UNIVERSITÀ DEGLI STUDI
DI TRENTO

DEPARTMENT OF INFORMATION ENGINEERING AND COMPUTER SCIENCE
**ICT International Doctoral School**

# Remote sensing-based channel modeling and deployment planning for low-power wireless networks

## Silvia Demetri

Advisors

Prof. Gian Pietro Picco

Università degli Studi di Trento

Prof. Lorenzo Bruzzone

Università degli Studi di Trento

September 2018

*A mio padre*

# Abstract

The deployment of low power wireless networks is notoriously effort-demanding, as costly in-field campaigns are required to assess the connectivity properties of the target location and understand where to place the wireless nodes. The characteristics of the environment, both static (e.g., obstacles obstructing the link line of sight) and dynamic (e.g., changes in weather conditions) cause variability in the communication performance, thus affecting the network operation quality and reliability. This translates into difficulties in effectively deploy, plan and manage these networks in real-world scenarios, especially outdoor. Despite the large literature on node placement, existing approaches make over-simplifying assumptions neglecting the complexity of the radio environment.

Airborne and satellite Remote Sensing (RS) systems acquire data and images over wide areas, thus enabling one to derive information about these areas at *large scale*. In this dissertation, we propose to leverage RS systems and related data processing techniques to *i) automatically* derive the static characteristics of the deployment environment that affect low power wireless communication; *ii) model* the relation between such characteristics and the communication quality; and *iii)* exploit this knowledge to *support the deployment* planning. We focus on two main scenarios: *a)* the deployment of Wireless Sensor Networks (WSNs) in forests; and *b)* the communication performance of Internet of Things (IoT) networks based on Long Range (LoRa) wireless technology in the presence of mixed environments.

As a first major contribution, we propose a novel WSN node placement approach (LaPS) that integrates remote sensing data acquired by airborne Light Detection and Ranging (LiDAR) instruments, a specialized path loss model and evolutionary computation to identify (near-)optimal node position in forests, *automatically* and *prior* to the actual deployment. When low-power WSNs operating at 2.4 GHz are deployed in forests, the presence of trees greatly affects communication. We define a processing architecture that automatically derives local forest attributes (e.g., tree density) from LiDAR data acquired over the target forest. This information is incorporated into a specialized path loss model, which is validated in deployments in a real forest, enabling fine-grained, per-link estimates of the radio signal attenuation induced by trees. Combining the forest attributes derived from LiDAR data with the specialized path loss model and a genetic algorithm, LaPS provides node placement solutions with higher quality than approaches based on a regular placement or on a standard path loss model, while satisfying the spatial and

network requirements provided by the user. In addition, LaPS enables the exploration of the impact of changes in the user requirements on the resulting topologies in advance, thus reducing the in-field deployment effort.

Moreover, to explore a different low-power wireless technology with starkly different trade-offs, we consider a LoRa-based IoT network operating in *i)* a free space like communication environment, i.e., the LoRa signal is transmitted from an high altitude weather balloon, traverses a free-of-obstacles space and is received by gateways on the ground; and *ii)* a mixed environment that contains built-up areas, farming fields and groups of trees, with both LoRa transmitters and receiving gateways close to the ground. These scenarios show a huge gap in terms of communication range, thus revealing to which extent the presence of objects affects the coverage that LoRa gateways can provide. To characterize the mixed environment we exploit detailed land cover maps (i.e., with spatial grain $10\times10\text{m}^2$) derived by automatically classifying multispectral remote sensing satellite images. The land cover information is jointly analyzed with LoRa connectivity traces, enabling us to observe a correlation between the land cover types involved in LoRa links and the trend of the signal attenuation with the distance. This analysis opens interesting research venues aimed at defining LoRa connectivity models that *quantitatively* account for the type of environment involved in the communication by leveraging RS data.

# Acknowledgements

My PhD experience has been a challenge and a unique learning opportunity. I would like to express my sincere gratitude to all the people who have contributed in various way to its completion. First and foremost, I want to thank my family for the constant and unconditional support they have offered to me all along the way.

My sincere thanks go to both my supervisors Gian Pietro Picco and Lorenzo Bruzzone for giving me the opportunity to take part in this journey and, especially, for providing valuable feedback and advice throughout the whole period of doctoral studies, while at the same time leaving me free to experiment and explore things my own way, mistakes included. Their passion for research and their vast knowledge have been both model and inspiration. I am also grateful to have had the chance to attend an international conference and to spend a research period abroad.

Many thanks go to Marco Zúñiga, Fernando Kuipers, Koen Langendoen and to all the members of the Embedded Software group of TU Delft, for nicely hosting me as a visiting PhD student. A special acknowledgment goes to Marco for the insightful discussions, constructive suggestions and helpful advice he has offered during this period, and after.

I am also grateful to all the members of the RSLab and of the D3S group, as well as to all the colleagues and friends who have given me suggestions, help and encouragement so many times. The nice times I have shared with them, including a number of coffee breaks, lunches and adventures in the mountains running experiments, have been a precious gift. Thank you Claudia, Andrea, Massimo, Leonardo, Francesca, Daniele, Davide, Ramona, Giuliano, Timofei, Dmitrii, Mattia, Francesco, Sara, Belma, Eric, Stef and all the people who have shown their support and friendship in many ways during these years.

This thesis is dedicated to my father, for the invaluable lessons he taught me by example and with love.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Remarkable advancements in sensing technologies led to the current availability of a variety of forefront monitoring instruments ranging from distributed networks of in-situ sensors to remote sensing systems for Earth observation. Both of them are applied to an ever-increasing number of domains as environmental monitoring [135, 145, 155, 97, 51, 133], disaster management [49, 67, 146, 48], surveillance [148, 66, 34, 94], and smart spaces [41, 56]. On the one hand, distributed networks of embedded devices as Wireless Sensor Networks (WSNs) perform *direct* and *in-situ* measurements of physical parameters (e.g., temperature) while exploiting low-power wireless communication to self-organize as a network and accomplish cooperative data collection tasks. On the other hand, Remote Sensing (RS) satellite/airborne systems acquire data (mainly images) over wide areas performing *remote* and *indirect* measurements of the properties of objects, areas or phenomena, by exploiting the propagation and reflection properties of electromagnetic radiation. In this dissertation we focus on the challenges posed by real-world outdoor environments on the deployment of low-power networks (e.g., WSNs) and exploit Remote Sensing systems and data to model low-power communication performance in outdoor scenarios and support the deployment task.

The physical layer at the base of the protocol stack of WSNs is defined by *low-power wireless* communication techniques operating in the unlicensed ISM bands. These techniques are key in rendering the system flexible (i.e., no wired connection is required) and able to operate for relatively long periods of time (low power consumption). Moreover, the wireless communication itself is often exploited as a source of information in applications as localization [12], target tracking [148] or wildlife monitoring [121]. Low-power wireless technology is indeed a constant presence in the evolution of distributed and pervasive embedded systems, including the emerging contexts of Cyber-Physical-Systems (CPS) and Internet of Things (IoT). In fact, the physical input/output in CPS, and the "things" in IoT, are connected to the Internet mostly via low-power wireless communi-

cation, both short-range (e.g., as defined by the IEEE 802.15.4 standard) and long-range (e.g., LoRa™ in the LoRaWAN specification). However, low-power wireless communication carries also some criticality, in that the link behavior is significantly affected by the surrounding environment, thus rendering the effective deployment of low-power networks extremely challenging.

**Motivation.** Almost two decades of developments in the context of WSNs enabled their successful application in a number of real world deployments [145, 22, 96, 154, 92]. However, real-world deployments often reveal difficulties in effectively deploying and managing such *low-power networks*, especially in outdoor scenarios [155, 87, 157, 57]. This is mostly due to the variability and unpredictability of the link behavior. Due to the link unpredictability costly trial-and-error pilot campaigns are required to understand where and how to place the wireless nodes, in order to achieve a satisfactory deployment.

The variability of short-range low-power links in real-world deployments has been extensively observed [153, 142, 154, 92, 105, 125] and significant effort has been devoted to its analysis [160, 161, 140, 43, 159, 88, 128, 117, 15]. This variability is imputable to several reasons, mostly related to the fact that the communication quality is strongly affected by the surrounding environment [69, 99, 144]. More precisely, the communication performance is affected by both *static* and *dynamic* components. The static components are the communication distance *and* the fixed characteristics of the deployment environment, namely the presence and position of objects with size comparable to the radio wavelength. These objects may obstruct the radio propagation and/or cause reflections, which in turn can constructively or disruptively affect the line-of-sight signal (i.e., multipath effect). The dynamic components are due to variations in the weather conditions (e.g., rain and temperature) [26, 27, 28, 18, 37, 9], multipath effects caused by the presence of moving reflectors or screens (e.g., human body), and interference due to concurrent transmissions [25, 16]. The static components determine a baseline on the communication performance, on top of which the dynamic components act, i.e., given a strong and a weak link in static conditions, the former is likely stronger than the latter also in the presence of comparable dynamics. Therefore, understanding and modeling the static components (i.e., the focus of this dissertation) can significantly support the planning of outdoor deployments, which still remains a challenging task to accomplish in practice.

In particular, the effectiveness of the device positioning, in terms of achievable communication quality, is largely determined by the static elements that are present in the environment. For instance, the deployment of a WSN in a forest (e.g., for a forest fire detection application) is affected by the presence of trees, their position and their structural characteristics (e.g., trunk diameter). Similarly, LoRa long-range communication perfor-

mance are different in a build-up urban area and in open farming fields. The deployment, to be effective, should be tailored to the target environment. In this respect, both tools for the in-field connectivity assessment [74, 55, 141] and deployment approaches based on pilot connectivity measurements collected on-site [85, 129] are available. However, their application requires costly in-field effort. Other planning approaches are aimed at identifying a priori optimal device placements [83, 23], but they mostly rely on oversimplified radio channel models and, therefore, perform poorly in real-world applications. We survey the related work on the challenges posed by specific types of environments on the different low-power wireless technologies we consider in this thesis in the correspondent chapters. In general, low-power wireless models accounting for the static and locally specific characteristics of outdoor deployment environments are lacking.

**Combining remote sensing and low-power networks.** In this dissertation, we propose to fill this gap exploiting Remote Sensing (RS) systems and related data processing techniques. In particular, RS data can provide *detailed* and *large-scale* information about the static characteristics of the target deployment area, as for instance position and density of trees in a forest and/or fine-grained maps representing the land cover type (e.g., buildings or farming fields). These characteristics can be extracted and exploited in an *automatic* fashion, thus enabling a systematic analysis aimed at *i) understanding* and *modeling* the impact of such characteristics on the communication performance; and *ii)* developing effective *network planning strategies*. RS data are therefore intended by us as a mean to derive prior knowledge about both the characteristics of the target deployment environment and their effect on the communication. This prior knowledge may provide insights on the reliability of the deployment, as well as models for simulation and guidelines to support it.

Different RS systems exist, providing different types of data and information. Passive RS systems (e.g., multispectral and hyperspectral sensors) exploit the sun as source of radiation that illuminates the scene under investigation and detect and measure the radiation that is reflected by the objects in the scene in different spectral channels. By processing the images generated by passive systems, information about the composition of materials and the horizontal structure of the scene can be retrieved (e.g., land cover classes can be identified by analyzing their spectral signature, up to the classification of different species of trees [45, 53]). Active RS systems, e.g., Light Detection and ranging (LiDAR) and Radio Detection and Ranging (RADAR), generate the source radiation themselves and capture the geometrical characteristics of the scene, i.e., horizontal and vertical structure, as well as the backscattering properties. Airborne LiDAR, for instance, provides detailed 3D information about the structure of forests and their geometry [97, 116, 17, 68, 111], allowing the estimation of tree metrics at the level of both single

tree (e.g., trunk diameter [68, 118]) as well as group of trees (e.g., tree density [108, 111]).

RS systems and WSNs have a huge potential for interplay. Indeed, they are often applied to similar applications (e.g., environmental monitoring and surveillance). In addition, their fusion has been investigated in several studies (e.g., [114, 137]), which, however, mainly focus on combining data and/or information provided by the two technologies about the investigated phenomena (e.g., in-situ measurements of physical parameters and information derived from RS data). In contrast, their combination for the modeling of low-power wireless communication w.r.t. the environment characteristics, which is the subject of this thesis, has been only marginally explored [78, 6, 113].

**Goal and contributions.** This dissertation explores the potential of RS systems and data to support the deployment of low-power networks in outdoor environments. The general goal is:

> *leveraging RS data to model low-power wireless connectivity as a function of local characteristics of the deployment environment, estimate those characteristics within an automatic and scalable framework, and develop effective strategies for the deployment planning.*

More specifically, the main contributions of this dissertation can be summarized as:

1. the definition of a **conceptual framework** aimed at exploiting RS data to assess a priori the local characteristics of a target deployment environment, analyze, model and predict their impact on low-power wireless communication over possibly large and differing target areas, and define effective deployment strategies by enabling one to explore the connectivity properties in an area of interest without having to run pilot campaigns;

2. the definition of a **specialized radio attenuation model** for low-power short range communication (i.e., IEEE 802.15.4 at 2.4 GHz) in forest environments exploiting a LiDAR based representation of the forest structure;

3. the development of an **automatic node placement approach** for WSNs in forest environments combining a LiDAR-based forest representation, a specialized radio model and an evolutionary optimization algorithm to derive effective node placement, prior to the deployment, accounting for both the structure of the target forest and the user requirements;

4. the joint **analysis** of LoRa communication traces and land cover maps derived from satellite multispectral images aimed at deriving LoRa connectivity models that account for the characteristics of the communication environment (e.g., presence of buildings and farming fields along the communication path).
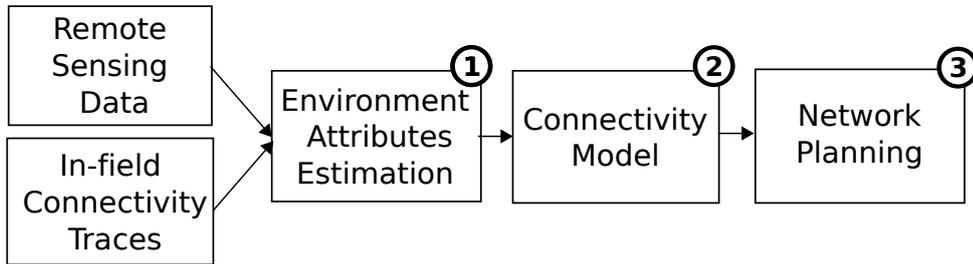
Figure 1.1: From RS data and in-field connectivity traces to deployment planning.

The proposed **conceptual framework** is depicted in Figure 1.1, which summarizes the three main successive steps leading to the achievement of the goal. We aim at exploiting RS data as the base of this framework, leveraging their wide coverage and detail. The research field of remote sensing systems and data processing is rather broad, we therefore provide an high level overview of the RS systems that we use in this context in Chapter 2, while the specific processing we apply for information extraction is detailed in the following chapters. By processing the raw RS data (e.g., LiDAR data and multispectral images) we can automatically estimate the attributes that characterize the deployment environment, e.g., tree density and land cover types (step 1). Such attributes can then be compared and analyzed together with real connectivity traces collected in the field, so as to enable a systematic evaluation of their effects on links leading to understand their role and importance in affecting communication performance (i.e., identifying which attributes mainly affect low-power wireless links, to which extent, and how). The attributes that result more significant in this perspective are then combined and exploited to define and calibrate a radio propagation model, thus allowing the estimation of the expected communication performance at specific locations (step 2). This model is then applied to predict the link quality, according to a given placement of transmitters and receivers, or to identify a favorable node placement among those possible in the target site, and, more in general, to develop RS-based tools and strategies for deployment planning (step 3).

