliage scattering, as described in the study [182]. However, in the case of line-of-sight transmission, a maximum coverage of 1 km is obtained with an RSSI of approximately $-100\,dBm$ and

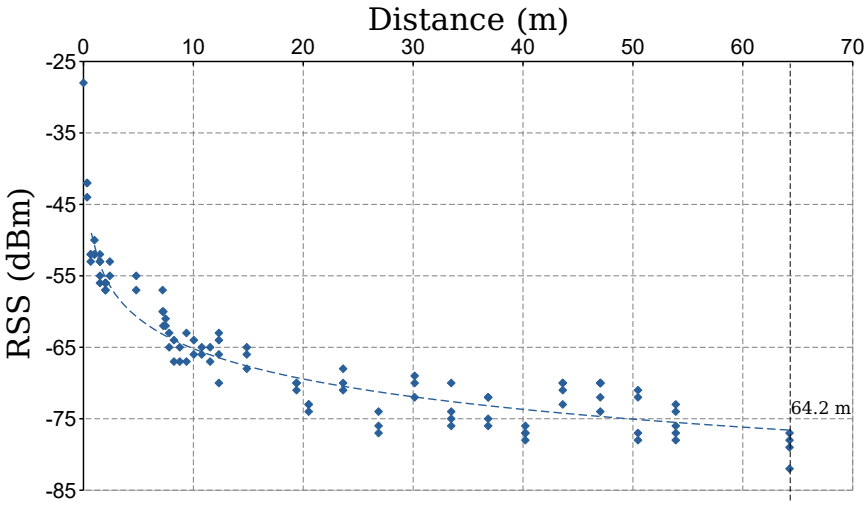


Figure 3.17: Received Signal Strength (RSS) in relation to the distance between the EAWPSN and the base station for an indoor scenario.

a Packet Error Rate (PER) close to 0%, as described in the technical report of the wireless module [224].

3.7 Summary

In this chapter, we have successfully developed and evaluated an energy-autonomous wireless piezoresistive sensor with integrated Time Domain to Digital Conversion (TDDC) capabilities. The proposed solution lays the foundation for widespread deployment in mass-scale agricultural IoT scenarios. The WSN's energy harvesting system efficiently converts ambient light into electrical energy, enabling self-sustaining and maintenance-free operation without needing an external power source, such as batteries. The piezoresistive device showed high sensitivity and accuracy in strain measurements, making it a reliable choice for various monitoring agricultural applications. The integration of TDDC proved to be very effective in reducing power consumption, as it eliminated the need for power-hungry and costly analog-to-digital converters. The time conversion into digital values enhanced the WSN's energy efficiency and simplified the signal processing pipeline, resulting in a more compact and cost-effective IoT design. The BLE wireless data transmission module supplied seamless data transmission to a central Base Station (BS), making the WSN suitable for remote and inaccessible environments. The combination of energy autonomy and wireless connectivity makes it an ideal solution for many agricultural applications,

such as real-time structural health monitoring, and more, where continuous, long-term data collection is crucial for proactive maintenance and fault detection. In conclusion, the energy-autonomous wireless piezoresistive sensor with integrated TDDC capabilities presented in this study has the potential to revolutionize the field of agricultural sensing and enable new applications that were previously impossible due to power constraints. Further research could explore optimization to expand the IoT sensor's capabilities and applications in wireless sensing and agriculture monitoring.

Chapter 4

Adaptive Algorithms for Battery-less LoRa Sensors in Agricultural Settings

4.1 Overview

Ambient energy-powered sensors are becoming increasingly crucial for the Internet-of-Things (IoT) sustainability in agriculture. In particular, battery-less sensors are a cost-effective solution that requires no battery maintenance, lasts longer and has greater weatherproofing properties due to the lack of a battery access panel. In this chapter, adaptive transmission algorithms will be studied to enhance the performance of battery-less IoT sensors based on the LoRa protocol. First, we characterize the device power consumption during sensor measurement and/or transmission events. Next, we explore various rural agricultural scenarios and dynamically adjust key network parameters, including inter-packet transmission time, data redundancy, and packet size, to optimize device performance. We design appropriate capacity-based storage, considering a renewable energy source (e.g., photovoltaic panel), and we analyze the probability of energy failures by exploiting both theoretical models and real energy traces. Finally, the results can be used as feedback to re-design the device with appropriate energy storage and meet certain reliability constraints.

4.2 Introduction

The rise of the Internet-of-Things (IoT) demands for efficient and sustainable power sources, and therefore ambient energy-powered sensors have attracted significant attention due to their environmental and economic benefits. By capturing, converting,

and storing energy from the immediate environment (such as solar radiation, thermal gradients, and mechanical vibrations), autonomous sensors are opening new perspectives in different fields, especially where power supply from the grid is absent, e.g., in agriculture [225, 182]. Moreover, new battery-less sensors offer numerous advantages compared to conventional alternatives, making them an attractive choice for IoT applications. Indeed, a battery-less design allows powering IoT devices with improved cost-effectiveness, recyclability, longevity, and weatherproofing, due to the elimination of batteries. Battery-less sensors find application in a wide range of IoT scenarios, such as agriculture (e.g. soil monitoring systems), healthcare (e.g. wearable or intra-body devices), and smart cities (e.g. solar-powered streetlights and traffic signals, surveillance cameras, etc.). As IoT technologies are becoming increasingly integrated into daily life, adopting battery-less sensors can significantly extend device lifespan, reduce carbon footprints and contribute to a more environmentally conscious society.

In general, an energy-autonomous system combines energy harvesting (EH) equipment with an energy storage system to ensure the continuity of the supply. The total cost depends on the level of reliability required: a lower probability of failure results in a higher cost. The device sensors and the data transmission system represent a variable load whose supply must be provided by the harvesting and storage sub-system [226], [227]. Indeed, the energy autonomy requirement implies the correct sizing and management of these three components, i.e., sources, load, and storage. On the one hand, the size of harvesting and storage cannot exceed certain limits (mainly due to cost and size issues); on the other hand, it is necessary to avoid energy failures and to ensure the transmission of sensor data [228]. Providing solutions to avoid the use of batteries by harvesting energy from the environment would encourage the deployment of IoT devices. The design of battery-less systems is therefore a complex task which includes the correct sizing of both the harvesting and of the storage systems and the energy management optimization. Many energy sources can be considered in the design: e.g., solar cells, vibration-based systems, thermoelectric, and solar thermoelectric methods [229].However, the highest power density (about 15 mW/cm^2) is provided by solar cells [230], and for this reason, in this chapter, we focus mainly on this energy source, which is also ideal for smart agriculture. For energy storage, rechargeable batteries or supercapacitors are usually adopted. Batteries offer a higher energy density compared to supercapacitors; however, supercapacitors have very low internal impedance, allowing higher current pulse without efficiency detrimental [231]. In the charge/discharge process, energy flows twice in the same circuit path of the accumulator, and this consideration suggests moving toward low-impedance storage

systems, such as super-capacitors, to improve the efficiency of the storage system. Moreover, battery-less sensors are more cost-effective and recyclable, last longer, require no battery maintenance, and have greater weatherproofing due to the lack of a battery access panel [232].

These benefits motivate the present work. However, there are still several problems and hurdles that must be tackled. For instance, energy harvesting power sources with low and intermittent output, energy storage capacitors, wireless interference, and intermittent random access transmissions can frequently cause power and/or communication failures.

In the literature, various energy-aware transmission algorithms have been proposed to optimize the performance of battery-free IoT sensors [233–236]. Indeed, to maximize the efficiency and reliability of these sensors, adaptive transmission algorithms play a crucial role [237, 238]. The aim is to balance energy consumption with communication requirements, considering parameters such as transmission power, data rate, modulation scheme, and duty cycle [239, 240]. By adapting the data transmission to the available energy and specific application needs, these algorithms improve the overall performance, reliability, and lifespan of battery-free IoT devices. For instance, in [233], it is possible to observe the implementation of specific wake-up policies through optimized algorithms for the best sampling frequency implementation based on the power received from the source. Furthermore, the use of adaptive algorithms allows addressing the challenge of incorporating intelligence into small battery-free sensors subject to constraints of limited energy resources and dynamically adapting computation conditions based on the unpredictable nature of harvested energy [241–243]. Adaptive protocols have been proposed in [244] to promote the coexistence of different battery-free devices, with varying transmission requirements, to dynamically allocate transmission slots for 40 devices without requiring prior knowledge of the environment. Furthermore, adaptive algorithms can decide whether to use the harvested energy immediately for transmission or to store it for future communications. This approach optimizes energy management and ensures efficient resource allocation, reducing, for example, transmission latency [235] and improving coexistence [244]. All this further highlights the importance of adaptive algorithms for battery-free IoT sensors.

In this chapter, we will analyze two commercial LoRa devices: a FiPy module equipped with Pytrack expansion board and a TTGO T-Beam ESP32 board. These devices are based on the ESP32 System on Chip (SoC) and feature an SX1276 LoRa transceiver. We believe that an approach based on real measurements is more realistic than simulation-based studies. Moreover, to the best of our knowledge, we are the

first to optimize specific network parameters for improved energy management, such as packet size or payload redundancy, in the context of battery-less LoRa devices. We thus propose a general approach that models both the energy source and storage, together with optimized energy management for the transmission system, tuned in real-time to comply with the available energy. Finally, results are verified on a realistic rural scenarios, analyzing different performance requirements such as the loss of energy probability.

- an architectural design and energy consumption analysis of the battery-less device, based on two commercial LoRa transmitters;
- the optimization of data transmission using adaptive scheduling and redundancy schemes;
- the validation of our model with a real dataset, evaluating the probability of energy failure and cost analysis for energy storage and harvesting.

The rest of the chapter is organized as follows: section 4.3 presents state-of-art of battery-less IoT sensors, and section 4.4 describes the general architecture of the proposed system and models the device energy consumption, including generation and optimization issues. Results are presented in section 4.5, varying storage size and analyzing the energy production in different seasons of the year. Finally, section 4.6 discusses the performance and section 4.7 concludes the chapter.

4.3 Background and Related work

In this section, background information and relevant literature are presented, focusing on the development of battery-less devices and the optimization of transmission parameters based on available energy from EH sources.

4.3.1 Battery-less devices for IoT

Making a device completely energy-neutral requires a thorough analysis of power consumption in different working states [245]. Several works developed theoretical models for battery-less devices using emulated environments [245–247]. For example, Delgado et al. [245] provide a Markov model to characterize the performance of battery-free LoRaWAN devices for uplink and downlink (UL/DL) transmissions and assess their performance in terms of the model parameters (i.e., device configuration, application

behavior, and environmental conditions). The study demonstrates that a 47 mF capacitor can handle 1 Byte SF7 transmissions every 60 s at a 1 mW energy harvesting rate. Indeed, the work shows that battery-free LoRaWAN communications are possible with the correct setup (i.e., capacitor size and turn-on voltage threshold) for various application behavior (i.e., transmission interval, packet sizes, energy harvesting rate). Furthermore, in [247], the effectiveness of battery-free LoRa networks powered by ambient EH sources has been studied, assuming random transmission schemes. By using methods from stochastic geometry and Markov chain analysis, a mathematical model for each of the system's components was built, and the likelihood of an energy and communication outage was analytically computed. The study has shown that LoRa networks' adaptive data rate (ADR) can result in energy outages when employing higher spreading factors, and suggests adaptive charging time schemes as a successful remedy. In [246], authors investigated the optimal parameters to schedule application tasks on battery-less IoT sensor devices. Using an environment emulator and a SODAQ ExpLoRer board, the authors validated a mathematical model for choosing the optimal parameters, in terms of minimum application cycle completion time, at which to perform sensing and transmission, considering different device and environmental conditions. The analysis shows that a device using LoRaWAN Class A, equipped with a capacitor of 10 mF, can measure the temperature and transmit its data at least once every 5 s, and can harvest at least 50 mW (10 mA of current). Finally, the work in [239] developed an energy-aware system model to operate battery-free IoT devices that include several wireless communication protocols. To assess the total energy efficiency of the IoT network, simulations based on a probabilistic sensing model are used. According to the results, to achieve self-sustainability in a heterogeneous short- and long-range network and enhance energy efficiency, an energy harvesting device combining a solar panel with a 270 F lithium-ion super-capacitor must be utilized as a power storage device.

However, these theoretical results, and in particular [245, 247], might be overly optimistic in terms of inter-arrival times between packets and the size of the storage capacitor, requiring to be confirmed in real-world implementation. The studies [248, 249] proposed some prototypes of battery-less nodes based on LoRa technology. Specifically, Orfei et al. [248] demonstrated the performance of a battery-less sensor for monitoring road traffic and bridge conditions, powered by a low-cost electromagnetic EH device, which employs an array of permanent magnets to improve energy efficiency. The collected energy is stored in a supercapacitor and powers an ARM Cortex M0+ microcontroller and a LoRa radio module to transmit information. On

the other hand, Boitier et al. [249] introduced a self-contained LoRa sensor with a photovoltaic power source and a pair of 25 F supercapacitors for energy storage. This solution assures 11 days of storage life in the absence of light. The proposed system also includes a circuit for energy management and troubleshooting on the first startup.

4.3.2 Energy management optimization

One of the challenges for battery-less IoT devices is their limited energy availability and reliance on the surrounding environment. If a device runs out of energy, it cannot perform its functions until the harvesting system recharges its energy storage [107]. Generally, the main parameters to optimize energy consumption on wireless nodes are the data transmission interval T (i.e., the time between two consecutive packets) and data overhead/redundancy NR (which can be implemented in several ways, e.g., through coding or transmission repetition). Therefore, a good policy for choosing the periodic interval to transmit information and tuning the amount of data redundancy is of paramount importance. Due to the imperfect predictability of real-world events that affect energy sources, the optimization strategy must strike a balance between capturing these events and consuming all available energy.

To achieve this, recent studies have suggested using artificial intelligence (AI) to learn such policies for battery-free sensors. In particular, it has been shown that data-driven strategies, such as reinforcement learning, can be exploited because the amount of available energy changes in patterns similar to other close environments. [250, 251, 240, 252].

In other words, energy availability could be predicted to make better decisions implementing proper energy resource management strategies in battery-less devices [237, 253]. However, the unpredictable nature of the energy sources requires large datasets to train these AI-based systems and optimize the device parameters in relation to the energy storage process. Moreover, these strategies should be implementable on low-complexity hardware, which is usually different for AI algorithms.

In [238], several scheduling algorithms are applied to battery-less LoRaWAN nodes, analyzing their performance in various simulated scenarios. Based on real-world EH measurements gathered from a testbed, the work studies the impact of energy-aware schemes on the number of transmitted packets and the mean packet interval. The results demonstrate that energy-aware algorithms can significantly enhance the performance of battery-less LoRaWAN nodes. However, the presented results are strongly influenced by the harvesting capabilities of the nodes. The works [254, 255] suggest a simple approach for network optimization, exploiting a revised sigmoid function that

can be easily computed on low-cost hardware. The operating strategy can be adjusted, based on a small subset of the most recent energy information or even the last two samples. Specifically, these methods reduce energy consumption by modifying the sample rate based on the remaining battery level [254] or the harvested ambient energy [255]. In particular, authors in [255] developed an objective function to optimize the transmission period T, with dynamic sampling adaptation schemes which can be classified into two possible categories: a) Threshold-based sampling adaptation (T-ASA), which basically adapts the sampling time based on energy thresholds, or b) Data-Driven Adaptive Sampling Algorithm (DDASA) by exploiting a sigmoid function to adapt the sampling time dynamically. Moreover, this scheme can be adjusted to the variability of the ambient energy, requires less computational capacity compared to complex AI or Markov chains schemes, and can be easily implemented on ultra-low power boards, such as battery-less LoRa nodes [256]. However, both these studies consider the use of batteries to supply wireless sensors, and in particular, the first one used a battery-powered sensor without EH capabilities [254], while the second one used a photovoltaic source and a battery as a storage system [255]. Conversely, this work investigates for the first time the possibility of using the revised sigmoid function algorithm to optimize the transmission parameters of a battery-less LoRa node powered by a solar panel.

In particular, we adopted DDASA to dynamically set the transmission time of the device and compute the optimal value with a low-complexity algorithm that can be implemented on low-cost hardware. Moreover, a sigmoid-based approach can also be applied to adjust the data redundancy according to the environmental conditions (e.g., state of charge or solar radiation). This consideration is also essential for the correct design of the nodes: indeed, from the literature above described, it is clear that an appropriate choice of electronic components, along with the correct optimization of transmission parameters (T, NR, packet size, etc.), is essential to achieve the right balance between energy and economic efficiency. For example, regarding the storage system capacity, very different values have been used in the state of the art, ranging from 10 mF [246] to 270 F [239]; for this reason, we believe it is necessary to design the system taking into consideration not only transmission parameters and energy efficiency but also the costs of the entire system. To achieve this goal, our work considers the loss of energy probability (LoEP). Lastly, to the best of our knowledge, this study represents the first characterization of battery-less LoRaWAN sensors, powered by solar energy, using adaptive algorithms to optimize transmission parameters and compute the LoEP to minimize the costs of the sensor node architecture.

4.4 System Architecture

The main components of the proposed adaptation system is presented in figure 4.1. The battery-less device acquires energy from a renewable energy source (e.g., solar energy), stored in a supercapacitor. Physical components, such as the solar panel and supercapacitor, must be designed to provide enough energy in normal operating conditions. It represents a significant challenge because the device design is critical to optimize the amount of energy harvested from renewable sources to perform sensing and transmission operations. Regarding data processing, when the device has acquired new data from the sensors and has enough energy, it transmits a new packet according to two parameters: i) the transmission interval, i.e., the time T between two consecutive packets, ii) the packet data size, with redundancy parameter NR.

In our implementation, these parameters are either fixed or tuned dynamically according to the algorithm described in subsection 4.4.2, which takes into account the energy stored in the device.

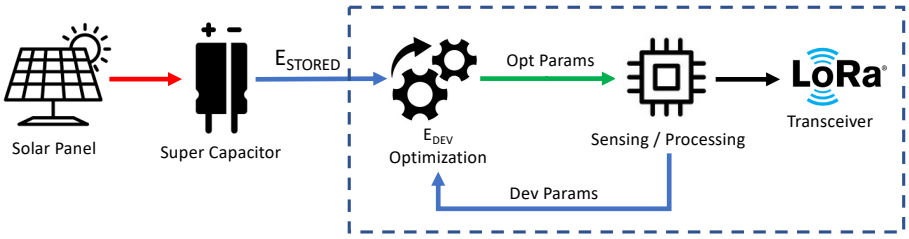


Figure 4.1: System Architecture and logical components.

As a renewable energy source, we assumed a photovoltaic (PV) panel as an energy source. In particular, we considered both a theoretical model and a real radiation data set as input, analyzing the impact of the EH source on the device and the effect of possible energy failures. Regarding the theoretical model, we used Duffie's radiation model [257], which is detailed in subsection 4.4.4. We employed the open data provided in [258] and [259] for the real radiation dataset.

4.4.1 Device energy model

The Energy Harvesting-based device has been designed taking into account following parameters:

- Device consumption profile (E_{DEV});
- Solar Panel Size (S_{PV});

- Energy Storage Capacity (C);
- Transmission interval (T);
- Data redundancy (NR).

For simplicity, the proposed model considers discrete time samples $n = k \cdot T_{PV}, k \in \mathbb{N}$, where T_{PV} is the energy source measurement sampling time, e.g. 1 hour or 15 min. Without loss of generality, in our scenario we take into account a PV panel as energy source, which can be modeled as:

$$E_{PV}(n) = \eta \cdot S_{PV} \cdot G_{solar}(n) \tag{4.1}$$

where η represents the PV efficiency, S_{PV} is the panel size and $G_{solar}(n)$ is the solar radiation in the time interval between $[n-1, n]$, i.e. $[k-1, k] \cdot T_{PV}$. The efficiency η allows taking into account different materials for the PV cell (mono, polycristalline or others).

The energy stored in the system is modeled using the following equation:

$$E_{ST}(n) = \min[E_{PV}(n) + E_{ST}(n-1) - E_{DEV} \cdot f_{TX}(n), E_{max}] \tag{4.2}$$

where $E_{ST}(n-1)$ is the energy stored by the system at time $n-1$, $E_{PV}(n)$ is the energy provided by the source, E_{DEV} is the energy consumed by the device and $f_{TX}(n)$ is a function which is equal to 1 if there is a transmission in the last interval (zero otherwise), as formalized in equation 4.4. The system energy storage is mainly composed of a supercapacitor which allows high-speed charge accumulation, up to the E_{max} limit (function of C). In our analysis, we consider that the charging time of the capacitor is very short compared to T_{PV}, so we can neglect the capacitor transient times. Finally, the energy consumed by the device E_{DEV} is computed as the sum of three main components as:

$$E_{DEV} = E_{SENS} + E_{TX}(NR) + E_{RX} \tag{4.3}$$

where E_{SENS} is the energy required to acquire and process data from the different sensors, $E_{TX}(NR)$ is the energy spent by the device to transmit a packet over the air and E_{RX} is the energy spent to perform the receiving operation.

Note that E_{TX} depends not only on the parameters of the LoRa transmission protocol but also on the payload size. Such payload can change dynamically to increase/decrease the number of sensor data sent in a single frame or to implement an

additional redundancy mechanism to improve the reliability of LoRaWAN transmissions (tuned by the overall NR parameter). A simple way to implement this solution consists in introducing temporal redundancy, i.e., to retransmit the same measurement data into multiple transmissions, thus increasing the probability that at least one of the packets is received. This approach is already exploited in commercial applications such as the Sensing Labs platform [260].

Thus, we implemented a sliding window mechanism that transmits the last NR measurements, with higher NR values yielding greater success probability. Figure 4.2 illustrates the shift-memory structure of NR, which updates the values whenever a new sensor measurement is performed.

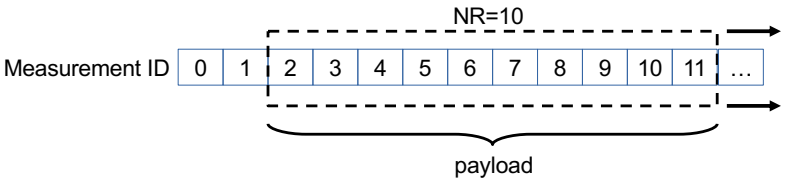


Figure 4.2: Sliding window scheme for data redundancy NR.

Finally, the device can transmit only if E_{ST} is high enough to perform a complete packet transmission. We model this functionality using the following activation function:

$$f_{TX}(n,T) = \begin{cases} 1 & E_{PV}(n) + E_{ST}(n-1) \geq E_{DEV} \quad \wedge \quad n = k \cdot T \\ 0 & otherwise \end{cases} \tag{4.4}$$

where the time interval T is the interval between two consecutive transmissions (for simplicity, we assume n to be a multiple of T). Clearly, the transmission delay decreases reducing T, in contrast, the amount of energy needed for the transmissions increases (and vice-versa). Next, we discuss how to tune T and NR parameters based on the energy received from the renewable source.

4.4.2 DDASA-based Transmission Algorithm

We defined a DDASA-based algorithm to adapt the transmission period of the sensor data dynamically. Indeed, as discussed in sec. 4.3, the DDASA algorithm [255] can be employed to adapt sampling and data transmission to optimize the resource utilization of the device. In particular, we considered this algorithm to increase or reduce the transmission interval and the data redundancy according to the amount of

Algorithm 1 DDASA-based Transmission Algorithm

$T = T_{\min}$ ▷ Initialize variables
$NR = NR_{\max}$
$m = 0$
$n = 0$
while true **do**
  if $(m \cdot T_{PV} \geq T)\&\&(E_{ST}(n) \geq E_{DEV})$ **then**
    $run(TxLoRaFrame)$ ▷ Transmit data
    $b_1 = (E_{ST}(n) - E_{ST}(n-m))/mean(E_{ST})$ ▷ Energy variation
    $m = 0$ ▷ reset time counter
    $b_2 = (E_{ST}(n) - E_{DEV})/E_{\max}$ ▷ energy gap
    $v_1 = \frac{2}{1+e^{-b_1}}$ ▷ Sigmoid function 1
    $v_2 = \frac{2}{1+e^{-b_2}}$ ▷ Sigmoid function 2
    $T = \min(T_{\max}, \max(T_{\min}, \frac{T}{v_1 \cdot v_2}))$ ▷ Update T
    $NR = \min(NR_{\max}, \max(NR_{\min}, NR \cdot v_1 \cdot v_2))$ ▷ Update NR
  $m++$
  $n++$

available energy to optimize energy consumption and minimize transmission failures. The adaptation is represented by a revised sigmoid function, which is expressed by $y(x) = \frac{2}{1+e^{-x}}$, where x in our case is the difference in stored energy computed between two consecutive transmissions.

4.4.3 Device Consumption Measurements

In our analysis, we conducted experiments to evaluate the current consumption in four different operating modes (sensing, transmitting, receiving, and sleeping). As shown in figure 4.3, the experimental setup was composed by a PV panel, a protection circuit, a supercapacitor and the LoRa device, with transmission settings summarized in table 4.1. In particular, we evaluated the power consumption of the LoRa FiPy and TTGO nodes, which include the ESP32 SoC and the SX1276 LoRa transceiver. We used a Tektronix MSO 2024B oscilloscope with TCP0020 current probe to measure the PV Panel Voltage, source voltage (from the supercapacitor) and the current absorbed by the device. For example, the power consumption of the TTGO device is depicted in figure 4.4 during different node activities. Specifically, the figure shows the device consumption during sensing, the transmission of a LoRa packet (LoRa TX), and the subsequent receive windows (LoRa RX1 and RX2) when the device listens for responses, acknowledgments, or downlink messages from the network server.

The average power consumption of the two considered devices in all the four mentioned states is outlined in table 4.2. In particular, the table shows that in the three active states (transmit, receive, sensing), the power consumption of the TTGO device is lower than the FiPy device. In contrast, the TTGO device initially had higher power consumption in the sleep state (about 10 mW), which was reduced at 0,15 mW by making the hardware changes suggested in [261]. Based on both devices' energy consumption characterization outcomes, it was decided to proceed with the analysis focusing solely on the TTGO device, as it shows lower energy consumption in all the measured conditions.

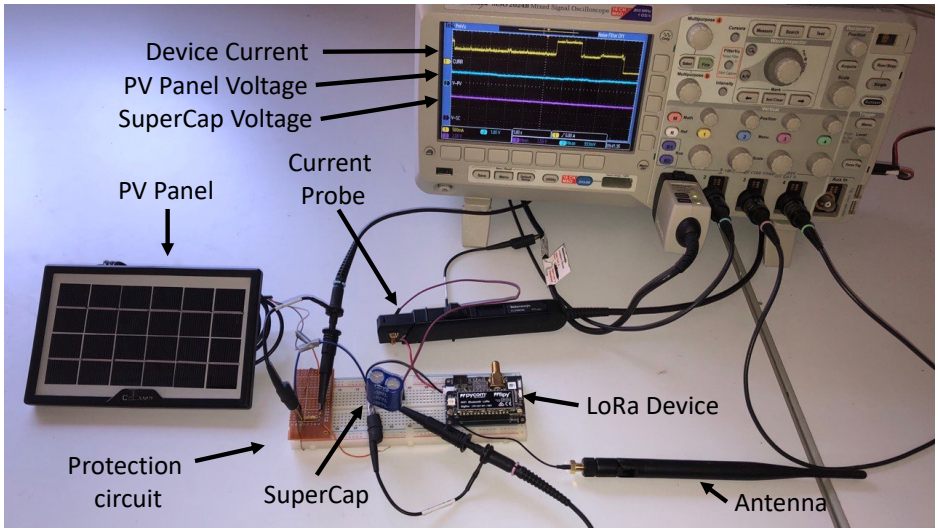


Figure 4.3: Experimental testbed setup used to measure the power consumption of the devices.

Table 4.1: Default values used for energy consumption measurements.

Setting	Value
Supply Voltage	3.7 V
Frequency	channel hopping in 868 MHz band
Spreading Factor	12
Forward Error Correction	4/5
CRC	Enabled
Payload length	2 bytes
Preamble Length	8 symbols

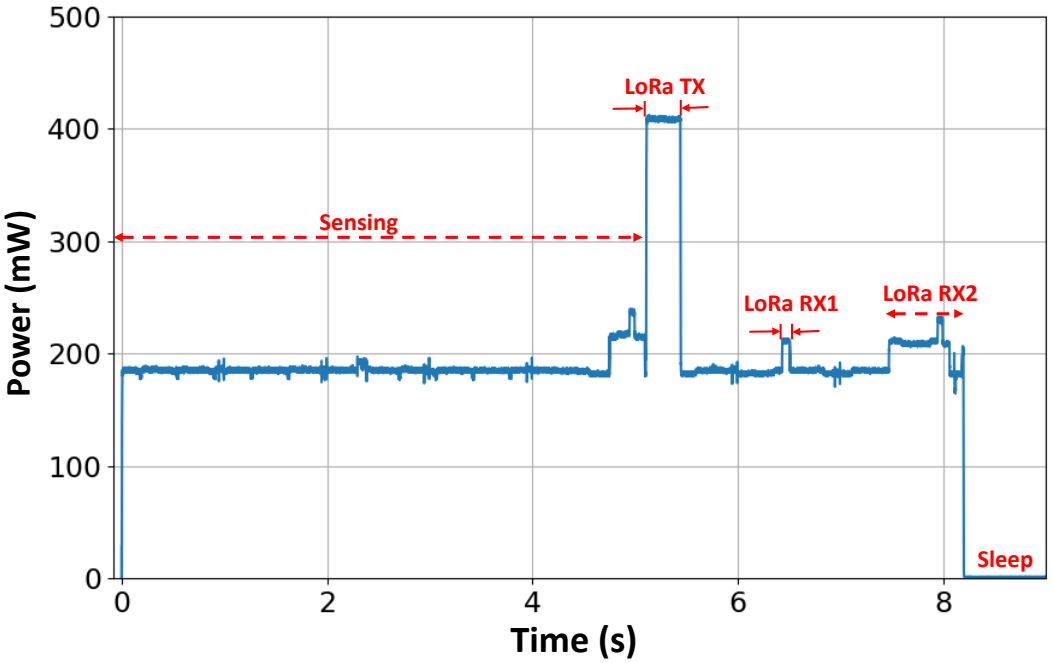


Figure 4.4: TTGO node power consumption in different working states.

Table 4.2: Power consumption of the two devices in the different operating modes.

Device	Transmit	Receive	Sensing	Sleep
FiPy	820 mW	600 mW	560 mW	0,225 mW
TTGO	420 mW	228 mW	195 mW	0,15 mW

4.4.4 Solar energy production model

In this subsection, we model the solar energy production based on the radiation theoretical model explained in chapter 1 of [262]. In particular, solar radiation can be predicted based on the day of the year, latitude, and azimuth. Energy reaches the Earth, gets partially scattered by the atmosphere, and can be converted by photovoltaic devices into electrical energy. Generally, the solar radiation overall captured by a solar panel comprises three components: direct, diffuse, and reflected. In our analysis, we selected the correlations by Erbs et al. [257]: based on measurements taken at various locations in the United States, an experimental data regression model is proposed for the luminosity index to estimate the diffused solar radiation.

The next step involves estimating the power generated by a solar panel to predict hourly production and determine the energy flows exchanged with the LoRa device

and supercapacitor. Precisely, the power delivered from photovoltaic panel was calculated considering the photoelectric conversion efficiency, panel area, solar radiation incident on the panel, and operating temperature, using the empirical relationship expressed in [263]. We then compared the average monthly power of the theoretical model with the real dataset provided in [259, 258], at two reference locations: the first, at latitude 38.132° in the region of Sicily, characterized by citrus, vineyard, and olive cultivation, and the second, at latitude 45.45° in Lombardy, primarily known for cultivating cereals (such as barley, rice, and corn), vegetables, forage, and potatoes. These datasets are provided by the National Solar Radiation Database (NSRDB) of the USA and SIAS (Servizio Informativo Agrometeorologico Siciliano), Italy, respectively. The first dataset is a comprehensive and publicly available source of solar radiation and meteorological data offered by the National Renewable Energy Laboratory (NREL) in the United States. The NSRDB provides high-resolution solar irradiance data and covers an extensive period, from 1991 to the present, with regular updates. The NSRDB offers high temporal (hourly or sub-hourly) and spatial (4 km x 4 km) resolution data, making it suitable for detailed analyses and accurate energy production simulations. The dataset includes various solar radiation parameters such as global horizontal irradiance (GHI), direct normal irradiance (DNI), and diffuse horizontal irradiance (DHI). Additionally, it provides essential meteorological parameters like temperature, humidity, wind speed, and precipitation. The NSRDB combines measurements from ground-based stations, satellite-derived data, and advanced atmospheric models to ensure accuracy and reliability. The NSRDB database is freely accessible through NREL's online platform or APIs, allowing users to download data for specific locations, time periods, and parameters.

Regarding the second dataset provided by SIAS (an agrometeorological information service in Sicily, Italy), it is based on a real-time weather monitoring system employing a network of meteorological stations distributed across the island of Sicily. The stations have wind speed sensors at 2 meters and 10 meters height, utilizing Robinson cups and optoelectronic transducer technology. The wind direction sensors at the same heights employ vane and optoelectronic transducer technology. Other sensors include a global radiation sensor (measuring cumulative solar radiation), air temperature sensor, relative humidity sensor, precipitation sensor (tipping bucket rain gauge), leaf wetness sensor, atmospheric pressure sensor, and snow depth sensor. The stations are synchronized in time to align with the forecasting models, and a weekly time check ensures station accuracy. The datalogger used is the MTX WST1800 [264]

model, featuring a single-board CMOS microprocessor with 128 KB of RAM and 64 KB of EPROM memory.

The SIAS dataset provides data acquired since 2003, at an hourly time resolution.

Comparing the theoretical model with the real datasets, figure 4.5 reveals that for both the rural locations considered, the theoretical model closely follows the actual solar radiation in the autumn and spring seasons; conversely, it tends to underestimation in the summer months and to overestimation during winters. The theoretical model performance is summarized in table 4.3, in terms of Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Bias Error (MBE). In particular, the analysis of MBE is used to estimate the average bias of the model. The obtained results show low bias. According to RMSE, the model has a significant impact on outliers, and this is reasonable because the model does not account for unpredictable radiation fluctuations due to atmospheric events. Considering a maximum normal surface radiance of around 1000 W/m^2 at sea level on a clear day, we can assume a normalized RMSE of approximately 10-13%. Finally, the R^2 values obtained are greater than 70%, which is an acceptable model fit.

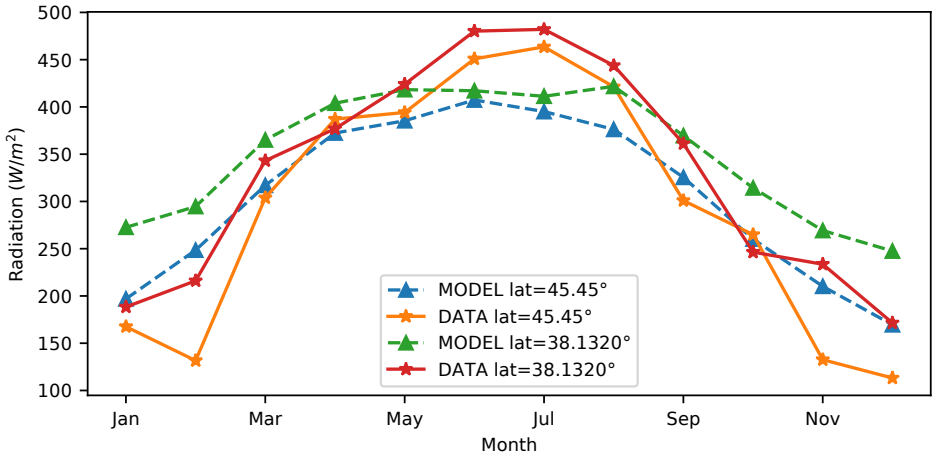


Figure 4.5: Monthly average solar radiation of the theoretical model and real dataset

Table 4.3: Difference between the theoretical model and real dataset at two latitudes.

Latitude	R^2	$RMSE\ (W/m^2)$	$MBE\ (W/m^2)$
45.45°	0.71	135.57	-1.63
38.132°	0.83	109.03	1.01

4.4.5 Solar panel parameters

Maximizing the efficiency of converted solar energy is strictly tied to solar panel selection. In particular, the power conversion efficiency (PCE) and maximum power point (MPP), as well as the size, are essential features for the design of the panel [187]. In our experiments, we employed a commercial silicon solar panel (sized 7.5 x 14 cm) . Therefore, a mathematical model, given in [265], was used to characterize the PV panel and identify the electrical parameters. In Figure 4.6, the PV panel's I–V and P–V characteristics are reported, assuming 820 W/m^2 of solar radiation. In the figure, I_{sc}, V_{oc}, I_{mp}, and V_{mp} denote short-circuit current, open circuit voltage, the maximum power point current, and the maximum power point voltage, respectively. By extracting these four parameters, we can identify the MPP operating condition in which the power transferred from the source to the load is maximized. Thanks to this characterization of the panel, the proposed dynamic algorithm can be tuned to work on an operating point closest to the V_{mp} voltage, reducing the system's sensitivity to current at the MPP point. From the figure, it is clear that the operating voltage value (V_{ope}) must fall within the supply voltage range of [3.3 - 5.2] V.

4.5 Evaluation and results

In order to evaluate the proposed system, we exploited the consumption model obtained from the device characterization, discussed in subsection 4.4.1, to simulate the behavior of the system by changing the design parameters such as capacity and solar panel size. We also employed the adaptation algorithm, based on energy-aware DDASA, to dynamically adapt the transmission period T and payload size as a function of NR, as described in algorithm 1 of the subsection 4.4.2. Finally, we conducted simulated experiments for one year using both the energy radiation model and real-world radiation datasets, described in subsection 4.4.4.

4.5.1 Performance of the adaptation algorithm

As a first experiment, we validated the adaptation algorithm capabilities to optimize the transmission parameters (i.e., transmission interval and NR). In particular, we measured the number of correct transmissions (#TX_DONE), the number of transmissions that failed due to insufficient energy (#TX_FAILED), and the number of bytes transmitted per packet. A transmission is considered failed when the stored energy is not sufficient to perform the overall procedure of sensing, processing, and

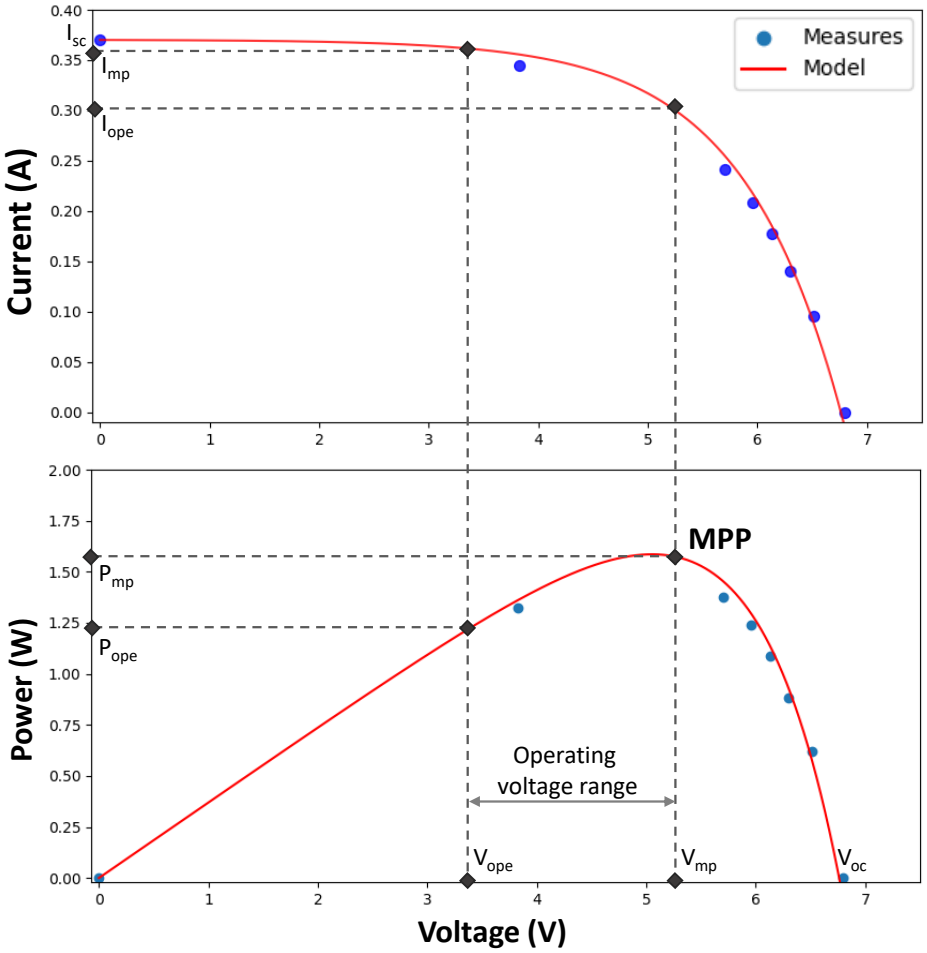


Figure 4.6: Solar panel characterization under 820 W/m^2 of solar exposure.

Table 4.4: Results obtained per season using the theoretical model with C=4.5 F and PV=0.01 m^2.

SEASON	Average T (min)	Average NR	#TX_DONE	#TX_FAIL	LoEP (%)	LoIP (%)	Av. pkt size (B)
WINTER	170	8.31	691	104	13.08	4.57E-06	16.62
SPRING	110	9.20	1141	72	5.94	5.18E-10	18.40
SUMMER	111	9.19	1164	78	6.28	8.96E-10	18.38
AUTUMN	170	8.27	676	132	16.34	3.12E-05	16.54
YEAR	140	8.74	3672	386	9.51	1.17E-07	17.48

transmission. Moreover, we compute the Loss of Energy Probability (LoEP), which directly translates to a packet loss, as $\frac{\#TX_{failed}}{\#TX_{failed}+\#TX_{done}}$. We analyze the impact of data redundancy defining a Loss of Information Probability (LoIP), i.e., the probability of losing NR consecutive packets, which can be computed as $LoIP = LoEP^{NR}$.

Tables 4.4 and 4.5 summarize the results obtained using the theoretical model or the real dataset, respectively. For the experiments, we consider $C = 4.5$ F and

Table 4.5: Results obtained per season using the real dataset, with C=4.5 F and PV=0.01 m^2.

SEASON	Average T (min)	Average NR	#TX_DONE	#TX_FAIL	LoEP (%)	LoIP (%)	Av. pkt size (B)
WINTER	213	7.64	540	127	19.04	3.15E-04	15.27
SPRING	105	9.28	1193	81	6.36	7.83E-10	18.56
SUMMER	99	9.27	1291	85	6.18	6.17E-10	18.54
AUTUMN	194	7.94	581	138	19.19	2.02E-04	15.89
YEAR	153	8.53	3605	431	10.68	5.13E-07	17.07

$PV = 0.01\,\mathrm{m}^2$, and we assume an EH efficiency of 15 $\mathrm{mW/cm}^2$ [230] and a latitude of 38.132°. The tables report the average TX interval T and NR parameters computed by the adaptive algorithm, the LoEP and LoIP probabilities, and the average packet size, which is directly influenced by NR and the number of sensors on the device (we assume a total sensor data of 2 Bytes, i.e., $2 \cdot NR$ bytes). Values show that during higher solar radiation periods (mainly summer and spring), the adaptation algorithm lowers T and increases NR, obtaining a good number of successful transmissions and resulting in a low LoEP compared to other seasons with lower radiation intensity (autumn and winter). For example, 1167 successful transmissions were obtained during the summer season using the theoretical radiation model, while in the winter, only 691 frames were successfully transmitted. Similar numbers were obtained using the real data. Note that with the specific values selected for C and PV size, the LoEP for the theoretical model is generally lower than the real dataset. This is due to abrupt weather events (e.g., cloud obfuscation), which are not included in the theoretical model, but are recorded in the real dataset. In any case, the tables show that LoIP is extremely low, even with LoEP as high as 20% (winter and autumn seasons in table 4.5). This demonstrates the importance of the data redundancy NR, which helps recover lost packets by retransmitting data multiple times.

Table 4.6: Comparison of packet sizes in LoRa technology for agricultural applications.

Ref	Location	Crop	Main application	Total pkt size (B)
[79]	Bucharest, Romania	Walnut	Temperature, humidity and pressure monitoring	16
[62]	-	-	Drip irrigation system	9
[55]	Hsin-Chu county, Taiwan	Vegetable	Irrigation and greenhouse monitoring	40
[76]	Depok, Indonesia	Starfruit	pH and soil moisture monitoring	5
[100]	Indonesia	-	Livestock monitoring system	20
[85]	Romney, USA	Tree farm	Humidity, temperature and sunligth monitoring	9
[266]	Salcedo, Ecuador	Potatoes, corn, onions	Humidity, temperature and soil moisture monitoring	16
[3]	Tianjin, China	Various crops	Greenhouse monitoring	9
[267]	Edirne, Turkey	-	Sensor and actuator controller	5
[14]	Osnabruck, Germany	-	Conductivity and soil temperature monitoring	2

The table 4.6 presents a comparison of packet sizes (pkt size) utilized in LoRa technology across various agricultural applications. Each row delineates a distinct study or project, detailing the location, specific crop or livestock monitored, and the

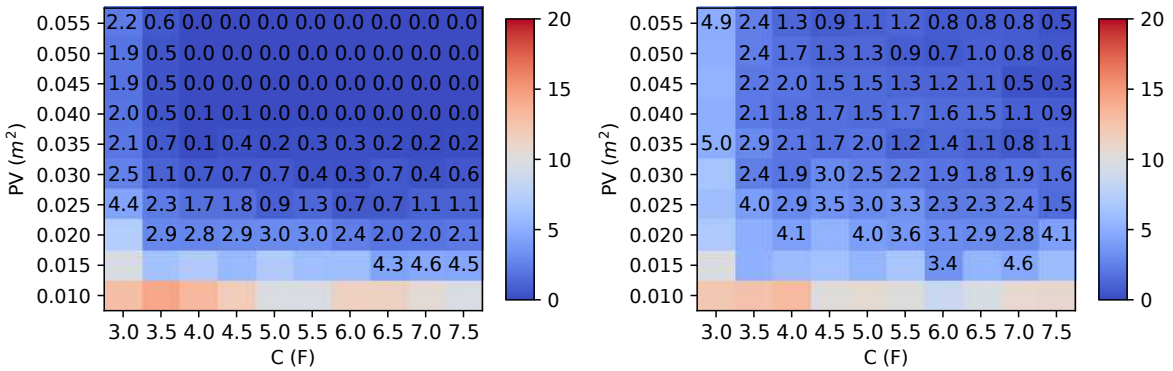
primary application of LoRa technology. From Bucharest, Romania, where walnut crops undergo temperature, humidity, and pressure monitoring, to Salcedo, Ecuador, where potatoes, corn, and onions are subjected to humidity, temperature, and soil moisture monitoring, the table encapsulates a diverse array of agricultural contexts. The packet sizes range from as low as 2 bytes to as high as 40 bytes, showcasing the variability in data transmission requirements across different agricultural setups. The average pkt size in the table 4.6 is 13 bytes. This comparative analysis provides valuable insights into the data transmission sizes in agriculture. It's noteworthy that the average total packet size is comparable to the average packet sizes generated by the DDASA algorithm for LoRa battery-free sensors, as shown in table 4.5. This suggests the feasibility of employing the DDASA algorithm in smart agriculture scenarios. However, it's crucial to carefully evaluate packet size in relation to redundancy and useful information transmission, adapting strategies on a case-by-case basis. Thanks to its flexibility, the DDASA algorithm could prove to be an effective solution for optimizing packet sizes in agricultural monitoring systems.

4.5.2 Impact of supercapacitor and PV panel size

A second set of experiments was conducted by varying the values of C and PV and analyzing the overall LoEP obtained using the proposed adaptive algorithm. The results are summarized in Figures 4.7 and 4.8, considering a one-year observation period and for two different latitudes. The heatmap shown in the figures depicts the average LoEP as a function of the capacity and solar panel size, respectively, on the x- and y-axes. Moreover, we highlight values less than or equal to 5% of LoEP by explicitly indicating the obtained value in the plot. These figures are helpful for establishing some component constraints for an optimal design of the device, as discussed in the following section.