This process is entirely applied (from step 1 to step 3) to static WSN deployments in forest environments. In this case, short range low-power communication at 2.4 GHz, according to the IEEE 802.15.4 standard, is considered, with TMote Sky as reference hardware platform. As we describe in Chapter 3, we leverage airborne Light Detection and Ranging (LiDAR) instruments and related automatic data analysis systems to determine local forest attributes (e.g., tree density) that, once factored into a **specialized radio path loss model**, enable accurate estimation of the received signal power. The approach, which is validated on connectivity traces collected in the field, is *automatic*, i.e., it does not require in-field campaigns, and *fine-grained*, i.e., it enables per-link estimates.

Then, we develop an **automatic node placement approach**, called LaPS, to identify

effective WSN node placements in the forest accounting for *i)* the forest characteristics; *ii)* the related effect on communication; and *iii)* the user requirements. Exploiting in combination a LiDAR-based representation of the forest structure, the specialized radio path loss model and an evolutionary optimization method, LaPS provides placement solutions automatically and prior to the actual deployment. Specifically, LaPS exploits genetic algorithms to efficiently explore the space of possible placement solutions that satisfy the requirements, evolving towards an optimized placement in the target forest area. This is done by evaluating the fitness of placement configurations in terms of quality and number of the communication links available to network nodes, while honoring the spatial and network requirements specified by the user. LaPS and the companion toolchain is presented in Chapter 4.

In Chapter 5, we consider LoRa low-power long-range wireless technology and, unlike 802.15.4, we implement only the first steps of the conceptual framework, namely step 1 and partly step 2, taking advantage of existent radio path loss models. In particular, we target the **analysis** of end-device to gateway communication in a mixed environment, which presents build-up areas, farming fields and groups of trees. Multispectral images, provided by the Sentinel-2 satellite constellation, are automatically processed to derive information about the land-cover types that characterize such a diversified environment, i.e., maps representing the land-cover classes (e.g. buildings, trees or farming fields) with 10 m spatial resolution. The land cover information is analyzed in combination with a set of connectivity traces collected outdoor, through controlled experiments in The Things Network (TTN), i.e., a LoRa-based open IoT network. Several insights about the impact of the land covers on the communication performance are derived, which constitute the premise for the definition of land-cover aware connectivity models for LoRa.

Finally, we conclude and explore possible venues for future research in Chapter 6.

We argue that the contribution of this thesis may significantly help the definition of realistic low-power connectivity models for outdoor environments and support the development of effective deployment strategies.

# Chapter 2

# Remote Sensing in a Nutshell

Remote Sensing (RS) systems measure the properties of surfaces and objects without coming in direct contact with them. This is achieved by illuminating the scene under investigation with a source of electromagnetic radiation and measuring, through sensors, the radiation that is reflected by the objects in the scene. Sensors are usually mounted on satellites, airplanes or UAVs, therefore acquiring data (mostly images) over wide areas. By processing RS data many properties of the reflecting objects can be retrieved at a *large scale* and *automatically*, and therefore with a limited cost in comparison to ground campaigns. Different properties can be derived, depending on the type of system and its specific characteristics [91, 34, 134]. Passive systems (i.e., optical systems) mostly use the sun as source of radiation and they capture the spectral response of the objects (spectral signature) in different spectral bands. The spectral response depends on the physical properties of the objects (e.g., their material). Active systems as Light Detection and Ranging (LiDAR) generate the source radiation themselves and mainly capture the geometry of the scene under investigation, as well as the backscattering properties (in terms of Laser Cross Section).

RS systems and data are exploited for a number of applications including forest analysis and land cover mapping, i.e., the application domains we focus on in this dissertation. In particular, we leverage airborne LiDAR data for forest analysis, in that it enables high-precision estimation of tree and forest features [97, 116, 139, 68, 13, 82], and multispectral satellite images for land-cover mapping, since by automatically classifying multispectral images fine-grained land-cover maps can be derived [71, 103, 106]. We concisely summarize the main characteristics of LiDAR and passive optical systems, and then motivate their selection in the context of this work.

**LiDAR systems.** LiDAR is an active system where the source of radiation (i.e., laser pulses) is generated by the system itself. A laser scanner, usually mounted on an airplane, transmits pulses with a nadir-looking geometry (i.e., towards the scene under investigation). Each pulse hits objects (e.g., trees) at different heights during its propagation, generating a reflection (called *return*) at every hit, that backpropagates towards the scanner.
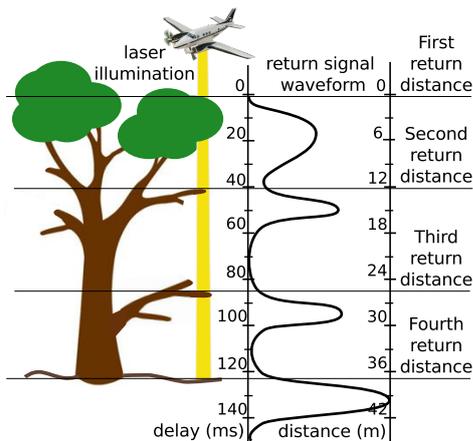


Figure 2.1: Airborne LiDAR acquisition mechanism.

The scanner measures the time elapsed between the transmission of the pulse and the reception of its reflected component, as well as the intensity of the reflection. The time interval is converted into distance, yielding a 3D cloud of point measurements (see Figure 2.2). This mechanism, shown in Figure 2.1, enables the 3D reconstruction of the trees in a forest. The representation can be very precise in both the horizontal and vertical direction, thus enabling the retrieval of information about shape, size, and position of trees, whose precision is strictly dependent on the spatial density of the emitted laser pulses in the data acquisition phase.

LiDAR data have been extensively used for the estimation of forest structural parameters both at stand level (i.e., by considering groups of trees) [72, 108] and single-tree level [68, 82] (e.g., tree height or trunk diameter at breast height). The most accurate estimates of forest parameters are achieved with high-density LiDAR data (i.e., $>5$ points/m$^2$). An example of high density LiDAR data is shown in Figure 2.2.
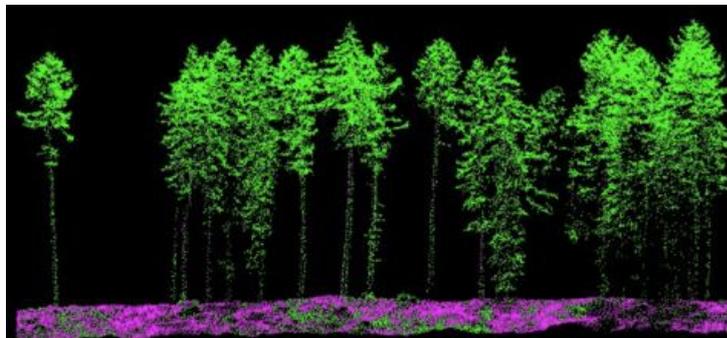


Figure 2.2: Sample of high density LiDAR data acquired in Trentino, Italy.

**Passive systems.** Passive systems (e.g., multispectral and hyperspectral scanners) are passive sensors that exploit the sun as source of electromagnetic radiation. They measure the radiation that is reflected by the surface under investigation in different spectral bands. Objects with different physical properties have different reflection characteristics in the different bands (i.e., different spectral signatures). As a consequence they can be detected and identified by analyzing the spectral behavior of their reflections. Multispectral systems measure the reflected radiation in 3 to about 13 spectral bands and such measurement is represented in multispectral images (see Figure 2.3). In contrast, hyperspectral systems measure the reflection in hundreds of narrow bands providing very high spectral resolution (i.e., very precise spectral signatures) at the cost of reduced spatial resolution. For this reason, in this work we focus on high spatial resolution multispectral images, which can be acquired over wide areas at a low cost. In particular, the Sentinel-2 (S2) satellite constellation provides multispectral global coverage with very high revisit frequency (i.e., 5 days at the equator) and high spatial resolution (i.e., 10 m), with the Copernicus open access platform making available the images for free. An example of Sentinel-2 multispectral image is shown in Figure 2.4. In particular, Figure 2.4a represents the true color composition of the bands Red, Green and Blue and Figure 2.4b represents the false color composition of the bands NearInfraRed, Red and Green. Their comparison shows how different materials and types of objects display different reflection properties in the different spectral channels. Multispectral satellite images are widely applied to large-scale monitoring of the Earth surface and they are exploited for a variety of services as land-cover and land-use mapping [71, 103, 51] and detection of changes [94, 119, 91].
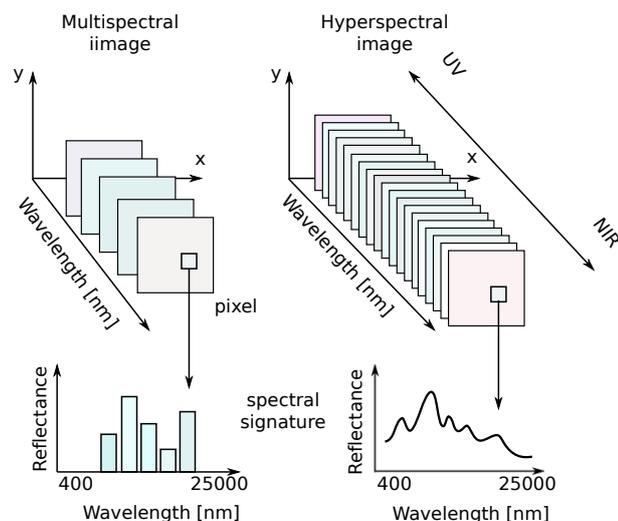


Figure 2.3: Multispectral and hyperspectral imaging.

9

Figure 2.4: Sentinel-2 multispectral image acquired over The Netherlands. True color composition of the bands Red, Green and Blue (a) and false color composition of the bands NearInfraRed, Red and Green.

**Selection of Remote Sensing data.** In this work, we aim at leveraging RS data to model: *i)* the effect of the structural characteristics of forests on low-power short range wireless communication; and *ii)* the effect of mixed environments (e.g., urban and rural) on low-power long range LoRa communication.

In the context of forest analysis, the RS techniques most commonly used are passive optical systems and active LiDAR systems [97, 116, 139, 68, 13, 82]. Optical systems (either multispectral or hyperspectral) represent mostly the horizontal structure of forests, and are therefore suited for the identification of forest areas and for the classification of tree species. In contrast, airborne LiDAR provides detailed three-dimensional information about the forest structure [97, 116, 68, 82], thus enabling the accurate estimation of the structural parameters of trees. For this reason, we focus on LiDAR only to characterize the forest environment. However, optical instruments, despite providing little information about the vertical structure of the forest, can be an alternative to LiDAR data, since their lower precision is compensated by a lower cost. their free availability. Further, the two technologies can be seen as complementary. The spectral information provided by passive systems, depending on the spatial and spectral resolution, may allow us to identify forest areas, classify tree species, or quantify the amount of green leaves in the canopy [53, 45]. The work in [115] explores the joint use of the two technologies, studying the tradeoffs between cost of data and accuracy of the forest parameter estimation. In this work we consider only LiDAR due to the high level of detail it provides w.r.t. forest structural description (both horizontal and vertical); the possibilities opened by its integration with optical data is part of our planned future work.

To characterize the impact of the environment on LoRa communication, we take into account that LoRa communication range is in the order of kilometers. Therefore, the communication links traverse different types of area (e.g., built-up areas, farming fields and groups of trees), which affect differently the communication performance. In this context, the availability of Sentinel-2 multispectral images at global scale and every few days constitutes an asset, in that by classifying multispectral images we can automatically identify different land-cover classes (e.g., Building, Trees and Fields), exploiting their different spectral response [71, 103, 131, 134, 34]. Accurate land cover maps can be derived to represent the type of environment which is traversed by LoRa communication links, thus characterizing with high spatial detail (i.e., with spatial granularity $10{\times}10$ m$^2$) the diversified LoRa communication environment.

# Chapter 3

# LiDAR to Estimate Signal Attenuation in Forests

The last decade witnessed a surge of applications where low-power wireless communication is employed to monitor natural ecosystems. In most cases, wireless sensor networks (WSNs) are deployed to either study a given environment [22] or habitat [96], or to protect it by detecting hazards such as forest fires [67]. Recently, however, low-power wireless has been employed also as a means to detect proximity among the wildlife that dwells in the natural environment [123, 121].

To these scenarios, low-power wireless communication brings unprecedented flexibility. However, it also brings a great deal of complication, due to the fact that its performance, critical to system operation, are strictly dependent on the specific environment. For instance, in the former case of a WSN with fixed nodes, its planning (number and position of nodes) must take into account the target environment and its effect on network performance. However, a similar issue arises also in the aforementioned proximity detection studies where the animal-borne low-power wireless nodes act as a "proximity sensor", recording beacons exchanged among nodes. Again, the distance at which detection may occur strongly depends on the environment, and changes based on the movement patterns of the animal; knowledge about its impact is key to enable a correct scientific inference from proximity data.

These problems are exacerbated in the forest environment, our focus in this chapter, where the heterogeneous vegetation creates local effects that amplify complexity. This long-standing issue is a barrier to environmental scientists willing to adopt WSN

technology, leaving them with the sole option of engaging in effort-demanding pilot deployments [62, 99, 92].

Indeed, despite the large literature on radio propagation in the presence of vegetation, the problem of deriving accurate estimates for low-power wireless is still largely unsolved. The majority of existing approaches rely on distance as the main parameter affecting performance, implicitly assuming a homogeneous environment in the target area. However, it is well-known that this assumption does not hold in general for low-power wireless [140]. In the case of forests, vegetation and its heterogeneity (in species and/or density) are the dominant factor determining performance, as confirmed empirically by a few studies. Marfievici et al. [99] report significant differences among species, assessed by directly immersing the same WSN deployment (i.e., same nodes and topology) in different forests. Liu et al. [92] observe, in a single large-scale deployment, that despite the regular placement of nodes, the irregularity of the forest makes the radio signal propagation irregular, and the network behavior largely unpredictable.

These aspects are only partly accounted for in existing radio propagation models tailored to forests, which carry two main limitations:

1. they assume *a priori* knowledge of vegetation attributes, whose estimation process remains often undefined, and in practice is commonly performed by means of costly in-field observations—therefore, the problem is simply moved to a different layer;

2. the estimation of vegetation attributes is typically coarse-grained, spanning relatively large areas; in principle, instead, a fine-grained estimate on a per-link basis would be desirable, given the high level of variation observed in the forest environment.

In this chapter, we propose a solution to both problems based on *remote sensing*, a technology already applied to forest monitoring. Remote sensing systems, typically satellite-based or airborne, acquire data (e.g., images) over wide areas and, through signal processing techniques, enable *fine-grained, automatic* estimation of tree and forest attributes. Here, we focus on a specific technology, Light Detection and Ranging (LiDAR), whose characteristics we concisely summarized in Chapter 2. The use of LiDAR data allows us to:

1. estimate forest attributes through *automatic* data analysis, therefore removing the need for in-field campaigns;

2. derive estimates that are very precise, to the point of identifying, e.g., the position of each individual tree and its diameter. We exploit this rich information to enable fine-grained estimates on a *per-link* basis, i.e., accounting for the presence of trees on each link.

In Section 3.1, we concisely survey the state of the art on radio propagation models accounting for the presence of vegetation. Our approach is described in Section 3.2. The

starting point is an existing path loss model [14], itself an extension of the common log-normal one, that *i)* takes explicitly into account the impact of trees on communication, and *ii)* assumes that the latter occurs at trunk level—a common choice in several of the aforementioned WSN deployment scenarios. The model in [14] relies on a *vegetation index (VD)* that depends on the *average* tree density and diameter throughout the area of interest, that are assumed to be known a priori. Our first contribution is the definition of a processing chain that, based on LiDAR data, *automatically* and accurately estimates these forest attributes. However, we show that, based on these attributes, significantly better estimates of the received signal power can be achieved. Instead of averaging these attributes over a macro-area, we perform a fine-grained analysis where the impact of trees is ascertained only *around each individual link*, therefore enabling more accurate, per-link estimates. In particular, this allows us to determine whether a link enjoys a clean line of sight, and therefore a free space path loss model is a better fit, or instead it is obstructed by trees.

The accuracy of our approach in general, and of per-link estimates in particular, is evaluated in Section 3.3 based on small-scale WSN deployments where we gather RSSI (Received Signal Strength Indicator) traces in an area where both LiDAR data and human-derived ground truth are available. Our results show that the accuracy we obtain with our per-link approach is significantly superior to existing approaches, including the model in [14], automatically tuned with our LiDAR-based approach. Specifically, accuracy is largely within $\pm 6$ dBm—the accuracy of RSSI readings from the radio transceiver.

## 3.1 Related Work

The potential of low-power wireless communications to enable unobtrusive and dense monitoring led to a number of real-world deployments in natural scenarios [22, 96, 121]. However, the connectivity assessment before deploying the network proved to be extremely difficult in harsh environments, with both pre-deployment network planning and data interpretation practices lacking of proper support. Although the node placement problem has been addressed by the WSN community [158, 84, 6], the proposed approaches are often based on non-realistic assumptions (e.g., isotropic communication range), rarely satisfied in real outdoor settings. Several studies report about the experimental evidence of the environment effect on WSN links [99, 92, 154, 40], including the impact of trees on communication performance when forests are considered. This experimental evidence, together with the lack of methods for a priori connectivity assessment suitable in forest environments, motivate our investigation.

Over the last decades, several radio propagation models accounting for the presence of vegetation have been presented. The theoretical approach at the base of mechanistics

models [130] involves the solution of Maxwell's equations with boundary conditions for each source of scattering along the propagation path. This approach is complex and often not applicable in practice. Therefore, several approximate and simpler empirical models, including the Weissberger [152] and COST 235 [101] models, which we describe in Section 3.3.4, have been developed. These models, however, are developed for scenarios in which communication links are distant from the ground and traverse the canopy. Therefore, they are not appropriate for the aforementioned common WSN applications, which rely on links that are closer to the ground and traverse the forest at the level of its trunks, as we show in Section 3.3.4.

Another conventional approach is based on the log-normal path loss model [126]. As we discuss in Section 3.2.1, the critical aspect in its application is the estimation of its parameters, whose values are strictly dependent on the specific environment being considered. These parameters are usually determined empirically by regression analysis of in-field measurements. As a consequence, results are site-specific and suitable only for environments very similar to those where measurements were performed [143]. This approach has been applied for instance in [58], which specifically focused on the ISM radio bands used by WSNs, and considers propagation paths relatively near to the ground, mainly affected by trunks. The authors explicitly consider situations where trees are obstructing the line of sight between transmitter and receiver nodes, deriving distinct models for different obstruction configurations. Nevertheless, these models still lack generality, being based on regressions from location-specific measurements.

To the best of our knowledge only two works [6, 78] mention explicitly the use of remote sensing in support to WSN deployments. In [6], LiDAR is mentioned as a source of information to characterize forests and trees. However, the whereabouts of information extraction are entirely neglected; LiDAR is simply one of the possible inputs to the 3D grid-based algorithm for deploying relay nodes, which is the focus of the paper. A simplistic radio model is considered for simulations, where nodes have a fixed and isotropic transmission range that is assumed to depend on the average tree height over the area of interest.

The analysis presented in [78], instead, is closely related to our work since it focuses on RS-based techniques for path loss prediction. The authors investigate the relationship between the path loss exponent $n$ and vegetation indexes derived from Landsat 8 satellite multispectral images. These indexes (e.g., the Normalized Difference Vegetation Index, NDVI) are "greenness indicators" denoting the amount of live green vegetation. The "green" part of plants absorbs the solar radiation in the visible (red) spectral bands and reflects it in the near-infrared band; NDVI is computed as the normalized difference between the two. The spatial granularity of the analysis is inherently determined by the

geometrical resolution of the images used, which represents the side of the ground area covered by each image pixel. This resolution is 30 m for Landsat 8—a far cry from the 50 cm used in our work. The work also partly uses a dataset estimated from even lower-resolution (250 m) MODIS images. The correlation between NDVI and path loss exponent is obtained by regression from RSSI measurements collected in a WSN deployment in an aspen boreal forest. However, this correlation appears to hold only when trees are in-leaf. Arguably, during the out-of-leaf period the amount of green leaves of the canopy is no longer the main factor affecting the communication, especially in a site characterized by deciduous trees and dense understory. In these conditions other factors predominate, which cannot be captured by the proposed greenness indicators.

## 3.2 Approach

Several models for radio propagation through vegetation have been proposed, both mechanistic [130] and empirical [101, 152, 143, 14, 58], as discussed in Section 3.1. The work we describe here is based on the empirical model in [14], as it shares our focus on communication at the trunk level. This model is an extension of the *log-normal path loss* model, where its parameters are expressed as a function of local forest attributes, based on a large set of measurements in forests with different species. However, this model (and most of the literature) assumes that vegetation is uniform in the area where radio attenuation must be estimated—an overly optimistic assumption that leads to imprecise estimates of the connectivity among individual links.

In the rest of this section we illustrate our remote sensing approach, based on LiDAR data and signal processing techniques. We overcome the two limitations above by:

1. estimating the model coefficients via *automatic* forest attribute extraction, which can be reliably performed over large areas without in-field campaigns;
2. providing fine-grained, per-link analysis of radio signal attenuation, greatly improving the overall accuracy of the resulting estimates, as shown quantitatively in Section 3.3.

### 3.2.1 Radio Signal Propagation in a Forest

The widely adopted *log-normal path loss* model [126] is defined in Equation (3.1) and

$$PL[dB] = PL(d_0) + 10 \cdot n \cdot log\left(\frac{d}{d_0}\right) + X_\sigma \tag{3.1}$$

describes both the logarithmic decay of the average signal power as a function of the distance $d$ from a transmitter and the random variation of the received power around the average. The level of attenuation in the signal power is modeled as a function of three main parameters: *i)* $PL(d_0)$, the path loss at a known reference distance $d_0$ in the

far field; *ii)* $n$, the path loss exponent representing the attenuation rate w.r.t. distance; *iii)* $\sigma$, the standard deviation of a zero-mean Gaussian random variable $X$ representing the variation around the average.

These parameters are strictly dependent on the environment at hand. Their value is difficult to determine in general due to the wide variability of the characteristics of the environment itself. In our case, trees are known to cause scattering, reflection, diffraction and absorption phenomena, depending on the radio frequency, the size of trunks, branches and leaves, and the path interception materials. Therefore, these parameters are usually estimated empirically, by curve fitting on large amounts of measurements collected in the target location [58].

Azevedo et al. [14] observed a linear relation between the path loss model parameters and the local vegetation characteristics. For instance, for our frequency of interest $f = 2.4$ GHz, the parameters of the log-normal path loss model take the form:

$$
\begin{aligned}
PL(d_0) &= -0.82 \cdot VD + 40.1 \\
n &= 0.1717 \cdot VD + 2.2043 \\
\sigma &= 4.4
\end{aligned}
\tag{3.2}
$$

where path loss parameters are expressed as a linear function of a *vegetation index VD* defined *exclusively as a function of detailed local vegetation attributes*. Indeed, *VD* is defined as

$$
VD = TD \cdot D
\tag{3.3}
$$

where $TD$ [trees/m$^2$] is the average density of trees, and $D$ [cm] is the average diameter of their trunks.

This is a significant advancement w.r.t. the problem of predicting the impact of vegetation on wireless communication. However, one key piece of information enabling the *practical* application of this model is missing in [14]: how to estimate the vegetation index *VD* reliably and cheaply. This is precisely what we address next, as one of the contributions of our work.

### 3.2.2 Determining Forest Attributes with LiDAR

We estimate the average diameter $D$ and density $TD$ of trees in an automatic fashion by processing high-density raw LiDAR data. The latter are actually LiDAR returns, which can be visualized as three-dimensional point clouds. An example is shown in Figure 3.2a, where each point represents the height at which the forest trees or ground were hit by the laser pulse.

These raw LiDAR data points are processed automatically, yielding a map in which each individual tree is represented, along with its attributes of position and trunk diam-

eter. The processing, shown in Figure 3.1, unfolds through three main phases, described next:

1. *pre-processing*: it extracts a digital Canopy Height Model (CHM), i.e., a high resolution raster dataset that maps the tree height as a discrete surface;

2. *single-tree identification*: the crown of each tree is determined by properly segmenting the CHM raster image;

3. *estimation of forest attributes*: the position and dendrometric attributes (i.e., height, crown radius, and trunk diameter) of each tree are determined.

**Preprocessing.** The LiDAR raw data are pre-processed in two steps. First, the Digital Elevation Model (DEM) is subtracted from each raw data point. The DEM represents the terrain and its morphology, mapping each pixel to the height of the terrain, with a precision that depends on the resolution of the technique used. The DEM is derived by the LiDAR data according to a standard technique [13]. This operation allows us to adjust and correct the raw data by extracting the actual elevation from the ground of each point. Second, from these corrected data points we generate a raster image with a geometrical resolution of 50 cm (i.e., each pixel represent a $50 \times 50$ cm$^2$ ground area) by assigning to each pixel the maximum height value of the points (as obtained from the previous step) belonging to the corresponding area. The output is the CHM raster image.

**Single-tree identification.** The next step is to detect each tree in the site under investigation, and to delineate its crown. To this end we apply a segmentation technique similar to [72]. We first perform a convolutional prefiltering on the CHM, to emphasize local maxima and tree crowns. Then, we apply the set level method [82] to detect local peaks and identify the tree tops. The latter represent the seeds we use to initialize a segmentation procedure based on the region growing technique. In each step, the region around each seed is expanded by including all the neighboring pixels. The process is iterated as long as the canopy height value of those pixels is higher than a predefined threshold (i.e., the height is decreased less than 80% w.r.t. the seed) and the region diameter does not exceed a maximum acceptable value, 15 m in our case.

The output of this processing stage is a *tree map* representing the dominant layer of
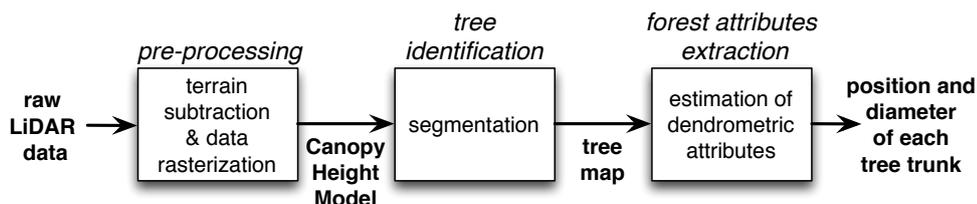


Figure 3.1: Extracting tree attributes from raw LiDAR data.

19
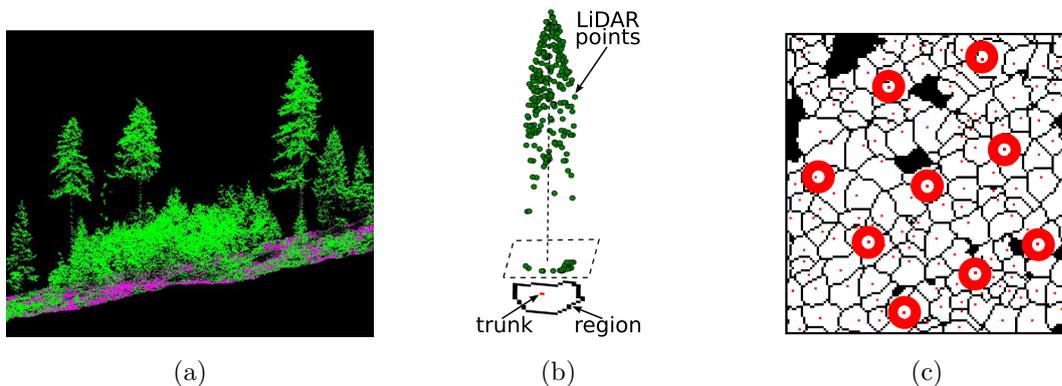
Figure 3.2: (a) High-density LiDAR data sample (b) LiDAR points corresponding to a single tree (c) Tree map example

the forest with a 50 cm spatial resolution. Each tree crown in our scene is delineated and the position of the trunk, corresponding to the position of the tree top, is estimated. An example showing the relation between the source raw LiDAR data and the resulting region delineating the crown is shown at the bottom of Figure 3.2b; the position of the tree trunk is the dot inside the region. Figure 3.2c shows an example tree map containing several of these regions (i.e., crowns and trunks). The area shown is actually one we used in the experiments we describe in Section 3.3; the bold circles represent the position of WSN nodes.

**Estimation of forest structural attributes.** The tree map allows us to determine the average density of trees $TD$, as number of trees per square meter. Moreover, it provides information about the position of those trees, which is exploited next to derive our per-link estimates of the radio signal power.

However, a last processing step is necessary to extract the diameter $D_t$ of each tree $t$. We estimate the Diameter at Breast Height (DBH) of each tree trunk using the approach presented in [68]. We model the diameter as a function of the tree height $H_L$ and crown radius $K_L$ as follows:

$$D_t = b_0 + b_1\,H_L + b_2\,K_L + b_3\,H_L^2 + b_4\,K_L^2 \tag{3.4}$$

The tree height $H_L$ can be easily derived directly from the CHM value corresponding to the tree top pixel (or seed). Indeed, in the previous step we directly map this value to the maximum height of the pre-processed LiDAR points belonging to the corresponding $50\times50$ cm$^2$ area. The crown's horizontal area is approximated by the area covered by all the pixels belonging to the corresponding region, from which the crown radius $K_L$ is easily computed as the radius of the circle whose area is equivalent to the region area.

We determine the coefficients $\{b_0, ..., b_4\}$ by applying a multi-linear regression that relates a small set of ground truth measurements (i.e., trunk diameter), collected by the local forest service in the location where we performed our experiments, with the tree height and crown radius, namely, with $H_L$, $K_L$, $H_L^2$, and $K_L^2$. We can then apply these coefficients to Eq. (3.4) and estimate the diameter of all the trees belonging to the entire target forest site.

It is worth noting that in-field tree measurements are not strictly necessary, as other diameter estimation approaches exist that do not rely on them. One prominent example are the height-diameter allometric equations widely adopted for forest inventories, which represent the relation between the diameter and the other tree dimensions according to the tree species at hand [97].