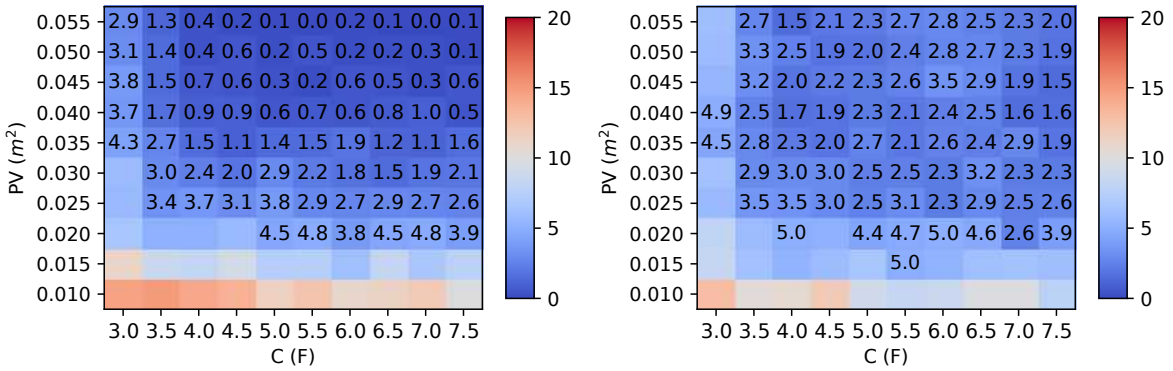
4.6 Design parameters and LoEP analysis

Fixing a LoEP (or LoIP) threshold involves the definition of acceptable loss probability and, in the design of a battery-less device, can help identify appropriate design parameters (such as capacity and solar panel size) that achieve this threshold value. Together with these parameters, the design optimization should also consider other factors, such as cost. Concerning Figures 4.7 and 4.8, values of C and PV can be identified so as to satisfy the constraints on the LoEP (e.g., less than or equal to 5%);



(a) Percentage of LoEP for Duffie's Model (b) Percentage of LoEP for real dataset

Figure 4.7: Loss of Energy Probability (LoEP) results for agricultural scenarios at latitude = 38.132°.



(a) Percentage of LoEP for Duffie's Model (b) Percentage of LoEP for real dataset

Figure 4.8: Loss of Energy Probability (LoEP) results for agricultural scenarios at latitude = 45.45°.

in this way, the choice of components for given LoEP can be made by optimizing the cost. However, the optimal choice of source and storage size must take into account the availability on the market. The required values can be obtained directly by a single unit or by composing multiple smaller components until the desired value is reached. In particular, concerning the size of the PV source, the choice can start from a single cell, but also array can be considered. Similarly, the capacitance of the storage system can be composed by parallel connection of two or more capacitors, if necessary. It can be noted that the total cost usually does not rise linearly; indeed, the market often proposes cheaper components with higher performances of the corresponding elementary units to be connected.

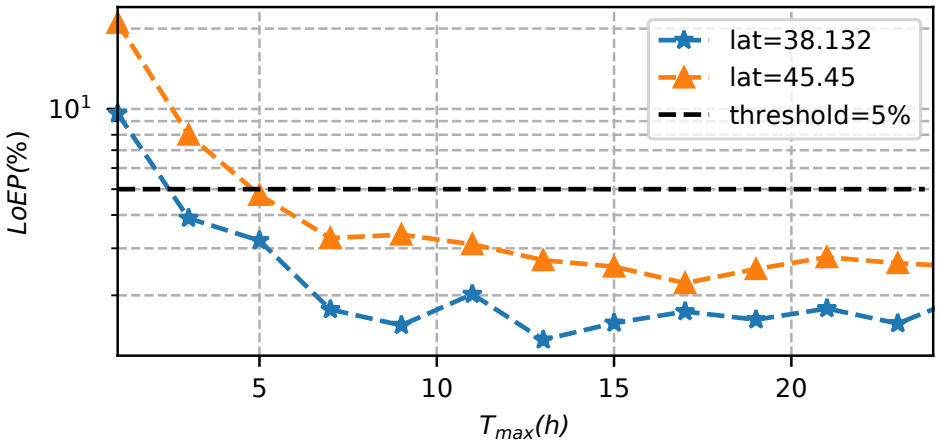


Figure 4.9: LoEP as a function of T_{max}, with $C = 5\,\mathrm{F}$, $S_{PV} = 0.035\,\mathrm{m}^2$, $NR_{max} = 10$.

Different potential commercial solutions involving supercapacitors and solar panels have been examined to elucidate this phenomenon. The paper [268] provides detailed explanations of their characteristics and costs. Subsequently, all conceivable combinations of supercapacitors and photovoltaic cells were assessed to determine the most cost-effective solution while maintaining a Loss of Exceedance Probability (LoEP) of 5 percent or less. The evaluation focused on a Duffie radiation model tailored for rural scenarios at a latitude of 38.132°.

Once the optimal system design has been obtained, the LoEP and LoIP can be re-analyzed with these new values. In particular, in this last experiment, we measure the loss probabilities when varying T and NR (in particular, by tuning T_{max} and NR_{max}), with a fixed $C = 5$ F and $S_{PV} = 0.035$ m^2. Figure 4.9 shows the performance of the adaptation algorithm in terms of LoEP, as a function of T_{max}. The figure clearly shows that the LoEP limit of 5% is satisfied when T_{max} is greater than 5 hours. Instead, with a T_{max} fixed to 5 hours, figure 4.10 shows the LoIP when varying NR_{max} between 1 and 10. In this case, we can highlight an exponential reduction of information loss probability, which can assure data transmission even with small values of NR.

4.7 Summary

In this chapter, LoRa-based battery-less devices were analyzed, with optimization of the transmission parameters achieved through an energy-aware adaptation algorithm. By characterizing and modeling the energy consumption of two battery-less sensors and employing theoretical energy models alongside real radiation traces obtained from

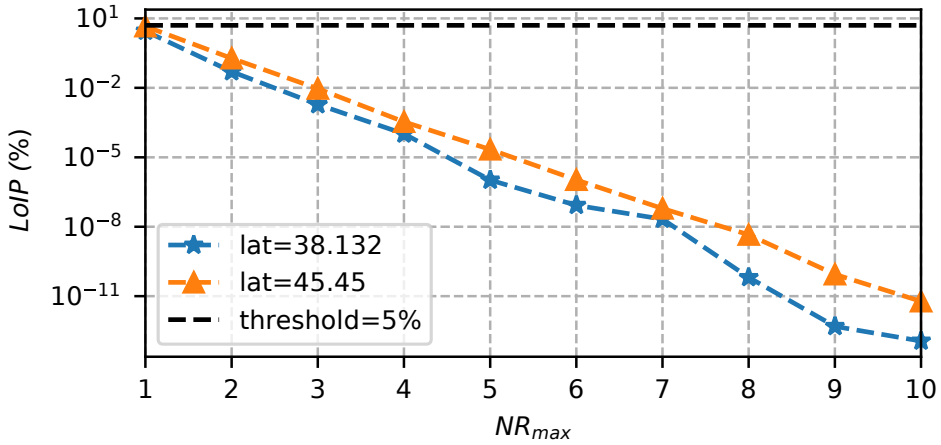


Figure 4.10: LoIP as a function of NR_{max}, with $C = 5$ F, $S_{PV} = 0.035\text{m}^2$, $T_{max} = 5$ h.

two distinct agricultural scenarios, the study explores the dynamic adjustment of transmission intervals and payload sizes to accommodate the available energy. Finally,a generalized approach for designing capacity-based storage and PV panel size is presented, yielding a cost-effective methodology for devising battery-less solutions tailored to ambient energy-powered LoRa sensors. The results showed that, for a given threshold probability of power failure, the proposed approach can successfully optimize the device's energy consumption by automatically setting the relevant transmission parameters. Moreover, information loss can be dramatically reduced simply by repeating the data transmission in multiple packets.

From the presented results, a system composed of a 5 F supercapacitor and a PV panel of $0.035\,m^2$ is capable of transmitting data packets every 5 hours, with a redundancy of 3 and a LoIP of 10^{-2}. Finally, the LoEP results obtained with the proposed algorithm, combined with an in-depth cost analysis, has allowed selecting the most economical solution for the dimensioning of the PV panel and supercapacitor, balancing implementation costs and energy failure probability.

The proposed approach is applicable to different IoT applications that require autonomous energy systems, such as smart agriculture. In future research activities, we will test the proposed algorithm on a large-scale deployment and study the impact of networking and modulation parameters, such as resource allocation of different Spreading Factors and Adaptive Data Rate algorithms.

Part 2: Specific Application scenarios for Sustainable Agriculture

Chapter 5

Optimizing Irrigation Networks Through Sustainable and Accurate IoT Flow Monitoring

5.1 Overview

Water scarcity is nowadays a critical global concern and an efficient management of water resources is paramount. This chapter presents an original approach for Smart Water Grids through Internet of Things (IoT) that involves the integration of multiple sensors placed across the irrigation consortium networks to accurately measure water flow. In order to enhance energy efficiency for the monitoring and communication processes, with an emphasis on sustainability, the power of graph theory and graph signal processing has been harnessed. This approach enables the representation of water flow in a highly adaptable and accurate manner, while concurrently minimizing the requirement for a large number of IoT sensors to transmit measurement data. A graph-based model is introduced, where the water flow is depicted as a signal on the graph. Additionally, an algorithm, named SmartGraph, is proposed to reconstruct the graph signal even in cases where certain measurements are unavailable. This framework is applied on a synthetic realistic environment within the context of LoRaWAN (Long Range Wide Area Network), an infrastructure and protocol designed for ultra-low-power IoT devices. The findings indicate that the implementation of SmartGraph yields a notable reduction in energy consumption while concurrently ensuring precise flow estimation. This research underscores the considerable potential for energy-efficient and accurate monitoring of water flow, thereby facilitating advancements in the management of wa-

ter grids. Moreover, it enables water operators to effectively address challenges related to water scarcity.

5.2 Introduction

Water scarcity is an ever-growing concern that affects communities and ecosystems worldwide [269, 270]. In fact, 70% of the world's freshwater withdrawals are now used for agriculture, and this percentage is expected to increase in order to meet the population's growing demand for food and energy [271]. According to government reports and available data [272] on the wasted water, the average result provided by the European Commission for the Europe region is approximately 26%, with a spike of 45% in some specific areas [273].

In recent decades, among the various practices implemented by irrigation managers to mitigate the effects of water scarcity, traditional irrigation systems have been increasingly replaced by pressurized water networks connected to Water Supply Systems [274]. These networks operate either on demand or on a rotational delivery schedule [271]. Therefore, methodologies are needed to design or restore these irrigation networks effectively. The aim is to achieve highly effective networks capable of meeting the desired water demands and service pressure with limited installation costs [271].

The efficient and effective management of Water Supply Systems (WSSs) has become essential to ensure a sustainable and resilient water supply. A WSS comprises water sources, water treatments, pumping stations, a water distribution network (WDN), and finally, the end users. A WDN encompasses junctions, tanks and reservoirs pipes, pumps, and valves. The junctions can serve as water supply or demand points to and from the network, while the tank(s) or reservoir(s) inject water into the system and represent finite or virtually infinite water sources. Nowadays there is an increasing interest in Smart Water Grids (SWGs) that are advanced water supply systems that incorporate digital technology, sensors, and data analytics to improve the efficiency, reliability, and sustainability of water supply and distribution. Estimating the precise volume and flow of running water and the localization of a possible leakage, is challenging due to variations in data collection methods, reporting practices, and the complexity of WSS across different countries. Traditional approaches have often relied on manual inspections or simplistic models, which can be time-consuming, labour-intensive, and prone to errors [275].

To this aim, water sensors can be distributed along the WDN for leak detection and flow monitoring and they can be abilitated to communicate this information to

a central system or device. However, WSS operators often operate within budget constraints, limiting the number of sensors that can be deployed to reduce the energy impact of the monitoring system; finding the right balance between cost-effectiveness and comprehensive coverage of the WDN is essential. Finally, the quality and resolution of the data sensors collected depend on their placement. Placing sensors at key nodes and junctions with high flow rates or historical leak occurrences improves the chances of timely detection. However, ensuring sufficient coverage across the entire network is crucial.

Concerning the sensor technology, there is no universal technology for WSS monitoring. IoT chips with low power consumption and long-distance wireless communication capability are ideal for these purposes, like WiFi-based, cellular networks or LPWAN networks. LPWAN devices are expected to dominate the field [276], with different infrastructure requirements: i) cellular infrastructure-dependent, such as NB-IoT; ii) third-party infrastructure reliant, like SigFox; and iii) Autonomous LPWANs, such as LoRaWAN [277]. Cellular technology-based LPWANs offer wide coverage, capacity, battery life, quality of service, and security but are not cost-effective due to subscription fees and dependence on commercial networks. SigFox [138], a patented network, spurred rapid innovation by increasing competition among LPWAN technologies. LoRaWAN offers numerous benefits, including low power consumption, extensive coverage, simplicity, and easy management due to its characteristics. However, it faces potential scalability issues in large-scale scenarios. This study targets the LoRaWAN technology, enabling real-time analysis and processing of the vast amounts of data generated by connected devices within the network edge. As shown in Fig. 5.1, the LoRaWAN architecture relies on a star-of-stars topology, where end devices or sensors connect with one or more Gateways (GWs) that forward packets to the Network Server (NS) and to the corresponding IoT applications server for what concern to the application data. According to the considered scenario, LoRaWAN sensors are strategically placed within the SWG elements to gather physical data information of the hydraulic resources.

In this context, energy efficiency represents a relevant requirement in water monitoring using IoT devices. To this goal, a key contribution to the improvement of SWGs management is to develop accurate water flow reconstruction methods that consider the energy efficiency of IoT measurement scenarios. In addition, precise flow control within WDNs is crucial for optimizing irrigation efficiency, preventing water wastage, and ensuring crops receive the necessary amount of water tailored to their specific requirements [271].

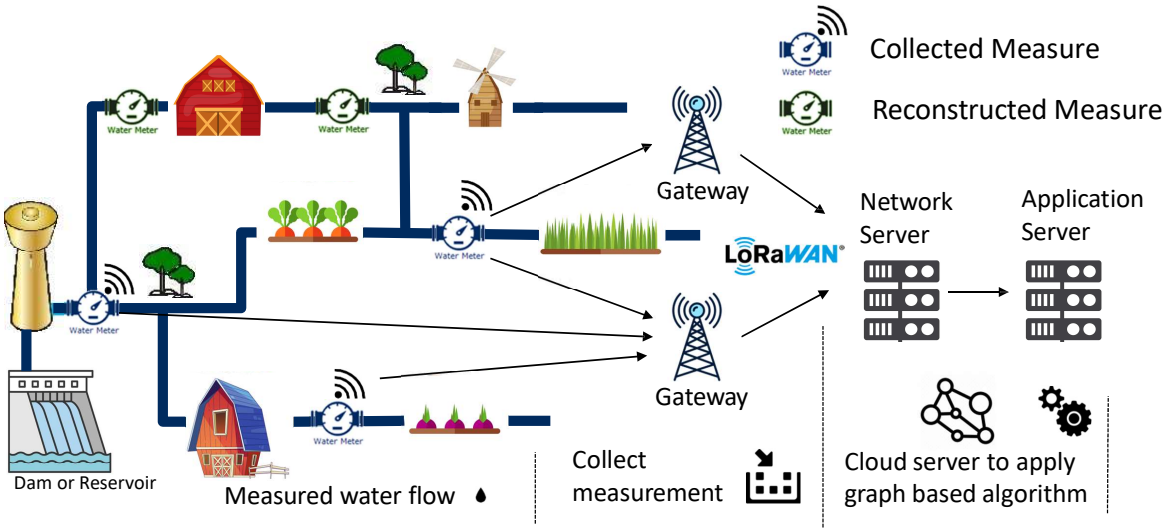


Figure 5.1: A LoRaWAN architecture, based on a star-of-stars topology. LoRaWAN sensors are positioned within the WSS to collect data from meter nodes.

This chapter focuses on the critical issue of energy efficiency within SWGs and investigates the potential of finding the optimal sensors' placement for water flow reconstruction. The proposed methodology, named GraphSmart, introduces a graph-based approach aimed at enhancing the precision of water flow reconstruction and improving energy efficiency within the SWG. Leveraging principles from graph theory, the topology of a distribution network is depicted and analyzed, followed by a graph transformation designed to extract the minimal set of nodes necessary for accurate network flow reconstruction. Additionally, an extended model is proposed to incorporate the potential for nodes to request water from the irrigation network, denoted as demand values.

Initially, the methodology is applied to a specific use case to thoroughly comprehend the functionality of the GraphSmart approach across various scenarios, including those with and without demand values. Then, the proposed approach is applied to LoRaWAN technology, identified as the most suitable IoT solution for the specific scenario. We developed a model that incorporates considerations for sensor energy consumption due to their communication ability, acknowledging the non-uniform energy usage of individual nodes due to variations in wireless coverage and radio parameters. This model accounts for the peculiarity of single-node energy consumption within the context of the wireless environment. In order to perform the simulation of GraphSmart in a realistic setting, we use the EPANET (US Environmental Protection Agency water NETwork) software and the NS-3 simulator. These tools are widely recognized

in the literature as the most suitable for evaluating WDN and LoRaWAN scenarios, respectively.

An irrigation consortium network was analyzed, and its corresponding LoRaWAN integration was presented, with considerations given to optimal GWs placement and radio coverage parameters. Additionally, an in-depth analysis of the energy consumption profile of LoRaWAN devices was provided, which served to calibrate the simulation parameters.

In this scenario, additional analyses were conducted to merge considerations regarding the accuracy in reconstructing the flow from a subset of measurements and the energy consumption of the communication network. The findings demonstrate the potential and innovation of GraphSmart, in achieving green and accurate IoT water monitoring.

Taken together, the key contributions of this chapter are:

- A graph model tailored for Water Supply Systems (WSSs) is introduced, drawing upon Graph Signal Processing (GSP) theory specifically crafted for water flow monitoring. This model facilitates the interpretation of water flow as a signal on a graph and is extended to accommodate demand values, allowing nodes to request water from the network.

- A method, denoted as *GraphSmart*, is presented to establish a sensor ranking that incorporates node centrality measures. Additionally, a green flow reconstruction algorithm is devised with the aim of reconstructing water flow using the minimum available measurements. The ranking procedures also suggest optimal sensor placement locations.

- The proposed method is applied to a realistic Water Distribution Network (WDN), demonstrating its efficacy in ensuring both accurate flow monitoring and reduced IoT energy consumption. This dual benefit translates to cost savings and enhanced environmental sustainability.

The rest of the chapter is organized as follows. Section 5.3 discusses related work on graph-based methods and LoRaWAN applications for SWGs. In section 5.4 we introduce the GraphSmart method, while in section 5.5 we apply it on several synthetic networks emulating different scenarios. In section 5.6 we introduce a study of the LoRaWAN network, including radio parameters and an energy model. In section 5.7 we report results of the proposed method on a realistic WDN. Section 5.8 concludes the chapter.

5.3 Related Work

In this section, an overview of the current state of research on graph-based approaches tailored specifically for WSS is presented, along with recent developments in communication technologies relevant to this particular application.

Graph Signal Processing (GSP) is an emerging research domain recently proposed to analyze data defined on non-Euclidean domains, such as graphs. According to the GSP framework, the structure of a generic graph, that is constituted by nodes and edges, is integrated by signal, that results associated to each graph node. With this position, the typical operations defined for signals have been formulated on the graph, such as filtering, compressing, frequency transformation [278].

There is a growing interest in exploring graphs and this is achieved by different perspectives, such as network theory [279] and signal processing [280]. Researchers have recently developed methods to capture essential graph characteristics, such as dynamic correlations [280] or non-linearities in the interacting system, that can be described as the succession of states possibly modeled as a Markov model [281] or to multi-kernel learning approaches [282]. The majority of the proposed approaches are data-driven [283], which makes those methods particularly able to fit the complexity of real data, but they have the intuitive limitation of less generalizability. Taking into consideration the time variability of the majority of physical phenomena modeled as graphs, a fundamental topic is modelling dynamic networks [284]. The study of time-varying graph requires sophisticated mathematical tools and often strong assumptions are imposed, such as stationarity [285].

Graph signal processing theory has been employed in real-world scenarios, where graphs naturally depicts the structure of the data, such as in biological [286] and sensor networks [287]. WSSs are complex infrastructures composed of interconnected pipelines facilitating the flow of water, and nodal junctions, which may be outfitted with advanced sensor technology for monitoring and control purposes. This configuration can be intuitively mapped into graphs, which are well known mathematical structures made by links and nodes. Thus, WSSs can be modelled as graphs, wherein the basic configuration graph links correspond to physical connections of the WSSs, while nodes correspond to the junctions [288, 289].

Recent research has applied graph-based methods to address the problem of leakage detection. Graph-based methods have been applied in the context of water distribution networks in sectorization projects [290], with particular attention to assess their resilience [291]. A large part of the research activity concerning graph theory has

mainly focused on the detection of water leakages using modelling tools [292, 293] and graph neural networks.

In the context of agricultural irrigation systems, graph signal processing has recently proved its capabilities in graph learning [294], in clustering to monitor the flow effectively [295] and in optimal deployment of sensors [296, 297].

Although graph-based approaches have been explored in the context of water supply systems, many unsolved questions persist. Indeed, accurate and stable methods for WSS monitoring are still missing and other problems related to the IoT application, such as the number of sensors and their location needed to reconstruct the water flow in the network. In this chapter, we deal with the problem of network flow reconstruction in an optimal way.

Authors in [276] show a comparative analysis of the available IoT technologies in the WSS field. After a presented analysis of vast literature, Wi-Fi was chosen main of the time, followed by LoRaWAN and cellular IoT. Wi-Fi is the most popular choice for IoT applications, but only because the considered literature includes old research scenarios. Indeed, three factors will make the LoRaWAN technology dominant in the future WSS scenario. The first is the operating frequency, indeed the higher RF frequencies are blocked by walls, trees, and other obstructions, but lower RF frequencies are less susceptible to this kind of issue. That's why cellular and LoRaWAN have lower operating frequencies with respect to Wi-Fi. Moreover, if the device is only a sensor node and only requires periodic transmission, then LoRaWAN is a good choice. Another two crucial factors are power consumption and the range of the device. LoRaWAN and cellular have a similar range of 10 km, but the cellular technology has a higher data rate.

In addition, energy efficiency is not typically a major concern for measurements taken at WSS tanks or pumping stations, as these usually receive energy from a power grid. However, the situation changes significantly when considering smart meters and flow monitoring sensors within the WSS, as these are usually battery-powered [298]. In these scenarios, sensors with minimal power consumption need to be paired with high-energy-efficient communication technologies to enable the sensors to have energy autonomy for several years. By combining long-range wireless communication with low power consumption, LoRaWAN technology allows for extended battery life [299]. This is particularly crucial for flow monitoring, where sensor locations are often at the periphery of the WDS without the possibility of connecting their power supply to a power grid.

5.4 The GraphSmart method for water monitoring

In this section, an initial overview of the proposed approach is provided, followed by a detailed examination of its specific implementation steps. Fig.5.2 shows the workflow

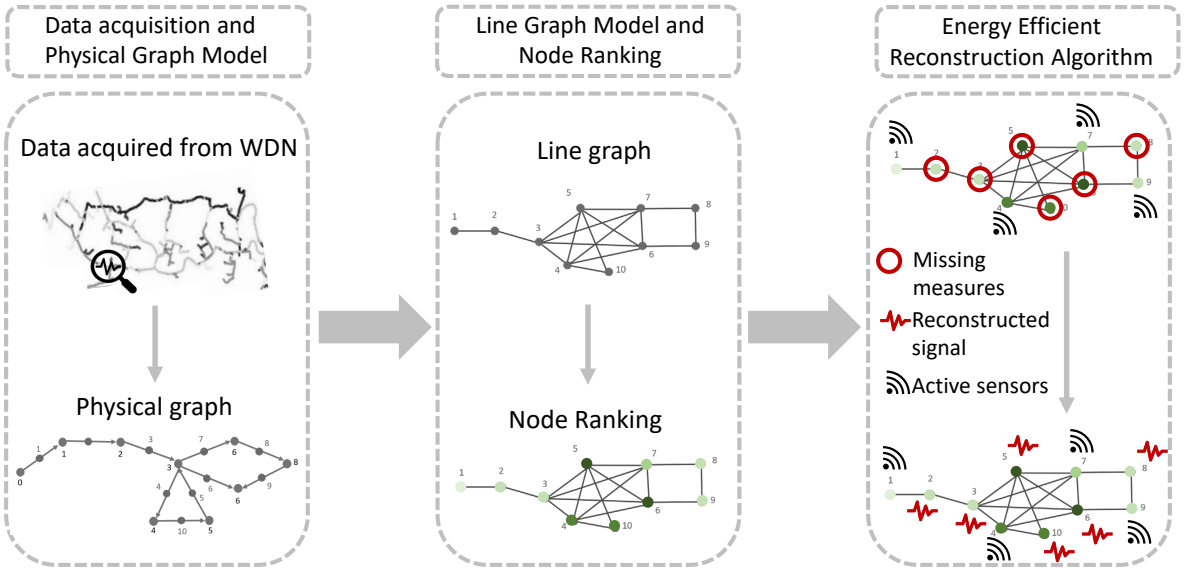


Figure 5.2: The figure outlines the proposed method: the first step consists in the generation of a graph related to the physical structure of the network. The second step is the development of an alternative representation through the line graph where flow information is associated to nodes. The we propose a node ranking by their reconstruction relevance. Finally, we have a reconstruction algorithm that leverages graph topology and selects a minimum set of flow measurements for reconstruction. Specifically, according to the ranking found in the second phase, the measurements of nodes highlighted by red circles (nodes 2,3,5,6,8,10)are not collected, but reconstructed from other nodes (nodes 1,4,7,9) equipped with active sensors.

of the GraphSmart method, where: (i) the flow data acquisition is modeled by a graph representation $\mathcal{G}$ of the physical network; (ii) an alternative graph representation $\mathcal{G}_L$ is introduced, associating the flow information with the graph nodes, and nodes are ranked based on their relevance for flow reconstruction; (iii) the reconstruction algorithm is provided based on the knowledge of the graph topology and a reduced number of flow measurements. The three steps depicted in the figure 5.2 are elaborated upon in the subsequent subsections.

5.4.1 Data Acquisition and Physical Graph model

The initial step of the algorithm represented in Fig. 5.2 is the mapping of a WDN into a graph to capture the complex structure and dynamics of those systems. According to this representation, the network is translated into a graph $\mathcal{G}=(\mathcal{V}, \mathcal{E})$, that is a

mathematical model represented through its N nodes (or vertices) and N_e links (or edges). The adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ element a_{ij} represents the weight of the link connecting the nodes i and j. In the context of water distribution networks, the standard way to build the graph consists in associating links to pipes and nodes to points where the pipes cross and that can be usually equipped by sensors [300]. This graph carries information about the physical layout and functionality of the WDN. The graph nodes correspond to essential components of the distribution system. The edges model the physical connections, typically pipes, allowing the water flow between different points in the network. The graph reflects the topological intricacies of the water distribution network, providing a mathematical representation that aids in understanding the network's architecture. Advanced techniques, such as Signal on Graph (SoG), can be employed within this framework to learn complex representations of the WDN. SoG have the capability to capture spatial and temporal dependencies, offering a more sophisticated understanding of the network's behaviour and enhancing the accuracy of predictive models.

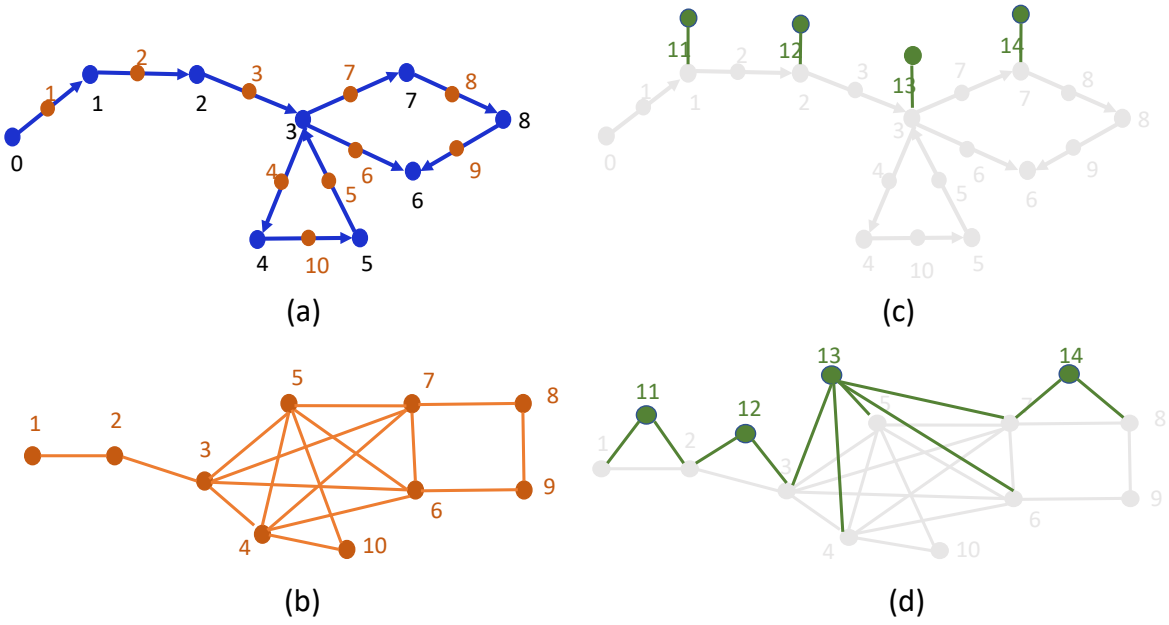


Figure 5.3: Example of a line graph representation. In panel (a) we have the original directed graph, with $N = 9$ nodes and $N_e = 10$ links. The orange numbers associated to each link enumerates graph edges and they visually constitute the nodes for the line graph. In panel (b) we have the corresponding line graph made by N_e nodes. Each node correspond to a link of the original graph. In panel (c) we have the extension to the demand values. Indeed, some graph nodes of the original graph require water from the network. It is modeled by an outgoing virtual flow. The virtual flows in original graph generates virtual nodes in the line graph, that are represented in panel (c)

In Fig. 5.3 (a) we have an example of graph with $N = 9$ nodes and $N_e = 10$ edges (both represented in blue) that will be used in the following subsection to present the approach.

5.4.2 Line Graph model transformation and Node Ranking

In this subsection, we will present the following steps of the GraphSmart method presented in (Fig.5.2), that involves a graph transformation and the identification of the node ranking.

With the aim to identify an optimal edge monitoring strategy, we work on an alternative domain where the flow values are associated to a node rather than a link, and the theoretically grounded reconstruction techniques are available [301]. Specifically, we resort to the line graph $\mathcal{G}_L$=($\mathcal{V}_L$, $\mathcal{E}_L$) associated to the original network graph $\mathcal{G}$. Each node $v \in \mathcal{V}_L$ of the line graph $\mathcal{G}_L$ is associated to one and only one link $e \in \mathcal{E}$ in the original graph $\mathcal{G}$; besides, two nodes $v_1, v_2 \in \mathcal{V}_L$ are connected if and only if the corresponding links $e_1, e_2 \in \mathcal{E}$ of the original graph originate from a common vertex $v \in \mathcal{V}$ [302–304].

We provide an example of deriving a line graph in Fig. 5.3(b). Here, the line graph consists of N_e nodes corresponding to the original graph shown in panel (a). It is visually evident that each link in panel (a), represented by orange points, corresponds to a node (orange node) in panel (b).

After associating the original link and its flow value with a node in the line graph, we can apply GSP to the line graph and reconstruct the flows from a limited set of measurements. Indeed, in the original graph representation of a WDN, the information of interest corresponds to the hydraulic flow associated with each link (or pipe). However, to apply SoGs method at the node level, a line graph transformation of this original graph representation is needed. Specifically, this redefines the graph such that the nodes, instead of the links, become the primary carriers of the hydraulic flow information. This adjustment allows for the application of the SoG method directly on the nodes of the graph.

The adjacency matrix of the line graph $\mathbf{A}_L \in \mathbb{R}^{N_e \times N_e}$ depends on the adjacency matrix of the original graph $\mathbf{A}$:

$$\mathbf{A}_L = \mathbf{B}\mathbf{B}^T - 2\mathbf{I} \tag{5.1}$$

where $\mathbf{B} \in \mathbb{R}^{N \times N_e}$ is the incidence matrix of the original graph, whose generic binary element b_{ij} indicates if the node i is implied in the $j-th$ link.

We propose a node ranking strategy on the line graph, which involves sorting the links (i.e., junctions) in the original graph. Specifically, we rank the nodes of the line graph $\mathcal{G}_L$ based on their predictability from neighboring nodes. This ranking enables us to determine which flow values can be readily predicted and which ones require measurement for accurate flow reconstruction. The rationale behind this ranking is that central nodes can be predicted from their neighbors more effectively than less interconnected ones.

Therefore, when only a reduced number of measurements can be collected, these should correspond to the least interconnected nodes of the line graph $\mathcal{G}_L$. Remarkably, these nodes in $\mathcal{G}_L$ correspond to less interconnected links of the original graph $\mathcal{G}$, such as links across different node clusters. This is coherent with recent research findings [305], indicating that such links significantly impact the network's algebraic properties. In line with this approach, we focus on node centrality [306] to capture its interconnections within the network.

The centrality of the nodes $\mathcal{V}_L$ within the line graph $\mathcal{G}_L$ is evaluated using various centrality estimators [306], including the "betweenness" $c_n^{(b)}$, "closeness" $c_n^{(c)}$, and "page-rank" $c_n^{(p)}$ centrality metrics.

The *"betweenness"* centrality metric $c_n^{(b)}$ of the n-th node $v_n \in \mathcal{V}_L$ depends on how many times this latter is included in the shortest paths $\mathcal{P}_{ij} = \{v_i, \cdots, v_j\}$, $i, j = 0, \cdots N-1$ between each and every graph node pair $(v_i, v_j) \in \mathcal{V}_L$, and it is computed as follows:

$$c_n^{(b)} = \sum_{\forall (v_i, v_j) \in \mathcal{V}_L} \frac{|\{\mathcal{P}_{ij} \; s.t. \; v_n \in \mathcal{P}_{ij}\}|}{|\mathcal{P}_{ij}|}$$

where $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$.

The *"closeness"* centrality metric $c_n^{(c)}$ of the n-th node $v_n \in \mathcal{V}_L$ depends on its distance from each and every other graph node $v_j \in \mathcal{V}_L$, $j \neq i$, and it is computed as follows:

$$c_n^{(c)} = \frac{1}{\sum\limits_{\forall (v_j) \in \mathcal{V}_L, j \neq i} d_{ij}}$$

where d_{ij} is the distance between nodes i and j.

The *"page-rank"* metric $c_n^{(p)}$ of the n-th node $v_n \in \mathcal{V}_L$ depends on how many times this latter is included in random walks. This algorithm determines the centrality based on how much each node appears within the shortest paths of the network. [1] $\mathcal{W}_{ij} = \{v_i, \cdots, v_j\}$, $i, j = 0, \cdots N-1$ between each and every graph node pair (i, j),

[1] The random walks are generated by selecting the successor of a node with probability 0.85 out of its neighbours and with probability 0.15 among all the other remaining network nodes.

$i, j = 0, \cdots N-1$

$$c_n^{(p)} = \sum_{\forall (v_i, v_j) \in \mathcal{V}_L} \frac{|\{\mathcal{W}_{ij} \; s.t. \; v_n \in \mathcal{W}_{ij}\}|}{|\mathcal{W}_{ij}|}$$

5.4.3 Green Reconstruction algorithm

In this subsection, the proposed reconstruction algorithm is introduced, which is graphically represented as the final step of Fig. 5.2. According to the ranking proposed in the previous subsection phase, the measures at the most central nodes (highlighted by red circles) are not collected, but reconstructed from adjacent, relevant nodes. This is illustrated in the right panel of Fig.5.2, where active sensors correspond to nodes 1,4,7, 9 , while missing measures, that are associated to missing or sleeping sensors are 2,3,5,6,8,10.

It means that we discard measurements at the most central nodes in $\mathcal{G}_L$, while keeping those at the least central ones, which are less predictable from their neighborhood. The centrality based ranking will identify the nodes that can be discarded without affecting the flow reconstruction accuracy.

To develop the reconstruction algorithm, we rely on the well-established WSS model commonly employed in the literature, such as the Hardy-Cross method [307]. Within GraphSmart model, we presume that water flow conforms to a distribution pattern where each junction adheres to the principle of continuity. According to the continuity equation, the sum of flow rates in pipes converging at a node, along with any external flows, must equate to zero. This requirement for continuity extends to every junction, wherein the algebraic sum of flow rates in the connecting pipes, alongside any demand flows, must also balance to zero. Thus, there exists a condition on water flow ensuring that the sum of inflows equals the outflows for each node in the graph, under the assumption of no leakage:

$$\mathbf{A_{FLOW}} \cdot \underline{1} = \underline{0} \tag{5.2}$$

From Eq. (5.1) it is possible to rewrite this condition in terms of incidence matrix $\mathbf{B_{FLOW}}$ computed from $\mathbf{A_{FLOW}}$:

$$\mathbf{B_{FLOW}} \cdot \underline{1} = \underline{0} \tag{5.3}$$

A different perspective can be obtained by considering the incidence matrix of the physical graph $\mathbf{B_{PHY}}$, i.e. when the adjacency matrix $\mathbf{A_{PHY}}$ is a binary symmetric

matrix that represents the presence of a pipe between two nodes. In this manner, the condition expressed in Eq.(5.3) can be formulated as follows:

$$\mathbf{B_{PHY}} \cdot \underline{f} = \underline{0} \tag{5.4}$$

We assume to know only N_{e^K} values of the SoG $\underline{f}$, while the remaining N_{e^U} are missing. For sake of clarity, we identify two components $\underline{f}^K$ and $\underline{f}^U$ that correspond respectively to the known and unknown values of the signal $\underline{f}$.

$$[\mathbf{B}^U_{PHY}|\mathbf{B}^K_{PHY}] \left[\frac{\underline{f}^U}{\underline{f}^K}\right] = \underline{0}$$

where $\mathbf{B}^U_{PHY}$ and $\mathbf{B}^K_{PHY}$ are the columns of the matrix $\mathbf{B}_{PHY}$ that are multiplied for $\underline{f}^U$ and $\underline{f}^U$ respectively. Let us remark that all the elements of $\mathbf{B}_{PHY}$ are known since they correspond to the geometry of the connections on the network. Following the development of the equations, it is possible to obtain a mathematical expression to compute $\underline{f}^U$, i.e. the missing values of the SoG representing the water flow in the network:

$$\underline{f}^U = \mathbf{B}^{U+}_{PHY} \cdot \mathbf{B}^K_{PHY} \cdot \underline{f}^K \tag{5.5}$$

where $\cdot^+$ is the pseudo-inverse of the original matrix.

Extention to the case of water demands. The original WSS network model encompasses infrastructure elements and consumer endpoints. The associated line graph allows us to effectively estimate the flow within the network using a reduced number of sensors. Let us now delve into the representation of consumer endpoints. Each consumer endpoint involves an incoming flow, i.e. the water supplied by the network, and an outgoing flow, i.e. the water demanded by the consumer, that we refer to as demand value. The incoming water flow is associated to the physical connection of the endpoint with the infrastructure, and hence to a network edge. The outgoing flow is not directly tied to a physical connection. This notwithstanding, we can represent it by introducing in $\mathcal{G}$ an additional edge that connects the endpoint to a virtual boundary node, which does not correspond to a physical point of the network. This extended graph $\mathcal{G}_{\restriction \S \sqcup}$ contains N_e connections related to the physical pipes and N_e associated to demand values With this extension, the adjacency matrix of the physical graph including demand values is $A_{ext} \in \mathbb{R}^{2N \times 2N}$:

$$A_{\text{ext}} = \left[\begin{array}{c|c} A & I \\ \hline 0 & 0 \end{array}\right] \tag{5.6}$$

where I is the identity matrix $I \in \mathbb{R}^{N\times N}$.

With this considerations, the adjacency matrix can be organized in blocks where we recognize the original adjacency matrix A, that models the physical structure of the network. Then we have another block $I \in \mathbb{R}^{N\times N}$ that reflects the possibility that each of the N nodes has a virtual outflow towards virtual boundary node. All the other elements of A_{ext} are zero since the virtual nodes cannot be connected in other ways.

With this position, the reconstruction algorithm applies regardless of whether there are outgoing flows at the consumer endpoints. Indeed, we seamlessly manage the water flows exiting the network at consumer endpoints, treating them akin to flows within the network itself. These outgoing flows may be either known (e.g., measured spilled water from consumer utility meters) or unknown (e.g., due to potential leakage within the WSS infrastructure), contingent upon the specific characteristics of the observed WSS network. We visually depict the presence and impact of the addition of virtual nodes in Fig.5.3(c). Here, the green links represent the flows at endpoints in panel (c), with the associated line graph shown in Fig.5.3(d), where the corresponding nodes are highlighted in green within the transformed line graph.

5.5 GraphSmart flow reconstruction accuracy

In this section, we focus on realistic WDN networks and we study the accuracy of the reconstruction, quantified by the Mean Square Error (MSE) between the ground truth $\underline{f}_j, j = 0, \cdots N-1$ and the reconstructed flow $\hat{\underline{f}}_j, j = 0, \cdots N-1$, computed as: $\sum_{j=0}^{N-1}(\hat{\underline{f}}_j - \underline{f}_j)^2$. The relationship between accuracy and the number of available sensors is established, demonstrating that the approach yields an energy-efficient monitoring system. The performance of the proposed method is presented in two distinct cases.

In order to better understand the key ideas and the algorithm steps, the first case shows results on an example network with a limited number of links, with and without demand values. Then, the second case presents the results of the proposed approach in a complex WDN scenario, extracted from a real infrastructure and obtained by the EPANET tool, which is a software able to generate WDN with real physical constraints associated with the water flow.

5.5.1 Results on small WDN

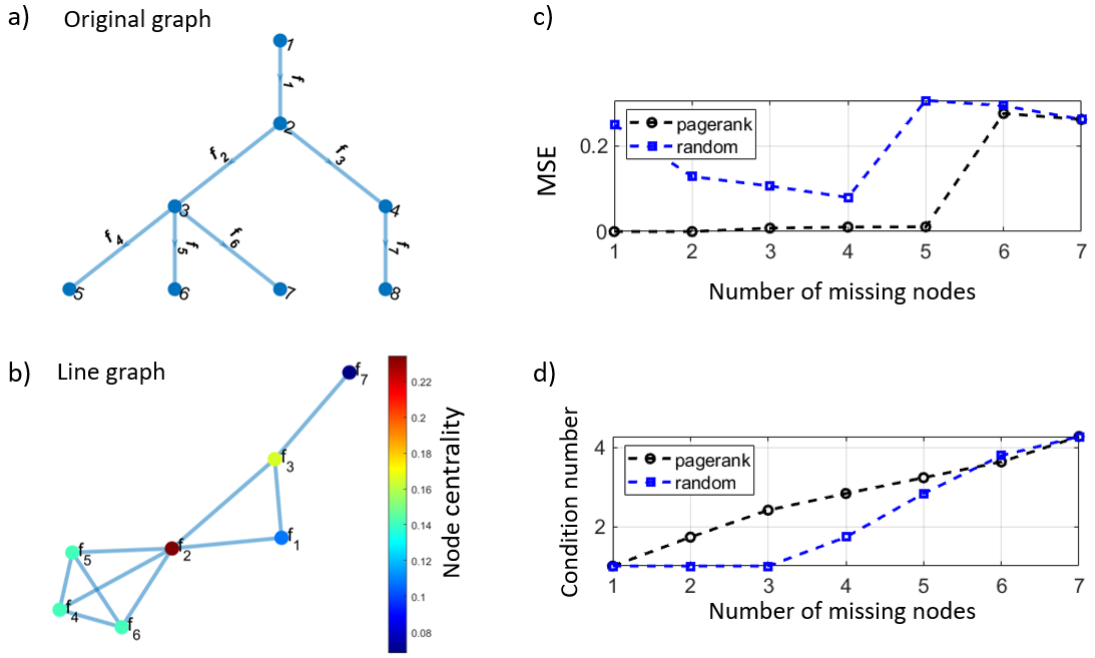


Figure 5.4: Example use case representation. In (a) we have the original graph represented by graph links. In (b) we have the corresponding line graph. The color scale associated with the nodes of the line graph corresponds to their centrality, as measured by the page-rank algorithm. In (c) we report the MSE between ground truth and reconstructed flow values of the line graph as a function of the number of removed sensors. In the black line, the sleeping nodes are ranked according to their descending centrality, while in the blue line they are randomly sorted. In (d) we have the condition number for inversion, which measures the quality of the inversion in Eq.(5.5).

In this subsection, results from a case study are presented with the aim of comprehending the functionality of the proposed approach. This analysis numerically illustrates the different stages of the proposed algorithm within a controlled scenario, as depicted in Fig. 5.4. Specifically, we consider the graph of a smal WDN with $N = 8$ nodes $1, 2..N$ and $N_e = 7$ edges $f_1, f_2, ..., f_{N_e}$ depicted in Fig. 5.4 (a). Then, we derive the associated line graph as in Eq.(5.1) that is made by N_e nodes, where each node corresponds to a link in the original graph, represented in Fig. 5.4 (b). The node centrality of the line graph is calculated using the page-rank algorithm as outlined in Sec. 5.4. The node centrality is represented by the color of the nodes in Fig. 5.4(b), with color scale levels indicated in the figure legend. As illustrated in the figure, the node with the highest centrality value is situated at the center of the graph ($N_e = 2$), which corresponds to node $N = 2$ in the original graph.

According to the proposed approach, this node is a strong candidate for removal as it represents a location where the flow can be more readily reconstructed. The key

hypothesis is that the first nodes, i.e. the most central nodes, are less important in the reconstruction algorithm.

In order to test this idea, we start to eliminate the values associated with the first ranked node and we apply the reconstruction algorithm. After the reconstruction algorithm, we compute the MSE between the reconstructed and ground truth node signal value. Let us remark that this operation corresponds in the original graph, to the reconstruction of the missing water flow value through the others.

These steps are replicated after eliminating other node values based on their centrality, and the results are depicted by the black line in Fig. 5.4(c). Results show that we can switch-off the 5 most central nodes and perfectly reconstruct their values, that means that only the two less central nodes, i.e. f_7 and f_1 (depicted in blu according to the color scale), are needed to completely recover the node signals.

To assess whether the ranking has an impact on the algorithm's ability to reconstruct the network, we replicate the same procedure but with a random ranking of the nodes (blue line in the figure). In this scenario, the performance deteriorates because there's a higher probability of removing nodes with lower centrality first. This indicates that the order of node removal, based on their centrality, plays a significant role in the successful reconstruction of the network flow.

Finally, in Fig. 5.4(d) we report the condition number for inversion C defined as $C(B^U_{PHY}) = ||B^U_{PHY}||||B^{-1U}_{PHY}||$. This metric is frequently used in literature to gauge the reliability of algorithms, particularly those involving matrix inversions [308]. A small condition number for inversion implies that the problem is well-conditioned, meaning the solution is stable and reliable. Conversely, a large condition number signifies an ill-conditioned problem. In such a case, minor changes in the input can lead to significant changes in the output, making the solution less reliable and more sensitive to input variations.

In this applications, a large C means that the quality of the inversion of the matrix B^U_{PHY} is poor. The analysis of the condition number shows that when all the node measurements, corresponding to the water flows, are available, C is low, meaning that the inversion of B^U_{PHY} is well performed. As we eliminate node values and reconstruct them through the others, the condition number increases and the problem becomes ill-conditioned making impossible the correct reconstruction of the node values.

To assess the case where physical demands are present in the network, we now perform the analyses in a modified version of the previous graph. In this modified version, additional edges are introduced to represent demand values, signifying the potential for nodes in the network to have additional outflows. This is represented in

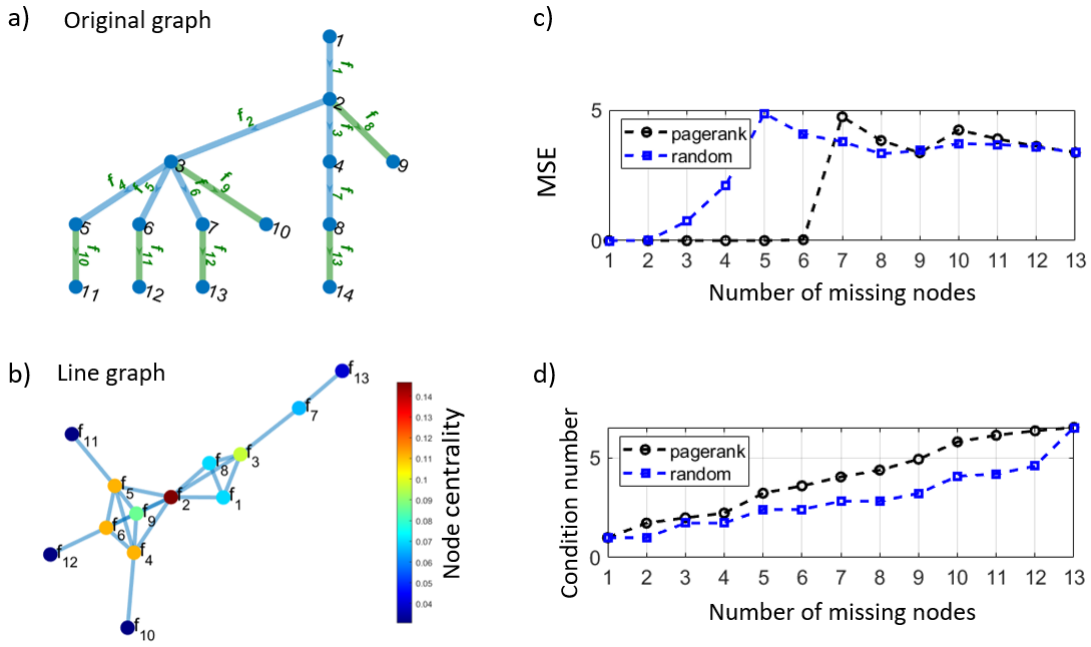


Figure 5.5: Example use case representation in the presence of demand values. In (a) we have the original graph represented by graph links coloured in blue, and demand values, represented by additional nodes and links coloured in green. In (b) we have the associated line graph. The color associated to the nodes of the line graph corresponds to their centrality, as measured by the page-rank algorithm. In (c) we report the MSE between ground truth and reconstructed node values of the line graph as function of the number of sleeping nodes. In the black line, the sleeping nodes are ranked according to their descending centrality, while in the blue line, they are randomly sorted. In (d) we have the condition number for inversion, which measures the quality of the inversion in Eq.(5.5).

Fig.5.5(a). According to the proposed model, anticipated in Sec. 5.4, we represent the water demand as an additional outgoing flow from each node of the network that demands water to the network. The virtual edges representing demand values are in green (f_8, f_9, f_{10}, f_{11}, f_{12}, f_{13}) in Fig.5.5(a) and they connect physical nodes,i.e. 1, 2, 3, 4, 5, 6, 7, 8, with virtual additional nodes, i.e. 9, 10, 11, 12, 13, 14.

From the original graph with those extensions, we derive the line graph in Fig. 5.5(b) according to Eq.(5.1) and we compute the node centrality with page-rank algorithm. The node centrality is represented as the color of the nodes. Following the same approach explained for the previous analysis, we derive the MSE between the reconstructed and the ground truth node values as a function of the number of sleeping nodes, reported in Fig. 5.5(c).

The results indicate that when the nodes of the line graph are appropriately ranked, the reconstruction error remains low (i.e., almost zero) for a larger number of sleeping nodes compared to the absence of ranking. Specifically, we can eliminate 6 out of 13 node values and accurately recover the signal. It's worth noting that the introduction

of demand values has resulted in a decrease in the percentage of eliminable node values. Indeed in the absence of demand values, we can switch-off a percentage of 71% of total nodes while in the presence of demand values, this percentage falls to 46%. This point is in line with the model since we do not change the topology of the network but we just have additional virtual nodes that relate to the original nodes of the graph and, consequently, the number of sleeping nodes that we can allow for a specific topology does not significantly change.

5.5.2 Results on irrigation consortium network

A predefined network model, extracted from a real infrastructure accessible in the Open Water Analytics community public repository [309], was utilized in irrigation consortium networks to evaluate GraphSmart. The hydraulic data desired for the study were generated using EPANET and the Water Network Tool for Resilience (WNTR) [310]. WNTR is a Python package designed to simulate real-time WDN. WNTR interfaces EPANET, which is an open source software to model hydraulic and quality dynamics of WDN or irrigation networks [311, 312]. EPANET takes into account the topological structure of the pipeline system along with a set of initial conditions (e.g. pipe diameter) and rules of how the system is operated so that it can compute flows and pressures throughout the network for a specific period of time. The simulator can add leaks to the network using a leak model [313].

The tools analyze the geometric structure of a pipeline system, taking into account various initial conditions such as pipe roughness and diameter, as well as operational rules. By doing so, it is able to calculate flows, pressures, and water quality parameters (such as disinfection concentrations and water age) throughout the network for a specific time period. By using these tools we generate a suitable dataset to apply the proposed method. To facilitate data analysis, we have chosen to export physical WDN values evolving during the time into two "Comma-Separated Values" (CSV) files, the first one which is related to the WDN junctions (nodes) and includes the physical objects that constitute the distribution system as well as its running parameters and the second one which is related to the WDN pipes (links). Table 5.1 and Table 5.2 contain the comprehensive set of features and fields associated with the links and nodes, respectively, as reflected in the provided dataset, that is publicly available on our GitHub page [314].

Element	Description	Unit
hour	A timestamp representing the time-interval we are currently watching in the simulation	H:M:S
linkID	Unique ID of a link inside the network	Not applicable
start_node	The source node of the link	Not applicable
end_node	The target node of the link	Not applicable
flowrate	The flow rate of the water inside the pipe at the current timestamp	GPM: gal/min
velocity	The velocity of the water inside the pipe at the current timestamp	ft/s

Table 5.1: The table showcases the attributes of the pipes. The flow direction within these pipes is determined by the difference in hydraulic head (the internal energy per weight of water or pump effect), with water flowing from the end with higher hydraulic head to the end with lower head.

Element	Description	Unit
hour	A timestamp representing the time-interval we are currently watching in the simulation	H:M:S
nodeID	Unique ID of a node inside the network	Not applicable
demand	Rate of water withdrawal from the network. A negative value is used to indicate an external source of flow into the junction	GPM: gal/min
head	Hydraulic head (i.e., elevation + pressure head) of water in the node of the WDN	ft
pressure	Measured pressure in the node of the WDN	psi
x_pos,y_pos	Coordinates of the node	meters
node_type	A string which tells the type of the node (i.e., "Junction", "Reservoir", "Tank")	Not applicable

Table 5.2: The table showcases the attributes of the nodes. Each node can be configured with a specific base demand pattern which represents the water request of the user during the whole simulation changing at a step size of an hour

Specifically, attention was directed towards the irrigation consortium network illustrated in Fig. 5.6. The network can serve an area of approximately 3200 hectares and comprises a total of $n = 83$ nodes and $m = 1$ reservoir. The area served by the consortium is divided into 83 sectors, each corresponding to one of the nodes within the network. Each node can be customized with a specific model of *base demand*, rep-

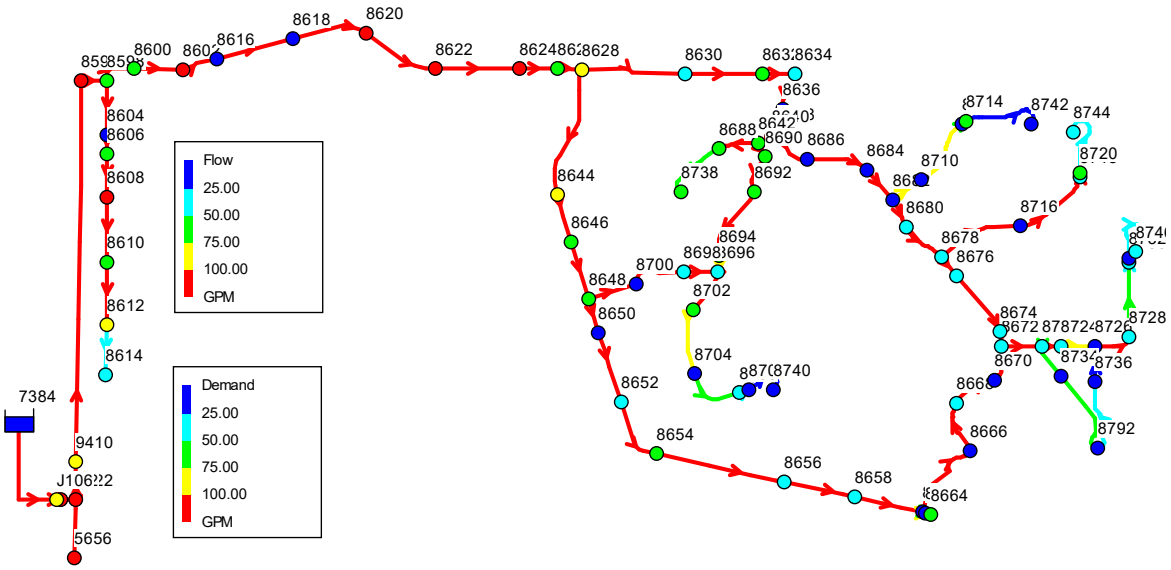


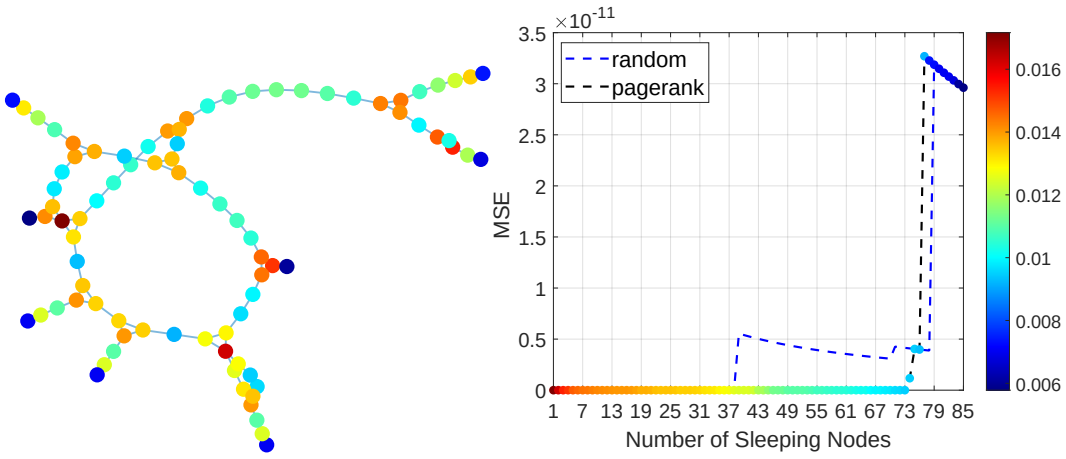
Figure 5.6: Selected irrigation consortium network with 83 junctions and 1 reservoir. Junctions demands and pipes flows measurements are represented in the figure by color depicted in the two legends scale.

resenting the water requirements of the irrigation sector throughout the simulation. The demand pattern changes at an hourly interval and is evenly distributed within a given range. In this study, two scenarios were examined. The first scenario involved the exclusion of demand configuration for the nodes, enabling water flow throughout the network while preventing any water spillage. The second scenario entailed configuring the nodes to have a demand evenly distributed within a specified range.

The figure employs two color scales to represent the demand values of the nodes (indicated by the color of the circle markers) and the flow within the pipes (indicated by the color of the connection segments between nodes). Both were measured in GPM: gal/min. The reservoir, identified by the label "7384," is positioned on the left side of the network, with the predominant flow direction being from left to right. As we move away from the reservoir, both the total demand value and the pressure at the junctions decrease due to the demand from the junctions on the left side. The links correspond to pipes and they are represented with an arrow in order to show the direction of the water flow.

A simulation spanning a one-month running experiment was conducted. To ensure the reproducibility of our experiments, the dataset for this experiment can be found in the publicly available repository provided in [314].

The proposed method is now applied to this irrigation consortium network. We firstly derive the binary adjacency matrix whose generic element $\mathbf{A}_{ij}$ is equal to 1 if



(a) (b)

Figure 5.7: Results of GraphSmart in the irrigation consortium network. In panel (a) we have the line graph where the color of each node corresponds to its centrality, measured by page-rank algorithm. In panel (b) we report the MSE between reconstructed and ground truth node values as function of the number of sleeping nodes in two cases, i.e. the ranking with page-rank centrality (in black line with nodes highlighted according to their centrality) and random sorting (in blue).

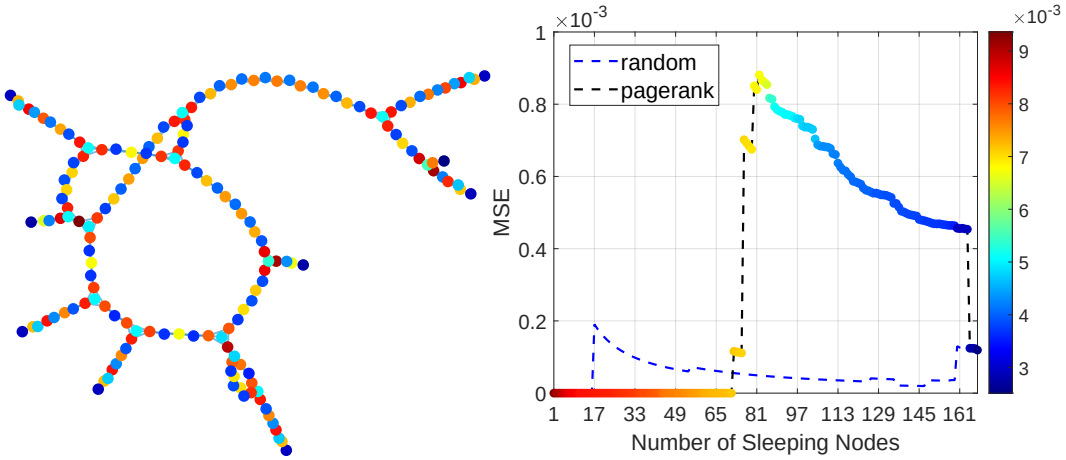
there is a link connecting the node i to the node j. Then we derive the incidence matrix $\mathbf{B}$ and the adjacency matrix of the associated line graph $\mathbf{A}_L$ as in Eq. (5.1) and rank the nodes in this transformed space according to their centrality, measured by the page-rank algorithm.

In Fig. 5.7a, the line graph associated with the irrigation consortium network is depicted, with each node's color corresponding to its centrality. It is noteworthy to focus on the least central nodes, depicted in blue, which correspond to physically isolated nodes and are crucial for the reconstruction process.

In Fig. 5.7b, the MSE between reconstructed and ground truth node values is presented as a function of the number of sleeping nodes in two scenarios: one with node ranking based on page-rank centrality and the other with random sorting. The findings demonstrate that by selecting the appropriate sorting method for eliminating node values, it is feasible to have up to 73 out of 85 sleeping nodes. In other words, only 12 sensors can be placed on this network consisting of 85 nodes, yet the network can still be fully reconstructed. It is important to note that if the sequence of sleeping nodes is randomly chosen, the MSE significantly increases, even with the elimination of only 37 sensor values.

We now consider the presence of demand values, signifying the addition of the possibility that nodes may demand water from the network. We compute the adjacency matrix in its extended expression $\mathbf{A}_{ext}$ as shown in Eq. (5.6). In this scenario, the

total number of links is equal to 168, consequently resulting in the line graph having 168 nodes, as depicted in Fig. 5.8a.



(a) (b)

Figure 5.8: Results of GraphSmart to the irrigation consortium network in presence of demand values. In panel (a) we have the line graph where the color of each node corresponds to its centrality, measured by page-rank algorithm. In panel (b) we report the MSE between reconstructed and ground truth node values as function of the number of sleeping nodes in two cases, i.e. the ranking with page-rank centrality (in black line with nodes highlighted according to their centrality) and random sorting (in blue).

The nodes of the line graph are ranked according to their centrality, and this ranking is represented by the color of the nodes in Fig. 5.8 (a). The MSE analysis is then replicated in this scenario, demonstrating that our ranking enables us to have 71 sleeping nodes out of the 168 proposed locations, in contrast to only 17 when randomly eliminating node values.

A few remarks on the presented results are in order. Firstly, adding virtual links (and nodes) to model water demand values does not change the graph topology and for this reason, the number of sleeping nodes, given the topology of the network, does not significantly change in the presence of demands. Secondly, the MSE is a step function, meaning that when the configuration of available node values enables the algorithm to reconstruct the water flow, the MSE is close to zero. Then, as soon as a critical node for the reconstruction is removed the reconstruction algorithm breaks and the MSE is on the same order of magnitude of the water flow. This consideration motivates the oscillations of the MSE around the maximum value since the reconstruction algorithm completely fails.

An integral aspect of the graph network is the integration of real-time sensor data. Each node in the graph can be associated with a specific sensor location, providing a

detailed and up-to-date understanding of the irrigation consortium network's performance. Parameters such as flow rates, pressure levels, and water quality measurements can be linked to their corresponding nodes, facilitating comprehensive monitoring.

5.6 IoT monitoring scenario for GraphSmart

This approach is based on an IoT system of water sensors that are placed at the pipe of the irrigation consortium or WDN to collect measurements of the flow in the network. Considering the growing emphasis on energy efficiency, the graph network can be harnessed to suitable design of sensor placement and data transmission, thereby minimizing energy consumption. The GraphSmart approach preserves the accuracy of network reconstruction while lowering energy consumption. With the aim of achieving this objective, the process involves the identification of critical nodes that offer essential information for flow reconstruction, while simultaneously minimizing energy consumption by sensors. To evaluate the GraphSmart approach in a LoRaWAN network communication scenario, we thoroughly examine its energy-related characteristics.

From the power consumption aspect, an ideal IoT chip needs to have very low power consumption during operation, ultra-low deep sleep current, and long-distance communication capability. The advent of LoRaWAN technology solves the problem of transmission distance, network signal and power consumption, and is the most cost-effective, which is exactly what we are looking forward to in the field of water meter reading applications [276]. Especially, in this chapter we focus on a flow meters deployment.

In the subsequent sections, the definition of an energy model for the LoRaWAN flow meter device is presented. Subsequently, the chosen simulation framework for conducting performance evaluation is introduced. Building upon this selected simulation framework, a precise scenario is established to align with the WDN deployment discussed in the preceding section, as well as the defined energy model.

5.6.1 Coverage model and simulation

Computer modelling and simulation are valuable methods for exploring system performance and evaluating strategies without costly implementation. Simulation becomes particularly valuable in IoT scenarios characterized by numerous nodes and large geographical areas. In the context of LoRaWAN networks, simulation enables the design and evaluation of LoRa-based applications prior to actual deployment. While various

LoRaWAN simulation tools exist in the literature, a study referenced in [315] identifies the NS-3 network simulator as the most suitable option.

NS-3 not only supports LoRaWAN technology but also offers extensive libraries [316]. NS-3 is an open-source discrete-event network simulator written in C++ and Python. The NS-3 simulator supports a wide variety of protocols such as Wi-Fi, LTE, IEEE 802.15.4, SigFox, LoRaWAN, and further networks.

The integrated LoRaWAN module in NS-3 is designed to meet the requirements of Class A devices. Class A devices are known for being the most power-efficient among the three available classes in LoRaWAN (Class A, Class B, and Class C). Key features of this module include installation of the network server, support for Adaptive Data Rate (ADR), confirmed messages, and multi-gateway support.

LoRaWAN NS-3 module considers N LoRaWAN sensors placed in a 2-dimensional space around M gateways. In the simulator, the Receiver Signal Strength Indicator (RSSI) value, associated with the sensor, depends on the distance between the device and the gateway according to the $L_{pl}(d)$ path loss model:

$$L_{pl}(d) = \overline{L_{pl}}(d_0) + 10\eta \log(\frac{d}{d_0}) + \chi_\sigma \quad [dB] \tag{8}$$

where $\overline{L_{pl}}(d_0)$ is the mean path loss at the reference distance d_0, η is the path loss exponent and $\chi_\sigma \sim \mathcal{N}(0, \sigma^2)$ is the normal distribution with zero mean and σ^2 variance to account for shadowing.

Finally, we enable the channel capture effect. This assumption is reasonable when the cell works in stable conditions, and collisions involving multiple overlapping frames are rare or have a dominant contribution in the interfering power as introduced in [317], [318]. Unless otherwise specified, Table 5.3 shows the scenario parameters used in our simulations.

LoRaWAN employs Frequency Shift Chirp Modulation to enable long-distance, low data rate communication within the sub-1GHz ISM bands. In our research, we run experiments within a simulation tool to assess performances across different selected rankings. By profiling real device energy consumption, we accurately depict the energy usage of individual device components.

We focus on a LoRaWAN Class A network configuration, characterized by end devices always initiating transmissions in a completely asynchronous manner. One of the key system parameters in our scenario is the reporting periodicity denoted as T, where each end device is allocated a unique initial reporting delay, following which it generates a new packet every T seconds. In this study, we do not account for Downlink

Parameter	Value
Carrier Frequency (MHz)	863.0
Bandwidth (kHz)	125
Code Rate (CR)	4/5
Message size [byte]	20
Message Period	1 packet every 300 seconds
Number of gateway	2
Number of nodes	85
TXPower	14 dBm
Path loss values	$\eta = 2.9$, $\sigma^2 = 0$, $\overline{L_{pl}}(40m) = -66\ dB$

Table 5.3: LoRaWAN scenario simulation parameters for the realistic WDN considered in this work.

(DL) transmissions, i.e., messages sent from gateways to end devices, we consider this limitation to be relatively minor since the majority of traffic in the selected scenario will be in the Uplink (UL) direction. Communication between sensors and GWs is spread out on different frequency channels and data rates. LoRaWAN uses up to 6 different programmable Spreading Factors (SFs): 7, 8, 9, 10, 11, 12. Furthermore, the adopted bandwidth can be configured: 125 kHz, 250 kHz and 500 kHz (typically 125 kHz for the 868 ISM band). LoRaWAN is based on the LoRa modulation. LoRa uses frequency chirps with a linear variation in frequency over time in order to encode information. It offers a sensitivity of the order of $-130\ dBm$. The Data Rate (DR) depends on the Bandwidth (BW) in Hz, the spreading factor SF and the Coding Rate (CR) [317] as:

$$DR = sf \cdot \frac{BW}{2^{sf}} \cdot CR \tag{5.7}$$

where the symbols/s are given by $BW/2^{sf}$ with $sf \in \{7-12\}$ and the channel coding rate CR is $4/(4+RDD)$ with the number of redundancy bits $RDD = 1, \cdots, 4$.
The symbol duration (sec) is calculated as follows:

$$T_{sym} = 2^{sf}/BW. \tag{5.8}$$

LoRa devices use a higher SF when the signal is weak or there is strong interference in the used channel. If a sensor is far away from a gateway, the signal gets weaker and therefore needs a higher SF. Using a higher SF means a longer symbol duration so a longer ToA, i.e., the total transmission time of a LoRa packet with a consequently higher impact on the power consumption. The selection of the data rate is a trade-off

Device	Transmit	Receive	Sensing	Sleep
FiPy	170 mA	120 mA	110 mA	50 μA
TTGO	111 mA	61 mA	51 mA	30 μA
STM32WL55JC	28 mA	11 mA	1.4 mA	1.5 μA

Table 5.4: Current consumption of the devices in different operating modes.

between communication range and packet duration or power consumption. Moreover, packets transmitted with different SFs, in principle, do not interfere with each other.

To maximize both the battery life of sensors and overall network capacity, LoRaWAN can manage the data rate and RF output for each sensor individually by means of an ADR scheme [319]. This mechanism determines the transmission parameters (SF and transmit power) of the device based on the estimation of the link budget in the uplink and the threshold RSSI for decoding the packet correctly at the current data rate. When the data rate is not achievable the ADR reduces it (by increasing the SF) to provide connectivity to the device. Finally, the network will be optimized to use the fastest data rate possible for each sensor.

In the applied scenario, at the onset of the simulation, each sensor receives an SF assignment according to the following procedure. Initially, we calculate the power level received by each gateway from the sensor. Subsequently, we select the gateway that records the highest received power and determine the SF based on that power level. In this assignment, we ensure that the sensor is assigned the lowest SF that surpasses the sensitivity of the gateway. We refer to this procedure as ADR in the next part of our work. The goal is to optimize communication between the sensor and the gateway while maintaining efficient data transmission.

For what concern the energy model, a LoRaWAN library for Class A in NS-3 has been previously developed by [320], [321] and [322].

In this study, an extensive examination of LoRaWAN sensors regarding energy consumption was conducted. Field experiments were performed to evaluate the current consumption across four operating modes: sensing, transmitting, receiving, and sleeping. Specifically, the power consumption of three devices, including the SX1276 LoRa transceiver, was estimated. Current measurements were carried out using a Tektronix MSO 2024B oscilloscope with a TCP0020 current probe. The supply voltage was set at 3.3 V, and the devices' transmission power was configured at $25; mW$ $(14; dBm)$.

The average current consumption of the three considered sensors in all the four mentioned states is outlined in Table 5.4. In particular, the table shows that in the three active states (transmit, receive, sensing), the power consumption of the

STM32WL55JC device is lower than the FiPy and TTGO devices. The reason for this difference is that the first two are systems specifically designed for prototyping. As such, they include additional components that enhance consumption. After evaluating the energy consumption characteristics of the devices, the decision was made to concentrate the analysis exclusively on the STM32WL55JC device [256]. Indeed, this ultra-low power device is highly suitable for large-scale green IoT deployment [323]. Thus, we use the results to adapt the available energy consumption module in NS-3 to enable the evaluation of LoRaWAN networks in terms of energy efficiency.

5.7 GraphSmart Integrated Performance evaluation

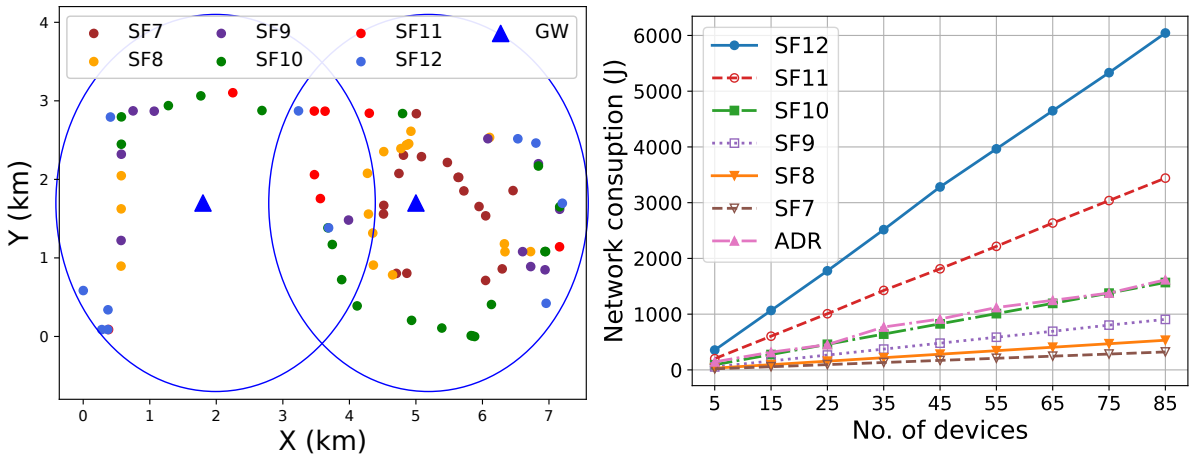
In this section, we present the advantage results of the GraphSmart approach, where we include the energy aspect, on the synthetic WDN presented in Sec.5.5.2. To this goal, we leveraged the NS-3 simulator module and conducted a series of assessments to gauge the performance metrics of the IoT network. These tests encompassed evaluations of network performance, GWs coverage and energy consumption.

5.7.1 WDN LoRaWAN monitoring scenario

The primary objective of the initial simulation campaign was to assess the impact of the selected SF on energy consumption. Energy consumption was determined by monitoring the battery level of the devices, initially set at 10,000 joules. Each device transmitted data at an average interval of 5 minutes and the battery level of each device was observed over a 24-hour experimental period. The exact interval time was uniformly distributed within the range of 4 to 6 minutes, with the value extracted at the end of each transmission. Finally, we calculated the total energy expended across the entire LoRaWAN network by summing up the energy consumption of the network devices.

Our focus in this study revolves around monitoring the flow rate within each pipe. To achieve this, we position a sensor at the end of every link. As a result, the number of devices exceeds the number of junctions, since multiple pipes may terminate at the same junction. In such cases, multiple devices are positioned at the same location to monitor the flow at the end of multiple pipes. In the end, we placed 85 monitoring LoRaWAN devices.

Lastly, based on the chosen network topology, we determine the optimal placement of GWs to efficiently collect the monitoring data. The considered GWs positions are shown in Fig. 5.9 (a) (triangular markers are the GWs) together with the coverage area when SF=12 (maximum coverage area) is configured (represented by big blue circles). According to the coverage area, each GW is able to receive only a subset of the total devices measurements. Fig. 5.9 (a) also shows the location of the nodes and



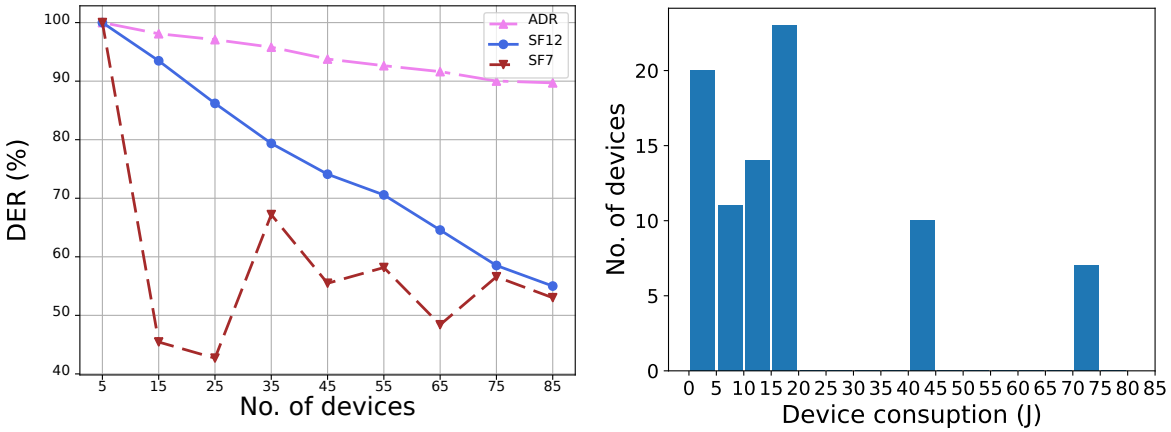
(a) (b)

Figure 5.9: (a) Devices deployment and SF values assigned by ADR in the proposed irrigation consortium network. (b) Total energy consumption in joule of the network as a function of the number devices present.

the SF values assigned by the ADR. For instance, the devices represented by the black dots had an SF of 7 and were placed near the GWs, while it assigns higher SF values to devices farther from the GWs, such as the devices represented by the yellow dots, which had an SF of 12. In the figure, the legend exactly depicts the color assigned to each SF. Furthermore, Fig. 5.9 (b) presents an analysis of the network's energy consumption as a function of the number of sleeping devices, considering various scenarios where ADR can be enabled or disabled, when ADR is disabled devices are configured to use the same SF. It was conducted multiple simulations involving devices with a constant SF ranging from 7 to 12, and with ADR enabled. The objective was to evaluate the impact of the selected SF on the energy efficiency of the entire network. In order to conduct the experiment, we randomly removed devices following their initial order, with a regular interval of 5 steps.

Higher network energy consumption was observed with $SF12$ configuration for the devices, while lower consumption was observed with $SF7$, this is due to the transmission duration of each device, predominant when the SF is higher. The line depicted

by triangular markers corresponds to the scenario where ADR is enabled. In this case, the SF configuration of the devices is set as shown in Fig. 5.9 (a). It should be noted that this particular scenario does not represent the optimal energy efficiency. However, it is important to observe that low consumption with SF set to 7 (optimal energy efficiency) does not necessarily imply optimal network efficiency. Indeed, it is also crucial to evaluate the packet delivery ratio.



(a) (b)

Figure 5.10: (a) DER as a function of the number of EDs in three representative scenarios, enabled ADR, all devices configured with SF=12 and SF=7. (b) Frequency distributions of devices numbers based on their consumption in the presence of ADR.

To this end, we extract the probability of correctly receiving a packet, which is a typical metric considered for characterizing LoRaWAN systems (often called Data Extraction Rate - DER) of the main three configured scenarios. Specifically, Fig. 5.10 (a) illustrates the trend of DER as a function of the number of devices in the network in three different scenarios: in the presence of active ADR (up triangular markers), when all devices are set to SF7 (down triangular markers), and SF12 (circular marker). Based on these configurations, it can be observed that the DER parameter for SF7 and SF12 drops below 60% when the two gateways serve more than 75 devices. Both scenarios negatively impact the performance of the devices, albeit for different reasons. In the case of SF=7 configuration, a significant number of devices are left out of coverage, resulting in a lower number of packets being received from gateways. On the other hand, when SF=12 is configured, the longer duration of the packets leads to multiple collisions, thereby reducing the probability of successful reception.

Conversely, with ADR active, the DER remains above or equal to 90%, allowing for a maximum number of serviceable devices up to 85. By, comparing Fig. 5.9 (b)

and Fig. 5.10 (a), results indicate that the ADR increases data delivery success within the network while maintaining an acceptable energy consumption, thereby increasing network energy efficiency.

To better visualize the consumption of the network devices, Fig. 5.10 (b) shows the frequency distribution of network devices according to their 24-hour energy consumption in the presence of ADR. Specifically, it is evident that the majority of devices within the studied network exhibit a daily energy consumption of less than 20 joules, with these devices predominantly utilizing lower SF values. Conversely, a subset of devices, approximately 18 out of the total 85, exhibit higher daily energy consumption, attributed to their configuration with SF values of 11 and 12. We note that in LoRaWAN, the duration of a packet doubles as the SF value increases by one unit. This exponential increase in duration is the reason behind the corresponding exponential increase in power consumption.

The introduction of ADR once again demonstrated notable improvements in energy efficiency. For instance, if all devices had been consistently configured with an SF of 12, the entire set of 85 devices would have incurred an approximate daily energy consumption of 71 joules per device, corresponding to 6035 joules when considering the entire network. As a result, it can be concluded that the adoption of ADR represents the optimal trade-off between energy consumption and network throughput in this context. Nevertheless, energy efficiency can be further enhanced through the GraphSmart approach, as discussed in the following subsection.

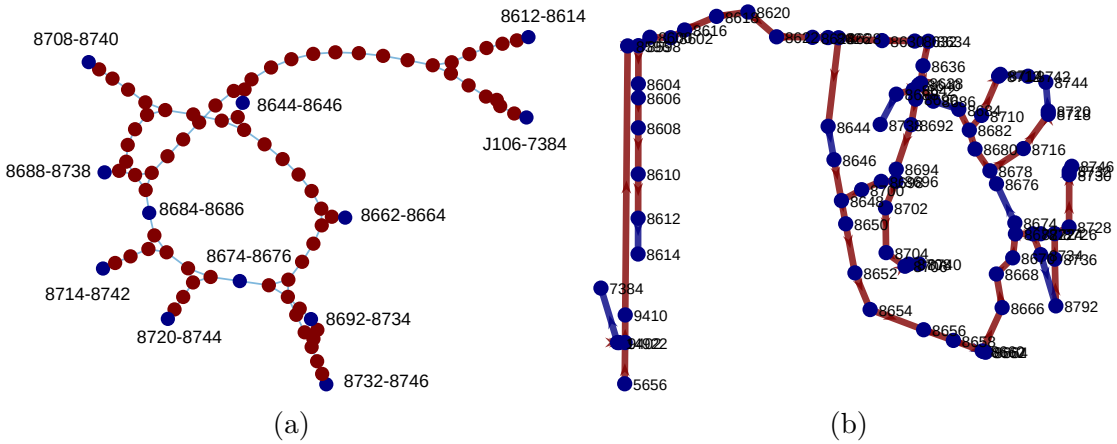


Figure 5.11: The sensor deployment configuration at the critical step. In panel (a) we have the line graph associated to the irrigation consortium network. In red we have sleeping nodes while in blue we have nodes corresponding to active sensors. Links needed to reconstruct the signal are identified through the start and end node. In panel (b) we report with the same colour code the configuration in the original graph and we observe the link position of the essential nodes (represented in blue) for the reconstruction algorithm.

5.7.2 GraphSmart energy saving

This subsection presents a comprehensive analysis of the four ranking strategies presented in Sec. 5.4.3. This analysis focuses on evaluating the performance of the GraphSmart strategies in terms of both flow reconstruction MSE and energy efficiency. The goal is to identify the most effective optimization strategy for selecting an optimal sensor deployment in the irrigation consortium network. The MSE is regarded as a measure of reconstruction performance. It has been demonstrated in Section 5.5.2 that it exhibits characteristics akin to a step function, with its values undergoing rapid fluctuations under critical conditions where a necessary node value is absent. To stress this point, we report in Fig. 5.11 the sensor deployment configuration at the critical step. It means that the presented figure is the configuration just before that the algorithm breaks. In panel (a), we observe the line graph, while in panel (b), the original graph is depicted. Active sensor nodes are represented in blue, while missing measurements slated for reconstruction are denoted in red. Notably, the link removal that disrupts the algorithm occurs between nodes 8644 and 8646.

The results of four ranking strategies are compared, analyzing their effects on two primary metrics: MSE and energy consumption. Through this evaluation, we endeavor to offer insights into identifying the most appropriate approach for attaining an optimal deployment. To streamline the analysis, we provide a visual representation in the form of a figure illustrating the MSE and energy consumption outcomes for each ranking strategy.

Fig. 5.12 illustrates a comparison between the MSE of flow reconstruction and energy consumption for the four rankings, *pagerank*, *closeness*, *betweeness* and random. Two subplots sharing the same x-axis are represented in the figure. The first subplot shows the MSE error of the network flow reconstruction versus the network energy consumption; the MSE is expressed as a logarithmic scale and the error is reduced when the energy consumption increases. The network energy consumption information is reported in the sharing axis.

The second subplot provides the number of sleeping devices as function of the energy consumption. Intuitively, the energy consumption is high when few sensors are removed, while it decreases when the number of sleeping nodes increases. Both the subplots show four lines, one for each ranking modality applied to remove sensors from the network.

With the aim to better understand the energy-reconstruction performance of GraphSmart we can discuss Fig. 5.12 in its entirety. In the initial subplot, a tolerated value for the MSE can be set, and the corresponding logarithmic value can be identified

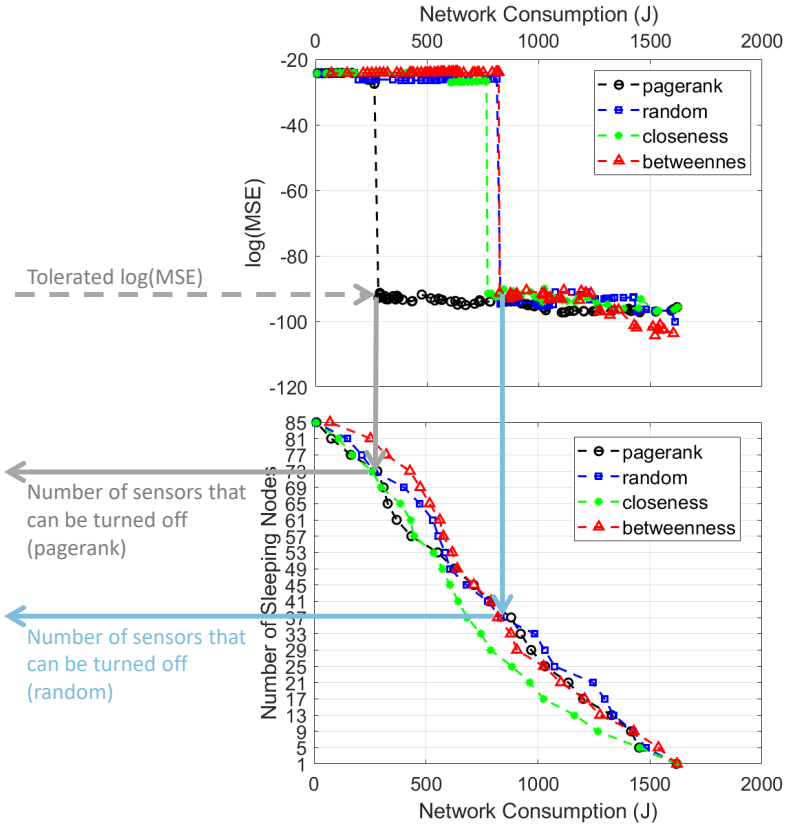


Figure 5.12: Integrated results of efficient and accurate GraphSmart. The top panel shows the MSE in logarithmic scale vs. the energy consumption of the network. The bottom panel presents the number of sensors that can be removed as function of the network energy consumption. The two figures share the same x-axis and they can be read together. We can select the tolerated MSE (represented with the dot line) and derive the network consumption associated to the centrality selected algorithm. Following the arrows on the left we obtain the number of sensors that can be removed to obtain the desired reconstruction error. In grey we have the pagerank and in blue the random cases.

utilizing the provided arrows in the top figure. Subsequently, the network energy consumption linked with the specified error value for the particular ranking procedure being examined can be determined. Finally, the number of sensors that can be deactivated can be read on the y-axis of the bottom figure, aligning with the derived network consumption.

Ranking	**Installed devices**	**Energy saving (%)**
Random	48	48.81
Betweenness	48	49.35
Closeness	48	58.11
Page-rank	20	73.01

Table 5.5: Ranking comparison with associated energy saving percentages.

The findings show that page-rank is the method that produces the best-sorted devices list. Indeed, after fixing the value of the tolerated error, the page-rank sorting algorithm enables us to turn off the largest number of devices. This indicates that using page-rank as an optimization strategy for choosing the optimal sensors deployment in the irrigation consortium network yields the best results in terms of both MSE and energy consumption. In particular, Table 5.5 highlights that the page-rank algorithm stands out as the best in terms of energy savings. This ranking is capable of reconstructing the irrigation network flow with only 20 installed devices and remarkable energy savings of 73.01%.

Taken collectively, it is demonstrated that GraphSmart, represents an effective approach for optimizing the deployment of resources within an irrigation consortium network, adept at balancing flow reconstruction accuracy and energy efficiency.

5.8 Conclusion and Future Work

The main purpose of this chapter is to introduce a novel method to optimize the monitoring of smart irrigation consortium network by an effective IoT sensor placement. The irrigation consortium network consist of physical pipes enabling the water flow and IoT sensors measuring features such as flow rates, pressure levels and water quality parameters. The hydraulic connectivity between nodes and pipes can be modeled as a graph by possibly adding other physical properties, such as the resistance and the diameter of pipes. In this scenario, sensor measurements acquired by an IoT network can be modelled as signals defined in a networked domain.

In this chapter, the line graph corresponding to the physical graph is analyzed, and a graph signal processing-based approach is introduced for flow reconstruction utilizing a reduced set of sensors. Specifically, node centrality measures computed in the line graph are employed to determine sensors that can be deactivated without compromising accuracy. The proposed method, referred to as GraphSmart, is evaluated through numerical simulations conducted in a realistic LoRaWAN network scenario. It is demonstrated that accurate flow reconstruction is achievable through strategically positioning a reduced number of sensors, resulting in substantial energy savings, approximately 73%.

The GraphSmart method has been applied to an irrigation consortium network, assuming that the IoT measurements are repeated at a pace related to the WSS stationarity interval. In future work, we can further to reduce the number of IoT sensors by designing joint subsampling of the measurements both in the time and

space domains. This can be accomplished by a multilayer vertex-time representation of the irrigation consortium network, where it is possible to jointly select samples in the vertex and time domain. This would extend the analysis on GraphSmart sensor displacement to identify the time resolution of IoT devices readings. Future work will address the analysis of a multilayer network representing the physical flow constraints along time, in order to identify the location and timing of IoT readings and improve the WSS flow reconstruction accuracy.

Taken together, the GraphSmart method offers versatile and insightful mapping of water distribution networks. By encompassing topology, hydraulic characteristics and advanced optimization techniques, our proposed approach contributes to a holistic understanding of the network's complexities, supporting endeavors to enhance efficiency, accuracy, and energy conservation within water distribution systems. In conclusion, GraphSmart represents a potentially useful tool for water operators by optimizing network node positions, which in turn can lead to improvements in water use efficiency, thanks to flow reconstruction.

Chapter 6

Machine Learning Models to Predict Daily Actual Evapotranspiration of Citrus Orchards

6.1 Overview

Precise estimations of actual evapotranspiration (ET_a) are essential for various environmental issues, including those related to agricultural ecosystem sustainability and water management. Indeed, the increasing demands of agricultural production, coupled with increasingly frequent drought events in many parts of the world, necessitate a more careful evaluation of crop water requirements.

Artificial Intelligence-based models represent a promising alternative to the most common measurement techniques, e.g. using expensive Eddy Covariance (EC) towers. In this context, the main challenges are choosing the best possible model and selecting the most representative features. The objective of this research is to evaluate two different machine learning algorithms, namely Multi-Layer Perceptron (MLP) and Random Forest (RF), to predict daily actual evapotranspiration (ET_a) in a citrus orchard typical of the Mediterranean ecosystem using different feature combinations. With many features available coming from various infield sensors, a thorough analysis was performed to measure feature importance, scatter matrix observations, and Pearson's correlation coefficient calculation, which resulted in the selection of 12 promising feature combinations. The models were calibrated under regulated deficit irrigation (RDI) conditions to estimate ET_a and save irrigation water. On average up to 38.5% water savings were obtained, compared to full irrigation. Moreover, among the differ-

ent input variables adopted, the soil water content (SWC) feature appears to have a prominent role in the prediction of ET_a. Indeed, the presented results show that by choosing the appropriate input features, the accuracy of the proposed machine learning models remains acceptable even when the number of features is reduced to only 4. The best performance was achieved by the Random Forest method, with seven input features, obtaining a root mean square error (RMSE) and a coefficient of determination (R^2) of 0.39 mm/day and 0.84, respectively. Finally, the results show that the joint use of SWC, weather and satellite data significantly improves the performance of evapotranspiration forecasts compared to models using only meteorological variables.

6.2 Introduction

According to the recent global report on water use published by UNESCO, irrigation represents about 70% of the global consumption of available freshwater [324]. Therefore, adopting sustainable agriculture is of paramount importance to minimize water consumption. In this context, pushing the agricultural system as a whole toward ecologically sustainable solutions is a major challenge given the increasing insufficiency of water availability [325]. Recent research suggests that farmers should be encouraged to adopt new solutions, particularly in drought-prone regions, in order to optimize both water quantity and quality, and ensure less water consumption for a more environmentally friendly future for the next generations [326].

In particular, in the Mediterranean ecosystems, which are characterized by dry and hot summers and rainfall mainly occurring in fall and winter, it is important to adopt sustainable irrigation strategies to increase water use efficiency and preserve water resources. The optimization of irrigation water employment in agriculture can be obtained, from the one hand, by estimating the exact crop water requirement and, from the other, through the application of dynamic irrigation strategies, such as the regulated deficit irrigation (RDI). The main goal of these strategies is to save water by controlling in real time the field irrigation as a function of the water status of soil and plants. This can be quantified by means of heterogeneous sensors and data aggregation techniques, in order to identify conditions for reducing the water consumption without affecting the crop growth. For example, it has been proved that RDI can be applied to citrus orchards during the stage II of crop without determining significant impact on crop yield [327–332]. However, the application of RDI strategies requires accurate monitoring to avoid severe crop water stress which can produce a decline in yield and/or irreversible effects on crop growth.

From field sensors to water needs. In order to monitor the crop water needs in real-time, several data sources for estimating the field status can be considered, among which meteorological data, soil water contents, drill and drop sensors, satellite images and vegetation indices. Different models for studying the soil water balance (SWB) are then applied for mapping infield data into a precise estimation of the water needs [333]. An important parameter for this estimation is the evapotranspiration (ET_a), which is the combination of two distinct processes: water evaporation from the soil surface, E_a and water transpiration from the plant canopy, T_a, depending on the climate and soil water status [334].