### 3.2.3 Automatic Model Tuning and Link-Level Estimate

We now describe how we exploit the automatic processing just described towards building estimates of the radio signal power in the target forest site. We distinguish two cases, we hereafter refer to as AREA and LINK. AREA refers to estimates derived using the original model by Azevedo et al. [14], which considers a single vegetation index $VD$ across the area at hand (i.e., the entire area being considered for the deployment). The contribution we put forth here is the automatic computation of $VD$. LINK, instead, refers to our own adaptation of this model, enabling more accurate per-link estimates, for which the vegetation index $VD$ is individually computed.

**AREA model.** We exploit the output of the processing of LiDAR data to *automatically* compute the index $VD$, necessary to derive estimates based on the model in [14].

We compute the tree density $TD$ by simply counting the overall number of trunks determined in the previous step and dividing by the area of interest. Similarly, we easily compute the average diameter $D$ based on the individual diameter estimates $D_t$. The vegetation index $VD$ is simply the product of $TD$ and $D$, as per Eq. (3.3); substituting the value of $VD$ in Eq. (3.2) yields the value of path loss coefficients; substituting the latter in Eq. (3.1) yields the expected path loss $PL$ in the target area, i.e., the expected amount of attenuation in the signal power depending on the distance.

The interesting quantity from an engineering point of view, however, is the expected received power $P^{rx}$[dB]. This can be computed as a function of path loss, at given distance and for a given frequency, by the following

$$P^{rx} = P^{tx} + G^{tx} + G^{rx} - PL \tag{3.5}$$

where $P^{tx}$ is the transmission power, and $G^{tx}$ and $G^{rx}$ the receiving and transmitting antenna gains, respectively.

In other words, Eq. (3.5) enables us to *predict*, in a given forest area, the received power as a function of distance.

**LINK model.** The fact that we are able to obtain *automatically* predictions of received power using AREA is already per se a significant advancement of the state of the art. Nevertheless, the AREA model is rather coarse, as it assumes that the influence of vegetation is homogeneous across the target area. Unfortunately, it is well-known that typical forest vegetation can not be treated as a single homogeneous dielectric mean [143] and that vegetation irregularity causes, in turn, signal propagation to be irregular [92], leading to non-uniform link connectivity even with a uniform placement of nodes.

Interestingly, AREA does not exploit all the information that can be extracted from the processing of raw LiDAR data. Only the *average* tree density and *average* tree diameter are considered, as prescribed by Eq. (3.3), while our processing actually yields the individual *position* of each tree $t$, along with its *diameter $D_t$*. Next, we show how this information can be exploited into a link-level model which takes into account the forest attributes on a *per-link* basis, i.e., between each node pair, leading to the significant accuracy improvements we discuss in Section 3.3.

The idea behind our LINK model is very simple, yet very effective. It exploits the detailed knowledge about the position and diameter of each tree to compute a vegetation index $VD_{ij}$ specific to the link between node $i$ and $j$:

1. if $VD_{ij} = 0$, the link enjoys a clean line of sight: therefore, the link behavior is better approximated by the *free space* path loss model

$$PL_{free}[dB] = 20\,log(d) + 20\,log(f) - 27.55 \tag{3.6}$$

   where $d$ [m] is the distance and $f$ [MHz] the frequency.

2. otherwise, if $VD_{ij} \neq 0$, the behavior of the link is better estimated by a "localized" version of Eq. (3.5) where the area-specific $VD$ is replaced by the link-specific $VD_{ij}$.
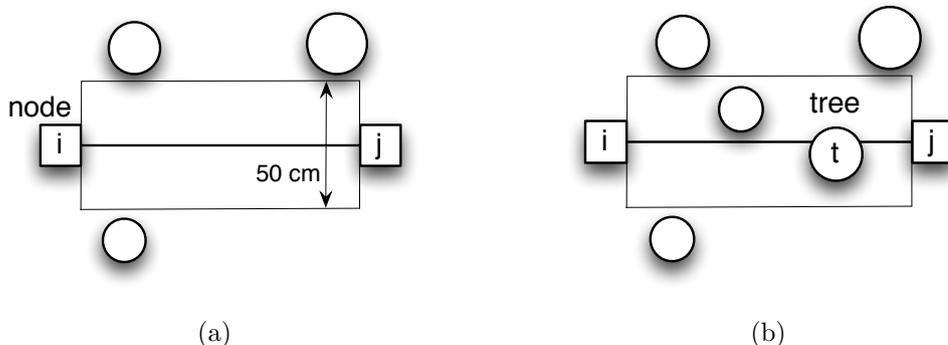


Figure 3.3: Determining the presence of trees on the line of sight.

Figure 3.3 illustrates the concept. $VD_{ij} = TD_{ij} \cdot D_{ij}$ is computed only within the rectangular area (hereafter called the *link area*) whose length is the line connecting the nodes $i$ and $j$ composing the link, and whose height is the pixel resolution used for processing, i.e., 50 cm in this paper. $TD_{ij}$ is trivially computed by dividing the number of trees in the link area by the area of the latter, while $D_{ij}$ is simply the average diameter computed over the trees in the link area. If there are no trees in the link area ($TD_{ij} = 0$), the link enjoys a clean line of sight (clean-LOS) as shown in Figure 3.3a. Otherwise, if $VD_{ij} \neq 0$, some trees are contained in the aforementioned area, and the link is obstructed (obstructed-LOS) as shown in Figure 3.3b. Since the notion of obstruction is defined by considering the entire link area, a link may be obstructed even if no tree is actually sitting on the line connecting the two nodes, as in the case where tree $t$ did not exist in Figure 3.3b. The actual processing to classify links based on line of sight is very simple, and simply consists of checking whether the set of pixels belonging to the link area intersects the set of pixels in which the positions of trunks are mapped.

## 3.3 Validation

We validate our LiDAR-based approach for the automatic assessment of low-power radio signal attenuation in forests by comparing our predictions with RSSI traces we collected in small-scale WSN deployments in a real forest. First, we present the selected location and describe our WSN deployments, the LiDAR data set we used, and how we acquired in-field RSSI traces. Then, we report and discuss our experimental results.

### 3.3.1 Experimental Location

Our study area is a typical alpine forest site in Val di Sella (Trentino, Italy), 1000 m above sea level. It is a mixed forest with trees belonging to the European beech (*Fagus sylvatica*) and Norway spruce (*Picea Abies*) species. No understory is present, the green-leaved part is composed by only a single layer of canopy. We selected this location because it is of particular interest for the local forest service. Therefore, this choice allowed us to exploit the availability of both high-density LiDAR data and, for validation, ground truth data collected in-field by the forest service itself. We consider two different land plots in our study, shown in Figure 3.4a, each with an extension of $\sim$1700 m². One is located in the middle of the forest, and the geographical position of its central point is 46° 0′ 53.64″ N, 11° 22′ 2.51″ E. We refer to this as *internal* plot. The second



Figure 3.4: (a) Experimental location and network topologies. (b) Node setup.

site, which we refer to as the *edge* plot, is on the edge of the forest, close to a clearing, with center at $46° 0' 47.45'' \, \mathrm{N}$, $11° 25' 52.16'' \, \mathrm{E}$.

## 3.3.2   LiDAR Dataset and Related Processing

The LiDAR dataset we use has been acquired by an Optech ALTM 3100EA sensor mounted on an airborne platform. The laser scanner emits pulses with wavelength 1064 nm. For each pulse, up to four returns (i.e., points associated up to the fourth object hit by the laser pulse) were recorded, providing an average point density $>5 \, \mathrm{pt/m^2}$. This dataset has been acquired over our target location on September 4, 2007. Given that our target is a dense old-growth forest, it is reasonable to assume that its structure at trunk level is not changed significantly between the LiDAR data acquisition and our RSSI traces collection. This assumption has been confirmed also by the local forest service.

By applying to the raw data the processing chain described in Section 3.2.2 we obtain a tree map for the whole area comprising both the *internal* and *edge* plots. We visually evaluate the accuracy of the obtained map by overlapping it both with the CHM and with an ortophoto, shown in Figure 3.4a, representing the same area with the same resolution. We assessed a good agreement between the crowns delineated in our tree map and the trees discernible in the other images.

We quantitatively evaluate the precision of the tree map and of the extracted tree attributes by exploiting the availability of a set of measurements gathered in our internal study site by the local forest service, which performed forest inventory by surveying trees in sample locations. One of these surveys targets a circular land plot with a 20 m radius, enclosed in our internal site. The forest service mapped 35 trees in this plot, noting their species along with height and trunk diameter at breast height. In comparison, in the same circular plot our processing chain detected 37 trees, of which 32 correspond to an actual one (i.e., 91% of trees are correctly detected). More precisely, we obtained 5 wrong detections or false positives (i.e., tree tops which actually do not correspond to trunks) and 3 missed detections or false negatives (i.e., missed tree tops/trunks). For the same set of trees, the multilinear regression we applied for the estimation of trunk diameters gives a coefficient of determination $R^2 = 0.7$ and a Root Mean Squared Error $RMSE = 2.75$ cm w.r.t. ground truth values.

As a consequence, the difference between the estimated and ground truth values of $TD$, $D$ and $VD$ is 0.0016 trees/m$^2$, 0.45 cm, and 0.02, respectively. This translates into a difference in the estimated received signal power $<0.05$ dBm for link distances in the range $[0, 60]$ m. This difference can be safely considered negligible in our context, as it is well below the RSSI reading accuracy of the radio chip. Therefore, we conclude that our raw LiDAR data allows us to estimate the relevant vegetation parameters with adequate precision.

### 3.3.3 Low-power Wireless RSSI Traces and Their Collection

The RSSI traces were collected during November 2013, with sunny weather. We deployed the same WSN in both forest plots and collected traces from 12:30 to 14:30 in the *internal* plot, and from 15:00 to 17:00 in the *edge* plot. During the experimental campaign we placed a LASCAR EL-USB-2+ data logger in the same site where we deployed the network to sample both temperature and relative humidity every 15 minutes, to accurately monitor the meteorological conditions experienced by the network. Temperature and humidity were almost stable during each time interval, ranging from -2.5℃ to -3℃, and from 35% to 30%, in the internal plot, and from -3℃ to -4℃ and from 30% to 25% in the edge plot, respectively.

**Node placement.** The WSN we deployed is composed of 9 nodes, placed in a 3×3 grid topology, to characterize the connectivity in a square portion—a sample "forest slice"—of our target environment. Nodes were powered by D-size batteries and placed in waterproof boxes. Care was taken to ensure the same vertical mounting (i.e., antenna orientation) for all nodes while placing them in boxes. Boxes were then latched onto tree trunks with elastic bands at 1.7 m from the ground, as depicted in Figure 3.4b. Since boxes were attached to trees, and these hardly ever form a perfect grid, the actual topology (Figure 3.4a) only approximated the intended one.

**Hardware platform.** We use the popular TMote Sky hardware platform, operating within the ISM 2.4 GHz frequency band according to the IEEE 802.15.4 standards. The radio module mounted on the platform is ChipCon 2420, which includes a digital Direct Sequence Spread Spectrum (DSSS) baseband modem coupled with a digital offset-QPSK modulator, providing an effective data rate of 250 kbps. We exploit the integrated inverted-F microstrip antenna, which is pseudo-omnidirectional with gain of 3.1 dBi.

**Software platform.** The in-field collection of RSSI traces was performed using TRIDENT [74], a tool developed in our group for the untethered execution of communication experiments and collection of connectivity traces. The tool automatically produces the TinyOS code to be installed on TMote Sky motes, based on the experiment configuration input by the user.

The communication links are probed by exchanging radio messages. More precisely, each mote broadcasts a message in a round-robin fashion, to avoid collisions, while all the others are listening to the radio channel. Each time a message is received the reception event is locally recorded by the receiving node, together with the corresponding RSSI value. In addition, nodes sample and locally store the ambient noise floor level. The same procedure is repeated until a user-configured number of messages is sent by each network node. The completion of this process defines a *round*. Each round is characterized by a set of parameters—time interval between two consecutive transmissions or Inter

Packet Interval (IPI), transmission power, radio channel—that can be set and configured in the experiment design phase. The results of the experiment (i.e., the sequence of packet receptions and their quality indicators) are stored in the local memory and can be recovered by means of both multi-hop forwarding among network nodes or direct USB connection.

**Experiment setup and execution.** Our experiment is composed of four 30-minutes rounds, in which each node of the network sends 200 messages with IPI equal to 9 s. All nodes play both sender and receiver roles. The result is a round-robin sending process in which at every second one node is transmitting a packet, while the others are listening. We choose to keep an interval of 1 s between transmissions to avoid possible clock drifts that can cause collisions among senders, given that we are using no Medium Access Control (MAC) protocol. Nodes always transmit on channel 18 (i.e., at 2.44 GHz with a bandwidth of 3 MHz). We alternate rounds with transmission power -1 dBm and -8 dBm, which we refer to as *high power* and *low power* respectively.

This experimental setting allows us to probe each $link_{ij}$ (i.e., the link from $node_i$ to $node_j$) every 9 s; we collect the raw packets, along with per-round and overall statistics. We expected to probe 72 links for each site, considering separately $link_{ij}$ and $link_{ji}$. However, the actual number of links we probed (i.e., 30 for each site) is lower than expected because some nodes malfunctioned halfway through the experiments, and were therefore excluded from the analysis. For each $link_{ij}$ the traces we collect consist of a sequence of records for each round. Each record represents the reception of a message along that link and contains RSSI and noise floor related to the reception of that message. The results in this paper are based on the analysis of the 41,794 data points we collected overall.

### 3.3.4 Results

In this section, we present our experimental results, validating our approach against the actual received signal power measured in our WSN deployment. Our goal is to assess to which extent: *i)* conventional empirical channel models for vegetated environments match the real RSSI traces we collected at trunk level; *ii)* LiDAR data allows us to automatically describe and represent local vegetation characteristics and enable received signal power predictions in a forest area; *iii)* we can improve the prediction accuracy by analyzing trees configuration and its impact at per link level.

The RSSI values in our traces represent the sum of the received radio signal power and the noise power. Therefore, we can convert those values to a received power indicator—comparable with the predictions of the models—and compute the corresponding $P^{rx}$ by subtracting, in Watt scale, the noise floor level from RSSI. This step, in addition, makes

our analysis independent from possible hardware differences, in terms of noise figure, between nodes.

**Shortcomings of conventional empirical channel models.** As a term of comparison, we first estimate the path loss $PL$ by applying both the Weissberger model [152] and the COST 235 model [101]. These conventional models estimate the *excess attenuation* due to vegetation as a quantity $L_{veg}$ [dB]. The overall path loss in this model is obtained as $PL_{free} + L_{veg}$, where the first term is the free space path loss as per Eq. (5.2). The excess attenuation $L_{veg}$ is expressed as a function of the radio frequency $f$, in GHz, and the depth of foliage $d$, in meters. For Weissberger, the excess attenuation is

$$L_{vegW}[dB] = \begin{cases} 0.45\, f^{0.284}d & 0 \leq d \leq 14 \\ 1.33\, f^{0.284}d^{0.588} & 14 < d \leq 400 \end{cases}$$

For COST 235, which distinguishes between in-leaf (IL) and out-of-leaf (OL) situations, excess attenuation is defined as:

$$L_{vegCOST}[dB] = \begin{cases} 15.6\, f^{-0.009}d^{0.26} & in-leaf \\ 26.6\, f^{-0.2}d^{0.5} & out-of-leaf \end{cases}$$

To perform our comparison, based on the above we compute the corresponding expected received power $P^{rx}$ (at given distance and for a given frequency) according to Eq. (3.5). For the COST 235 model we consider both the in-leaf and out-of-leaf settings, given the mixed nature (i.e., coniferous/deciduous) of our forest plots.