Recently, several data-driven models (artificial neural networks, K-nearest neighbors, random forest, etc.) have been proposed for predicting evapotranspiration even with a limited amount of infield sensors, i.e. by working only with some low-cost typologies of sensors or with a limited spatial granularity of measurements [335–338]. Models are customized not only for different climates, but also for different typologies of crops, such as potato crops [335, 339, 340], green pepper crops [341], or cereal crops [340, 342–346]. The models can be studied in combination with the irrigation policy, for minimizing the water consumption, as described in [347] for the case of a rice crop. However, these works focus on herbaceous crops or horticultures only. Other models for estimating the field water needs are based on the study of the surface energy balance, by processing land surface temperature (LST) [348]. These approaches have been demonstrated in the case of olive and pomegranate orchards and vineyards but, to the best of our knowledge, there is no previous literature focusing on the definition of models for citrus orchards, despite the fact that in the Mediterranean basins citrus is one of the most cultivated crops [349]. Citrus orchards have high water requirements [331, 350] and differ significantly from olives and pomegranates crops, which are characterized by a high capacity to resist in arid environments [351, 352].

Reference, potential and actual evapotranspiration. It is important to note that most existing literature employs machine learning (ML) models to predict the crop *reference* evapotranspiration (ET_o) or potential evapotranspiration (ET_c), whereas the focus of this chapter is on *actual* evapotranspiration (ET_a), which plays a key role in the quantitative evaluation of the actual crop water requirements, necessary for irrigation water management. While ET_a values are considered in ecosystem studies, for studying wetland conservation [353] and water cycles [354], the comprehensive reviews of evapotranspiration models presented in [355, 356] focus on ET_o only. In [355], which compares several papers from 2009 to 2021, the main goal of ET_o estimation is developing intelligent irrigation systems, while in [356] it is demonstrated how

Table 6.1: Literature comparison of models to predict **potential** evapotranspiration (ET_c) considering $RMSE$ and R^2

References	Model	Crop or vegetation	Length of dataset (days)	Number of input features	$RMSE$ (mmd^{-1})	R^2
[346]	DNN	Wheat	17531	3	0.15 - 0,42	0,94 - 0,97
[339]	kNN, ANN, AdaBoost	Potato	240	1 - 4	0,24 -1,01	0.68 - 0.96
[338]	kNN, SVM, RF, AdaBoost	Sugar beet	340	3 - 7	0,22 - 1,13	0.79 - 0.99
[335]	ANN, NNGA , MNLR	Potato	990	1 - 6	0,05 - 0,35	0.43 - 0.96
[344]	MLR, BP	Maize, wheat soybean	900	4 - 5	1,19 - 1,52	0.69 - 0.90
[345]	MLP, RBF	Barley	200	4	$NRMSE$ 0.23 - 0.31	0,89 - 0,93
[343]	ANN	Wheat, maize	250 - 430	4 - 8	$NRMSE$ 0,10 - 0.54	0.86 - 0.99

Notes: RF = Random Forests, SVM = Support Vector Machines, ANN = Artificial Neural Network, MLR = Multiple Linear Regression, kNN = k-nearest neighbors, MLP = Multi-layer Perceptron, AdaBoost = Adaptive Boosting, NNGA = Neural Network–Genetic Algorithm, MNLR = Multivariate Nonlinear Regression, BP = Back-Propagation neural network, DNN = Deep Neural Network, RDF = Radial Basis Functions, NRMSE = Normalized Root Mean Squared Error.

ET_o models developed from 2007 to 2019 can be used for a wide climatic range. Despite the number of proposed models, several issues for the application of these models are still open. In fact, evapotranspiration is a nonlinear and complex phenomenon, and its estimation is based on the availability of several climatic and crop parameters and their mutual interactions with each other. Consequently, the transition from ET_o to more significant quantities such ET_a is not straightforward. Indeed, while ET_o can be derived by using only climatic variables, ET_a depends on the specific crop, as well as on soil and plant conditions. Tables 6.1 and 6.2 summarize the main characteristics of related works on potential evapotranspiration and actual evapotranspiration estimation, respectively, by comparing the approach, crop typology, dataset size, number of input features and model accuracy.

Impact of hydrological factors. Recent papers like [357, 337, 359, 358] used ML to study ET_a only from a generic hydrology or atmospheric point of view, and do not consider irrigation strategies like RDI. For example, [360] models ET_a using ANN and Genetic Programming in a Canadian landscape covered by spontaneous vegetation; consequently, the study does not analyze water requirements in a specific crop and does not consider possible irrigation strategies. In [361] the authors quantified the impact of some hydrological factors on ET_a using Bayesian model averaging for forest, cropland, and grassland ecosystems. In addition, it was demonstrated that soil

Table 6.2: Literature comparison of machine learning models to predict **actual** evapotranspiration (ET_a) considering $RMSE$ and R^2

Reference	Models	Crop or vegetation	Length of dataset (day)	Number of input features	$RMSE$ (mmd^{-1})	R^2
[336]	M5P, BAGGING, RF, SVR	Pastures grass	1825	3 - 6	0.18 - 0.40	0.93 - 0.98
[337]	ANN	Different land cover	11713	2 - 5	0.39 - 0.78	0.98 - 0.99
[341]	ENN	Green peppers	800	10	0.35 - 0.61	0.86 - 0.97
[340]	RF, LSTM	Corn, soybeans, potatoes	6208	5 - 16	0.40 - 1.30	0.42 - 0.70
[357]	LSTM	Grassland, Forest Alpine meadow	300 - 500	4 - 8	0.21 - 1.06	0.16 - 0.80
[358]	LR - EFS	Grassland Open shrubland Barren vegetation	4017	5 - 18	0.42 - 1.81	0.15 - 0.82
[353]	RF, MLP, kNN, ARDS	Sawgrass	2069	3 - 7	0.42 - 1.06	0.42 - 0.90
[354]	RF	Grassland	Datasets of various sizes	21	0.28 - 0.73	0.64 - 0.89
[359]	RF	Different land cover	Datasets of various sizes	7	0.99 - 2.18	0.32 - 0.75
[360]	ANN, GP, SAS/STAT	Spontaneous flora	150	5 - 9	0.06 - 0.10	R: 0.71 - 0.88
[361]	BMA	Grassland Cropland Forest	Datasets of various sizes	7	0.32 - 0.67	R: 0.83 - 0.97
Present work (2023)	RF, MLP	Citrus orchard	576	4 - 10	0.39 - 0.57	0.69 - 0.84

Notes: M5P = Quinlan's M5 algorithm or M5P regression tree, RF = Random Forests, SVR = Support Vector Regression, ANN = Artificial Neural Network, ENN = Elman Neural Network, LSTM = Long short-term memory, LR = Linear Regression, EFS = Exhaustive Feature Selection, MLP = multilayer perceptron, k-NN = k-nearest neighbors, ARDS = Additive Regression of Decision Stumps, GP = Genetic Programming, SAS/STAT = Statistical model (multiple regression equation), BMA = Bayesian Model Averaging, Cub = Cubist package in R, Sin = sinusoidal models.

moisture does not contribute significantly to ET_a in forest areas because vegetative transpiration comprises a large portion of ET_a. Finally, the work in [341] combines ET_a predictions with RDI, but only for horticulture.

Research contributions. Taking into account the limitations of current ET_a models, the aim of this work is defining a data-driven model for predicting actual evapotranspiration of a citrus orchard under regulated deficit irrigation. To this purpose, we exploit a large dataset of infield measurements collected in a suburb of Palermo, in Italy, which includes meteorological data, Vegetation Indices (VIs), soil water contents, as well as direct measurements of ET_a (to be used as a ground-truth). A variety of input combinations are explored, analyzing feature importance and Pearson's correlation coefficient, and studying the performance of several ML models (namely, MLP and RF). Interestingly, the presented results show that the accuracy of the proposed ML models remains acceptable even reducing the input features down to four. To the

best of our knowledge, we are the first to estimate ET_a in a hot summer Mediterranean ecosystem under RDI, with an average water saving of up to 38.5%, compared to full irrigation management. These estimation models are fundamental to understand and reduce the impact of climate change and water scarcity, thus promoting sustainable agricultural irrigation solutions.

6.3 Background and motivation

The application of irrigation strategies such as RDI during specific stages of crop growth could result in the optimization of irrigation water use in agricultural ecosystems. The actual crop water requirement can be assessed, among others approaches, by the continuous monitoring of soil water content, as investigated by [362], or using soil water balance (SWB) models [334]. In the former, the quality of the results depends on the accuracy and acquisition time of the soil moisture sensors, while the latter requires the accurate estimation of actual crop evapotranspiration, ET_a. Direct measurements of ET_a can be acquired by weighing lysimeters [363], which however are characterized by high installation and maintenance costs. On the other hand, indirect methods to estimate ET_a include the widely used dual crop coefficient approach, suggested in the FAO-56 paper [364], and the application of the surface energy balance (SEB), such as the Surface Energy Balance Algorithm for Land (SEBAL) [365], and Two-Source Energy Balance Model (TSEB) [366, 367]. According to the dual crop coefficient approach [364] the maximum ET_a can be estimated as the product of crop reference evapotranspiration (ET_o) and the term ($K_{cb} + K_e$), being K_{cb} the basal crop coefficient and K_e the soil evaporation coefficient. The K_{cb} coefficients for the different crops and stages of vegetative growth, tabulated in the FAO-56 paper [364], have been recently updated by [368] to account for the research published in the last 20 years. ET_o represents the atmospheric evaporative demand, whose values can be computed based on air temperature, solar radiation, wind speed and relative air humidity acquired by a standard weather station [369]. When the soil water content is limiting evapotranspiration, it is necessary to include two additional coefficients, K_s and K_r, with values ranging between 0 and 1, to reduce crop transpiration and soil evaporation, respectively. In a recent review, [370] showed the possibility to derive the basal crop coefficient, K_{cb}, based on linear and non-linear relationships employing, as independent variables, vegetation indices (VIs) derived by remote sensing data in the visible and near-infrared domains, whereas earlier, [371] proposed to estimate K_{cb} based on a linear function of VIs obtained in the shortwave region.

The SEB models are used to estimate latent heat flux (LE) as the residual term of the energy balance equation. The values of LE measured by the Eddy Covariance (EC) tower have been largely used to assess ET_a. The common configuration of EC towers consists of an infrared open-path gas analyzer to measure H_2O vapor and CO_2 concentration in the atmosphere, a 3D-sonic anemometer to measure the sonic air temperature and the three components of wind speed, a 4-components net radiometer to measure the net radiation and a flux plate to measure the soil heat flux. Although both the water and energy balance models have been largely used to estimate ET_a, the first one requires a large number of meteorological variables to evaluate ET_o (not always available and with acquisition times not enough adequate for the purpose [339]), whereas the second, despite the smaller number of required variables, needs the availability of an EC tower equipped with quite expensive instruments. Furthermore, the quality and temporal continuity of the acquisitions can be influenced by the different measurement conditions from the theoretical assumptions, as well as by the malfunction of the instruments caused by the meteorological conditions and also by the imperfect calibrations of the sensors [372].

The limitations of the classical methods described above indicate the need to assess alternative procedures to increase the availability and accessibility of ET_a measurements. For this reason, ML algorithms can be powerful tools for predicting actual evapotranspiration. Only recently, researchers applied ML to predict two important quantities, ET_c and/or ET_a, which are of paramount interest for crop irrigation management. Indeed, approximately 90% of the papers listed in tables 6.1 and 6.2 have been published within the past four years. Specifically, table 6.1 provides a summary of the research using ML models to estimate potential evapotranspiration (ET_c), which is considered a good approximation of ET_a for healthy crops (no diseases) and without water stress. From the table, the ET_c estimation accuracy, measured in terms of $RMSE$ and R^2, ranges on average from 0.37 mmd^{-1} to 0.89 mmd^{-1} and between 0.75 and 0.96, respectively. Similarly, table 6.2 reports papers focusing on ET_a, with average ranges of $RMSE$ and R^2, from 0.37 mmd^{-1} to 0.97 mmd^{-1} and between 0.55 and 0.86, respectively. However, these results were mostly obtained in forest, cropland and grassland ecosystems, characterized by an homogeneous soil cover condition. Instead, in this work we focus on the prediction of ET_a in a citrus orchard under regulated deficit irrigation, in a Mediterranean ecosystem. To the best of current knowledge, machine learning models have been employed for the first time to predict ET_a in tree orchards.

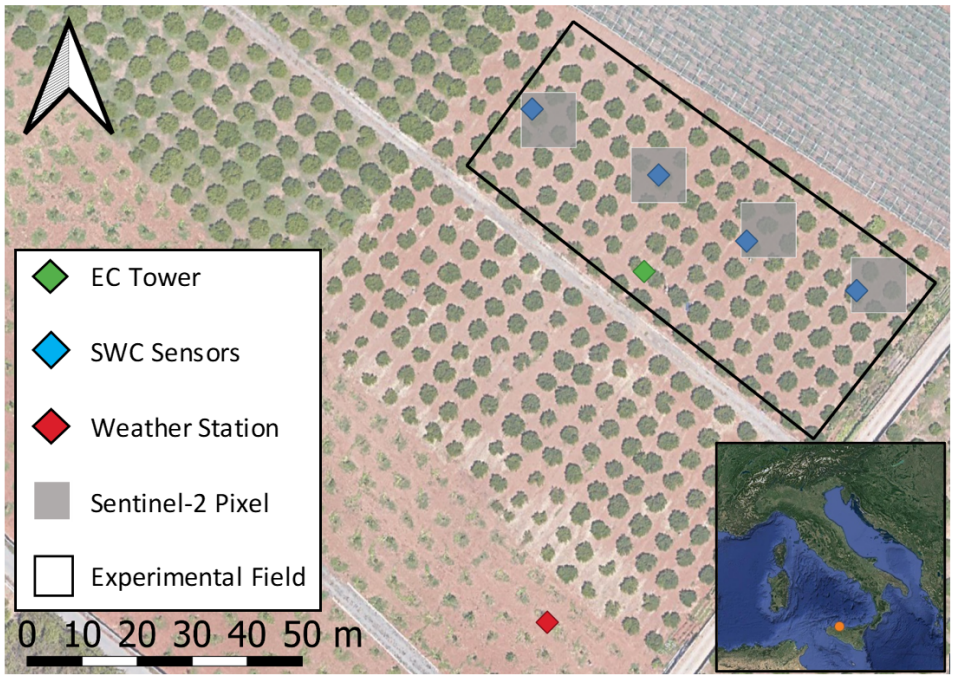


Figure 6.1: Map of the experimental site showing the location of the weather station (WD), flux tower (EC), and drill and drop soil water content sensors; the four Sentinel-2 (L2A/L2B) pixels are also shown.

6.4 Materials and methods

6.4.1 Description of the study area and experimental field

The experiment was carried out for four years (2018-2021) in a citrus orchard (Citrus reticulata Blanco, cv. Mandarino Tardivo di Ciaculli) located in a suburb of Palermo, Italy (38°4' 53.4" N, 13° 25' 8.2" E), as shown in Fig. 6.1. The field extension is about 0.4 ha, with trees planted at a spacing of 5.0×5.0 m and an average height of about 2.5 m. The average fraction cover is about 0.48. The field is generally irrigated with a subsurface drip system, operating from 2018, with two lateral pipes, which contain co-extruded emitters discharging 2.3 l/h at a pressure of 100 kPa with a spacing of 1.0 m (i.e.,10 $emitters/tree$), per plant row, one on each side of the tree, at 1.1 m from the trunks, installed at 0.30 m depth. However, the old micro-sprinklers irrigation system, operating until 2017, is occasionally activated mainly before weeding. The regulated deficit irrigation (RDI) strategy is described in detail in [373]. The climate is Mediterranean, with rainfall generally concentrated in fall and

winter, and high temperatures in summer. According to the recent version of Köppen climatic classification, the zone has a hot summer Mediterranean climate (Csa) [374].

A standard WatchDog 2000 weather station (WS) (Spectrum Technologies, Inc., Aurora, IL, USA) is installed near the field (see Fig. 6.1) to collect, every 30 minutes, the values of air temperature, $T\,(^{\circ}C)$, global solar radiation, $R_s\,(MJm^{-2}d^{-1})$, relative air humidity, $RH\,(\%)$, wind speed and direction at 2 m height, $U_2\,(ms^{-1})$, and rainfall height, $P\,(mm)$. Since March 2019, measurements of actual evapotranspiration, $ET_a\,(mm)$ have been acquired by an Eddy Covariance (EC) flux tower (Fig. 6.1). The tower is equipped with a 4-components net radiometer (CNR4, Campbell Scientific Inch., Logan, Utah) installed at 3.0 m height, a three-dimensional sonic anemometer (CSAT3-D, Campbell Scientific Inch., Logan, Utah) and an infrared open patch gas analyzer (Li-7500, Licor bioscience inch., Lincoln, Nebraska) to measure, respectively, the net radiation, $Rn\,(Wm^{-2})$, with a frequency of 30 min, as well as the 3D-components of wind speed and the concentrations of H_2O vapor and CO_2 in the atmosphere with a frequency of 20 Hz. All the high and low-frequency data are stored in a CR3000 datalogger (Campbell Scientific Inch., Logan, Utah) equipped with a 2GB memory card. Sensible, $H\,(Wm^{-2})$ and latent, $\lambda ET\,(Wm^{-2})$ heat fluxes were evaluated as:

$$H = \rho c_p \sigma_{WT} \tag{6.1}$$

$$\lambda ET = \lambda \sigma_{WQ} \tag{6.2}$$

where $\rho\,(g/m^3)$ is the air density, $c_p\,(Jg^{-1}K^{-1})$ is the air specific heat capacity at constant pressure, $\sigma_{WT}\,(mKs^{-1})$ is the covariance between vertical wind speed and air temperature, $\lambda\ (Jg^{-1})$ is the latent heat of vaporization and $\sigma_{WQ}\,(gm^{-2}s^{-1})$ is the covariance between vertical wind speed and the water vapour density.

To estimate daily crop reference evapotranspiration, $ET_o\ (mmd^{-1})$, the FAO-56 Penman-Monteith (PM) equation [364] was used:

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma(\frac{900}{T_a} + 273)(U_2(e_s - e_a))}{\Delta + \gamma(1 + 0.34U_2)} \tag{6.3}$$

where $\Delta\,(kPa/^{\circ}C)$ is the slope of saturation vapour pressure curve, $R_n\,(MJm^{-2}d^{-1})$ is the net radiation at the crop surface, $G\ (MJm^{-2}d^{-1})$ is the soil heat flux density, $e_s - e_a\ (kPa)$ is the actual vapour pressure deficit, $\gamma\,(kPa^{\circ}C^{-1})$ is the psychrometric constant and $U_2\,(ms^{-1})$ is the wind speed measured thus at 2 m height. When considering the single crop coefficient approach, the values of ET_a can be obtained as:

$$ET_a = (K_c K_s) ET_o \tag{6.4}$$

where ET_o is the crop reference evapotranspiration, K_c is the crop coefficient and K_s is the water stress coefficient. The contribution of soil evaporation and crop transpiration is represented by a single K_c accounting for the difference between the reference crop and the considered crop in terms of biophysical characteristics such as canopy properties, ground cover and aerodynamic resistance. The water stress coefficient, K_s, ranging between 0 and 1, is introduced as a multiplicative factor to take into account the actual soil water status. Based on the irrigation strategy adopted in the field, the value of K_s resulted generally equal to 1.0, except during phase II of fruit growth (beginning of July-mid August), when $K_s < 1$ as a consequence of the limited water application.

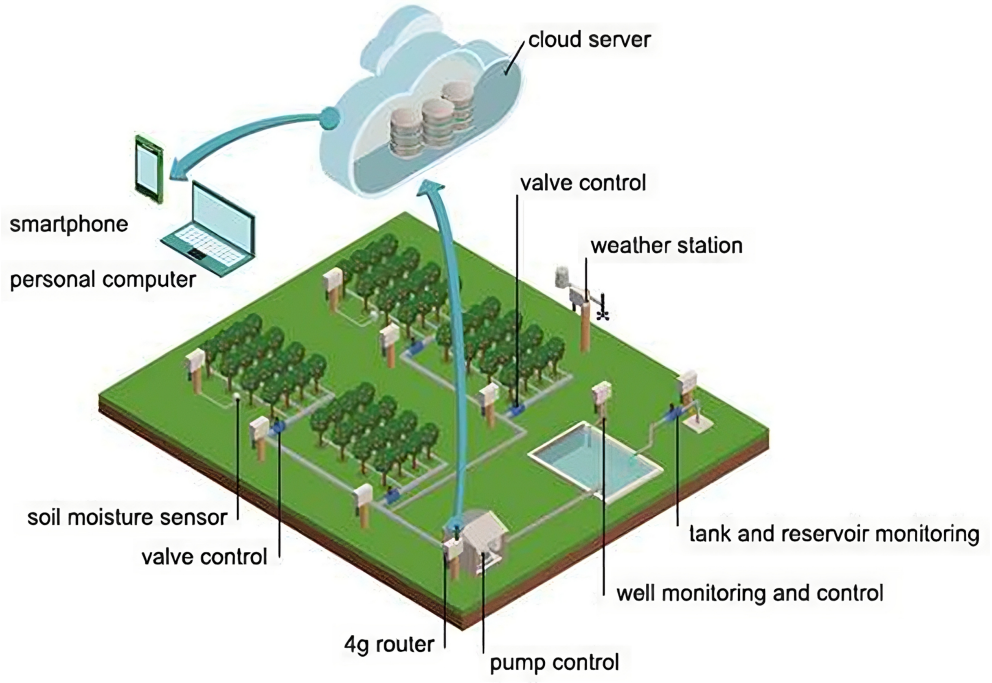


Figure 6.2: IoT monitoring system deployed in the citrus orchard.

The dynamic of soil water content (SWC) is monitored by four drill and drop sensors (Sentek Pty Ltd, Stepney, Australia) installed 0.8 m far from the tree trunks, which provide the measurements up to 0.6 m depth (in steps of 0.1 m), with a time resolution of about 30 minutes. All soil moisture sensors are interfaced with electronic boards that utilize license-free communication to transmit data to a gateway, which in turn enables Internet connections through a 4G router, as shown in Figure 6.2. The connection is used to log in as a client to a TCP/IP server, which decodes packet transmission and stores the data in a MySQL database.

The Spatio-temporal variability of vegetation indices based on reflectance data in the visible (VIS), near infrared (NIR) and shortwave infrared region (SWIR) of the electromagnetic spectrum was investigated based on the images acquired by Sentinel-2 twin satellites (2A and 2B), characterized by a temporal resolution of about 5 days. In particular, the multi-spectral images (MSI) level 2A provide high-resolution data with spatial resolutions of 10 m, 20 m and 60 m, calibrated in reflectance at the bottom of the atmosphere (BoA), orthorectified and corrected for the atmospheric effects (ESA, https://scihub.copernicus.eu/) [375]. The available scenes were used to calculate the normalized difference vegetation index (NDVI) [376, 377] and the normalized difference water index (NDWI) [378] as:

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \tag{6.5}$$

$$NDWI = \frac{\rho_{nir} - \rho_{swir}}{\rho_{nir} + \rho_{swir}} \tag{6.6}$$

where ρ_{nir}, ρ_{red} and ρ_{swir} are the near-infrared, red and shortwave reflectance, respectively. The corresponding central wavelength, for Sentinel-2A and -2B satellites, are 664.6 nm and 664.9 nm for the red band (B4), 832.8 nm and 832.9 nm for NIR (B8), and 1613.7 nm and 1610.4 nm for SWIR (B11).

Table 6.3: Variables used in the analysis and number of records available in the investigated years.

			2018	**2019**	**2020**	**2021**
Instruments	**Variable**	**Units**	365	365	366	334
Drill and Drop	SWC	(cm^3cm^{-3})	355	344	355	320
WatchDog 2000	ET_o	(mm d^{-1})	363	365	366	333
EC Tower	ET_a	(mm d^{-1})	-	193	120	263
Sentinel-2	Images	-	44	75	74	58

6.4.2 Dataset description and pre-processing

The available database includes weather data recorded by the weather station from January 2018 to November 2021, as well as the micro-meteorological data acquired by the EC tower from March 2019 to November 2021, in both cases registered at sub-hourly time-steps. The former dataset was used to estimate ET_o, by equation 6.3. The latter dataset was processed, using a specific software developed by [379], excluding all the records acquired in days where the rainfall height was higher than

2.5 mm. The suitability of the records acquired by the EC tower was assessed based on the energy balance closure quantified by the closure ratio (CR) [380], representing the slope of the regression line of the turbulent heat fluxes against available energy, evaluated as:

$$CR = \frac{LE + H}{R_n - G} \tag{6.7}$$

where LE and H are the latent and sensible heat fluxes, R_n is the net radiation and G is the heat flux into the soil. Sub-hourly data was then aggregated at daily time steps. However, in March 2020 the COVID-19 pandemic lockdown caused the impossibility to visit the field and fix the EC tower, with consequent acquisition failure. Moreover, a single daily value of SWC, representative of the entire field, was calculated as the average of the available data acquired in the layer 0-0.60 m by the four drill and drop probes. Finally, for the four years considered, 251 images, under clear sky conditions, from Sentinel-2 satellite were downloaded and pre-processed using the R library named "sen2r"[381]. Using equations 6.5 and 6.6, for the whole field the vegetation indices were calculated by a specific script implemented in Matlab® R2019b and then exported in QGIS (release 3.4.3) environment to view the maps of NDVI and NDWI. To determine NDWI at the same spatial resolution as NDVI (10 m), the value of a single-pixel reflectance in the SWIR domain (20 m) was associated with the four reflectance values corresponding to the NIR spatial resolution. For both vegetation indices, a single representative value for the entire plot was obtained by averaging the four values calculated in the pixels containing the drill and drop probes. Table 6.3 summarizes the source and the size of the available dataset for the four investigated years (2018-2021).

Overall, the complete dataset has a length of 1430 days and contains 12 features (U_2, R_s, RH_{min}, RH_{max}, T_{min}, T_{max}, ET_o, SWC, NDVI, NDWI, DOY, ET_a), where DOY is the Day Of the Year and the actual evapotranspiration ET_a represents the variable of interest. This feature has 854 missing values, and for this reason, the proposed neural network models are exploited for ET_a predictions and gap-filling. The climate variables and consequently the ET_o feature have only three missing values in the dataset, while the SWC, and VIs features have 56 and 1179 missing values, respectively, over the four years of observation.

6.4.3 Machine Learning models

This section describes the design and implementation of the examined ML models. As a starting point, four models were initially tested: Linear Regression (LR), Support

Vector Machine (SVR), Random Forest (RF), and Multi-Layer Perceptron (MLP). From preliminary results, omitted here for the sake of brevity, only MPL and RF models were selected for their ability to model the nonlinear evapotranspiration phenomena. The models were implemented using the scikit-learn library [382], an open-source ML library, for the python programming language. Furthermore, the code developed in this work is released open source to the scientific community [383]. This will allow repeatability and ease future research efforts in the development of environmentally sustainable irrigation solutions. In the following, an introduction to Artificial Neural Networks (ANNs) and RF algorithms is presented, followed by the description and validation of the proposed MLP-based and RF predictors.

ANNs and Multi-Layer Perceptron

ANNs are a class of powerful ML tools that can be used to solve classification and regression problems. ANNs can be distinguished in two types of architectures, depending on the types of connection between neurons. In the feedback architectures, the presence of connections between neurons of the same layer or between neurons of the previous layer realizes a feedback connection. In the feedforward architectures, the connections between the neurons do not allow feedback between layers, and the signal is transmitted only to the neurons of the next layer.

A widely used feedforward ANN is the Multi-Layer Perceptron (MLP), constituted by one input layer, one or more hidden layers, and a layer of output neurons. The neurons are connected to the ones of the next layer with a certain weight and, in each neuron, the weighted sum of input variables is transformed into an output value through an activation function, defined as:

$$\mathrm{Y} = \psi \left(\sum w_i * x_i + b \right) \tag{6.8}$$

where w_i is the weight, x_i is the neuron input, b is the neuron bias, and ψ is the activation function.

The training of the network is usually done with a backpropagation algorithm, which is divided into two phases. In the first phase (forwarding), controlled inputs are applied to the network, pushing the activation of the input layer neurons. The signal propagates to the next layers, finally reaching the output neurons. The error between the desired output and the obtained result is then calculated for each neuron. In the second phase (backwarding), the error value is propagated backward and the weights of each link are accordingly modified with an optimization method, which

aims to minimize the output error. Finally, the network "model selection" is achieved by choosing a set of hyperparameters (i.e. number of hidden layers, number of neurons in each layer, learning rate, solver weight optimization, epoch scale, activation functions, etc.) which characterize the architecture of the MLP model [384]. The best model is selected comparing the performance scores of all possible combinations of hyperparameters.

Random Forest

A Random Forest is a particular classifier/regressor formed by a set of decision trees represented as independent and identically distributed random vectors. This technique is part of ensemble learning that has made significant improvements in learning accuracy for classification and regression tasks [385]. A random subset of the features is chosen at each candidate split during the learning process when using random forests, which employ a modified tree learning algorithm [386]. Each decision tree within the RF is constructed and trained from a random subset of the data in the training set. Therefore, the trees do not use the complete set, and at each node the best attribute is chosen from a randomly selected set of attributes (thus, not necessarily the absolute best attribute). For example, given the training dataset (X, Y), with each element $x_i \in X \in R$, $y_i \in Y \in R$, one can train M different trees on different subsets, chosen randomly with replacement, and then compute the ensemble average:

$$f(x) = \sum_{m=1}^{M} \frac{1}{M} f_m(x) \tag{6.9}$$

where f_m is the m'th tree. This technique is named bagging, which stands for bootstrap aggregating [387]. The basic idea behind bagging is to average models containing errors but approximately unbiased, so as to reduce the variance of an estimated forecast function. Decision trees are ideal candidates for bagging because they can capture complex interaction structures present in the data and, if grown with sufficient depth, have relatively low bias.

Since decision trees are known to be error-prone, they can benefit in important ways from their averaging. In other words, decision trees lower the model variance without raising bias and, thus, this bootstrapping method improves the model's performance. Indeed, while single tree predictions are very sensitive to noise in the training set, an average of several trees reduces this sensitiveness, provided that the trees are uncorrelated. Highly connected trees result from merely training several trees on a

single training set. Through the use of various training sets, bootstrap sampling can de-correlate trees. Thus, randomness is a factor that becomes part of the construction of the RFs, and is intended to increase their diversity and thus decrease their correlation. In the case of a regression, the final result returned by the RF is the average of the numerical result by the different decision trees.

Standardization and Performance metrics

In this work, the dataset was standardized to reduce the influence of outliers. All features were normalized and scaled so that they have a similar range. In particular, a standard score based on the following equation was used:

$$z = \frac{x - \mu}{\sigma} \tag{6.10}$$

where x is the real value of the sample, μ is the mean of the population, and σ is the standard deviation. Using the standard score all features have a zero mean and a unit variance.

The model performance was evaluated based on three statistical indicators: coefficient of determination (R^2), root mean square error ($RMSE$), and mean bias error (MBE), calculated according to the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \tag{6.11}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n}} \tag{6.12}$$

$$MBE = \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)}{n} \tag{6.13}$$

where y_i represents the measured evapotranspiration ET_a value of the i-th sample, $\widehat{y}_i$ is the corresponding predicted value, and $\overline{y}_i$ is the mean measured data for a total n observations.

Model hyperparameters selection

For both MLP and RF predictors, the grid search technique was applied to compute the optimum values of hyperparameters. Regarding the MLP, the network was implemented with an input layer, 3 hidden layers, and an output layer. In the model selection phase, the performance obtained using different hyperparameters was compared. Specifically, the following hyperparameters were tested:

1. Solvers: Limited-Broyden–Fletcher–Goldfarb–Shanno (L-BFGS), Adam, Stochastic Gradient Descent (SGD) with constant learning rate, SGD with adaptive learning rate;
2. Number of neurons in the hidden layers: from 1 to 100;
3. Regularization factor "alpha" (L2 penalty): 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4};
4. Activation function: identity, logistic, tanh, ReLU;
5. Learning Rate: constant, invscaling, adaptive.

The following configuration optimized the performance of the model and was accordingly adopted for analysis: Adam solver [388], 10 neurons per hidden layer, alpha $= 10^{-4}$, constant learning rate, and rectifier activation function, also called Rectified Linear Unit (ReLU) activation function, defined as:

$$\mathrm{ReLU}(k) = \begin{cases} k, \text{ if } k > 0; \\ 0, \text{ if } k \leq 0. \end{cases} \tag{6.14}$$

In the case of the RF, the hyperparameters include the number of decision trees in the forest, the maximum depth of the decision tree, the number of features considered by each tree when splitting a node, etc. This set of hyperparameters was tested using the grid configuration shown below:

1. Number of decision trees: from 100 to 1000 (in steps of 100);
2. Number of features to consider at every split (max_features): *auto*, *sqrt*, *log*2, *None*;
3. Maximum number of levels in decision tree: None, or from 10 to 100 (in steps of 10);
4. Minimum number of samples required to split a node (samples_split): 2, 5, 10;

5. Minimum number of samples required at each leaf node (samples_leaf): 1, 2, 4;
6. Method of selecting samples for training each tree (bootstrap): True or False;
7. Parameter for minimal cost-complexity pruning (cpp_alpha): from 0 to 0.06 (in steps of $4 \cdot 10^{-4}$).

The following setup was chosen for investigation since it achieved the best prediction accuracy: 1000 trees, max_features=auto, 50 levels, samples_split=2, samples_leaf=2, bootstrap=True, cpp_alpha=0.

Finally, feature analysis was performed to test different combinations of the input features among a set of 12 different possibilities. Moreover, ML algorithms were also exploited to compensate for missing values in the dataset, as detailed in the following Sec. 6.4.3.

Prediction algorithms and gap filling

This work exploits ML models also to perform gap-filling procedures of missing data. Gaps are present in the VIs because samples are not available with the same time resolution of the other variables. Few data samples were absent in other input features as well (SWC, ET_o, and climate data), as detailed in Sec. 6.4.2. Thus, for the missing input data described earlier, the KNN-Imputer was utilized to fill in missing values using the k-Nearest Neighbor approach [389, 390]. However, when the number of missing values is greater than the number of observed ones (the case of VIs), iterative imputation was used [391]. Note also that the dataset misses several ET_a measurements, with 542 complete records out of a total number of 1430 records. Being ET_a the output of the proposed data-driven models, the previous approaches for filling in the missing data were not used. Nevertheless, MLP and RF models were trained only on the sub-set of complete data and the best models can be selected as detailed in the next section. Cross-validation was also used to assess models reliability and avoid over-fitting. Obviously, the trained models can be used for future predictions, as well as for predicting the missing ET_a values of the dataset.

6.5 Feature Analysis and Results

For the four years of observation, figure 6.3 (a-d) presents the temporal dynamics of the daily weather variables acquired by the weather station. The annual pattern of daily solar radiation, R_s, is similar across the four years, with maximums values

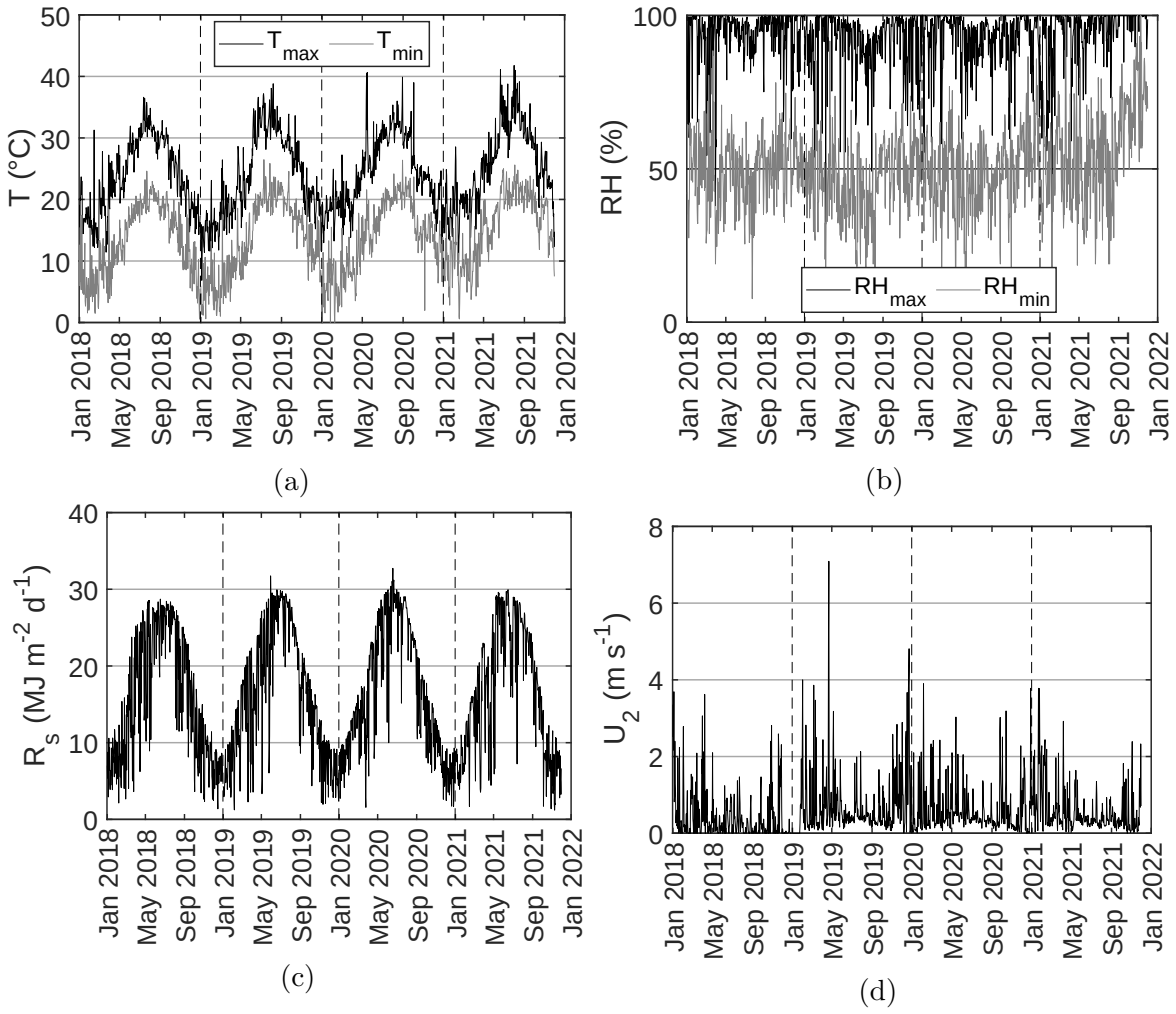


Figure 6.3: Temporal dynamic of daily climatic variables registered between 2018 and 2021: (a) maximum and minimum air temperatures, (b) maximum and minimum relative air humidity, (c) global solar radiation and (d) wind speed.

generally slightly lower than 30 $MJm^{-2}d^{-1}$ in summer and minimums lower than 10 $MJm^{-2}d^{-1}$ in winter. The annual dynamic of air temperature, T, follows that of R_s, with the maximum ranging between approximately 13 and 40°C and minimums ranging between 1 and 25°C, respectively in winter and summer. The average relative air humidity ranged between 48.0 and 92.4%, whereas wind speed typically remained lower than 2 ms^{-1}, even though some daily peaks of about 4 ms^{-1} or higher were also recorded.

The temporal dynamic of daily air temperature, net radiation and wind speed, as well as the frequency of wind direction recorded by the EC tower in 2019 and 2021 are shown in figures 6.4 (a-d). Compared to the values registered by the weather

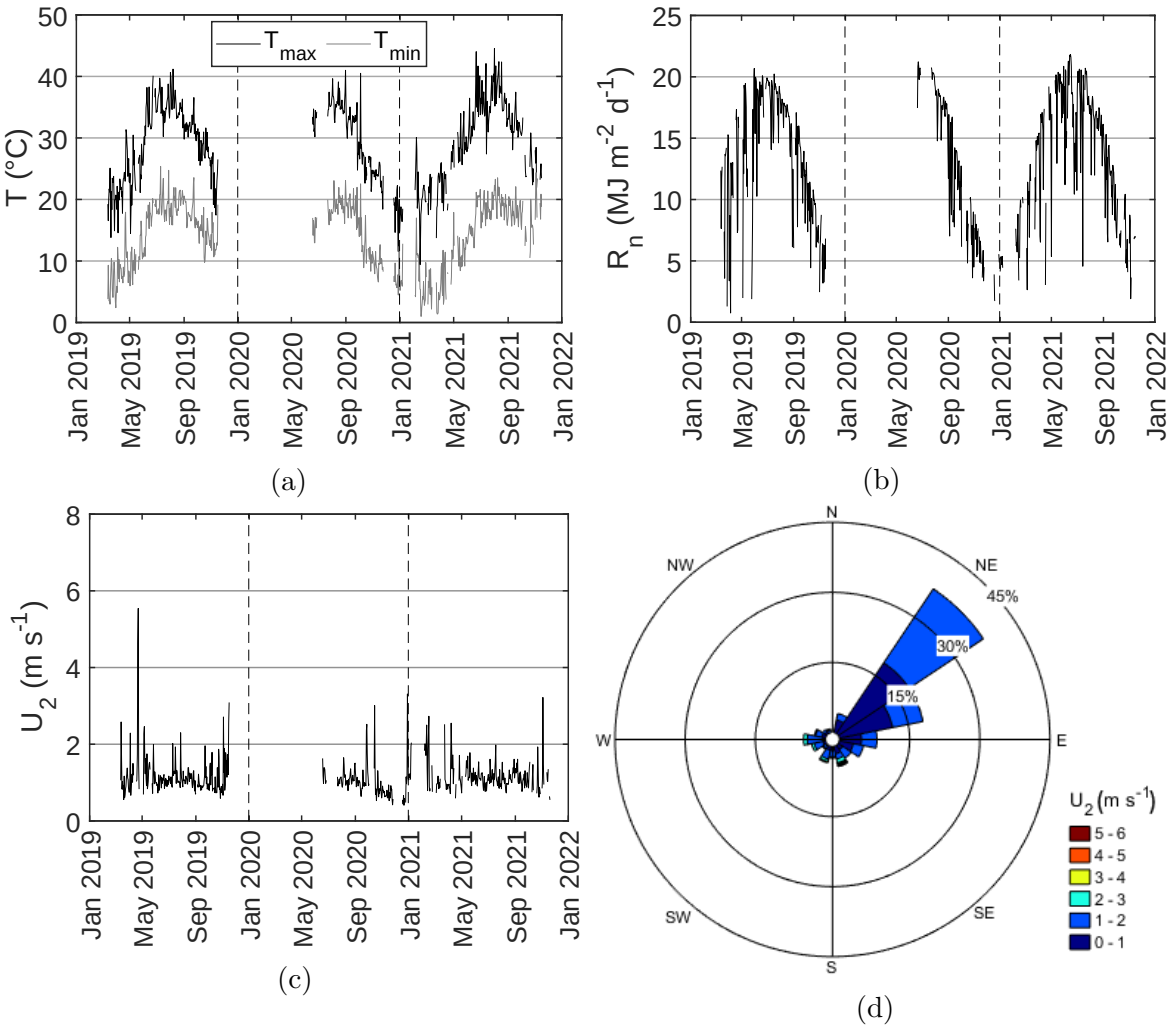


Figure 6.4: Temporal dynamic of (a) maximum and minimum air temperatures, (b) net radiation, (c) wind speed and (d) frequency of wind direction as registered by the EC tower in 2019, 2020, and 2021.

station, the EC tower slightly underestimated the minimum sonic air temperature and slightly overestimated the maximum sonic air temperature, mainly in summer when the values consistently exceeded $30°C$. The sonic wind speed was generally higher than the corresponding measurements taken by the cup anemometer installed in the weather station, due to the different installation heights and to the greater sensitivity of the CSAT3-D sonic anemometer. The dominant wind speed resulted in 1.7 ms^{-1} along the direction of 45°(NE). Figure 6.5 (a-c) shows the relationships between the turbulent heat fluxes, $H + LE$, and the available energy, $Rn - G$, measured by the EC tower in 2019, 2020 and 2021. The slope of the regression line, which represents the

closure ratio (CR), resulted equal to 0.98 for 2019, 0.88 for 2020 and 1.03 for 2021, indicating the suitability of the estimated energy balance components [392, 393].

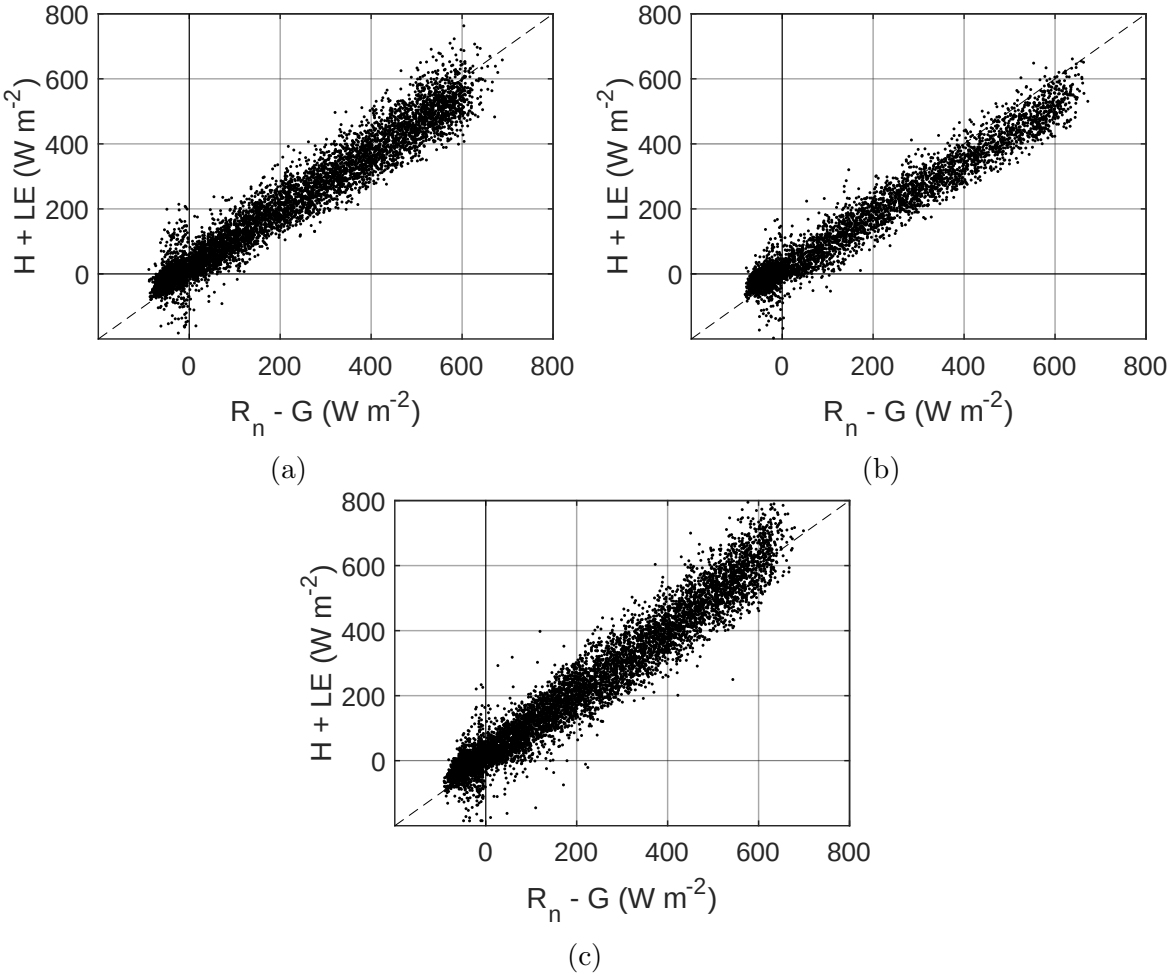


Figure 6.5: Relationships between turbulent heat fluxes, H+LE, and available energy, Rn-G, measured by the EC tower for 2019 (a), 2020 (b), and 2021 (c).

The temporal dynamics of daily crop reference evapotranspiration, ET_o, actual evapotranspiration, ET_a, and precipitation, P, are depicted in figure 6.6. The annual dynamic of ET_o follows that of air temperature and global solar radiation, with annual values ranging between about 1 mmd^{-1} in winter and 6 mmd^{-1} in summer; the peak of 8.8 mmd^{-1} registered on May 14, 2020, was due to the simultaneous occurrence of high air temperature, relatively low minimum relative air humidity and high wind speed. The values of ET_a in the considered period were generally lower than daily ET_o, although they resulted occasionally higher during or immediately after rainy days, as a consequence of the relatively greater contribution of evaporation from soil

and/or leaf surface. For this reason, the values of ET_a recorded on days characterized by rainfall heights higher than 2.5 mm were not included for further analysis.

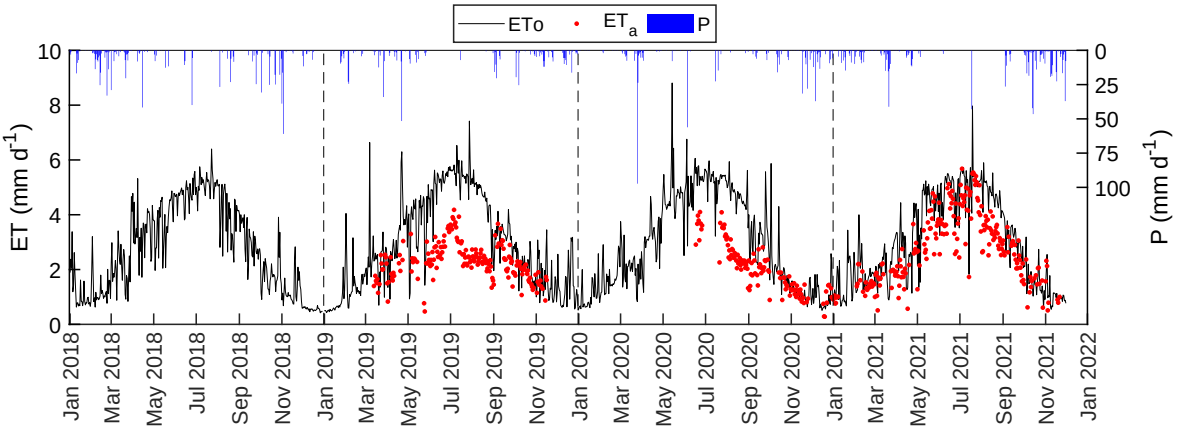


Figure 6.6: Temporal dynamic of crop reference evapotranspiration, ET_o, actual crop evapotranspiration, ET_a, and precipitation, P.

Figure 6.7 illustrates the temporal dynamics of daily soil water contents (SWC) obtained by averaging the values acquired by the four probes from the soil surface up to 0.6 m depth. The bottom of figure 6.7 also shows the daily SWC profile. As can be observed, for the four years considered, the average SWC ranged from 0.18 cm^3cm^{-3} to 0.34 cm^3cm^{-3}. The temporal dynamic of SWC is affected by the occurrence of rain events which determines the rapid increase of soil water contents. However, the rise of SWC observed in February, May and June 2020 was associated with wetting events of the upper soil surface operated by the micro-sprinklers irrigation system after long drought periods. The standard deviation associated with daily SWC in February, May and June 2020 as well as in April, May and June 2021 ranged between 0.01 cm^3cm^{-3} and 0.08 cm^3cm^{-3}, with the relatively higher values obtained during the irrigation seasons (from mid of May to the end of September), when the values of SWC at 30-50 cm depths resulted considerably higher than those of the upper soil layers, not wetted by the subsurface drip irrigation system.

The temporal dynamic of the average NDVI and NDWI for the four years, illustrated in figure 6.8, shows quite similar patterns. The values of NDVI ranged from 0.40 to 0.90, while the values of NDWI resulted variable between 0.01 and 0.60. Both VIs resulted maximum in winter ($NDVI_{mean} = 0.90$, $NDWI_{mean} = 0.36$) and minimum during summer ($NDVI_{mean} = 0.59$, $NDWI_{mean} = 0.10$), with quite low standard deviations as a consequence of the limited variability in the spectral response of the field surface.

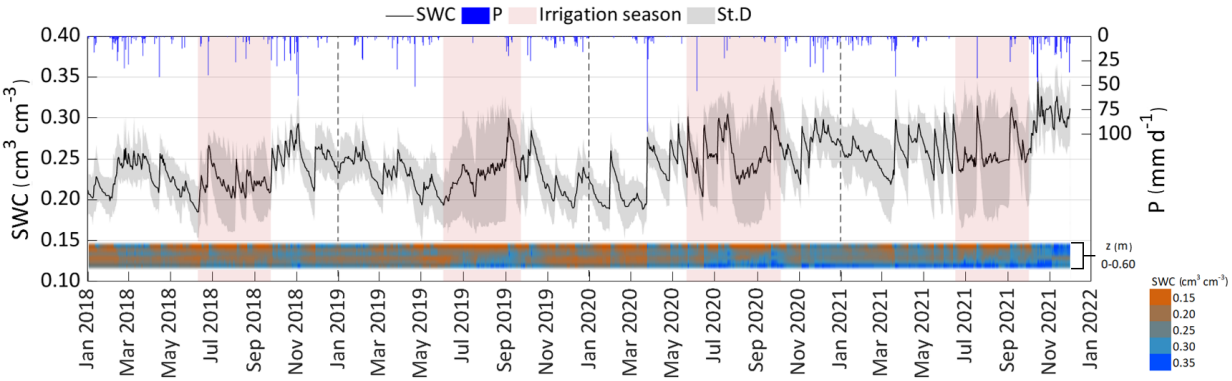


Figure 6.7: Temporal dynamics of daily average soil water content, SWC, valid for the entire field and precipitation, P. The pattern of the average SWC profile from 0 to 0.60 m depth is also shown at the bottom of the figure.The box indicates the irrigation season.

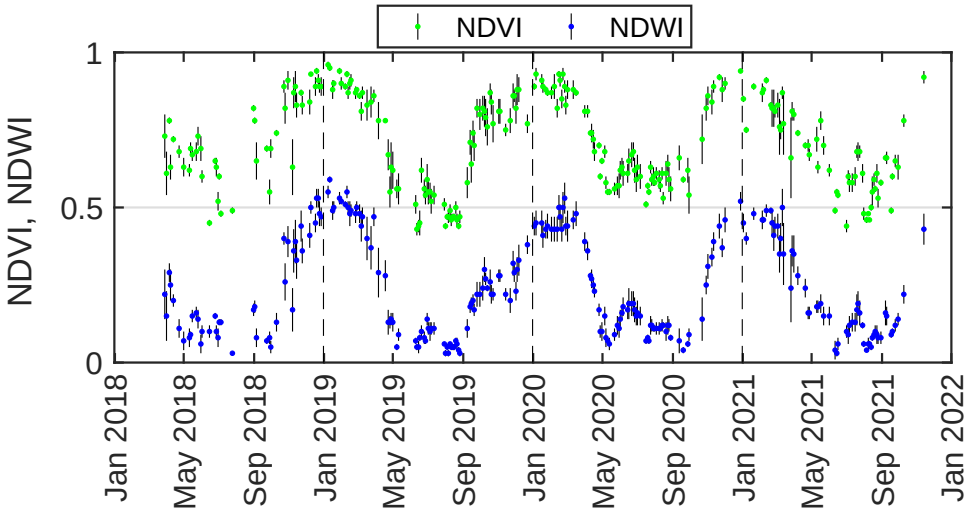


Figure 6.8: Temporal dynamics of average NDVI and NDWI with the corresponding standard deviation.

6.5.1 Feature analysis

Since ML models usually portray the relationship between input and output features, a preliminary feature analysis was performed. Indeed, this analysis plays a key role in reducing the dimensionality of the problem by favoring the application of ML models with a limited number of features, reducing the complexity, implementation and deployment costs in the field.

For example, the matrix in figure 6.9 shows at a glance the relationship between each two features: the scatter plot at row n and column m shows the n-th feature as a function of the m-th feature. This is helpful to spot correlations in the dataset, since

the more sparse are points in a plot, the more uncorrelated are the two corresponding features. In particular, it is clear that ET_o and R_S have a higher correlation with ET_a than other features. Finally, the plots in the diagonal represent the kernel density estimation of each feature, which is an estimate of their probability density function.

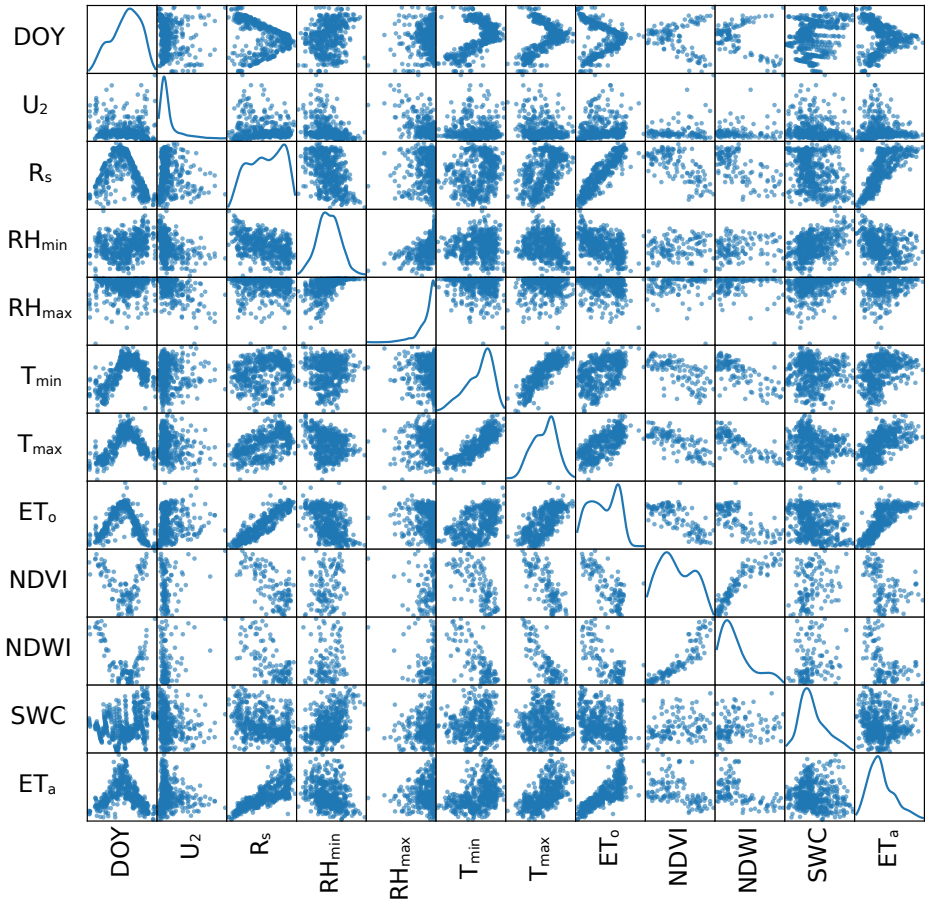


Figure 6.9: Scatter matrix used for overview and features analysis in ML models implementation.

Figure 6.10 shows the Pearson correlation coefficient, which measures the linear correlation of each couple of features. It ranges between 1 (dark red) and -1 (dark blue), were darker colors indicate higher absolute values, i.e. the two features are strongly correlated (the sign of the coefficient reflects the slope of the linear relation). In this case, the actual evapotranspiration ET_a has a positive Pearson correlation coefficient with the following features: ET_o, R_s, T_{max}, T_{min}, with a value of 0.8, 0.78, 0.59, 0.38, respectively. On the other hand, the features $NDVI$, $NDWI$, show a strong negative Pearson correlation coefficient with ET_a.

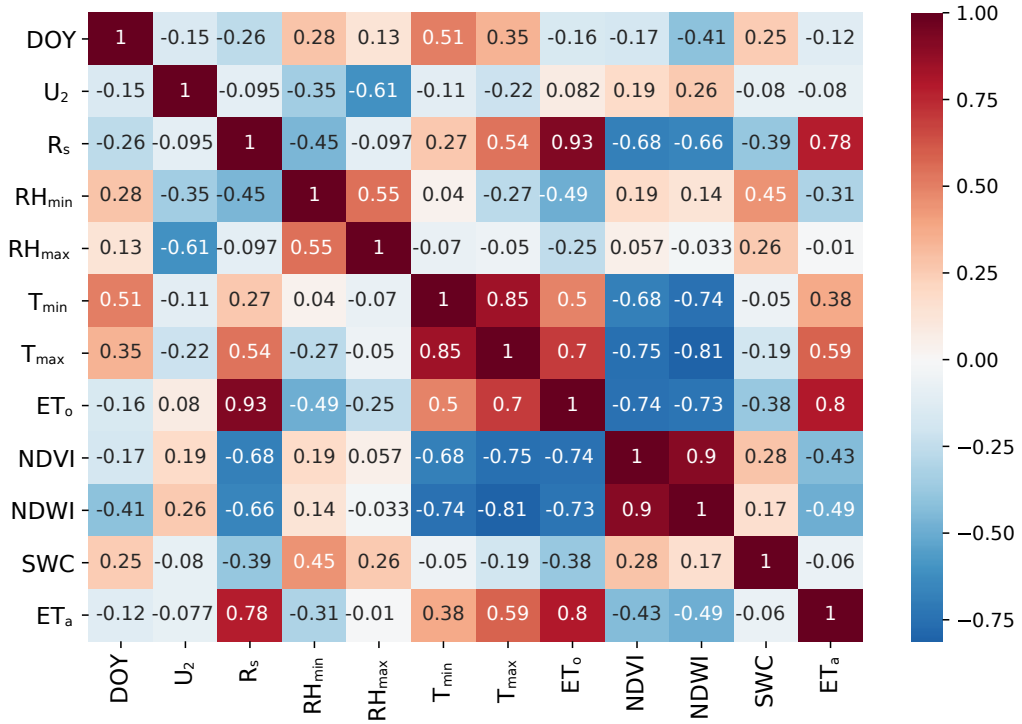


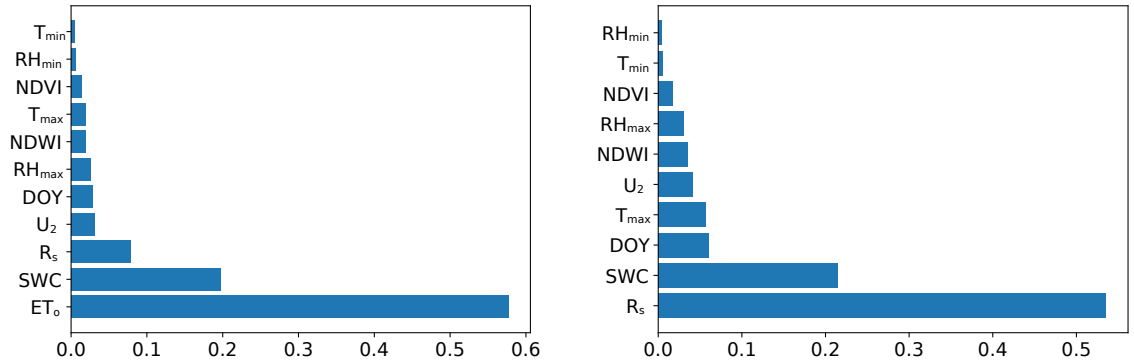
Figure 6.10: Pearson correlation coefficient between each pair of features.

Figure 6.11a shows the feature importance scores (FIS) of each variable in the dataset for the prediction of ET_a. The higher the FIS, the more important the feature. The importance of a feature is computed as the (normalized) total reduction of the criterion brought by that feature and is also known as the Gini importance [394]. In this work, FIS is computed using the Gradient Boosting Regressor [395]. Among all features in the dataset, the three most important ones are ET_o, SWC, and R_s, with a FIS of 0.577, 0.197, and 0.078, respectively.

Note that this result includes the reference evapotranspiration ET_o in the input feature analysis. However, since ET_o is a climate dependent parameter calculated from other measured climate variables, it is interesting to investigate how FIS changes when ET_o is excluded from the dataset. Indeed, from figure 6.11b, it is clear that excluding ET_o from the dataset sharply increases the FIS of R_s from 0.078 to 0.556 compared to the previous feature analysis. In this case, the most important features become R_s, SWC, and DOY.

6.5.2 Performance Evaluation and Results

Based on the feature analysis, it was possible to choose the features contained in the entire dataset or eliminate some less relevant features that could compromise the



(a) (b)

Figure 6.11: Feature importance score (FIS) of each feature of the dataset in the ET_a prediction (a), FIS when ET_o is excluded from the dataset (b).

generalization of the model, focusing the attention to the most relevant ones. More into details, table 6.4 shows the input feature combinations used to implement the developed models. In particular, for both MLP and RF, combination 1 contains all available features except ET_o. From combination 2 to 7 other features with less importance are gradually removed, while combination 8 uses only the variables measured by the weather station.

Finally, combinations 9 and 10 do not use the weather variables, but consider the ET_o feature instead, and combinations 11 and 12 are the same of models 9 and 10, except that ET_o is substituted by R_s, which is the most important feature among weather variables, both in terms of Pearson correlation coefficient and feature importance score.

Table 6.4: Input features combination used in input for the MLP and RF models.

Nr.	Input features selected
1	R_s, U_2, RH_{min}, RH_{max}, T_{min}, T_{max}, SWC, $NDVI$, $NDWI$, DOY
2	R_s, U_2,RH_{max}, T_{min}, T_{max}, SWC, $NDVI$, $NDWI$, DOY
3	R_s, U_2,RH_{max}, T_{max}, SWC, $NDVI$, $NDWI$, DOY
4	R_s, U_2,RH_{max}, T_{max}, SWC, $NDWI$, DOY
5	R_s, U_2, T_{max}, SWC, $NDWI$, DOY
6	R_s, U_2, T_{max}, SWC, DOY
7	R_s, T_{max}, SWC, DOY
8	R_s, U_2, RH_{min}, RH_{max}, T_{min}, T_{max}
9	ET_o, SWC, $NDVI$, $NDWI$, DOY
10	ET_o, $NDVI$, $NDWI$, DOY
11	R_s, SWC, $NDVI$, $NDWI$, DOY
12	R_s, $NDVI$, $NDWI$, DOY

The results obtained by the different models in terms of RMSE, R^2, and MBE are detailed in table 6.5. Specifically, the table shows that both the MLP and RF models perform well, although generally, RF performed better than MLP independently of the input features considered in the model. In particular, the Multi-Layer Perceptron method showed that models MLP1 and MLP5 were the best choices for ET_a estimation, with the lowest RMSE (0.44 mmd^{-1}) and highest R^2 (0.82) score compared to the other input feature combinations. On the other hand, the best scores for the Random Forest predictor are obtained with models RF2, RF3 and RF4, showing an RMSE of 0.39 mmd^{-1} and an R^2 of 0.84. Moreover, low MBE values indicate that ET_a predictions do not deviate much from the corresponding measured values (for all scenarios examined, MBE was between -0.025 and 0.041 mmd^{-1}). Note that, in some cases, a reduction in the number of feature produces equal or even better performances than models including all available features. For example, in the case of the MLP predictor, MLP1 and MLP5 show the same performance, but MLP1 used all the ten features available while MLP5 used only six features (the four less important features were removed, i.e. RH_{min}, RH_{max}, T_{min}, $NDVI$). Similarly, for the RF predictor, performance improved after removing features RH_{min}, T_{min}, and $NDVI$ (i.e. moving from RF1 to RF4). Finally, the RF predictor is also consistent with the reduction of input features: for example, in RF6 and RF11, i.e. using only five features, the RF predictor shows very good scores ($R^2 = 0.82$ and RMSE = 0.42 mmd^{-1}, close to the best ones), where RF6 used the first five features with the highest feature importance score (R_s, U_2, T_{max}, SWC, DOY), while RF11 used only weather variable R_s and features SWC, $NDVI$, $NDWI$, DOY. Indeed, from table 6.5 it is clear that the joint use of SWC, NDVI, NDWI, and DOY features with ET_o (MLP9 or RF9) improves the performance of ET_a predictions compared to models using only weather variables (MLP8 or RF8).

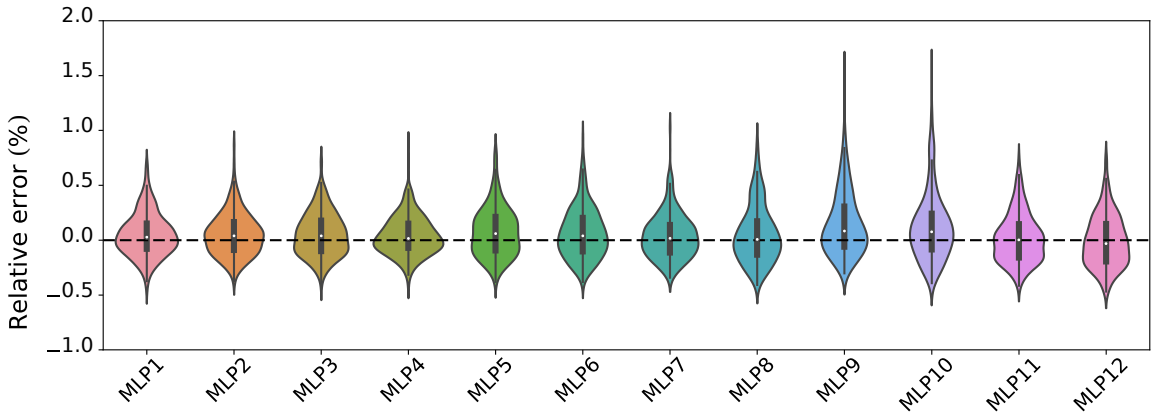
The accuracy of the different models was explored by analyzing the violin plots shown in figure 6.12. These plots show the relative error distributions of all MLP and RF models during irrigation seasons. In the middle of each violin plot, there is a small box plot, with the rectangle representing the ends of the first and third quartiles, and a central white dot for the median. The concentration of the relative error is close to zero, highlighting the good performance of the analyzed models. In particular, figure 6.12a shows that the errors of most MLP models (MLP1 to MLP11) have a median greater than zero (in the range $[0.002, 0.076]\%$), which means that these models slightly overestimated the ET_a values. Only the MLP12 model had a median error below zero (-0.031%).

Table 6.5: Results and model comparison

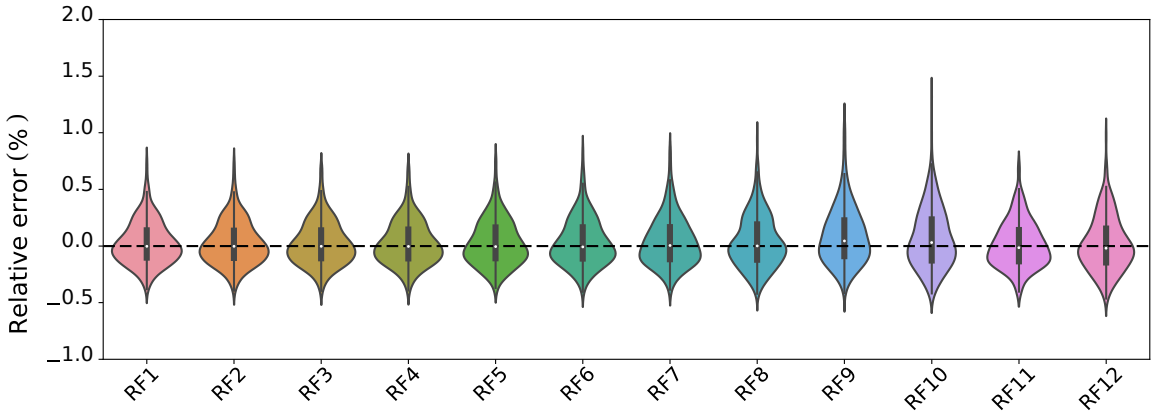
Predictor	**Model name**	**Nr. of input features**	**Input features**	R^2	$RMSE$ $[mmd^{-1}]$	MBE $[mmd^{-1}]$
Multi Layer Perceptron (MLP)	MLP1	10	R_s, U_2, RH_{min}, RH_{max}, T_{min}, T_{max}, SWC, $NDVI$, $NDWI$, DOY	0.82	0.44	0.017
	MLP2	9	R_s, U_2,RH_{max}, T_{min}, T_{max}, SWC, $NDVI$, $NDWI$, DOY	0.80	0.45	-0.025
	MLP3	8	R_s, U_2,RH_{max}, T_{max}, SWC, $NDVI$, $NDWI$, DOY	0.80	0.46	-0.024
	MLP4	7	R_s, U_2,RH_{max}, T_{max}, SWC, $NDWI$, DOY	0.81	0.44	0.008
	MLP5	6	R_s, U_2, T_{max}, SWC, $NDWI$, DOY	0.82	0.44	0.005
	MLP6	5	R_s, U_2, T_{max}, SWC, DOY	0.78	0.47	0.015
	MLP7	4	R_s, T_{max}, SWC, DOY	0.78	0.48	-0.016
	MLP8	6	R_s, U_2, RH_{min}, RH_{max}, T_{min}, T_{max}	0.69	0.57	0.001
	MLP9	5	ET_o, SWC, $NDVI$, $NDWI$, DOY	0.75	0.50	-0.002
	MLP10	4	ET_o, $NDVI$, $NDWI$, DOY	0.70	0.56	-0.009
	MLP11	5	R_s, SWC, $NDVI$, $NDWI$, DOY	0.74	0.53	0.041
	MLP12	4	R_s, $NDVI$, $NDWI$, DOY	0.72	0.54	0.001
Random Forest (RF)	RF1	10	R_s, U_2, RH_{min}, RH_{max}, T_{min}, T_{max}, SWC, $NDVI$, $NDWI$, DOY	0.83	0.40	0.013
	RF2	9	R_s, U_2,RH_{max}, T_{min}, T_{max}, SWC, $NDVI$, $NDWI$, DOY	0.84	0.39	0.011
	RF3	8	R_s, U_2,RH_{max}, T_{max}, SWC, $NDVI$, $NDWI$, DOY	0.84	0.39	0.009
	RF4	7	R_s, U_2,RH_{max}, T_{max}, SWC, $NDWI$, DOY	0.84	0.39	0.007
	RF5	6	R_s, U_2, T_{max}, SWC, $NDWI$, DOY	0.83	0.40	0.008
	RF6	5	R_s, U_2, T_{max}, SWC, DOY	0.82	0.41	0.008
	RF7	4	R_s, T_{max}, SWC, DOY	0.81	0.42	-0.005
	RF8	6	R_s, U_2, RH_{min}, RH_{max}, T_{min}, T_{max}	0.72	0.52	0.004
	RF9	5	ET_o, SWC, $NDVI$, $NDWI$, DOY	0.80	0.42	-0.003
	RF10	4	ET_o, $NDVI$, $NDWI$, DOY	0.72	0.52	-0.012
	RF11	5	R_s, SWC, $NDVI$, $NDWI$, DOY	0.82	0.42	-0.004
	RF12	4	R_s, $NDVI$, $NDWI$, DOY	0.76	0.48	0.005

Figure 6.12b, instead, shows that most of the RF models (9 out of 12) slightly underestimated the ET_a values, with a negative median error (the worst case was RF12, with -0.019%). On the other hand, RF7, RF10, and RF11 had a median above zero (0.004%, 0.044%, and 0.030%, respectively), therefore overestimating ET_a. The violin plots show that using features set 9 and 10 both MLP and RF had a longer upper tails, indicating that in some occasions these models provided larger overestimates. In addition, the relative error distribution was asymmetrical (positive skew) for most of the considered models. As expected, the RF models had a median error closer to zero, compared to MLP models.

Figures 6.14 and 6.15 show a comparison between observed and predicted ET_a values for the MLP and RF models, respectively, considering all 12 input feature combinations. The figures also show the linear robust regression interpolation forced through the origin (blue line) and the identity line (red dashed), from which it is clear that most of the models have a tendency to slightly underestimate the measured actual evapotranspiration values when $ET_a > 4.5\ mmd^{-1}$. On the other hand, a slight overestimation is present for $ET_a < 1\ mmd^{-1}$. This behavior based on the ET_a values to be predicted was also noted by [353]. Overall, the slope of the regression line for MLP models ranges between 0.951 and 1.031, while, for RF models, the slope varies between 0.941 and 1.005, which means that the latter have slightly less variance than the former. The slope closest to 1.0 for MLP were observed for the models MLP2 (fig.6.14b), MLP7 (fig.6.14g), and MLP9 (fig.6.14i) characterized by regression slope



(a)



(b)

Figure 6.12: Violin plots of the ET_a relative errors in predictions obtained using (a) MLP models and (b) RF models.

of 0.992, 1.009, and 1.003 respectively. On the other hand, the best slopes of RF models had values of 1.006, 0.992, and 1.005, for RF1 (fig.6.15a), RF4 (fig.6.15d), and RF9 (fig.6.15i), respectively.

Finally, a hydrological analysis (Fig. 6.13) was conducted in terms of cumulative rainfall, reference crop evapotranspiration, and the volumes of irrigation water supplied to the field to assess the impact of the RDI strategy on optimizing water use. Considering the annual cumulative precipitation, a total of 924 mm was recorded in 2018, and only 551 mm and 577 mm in 2019 and 2020. Furthermore, at the end of November 2021, the cumulative precipitation was equal to 776 mm. On the other hand, the annual crop reference evapotranspiration was 931 mm in 2018, 1069 mm in 2019 and 1076 mm in 2020, as a consequence of the relatively higher daily ETo values registered in 2019 and 2020. Finally, at the end of November 2021, the cu-

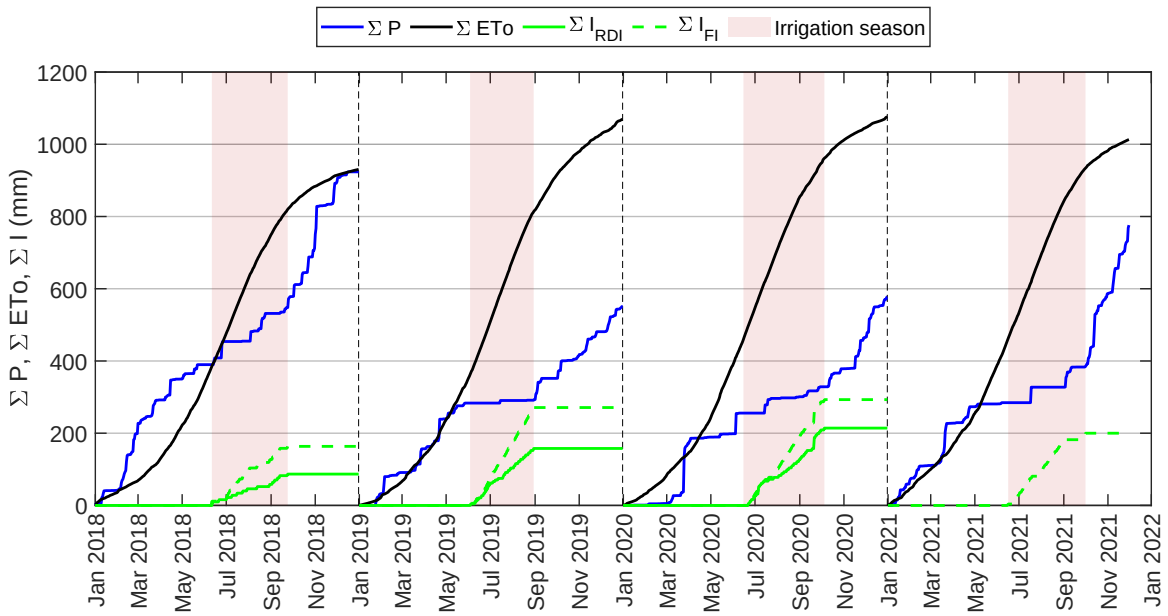


Figure 6.13: Cumulative precipitation, ΣP, crop reference evapotranspiration, ΣETo, and irrigations, ΣI, distributions during 2018–2021. The pink box indicates the irrigation season.

mulative crop reference evapotranspiration was equal to 1013 mm. The differences in the cumulative values of irrigation were due to the pluviometric deficit ($P - ETo$) values at the beginning of the irrigation season, equal to 0 mm in 2018, -84 mm in 2019, -243 mm in 2020, and -178 mm in 2021. The applied RDI strategy allowed an average water saving of about 38.5%, in 2018, 2019 and 2020, compared to the full irrigation (FI) management which was followed by the farmer in the rest of the field. The extraordinary high temperatures that were reached during the summer 2021 and the crop conditions before the water stress application period, induced the farmer not to apply the RDI strategy in this year. The maximum water saving was achieved in 2018 with 46.9%, as shown in Table 6.6. However, considering the entire observation period from 2018 to 2021, the percentage of water saved drops to 28.9%, which is still a significant amount considering the four-year observation period.

6.6 Discussion

The results presented in this work suggest that in general both the MLP and RF models were able to accurately describe the nonlinear relationships between weather and crop parameters, providing accurate predictions of ET_a under regulated deficit irrigation. These ML models could be used to optimize irrigation strategies according

Variable	2018	2019	2020	2021	Mean
$\Sigma I_{FI}\,(mm)$	163.6	271	293	200	231.9
$\Sigma I_{RDI}\,(mm)$	86.74	157.75	214.3	200	164.7
ΣI_{FI} - $\Sigma I_{RDI}\,(mm)$	76.86	113.25	78.7	0	67.20
Water saving(%)	46.9	41.8	26.9	0	28.9

Table 6.6: Comparison of cumulative irrigation, ΣI, between full and deficit irrigation strategies.

to the availability of water resources as well as to adopt RDI strategies, saving up to 38.5%, compared to full irrigation management. Therefore, this work offers a contribution to the creation of a sustainable agricultural environment, according to recent work in [325]. Moreover, since direct measurements of the actual values of ET_a (e.g. by means of EC towers) can be expensive and complex, in this work we use these values only for training a field-specific data model. Then, the models are used to provide a prediction of ET_a for the examined ecosystem based on low-cost sensor data only, with little impact on estimation accuracy. The results obtained by the models used in this chapter are summarized in the last row of table 6.2. Indeed, the results were generally consistent with the performance of previous studies in the literature when taking into account the average ranges of $RMSE$ and R^2 ([0.37, 0.97] mmd^{-1} and $[0.55, 0.86]$, respectively), while the performance of the proposed models varied between 0.39 mmd^{-1} and 0.57 mmd^{-1} for $RMSE$ and between 0.69 and 0.84 for R^2.

Furthermore, many of the papers listed in table 6.2 estimate ET_a in forest, cropland and grassland ecosystems, characterized by a homogeneous soil cover condition. However, in these scenarios, ET_a is rarely affected by the evaporative processes from the soil [367, 361]. Recently the possibility to estimate ET_a in tree crops, through ML models, was assessed by [348], using LST as input with a spatial resolution of 10 m, obtained from the downscaling of Landsat-8 LST data, by means of two different models (cubist and RF). In particular, the authors estimated ET_a through the Operational Simplified Surface Energy Balance Model (SSEBop) of olives and pomegranates tree crops, obtaining results slightly worse than the ones presented in this paper. However, it is important to highlight that the ET_a predictions presented in [348] are derived from the application of the SSEBop model and not from an ML model trained using ET_a in situ measurements. Furthermore, the olives and pomegranates crops are characterized by a high capacity to resist arid environments [351, 352], while citrus orchards are characterized by higher water requirements [331]. In this sense, it is possible to affirm that this work represents the first attempt to exploit ML models to

predict ET_a in a tree crop under RDI, where nonlinear processes, such as transpiration from the canopy and evaporation from the soil, must be taken into account.
It is important to note that the results obtained in this work are best when the SWC feature is included in the input variables. Indeed, from the results shown in table 6.5 and the feature importance analysis (Fig.6.11a and 6.11b), it is clear that SWC has a significant impact on the ET_a prediction. This evidence is supported by other studies in the literature showing that soil moisture becomes a significant factor for modeling ET_a when soil water supply is insufficient [396, 358, 357, 336]. Moreover, other works in ecological fields [397–399] confirmed that RF models are able to capture nonlinear dependencies among different variables, supporting the results presented in table 6.5 and fig. 6.12, in which the RF models appear to best predict the nonlinear phenomena of ET_a.

Finally, other hydrological features such as precipitation (P) and irrigation (I) were also considered as inputs to our ML models. However, these variables were eventually excluded for the following reasons. First, precipitation values are localized measurements and could limit the possibility of generalizing the obtained ML models, as discussed in [358, 344]. Second, in presence of subsurface drip irrigation systems, irrigation times are weakly correlated with the ET_a measurements (the obtained Pearson correlation and feature importance score were 0.150 and 0.003, respectively). This can be due to both negligible soil surface moisture during the irrigation season and failure to capture flows through the root zone, as explained for example in [340]. Note that, the two variables P and I are indirectly contained in the SWC feature used in our models. In particular, the SWC values were averaged considering four different positions in the field and six different measurement depths, from the surface down to 0.6 m depth. These measurements are thus more accurate for modeling ET_a under deficit irrigation.

6.7 Summary

This chapter explored the possibility of using ML models for accurately predicting actual evapotranspiration values in Mediterranean citrus orchards, where an RDI strategy is applied. Evapotranspiration data is very relevant to increase the ecosystems resilience, including the sustainability of the irrigation process. Different choices are possible for retrieving this data: deploying the instrumentation devised to perform direct measurements of ET_a (e.g. EC towers) or using this instrumentation only

temporarily for training a field-specific data model, which will be able to provide an indirect estimation of ET_a for the examined ecosystem, based on sensors data only.

Since measuring the actual values by means of EC towers can be expensive and complex, the objective of this research was to assess the second choice and study the impact of different features and ML approaches (namely, MLP and RF) on estimation accuracy. To identify the most relevant input data, an in-depth analysis of feature importance and correlation was performed on a dataset based on extensive in-field measurements collected in a suburb of Palermo, in Italy. The dataset includes meteorological data, vegetation indices, soil water contents, as well as direct measurements of ET_a (to be used as a ground-truth), acquired respectively by means of a weather station, satellite images, drill and drop sensors, and an EC tower. The results obtained are promising: data-driven models can achieve an accuracy higher than 82% for predicting ET_a values, especially in the case of RF models.

Moreover, reducing the number of input features from 10 to 5 does not harm the prediction accuracy significantly, thus enabling the possibility of opportunistically selecting in-field sensors. In general, the experiments demonstrated that ET_a is better predicted with agro-meteorological data of high importance (R_s, T_{max}, SWC) combined with vegetation indices, compared to models using weather variables only (i.e. R_s, U_2, RH_{min}, RH_{max}, T_{min}, T_{max}). Indeed, both soil water content and solar radiation were significant factors to predict ET_a accurately.

Results show that the prediction of ET_a in a tree crop under RDI strategy, allows an average water saving up to 38.5%, compared to full irrigation, reducing the impact of climate change and water scarcity in the global environmental equilibrium. Furthermore, the approaches proposed in this study, extended to different crops and climate conditions, can be a useful solution to improve the estimation of effective crop water requirements, with the aim to choose the best irrigation strategies for sustainable water management of productive ecosystems. In this context, the availability of continuous daily actual evapotranspiration time series is essential to implement, at local and regional scale, agro-hydrological models to support policy decisions and to better optimize agricultural practices.

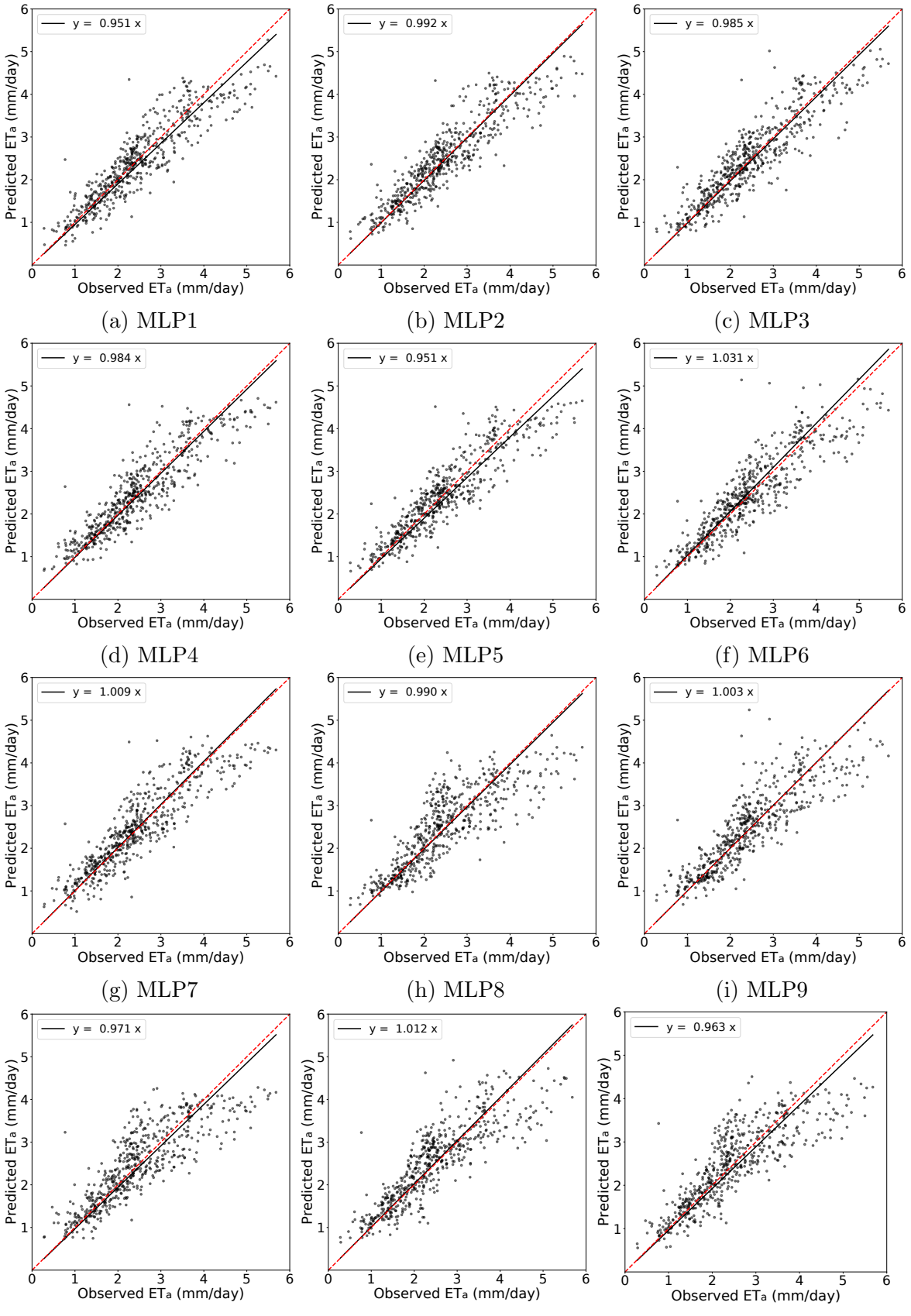


Figure 6.14: Comparison of predicted daily ET_a values by MLP methods with observed values for the testing subsets.

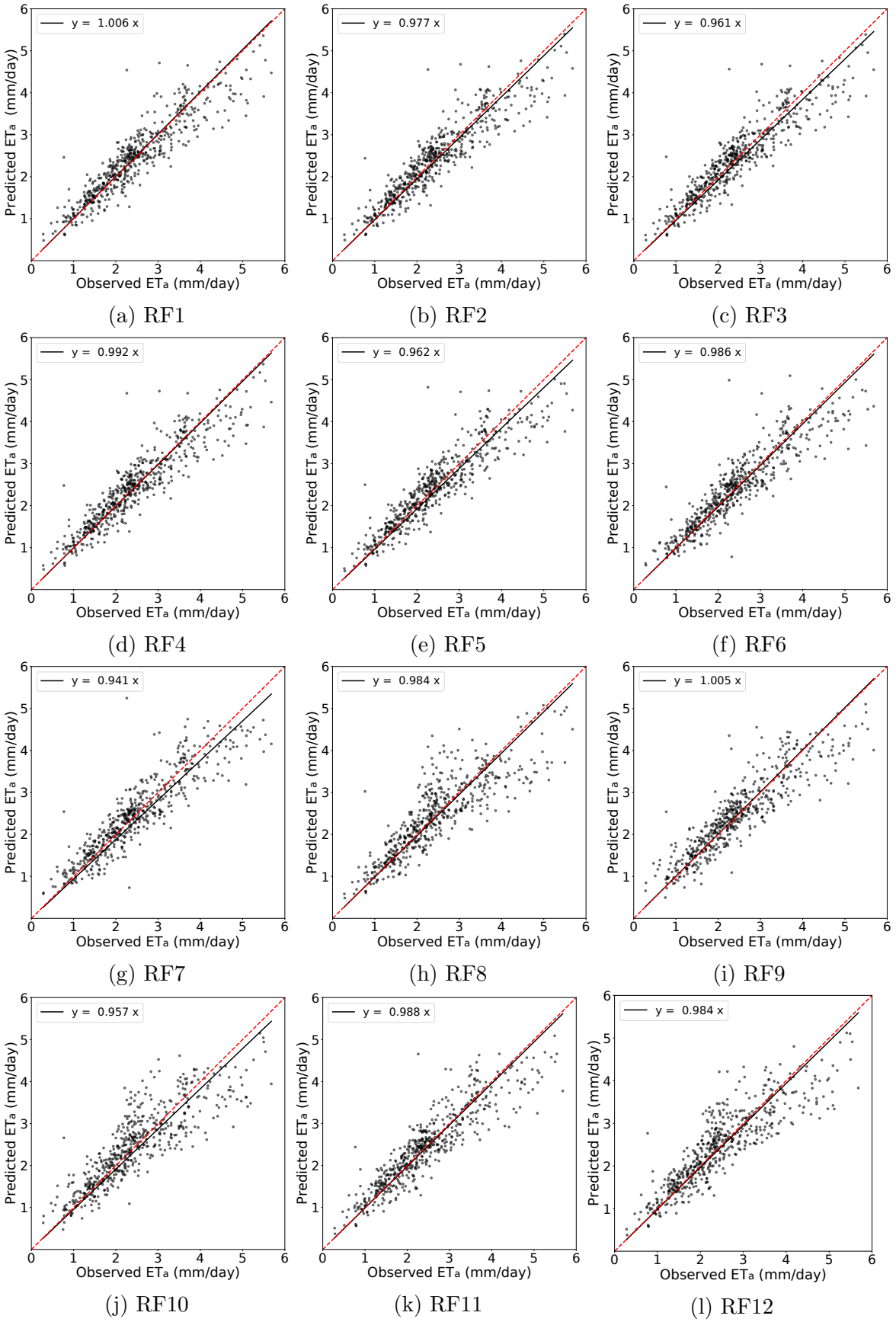


Figure 6.15: Comparison of predicted daily ET_a values by RF methods with observed values for the testing subsets.

Chapter 7

A Machine Learning Framework to Estimate Crop Coefficient Dynamics of Citrus Orchards

7.1 Overview

Obtaining crop coefficient (K_c) values quickly and accurately is essential for precision irrigation in field crop management, especially in a Mediterranean climate scenario, where citrus productivity strongly depends on effective irrigation strategies. Indeed, it is well-documented in the literature that using specific crop coefficient values during the growth phase of the crop could be a valuable solution for precise water irrigation, meeting crop water demands without compromising production. However, the lack of spatial and temporal resolution agro-metereological data (or the prohibitive costs to obtain such data on the fields), hampers the ability to determine the appropriate K_c values for optimal irrigation practices. This study is aimed to develop a framework that harnesses the potential of machine learning (ML) algorithms in conjunction with the Seasonal-Trend decomposition (STD) algorithm to effectively estimate the dynamics of K_c time series. First, three different ML algorithms are employed, namely Multi-Layer Perceptron (MLP), Random Forest (RF), and k-Nearest Neighbors (kNN), to predict actual evapotranspiration (ET_a) in Mediterranean citrus orchard. Then, having predicted ET_a, K_c values were derived using FAO-56 guidelines, and its accurate dynamics was estimated using Isolation Forest and STD algorithms. In order to evaluate and discuss the results of the proposed framework, four different baseline models were used. The RF model provided the best performance, obtaining

a root-mean-square error (RMSE) of 0.13, even when using only agro-meteorological variables. The proposed framework does not employ satellite images and requires less data and computational capacity compared to traditional methods, thus representing an excellent solution for implementing K_c estimation in smart agriculture.