Figure 3.5 compares the per-distance average of the real measurements collected in-field against the estimated received power according to Weissberger and COST 235 models, as a function of the link distance. We show only the plots for the low-power setting, because the plots for the high-power setting show similar results. COST 235 estimations, both in-leaf and out-of-leaf, do not fit well our data. Although the Weissberger model behaves slightly better, it still significantly *underestimates* the received power for most of the links. As a consequence, its application for network planning and deployment in this forest would likely lead to *overprovisioning*.

The reason behind the discrepancy between these models and the real traces is that the former are not sufficiently representative of our specific forest environment. This is not surprising, as these models are intended as generic, "one-size-fits-all" solutions, which account for vegetation by making the assumption that it is homogeneous, and as such they fail to capture the intrinsic variability displayed by the complex forest environment. In addition, these models were built for communication links that traverse the canopy (as most of the empirical models developed in the past), which is not the scenario we are considering. Finally, they have been derived empirically by curve fitting and, even

though the measurement sets on which they are based is large, they clearly cannot cover all possible forest configurations at the same time.

**LiDAR-based estimates: AREA.** We now show that our LiDAR-based *automatic* approach for estimating the parameters of the model described in Section 3.2 provides more accurate predictions. Table 3.1 shows the vegetation index and its constituents for both our deployments, as estimated by the processing chain described in Section 3.2.2. Based on these values, we compute the path loss coefficients as a function of *VD* as per Eq. (3.2) and, by applying Eq. (3.5) and Eq. (3.1), we define the curve of the expected received signal power as a function of distance.

| **site** | $D$ [cm] | $TD$ [trees/m$^2$] | $VD$ |
|----------|----------|--------------------|------|
| internal | 24.2 | 0.0355 | 0.8598 |
| edge | 25.2 | 0.0372 | 0.9366 |

Table 3.1: Vegetation parameters in our deployment sites.

The results are shown in Figure 3.5. We can visually assess that the trend of our experimental data is well captured by such curve. We quantitatively evaluate the performance by comparing our in-field data and the estimated received power for the reference distances of our links. In particular, we consider the average received power computed on our low power and high power traces, in the internal and edge sites, separately. For instance, the average estimation error for the internal site at low power is 4.06 [dBm] with a standard deviation of 3.31, minimum 0.25 and maximum 12.1. A complete account of the estimation errors for both sites and power settings is shown in Table 3.2.
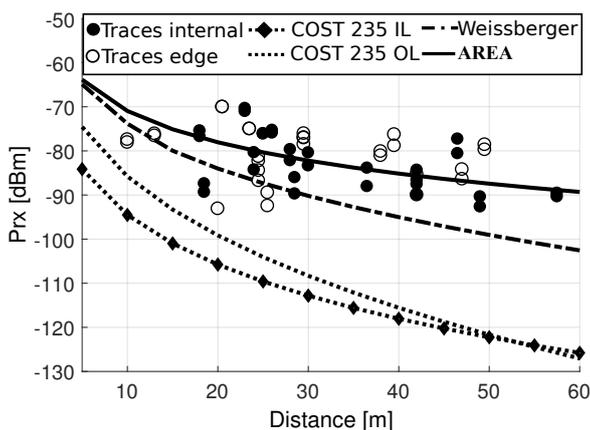


Figure 3.5: Received signal power according to real RSSI traces and various prediction models (low power).
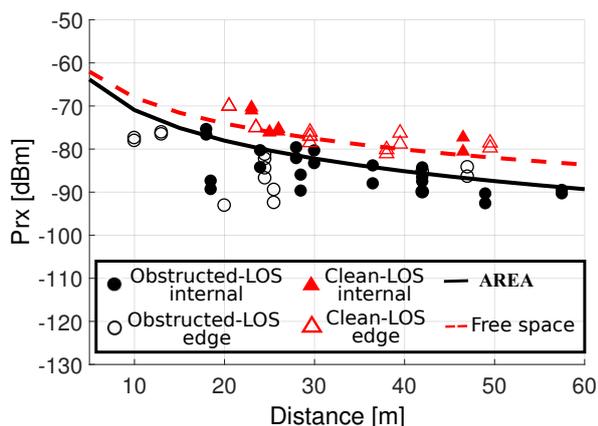


Figure 3.6: Analysis of clean-LOS and obstructed-LOS links using AREA and the free space path loss model (low power).

**LiDAR-based estimates: LINK.** Despite the promising result, there are still significant discrepancies between the real data and the AREA prediction model. Next, we show that the LINK model we introduced in Section 3.2.3, thanks to its ability to select the most appropriate path loss model, leads to further and significant improvement in estimation accuracy.

Figure 3.6 shows the real power received, as derived from our RSSI traces, for both clean-LOS (triangles) and obstructed-LOS (dots) links, together with the estimation curves given by AREA and the free space path loss model. From the chart, it is evident that the latter can better predict the behavior of clean-LOS links for which, across both deployment sites, our diversified estimation strategy reduces the average error from 6.22 dBm to 1.86 dBm at low power, and from 14.21 dBm to 2.71 dBm at high power. Table 3.2 offers the complete error statistics (average, standard deviation, minimum and maximum) of this "dual-model" LINK prediction technique vs. AREA, across different deployment sites, power settings, and line-of-sight situations (clean vs. obstructed). By automatically identifying clean-LOS and obstructed-LOS links and applying our diversified strategy accordingly, we significantly and systematically reduce the estimation error for all the locations and power settings we considered in our experiments.

Figure 3.7 offers a different, graphical view of the comparison by plotting the real received power against the one predicted by AREA and LINK. The diagonal on the plot depicts the perfect estimate with zero error; the closer a point is to this line, the more accurate the corresponding prediction is. Moreover, the charts also show the ±6 dBm error band; we chose this value as this is the accuracy of the RSSI readings of the CC2420 radio chip our WSN nodes are equipped with. Figure 3.7a compares the performance of AREA and LINK w.r.t. clean-LOS and obstructed-LOS links, represented by black and white dots, respectively. It is interesting to note that AREA consistently *understimates* the received power for clean-LOS points. In LINK, these points are effectively "shifted" closer



(a) Line of sight: Clean vs. obstructed.    (b) Deployment site: Internal vs. edge.
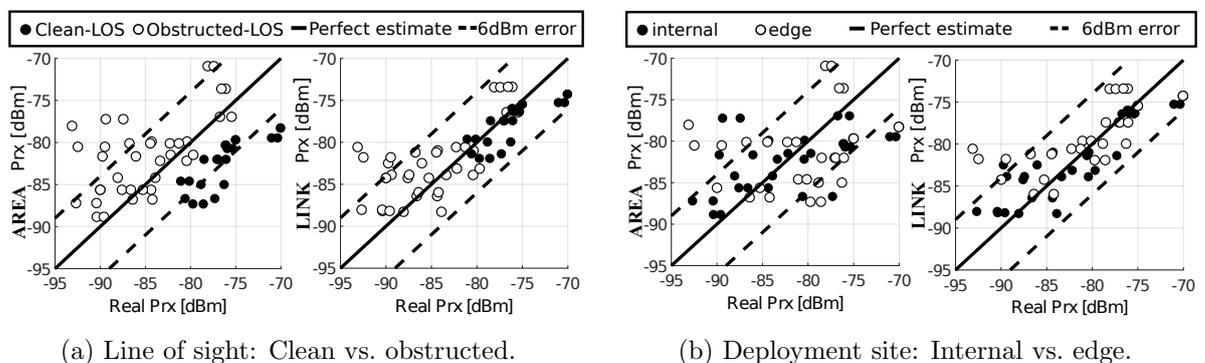
Figure 3.7: Prediction accuracy of AREA and LINK w.r.t. a ±6 dBm error band.

| | site | avg error | | std dev | | min | | max | | % in ±6 dBm | | % in ±1 dBm | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AREA | LINK | AREA | LINK | AREA | LINK | AREA | LINK | AREA | LINK | AREA | LINK |
| low power | **internal** | 4.06 | 2.52 | 3.31 | 1.99 | 0.25 | 0.03 | 12.10 | 7.20 | 76.64 | 96.67 | 20.00 | 33.33 |
| | clean-LOS | 6.48 | 1.97 | 2.18 | 2.08 | 4.21 | 0.03 | 9.40 | 4.94 | 50.00 | 100.00 | 0.00 | 50.00 |
| | obstructed-LOS | 3.18 | 2.72 | 3.25 | 1.97 | 0.25 | 0.14 | 12.1 | 7.20 | 86.36 | 95.54 | 27.27 | 27.27 |
| | **edge** | 5.57 | 3.11 | 3.32 | 3.02 | 0.39 | 0.40 | 15.04 | 12.50 | 56.67 | 86.67 | 6.67 | 26.67 |
| | clean-LOS | 6.07 | 1.80 | 1.91 | 1.48 | 3.55 | 0.39 | 8.77 | 4.27 | 57.14 | 100.00 | 0.00 | 35.71 |
| | obstructed-LOS | 5.11 | 4.33 | 4.27 | 3.58 | 0.39 | 0.39 | 15.04 | 12.50 | 56.25 | 75.00 | 12.54 | 18.75 |
| | **all** | 4.81 | 2.81 | 3.38 | 2.55 | 0.25 | 0.03 | 15.04 | 12.50 | 66.67 | 90.00 | 13.33 | 30.00 |
| | clean-LOS | 6.22 | 1.86 | 1.97 | 1.68 | 3.55 | 0.03 | 9.40 | 4.94 | 54.55 | 100.00 | 0.00 | 40.91 |
| | obstructed-LOS | 3.29 | 3.37 | 3.76 | 2.81 | 0.25 | 0.14 | 15.04 | 12.50 | 73.68 | 86.84 | 21.05 | 23.68 |
| high power | **internal** | 8.20 | 2.73 | 4.72 | 1.69 | 0.77 | 0.26 | 17.58 | 5.49 | 36.67 | 100.00 | 6.67 | 13.33 |
| | clean-LOS | 14.49 | 2.94 | 2.08 | 2.06 | 12.33 | 0.83 | 17.58 | 5.49 | 0.00 | 100.00 | 0.00 | 12.50 |
| | obstructed-LOS | 5.91 | 2.65 | 2.97 | 1.58 | 0.77 | 0.26 | 10.19 | 5.39 | 50.00 | 100.00 | 9.09 | 13.64 |
| | **edge** | 10.18 | 4.15 | 4.89 | 4.97 | 0.73 | 0.43 | 17.11 | 14.12 | 30.00 | 83.33 | 3.33 | 23.33 |
| | clean-LOS | 14.06 | 2.57 | 1.94 | 1.66 | 11.00 | 0.43 | 17.11 | 5.23 | 0.00 | 100.00 | 0.00 | 28.57 |
| | obstructed-LOS | 6.79 | 5.53 | 4.08 | 6.40 | 0.73 | 0.42 | 14.60 | 14.12 | 56.25 | 68.75 | 6.25 | 18.75 |
| | **all** | 9.19 | 3.44 | 4.87 | 3.75 | 0.73 | 0.26 | 17.58 | 14.12 | 33.33 | 91.67 | 5.00 | 18.33 |
| | clean-LOS | 14.21 | 2.71 | 1.96 | 1.77 | 11.00 | 0.43 | 17.58 | 5.49 | 0.00 | 100.00 | 0.00 | 22.73 |
| | obstructed-LOS | 6.28 | 3.86 | 3.46 | 4.48 | 0.73 | 0.26 | 14.60 | 14.12 | 52.63 | 86.84 | 7.90 | 15.79 |

Table 3.2: Experimental results

to the diagonal, as the effect of the attenuating *VD* coefficient is removed from the path loss model. *All* of the clean-LOS points are within the ±6 dBm error band. On the other hand, AREA appears to partly *overestimate* the effect of vegetation. Moreover, the white points in LINK are in general less spread and much closer to the diagonal. Figure 3.7b compares the performance of AREA and LINK w.r.t. the deployment site, i.e., internal vs. edge. In this case, LINK provides in general better predictions in both cases.

The rightmost part of Table 3.2 shows the fraction of predictions falling within the reference ±6 dBm error band for both AREA and LINK, for all the combinations of deployment sites, power settings, and link types. The quantitative data confirm that LINK systematically improves over AREA; apart from the dramatic improvement for clean-LOS links, obstructed-LOS ones have a significant improvement in the case of high power. We also show data for the narrower error band of ±1 dBm, to assess how close the two models approximate the perfect estimate. Once again, LINK gets systematically closer; the best result is for clean-LOS, low-power, internal plot, where 50% of the links are within ±1 dBm of the real value with LINK, and 0% with AREA.

Although LINK represents an improvement in estimation performance, there are still discrepancies that the current model cannot account for. We argue that these are mainly due to the effect of obstructing trunks which are very close to the node, e.g., the trunks on which nodes are latched onto. This is addressed in Chapter 4.

## 3.4 Conclusions

The goal of accurately predicting the behavior of low-power wireless communication is still a rather elusive one, especially in forests, where the irregularity of vegetation exacerbates

complexity. In this chapter, we presented a LiDAR-based approach that is *automatic*, i.e., it does not require the in-field campaigns commonly adopted to perform radio (or vegetation) surveys, and *fine-grained*, i.e., it enables received signal power estimates on a per-link basis. We validated our approach on real RSSI traces from two small-scale WSN deployments in a forest, and shown that it achieves unprecedented accuracy in estimating the received signal power.

Despite the encouraging results, however, further work is required to explore how the presented approach is affected by other, complementary variables (e.g., forests with different foliage or a denser understory, and environmental conditions as rain or variations in temperature). In addition, further research is needed to investiagte the impact of other forest attributes—e.g., the distance between a node and a tree trunk, to assess their actual impact on connectivity and consequently on our estimates. Finally, a practical use of our technique would be its integration in a network planning tool where, starting from an initial seed placement of nodes, the optimal placement is automatically derived. Both the impact of the distance between a node and a tree trunk and the integration of this approach in a network planning tool are addressed in Chapter 4.

# Chapter 4

# LaPS: LiDAR-assisted Placement of WSNs in Forests

Wireless Sensor Networks (WSNs) enable in-situ, unattended monitoring of outdoor natural environments with unprecedented density and flexibility. In the context of forests, which are the main focus of this chapter, they have been employed in several applications with different goals, e.g., including forest fire detection, microclimate monitoring, wildlife monitoring [67, 121, 145, 49, 138]. In this context, a major challenge is to achieve a proper node placement, as it strongly affects the behavior of communication links [159, 44, 92, 113, 63], and therefore the connectivity of the deployed network and the application performance and reliability at large. However, deploying WSNs in the real world remains a very challenging task, especially in outdoor environments [142, 63, 96]; forests further exacerbate the challenge due to the presence of trees and vegetation impairing low-power radio communication [99, 40, 154].

In principle, a large literature on network design and node placement optimization exists [158, 23, 122, 83, 5, 77], surveyed in Section 4.1. However, the modeling of low-power wireless communication in these approaches lacks realism, as they often neglect the specific features of the real-world target scenario. Since the characteristics of the target environment may vary wildly, they are likely to disrupt the model assumptions, yielding estimates that are unrealistic and of little practical use.

As a consequence, the placement problem is often tackled directly in-field, by means of effort-demanding experimental campaigns. A common approach is to define an initial placement "guess" based on the spatial (e.g., node density) and network (e.g., expected signal strength or number of neighbors) requirements germane to the application, and

---

on the nominal data found in datasheets and derived by idealized radio communication models. Next, the quality of such placement is evaluated in-field by means of connectivity tests, for which several supporting tools exist [74, 42, 141]. However, due to the aforementioned peculiarity of the target environment, it is rarely the case that this initial guess is satisfactory; the position of nodes typically must be nudged based on the outcome of the tests, which must therefore be re-executed, leading to a trial-and-error cycle that repeats until a satisfactory network configuration is found. The effort required by this process obviously increases with the scale of the network and the complexity of the target environment.

In this respect, the forest environment represents both a challenge and an opportunity. The challenge is the fact that, as already pointed out, trees and vegetation impair the radio signal. However, the opportunity is that a significant fraction of this impact is induced by trees, and is therefore *permanent*, i.e., not time-variant, at least not on a short time scale; in other words, the attenuation they induce could be in principle estimated beforehand, and form the basis for determining a satisfactory node placement. This is precisely the goal of this chapter.

**Approach and contributions.** We present an *automatic* node placement approach and companion tool, called LaPS (L̲iDAR-a̲ssisted P̲lacement for wireless S̲ensor networks), that optimizes the positions of WSN nodes *prior to deployment* by taking into account the *real characteristics of the target forest.*

The node placement identified by LaPS is subject to a set of simple user-defined spatial and network requirements (Section 4.2) specifying desired properties about the placement of nodes and the resulting network connectivity. In this work, the quality of the network layout output by LaPS is assessed in terms of the overall number of communication links and their average expected receive power; however, alternative formulations can be easily encoded, thanks to our flexible design.

A second input to LaPS is the information about the forest structure. To this end, we rely on remote sensing and specifically airborne Light Detection and Ranging (LiDAR) technology (introduced in Chapter 2), as it has been extensively applied to forest analysis and management [97, 116, 139, 68, 13, 82] for the accurate estimation of forest attributes (e.g., tree position, trunk diameter, tree density) and as it is becoming increasingly available for forestry applications.

User requirements and raw LiDAR data are fed to the LaPS toolchain, for which we provide an overview in Section 4.3. The first component of the toolchain transforms the raw LiDAR data into a higher-level representation of forest attributes we call a *tree map* (Section 3.2.2). This information provides the crucial parameters of a specialized radio model that enables accurate per-link estimates of communication quality by taking into

account the attenuation induced by trees, as we described in Chapter 3.

The tree map representing the target forest area and the radio model configured with this information are both input, along with user requirements, to an evolutionary optimization method (Section 4.4) that constitutes the last component of the toolchain. Specifically, we exploit genetic algorithms to explore the space of possible placement solutions and evolve towards an optimized placement. This is done by evaluating the *fitness* of placement configurations in terms of quality and number of the communication links available to network nodes, while honoring the spatial and network requirements specified by the user.

To the best of our knowledge, we are the first to exploit LiDAR information with the goal to *i)* characterize the attenuation induced by trees, and *ii)* exploit this and the acquired knowledge about tree positions to identify an optimal placement of WSN nodes.

Before moving to the evaluation of our placement approach, we discuss the configuration of the LaPS tool (Section 4.5) we use for it. This entails not only providing specific examples of user constraint, but also the identification of two parameters, the expected received power and the distance of nodes from trunks, whose minimum value has a relevant impact on the quality of the output solutions and whose value must be determined experimentally.

Finally, we evaluate the quality of the node placements output by LaPS (Section 4.6). We analyze various performance metrics, including the overall average expected received power and number of links in the network, along with topological properties quantifying its degree of connectivity and therefore intrinsic robustness. We compare against *i)* a grid-based approach that "blindly" places nodes in each spatial cell without taking into account the presence of trees, and *ii)* a line-of-sight approach that uses the first portion of our LaPS toolchain to determine the links that are unencumbered by trees and for which the standard path loss radio model applies. In essence, the first baseline disregards both the information about the forest structure and the corresponding customized radio model, while the second disregards only the latter. By showing that the placements output by LaPS are significantly better than both baselines we confirm that the additional complexity introduced by LaPS is indeed worthwhile. This becomes even more evident when we introduce a four-fold scale-up of the test network, in which the baselines above cannot even find an acceptable solution.

We conclude our evaluation by quantifying the improvement in quality between the placement solution output by LaPS and the average one. The latter effectively provides a measure of the placement that in-field campaigns are likely to identify. Our results confirm that the improvement attainable by LaPS is significant, thanks to its ability to quickly explore several alternate placements. Therefore, LaPS can provide solutions

that *i)* are of higher quality, and *ii)* without requiring the significant effort of in-field campaigns. Further, and along the same lines, LaPS enables the exploration of slightly different network and/or performance parameters, to an extent that would be simply prohibitive if performed in-field.

The chapter ends in Section 4.7 with brief concluding remarks including opportunities for future work.

## 4.1 Related Work

**Genetic algorithms and WSN node placement.** In the last two decades, several studies exploited genetic algorithms towards the planning of mobile networks, in particular for cellular base stations placement and configuration [104, 90, 98]. The effectiveness of these approaches inspired the application of this technique also in the context of WSNs placement optimization.

However, many of these approaches (e.g., [54, 23, 24, 6]) adopt a grid or some other regular pattern (e.g., a tessellation) as the reference network layout, and select candidate positions for the optimal node placement only among the vertexes of this regular pattern (e.g., at the crossing of grid lines). This intuitive abstraction simplifies the mathematical treatment of the problem and reduces the search space. However, from a practical standpoint, while it is meaningful for some applications—and indeed we consider it ourselves—it is at odds with several real scenarios, as it neglects specific (and often irregular) spatial requirements of applications. In contrast, our approach supports arbitrary spatial requirements where nodes must be deployed in designated areas; the requirements we consider include the common case where nodes must be fastened to trees, yielding an intrinsically irregular structure of the resulting network and search space.

Another significant idealization of the placement problem concerns the communication range, which is typically assumed perfectly isotropic. In part, this is a consequence of the fact that the majority of approaches in the WSN literature aim at jointly optimizing energy consumption and sensing coverage, i.e., minimizing the former without compromising the latter [122, 23, 79]. In this respect, modeling the range of both communication and sensing as a perfect circle is a natural abstraction that, again, greatly simplifies the mathematical treatment. The price to pay, however, is the inability to transfer these approaches in the real world, where communication range is known to be far from isotropic [159].

For instance, the authors of [23] exploit genetic algorithms to optimize a cluster-based approach by determining the best clustering scheme, the operational mode of nodes (e.g., active vs. inactive, slave vs. cluster-head), and their transmission power. However, candidate node positions are restricted to those belonging to a regular grid layout. Communication range is one of the design parameters considered in the optimization; however, it is

assumed to depend only on the transmission power, and the impact of the deployment environment is neglected. Similarly, the multi-objective approach presented in [79] optimizes sensor coverage and lifetime, with the additional assumption of a fixed communication and sensing range. Another multi-objective optimization technique is presented in [54] in the context of precision agriculture. Again, the focus is on the selection of node status by taking into account application-specific requirements related to the operation mode, but no effect of the agricultural field is considered in evaluating the connectivity of the network. Analogous considerations hold for the approach in [151], where the optimization of sensing coverage and network connectivity is centered around the sleep intervals scheduling for energy conservation while providing different degrees of coverage through their dynamic reconfiguration. However, the connectivity estimation is, once again, unrealistic. The same holds for [122] that, after explicitly mentioning the challenges of real deployments and the difficulty of acquiring prior information about the environment, proceed to optimize radio coverage and energy consumption based on a circular communication range of fixed radius.

In contrast, Krause et al. [84, 85] propose a placement approach that simultaneously optimizes sensing and communication quality, avoiding the simplistic assumption of fixed communication range. They take into account the variability of the link quality and define the communication cost by probabilistic models that are learned in an initial small pilot deployment. However, it is hard to learn a realistic probability distribution of the link quality in a forest through a small pilot deployment, due to the intrinsic inhomogeneity of such environment. The effective application of this approach in forests would require significant, therefore costly, pilot campaigns to learn representative models.

The common idealistic assumption of fixed and isotropic communication range is shared also by the multi-objective optimization of node position and transmission power presented in [83]. However, this work has an additional point of contact with ours, in that it introduces $k$-connectivity constraint on the resulting network, aimed at ensuring some degree of robustness. In our approach, we consider this and other networking constraints, and analyze their impact on a wider notion of robustness based both on topological properties and consideration common in the design of WSN networking protocols.

Finally, and more generally, LaPS sharply departs from the aforementioned strongly idealized settings by *i)* taking into account the peculiarity of communication range in WSN with a *specialized* attenuation model *ii)* whose parameters are directly informed by the characteristics of the forest area targeted by the deployment, automatically derived via our LiDAR-based toolchain.

**Modeling the non-isotropic communication range of WSNs.** As already mentioned, the problem of radio range irregularity is anyhow well known, and it has been

addressed in the recent literature. However, it is treated mainly in statistical terms, specifically: *i)* in terms of Degree of Irregularity (DOI), originally defined in [65] as the maximum range variation per unit degree change in the direction of radio propagation; or *ii)* refined and incorporated in the Radio Irregularity Model (RIM) presented in [162], where the DOI is improved by introducing a random weight for each direction, according to a Weibull distribution, to represent the random variance experimentally observed at each direction; or *iii)* by describing the statistical behavior of the reception rate [159, 43]. Cerpa et al. [43] characterize the communication links using non-parametric statistical models that describe the reception likelihood and its confidence interval considering features of links and groups of links (e.g., distance, asymmetry, uniformity of transmitters and receivers). Zuniga et al. [159] identify three different regions (i.e., connected, transitional and disconnected), which present different communication characteristics according to the link distance. Links belonging to the connected region are stable and symmetric, featuring high reception rate; links in the disconnected region are almost completely lossy and unreliable, whereas links in the intermediate transitional region are highly variable and unpredictable. However, both these studies, despite considering different deployment environments (e.g., indoor and outdoor), do not explicitly take into account the specific characteristics of such environments in the analysis. In other words, although these formulations are derived from experimental evidence, they represent an attempt to generalize a behavior which is intrinsically determined by the specificity of the deployment scenario.

In LaPS, we make a step forward in reconciling the intrinsic generality of modeling and the specificity of the environment at hand, by exploiting the peculiarity of the forest setting. The specialized radio propagation model we defined in Chapter 3 is the cornerstone of our approach. Although we cannot reproduce every specific aspect of the target environment affecting communication (e.g., temperature, humidity), we exploit the fact that its quality is severely impacted by the presence of trees, whose position does not change over a short time scale. As a consequence, the attenuation they introduce can be itself modeled and taken into account when determining the optimal node placement.

**Exploiting digital models of the environment.** The idea of exploiting digital models of a target area and incorporating them in network planning approaches to increase their realism has been applied almost two decades ago by Krzanowsky et al. [86]. Digital Elevation Models (DEMs) and land cover maps were included in a genetic process to tune the expected signal attenuation and compute a more realistic cell coverage, therefore improving the positioning of base stations in wireless networks. More recently, the authors of [113] presented a connectivity model that takes into account topographic and vegetation features, similarly derived from a DEM and land cover maps. However, this model relies on a machine learning algorithm whose training requires the collection of a significant

amount of in-field connectivity measures in the target deployment environment.

DEMs are often obtained from satellite or airborne remote sensing systems, whose capabilities have also been explored in this context. In [132] the authors describe how to extract building footprints from LiDAR data with the goal of supporting the design of wireless communications systems in urban areas. These urban features are incorporated in microcell ray tracing models and exploited to assess the visibility status between transmitter and receiver (i.e., visible or obstructed), therefore enabling the selection of a proper radio model (i.e., free space loss vs. single-knife edge diffraction). Similarly, in [89] the line of sight between the satellites and the receiver in a GPS system is evaluated using LiDAR and 3D ray tracing, to assess the positioning accuracy. The work we present here is inspired by these works, in the sense that we similarly exploit the high accuracy of LiDAR to build a detailed model of the environment, and incorporate it into a propagation model. However, these approaches are not directly applicable to our case, as they focus on an urban setting instead of a forest one; further, the peculiarity of the latter, and specifically the need to take into account the attenuation induced by *each* tree, requires a much more fine-grained approach than these works offer.

On the other hand, only few studies in the WSN literature attempted to increase the degree of realism by considering the target scenario and its impact on the effectiveness of node placement. In [5], DEM models are used to estimate the line of sight of PIR sensors in a mountainous region, and determine via an evolutionary approach the optimal placement maximizing sensing coverage. The approach in [77] similarly takes into account coarse-grained elements of the environment (e.g., the presence of vegetation or rivers) known to degrade the sensing capabilities by a given percentage. Another attempt to explicitly reckon with the specific environment where the WSN is deployed is presented in [10], which also includes a radio model that *i)* estimates the range on quantized directions around the node by accounting for the presence of obstacles, and *ii)* introduces time-variant environment-dependent components (e.g., a climate factor and an environmental attenuation factor). The results obtained through simulations show that the average range indeed varies based on these factors. However, these are not quantified and the estimation of the corresponding coefficients not discussed; global knowledge about the environmental factors and their impact on communication is instead implicitly assumed.

In contrast, the radio attenuation model at the core of LaPS is validated by in-field experiments; further, to the best of our knowledge, we are the first to exploit remote sensing and specifically LiDAR to derive the fine-grained model of the environment enabling an accurate estimation of attenuation and, ultimately, a significantly better node placement.

## 4.2   Requirements and Goals

We focus on the deployment of WSNs in forests, typically to directly monitor their environment or the animals dwelling in it. Given that the target location is typically harsh and not easy to reach, simplifying the WSN deployment process is crucial, as discussed at the beginning of this chapter.

Two factors determine the effectiveness of a WSN deployment in our context. On one hand, *spatial* requirements must be honored, determining how nodes should be distributed in the environment to fulfill the application needs. On the other hand, a node placement that satisfies only these constraints is not sufficient; *network* requirements must also be fulfilled, to ensure good connectivity among nodes and other properties to be exploited by network protocols towards reliable and efficient communication.

In principle, both spatial and network requirements strongly depend on the application at hand; our approach and associated tools are general enough to accommodate a wide spectrum of requirements. However, in this dissertation, we focus on common requirements we informally state next in Section 4.2.1 and Section 4.2.2.

### 4.2.1   Spatial Requirements

**Global: REGULAR vs. IRREGULAR.**  In some applications, e.g., including forest fire detection and microclimate monitoring [49, 67, 63], it is desirable to place nodes in a way that guarantees a spatially uniform sensing with a desired and controlled density. In principle, this could be achieved by organizing nodes in a regular grid, which is indeed a popular choice in placement approaches [158]. In practice, however, an *exact* grid is often not feasible (e.g., due to irregularity of the target environment) or even desirable (e.g., network concerns may suggest a slightly different placement). For this reasons, deployments typically exploit a tessellation of the target area [158, 24, 6] into tiles (e.g., triangles or squares); each node must be contained in one of the tiles. Hereafter, we refer to this deployment scenario as REGULAR.

Other applications are driven by different spatial requirements. For instance, in wildlife monitoring [123, 124] biologists are often interested in monitoring specific sites in the target area (e.g., close to water and feeding sources, or where animal traces are frequently found). Nodes are typically required to be placed near these sites (e.g., within a given radius around a position); in some cases, however, a node may be required to be placed in a designated, geo-referenced position. Moreover, extra relay nodes may be required, whose position is not subject to strict spatial requirements. We refer to this deployment scenario as IRREGULAR.