7.2 Introduction

According to the Word Population Prospects 2022 [400], recently released by the United Nations Educational, Scientific and Cultural Organization (UNESCO), although in 2020 the growth rate of the global population fell under 1%, the projections suggest that the world's population could grow to around 8.5 billion in 2030, reaching a peak of 10.4 billion in 2080. This demographic growth could produce an increase in natural resource use, such as food and freshwater. Since freshwater resources are limited and about 70% are used for agricultural uses [324], accurate estimations of crop water requirements are essential to improve irrigation management and water use in agriculture. Specifically, in the Mediterranean semi-arid regions, characterized by very hot summer and warm winter, citrus is one of the most cultivated crops [349] and is characterized by very high water needs [350]. Furthermore, 90% of the production is freshly consumed while only 10% goes to transformation processing. In the climatic contest of the Mediterranean environment, the performance of citrus productivity depends on adequate irrigation strategies [401]. Therefore, in order to improve the productivity of the citrus groves it is essential to manage the available water resources [402].

In particular, [403] demonstrated that accurately monitoring specific values of the crop coefficient, K_c, during the crop's growth stage, could be a useful solution to provide precise water applications to satisfy crop water demand, without affecting the production. These values of K_c permit the identification of physical and physiological differences among the crops. The K_c knowledge is also necessary to implement the FAO56 agro-hydrological model [404]. According to the single crop coefficient approach, crop potential evapotranspiration, ET_c, is calculated as the product of crop reference evapotranspiration, ET_o, which represents the meteorological evaporative demand, and the crop coefficient, K_c. As suggested by [404], the K_c values change with the crop's growth stage, therefore it is correct to consider different values for each phenological stage. The first version of the FAO56 paper reported values of crop coefficients [404], which were recently revisited by [368]. Several authors have studied the possibility to optimize irrigation water use by means of accurate estimation

of K_c [373, 405–407]. Specifically, the cited methods are based on specific functional relationships between remote sensed vegetation indices and crop coefficients. However, complex analysis and satellite images are often necessary to implement these techniques, which are not always accessible. For this reason, these approaches are not easily implementable for a large number of operators. Therefore, it is very important to develop alternative procedures to estimate the crop coefficient.

In this chapter, the main objective is to develop a machine-learning framework to estimate the temporal dynamic of the crop coefficient in a typical citrus orchard of the Mediterranean region. The framework does not require the use of satellite images but is based only on agro-meteorological variables. This approach requires less data and computational capacity, making it easily implementable on a mini-computer such as a Raspberry Pi or a NUC (Next Unit of Computing). This, in turn, enables the implementation of edge computing in the wireless sensor network deployed in the agricultural field. Additionally, local data processing allows for a reduction in latency and data transmission on the backhaul, avoiding the need of complex communication infrastructure. Finally, the proposed framework facilitates the creation of intelligent self-managed sensor networks, capable of estimating the crop coefficient value even if the Internet connection is absent.

7.3 Literature review

According to the single crop coefficient approach suggested by [404], the product between the crop coefficient, K_c, and the stress coefficient, K_s, could be expressed as the ratio between actual evapotranspiration, ET_a, and crop reference evapotranspiration, ET_o:

$$(K_c K_s) = \frac{ET_a}{ET_o} \tag{7.1}$$

where ET_a represents the evapotraspirative flux referred to the real soil water condition, whereas ET_o represents the evapotranspiration of a reference hypothetical crop (i.e., alfalfa), actively growing, adequately watered and characterized by defined values of height, surface resistance, and albedo, which is, therefore, associated with the meteorological evaporative demand. The most widely used method to estimate ET_o is represented by the FAO56 Penman-Monteith (PM) equation, where the variable is function only of the climate variables [404]:

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma(\frac{900}{T_a} + 273)(u_2(e_s - e_a))}{\Delta + \gamma(1 + 0.34u_2)} \tag{7.2}$$

where $\Delta\,(kPaC^{-1})$ is the slope of saturation vapor pressure curve, $Rn\,(MJm^{-2}d^{-1})$ is the net radiation at the crop surface, $G\,(MJm^{-2}d^{-1})$ is the soil heat flux density, e_s–e_a (kPa) is the actual vapor pressure deficit, $\gamma\,(kPaC^{-1})$ is the psychrometric constant and $u_2\,(ms^{-1})$ is the wind speed measured at 2 m height.

The Eddy Covariance (EC) system is the method recognised by several authors as the most used and reliable to retrieve direct measurement of ET_a [408–411]. For a large scale use, this method could be time-consuming and very economically expensive. For these reasons, indirect methods have been developed to estimate crop coefficients [410].

In the past three decades, remote sensing (RS) techniques, based on multispectral and multi-platform remotely sensed images, have been used to develop functional relationships between vegetation indices (VIs) and K_c, to estimate the spatio-temporal variability. This approach is based on the assumption of a direct relationship between K_c and various vegetation spectral indices (NDVI, NDWI, SAVI, etc.) derived from reflectance in the visible (VI), near-infrared (NIR), and shortwave infrared (SWIR) regions of the electromagnetic spectrum. [412] and [413] showed the empirical evidence of the relationship between crop coefficient values and VIs. Following this approach, several authors have proposed functional relationships (K_c-VIs) for different herbaceous crops such as garlic, bell pepper, broccoli, and lettuce [414]. An overview of functional relationships K_c-VI developed to assess crop coefficient was proposed by [415], where the authors demonstrate that high spatial and temporal resolution allows monitoring of crop biophysical parameters. In a recent review, [370] show the potential of RS to derive the K_c-VI relationships. In particular, a detailed list of K_c-VI relationships developed for a variety of crops was presented with their advantages and limitations based on a SWOT analysis. The reliability of the predictive relationships for the estimation of evapotranspiration, was expressed in terms of root-mean-square errors (RMSE), with values ranging between 0.40 and 0.72 mmd^{-1}. However, only a few investigations have been carried out on arboreous crops.

In Mediterranean regions, citrus orchards are extensively diffused and have a relevant economy importance. To this aim, [406] studied the reliability of the ArcDualKc model [416] to estimate spatially distributed values of K_c in an orange orchard located in the island of Sicily (Italy). They found that the ArcDualKc model joint with VI retrieved from Sentinel-2 multi-spectral images (MSI) could be a useful tool to optimize water use in agriculture in the context of deficit irrigation. A similar study was carried out by [373] in a citrus orchard located in Sicily. The authors found that a non-linear relationship between crop coefficient and a combination of two vegetation

indices (NDVI and NDWI) allows estimating the values of crop coefficient in a citrus orchard managed under deficit irrigation strategies with good accuracy. These methods are based on the correlation between VIs and K_c, which is widely recognized as a reliable estimation technique. However, they require a large amount of data, needed to acquire RS products and pursue complex image processing. For this reason, these approaches are not easily implementable.

Recently, some works have adopted machine learning algorithms to estimate K_c for different crops. For example, [346] estimated the monthly maize coefficient values using an artificial neural network method. Specifically, the study showed the results of the most suitable combination of climate variables and the best neural network configuration for K_c prediction in the four main maize-producing geographical areas of Egypt. In addition, it was noted that the optimal neural network configuration was given by the integration of data such as minimum and maximum temperature and solar radiation in different layers of hidden neurons. Created models achieved high accuracy (RMSE = 0.25) when compared with the classic FAO CROPWAT model [404]. [417] presented an unmanned aerial vehicle-based technique to estimate K_c at the individual vine level in order to capture the spatial variability of water requirements in a commercial vineyard located in South Australia. Spectral and structural information, captured at various phenological stages of the vine through two seasons, were used to model crop coefficients using a framework of univariate, multivariate, and machine learning models (convolution neural network and random forest). In particular, the combination of spectral reflectance and vine structural features as input variables in the random forest (RF) model provided accurate K_c estimation (RMSE = 0.06) at the beginning and end of the season. The [418] study introduces a spatially unconstrained artificial intelligence model to estimate the K_c of tomato crops, collecting data from 600 tomato fields in various locations around the world. The dataset included weather data, Normalized Difference Vegetation Index (NDVI) from Sentinel-2 and Landsat-8 satellites, and soil properties such as season start date and country. The model adopted was Extreme Gradient Boosting, which produced an RMSE equal to 0.15 to estimate K_c for the entire season. The work conducted by [419], explored estimation models for daily maize K_c derived from two machine learning algorithms, random forest regression (RFR) and multiple linear regression (MLR), based on the use of leaf area index (LAI) and six types of multispectral vegetation indices (VIs) obtained from an unmanned aerial vehicle (UAV). In addition, it was noted that water stress strongly impacted the accuracy of prediction models. In particular, the performance decrease of MLR models was noted due to nonlinearity between selected features (LAI, VI)

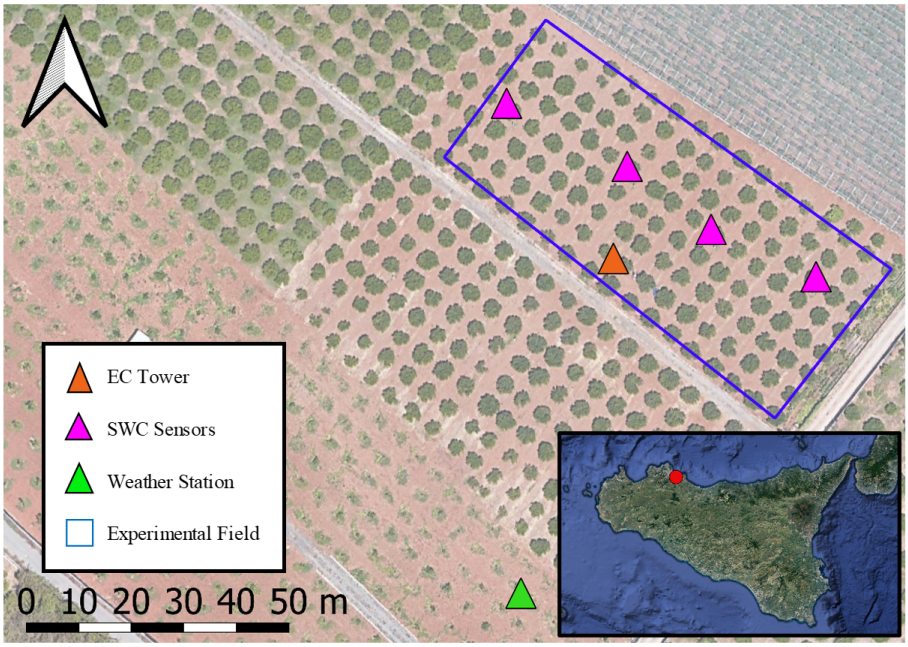


Figure 7.1: Map of the experimental field showing the location of the weather station, EC tower, and drill and drop SWC sensors.

and K_c under different water stress conditions. In contrast, better results associated with RFR algorithm were obtained because it adequately reflected the nonlinear relationship between K_c model and its input features. An extension of this work was reported in [420], where the role of RS variables based on multispectral, RGB, and thermal infrared information and their combinations was explored to develop prediction models using six ML algorithms (linear regression-LR, polynomial regression-PR, exponential regression-ER, random forest regression-RFR, support vector regression-SVR, and deep neural network-DNN). Again, the RFR model achieved the highest accuracy (RMSE = 0.10) and it was recommended for estimating the maize K_c.

However, all these techniques require the use of satellite or UAV images while the framework proposed in this chapter is based only on agro-meteorological variables. This approach requires less data and computational capacity, making it easily implementable even on a Raspberry Pi and deployed directly in the agricultural field.

7.4 Materials and methods

The experiment was carried out in a typical citrus orchard (Citrus reticulata Blanco, cv. Mandarino Tardivo di Ciaculli) of the Mediterranean region, located near Villabate, Sicily (Italy) ($38°4' 53.4'' 4'N$, $13°25' 8.2''E$) in the period from 2018 to 2022.

The experimental field (Fig. 7.1) is equipped with a WatchDog (WD) 2000 series (Spectrum Technologies, Inc., Aurora, IL, USA) standard weather station to collect the main agro-meteorological variables, four 0.60 m long "drill & drop" probes (Sentek Pty Ltd, Stepney, Australia) to monitor the temporal dynamic of the soil water content (SWC) and an Eddy Covariance (EC) flux tower to measure net radiation R_n (Wm^{-2}), H_2O and CO_2 concentrations and ground heat flux G (Wm^{-2}). The two concentrations are processed, using a specific software developed by [379], to estimate sensible and latent heat fluxes, H and LE, (Wm^{-2}). The turbulent heat fluxes, H+LE, as a function of the available energy, $R_n - G$, measured by the EC tower in the orchard at hourly time steps, for 2019, 2020 and 2021 were evaluated in chapter 6. The ET_a suitability was evaluated according to the Closure Ratio method (CR) proposed by [380]. The slope of the regression line passing through the origin represents the closure ratio, CR, whose values resulted satisfactory and equal to 0.98, 0.88, 1.03 and 0.88 for 2019, 2020, 2021 and 2022, respectively. For tree crops, [392] considered acceptable values of CR ranging between 0.80 and 0.90. However, values of the closure ratio, equal to 1.08 and 1.03, were obtained by [393] in two citrus orchards in south Morocco characterized by a semi-arid Mediterranean climate. More details regarding the sensors installed in the experimental field can be found in subsection 6.4.1 of chapter 6.

Table 7.1: Variables used and number of records available in the investigated years

			2018	**2019**	**2020**	**2021**	**2022**
	Variable	**Units**	365	365	366	365	365
Drill and Drop	SWC	(cm^3cm^{-3})	355	344	355	354	251
WatchDog 2000	ET_o	(mmd^{-1})	363	365	366	365	365
EC Tower	ET_a	(mmd^{-1})	-	193	120	265	291

7.4.1 Data description

The training database included climate data and SWC registered on ground, from January 1st 2018 to December 31st 2022, by WD weather station and drill & drop probes. Respectively, WD weather station provide half-hourly records of air temperature T_{min} and T_{max} (°C); relative air humidity RH_{min} and RH_{max} (%); global solar radiation R_s ($MJm^{-2}d^{-1}$) and wind speed measured at 2m height u_2 (ms^{-1}); while the drill & drop probes allowed monitoring the SWC at 0.1m intervals, up to a depth of 0.60 m, with a time-step of about 30 min. The average SWC between soil surface and 0.5m depth was assumed representative of the root zone.

The ET_a values are retrieved from the values of LE [404], and all data was aggregated at daily time-step. In order to account for the seasonal fluctuations in the weather data such as air temperature, T, and solar radiation, R_s, the day of year (DOY) were also considered as a training data [421]. Table 7.1 presents the variables that constitute the training database, the number of available records and the data source. In summary, the comprehensive dataset covers 1824 days and includes 10 features (u_2, R_s, RH_{min}, RH_{max}, T_{min}, T_{max}, ET_o, SWC, DOY, ET_a), from which actual evapotranspiration ET_a is the predicted variable used to determine the crop coefficient K_c.

7.4.2 Machine learning models

In this section we briefly describe the different ML algorithms employed, namely Multi-Layer Perceptron (MLP), Random Forest (RF), and k-Nearest Neighbors (kNN). These algortihms were implemented using the scikit-learn library [382], an open-source ML library for the Python programming language, and were used to predict the actual evapotranspiration (ET_a) and, in turn, the K_c values derived using the FAO-56 guidelines.

Multi-Layer Perceptron

Multi-Layer Perceptron belongs to the class of Artificial Neural Networks (ANNs), which learn by simulating the action of biological neurons. Each artificial neuron receives a signal as input, process it and send the new signal to connected neurons. In *feedforward* ANN architectures like the MLP, artificial neurons are arranged in layers and information is processed one-way from the starting raw input layer up to the final output layer [422]. Layers in between are collectively called "hidden layers". Signals from one layer of neurons propagate to the next layer with different weights and a constant bias term [423]. The result then passes through an *activation function* to add non-linearity, and the final output is then fed as input to the successive neuron layer. Iterating this forward feeding, the final output layer will be a set of complex functions of the inputs. To find the best weights and biases in the design of the MLP, a metric of similarity must be defined, called the *loss function*. The network is then trained on a known set of input features and outputs, optimized using the *backpropagation* algorithm to minimize the loss function [424].

Random Forest

A Random Forest is a collection of random Trees [385]. Specifically, Decision Trees are ML algorithms usually chosen for the simplicity of their approach and visual representation: starting from the complete set of features, each branch of the tree selects a feature and tries to classify the data point with respect to the target classes [425]. If a classification is possible, the branch ends in a leaf labelled by a specific class or a probability distribution among the classes. If the classification is not yet possible, the branch continues to a new branch with a new feature.

To avoid overfitting training data, more trees are used with the bootstrap aggregating technique, selecting a random sample and replacing the training set to fit the trees, so that no tree is trained with the complete set of data. More randomness is given selecting only a random subset of candidate features for each branch split. The final output is given then by majority voting among the output of each tree. The following setup was chosen for the implementation of the RF in Scikit-learn, since it achieved the best prediction accuracy: 1000 trees, max features = auto, 50 levels, samples split = 2, samples leaf = 2, bootstrap = True, cpp alpha = 0.

K-nearest neighbors

K-nearest neighbor (kNN) is a traditional non-parametric method that is part of the lazy learning field or instance-based learning [426]. This technique approximates the function exclusively within a local context, and all computations are postponed until the classification stage [427]. While this approach boasts simplicity, it proves highly effective, intuitive, and competitive, rendering it well-suited to addressing both classification and regression tasks involving dimensions of moderate scale. Despite its numerous advantages, the kNN method does possess certain limitations, especially in cases where the training dataset exhibits considerable size. Indeed, the kNN method can be time-intensive due to the necessity of calculating distances for every query instance when compared against the entirety of the training samples. Nevertheless, the employment of a KD-tree can expedite kNN searches when dealing with substantial datasets. In addition, this method shows a high degree of sensitivity to irrelevant and redundant features, and the optimal distance metric and feature set cannot be determined with certainty [428]. Consequently, it is imperative that these disadvantages be taken into account when determining classifications.

7.4.3 Statistical analysis metrics

The performance of each ML model was evaluated in two different steps of the framework. In the first phase, the models are used for the prediction of actual evapotranspiration. In this phase, the goodness-of-fit indicators are: (i) the Root Mean Square Error (RMSE), whose target value is zero when there are no differences between the predicted and measured values; (ii) the coefficient of determination (R^2), whose target value is one, indicating that the variance of the observed values is explained by the model [429].

In the second phase, the correspondence between crop coefficient values estimated by this framework and those measured or derived from other models was evaluated. In this case, in addition to the RMSE already mentioned, the following metrics were also adopted: (i) the Mean Bias Error (MBE), whose target value is zero. A positive MBE indicates that the simulated values are overestimated, while a negative value indicates that the values are underestimated [430]; (ii) the Mean Absolute Error (MAE), whose best value is zero; (iii) the percent bias (PBIAS), whose target value is zero. Positive values are associated with the model underestimation, while negative values indicate the model overestimation; (iv) finally, the Pearson′s Correlation Coefficient (PCC), whose target value is one. Values of PCC between 0.5 and 1.0 indicate a good model performance and strong correlation between predicted and observed values [431].

7.4.4 Framework description

Three ML methods were trained to perform gap-filling of the ET_a time series and then to estimate K_c time series. The flowchart shown in figure 7.2 represent the framework employed to estimate K_c dynamics during the observation period. The framework includes six main blocks explained in detail in the following subsections.

Pre-processing

In this phase, it is essential to pre-process the data in order to improve the performance of the models. First, missing data were imputed using an iterative imputation approach, followed by standardization. Iterative imputation involves utilizing a multivariate imputer to estimate each feature from all other features. It consists of modeling each feature with missing values as a function of other features in a round-robin fashion [432]. Moreover, the standard-scaler, or z-score, [433] is used to transform the values into a standard normal distribution with zero mean and unit variance. In this way, all input features were normalized and scaled so that they have a similar range.

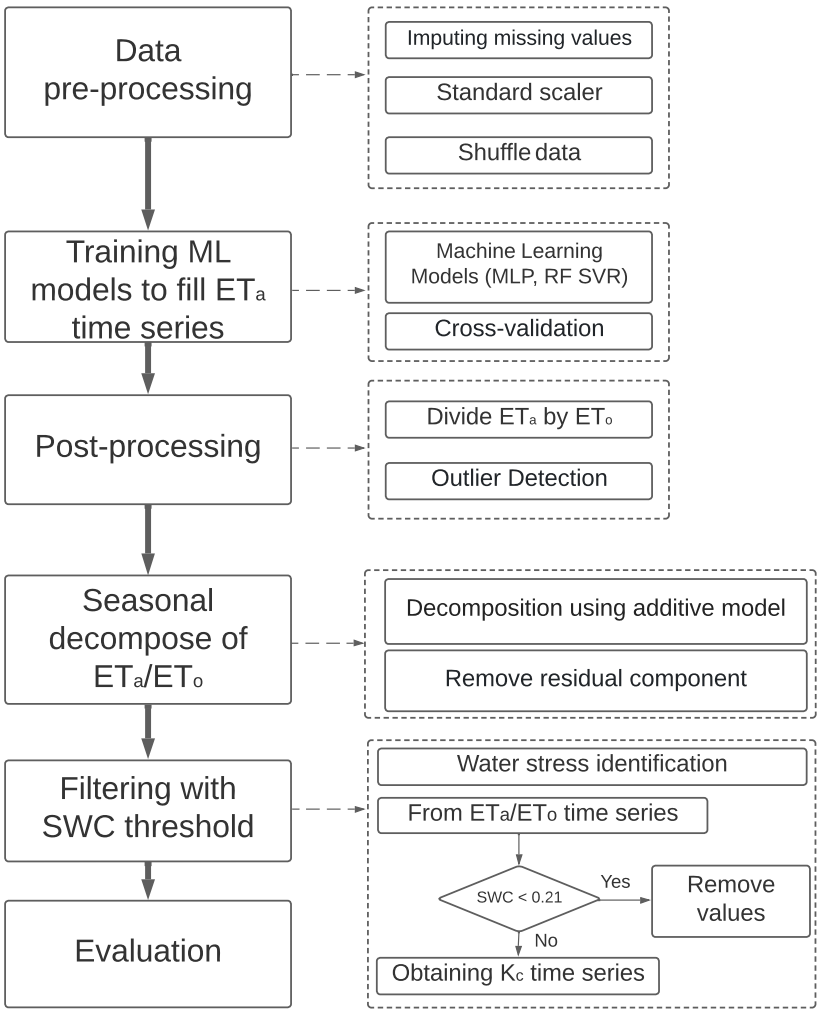


Figure 7.2: Flowchart of the proposed machine learning framework for estimating crop coefficient dynamics.

Lastly, time-series are shuffled to the purpose of reducing over-fitting and making sure that models remain general.

Gap filling in ET_a time series using machine learning

The number of daily ET_a measurements, acquired from January 2018 to December 2022, resulted in about 48% of the total days of the monitoring period; the missing data were caused by the malfunctioning of installed instruments. For instance, in March 2020 the COVID-19 pandemic lockdown caused the impossibility to visit the field to fix the EC tower, with the consequent failure of the acquisition. Moreover, all the records acquired in days when the rainfall height was higher than 2.5 mm were

excluded. Overall, actual evapotranspiration ET_a has 957 missing values, and for this reason, the machine learning models are employed for gap-filling. In particular, in this step of the framework, different machine learning models (MLP, RF, kNN) were evaluated and compared, evaluating their performance using cross-validation [434].

Post-processing

At this point, the developed models were used to determine the trend of the citrus crop coefficient for the five years of observation. In particular, eq. 7.1 was used to derive the (K_cK_s) time series for the examined crop. Subsequently, an automatic outlier detection model called Isolation Forest [435], a tree-based anomaly detection algorithm, was applied. The contamination hyperparameter was set at 0.01 to remove outliers in the (K_cK_s) time series that were not in agreement with the literature [405, 436, 373].

Seasonal decomposition

The ET_a/ET_o ratio is characterized by a strong seasonal component which deeply influences the (K_cK_s) time series. Under certain conditions (very hot summers, high water stress, presence of weeds, etc.) it can happen that the time series present high scattering. Thus, to obtain reliable (K_cK_s) time series, the ET_a/ET_o ratio was decomposed into three additive components: average trend, seasonality and residual. The decomposition is obtained using the statsmodels Python module [437] and it is based on a convolution filter and a moving average. In this way, the noise (or residual) can be easily removed. In addition, by summing the other two components (average trend and seasonality), a scattering reduction in the time series can be achieved.

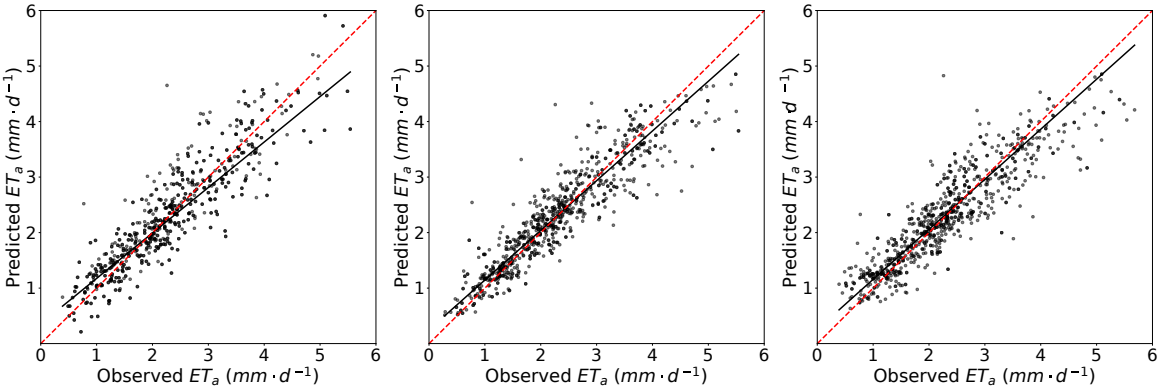
Filtering

At this stage, data filtering is applied to identify the presence of water stress. In fact, it is necessary to disregard from the dataset those days characterized by values of SWC lover than 0.21 cm^3cm^{-3}, which was identified as the threshold of SWC below which crop water stress occur and therefore $K_s < 1$ [438]. As a result, equation 6.4 is considered only in the case where $K_s = 1$ and consequently the ET_a/ET_o ratio correctly provides the dynamics of the crop coefficient.

Evaluation

The last step consists of evaluation of the K_c values estimated by the framework, according to the metrics presented in section 7.4.3. Specifically, the results obtained are compared against: (i) field measurements, (ii) two different theoretical models [404, 368], and (iii) an empirical model [373]. The two theoretical models are different because the first considers the presence of ground weeds, while the second their absence.

7.5 Results



(a) MLP (b) RF (c) kNN

Figure 7.3: Comparison of predicted daily ET_a values with observed values for the testing subsets.

Figure 7.3 presents a comparison between observed and predicted ET_a values for the MLP, RF, and kNN models, respectively. In the three graphs, the robust linear regression interpolation (black line) and the identity line (dashed red) are also displayed. The results obtained from the models demonstrate accurate ET_a predictions. The MLP model exhibits a regression line with the equation $y = 0.82x + 0.36$, a coefficient of determination (R^2) of 0.89, and a root-mean-square error (RMSE) of 0.35 mmd^{-1}. These values indicate a good adherence of the model to the observed data, with a slight margin of error. The RF Model behaves similarly to the MLP model, with a regression line described by the equation $y = 0.90x + 0.25$, an R^2 of 0.88, and an RMSE of 0.36 mmd^{-1}. In this case, as well, the model displays adequate predictive capacity, although with a slightly higher error than the MLP model. Finally, the kNN model features a regression line with the equation $y = 0.90x + 0.26$, an R^2 of 0.85,

and an RMSE of 0.41 mmd^{-1}. Despite the slightly lower coefficient of determination compared to the other two models, the kNN model still demonstrates a good ability to predict ET_a in the analyzed scenario. In summary, all three models provide reliable predictions of current evapotranspiration, with minimal differences in their performance. The regression lines highlight a strong correlation between the observed and predicted values for each model, confirming the validity of the methods used in the study.

Table 7.2: Estimated crop coefficients during mid-seasons, initial and late seasons of the observation period

Method	$K_{C,mid}$						$K_{C,in}$–$K_{C,end}$					
	2018	**2019**	**2020**	**2021**	**2022**	**Mean**	**2018**	**2019**	**2020**	**2021**	**2022**	**Mean**
Kc-EC	-	0.54±0.09	0.56±0.09	0.77±0.21	0.58±0.13	0.63±0.15	-	0.73±0.13	0.97±0.30	0.79±0.17	0.75±0.22	0.80±0.23
Kc-VIs	0.59±0.07	0.53±0.03	0.60±0.04	0.59±0.07	0.47±0.03	0.56±0.07	0.82±0.19	0.92±0.18	0.91±0.21	0.90±0.17	0.75±0.23	0.87±0.20
Kc-MLP	0.67±0.10	0.65±0.09	0.65±0.09	0.66±0.09	0.67±0.08	0.66±0.09	0.80±0.12	0.77±0.10	0.80±0.11	0.80±0.10	0.81±0.11	0.79±0.11
Kc-RF	0.66±0.08	0.63±0.07	0.64±0.07	0.64±0.07	0.64±0.07	0.64±0.07	0.76±0.15	0.81±0.14	0.81±0.14	0.81±0.14	0.82±0.15	0.80±0.15
Kc-kNN	0.67±0.11	0.64±0.06	0.65±0.08	0.66±0.07	0.69±0.07	0.66±0.08	0.87±0.14	0.80±0.12	0.89±0.11	0.87±0.12	0.86±0.13	0.86±0.13

Table 7.2 summarizes the K_c values obtained with the examined models during mid-season (from mid-May to the end of September, $K_{c,mid}$), and initial/late season (from the beginning of January to mid-May, $K_{c,ini}$, and until the end of December, $K_{c,end}$), of the five years of observation, where the $K_{c,mid}$ values are those of most significant interest for irrigation scheduling. The table also includes the standard deviation of K_c during the specified periods. The values associated with each of the three ML models differed with the values obtained from the eddy covariance tower (Kc-EC) or satellite images (Kc-VIs), especially with the lower values recorded in 2019 and 2022 due to the relatively higher evaporative atmospheric demand. However, when the average values of the entire period were considered, the differences in K_c, mid obtained with the ML models and Kc-EC or Kc-VIs tended to decrease. For instance, the Kc-EC method highlights an average of 0.63 ± 0.15, while the Kc-MLP, Kc-RF, and Kc-kNN methods show average values of 0.66 ± 0.09, 0.64 ± 0.07, and 0.66 ± 0.08, respectively.

In 2021, the values of ET_a resulted generally higher than in the other years. This effect can be associated with the relatively higher air temperatures registered in 2021 compared with the other years, as well as with the transpiration of the growing weeds due to field mismanagement during the irrigation season, as can be observed in Fig.7.4. Consequently, during the 2021 irrigation season (from May to October), the $K_{c,mid}$ values obtained through the Kc-EC method, i.e., using measurements taken with the eddy-covariance (EC) tower, exhibited a notable dispersion within the range of 0.52 to 1.00. Specifically, a concentration significantly higher than the $K_{c,mid}$ values recom-

Figure 7.4: Field's photos for three different irrigation seasons.

mended by the [404] (0.55) and [368] (0.50) was observed. In this context, the combined use of ML and STL (Seasonal and Trend decomposition using Loess) models for the seasonal decomposition of time series allows for a reduction in weed transpiration contribution. Thereby, by removing the residual component from the time series, a 13% reduction of $K_{c,mid}$ can be observed in 2021 irrigation season (table 7.2). In addiction, it is interesting to note that the three machine learning models, Kc-MLP, Kc-RF, and Kc-kNN, exhibit a lower standard deviation compared to the EC and VI-based estimations. This observation can be attributed to the effective removal of the residual obtained through seasonal decomposition applied by the framework to these models.

Table 7.3: Performance comparison of the proposed framework with four baseline models.

Model	Reference	RMSE	MBE	MAE	PBIAS	PCC	R^2
Kc-MLP	Kc-EC	0.15	0.01	0.12	0.25	0.54	0.28
	Kc-VIs	0.16	0.03	0.13	5.34	0.52	0.22
	Allen	0.14	0.02	0.10	3.07	0.55	0.24
	Rallo	0.17	0.14	0.14	25.69	0.55	-6.04
Kc-RF	Kc-EC	0.15	-0.01	0.11	5.06	0.58	0.37
	Kc-VIs	0.16	0.03	0.13	4.51	0.57	0.30
	Allen	0.13	0.01	0.10	2.25	0.60	0.33
	Rallo	0.17	0.14	0.14	24.72	0.60	-5.44
Kc-kNN	Kc-EC	0.17	0.03	0.13	5.06	0.50	0.17
	Kc-VIs	0.16	0.07	0.13	10.41	0.60	0.22
	Allen	0.14	0.05	0.10	8.10	0.67	0.28
	Rallo	0.20	0.18	0.18	31.86	0.67	-8.92

Table 7.3 shows a performance comparison of the proposed framework with empirical measures or reference models. The results are provided for three different models: Kc-MLP, Kc-RF, and Kc-kNN. For each model the outcomes are reported concerning four distinct references: Kc-EC, Kc-VIs, Allen [404], and Rallo [368]. As an example, the Kc-MLP model compared against Kc-EC as reference, the RMSE is 0.15, the MBE is 0.01, the MAE is 0.12, the PBIAS is 0.25, the PCC is 0.54, and the R^2 is 0.28. In general, models such as Kc-MLP and Kc-RF seem to have relatively low RMSE values, indicating good prediction accuracy with respect to the observed data. On the other hand, Kc-kNN shows slightly higher, but still acceptable RMSEs. Kc-RF has MBE values closer to zero, indicating an ability to predict without significant bias. In contrast, Kc-kNN possesses slightly higher MBE values, indicating a tendency towards positive bias in the predictions. The MAE values for Kc-MLP and Kc-RF are relatively low, which in the context of saving water can be an important feature of predictions. All the models exhibit a positive PCC correlation between their forecasts and the reference data. The majority of the R^2 values are positive, signifying a good fit to the data for the models. When considering the Kc-VIs as benchmark, the results indicate that all the developed models perform similarly in terms of RMSE, MBE, and MAE. However, the Kc-kNN model has a considerably higher PBIAS value compared to the other two models, suggesting a tendency to overestimate the reference data. In summary, it can be observed that the Kc-RF model outperforms the other two

Figure 7.5: Field's photos for three different rainy seasons.

models in terms of RMSE, MBE, MAE, and R^2. Nevertheless, the developed models encounter challenges in accurately predicting the Rallo reference data, as evidenced

by the negative values of R^2 for this reference. The results associated with Rallo show these values due to the field usually being full of ground weeds during the rainy season as reported theoretically in [373] and shown in Figure 7.5.

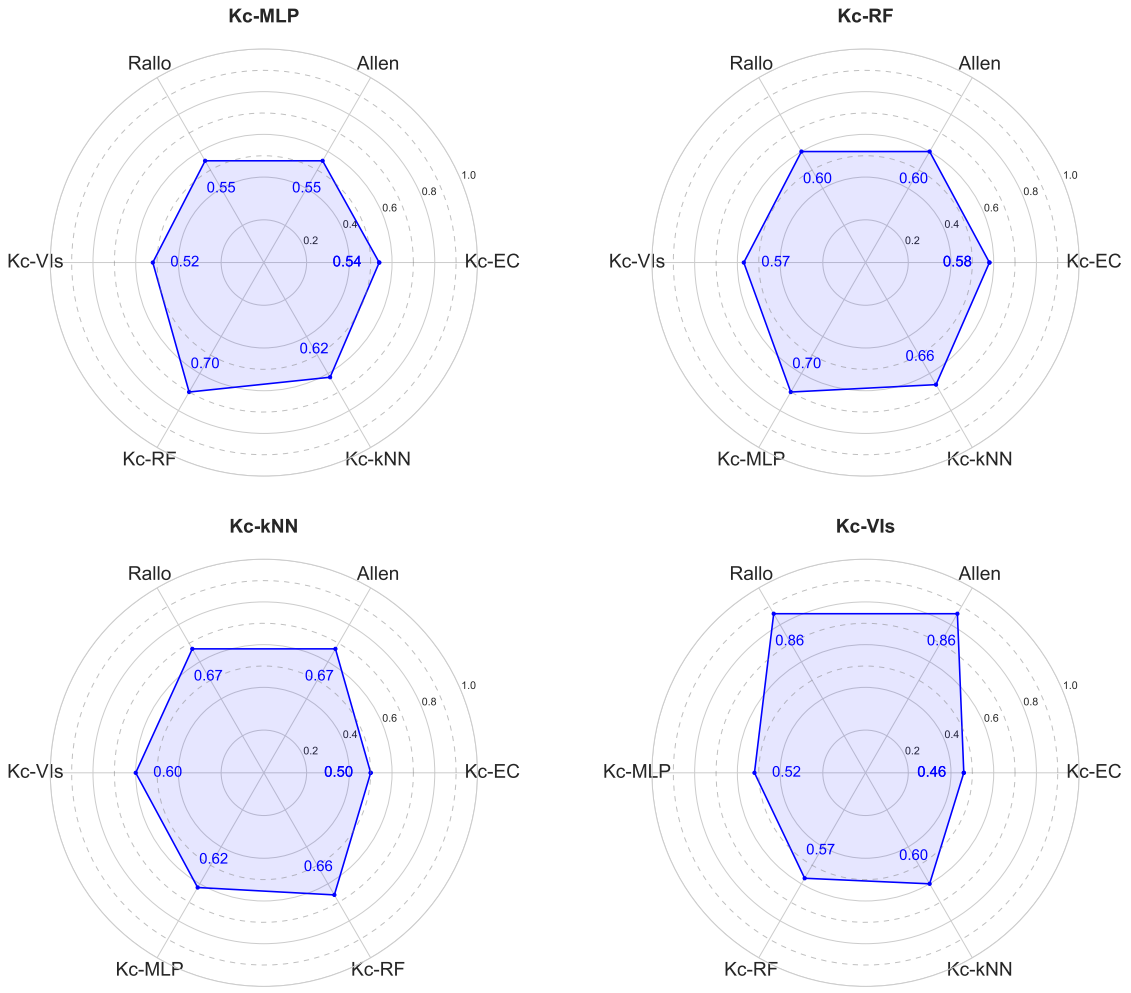
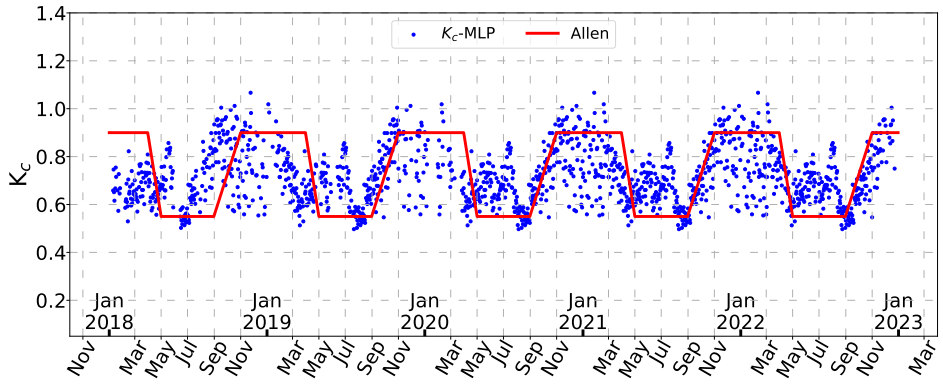


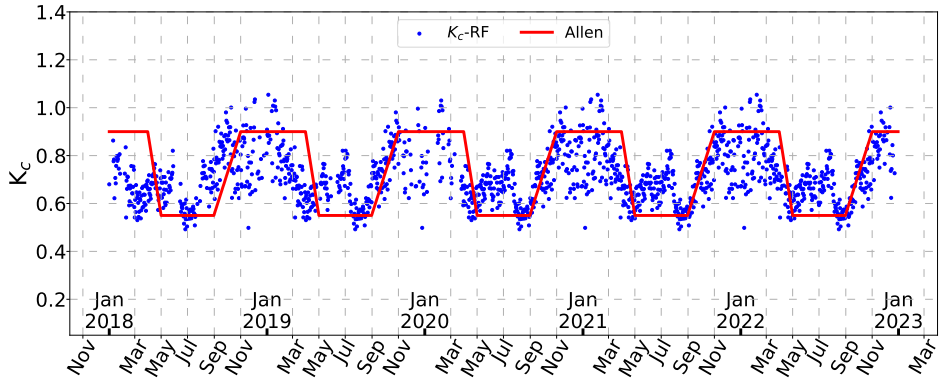
Figure 7.6: Radial graphs of Pearson correlation coefficient between each pair of models.

The Pearson coefficient in Figure 7.6 shows some interesting correlations between the reference models (Allen, Rallo, Kc-EC, Kc-VIs) and the developed ML models (Kc-MLP, Kc-RF, Kc-kNN). In particular, we note that the correlation coefficients between the reference models tend to be higher than the correlations with the developed models. For example, the Pearson correlation coefficients of the Allen and Rallo theoretical models have a value of 0.86 with the empirical model Kc-VIs presented in [373], indicating a strong positive correlation between them. On the other hand, the correlations between the reference models and the developed ML models are generally

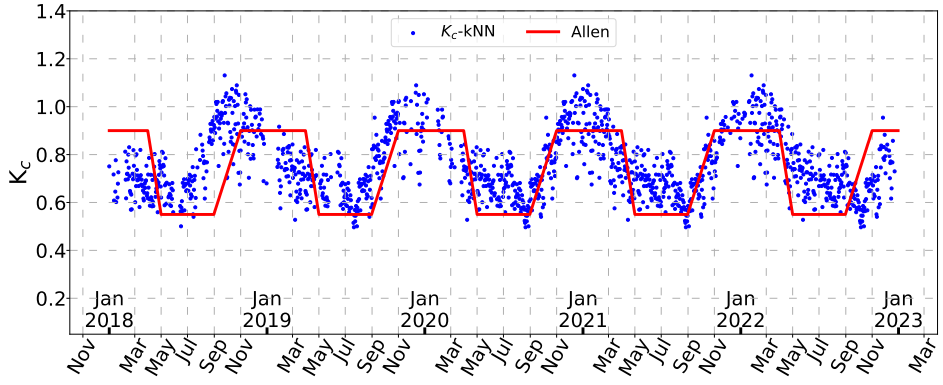
lower. For example, the correlation coefficients between Kc-MLP and the reference models vary between 0.52 and 0.55. Similarly, the correlations for the proposed Kc-RF and Kc-kNN models vary between 0.50 - 0.67. Although not as high as for the Kc-VIs, ML models show positive and significant correlations with reference models. Furthermore, since reference models account for the general approximate growth cycle of the



(a) Kc-MLP



(b) Kc-RF



(c) Kc-kNN

Figure 7.7: Values of K_c estimated on several days of the study period with: (a) the Kc-MLP model; (b) the Kc-RF model and (c) the Kc-kNN model.

crops, positive correlations with the ML predictions mean that these are modeling correctly the process.

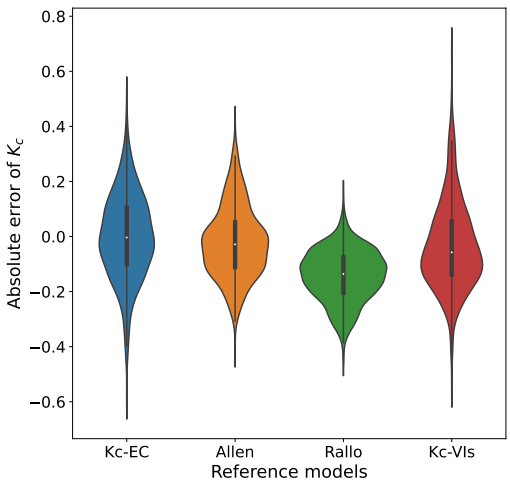
Figure 7.7 shows the time evolution of the crop coefficient K_c estimated with three different ML models (blue dots) at different periods of the year from January 2018 to December 2022, as well as the trend of the theoretical Allen model (red curve) used as a comparison. It is possible to identify the different stages of crop development despite the variability among ML models. In fact, the seasonal dynamics of K_c in the ML models align with the characteristic pattern observed in citrus orchards, where mid-season values are lower than those in the early and late seasons due to stomatal closure during periods of high atmospheric evaporative demand [404].

Figure 7.8 shows the violin plots distribution of the absolute error for three ML models calculated against the four reference models. In the middle of each violin plot, there is a small box plot, with the rectangle representing the ends of the first and third quartiles, and a central white dot for the median. In the case of the Kc-MLP model, a small underestimation appeared both using Kc-EC or Allen as reference, with median errors of -0.004 and -0.028, respectively. The best performance was obtained using the Kc-RF model, with median absolute errors notably low: for instance, with Kc-EC as reference, the median was as low as -0.001, while with Allen an appreciable precision was obtained with a median of -0.022.

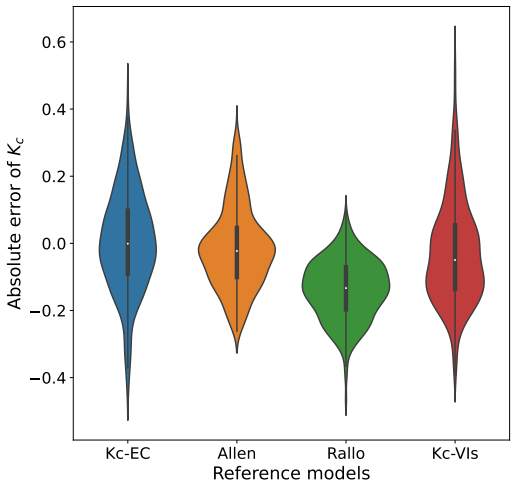
7.6 Discussion

While the proposed framework was developed and tested in a citrus orchard, it can be easily extended to different types of tree crops in order to develop production-efficient irrigation strategies. Furthermore, in order to increase sustainability and lower the costs for monitoring in agriculture, it is possible to think of solutions that use the more expensive and complex instruments (e.g. EC towers) only temporarily for training a field-specific ML model, which will be able to provide an indirect estimate of K_c, based on data from the less expensive sensors (e.g. weather station). For instance, assuming that the models have already undergone training, Table 7.4 shows the differences in terms of data requirements between the different models used to estimate the K_c values during an observation month.

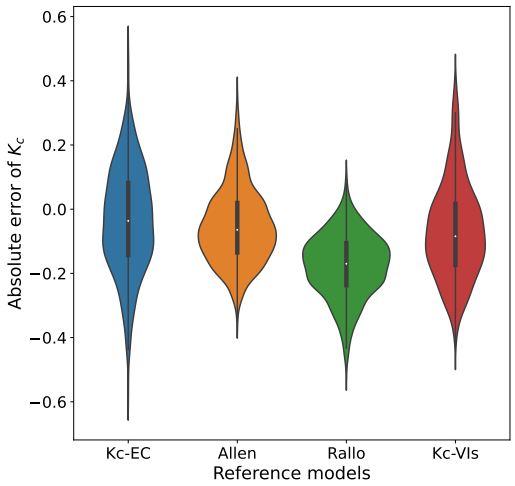
The Kc-EC model calculates the value of K_c based on the measurement of ET_a. Typically, this estimation is both complex and costly. In this case, ET_a is obtained indirectly through the measurement of latent heat flux. The EC tower yields approximately 2 GB of data for 30 days of monitoring. Furthermore, estimating the K_c values



(a) Absolute error distribution of Kc-MLP



(b) Absolute error distribution of Kc-RF



(c) Absolute error distribution of Kc-kNN

Figure 7.8: Violin plots of absolute error distribution of crop coefficient obtained using (a) Kc-MLP, (b) Kc-RF, and (c) Kc-kNN models.

Table 7.4: Comparison of required instrumentation and input data size among various models for estimating K_c values during one month of observation.

Model	Instruments	Input variable	Input data size
Kc-EC	EC Tower	Latent heat flux, ET_a	2 GB
Kc-VIs	Sentinel-2	Images, $NDVI$, $NWDI$	10 - 15 GB
Present models	WatchDog 2000 Drill and Drop	R_s, U_2, RH_{min}, RH_{max}, T_{min}, T_{max}, SWC, DOY	151 KB

requires around two hours for pre-processing this data. Regarding the Kc-VIs model, it is based on a non-linear relationship between the crop coefficient and a combination of two vegetation indices (NDVI and NDWI). The vegetation indices were investigated based on images acquired by the Sentinel-2 satellite with a temporal resolution of approximately 2-3 days [373]. In general, this model requires a large amount of data, ranging from 10 GB to 15 GB, and substantial computational capabilities for image processing. In contrast, the ML models adopted in the proposed framework (Kc-MLP, Kc-RF, Kc-kNN) were simply based on meteorological and environmental variables, such as solar radiation, relative humidity, temperature, SWC, etc. Since they needed only 151 KB of data, these models are much more lightweight in terms of resource requirements.

7.7 Summary

Crop coefficient plays a fundamental role in evaluating the performance of water resource management strategies to enhance water usage. This chapter presents a framework for modeling daily K_c values and establishes their dynamics at various crop growth stages. The proposed framework uses agro-meteorological data collected from field monitoring systems as input features, which allows the estimation of K_c without the need for expensive instrumentation (e.g. EC tower) and remote sensing imagery, thus reducing the resources required for the analysis. In order to accomplish this goal, machine learning models and algorithms for seasonal time series decomposition are combined. A total of three different ML models such as MLP, RF and kNN were adopted. In addition, the use of seasonal decomposition was essential as it reduced the component due to the presence of an actively transpiring ground cover. After post-processing, estimated K_c values were compared with known models and the best performance was achieved by the Kc-RF model with an RMSE of 0.13. Lastly, even though the study was conducted in a citrus orchard, its findings and methods can readily be applied to other irrigated agroecosystems to increase productivity while conserving water.

Chapter 8

Conclusion

In this thesis, some of the crucial challenges to make agriculture sustainable were addressed by leveraging wireless sensor networks, IoT technologies, and Artificial Intelligence. The conducted study proposes holistic and eco-friendly solutions for the digitization of agricultural businesses.

In the first part of the work, an analysis of the state of the art identified innovative aspects to be introduced in the field of wireless network, focusing on the use of LPWAN technologies. Through the comparison of path loss between SigFox and LoRaWAN, solutions were examined to improve transmission efficiency and minimize interference between the two technologies. For instance, by evaluating the effectiveness of protection distance to mitigate interference. Additionally, to enhance the environmental sustainability of agriculture, alternatives for sensor power supply were explored, aiming for sustainable monitoring in the Internet of Things through battery-free wireless sensors.

In the second part of this work, specific use cases are addressed, and alternative eco-sustainable solutions are proposed to reduce water wastage in irrigation systems, using machine learning algorithms and graph signal processing. In particular, machine learning models have been applied to predict the value of actual evapotranspiration (ET_a) in a citrus orchard, aiming to estimate this physical quantity in regulated deficit irrigation (RDI) systems, where the main objective is to save water without reducing agricultural production. Subsequently, machine learning models for ET_a prediction have been integrated into a framework for estimating and predicting the Crop Coefficient (K_c) under water stress conditions. Finally, a methodology for studying WDNs is presented, using the EPANET and NS3 simulators. This allows for the optimization and study of the smart irrigation consortium network from both water and wireless

perspectives. Indeed, thanks to the reconstruction of the water flow using graph theory, it is possible to obtain an accurate representation of the flow through strategic placement of a reduced sensor number, resulting in a substantial energy saving of approximately 73%.

Bibliography

[1] Chaojie Gu et al. Lora-based localization: Opportunities and challenges. *arXiv preprint arXiv:1812.11481*, 2018.

[2] Gaia Codeluppi et al. Lorafarm: A lorawan-based smart farming modular iot architecture. *Sensors*, 20(7):2028, 2020.

[3] Wenju Zhao et al. Design and implementation of smart irrigation system based on lora. In *IEEE Globecom Workshops*, pages 1–6, 2017.

[4] Antonio Valente et al. Low-cost lorawan node for agro-intelligence iot. *Electronics*, 9(6), 2020.

[5] JG Conijn et al. Agricultural resource scarcity and distribution: a case study of crop production in africa. *Report: Plant Research International 380*, 2011.

[6] FAO. The future of food and agriculture–trends and challenges. *Annual Report in Food and Agriculture Organisation Rome*, 2017.

[7] Angelita Rettore de Araujo Zanella et al. Security challenges to smart agriculture: Current state, key issues, and future directions. *Array*, 2020.

[8] Jinsong Wu, Song Guo, Huawei Huang, William Liu, and Yong Xiang. Information and communications technologies for sustainable development goals: State-of-the-art, needs and perspectives. *IEEE Communications Surveys & Tutorials*, 20(3):2389–2406, 2018.

[9] Olakunle Elijah et al. An overview of internet of things (iot) and data analytics in agriculture: Benefits and challenges. *IEEE Internet of Things Journal*, 5(5):3758–3773, 2018.

[10] Lei Shu et al. Guest editorial: Sustainable and intelligent precision agriculture. *IEEE Transactions on Industrial Informatics*, 17(6):4318–4321, 2021.

[11] Othmane Friha et al. Internet of things for the future of smart agriculture: A comprehensive survey of emerging technologies. *IEEE CAA J. Autom. Sinica*, 8(4):718–752, 2021.

[12] Tamoghna Ojha, Sudip Misra, and Narendra Singh Raghuwanshi. Internet of things for agricultural applications: The state of the art. *IEEE Internet of Things Journal*, 8(14):10973–10997, 2021.

[13] Ercan Avşar and Md Najmul Mowla. Wireless communication protocols in smart agriculture: A review on applications, challenges and future trends. *Ad Hoc Networks*, page 102982, 2022.

[14] Alexander Grunwald et al. Lorawan in a rural context: Use cases and opportunities for agricultural businesses. In *Mobile Communication-Technologies and Applications*, pages 1–6, 2019.

[15] Hicham Klaina, Imanol Picallo Guembe, Peio Lopez-Iturri, Miguel Ángel Campo-Bescós, Leyre Azpilicueta, Otman Aghzout, Ana Vazquez Alejos, and Francisco Falcone. Analysis of low power wide area network wireless technologies in smart agriculture for large-scale farm monitoring and tractor communications. *Measurement*, 187:110231, 2022.

[16] Esma aand others Kökten. Low-powered agriculture iot systems with lora. In *2020 IEEE Microwave Theory and Techniques in Wireless Communications (MTTW)*, volume 1, pages 178–183. IEEE, 2020.

[17] Sebastián Gutiérrez et al. Smart mobile lora agriculture system based on internet of things. In *2019 IEEE 39th Central America and Panama Convention (CONCAPAN XXXIX)*, pages 1–6. IEEE, 2019.

[18] Luiz FP Oliveira et al. Advances in agriculture robotics: A state-of-the-art review and challenges ahead. *Robotics*, 10(2):52, 2021.

[19] Badreddine Miles et al. A study of lorawan protocol performance for iot applications in smart agriculture. *Computer Communications*, 164:148–157, 2020.

[20] Mahammad Shareef Mekala et al. A survey: Smart agriculture iot with cloud computing. In *2017 international conference on microelectronic devices, circuits and systems (ICMDCS)*, pages 1–7. IEEE, 2017.

[21] Ye Liu et al. From industry 4.0 to agriculture 4.0: Current status, enabling technologies, and research challenges. *IEEE Transactions on Industrial Informatics*, 17(6):4322–4334, 2021.

[22] Zhaoyu Zhai et al. Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture*, 2020.

[23] A. O. Bernard et al. Low power long range transmitter. *European Patent EP 2 763 321 A1*, Aug. 6, 2014.

[24] Felipe S Dantas Silva et al. A survey on long-range wide-area network technology optimizations. *IEEE Access*, 9:106079–106106, 2021.

[25] Sandra Sendra et al. Lorawan network for fire monitoring in rural environments. *Electronics*, 9(3):531, 2020.

[26] Rida El Chall et al. Lorawan network: Radio propagation models and performance evaluation in various environments in lebanon. *IEEE Internet of Things Journal*, 6(2):2366–2378, 2019.

[27] Alessandro Andreadis et al. Low-power iot environmental monitoring and smart agriculture for unconnected rural areas. In *2022 20th Mediterranean Communication and Computer Networking Conference (MedComNet)*, pages 31–38, 2022.

[28] Dick Carrillo and Jorge Seki. Rural area deployment of internet of things connectivity: Lte and lorawan case study. In *2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pages 1–4. IEEE, 2017.

[29] Sebastian Sadowski and Petros Spachos. Wireless technologies for smart agricultural monitoring using internet of things devices with energy harvesting capabilities. *Computers and Electronics in Agriculture*, 172:105338, 2020.

[30] Kais Mekki et al. Overview of cellular lpwan technologies for iot deployment: Sigfox, lorawan, and nb-iot. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 197–202, 2018.

[31] Alexandru Lavric et al. Internet of things and lora™ low-power wide-area networks: a survey. In *International Symposium on Signals, Circuits and Systems (ISSCS)*, pages 1–5. IEEE, 2017.

[32] D. Croce et al. Performance of lora for bike-sharing systems. In *2019 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, pages 1–6, 2019.

[33] D Croce et al. *LoRa Technology Demystified: From Link Behavior to Cell-Level Performance*, volume 19, pages 822–834. IEEE Transactions on Wireless Communications, 2 2020.

[34] Abdollah Jabbari et al. A secure and lorawan compatible user authentication protocol for critical applications in the iot environment. *IEEE Transactions on Industrial Informatics*, 18(1):56–65, 2022.

[35] Muhammad Shoaib Farooq et al. A survey on the role of iot in agriculture for the implementation of smart farming. *IEEE Access*, 7, 2019.

[36] Hubert Zimmermann. Osi reference model-the iso model of architecture for open systems interconnection. *IEEE Transactions on communications*, 28(4):425–432, 1980.

[37] Andreas Kamilaris et al. A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, 143, 2017.

[38] Bam Bahadur Sinha and R Dhanalakshmi. Recent advancements and challenges of internet of things in smart agriculture: A survey. *Future Generation Computer Systems*, 126:169–184, 2022.

[39] Peter Salhofer. Evaluating the fiware platform. In *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

[40] Manlio Bacco et al. The digitisation of agriculture: a survey of research activities on smart farming. *Array*, 3:100009, 2019.

[41] Maria Angeles Rodriguez et al. Fiware open source standard platform in smart farming-a review. In *Working Conference on Virtual Enterprises*, pages 581–589. Springer, 2018.

[42] Pedro Corista et al. An iot agriculture system using fiware. In *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pages 1–6, 2018.

[43] Fendi Aji Purnomo et al. Development of air temperature and soil moisture monitoring systems with lora technology. In *Journal of Physics: Conference Series*, page 012029. IOP Publishing, 2021.

[44] Cayenne: The world's first drag-and-drop iot project builder, https://developers.mydevices.com/cayenne/features/ Accessed on 30/12/2022.

[45] Nuno Silva, , et al. Low-cost iot lora® solutions for precision agriculture monitoring practices. In *EPIA Conference on Artificial Intelligence*, pages 224–235. Springer, 2019.

[46] Loubna Hamami et al. Application of wireless sensor networks in the field of irrigation: A review. *Computers and Electronics in Agriculture*, 179:105782, 2020.

[47] John R Philip. Plant water relations: some physical aspects. *Annual Review of Plant Physiology*, 17(1):245–268, 1966.

[48] C Cammalleri et al. Combined use of eddy covariance and sap flow techniques for partition of et fluxes and water stress assessment in an irrigated olive orchard. *Agricultural Water Management*, 120:89–97, 2013.

[49] R K Srivastava, , et al. Effective crop evapotranspiration measurement using time-domain reflectometry technique in a sub-humid region. *Theoretical and Applied Climatology*, 129(3), 2017.

[50] Achilles D. Boursianis et al. Smart irrigation system for precision agriculture—the arethou5a iot platform. *IEEE Sensors Journal*, 21(16):17539–17547, 2021.

[51] Ravi Kishore aand others Kodali. Lora based smart irrigation system. In *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pages 1–5, 2018.

[52] Xiaohu Jiang et al. Investigation of soil wetting pattern in drip irrigation using lorawan technology. In *ASABE Annual International Virtual Meeting*, page 1, 2020.

[53] Haozhe aand others Zhang. Internet of things (iot)-based precision irrigation with lorawan technology applied to high tunnel vegetable production. In *2020*

ASABE Annual International Virtual Meeting, page 1. American Society of Agricultural and Biological Engineers, 2020.

[54] Aktham Hasan Ali et al. A smart monitoring and controlling for agricultural pumps using lora iot technology. *Indonesian Journal of Electrical Engineering and Computer Science*, 13(1):286–292, 2019.

[55] Yu Chuan Chang et al. A machine learning based smart irrigation system with lora p2p networks. In *2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, pages 1–4. IEEE, 2019.

[56] DAVID HENNA et al. Smart irrigation management system using lorawan based sensor nodes. *International Journal of Applied Engineering Research*, 15(11), 2020.

[57] Selvam Loganathan et al. Automated irrigation system based on lora and ml for marginal farmers. *International Journal of Sensors Wireless Communications and Control*, 10(3):345–353, 2020.

[58] Richard G Allen et al. Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *Fao*, 300(9):D05109, 1998.

[59] Mohamed Emharraf et al. Intelligent agriculture platform based on low energy and cost wireless sensors for efficient water irrigation. In *International conference on smart Information & communication Technologies*, pages 460–469. Springer, 2019.

[60] Zhang Xue aand others Yuan. Irrigation remote control system based on lora intelligence. In *Journal of Physics: Conference Series*, volume 1635, page 012067. IOP Publishing, 2020.

[61] Iván Froiz-Míguez et al. Design, implementation, and empirical validation of an iot smart irrigation system for fog computing applications based on lora and lorawan sensor nodes. *Sensors*, 20, 2020.

[62] Maksudjon Usmonov et al. Design and implementation of a lora based wireless control for drip irrigation systems. In *International Conference on Robotics and Automation Engineering*, pages 248–253, 2017.

[63] Van Anh Vu et al. Design of automatic irrigation system for greenhouse based on lora technology. In *2018 International Conference on Advanced Technologies for Communications (ATC)*, pages 72–77, 2018.

[64] N Ashika Barveenand others Nisa. Agriculture irrigation water demand forecasting using lora technology. *Int. Res. J. of Eng. and Technol*, 6(3):3050–3052, 2019.

[65] Mohammad Hossein Anisi et al. A survey of wireless sensor network approaches and their energy consumption for monitoring farm fields in precision agriculture. *Precision Agriculture*, 16(2):216–238, 2015.

[66] Robert Dwight Grisso et al. Precision farming tools: variable-rate application. *Virginia Cooperative Extension*, 2011.

[67] JE Sawyer. Concepts of variable rate technology with considerations for fertilizer application. *Journal of Production Agriculture*, 1994.

[68] Arif Behiccedil et al. Variable rate fertilizer application in turkish wheat agriculture: Economic assessment. *African Journal of Agricultural Research*, 5(8):647–652, 2010.

[69] Hassan S Chattha et al. Variable rate spreader for real-time spot-application of granular fertilizer in wild blueberry. *Computers and Electronics in Agriculture*, 100:70–78, 2014.

[70] Wayne Andrews et al. Remote visual monitoring system for tunnel grown crops. *MDPI*, 2021.

[71] Takaaki Kawai. Video slice: Image compression and transmission for agricultural systems. *Sensors*, 21(11):3698, 2021.

[72] Mookeun Ji et al. Lora-based visual monitoring scheme for agriculture iot. In *IEEE Sensors Applications Symposium (SAS)*, pages 1–6, 2019.

[73] Tonghao Chen et al. Efficient image transmission using lora technology in agricultural monitoring iot systems. In *International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data*, pages 937–944, 2019.

[74] Mohd Hakimi Zohari et al. Server monitoring based on iot using thingspeak. *Journal of Electrical Power and Electronic Systems*, 1(2), 2019.

[75] Redmond R Shamshiri et al. Development and field evaluation of a multichannel lora sensor for iot monitoring in berry orchards. *41. GIL-Jahrestagung, Informations-und Kommunikationstechnologie in kritischen Zeiten*, 2021.

[76] Achmad Fauzi Rachmani et al. Design of iot monitoring system based on lora technology for starfruit plantation. In *TENCON 2018 - 2018 IEEE Region 10 Conference*, pages 1241–1245, 2018.

[77] Nik Hisham Nik Ibrahim et al. Ir 4.0 using iot and lorawan to accelerate lentinula edodes growth. In *2018 2nd International Conference on Smart Sensors and Application (ICSSA)*, pages 28–32, 2018.

[78] Ritesh Kumar Singh et al. Leveraging lorawan technology for precision agriculture in greenhouses. *Sensors*, 20(7):1827, 2020.

[79] D I Sacaleanu et al. Monitoring walnut orchards with lora technology. In *2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, pages 76–79. IEEE, 2018.

[80] Davide Brunelli et al. Energy neutral machine learning based iot device for pest detection in precision agriculture. *IEEE Internet of Things Magazine*, 2(4):10–13, 2019.

[81] Yaqing He et al. Iot tree healthcare with improved sensor deployment scheme. In *2020 IEEE 2nd International Workshop on System Biology and Biomedical Systems (SBBS)*, pages 1–4, 2020.

[82] Riccardo Valentini et al. New tree monitoring systems: from industry 4.0 to nature 4.0. *ANNALS OF SILVICULTURAL RESEARCH*, 2019.

[83] J. Pedro Amaro et al. An energy study of a lorawan based electrical impedance spectroscopy module for tree health monitoring. In *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, volume 1, pages 2916–2921, 2019.

[84] Hicham Klaina et al. Implementation of an interactive environment with multilevel wireless links for distributed botanical garden in university campus. *IEEE Access*, 8:132382–132396, 2020.

[85] Daeun Yim et al. An experimental lora performance evaluation in tree farm. In *IEEE sensors applications Symposium (SAS)*, pages 1–6, 2018.

[86] Sunnyeo Park et al. Forestry monitoring system using lora and drone. In *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*, pages 1–8, 2018.

[87] André Granier. A new method of sap flow measurement in tree stems. *Annales des Sciences Forestières 42:193-200*, 1985.

[88] F C Do et al. Towards reduced heating duration in the transient thermal dissipation system of sap flow measurements. In *X International Workshop on Sap Flow 1222*, pages 149–154, 2017.

[89] Mackenzie J Dix et al. Recalibrating best practices, challenges, and limitations of estimating tree transpiration via sap flow. *Current Forestry Reports*, pages 1–7, 2021.

[90] Elisabeth Borges et al. Assessment of physiological states of plants in situ an innovative approach to the use of electrical impedance spectroscopy. *Citeseer*, 2013.

[91] E. Borges et al. Early detection and monitoring of plant diseases by bioelectric impedance spectroscopy. In *2012 IEEE 2nd Portuguese Meeting in Bioengineering (ENBENG)*, pages 1–4, 2012.

[92] Elisabeth aand others Borges. Bioimpedance parameters as indicators of the physiological states of plants in situ. *International Journal on Advances in Life Sciences*, 6:74–86, 2014.

[93] Muhammad Sohail Sardar et al. Experimental analysis of lora css wireless transmission characteristics for forestry monitoring and sensing. In *2018 International Symposium in Sensing and Instrumentation in IoT Era (ISSI)*, pages 01249–01254, 2018.

[94] Augustine Ikpehai et al. Low-power wide area network technologies for internet-of-things: A comparative review. *IEEE Internet of Things Journal*, 6(2):2225–2240, 2019.

[95] Xue-fen Wan et al. Lora propagation testing in soil for wireless underground sensor networks. In *2017 Sixth Asia-Pacific Conference on Antennas and Propagation (APCAP)*, pages 1–3. IEEE, 2017.

[96] Wataru Iwasaki et al. Iot sensors for smart livestock management. In Kohji Mitsubayashi, Osamu Niwa, and Yuko Ueno, editors, *Chemical, Gas, and Biosensors for Internet of Things and Related Applications*, pages 207–221. Elsevier, 2019.

[97] B R Dos Reis et al. A lora sensor network for monitoring pastured livestock location and activity. *Translational Animal Science*, 2021.

[98] Qi Li et al. A data collection collar for vital signs of cows on the grassland based on lora. In *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*, pages 213–217, 2018.

[99] Lorenzo Germani et al. An iot architecture for continuous livestock monitoring using lora lpwan. *Electronics*, 8(12):1435, 2019.

[100] Mukhammad Gufron Ikhsan et al. Mobile lora gateway for smart livestock monitoring system. In *2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS)*, pages 46–51. IEEE, 2018.

[101] ALAYKKA FOWLER. Cattle tracking challenge accepted: Connecting new mexico cows and desert ranches with abeeway. In *www.actility.com*, 2023.

[102] Madoune R. Seye et al. Cowshed: Communication within white spots for breeders. In *2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*, pages 236–238, 2019.

[103] Rachad Atat, Lingjia Liu, Jinsong Wu, Guangyu Li, Chunxuan Ye, and Yi Yang. Big data meet cyber-physical systems: A panoramic survey. *IEEE Access*, 6:73603–73636, 2018.

[104] IoT Analytics. 5 things to know about the lpwan market in 2021. *IoT analytics*, 2021.

[105] Nour El Hoda Djidi et al. How can wake-up radio reduce lora downlink latency for energy harvesting sensor nodes? *Sensors*, 21, 2021.

[106] F Aoudia et al. Long-short range communication network leveraging lora™ and wake-up receiver. microprocessors and microsystems".

[107] Faisal Karim Shaikh et al. Energy harvesting in wireless sensor networks: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 55:1041–1054, 2016.

[108] Hua Chao, Yu Chen, and Jinsong Wu. Power saving for machine to machine communications in cellular networks. In *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, pages 389–393, 2011.

[109] Sujesha Sudevalayam et al. Energy harvesting sensor nodes: Survey and implications. *IEEE Communications Surveys Tutorials*, 13(3):443–461, 2011.

[110] Paulo Lissa et al. Machine learning methods applied to building energy production and consumption prediction. In *AICS*, pages 236–247, 2019.

[111] Luisa Fernanda Jiménez Alvarez et al. Renewable energy prediction through machine learning algorithms. In *2020 IEEE ANDESCON*, pages 1–6. IEEE, 2020.

[112] Mohammad Riyaz Belgaum et al. Integration challenges of artificial intelligence in cloud computing, internet of things and software-defined networking. In *2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*, pages 1–5, 2019.

[113] Van-Truong Truong et al. System performance of wireless sensor network using lora–zigbee hybrid communication. *CMC-COMPUTERS MATERIALS & CONTINUA*, 68(2):1615–1635, 2021.

[114] Luca aand others Davoli. Hybrid lora-ieee 802.11 s opportunistic mesh networking for flexible uav swarming. *Drones*, 5(2):26, 2021.

[115] Atos. Fiware iot agent for lorawan protocol, https://fiware-lorawan.readthedocs.io/en/latest/index.html Accessed on 30/12/2022.

[116] Sjaak Wolfert et al. Big data in smart farming–a review. *Agricultural systems*, 153:69–80, 2017.

[117] Vishal Meshram, Kailas Patil, Vidula Meshram, Dinesh Hanchate, and SD Ramkteke. Machine learning in agriculture domain: a state-of-art survey. *Artificial Intelligence in the Life Sciences*, 1:100010, 2021.

[118] Xing Yang et al. A survey on smart agriculture: Development modes, technologies, and security and privacy challenges. *IEEE/CAA Journal of Automatica Sinica*, 8(2):273–302, 2021.

[119] Qian aand others Wang. Mr-iot: An information centric mapreduce framework for iot. In *2018 15th IEEE Annual Consumer Communications Networking Conference (CCNC)*, pages 1–6, 2018.

[120] Maurício F. Lima Pereira et al. Parallel computational structure and semantics for soil quality analysis based on lora and apache spark. In *IEEE 14th International Conference on Semantic Computing (ICSC)*, pages 332–336, 2020.

[121] Martin C Bor et al. Do lora low-power wide-area networks scale? In *Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2016.

[122] Rachel Kufakunesu et al. A survey on adaptive data rate optimization in lorawan: Recent solutions and major challenges. *Sensors*, 2020.

[123] Francesca Cuomo et al. Predicting lorawan behavior: How machine learning can help. *Computers*, 9(3), 2020.

[124] D Garlisi et al. Interference cancellation for lora gateways and impact on network capacity. *IEEE Access*, 9:128133–128146, 2021.

[125] Jeferson Rodrigues Cotrim et al. Lorawan mesh networks: a review and classification of multihop communication. *Sensors*, 20(15):4273, 2020.

[126] Kauê Vinicius de Oliveira et al. Wireless sensor network for smart agriculture using zigbee protocol. In *2017 IEEE First Summer School on Smart Cities (S3C)*, pages 61–66. IEEE, 2017.

[127] Durai Raj Vincent et al. Sensors driven ai-based agriculture recommendation model for assessing land suitability. *Sensors*, 19(17):3667, 2019.

[128] K. Lokesh Krishna, Omayo Silver, Wasswa Fahad Malende, and K. Anuradha. Internet of things application for implementation of smart agriculture system. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pages 54–59, 2017.

[129] Weishun Ma, Yan Wei, Fengshen Sun, and Yuanbao Li. Design and implementation of water saving irrigation system based on zigbee sensor network. In *IOP Conference Series: Earth and Environmental Science*, volume 252, page 052086. IOP Publishing, 2019.

[130] Zhilbert Tafa, Fatlum Ramadani, and Blerona Cakolli. The design of a zigbee-based greenhouse monitoring system. In *2018 7th Mediterranean conference on embedded computing (MECO)*, pages 1–4. IEEE, 2018.

[131] Giovanni Valecce et al. Nb-iot for smart agriculture: Experiments from the field. In *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, volume 1, pages 71–75. IEEE, 2020.

[132] Giancarlo Sciddurlo et al. Looking at nb-iot over leo satellite systems: Design and evaluation of a service-oriented solution. *IEEE Internet of Things Journal*, 9(16):14952–14964, 2021.

[133] Sakshi Popli et al. Adaptive small cell position algorithm (aspa) for green farming using nb-iot. *Journal of Network and Computer Applications*, 173:102841, 2021.

[134] German Castellanos et al. System assessment of wusn using nb-iot uav-aided networks in potato crops. *IEEE Access*, 8:56823–56836, 2020.

[135] Rubbens Boisguene et al. A survey on nb-iot downlink scheduling: Issues and potential solutions. In *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 547–551. IEEE, 2017.

[136] Brandon Foubert and Nathalie Mitton. Long-range wireless radio technologies: A survey. *Future internet*, 12(1):13, 2020.

[137] IoT Analytics. LPWAN Market Report 2018-2023. Accessed: 31/08/2022. https://iot-analytics.com/product/lpwan-market-report-2018-2023/, September 2018.

[138] SigFox. *https://build.sigfox.com/sigfox-device-radio-specifications.*

[139] Qualcomm Inc. "Narrowband IoT (NB-IoT)," RP-151621, 3GPP TSG RAN Meeting #69. Accessed: 31/08/2022. http://www.3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_69/Docs/RP-151621.zip, September 2015.

[140] Brecht Reynders, Wannes Meert, and Sofie Pollin. Range and coexistence analysis of long range unlicensed communication. In *2016 23rd International Conference on Telecommunications (ICT)*, pages 1–6, 2016.

[141] Tallal Elshabrawy and Joerg Robert. The impact of ism interference on lora ber performance. In *2018 IEEE Global Conference on Internet of Things (GCIoT)*, pages 1–5, 2018.

[142] ETSI. Technical characteristics for low power wide area networks chirp spread spectrum (lpwan-css) operating in the uhf spectrum below 1 ghz. Accessed: 31/08/2022. https://www.etsi.org/deliver/etsi_tr/103500_103599/103526/01.01.01_60/tr_103526v010101p.pdf, 2018.

[143] *SEAMCAT - Spectrum Engineering Advanced Monte Carlo Analysis Tool*, available at: https://cept.org/eco/eco-tools-and-services/seamcat-spectrum-engineering-advanced-monte-carlo-analysis-tool.

[144] Malek Al-Chalabi. Vertical farming: Skyscraper sustainability? *Sustainable Cities and Society*, 18:74–77, 2015.

[145] Mads Lauridsen, Benny Vejlgaard, Istvan Z. Kovacs, Huan Nguyen, and Preben Mogensen. Interference measurements in the european 868 mhz ism band with focus on lora and sigfox. In *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2017.

[146] Jetmir Haxhibeqiri, Adnan Shahid, Martijn Saelens, Jan Bauwens, Bart Jooris, Eli De Poorter, and Jeroen Hoebeke. Sub-gigahertz inter-technology interference. how harmful is it for lora? In *2018 IEEE International Smart Cities Conference (ISC2)*, pages 1–7, 2018.

[147] Ferran Adelantado, Xavier Vilajosana, Pere Tuset-Peiro, Borja Martinez, Joan Melia-Segui, and Thomas Watteyne. Understanding the limits of lorawan. *IEEE Communications Magazine*, 55(9):34–40, 2017.

[148] Daniele Croce, Domenico Garlisi, Fabrizio Giuliano, Alice Lo Valvo, Stefano Mangione, and Ilenia Tinnirello. Performance of lora for bike-sharing systems. In *2019 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, pages 1–6, 2019.

[149] Konstantin Mikhaylov, Juha Petäjäjärvi, and Janne Janhunen. On lorawan scalability: Empirical evaluation of susceptibility to inter-network interference. In *2017 European Conference on Networks and Communications (EuCNC)*, 2017.

[150] Naga Srinivasarao Chilamkurthy, Om Jee Pandey, Anirban Ghosh, Linga Reddy Cenkeramaddi, and Hong-Ning Dai. Low-power wide-area networks: A broad overview of its different aspects. *IEEE Access*, 2022.

[151] Yeon-Gyu Park, Eun-Young Chang, Il-Kyoo Lee, and Yan-Ming Cheng. Analysis of wi-fi halow device interference to lte user equipment. *Journal of Communications*, pages 58–64, 01 2020.

[152] Rogério Dionísio, Teodora Lolić, and Pedro Torres. Electromagnetic interference analysis of industrial iot networks: From legacy systems to 5g. In *2020 IEEE Microwave Theory and Techniques in Wireless Communications (MTTW)*, volume 1, pages 41–46, 2020.

[153] Lorenzo Vangelista, Luca Dell'Anna, and Ivano Calabrese. On the coexistence of lorawan and legacy short range devices in unlicensed bands in europe. In *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–5, 2019.

[154] Winfred Ingabire, Hadi Larijani, and Ryan M. Gibson. Performance evaluation of propagation models for lorawan in an urban environment. In *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pages 1–6, 2020.

[155] Yukimasa Nagai, Jianlin Guo, Philip Orlik, Takenori Sumi, Benjamin A. Rolfe, and Hiroshi Mineno. Sub-1 ghz frequency band wireless coexistence for the internet of things. *IEEE Access*, 9:119648–119665, 2021.

[156] Alexandru Lavric, Adrian I. Petrariu, and Valentin Popa. Long range sigfox communication protocol scalability analysis under large-scale, high-density conditions. *IEEE Access*, 7:35816–35825, 2019.

[157] ETSI. Short range devices (srd) operating in the frequency range 25 mhz to 1 000 mhz; part 2: 2014/53/eu. Accessed: 31/08/2022. https://www.etsi.org/deliver/etsi_en/300200_300299/30022002/03.01.01_60/en_30022002v030101p.pdf.

[158] Preben Elgaard Mogensen and Jeroen Wigard. Cost action 231: Digital mobile radio towards future generation system. Final Report. Section 5.2: On antenna and frequency diversity in GSM. Section 5.3: Capacity study of frequency hopping GSM network, 1999.

[159] P Series. Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 900 mhz to 100 ghz. *Recommendation ITU-R*, pages 1238–7, 2012.

[160] Weitao Xu, Jun Young Kim, Walter Huang, Salil S. Kanhere, Sanjay K. Jha, and Wen Hu. Measurement, characterization, and modeling of lora technology in multifloor buildings. *IEEE Internet of Things Journal*, 7(1):298–310, 2020.

[161] Rida El Chall, Samer Lahoud, and Melhem El Helou. Lorawan network: Radio propagation models and performance evaluation in various environments in lebanon. *IEEE Internet of Things Journal*, 2019.

[162] Juha Petajajarvi, Konstantin Mikhaylov, Antti Roivainen, Tuomo Hanninen, and Marko Pettissalo. On the coverage of lpwans: range evaluation and channel attenuation model for lora technology. In *2015 14th international conference on its telecommunications (itst)*, pages 55–59. IEEE, 2015.

[163] Pascal Jörke, Stefan Böcker, Florian Liedmann, and Christian Wietfeld. Urban channel models for smart city iot-networks based on empirical measurements of lora-links at 433 and 868 mhz. In *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017.

[164] Silvia Demetri, Marco Zúñiga, Gian Pietro Picco, Fernando Kuipers, Lorenzo Bruzzone, and Thomas Telkamp. Automated estimation of link quality for lora: A remote sensing approach. In *2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2019.

[165] Madoune R Seye, Bassirou Ngom, Bamba Gueye, and Moussa Diallo. A study of lora coverage: range evaluation and channel attenuation model. In *2018 1st International Conference on Smart Cities and Communities (SCCIC)*, pages 1–4. IEEE, 2018.

[166] Gianni Pasolini, Chiara Buratti, Luca Feltrin, Flavio Zabini, Cristina De Castro, Roberto Verdone, and Oreste Andrisano. Smart city pilot projects using lora and ieee802. 15.4 technologies. *Sensors*, 18(4):1118, 2018.

[167] Martin Stusek, Dmitri Moltchanov, Pavel Masek, Konstantin Mikhaylov, Otto Zeman, Martin Roubicek, Yevgeni Koucheryavy, and Jiri Hosek. Accuracy assessment and cross-validation of lpwan propagation models in urban scenarios. *IEEE Access*, 8:154625–154636, 2020.

[168] Radim Kalfus and Tomáš Hégr. Ultra narrow band radio technology in high-density built-up areas. In *International Conference on Information and Software Technologies*, pages 663–676. Springer, 2016.

[169] Philip A. Catherwood, Michael Little, Dewar Finlay, and James McLaughlin. Recovery of incapacitated commercial delivery drones using lpwan technology. *IEEE Intelligent Transportation Systems Magazine*, 2020.

[170] Lorenzo Germani, Vanni Mecarelli, Giuseppe Baruffa, Luca Rugini, and Fabrizio Frescura. An iot architecture for continuous livestock monitoring using lora lpwan. *Electronics*, 8(12):1435, 2019.

[171] Milos Stankovic. Nb-iot and lorawan performance testing in urban and rural environment, Dec 2020.

[172] Mike Oluwatayo Ojo, Davide Adami, and Stefano Giordano. Experimental evaluation of a lora wildlife monitoring network in a forest vegetation area. *Future Internet*, 13(5):115, 2021.

[173] Pablo Avila-Campos, Fabian Astudillo-Salinas, Andres Vazquez-Rodas, and Alcides Araujo. Evaluation of lorawan transmission range for wireless sensor networks in riparian forests. In *Proceedings of the 22nd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 199–206, 2019.

[174] SEAMCAT configuration files. Accessed: 31/08/2022. https://github.com/fabriziogiuliano/css-unb-seamcat.git.

[175] CEPT. Estimating the sustainability of shared use of iot lpwans and legacy srds within the band 863-870 mhz - (part of se24_wi69). Accessed: 31/08/2022. https://cept.org/Documents/se-24/48179/se24-18-069iot-lpwan-vs-legacy-srd-study-wi69-Copen-hagen, 2018.

[176] Daniele Croce, Michele Gucciardo, Stefano Mangione, Giuseppe Santaromita, and Ilenia Tinnirello. Lora technology demystified: From link behavior to cell-level performance. *IEEE Transactions on Wireless Communications*, 19(2):822–834, 2020.

[177] Domenico Garlisi, Stefano Mangione, Fabrizio Giuliano, Daniele Croce, Giovanni Garbo, and Ilenia Tinnirello. Interference cancellation for lora gateways and impact on network capacity. *IEEE Access*, 9:128133–128146, 2021.

[178] Antonino Pagano, Daniele Croce, Ilenia Tinnirello, and Gianpaolo Vitale. A survey on lora for smart agriculture: Current trends and future perspectives. *IEEE Internet of Things Journal*, 10(4):3664–3679, 2023.

[179] Osmel Martínez Rosabal, Onel L. Alcaraz López, Hirley Alves, and Matti Latva-Aho. Sustainable rf wireless energy transfer for massive iot: Enablers and challenges. *IEEE Access*, 11:133979–133992, 2023.

[180] Marco Gonzalez, Pengcheng Xu, Rémi Dekimpe, Maxime Schramme, Ivan Stupia, Thibault Pirson, and David Bol. Technical and ecological limits of 2.45-ghz wireless power transfer for battery-less sensors. *IEEE Internet of Things Journal*, 10(17):15431–15442, 2023.

[181] Yiqin Deng, Zhigang Chen, Xin Yao, Shahzad Hassan, and Ali MA Ibrahim. Parallel offloading in green and sustainable mobile edge computing for delay-constrained iot system. *IEEE Transactions on Vehicular Technology*, 68(12):12202–12214, 2019.

[182] Roberto La Rosa, Catherine Dehollain, Mario Costanza, Angelo Speciale, Fabio Viola, and Patrizia Livreri. A battery-free wireless smart sensor platform with bluetooth low energy connectivity for smart agriculture. In *2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON)*, pages 554–558. IEEE, 2022.

[183] Namanu Panayanthattaa, Giacomo Clementi, Merieme Ouhabaz, Mario Costanza, Samuel Margueron, Ausrine Bartasyte, Skandar Basrour, Edwige Bano, Laurent Montes, Catherine Dehollain, et al. A smart battery free system for wireless condition monitoring using piezoelectric energy harvester. In *2022 Smart Systems Integration (SSI)*, pages 1–4. IEEE, 2022.

[184] Mahantesh P Mattada and Hansraj Guhilot. Time-to-digital converters—a comprehensive review. *International Journal of Circuit Theory and Applications*, 49(3):778–800, 2021.

[185] Jakub Szyduczyński, Dariusz Kościelnik, and Marek Miśkowicz. Time-to-digital conversion techniques: a survey of recent developments. *Measurement*, page 112762, 2023.

[186] STMicroelectronics. Ultra-Low-Power 32-bit MCU Arm-Based Cortex-M0+, up to 16 KB Flash Memory, 8 KB SRAM, 2 KB EEPROM, USB,

ADC, DAC, STM32L031x4 STM32L031x6 datasheet. In *Available online: https://www.st.com/resource/en/datasheet/stm32l031f4.pdf (accessed on March 25th, 2023)*. STMicroelectronics Geneva,Switzerland, 2018.

[187] Roberto La Rosa, Catherine Dehollain, Andreas Burg, Mario Costanza, and Patrizia Livreri. An energy-autonomous wireless sensor with simultaneous energy harvesting and ambient light sensing. *IEEE Sensors Journal*, 21(12):13744–13752, 2021.

[188] Wei Zhang, Rui Ma, Xiayu Wang, Hao Zheng, and Zhangming Zhu. A high linearity tdc with a united-reference fractional counter for lidar. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 69(2):564–572, 2021.

[189] Kentaro Yoshioka, Hiroshi Kubota, Tomonori Fukushima, et al. A 20-ch tdc/adc hybrid architecture lidar soc for 240 × 96 pixel 200-m range imaging with smart accumulation technique and residue quantizing sar adc. *IEEE Journal of Solid-State Circuits*, 53(11):3026–3038, 2018.

[190] Maliang Liu, Haizhu Liu, Xiongzheng Li, and Zhangming Zhu. A 60-m range 6.16-mw laser-power linear-mode lidar system with multiplex adc/tdc in 65-nm cmos. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 67(3):753–764, 2019.

[191] Qiwei Huang, Hyobin Joo, Jinwoo Kim, Chenchang Zhan, and Jinwook Burm. An energy-efficient frequency-domain cmos temperature sensor with switched vernier time-to-digital conversion. *IEEE Sensors Journal*, 17(10):3001–3011, 2017.

[192] Chun-Chi Chen, Chao-Lieh Chen, Wei Fang, and Yen-Chan Chu. All-digital cmos time-to-digital converter with temperature-measuring capability. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 28(9):2079–2083, 2020.

[193] W. El-Halwagy, P. Mousavi, and M. Hossain. A 79db sndr, 10mhz bw, 675ms/s open-loop time-based adc employing a 1.15ps sar-tdc. In *2016 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, pages 321–324, 2016.

[194] A Jimenez-Irastorza, Juan F Sevillano, Fernando Arizti, I Rebollo, and Roc Berenguer. Comparative study on time-to-digital converters for low-power rfid tag sensors. *Microelectronics Journal*, 44(10):912–919, 2013.

[195] Seon-Kyoo Lee, Young-Hun Seo, Hong-June Park, and Jae-Yoon Sim. A 1 ghz adpll with a 1.25 ps minimum-resolution sub-exponent tdc in 0.18 μ m cmos. *IEEE Journal of Solid-State Circuits*, 45(12):2874–2881, 2010.

[196] Alberto Demarziani, Edoardo Bonizzoni, Franco Maloberti, and Alessandro D'Amato. Design of a low power time to digital converter for flow metering applications. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1646–1649. IEEE, 2015.

[197] José A Hidalgo-López and Julián Castellanos-Ramos. Two proposals to simplify resistive sensor readout based on resistance-to-time-to-digital conversion. *Measurement*, 213:112728, 2023.

[198] Alessandro Depari, Paolo Bellagente, Paolo Ferrari, Alessandra Flammini, Marco Pasetti, Stefano Rinaldi, and Emiliano Sisinni. Minimal wide-range resistive sensor-to-microcontroller interface for versatile iot nodes. *IEEE Transactions on Instrumentation and Measurement*, 71:1–9, 2022.

[199] Ponnalagu Ramanathan Nagarajan, Boby George, and V. Jagadeesh Kumar. Improved single-element resistive sensor-to-microcontroller interface. *IEEE Transactions on Instrumentation and Measurement*, 66(10):2736–2744, 2017.

[200] Zivko Kokolanski, Manel Gasulla, and Ferran Reverter. Differential inductive sensor-to-microcontroller interface circuit. In *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–5. IEEE, 2019.

[201] Zbigniew Czaja. Time-domain measurement methods for r, l and c sensors based on a versatile direct sensor-to-microcontroller interface circuit. *Sensors and Actuators A: Physical*, 274:199–210, 2018.

[202] Venkatesh Gaddikeri, Murtaza Hasan, Dilip Kumar, Arjamadatta Sarangi, and Wasi Alam. Performance analysis and measurement of soil moisture content by piezoresistive sensor. *Mapan*, 37(1):149–160, 2022.

[203] Huan Liu, Taipeng Wang, Junyao Wang, Qi Hou, Dingyi Ren, Guangze Gao, Yaqun Li, Jingran Quan, Jianxin Xu, and Hongxu Pan. Upcycling of agrowaste: Biobased piezoresistive sensor by exploiting the natural porous structure of cornstalk pith for detecting pressure signals. *ACS Sustainable Chemistry & Engineering*, 2024.

[204] Rajul S Patkar, Mamta Ashwin, and V Ramgopal Rao. Piezoresistive microcantilever based lab-on-a-chip system for detection of macronutrients in the soil. *Solid-State Electronics*, 138:94–100, 2017.

[205] Rajul S. Patkar, Madhuri Vinchurkar, Mamta Ashwin, Andrea Adami, Flavio Giacomozzi, Leandro Lorenzelli, Maryam Shojaei Baghini, and V. Ramgopal Rao. Microcantilever based dual mode biosensor for agricultural applications. *IEEE Sensors Journal*, 20(13):6826–6832, 2020.

[206] Chrysanthos Maraveas and Thomas Bartzanas. Sensors for structural health monitoring of agricultural structures. *Sensors*, 21(1):314, 2021.

[207] Rajul S. Patkar, Madhuri Vinchurkar, Mamta Ashwin, and V. Ramgopal Rao. A novel pet-based piezoresistive mems sensor platform for agricultural applications. *Journal of Microelectromechanical Systems*, 26(4):746–748, 2017.

[208] Borja Genoves Guzman, Javier Talavante, Dayrene Frometa Fonseca, Muhammad Sarmad Mir, Domenico Giustiniano, Katia Obraczka, Michael E. Loik, Sylvie Childress, and Darryl G. Wong. Toward sustainable greenhouses using battery-free lifi-enabled internet of things. *IEEE Communications Magazine*, 61(5):129–135, 2023.

[209] Jens F Peters, Manuel Baumann, Benedikt Zimmermann, Jessica Braun, and Marcel Weil. The environmental impact of li-ion batteries and the role of key parameters–a review. *Renewable and Sustainable Energy Reviews*, 67:491–506, 2017.

[210] PG Balakrishnan, R Ramesh, and T Prem Kumar. Safety mechanisms in lithium-ion batteries. *Journal of power sources*, 155(2):401–414, 2006.

[211] Core Specification Working Group. *Bluetooth Core Specification*. Bluetooth Special Interest Group (SIG), core-main@bluetooth.org, v5.2 edition, December 2019.

[212] Panasonic. Amorton solar cell catalog. In *Available online: https://www.panasonic-electric-works.com/eu/amorton-amorphous-silicon-solar-cells.htm (accessed on June 20th, 2020)*. Panasonic Corporation, Japan, 2019.

[213] Cheng-Hsin Chuang, Da-Huei Lee, Wan-Jung Chang, Wan-Ching Weng, Muhammad Omar Shaikh, and Chung-Lin Huang. Real-time monitoring via

patch-type piezoelectric force sensors for internet of things based logistics. *IEEE Sensors Journal*, 17(8):2498–2506, 2017.

[214] Xiaoqiong Li, Bei Qi, Xiao Wanv, Jingwen Zhang, Wei Yang, Yongjun Xiao, Fuwei Mao, Kailin Cai, Liang Huang, and Jun Zhou. Electret-based flexible pressure sensor for respiratory diseases auxiliary diagnosis system using machine learning technique. *Nano Energy*, page 108652, 2023.

[215] AS Fiorillo, CD Critello, and SA Pullano. Theory, technology and applications of piezoresistive sensors: A review. *Sensors and Actuators A: Physical*, 281:156–175, 2018.

[216] Abdo-Rahmane Anas Laaraibi, Gurvan Jodin, Damien Hoareau, Nicolas Bideau, and Florence Razan. Flexible dynamic pressure sensor for insole based on inverse viscoelastic model. *IEEE Sensors Journal*, 23(7):7634–7643, 2023.

[217] Liangqi Yuan, Hongwei Qu, and Jia Li. Velostat sensor array for object recognition. *IEEE Sensors Journal*, 22(2):1692–1704, 2021.

[218] Muhammad Usman Ali Khan, Raad Raad, Faisel Tubbal, Panagiotis Ioannis Theoharis, Sining Liu, and Javad Foroughi. Bending analysis of polymer-based flexible antennas for wearable, general iot applications: A review. *Polymers*, 13(3):357, 2021.

[219] Abdo-Rahmane Anas Laaraibi, Corentin Depontailler, Gurvan Jodin, Damien Hoareau, Nicolas Bideau, and Florence Razan. An innovative wearable sensing system based on flexible piezoresistive sensors to estimate upper body joint angle using a nonlinear autoregressive exogenous neural model. *IEEE Sensors Journal*, 23(21):26539–26550, 2023.

[220] Luigi Raiano, Joshua Di Tocco, Carlo Massaroni, Giovanni Di Pino, Emiliano Schena, and Domenico Formica. A pca-based method to select the number and the body location of piezoresistive sensors in a wearable system for respiratory monitoring. *IEEE Sensors Journal*, 21(5):6847–6855, 2020.

[221] Velostat. Film, velostat, 36 x 150 roll 67y8891. In *https://it.farnell.com/scs/1704-36x150/film-velostat-36-x-150-roll-67y8891/dp/3218528*, accessed: 16.02.2024.

[222] ACF. Ruban adhesif transfert conducteur 9703. https://fr.rs-online.com/web/p/rubans-metalliques/2209928. Accessed: 20.04.2022.

[223] IET. PRS-330 PRECISION PROGRAMMABLE RESISTANCE BOX & RTD SIMULATOR. https://www.ietlabs.com/prs-330-programmable-decade-resistor.html. Accessed: 16/07/2023.

[224] Saverio Grutta. Radio communication range estimation in ISM band. Technical Report Rev 2, STMicroelectronics, July 2022.

[225] Antonino Pagano, Daniele Croce, Ilenia Tinnirello, and Gianpaolo Vitale. A survey on lora for smart agriculture: Current trends and future perspectives. *IEEE Internet of Things Journal*, 2022.