**Node: FREE vs. TRUNK.** In addition to the global spatial requirements above, some applications pose also constraints w.r.t. the physical placement of nodes themselves. More precisely, in what we refer to as the FREE placement, the nodes can be placed anywhere in the target area, e.g., on the ground or atop poles. However, this option is not viable in some locations, e.g., to avoid that nodes are relocated by weather events or animals, or due to the danger of damaging the tree roots, respectively. Therefore, a common alternative is to fasten the nodes directly to the trees themselves; we refer to this placement as TRUNK. Clearly, these two options implicitly define two different search spaces for the node placement problem, since the TRUNK admits acceptable node positions *only* in correspondence of the tree trunks.

### 4.2.2 Network Requirements

Network requirements ultimately depend on the specific network stack adopted. However, the reliability of the latter is in turn directly affected by the *connectivity* of the network, which we capture by posing requirements on two fundamental metrics on each network node:

- *Minimum expected received power.* This metric effectively defines, in the most basic (and therefore general) way, the minimum acceptable quality of a link. By enforcing this requirement globally, we ensure that all links in the network are "good enough". We further elaborate on this notion in Section 4.5.1.

- *Minimum number of neighbors.* This metric builds on the previous one by considering, among all node neighbors, only those with good links. By guaranteeing that a minimum number of these neighbors are available, this metric ensures that each node has enough communication options—a critical factor for reliability, e.g., in routing protocols.

### 4.2.3 Goal: Optimal Node Placement

The goal of our approach is to find an *optimal* node placement, defined informally as a positioning assignment for all nodes of the network that:

- satisfies both the spatial and network requirements above, and

- maximizes the quality of the resulting communication network.

Again, several formulations are possible for the latter aspect. Hereafter, we rely on the same basic notions of received power and number of neighbors we exploited in Section 4.2.2, and identify the optimal placement as the one that maximizes:

- the total number of acceptable links in the network, and

- the average expected received power across all of these acceptable links.

Although these metrics are very basic, when their value is jointly optimized by our approach (Section 4.3) they bear a direct effect in shaping the topological characteristics of the network, e.g., how connected is the resulting network graph or how many source-to-sink paths exist in a potential routing topology, as we show in the evaluation (Section 4.6).

## 4.3   LiDAR-assisted Node Placement: An Overview

We now provide the reader with a bird's eye view of the approach we employ to achieve the goal of optimal placement, for which we exploit the availability of LiDAR information. Figure 4.1 offers a pictorial representation of the key components and their relationships.

We exploit in combination three main building blocks:

1. *LiDAR-based forest representation* (Section 3.2.2). It takes as input the geographical representation of the target forest area and the associated LiDAR data and determines the position of trees and the diameter of their trunks in the area of interest, encoded as a *tree map*;

2. *LiDAR-based radio attenuation model* (Section 3.2.3). It enables *a priori* estimation of the received power of the radio signal based on the aforementioned tree map.

3. *Evolutionary optimization* (Section 4.4). It is the last and most important step, whose inputs are the key parameters of the placement problem, namely, the number of nodes and the spatial and network requirements and constraints (4.2). These are combined with the above knowledge about tree positions and their impact on radio propagation and exploited by a genetic algorithm that efficiently explores the search space of feasible placements (i.e., those satisfying spatial and network requirements) and selects the best one.

We next describe in detail how the search for the optimal placement is implemented.
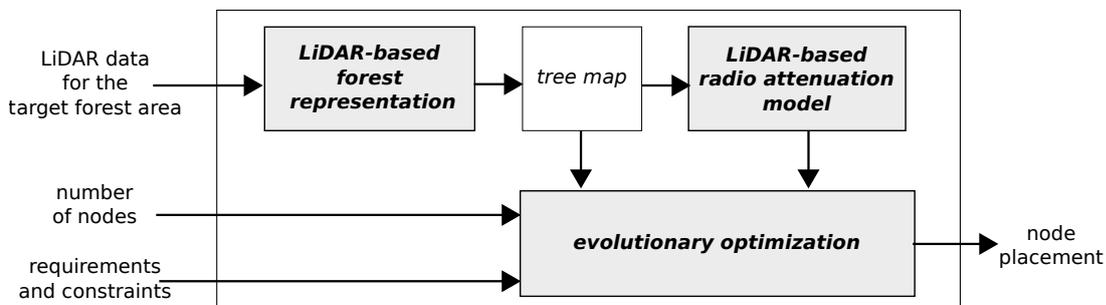


Figure 4.1: LiDAR-assisted node placement.

## 4.4 Evolutionary optimization

As shown in Figure 4.1, the evolutionary optimization combines the knowledge derived from LiDAR about the forest structure and its effect on the communication quality (Chapter 3) with user-defined requirements (Section 4.2) about the number and position of nodes, and outputs a (sub-)optimal placement satisfying these requirements in the target forest area.

The latter cannot be found simply by exhaustive search, which rapidly becomes unfeasible as the number of nodes and/or size of the target area increase. Therefore, we exploit genetic algorithms, a well-known class of numerical optimization procedures inspired by biological evolution that proved effective in optimization problems similar to ours [54, 23, 79, 83, 86, 77].

This class of approaches examines a search space by manipulating and evaluating a set of possible solutions, i.e., a *population* of *individuals*; in our case, an individual is a candidate placement configuration, represented by simply concatenating all node positions. We describe how the latter are determined and in general how the forest area is modeled in Section 4.4.1.

An individual can be considered as acceptable only if it satisfies *constraints* representing the problem at hand; Section 4.4.2 provides a formalization of the constraints for our problem, which descend from the spatial and network requirements outlined in Section 4.2. Further, an individual is associated with a *fitness* value, representing a measure of quality of the individual and determined by an application-specific function; we describe the one we use in Section 4.4.3.

The initial population is created by randomly generating a predefined number of individuals (30 in our current implementation). New populations are generated by applying genetic operators to selected individuals, identified by a *parent selection* procedure. The genetic operators are *crossover*, which recombines individuals, and *mutation*, which introduces random variations on individuals based on a predefined probabilistic criterion. Populations evolve iteratively through so-called generations; after each iteration, the fitness of each individual is recomputed and becomes the basis for parent selection in the next generation. Eventually, the process converges to a (near)-optimal solution compliant with the constraints; a global optimum is not guaranteed, but the effect of local optima is mitigated thanks to the random component in the parent selection and application of genetic operators, as described in Section 4.4.4.

Figure 4.2 illustrates our evolutionary optimization based on genetic algorithms.
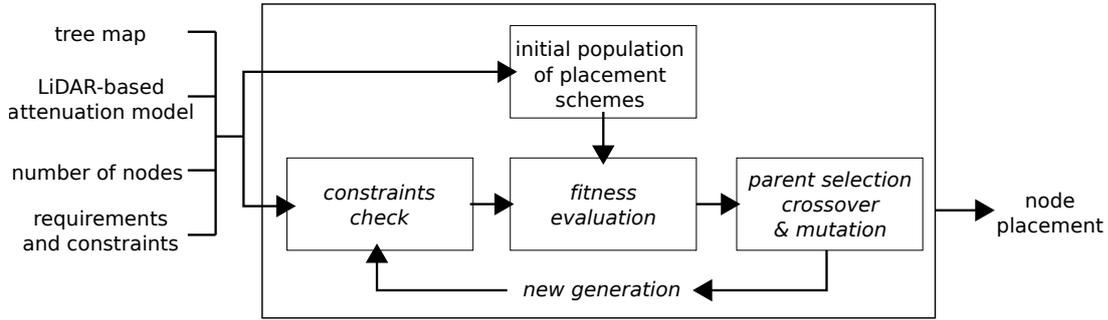
Figure 4.2: Evolutionary optimization of placement via genetic algorithms.

### 4.4.1 System model

We define a 2D optimization problem aimed at finding a (near-)optimal node placement in a target forest area. The positions of the nodes determine the communication quality, depending on the local forest characteristics. Even a small displacement of the nodes can significantly modify communication performance. By taking into account this observation, we propose an approach that is more fine-grained than those based on regular grids or patterns (e.g., regular triangle tessellation) [158, 24, 6].

The problem is defined starting from the deployment reference scenario (Section 4.2.1) and two initial variables: 1) the node set $\mathbb{N} = \{n_1, \cdots, n_N\}$, with cardinality $N = |\mathbb{N}|$, and 2) a target forest area $A$, for which LiDAR data are available.

By processing raw LiDAR data (Section 3.2.2), we first estimate position and diameter of tree trunks and then represent $A$ by a 2D *tree map* with size $h \times w$. A sample forest area $A$ and the corresponding *tree map* are shown in Figure 4.3a and 4.3b, respectively. A uniform quantization is applied to $A$, with quantization cells defined by the resolution of the tree map (i.e., $50 \times 50$ cm). The size of the quantization cell represents the spatial granularity we adopt in the analysis, with one quantization cell approximating one possible position where a node can be placed, if the constraints are fulfilled.

Then, we consider $N$ target spatial tiles, where a *tile* is an area in which one and only one node is allowed to be placed. The size of the spatial tile depends on the reference scenario and on the application requirements (Section 4.2.1):

- REGULAR: the target area is divided into $N$ equal-size square spatial tiles. Their size is application-dependent and chosen to cover the whole target area with $N$ nodes. Figure 4.3c shows an example of 9 tiles covering a sample target area, along with trunk positions derived from the tree map.

- IRREGULAR: the $N$ spatial tiles are sub-areas of $A$ arbitrarily shaped and sized according to the application requirements. An example is reported in Figure 4.3d.

(a) Forest area



(b) Tree map



(c) 9 equal-size square spatial tiles and trunk positions



(d) 9 arbitrary spatial tiles and trunk positions
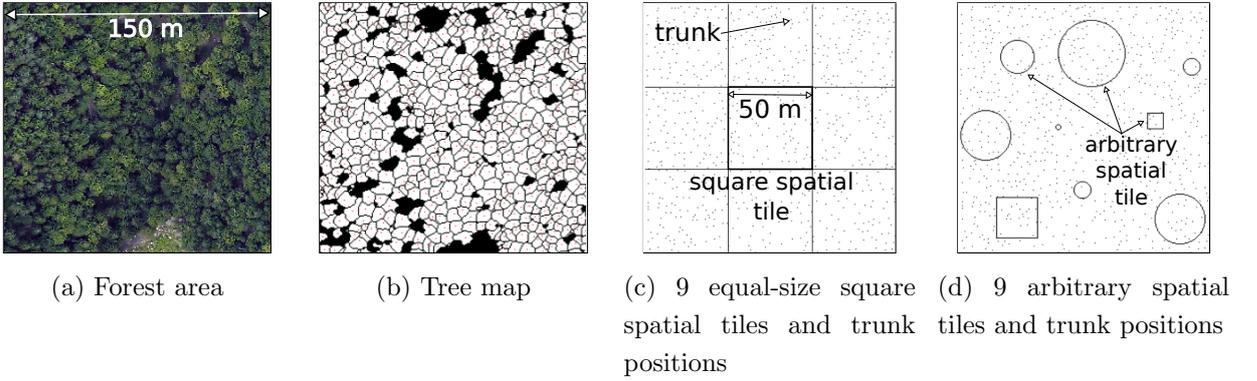
Figure 4.3: Division of a sample forest area into 9 spatial tiles (equal-size square spatial tiles for REGULAR scenario and arbitrary spatial tiles for IRREGULAR scenario).

The design variables are the positions of each network node in the target area. The set of these positions describes the overall node placement, encoded in a 2D scalar vector

$$\mathbb{P} = \{(x_1, y_1), (x_2, y_2), \cdots, (x_i, y_i), \cdots, (x_N, y_N)\}$$

where the position of a node $n_i$ is defined by its scalar coordinates $(x_i, y_i)$ in the target area.

A placement $\mathbb{P}$ implicitly identifies also a set of links $\mathbb{L}(\mathbb{P})$. A network link $l_{ij} \in \mathbb{L}(\mathbb{P})$ connects node $n_i$ in position $(x_i, y_i)$ with $n_j$ in $(x_j, y_j)$, $i \neq j$. $\mathbb{L}(\mathbb{P})$ represents all links connecting the nodes in $\mathbb{P}$, and therefore $|\mathbb{L}(\mathbb{P})| = \frac{N(N-1)}{2}$. In practice, however, only a subset of these links are interesting towards our problem, i.e., those whose nodes are in communication range. We denote with $P_{rx}(l_{ij})$ the expected received signal power for link $l_{ij}$, computed based on the LINK approach (Section 3.2.3) as a function of the distance between $n_i$ and $n_j$, the presence of trunks on the link line-of-sight, and the transmission power. $P_{rx}$ is at the core at one of the constraints of our placement problem, as described next.

### 4.4.2 Constraints

Along the lines of Section 4.2, we define separately the spatial and network constraints characterizing the placement problem. Unless otherwise noted, the constraints hold $\forall i, j \in \{1, \cdots, N\}$. Interestingly, these constraints are fundamental not only to capture application and system requirements, but also to narrow the search space and consequently reduce the computational overhead.

**Spatial constraints.** We impose the presence of *exactly one node in each target spatial tile*:

$$A_1: \quad (x_i, y_i) \in tile_i \tag{4.1}$$

regardless of the shape of the tile (i.e., a square for REGULAR, an arbitrary shape for IRREGULAR).

Apart from this fundamental constraint, others can be defined that further control the placement. For instance, we introduce the following constraints for REGULAR:

$$A_{21}: \quad dist((x_i, y_i), tile_i) \geq B \tag{4.2}$$

$$A_{22}: \quad dist((x_i, y_i), (x_j, y_j)) \geq D \tag{4.3}$$

where *dist* returns the Euclidean distance[1], and $B$ and $D$ are threshold values. Essentially, these constraints aim at preserving some degree of uniform coverage: $A_{21}$ avoids *node positions too close to the tile boundary*, and $A_{22}$ imposes a *minimum distance between nodes*.

Similar constraints can be defined for the IRREGULAR scenario, as shown in Figure 4.3d. However, their precise definition is not particularly interesting, and therefore omitted. In general, alternative and application-dependent constraints are possible. For example, in the context of a forest fire detection application it may be desirable to place temperature sensors where trees are relatively dense and, alternatively, it may be better to measure wind speed in clearings. These constraints can be easily formulated based on the tree map, and automatically accounted for in the companion tool.

**Network constraints.** We formalize the requirements in Section 4.2.2 with the following constraints:

$$C_1: \quad P_{rx}(l_{ij}) \geq P \tag{4.4}$$

$$C_2: \quad |\mathbb{L}_k| \geq M, \quad \mathbb{L}_k = \{l_{ij} \in \mathbb{L}_{acc}(\mathbb{P}) \mid i = k\}\} \tag{4.5}$$

$C_1$ sets the *minimum expected received power* $P$ that defines when a link yields acceptable communication quality; we investigate the appropriate value of $P$ in Section 4.5.1. $C_2$ defines the *minimum number $M$ of well-connected neighbors* required for each node.