[226] Sujesha Sudevalayam and Purushottam Kulkarni. Energy harvesting sensor nodes: Survey and implications. *IEEE Communications Surveys & Tutorials*, 13(3):443–461, 2011.

[227] George Y. Odongo, Richard Musabe, Damien Hanyurwimfura, and Abubakar Diwani Bakari. An efficient lora-enabled smart fault detection and monitoring platform for the power distribution system using self-powered iot devices. *IEEE Access*, 10:73403–73420, 2022.

[228] Stefano Buzzi, Chih-Lin I, Thierry E. Klein, H. Vincent Poor, Chenyang Yang, and Alessio Zappone. A survey of energy-efficient techniques for 5g networks and challenges ahead. *IEEE Journal on Selected Areas in Communications*, 34(4):697–709, 2016.

[229] Fan Wu, Christopf Rudiger, and Rasit Yuce. Real-time performance of a self-powered environmental iot sensor network system. *Sensors*, 2017.

[230] U. Alvarado, A. Juanicorena, I Adin, B. Sedano, and I. Gutierrez. Energy harvesting technologies for low-power electronics. *Trans. Emerging Tel. Tech.*, 2012.

[231] Gianpaolo. Vitale and Emiliano Pipitone. A regenerative braking system for internal combustion engine vehicles using supercapacitors as energy storage elements - part 1: System analysis and modelling. *Journal of power sources*, 2020.

[232] Hafiz Husnain Raza Sherazi, Muhammad Ali Imran, Gennaro Boggia, and Luigi Alfredo Grieco. Energy harvesting in lorawan: A cost analysis for the industry 4.0. *IEEE Communications Letters*, 22(11):2358–2361, 2018.

[233] Matteo Pizzotti, Luca Perilli, Massimo Del Prete, Davide Fabbri, Roberto Canegallo, Michele Dini, Diego Masotti, Alessandra Costanzo, Eleonora Franchi Scarselli, and Aldo Romani. A long-distance rf-powered sensor node with adaptive power management for iot applications. *Sensors*, 17(8):1732, 2017.

[234] Mohsen Karimi, Yidi Wang, and Hyoseung Kim. Energy-adaptive real-time sensing for batteryless devices. In *2022 IEEE 28th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, pages 205–211. IEEE, 2022.

[235] Kunyi Chen, Hong Gao, Zhipeng Cai, Quan Chen, and Jianzhong Li. Distributed energy-adaptive aggregation scheduling with coverage guarantee for battery-free wireless sensor networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 1018–1026. IEEE, 2019.

[236] Fan Yang, Ashok Samraj Thangarajan, Gowri Sankar Ramachandran, Bhaskar Krishnamachari, Wouter Joosen, Christophe Huygens, and Danny Hughes. Astar: Sustainable battery free energy harvesting for heterogeneous platforms and dynamic environments. In *Proceedings of the 2019 International Conference on Embedded Wireless Systems and Networks, EWSN 2019*, pages 71–82. ACM, 2019.

[237] Francesco Fraternali, Bharathan Balaji, Dhiman Sengupta, Dezhi Hong, and Rajesh K Gupta. Ember: energy management of batteryless event detection sensors with deep reinforcement learning. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 503–516, 2020.

[238] Martina Capuzzo, Carmen Delgado, Jeroen Famaey, and Andrea Zanella. Energy-aware packet schedulers for battery-less lorawan nodes. *ArXiv*, abs/2212.09453, 2022.

[239] Mehmet Erkan Yuksel and Huseyin Fidan. Energy-aware system design for batteryless lpwan devices in iot applications. *Ad Hoc Networks*, 122:102625, 2021.

[240] Francesco Fraternali, Bharathan Balaji, Dezhi Hong, Yuvraj Agarwal, and Rajesh K Gupta. Marble: collaborative scheduling of batteryless sensors with meta reinforcement learning. In *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pages 140–149, 2021.

[241] Abu Bakar, Tousif Rahman, Rishad Shafik, Fahim Kawsar, and Alessandro Montanari. Adaptive intelligence for batteryless sensors using software-accelerated tsetlin machines. In *Proceedings of SenSys*, 2022.

[242] Alessandro Montanari, Manuja Sharma, Dainius Jenkus, Mohammed Alloulah, Lorena Qendro, and Fahim Kawsar. eperceptive: energy reactive embedded intelligence for batteryless sensors. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 382–394, 2020.

[243] Abu Bakar, Rishabh Goel, Jasper de Winkel, Jason Huang, Saad Ahmed, Bashima Islam, Przemysław Pawełczak, Kasım Sinan Yıldırım, and Josiah Hester. Protean: An energy-efficient and heterogeneous platform for adaptive and hardware-accelerated battery-free computing. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 207–221, 2022.

[244] Gaia Maselli, Mauro Piva, and John A Stankovic. Adaptive communication for battery-free devices in smart homes. *IEEE Internet of Things Journal*, 6(4):6977–6988, 2019.

[245] Carmen Delgado, José María Sanz, Chris Blondia, and Jeroen Famaey. Batteryless lorawan communications using energy harvesting: Modeling and characterization. *IEEE Internet of Things Journal*, 8(4):2694–2711, 2020.

[246] Adnan Sabovic, Carmen Delgado, Dragan Subotic, Bart Jooris, Eli De Poorter, and Jeroen Famaey. Energy-aware sensing on battery-less lorawan devices with energy harvesting. *Electronics*, 9(6):904, 2020.

[247] Orestis Georgiou, Constantinos Psomas, Eleni Demarchou, and Ioannis Krikidis. Lora network performance under ambient energy harvesting and random transmission schemes. In *ICC 2021-IEEE International Conference on Communications*, pages 1–6. IEEE, 2021.

[248] Francesco Orfei, Chiara Benedetta Mezzetti, and Francesco Cottone. Vibrations powered lora sensor: An electromechanical energy harvester working on a real bridge. In *2016 IEEE SENSORS*, pages 1–3. IEEE, 2016.

[249] Vincent Boitier, Bruno Estibals, Florian Huet, and Lionel Seguier. Battery-free power supply for wireless sensor combining photovoltaic cells and supercapacitors. *Energy and Power Engineering*, 15(3):151–179, 2023.

[250] Fayçal Ait Aoudia, Matthieu Gautier, and Olivier Berder. Rlman: An energy manager based on reinforcement learning for energy harvesting wireless sensor networks. *IEEE Transactions on Green Communications and Networking*, 2(2):408–417, 2018.

[251] Francesco Fraternali, Bharathan Balaji, Yuvraj Agarwal, and Rajesh K Gupta. Aces: Automatic configuration of energy harvesting sensors with reinforcement learning. *ACM Transactions on Sensor Networks (TOSN)*, 16(4):1–31, 2020.

[252] Yubo Luo and Shahriar Nirjon. Smarton: Just-in-time active event detection on energy harvesting systems. In *2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 35–44. IEEE, 2021.

[253] Yubo Luo and Shahriar Nirjon. Spoton: Just-in-time active event detection on energy autonomous sensing systems. *Brief Presentations Proceedings (RTAS 2019)*, 9, 2019.

[254] Tongxin Shu, Min Xia, Jiahong Chen, and Clarence De Silva. An energy efficient adaptive sampling algorithm in a sensor network for automated water quality monitoring. *Sensors*, 17(11):2551, 2017.

[255] Elvina Gindullina, Leonardo Badia, and Xavier Vilajosana. Energy modeling and adaptive sampling algorithms for energy-harvesting powered nodes with sampling rate limitations. *Transactions on Emerging Telecommunications Technologies*, 31(3):e3754, 2020.

[256] STMicroelectronics. Ultra-low power multi-modulation wireless STM32WL5x microcontrollers. Accessed: 17/05/2023. https://www.st.com/en/microcontrollers-microprocessors/stm32wl5x.html.

[257] DG Erbs, SA Klein, and JA Duffie. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. *Solar energy*, 28(4):293–302, 1982.

[258] SIAS. Servizio informativo agrometeorologico siciliano. Accessed: 17/05/2023. http://www.sias.regione.sicilia.it/.

[259] National solar radiation database. Accessed: 17/05/2023. https://nsrdb.nrel.gov/.

[260] Sensing labs temporal redundancy. Accessed: 17/05/2023. https://sensing-labs.com/the-temporal-redundancy/.

[261] Adrian K. TTGO Tbeam Ultra Low Power Modification. Accessed: 07/12/2022. https://github.com/HikariBoy/TTGO-Tbeam-Ultra-Low-Power-Modification.

[262] Maria Carmela Di Piazza and Gianpaolo Vitale. *Photovoltaic sources: modeling and emulation*. Springer, 2013.

[263] Cai Tao, Duan Shanxu, and Chen Changsong. Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement. In *The 2nd International Symposium on Power Electronics for Distributed Generation Systems*, pages 773–777, 2010.

[264] Mtx s.r.l. Accessed: 21/06/2023. http://www.mtx.it.

[265] Maria Carmela Di Piazza and Gianpaolo Vitale. Photovoltaic field emulation including dynamic and partial shadow conditions. *Applied Energy*, 87(3):814–823, 2010.

[266] Edgar Fabián Rivera Guzmán, Edison David Mañay Chochos, Mauricio Danilo Chiliquinga Malliquinga, Paúl Francisco Baldeón Egas, and Renato Mauricio Toasa Guachi. Lora network-based system for monitoring the agricultural sector in andean areas: Case study ecuador. *Sensors*, 22(18):6743, 2022.

[267] Deniz TAŞKIN and Selçuk Yazar. A long-range context-aware platform design for rural monitoring with iot in precision agriculture. *International Journal of Computers Communications & Control*, 15(2), 2020.

[268] Fabrizio Giuliano, Antonino Pagano, Daniele Croce, Gianpaolo Vitale, and Ilenia Tinnirello. Adaptive algorithms for batteryless lora-based sensors. *Sensors*, 23(14):6568, 2023.

[269] David Seckler, Randolph Barker, and Upali Amarasinghe. Water scarcity in the twenty-first century. *International Journal of Water Resources Development*, 15(1-2):29–42, 1999.

[270] Junguo Liu, Hong Yang, Simon N Gosling, Matti Kummu, Martina Flörke, Stephan Pfister, Naota Hanasaki, Yoshihide Wada, Xinxin Zhang, Chunmiao Zheng, et al. Water scarcity assessments in the past, present, and future. *Earth's future*, 5(6):545–559, 2017.

[271] E Creaco, G Barbero, A Montanaro, and M Reduzzi. Effective optimization of irrigation networks with pressure-driven outflows at randomly selected installation nodes. *Scientific Reports*, 13(1):19218, 2023.

[272] K et al Urama. *Options for decoupling economic growth from water use and water pollution. A report of the Water Working Group of the UNEP International Resource Panel*. 03 2016.

[273] Directorate-General for Environment European Commission. Report from the commission to the european parliament and the council on the implementation of the water framework directive (2000/60/ec) and the floods directive (2007/60/ec), 2019.

[274] Nicolas Guyennon, Emanuele Romano, and Ivan Portoghese. Long-term climate sensitivity of an integrated water supply system: The role of irrigation. *Science of the total environment*, 565:68–81, 2016.

[275] Zheng Liu and Yehuda Kleiner. State of the art review of inspection technologies for condition assessment of water pipes. *Measurement*, 46(1):1–15, 2013.

[276] Mohammed Rezwanul Islam, Sami Azam, Bharanidharan Shanmugam, and Deepika Mathur. A review on current technologies and future direction of water leakage detection in water distribution network. *IEEE Access*, 10:107177–107201, 2022.

[277] N. Sornin, A. Yegin, and et al. LoRaWAN 1.1 Specification, October 2017. Accessed: 2020-04-16.

[278] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst. Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE*, 106(5):808–828, 2018.

[279] Mark Newman, Albert-László Barabási, and Duncan J Watts. *The structure and dynamics of networks*. Princeton university press, 2011.

[280] Georgios B Giannakis, Yanning Shen, and Georgios Vasileios Karanikolas. Topology identification and learning over graphs: Accounting for nonlinearities and dynamics. *Proceedings of the IEEE*, 106(5):787–807, 2018.

[281] Nitin J Williams, Ian Daly, and Slawomir J Nasuto. Markov model-based method to analyse time-varying networks in eeg task-related data. *Frontiers in computational neuroscience*, 12:76, 2018.

[282] GV Karanikolas and Georgios B Giannakis. Identifying directional connections in brain networks via multi-kernel granger models. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6304–6308. IEEE, 2017.

[283] Georgios V Karanikolas, Georgios B Giannakis, Konstantinos Slavakis, and Richard M Leahy. Multi-kernel based nonlinear models for connectivity identification of brain networks. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6315–6319. IEEE, 2016.

[284] Yanning Shen, Georgios B Giannakis, and Brian Baingana. Nonlinear structural vector autoregressive models with application to directed brain networks. *IEEE Transactions on Signal Processing*, 67(20):5325–5339, 2019.

[285] Daniel Romero, Vassilis N Ioannidis, and Georgios B Giannakis. Kernel-based reconstruction of space-time functions on dynamic graphs. *IEEE Journal of Selected Topics in Signal Processing*, 11(6):856–869, 2017.

[286] Weiyu Huang, Thomas AW Bolton, John D Medaglia, Danielle S Bassett, Alejandro Ribeiro, and Dimitri Van De Ville. A graph signal processing perspective on functional brain imaging. *Proceedings of the IEEE*, 106(5):868–885, 2018.

[287] Alexandre Ciancio, Sundeep Pattem, Antonio Ortega, and Bhaskar Krishnamachari. Energy-efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm. In *Proceedings of the 5th international conference on Information processing in sensor networks*, pages 309–316, 2006.

[288] Armando Di Nardo, Michele Di Natale, Giovanni F Santonastaso, Velitchko G Tzatchkov, and Victor H Alcocer-Yamanaka. Water network sectorization based on graph theory and energy performance indices. *Journal of Water Resources Planning and Management*, 140(5):620–629, 2014.

[289] S Mohan Kumar, Shankar Narasimhan, and S Murty Bhallamudi. State estimation in water distribution networks using graph-theoretic reduction strategy. *Journal of Water Resources Planning and Management*, 134(5):395–403, 2008.

[290] Velitchko G Tzatchkov, Victor H Alcocer-Yamanaka, and Víctor Bourguett Ortíz. Graph theory based algorithms for water distribution network sectorization projects. In *Water Distribution Systems Analysis Symposium 2006*, pages 1–15, 2008.

[291] Manuel Herrera, Edo Abraham, and Ivan Stoianov. A graph-theoretic framework for assessing the resilience of sectorised water distribution networks. *Water Resources Management*, 30:1685–1699, 2016.

[292] Aravind Rajeswaran, Sridharakumar Narasimhan, and Shankar Narasimhan. A graph partitioning algorithm for leak detection in water distribution networks. *Computers & Chemical Engineering*, 108:11–23, 2018.

[293] Jiheon Kang, Youn-Jong Park, Jaeho Lee, Soo-Hyun Wang, and Doo-Seop Eom. Novel leakage detection by ensemble cnn-svm and graph-based localization in water distribution systems. *IEEE Transactions on Industrial Electronics*, 65(5):4279–4289, 2017.

[294] Tao Bai and Pejman Tahmasebi. Graph neural network for groundwater level forecasting. *Journal of Hydrology*, 616:128792, 2023.

[295] LR Bindu, P Titus, and D Dhanya. Clustered wireless sensor network in precision agriculture via graph theory. *Intelligent Automation & Soft Computing*, 36(2), 2023.

[296] Hefei Gao, Naiyu Cui, and Wei Wang. Deployment optimization of intelligent wireless sensor network using graph similarity based on multi-granularity cross representation and matching. *IEEE/ACM Transactions on Networking*, pages 1–13, 2023.

[297] Hefei Gao, Qianwen Zhu, and Wei Wang. Optimal deployment of large-scale wireless sensor networks based on graph clustering and matrix factorization. *EURASIP Journal on Advances in Signal Processing*, 2023(1):1–17, 2023.

[298] Michael Pointl and Daniela Fuchs-Hanusch. Assessing the potential of lpwan communication technologies for near real-time leak detection in water distribution systems. *Sensors*, 21(1):293, 2021.

[299] Hafiz Husnain Raza Sherazi, Luigi Alfredo Grieco, Muhammad Ali Imran, and Gennaro Boggia. Energy-efficient lorawan for industry 4.0 applications. *IEEE Transactions on Industrial Informatics*, 17(2):891–902, 2020.

[300] Jonathan W Berry, Lisa Fleischer, William E Hart, Cynthia A Phillips, and Jean-Paul Watson. Sensor placement in municipal water networks. *Journal of Water Resources Planning and Management*, 131(3):237–243, 2005.

[301] Aamir Anis, Aly El Gamal, A Salman Avestimehr, and Antonio Ortega. A sampling theory perspective of graph-based semi-supervised learning. *IEEE Transactions on Information Theory*, 65(4):2322–2342, 2018.

[302] Lowell W Beineke. Characterizations of derived graphs. *Journal of Combinatorial theory*, 9(2):129–135, 1970.

[303] Dragoš Cvetkovic, Peter Rowlinson, and Slobodan Simic. *Spectral generalizations of line graphs: On graphs with least eigenvalue-2*, volume 314. Cambridge University Press, 2004.

[304] Tim S Evans and Renaud Lambiotte. Line graphs, link partitions, and overlapping communities. *Physical review E*, 80(1):016105, 2009.

[305] Elena Ceci and Sergio Barbarossa. Graph signal processing in the presence of topology uncertainties. *IEEE Transactions on signal processing*, 68:1558–1573, 2020.

[306] Francisco Aparecido Rodrigues. Network centrality: an introduction. *A mathematical modeling approach from nonlinear dynamics to complex systems*, pages 177–196, 2019.

[307] Hardy Cross. Analysis of flow in networks of conduits or conductors. Technical report, University of Illinois at Urbana Champaign, College of Engineering . . . , 1936.

[308] Nicholas J Higham. *Accuracy and stability of numerical algorithms*. SIAM, 2002.

[309] OpenWaterAnalytics. Accessed: 2023-08-28.

[310] Katherine A. Klise, Michael Bynum, Dylan Moriarty, and Regan Murray. A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study. *Environmental Modelling & Software*, 95:420–431, 2017.

[311] Jun Liu and Yu Guoping. Iterative methodology of pressure-dependent demand based on epanet for pressure-deficient water distribution analysis. *Journal of Water Resources Planning and Management*, 139:34–44, 01 2013.

[312] Modesto Pérez-Sánchez, Francisco Javier Sánchez-Romero, Helena M Ramos, and P Amparo López-Jiménez. Modeling irrigation networks for the quantification of potential energy recovering: a case study. *Water*, 8(6):234, 2016.

[313] Daniel A Crowl and Joseph F Louvar. *Chemical process safety: fundamentals with applications*. Pearson Education, 2001.

[314] GraphSmart applicated dataset. Accessed: 2023-10-28.

[315] Mukarram A. M. Almuhaya, Waheb A. Jabbar, Noorazliza Sulaiman, and Suliman Abdulmalek. A survey on lorawan technology: Recent trends, opportunities, simulation tools and future directions. *Electronics*, 11(1), 2022.

[316] Davide Magrin, Stefano Avallone, Sumit Roy, and Michele Zorzi. Validation of the ns-3 802.11ax ofdma implementation. In *Proceedings of the 2021 Workshop on Ns-3*, WNS3 '21, page 1–8, New York, NY, USA, 2021. Association for Computing Machinery.

[317] Domenico Garlisi, Ilenia Tinnirello, Giuseppe Bianchi, and Francesca Cuomo. Capture Aware Sequential Waterfilling for LoRaWAN Adaptive Data Rate. *IEEE Transactions on Wireless Communications*, 20(3):2019–2033, March 2021.

[318] Domenico Garlisi, Stefano Mangione, Fabrizio Giuliano, Daniele Croce, Giovanni Garbo, and Ilenia Tinnirello. Interference Cancellation for LoRa Gateways and Impact on Network Capacity. *IEEE Access*, 9:128133–128146, 2021.

[319] Dae-Young Kim, Seokhoon Kim, Houcine Hassan, and Jong Hyuk Park. Adaptive data rate control in low power wide area networks for long range IoT services. *Journal of Computational Science*, 22:171 – 178, 2017.

[320] Floris Van den Abeele, Jetmir Haxhibeqiri, Ingrid Moerman, and Jeroen Hoebeke. Scalability analysis of large-scale LoRaWAN networks in ns-3. *IEEE INTERNET OF THINGS JOURNAL*, 4(6):2186–2198, 2017.

[321] Brecht Reynders, Qing Wang, and Sofie Pollin. A LoRaWAN module for ns-3: implementation and evaluation. In *Proceedings of the 10th Workshop on ns-3 - WNS3 '18*, pages 61–68, Surathkal, India, 2018. ACM Press.

[322] Davide Magrin, Marco Centenaro, and Lorenzo Vangelista. Performance evaluation of LoRa networks in a smart city scenario. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–7, Paris, France, May 2017. IEEE.

[323] Roberto La Rosa, Lokman Boulebnane, Daniele Croce, Patrizia Livreri, and Ilenia Tinnirello. An energy-autonomous and maintenance-free wireless sensor

platform with lora connectivity. In *2023 12th International Conference on Renewable Energy Research and Applications (ICRERA)*, pages 461–464. IEEE, 2023.

[324] WWDR. The united nations world water development report 2021: valuing water. *UNESCO World Water Assessment Programme*, 2021.

[325] Sohini Gangopadhyay, Asish Saha, Subodh Chandra Pal, Rituparna Banerjee, Sunanda Batabyal, Nilanjan Das, Phani Bhusan Ghosh, Ayan Mondal, and Sudipto Mandal. A new methodological approach to the establishment of sustainable agricultural ecology in drought vulnerable areas of eastern india. *Ecological Informatics*, page 102013, 2023.

[326] Bui Quoc Lap, Huu Du Nguyen, Phi Thi Hang, Nguyen Quang Phi, Vinh Truong Hoang, Pham Gia Linh, Bui Thi Thanh Hang, et al. Predicting water quality index (wqi) by feature selection and machine learning: A case study of an kim hai irrigation system. *Ecological Informatics*, page 101991, 2023.

[327] JR Castel. Water use of developing citrus canopies in valencia, spain. In *Proceeding International Society Citriculture, IX Congress*, pages 223–226, 2000.

[328] P et al González-Altozano. Regulated deficit irrigation in 'clementina de nules' citrus trees. ii: Vegetative growth. *The Journal of Horticultural Science and Biotechnology*, 75(4):388–392, 2000.

[329] Stagno Consoli, F Stagno, G Roccuzzo, GL Cirelli, and F Intrigliolo. Sustainable management of limited water resources in a young orange orchard. *Agricultural water management*, 132:60–68, 2014.

[330] Fiorella Stagno, Francesco Intrigliolo, Simona Consoli, Alberto Continella, and Giancarlo Roccuzzo. Response of orange trees to deficit irrigation strategies: Effects on plant nutrition, yield, and fruit quality. *Journal of Irrigation and Drainage Engineering*, 141(10):04015014, 2015.

[331] Giovanni Rallo, Pablo González-Altozano, Juan Manzano-Juárez, and Giuseppe Provenzano. Using field measurements and fao-56 model to assess the ecophysiological response of citrus orchards under regulated deficit irrigation. *Agricultural Water Management*, 180:136–147, 2017.

[332] Àngela Puig-Sirera, Giuseppe Provenzano, Pablo González-Altozano, Diego S Intrigliolo, and Giovanni Rallo. Irrigation water saving strategies in citrus orchards: Analysis of the combined effects of timing and severity of soil water deficit. *Agricultural Water Management*, 248:106773, 2021.

[333] Jason Kelley et al. Using neural networks to estimate site-specific crop evapotranspiration with low-cost sensors. *Agronomy*, 9(2):108, 2019.

[334] LS Pereira, Paul Paredes, and Nebojsa Jovanovic. Soil water balance models for determining crop water and irrigation requirements and irrigation scheduling focusing on the fao56 method and the dual kc approach. *Agricultural water management*, 241:106357, 2020.

[335] Mohammad-Bagher Aghajanloo, Ali-Akbar Sabziparvar, and P Hosseinzadeh Talaee. Artificial neural network–genetic algorithm for estimation of crop evapotranspiration in a semi-arid region of iran. *Neural Computing and Applications*, 23(5):1387–1393, 2013.

[336] Francesco Granata. Evapotranspiration evaluation models based on machine learning algorithms—a comparative study. *Agricultural Water Management*, 217:303–315, 2019.

[337] Spencer Walls, Andrew D Binns, Jana Levison, and Scott MacRitchie. Prediction of actual evapotranspiration by artificial neural network models using data from a bowen ratio energy balance station. *Neural Computing and Applications*, 32(17):14001–14018, 2020.

[338] Sevim Seda Yamaç. Artificial intelligence methods reliably predict crop evapotranspiration with different combinations of meteorological data for sugar beet in a semiarid area. *Agricultural Water Management*, 254:106968, 2021.

[339] Sevim Seda Yamaç et al. Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. *Agricultural Water Management*, 228:105875, 2020.

[340] Ammara Talib, Ankur R Desai, Jingyi Huang, Tim J Griffis, David E Reed, and Jiquan Chen. Evaluation of prediction and forecasting models for evapotranspiration of agricultural lands in the midwest us. *Journal of Hydrology*, 600:126579, 2021.

[341] Jingran Liu, Xiankuan Meng, Yingjie Ma, and Xin Liu. Introduce canopy temperature to evaluate actual evapotranspiration of green peppers using optimized enn models. *Journal of Hydrology*, 590:125437, 2020.

[342] Dahua Tang, Yu Feng, Daozhi Gong, Weiping Hao, and Ningbo Cui. Evaluation of artificial intelligence models for actual crop evapotranspiration modeling in mulched and non-mulched maize croplands. *Computers and electronics in agriculture*, 152:375–384, 2018.

[343] Nazanin Abrishami, Ali Reza Sepaskhah, and Mohammad Hossein Shahrokhnia. Estimating wheat and maize daily evapotranspiration using artificial neural network. *Theoretical and Applied Climatology*, 135(3):945–958, 2019.

[344] Xin Han, Zheng Wei, Baozhong Zhang, Yinong Li, Taisheng Du, and He Chen. Crop evapotranspiration prediction by considering dynamic change of crop coefficient and the precipitation effect in back-propagation neural network model. *Journal of Hydrology*, 596:126104, 2021.

[345] Masoumeh Hashemi et al. Evaluation of artificial neural network and penman–monteith equation for the prediction of barley standard evapotranspiration in a semi-arid region. *Theoretical and Applied Climatology*, 139(1):275–285, 2020.

[346] Ahmed Elbeltagi, Jinsong Deng, Ke Wang, Anurag Malik, and Saman Maroufpoor. Modeling long-term dynamics of crop evapotranspiration using deep learning in a semi-arid environment. *Agricultural Water Management*, 241:106334, 2020.

[347] Lingyi Zhang, Runtao Wang, Changli Zhang, Shuwen Wang, and Tianlong Xu. Intelligent irrigation strategy based on regulated deficit theory and fuzzy control for rice in cold region. *Transactions of the Chinese Society of Agricultural Engineering*, 32(13):52–58, 2016.

[348] Luis A R. Rojas, Italo Moletto-Lobos, Fabio Corradini, Cristian Mattar, Rodrigo Fuster, and Cristián Escobar-Avaria. Determining actual evapotranspiration based on machine learning and sinusoidal approaches applied to thermal high-resolution remote sensing imagery in a semi-arid ecosystem. *Remote Sensing*, 13(20):4105, 2021.

[349] MAPA. Encuesta sobre superficies y rendimientos de cultivos en españa. *Subsecretaría de Agricultura, Pesca y Alimentación, Ministerio de Agricultura, Pesca y Alimentación, Madrid, España (2019)*, 2019.

[350] Abdellah El Hari, M Chaik, Nadira Lekouch, Azeddine Sedki, and Abderrahaman Lahrouni. Water needs in citrus fruit in a dry region of morocco. *Journal of Agriculture and Environment for International Development (JAEID)*, 104(3/4):91–99, 2010.

[351] Pierluigi Pierantozzi, Mariela Torres, Romina Bodoira, and Damián Maestri. Water relations, biochemical – physiological and yield responses of olive trees (olea europaea l. cvs. arbequina and manzanilla) under drought stress during the pre-flowering and flowering period. *Agricultural Water Management*, 125:13–25, 2013.

[352] Theresa Volschenk. Water use and irrigation management of pomegranate trees - a review. *Agricultural Water Management*, 241:106375, 2020.

[353] Francesco Granata, Rudy Gargano, and Giovanni de Marinis. Artificial intelligence based approaches to evaluate actual evapotranspiration in wetlands. *Science of The Total Environment*, 703:135653, 2020.

[354] Chen Zhang, Geping Luo, Olaf Hellwich, Chunbo Chen, Wenqiang Zhang, Mingjuan Xie, Huili He, Haiyang Shi, and Yuangang Wang. A framework for estimating actual evapotranspiration at weather stations without flux observations by combining data from modis and flux towers through a machine learning approach. *Journal of Hydrology*, 603:127047, 2021.

[355] Pradeep Hebbalaguppae Krishnashetty, Jasma Balasangameshwara, Sheshshayee Sreeman, Sujeet Desai, and Archana Bengaluru Kantharaju. Cognitive computing models for estimation of reference evapotranspiration: A review. *Cognitive Systems Research*, 70:109–116, 2021.

[356] Wang Jing, Zaher Mundher Yaseen, Shamsuddin Shahid, Mandeep Kaur Saggi, Hai Tao, Ozgur Kisi, Sinan Q Salih, Nadhir Al-Ansari, and Kwok-Wing Chau. Implementation of evolutionary computing models for reference evapotranspiration modeling: short review, assessment and possible future research directions. *Engineering applications of computational fluid mechanics*, 13(1):811–823, 2019.

[357] Xiaoshu Wang, Bing Gao, and Xu-Sheng Wang. Investigating the ability of deep learning on actual evapotranspiration estimation in the scarcely observed region. *Journal of Hydrology*, 607:127506, 2022.

[358] Josefina Mosre et al. Actual evapotranspiration estimates in arid cold regions using machine learning algorithms with in situ and remote sensing data. *Water*, 13(6):870, 2021.

[359] Xian Wang, Lei Zhong, Yaoming Ma, Yunfei Fu, Cunbo Han, Peizhen Li, Zixin Wang, and Yuting Qi. Estimation of hourly actual evapotranspiration over the tibetan plateau from multi-source data. *Atmospheric Research*, 281:106475, 2023.

[360] Zohreh Izadifar et al. Prediction of hourly actual evapotranspiration using neural networks, genetic programming, and statistical models. *Hydrological processes*, 24(23):3413–3425, 2010.

[361] Yuefeng Hao, Jongjin Baik, Hien Tran, and Minha Choi. Quantification of the effect of hydrological drivers on actual evapotranspiration using the bayesian model averaging approach for various landscapes over northeast asia. *Journal of Hydrology*, 607:127543, 2022.

[362] Gastao Lukangu, Michael J Savage, and Michael A Johnston. Use of sub-hourly soil water content measured with a frequency-domain reflectometer to schedule irrigation of cabbages. *Irrigation science*, 19(1):7–13, 1999.

[363] Frederik Schrader, Wolfgang Durner, Johann Fank, Sebastian Gebler, Thomas Pütz, Matthias Hannes, and Ute Wollschläger. Estimating precipitation and actual evapotranspiration from precision lysimeter measurements. *Procedia Environmental Sciences*, 19:543–552, 2013.

[364] Richard G Allen, Luis S Pereira, Dirk Raes, Martin Smith, et al. Fao irrigation and drainage paper no. 56. *Rome: Food and Agriculture Organization of the United Nations*, 56(97):e156, 1998.

[365] Wim GM Bastiaanssen, Massimo Menenti, RA Feddes, and AAM Holtslag. A remote sensing surface energy balance algorithm for land (sebal). 1. formulation. *Journal of hydrology*, 212:198–212, 1998.

[366] John M Norman, William P Kustas, and Karen S Humes. Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. *Agricultural and Forest Meteorology*, 77(3-4):263–293, 1995.

[367] Wei Yao, Min Han, and Shiguo Xu. Estimating the regional evapotranspiration in zhalong wetland with the two-source energy balance (tseb) model and landsat7/etm+ images. *Ecological Informatics*, 5(5):348–358, 2010.

[368] G Rallo, TA Paço, Paula Paredes, À Puig-Sirera, R Massai, G Provenzano, and LS Pereira. Updated single and dual crop coefficients for tree and vine fruit crops. *Agricultural Water Management*, 250:106645, 2021.

[369] O Alkaeed, C. Flores, K. Jinno, and A. Tsutsumi. Comparison of several reference evapotranspiration methods for itoshima peninsula area, fukuoka, japan. *Memoirs of the Faculty of Engineering, Kyushu University*, 66(1), 2006.

[370] I Pôças, A Calera, I Campos, and M Cunha. Remote sensing for estimating and mapping single and basal crop coefficientes: A review on spectral vegetation indices approaches. *Agricultural Water Management*, 233:106081, 2020.

[371] Huimin Lei and Dawen Yang. Combining the crop coefficient of winter wheat and summer maize with a remotely sensed vegetation index for estimating evapotranspiration in the north china plain. *Journal of Hydrologic Engineering*, 19(1):243–251, 2014.

[372] Daniele Masseroni, Chiara Corbari, and Marco Mancini. Limitations and improvements of the energy balance closure with reference to experimental data measured over a maize field. *Atmósfera*, 27(4):335–352, 2014.

[373] Matteo Ippolito, Dario De Caro, Giuseppe Ciraolo, Mario Minacapilli, and Giuseppe Provenzano. Estimating crop coefficients and actual evapotranspiration in citrus orchards with sporadic cover weeds based on ground and remote sensing data. *Irrigation Science*, pages 1–18, 2022.

[374] Markus Kottek, Jürgen Grieser, Christoph Beck, Bruno Rudolf, and Franz Rubel. World map of the köppen-geiger climate classification updated. *Meteorologische Zeitschrift, Vol. 15, No. 3, 259-263*, 2006.

[375] Magdalena Main-Knorn, Bringfried Pflug, Jerome Louis, Vincent Debaecker, Uwe Müller-Wilm, and Ferran Gascon. Sen2cor for sentinel-2. In *Image and Signal Processing for Remote Sensing XXIII*, volume 10427, pages 37–48. SPIE, 2017.

[376] JW Rouse, RH Haas, JA Schell, and DW Deering. Monitoring vegetation systems in the great plains with erts: Proceedings of the third earth resources technology satellite-1 symposium. *NASA SP-351*, pages 301–317, 1974.

[377] Ahmad Toosi, Farzaneh Dadrass Javan, Farhad Samadzadegan, Soroosh Mehravar, Alishir Kurban, and Hossein Azadi. Citrus orchard mapping in juybar, iran: Analysis of ndvi time series and feature fusion of multi-source satellite imageries. *Ecological Informatics*, 70:101733, 2022.

[378] Bo-Cai Gao. Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, 58(3):257–266, 1996.

[379] G Manca. Analisi dei flussi di carbonio di una cronosequenza di cerro (quercus cerris l.) dell'italia centrale attraverso la tecnica della correlazione turbolenta. *PhD Dissertation Thesis, University of Tuscia, Viterbo,*, 225 pp, 2003.

[380] JH Prueger, JL Hatfield, TB Parkin, WP Kustas, LE Hipps, CMU Neale, JI MacPherson, WE Eichinger, and DI Cooper. Tower and aircraft eddy covariance measurements of water vapor, energy, and carbon dioxide fluxes during smacex. *Journal of Hydrometeorology*, 6(6):954–960, 2005.

[381] Luigi Ranghetti, Mirco Boschetti, Francesco Nutini, and Lorenzo Busetto. "sen2r": An R toolbox for automatically downloading and preprocessing sentinel-2 satellite data. *Computers & Geosciences*, 139:104473, doi: 10.1016/j.cageo.2020.104473, 2020.

[382] Pedregosa et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[383] Git Repository. Machine learning models to predict daily actual evapotranspiration of citrus orchards. Accessed: 15/03/2023. https://github.com/fedesss98/ml-to-eta-of-citrus-orchards.git.

[384] MR Sowmya, MB Santosh Kumar, and Sooraj K Ambat. Comparison of deep neural networks for reference evapotranspiration prediction using minimal meteorological data. In *2020 advanced computing and communication Technologies for High Performance Applications (ACCTHPA)*, pages 27–33. IEEE, 2020.

[385] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[386] Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

[387] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[388] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[389] Olga Troyanskaya, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein, and Russ B Altman. Missing value estimation methods for dna microarrays. *Bioinformatics*, 17(6):520–525, 2001.

[390] S Thirukumaran and A Sumathi. Missing value imputation techniques depth survey and an imputation algorithm to improve the efficiency of imputation. In *2012 Fourth International Conference on Advanced Computing (ICoAC)*, pages 1–5. IEEE, 2012.

[391] Michael B Richman, Theodore B Trafalis, and Indra Adrianto. Missing data imputation through machine learning algorithms. In *Artificial intelligence methods in the environmental sciences*, pages 153–169. Springer, 2009.

[392] William P Kustas, John H Prueger, Karen S Humes, and Patrick J Starks. Estimation of surface heat fluxes at field scale using surface layer versus mixed-layer atmospheric variables with radiometric temperature observations. *Journal of Applied Meteorology*, 38(2):224–238, 1999.

[393] SALAH Er-Raki, Abdelghani Chehbouni, NOURA Guemouria, J Ezzahar, S Khabba, Gilles Boulet, and L Hanich. Citrus orchard evapotranspiration: comparison between eddy covariance measurements and the fao-56 approach estimates. *Plant Biosystems*, 143(1):201–208, 2009.

[394] Sowmya Mangalath Ravindran, Santosh Kumar Moorakkal Bhaskaran, and Sooraj Krishnan Nair Ambat. A deep neural network architecture to model reference evapotranspiration using a single input meteorological parameter. *Environmental Processes*, 8(4):1567–1599, Dec 2021.

[395] Akhil Kadiyala and Ashok Kumar. Applications of python to evaluate the performance of decision tree-based boosting algorithms. *Environmental Progress & Sustainable Energy*, 37(2):618–623, 2018.

[396] Kaicun Wang and Shunlin Liang. An improved method for estimating global evapotranspiration based on satellite determination of surface net radiation, vegetation index, temperature, and soil moisture. *Journal of Hydrometeorology*, 9(4):712–727, 2008.

[397] Marco Girardello, Matteo Griggio, Mark J Whittingham, and Stephen P Rushton. Models of climate associations and distributions of amphibians in italy. *Ecological research*, 25:103–111, 2010.

[398] Ritika Srinet, Subrata Nandy, and NR Patel. Estimating leaf area index and light extinction coefficient using random forest regression algorithm in a tropical moist deciduous forest, india. *Ecological Informatics*, 52:94–102, 2019.

[399] Francesco Piccioni, Céline Casenave, Meïli Baragatti, Bertrand Cloez, and Brigitte Vinçon-Leite. Calibration of a complex hydro-ecological model through approximate bayesian computation and random forest combined with sensitivity analysis. *Ecological Informatics*, 71:101764, 2022.

[400] UN DESA/POP. United nations department of economic and social affairs, population division (2022). world population prospects 2022: Summary of results. un desa/pop/2022/tr/no. 3. In *UNDESA*. Montreal, Canada, 2022.

[401] Giovanni Rallo, Pablo González-Altozano, Juan Manzano-Juárez, and Giuseppe Provenzano. Using field measurements and fao-56 model to assess the ecophysiological response of citrus orchards under regulated deficit irrigation. *Agricultural Water Management*, 180:136–147, 2017.

[402] Kaitlin Johnson, Sindhuja Sankaran, and Reza Ehsani. Identification of water stress in citrus leaves using sensing technologies. *Agronomy*, 3(4):747–756, 2013.

[403] Giovanni Piccinni, Jonghan Ko, Thomas Marek, and Terry Howell. Determination of growth-stage-specific crop coefficients (kc) of maize and sorghum. *Agricultural water management*, 96(12):1698–1704, 2009.

[404] Richard G Allen, Luis S Pereira, Dirk Raes, Martin Smith, et al. Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *Fao, Rome*, 300(9):D05109, 1998.

[405] Daniel A Segovia-Cardozo, Loris Franco, and Giuseppe Provenzano. Detecting crop water requirement indicators in irrigated agroecosystems from soil water

content profiles: An application for a citrus orchard. *Science of The Total Environment*, 806:150492, 2022.

[406] G. Longo-Minnolo, D. Vanella, S. Consoli, D.S. Intrigliolo, and J.M. Ramírez-Cuesta. Integrating forecast meteorological data into the arcdualkc model for estimating spatially distributed evapotranspiration rates of a citrus orchard. *Agricultural Water Management*, 231:105967, 2020.

[407] L. Mateos, M.P. González-Dugo, L. Testi, and F.J. Villalobos. Monitoring evapotranspiration of irrigated crops using crop coefficients derived from time series of satellite images. i. method validation. *Agricultural Water Management*, 125:81–91, 2013.

[408] X. LIU, S. YANG, J. XU, J. ZHANG, and J. LIU. Effects of soil heat storage and phase shift correction on energy balance closure of paddy fields. *Atmósfera*, 30(1):39–52, 2017.

[409] Z. XU, S. LIU, and M. HU. Measurements of evapotranspiration by eddy covariance system in the hai river basin. *In Proceedings of the 2009 International Symposium of HAIHE Basin Integrated Water and Environment Management, Beijing, China*, 2009.

[410] Dennis D. Baldocchi. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology*, 9(4):479–492, 2003.

[411] Tracy E Twine, WP Kustas, JM Norman, DR Cook, PRea Houser, TP Meyers, JH Prueger, PJ Starks, and ML Wesely. Correcting eddy-covariance flux underestimates over a grassland. *Agricultural and forest meteorology*, 103(3):279–300, 2000.

[412] JL Heilman, WE Heilman, and Donald G Moore. Evaluating the crop coefficient using spectral reflectance. *Agronomy Journal*, 74(6):967–971, 1982.

[413] Walter C Bausch and Christopher MU Neale. Crop coefficients derived from reflected canopy radiation: a concept. *Transactions of the ASAE*, 30(3):703–0709, 1987.

[414] Lee F Johnson and Thomas J Trout. Satellite ndvi assisted monitoring of vegetable crop evapotranspiration in california's san joaquin valley. *Remote Sensing*, 4(2):439–455, 2012.

[415] Alfonso Calera, Isidro Campos, Anna Osann, Guido D'Urso, and Massimo Menenti. Remote sensing for crop water management: From et modelling to services for the end users. *Sensors*, 17(5):1104, 2017.

[416] Juan Miguel Ramírez-Cuesta, José Manuel Mirás-Avalos, José Salvador Rubio-Asensio, and Diego S. Intrigliolo. A novel arcgis toolbox for estimating crop water demands by integrating the dual crop coefficient approach with multi-satellite imagery. *Water*, 11(1), 2019.

[417] Deepak Gautam, Bertram Ostendorf, and Vinay Pagay. Estimation of grapevine crop coefficient using a multispectral camera on an unmanned aerial vehicle. *Remote Sensing*, 13(13):2639, 2021.

[418] Ran Pelta, Ofer Beeri, Rom Tarshish, and Tal Shilo. Forecasting seasonal plot-specific crop coefficient (kc) protocol for processing tomato using remote sensing, meteorology, and artificial intelligence. *Precision Agriculture*, 23(6):1983–2000, 2022.

[419] Guomin Shao, Wenting Han, Huihui Zhang, Shouyang Liu, Yi Wang, Liyuan Zhang, and Xin Cui. Mapping maize crop coefficient kc using random forest algorithm based on leaf area index and uav-based multispectral vegetation indices. *Agricultural Water Management*, 252:106906, 2021.

[420] Guomin Shao, Wenting Han, Huihui Zhang, Liyuan Zhang, Yi Wang, and Yu Zhang. Prediction of maize crop coefficient from uav multisensor remote sensing using machine learning methods. *Agricultural Water Management*, 276:108064, 2023.

[421] Amro Negm, Mario Minacapilli, and Giuseppe Provenzano. Downscaling of american national aeronautics and space administration (nasa) daily air temperature in sicily, italy, and effects on crop reference evapotranspiration. *Agricultural Water Management*, 209:151–162, 2018.

[422] Dae-Hyun Jung, Hyoung Seok Kim, Changho Jhin, Hak-Jin Kim, and Soo Hyun Park. Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse. *Computers and Electronics in Agriculture*, 173:105402, 2020.

[423] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[424] Yann LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–50. Springer, 2002.

[425] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.

[426] David W Aha, Dennis Kibler, and Marc K Albert. Instance-based learning algorithms. *Machine learning*, 6:37–66, 1991.

[427] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[428] Sadegh Bafandeh Imandoust, Mohammad Bolandraftar, et al. Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *International journal of engineering research and applications*, 3(5):605–610, 2013.

[429] Joseph G Eisenhauer. Regression through the origin. *Teaching statistics*, 25(3):76–80, 2003.

[430] John B Kennedy and Adam M Neville. *Basic statistical methods for engineers and scientists*. Crowell, 1976.

[431] Oscar González-Recio, Guilherme JM Rosa, and Daniel Gianola. Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livestock Science*, 166:217–231, 2014.

[432] Stef Van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 45:1–67, 2011.

[433] Najmeddine Abdennour, Tarek Ouni, and Nader Ben Amor. The importance of signal pre-processing for machine learning: The influence of data scaling in a driver identity classification. In *2021 IEEE/ACS 18th International Conference on Computer Systems and Applications (AICCSA)*, pages 1–6. IEEE, 2021.

[434] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145, 1995.

[435] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 eighth ieee international conference on data mining*, pages 413–422. IEEE, 2008.

[436] Srinivasa Rao Peddinti and BVN P Kambhammettu. Dynamics of crop coefficients for citrus orchards of central india using water balance and eddy covariance flux partition techniques. *Agricultural Water Management*, 212:68–77, 2019.

[437] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, volume 57, pages 10–25080. Austin, TX, 2010.

[438] L Franco, A Motisi, and G Provenzano. Agro-hydrological models and field measurements to assess the water status of a citrus orchard irrigated with microsprinkler and subsurface drip systems. In *Acta Hortic. 1314*, pages 75–82, 2021.

Acknowledgements

At the conclusion of this doctoral thesis, I would like to express my gratitude to all the individuals who have supported me throughout my journey. Without their invaluable contributions, my work would not have been as comprehensive, and the path taken would have certainly been more challenging.

First and foremost, I would like to extend a heartfelt thank you to my supervisor, *Ilenia Tinnirello*, and my co-supervisors, *Daniele Croce, Gianpaolo Vitale*, and *Roberto La Rosa*, who have always been ready to guide me through every stage of thesis development. Thanks to all of you, I have expanded my knowledge and skills.

I thank the colleagues with whom I have actively collaborated over the past three years, and in particular, *Domenico, Fabrizio*, and *Federico* for their fundamental help in the various research activities conducted by our team. Along with them, I would also like to thank the remaining colleagues from the Doctoral course in ICT and Telecommunication Labs: *Gloria, Alessandra, Gennaro, Gabriele*, and *Lino*. They all made me feel like part of a family.

I wish to express my sincere gratitude to Prof. *Giuseppe Provenzano* (who sadly passed away in December 2022) for his invaluable contributions to a part of this thesis and for fostering outstanding human and professional relationships. Furthermore, I would like to thank his collaborators, *Dario and Matteo*, who embody the same vision and openness that distinguished the Professor, thus contributing to the advancement of scientific research.

I thank *Samuel Margueron and Mario Costanza* for their valuable assistance in conducting research during my time abroad at the FEMTO-ST research laboratory in

Besançon, France. I also extend my gratitude to all the members of the FEMTO-ST for welcoming me and making me feel at ease. I will always carry with me the cultural enrichment that this foreign experience has provided.

I would like to thank the entire WITS group, led by Professor *Francesca Cuomo*, who, together with colleagues from Sapienza University of Rome, involved me in interesting research activities, fostering my professional growth.

I deeply want to thank my girlfriend, *Simona*, for her constant support and love during this academic journey. Her encouragement and presence made the tough times lighter and made this research experience more meaningful. To my wonderful family, to my parents *Giuseppe* and *Rosalia*, and to my sister *Piera*, I cannot express enough gratitude for their endless support and affection. Their words of wisdom and encouragement gave me the strength to persevere and complete this work.

Finally, but not in order of importance, I would like to thank *God* for giving me the knowledge, ability, and opportunity to undertake and persevere in this journey. Without His blessings, this journey could not even have begun.