The definition of the latter constraint relies on the set of *acceptable links*, which in principle contains all links satisfying $C_1$, $\mathbb{L}_{acc}(\mathbb{P}) = \{l_{ij} \in \mathbb{L}(\mathbb{P}) \mid C_1\}$. In practice, however, an additional constraint is needed to set the *minimum distance $T$ of a node from a trunk, along the link line-of-sight*:

$$C_3: dist((x_i, y_i), (x_t, y_t)) \geq T \wedge dist((x_j, y_j), (x_t, y_t)) \geq T, \quad \forall(x_t, y_t) \in trees(l_{ij}) \tag{4.6}$$

---

[1]For the sake of simplicity we slightly abuse the notation by assuming that it also returns the (minimum) distance of $(x_i, y_i)$ from any of the sides of $tile_i$.

where $trees(l_{ij})$ returns the position of all tree trunks that intersect link $l_{ij}$. The set of acceptable links is accordingly defined as

$$\mathbb{L}_{acc}(\mathbb{P}) = \{l_{ij} \in \mathbb{L}(\mathbb{P}) \mid C_1 \wedge C_3\} \tag{4.7}$$

The additional constraint $C_3$ is necessary to exclude from the candidate placement situations where tree trunks are on the link line-of-sight and very close to one of the communicating nodes. In these extreme cases, communication quality is significantly degraded; further, this significant degradation is also less accurately captured by the propagation model. We further elaborate on these topics and provide a value for $T$ in Section 4.5.2.

Note how $C_3$ holds regardless of the node spatial placement, i.e., FREE vs. TRUNK, as it is a condition on the *link* and not on the *node*. Specifically, recall from Section 4.2.1 that in a TRUNK scenario nodes are latched on trees, and their distance from the trunk is therefore 0 m. As a consequence, given a node, all links "behind" the trunk it is attached to violate $C_3$ and are discarded, while all links "in front" are candidates for further evaluation.

### 4.4.3 Fitness Function

We now illustrate the definition of the fitness function $F(\mathbb{P})$ that is the basis for evaluating and comparing the candidate placement solutions fulfilling the constraints.

We define $F(\mathbb{P})$ based on two components, which descend from the requirements and goals in Section 4.2. The first one is the total number of acceptable links in the network

$$L_{acc} = |\mathbb{L}_{acc}(\mathbb{P})| \tag{4.8}$$

while the second component is the average expected received power across all acceptable links

$$\overline{P}_{rx} = \sum_{l_{ij} \in \mathbb{L}_{acc}(\mathbb{P})} \frac{P_{rx}(l_{ij})}{|\mathbb{L}_{acc}(\mathbb{P})|} \tag{4.9}$$

Maximizing the first value increases the chances to build a network that is connected as well as robust, while maximizing the second increases the overall communication quality. As we show in our evaluation (Section 4.6.2), their joint optimization yields significantly better placements w.r.t. using each component alone. Therefore, we take both components into account in the fitness function:

$$F(\mathbb{P}) = L_{acc} + \lambda \overline{P}_{rx} \tag{4.10}$$

Both terms are normalized between 0 and 1, based on the minimum and maximum values attainable. A tuning parameter $\lambda$ determines which term is predominant and how

strongly. A preliminary analysis showed that the best performance is achieved when $\lambda = 1$, yielding a range $[0, 2]$ of variation for the fitness value. Finally, the link connecting node $i$ with node $j$ is counted only once both in $L_{acc}$ and $\overline{P}_{rx}$, i.e., links $l_{ij}$ and $l_{ji}$ are not distinguished.

### 4.4.4 Parent Selection and Genetic Operators

An individual $(\mathbb{P}, F(\mathbb{P}))$ is represented by the combination of a candidate placement $\mathbb{P}$ and the corresponding value of the fitness function, $F(\mathbb{P})$. New populations are generated by iteratively applying a parent selection procedure and a genetic operator, and recomputing the fitness value.

The *parent selection* procedure selects parent individuals for reproduction, to generate offsprings. We apply the *binary tournament selection* [127] approach, where two individuals are selected at random and their fitness compared; the individual with better fitness is selected as parent. Tournament selection is executed until a pool of parents of predefined size (e.g., half the population) is selected.

The *genetic operator* performs *crossover* and *mutation* [127] on the pool of selected parents to produce offsprings. Specifically, we execute *one-point crossover* with probability 0.9 and *uniform mutation* with probability 0.1. The former randomly selects two different parents, randomly selects a crossover point (i.e., an index in the vector $\mathbb{P}$ of positions associated to an individual), and swaps the node positions beyond this point between the two parents. The latter applies the following mutation to a single parent chosen at random: each node position in its $\mathbb{P}$ is replaced, with probability $\frac{1}{N}$, with another position in the same spatial tile, uniformly chosen at random among acceptable ones.

These probability values favor the exploration of the possible combinations of node positions currently considered as parents before introducing new random ones. The offsprings are then checked against constraints; individuals that do not comply with the constraints are discarded, while the fitness value is recomputed for the others. The latter compliant offsprings are merged with their parents into an *intermediate population*, from which a number (equal to the population size) of individuals with the best fitness is selected, and a new generation is created. Elitism is included, i.e., the current best individual of each generation always survives into the next one.

## 4.5 Configuring Spatial and Network Constraints

The node placement approach in Section 4.3 is based on constraints modeling spatial and system requirements, which are in turn based on configuration parameters whose values we discuss here.

We devote particular attention to constraints $C_1$ and $C_3$ in Section 4.5.1 and Section 4.5.2, as determination of the corresponding thresholds $P$ and $T$ strictly depends

| constraint | parameter | value | definition |
|:---:|:---:|:---:|:---|
| $C_1$ | $P$ | $-85$ dBm | minimum expected received power at a node |
| $C_2$ | $M$ | 3 | minimum number of acceptable links per node |
| $C_3$ | $T$ | 5 m | minimum trunk-node distance |
| $A_{21}$ | D | 10 m | minimum distance of a node from the tile border |
| $A_{22}$ | B | 25 m | minimum link length |

Table 4.1: Configuring application and network constraints.

on experimental evidence. We offer quantitative considerations based on our own experimental setup that, due to the characteristics of the hardware and environment, is likely to cover a broad spectrum of applications. On the other hand, the methodology we describe can be used to replicate ad hoc examples for different hardware and/or forest environments, enabling one to easily determine the appropriate values for $P$ and $T$.

Finally, in Section 4.5.3 we discuss the other parameters, which instead depend on generic network and spatial requirements. The values chosen, used in the evaluation (Section 4.6), are meant solely to exemplify the flexibility and usefulness of our tool. Table 4.1 provides a summary of these values in the context of the constraints in which they are used.

### 4.5.1 Determining the Minimum Expected Received Power

Constraint $C_1$ relies on a threshold $P$ meant to filter out communication links likely to be too unreliable. In ideal conditions, the value of $P$ is simply the receiver sensitivity threshold. For example, the CC2420 datasheet specifies a value of $-94$ dBm; indeed, this was the minimum power level measured for received packets in the experimental campaign we describe in Section 4.5.2.

However, outdoor environments are far from ideal, due to several environmental factors (e.g., temperature and humidity or multipath effects) that affect the radio signal; using the value above is known to lead to unreliable results, as shown by several empirical studies (e.g., [99, 142, 15, 159]). Small signal variations can cause abrupt changes in the ability to receive packets, rendering links unstable. The metric commonly used in these studies is Packet Receipt Ratio ($PRR$), computed as the number of packets received on a link over the number of those sent. It has been shown [159] that when the reception power is close to the receiver sensitivity threshold, links belong to a transitional region with highly variable $PRR$; instead, when the reception power is well above threshold, links belong to a connected region where they exhibit high $PRR$. Hereafter, we aim at ensuring the highest communication quality, therefore retaining only links that belong to the connected region. Other, less conservative choices are easily supported by setting a different value for the threshold $P$.

49

The relation between *PRR* and *RSSI* has been analyzed in the literature in several experimental observations. In [80], for instance, a good *PRR* is observed for an average *RSSI* > −87 dBm. The measurements we report in Section 4.5.2 confirm these findings. Figure 4.4 shows the *PRR* as a function of the average received power, computed every 100 packets sent. The charts shows no packet loss for *RSSI* ≥ −85 dBm; we therefore select $P = -85$ dBm as the threshold for $C_1$.
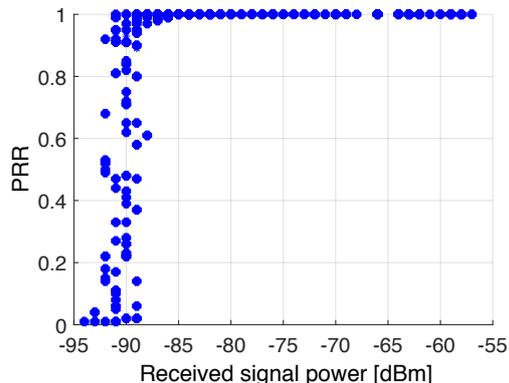


Figure 4.4: *PRR* vs. average received power.

### 4.5.2 Determining the Minimum Trunk-Node Distance

Constraint $C_3$ specifies the minimum distance $T$ between the nodes composing a link and the trees on the link line-of-sight. Indeed, when nodes are too close to a tree, the attenuation induced by the latter increases abruptly; further, this increase is not captured accurately by the model in Section 3.2.3.

To determine the threshold value $T$ beyond which these phenomena occur, we run dedicated in-field experiments in Monte Bondone, near Trento, using TMote Skymotes as in the validation of the LiDAR-based radio attenuation model in Section 3.3. The experimental setup is described in Figure 4.5. We select an isolated trunk and place two nodes at different distances on its opposite sides; the trunk is therefore on the line of sight of the communication link between the nodes. We consider a maximum link



Figure 4.5: Determining the minimum trunk-node distance $T$: experimental setup.

length of 60 m, and explore different positions of the trunk inside the link by varying *i)* the distance of the tree from one of the nodes, and *ii)* the link length, i.e., the distance between the two nodes. More precisely, the trunk "moves" w.r.t. one of the nodes in 5 m increments, from 0 to 30 m; at the latter distance, the trunk is in the center of the link. The link length varies from 5 to 60 m; this is achieved in 5 m increments, except for lengths > 40 m, for which the increment is 10 m. For each configuration of trunk distance and link length, 100 packets are sent by each node by alternating its packet transmissions (TX) at 1 pkt/s with the other node; overall, one packet is sent on the link every 500 ms. We repeated this process with two TX powers, −1 and −8 dBm, obtaining similar results; here, we report only those with the higher TX power. Moreover, we repeated experiments

(a) Trunk diameter: 20 cm.

(b) Trunk diameter: 40 cm.
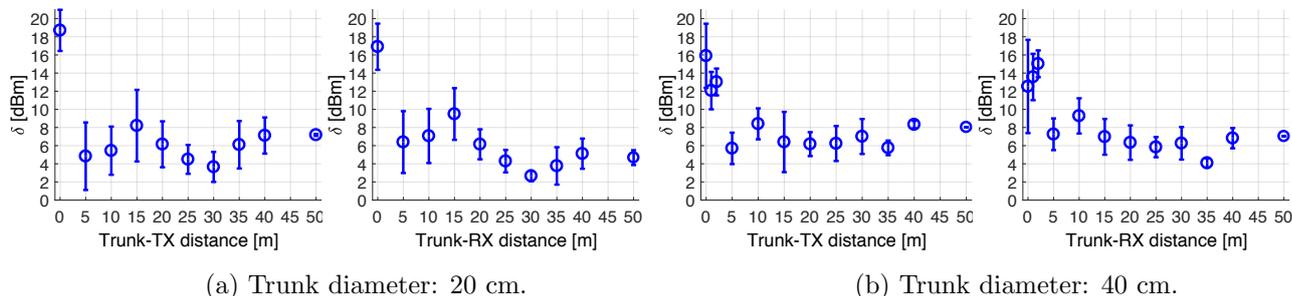
Figure 4.6: Average $\delta$ and standard deviation of the error between the attenuation model in Section 3.2.3 and our measurements when a tree trunk is near to one of the communicating nodes.

with two isolated trees of different diameter, 20 and 40 cm. Overall, we collected ≈29600 data points.

We report the average and standard deviation of the error $\delta$ between our measurements and the estimates of the LINK model in Section 3.2.3, as a function of the distance $d$ between the trunk and the communicating node; further, we separate the cases in which the latter node (which can be on either side of the tree) is transmitting or receiving. $\delta$ is computed for a given distance $d$ on *all* the link lengths involved; e.g., the value $\delta$ computed for $d = 5$ m in TX mode is the average of the values for all links in which *either* node is transmitting at 5 m from the trunk. However, each *individual* model estimate depends on the link length; therefore, we compare each link measurement against its corresponding distance-dependent estimate, and obtain the error averaged across all link lengths. The specific diameter at hand is similarly accounted by the model for individual estimates.

Figure 4.6 shows the results. We observe that the communicating status of a node does not affect significantly the estimation error: whether the trunk is at distance $d$ from a transmitter or a receiver, the effect is approximately the same. On the other hand, the trunk diameter affects the relationship between estimation error $\delta$ and distance $d$; however, the error $\delta$ is overall in line with the results we obtained in Section 3.3, except when $d < 5$ m. In this case, our model significantly overestimates the received power by failing to account for the severe attenuation induced by the trunk, regardless of its diameter. The experiments with trunk diameter 40 cm show that this strong attenuation holds not only when the trunk is immediately in front of the communicating node ($d = 0$ m) but also at distance 1 and 2 m; unfortunately, we were not able to gather additional measurements due to logistical reasons.

Based on these considerations, hereafter we conservatively set $T = 5$ m in constraint $C_3$.

### 4.5.3 Other Parameters

The other parameters in Table 4.1 are set as follows. The minimum number $M$ of acceptable links for a given node ($C_2$) is key to build a robust network, as it bears a direct impact on the connectivity of the network graph, analyzed quantitatively in Section 4.6. We set $M = 3$, based on the consideration that lower values may easily lead to the creation of several "branches" off the graph. However, in Section 4.6.4 we also show that relaxing this constraint may be convenient to cope with scale.

As for spatial constraints, we set in $A_{21}$ the minimum distance of a node from the tile border to $B = 10$ m, and in $A_{22}$ the minimum link length to $D = 25$ m. These values, respectively ⅕ and ½ of the tile side, are mostly meant to be illustrative, and clearly depend on the application at hand. Further, as discussed in Section 4.4.2, we show the constraint definition only for the case of a REGULAR placement; IRREGULAR ones can be similarly defined.

## 4.6 Evaluating the Quality of Node Placement

We now evaluate the performance of our toolchain, configured as described in Section 4.5. We consider spatial constraints stemming the uniformity of the deployment (REGULAR vs. IRREGULAR) and the mechanics of node positioning (FREE vs. TRUNK), as described in Section 4.2. We exploit the same LiDAR dataset acquired for the forest location described in Section 3.3, in which we consider the $150 \times 150$ m$^2$ area shown in Figure 4.3a, with center at N 46° 0′ 48.546″, E 11° 25′ 52.122″.

We define the performance metrics and comparison baselines in Section 4.6.1, followed in Section 4.6.2 by a quantitative analysis of the fitness function $F(\mathbb{P})$ confirming that the linear combination of $\overline{P}_{rx}$ and $L_{acc}$ indeed yields better results w.r.t. the independent use of these components. We then evaluate the performance of our approach first in a 9-node deployment inside the aforementioned area (Section 4.6.3), followed by a 36-node deployment that allows us to investigate the effect of scaling up the network (Section 4.6.4). Finally, we offer a summary of the evaluation and a discussion of the main benefits of LaPS (Section 4.6.5).

### 4.6.1 Performance Metrics and Comparison Baselines

**Performance metrics.** We analyze quantitatively the networks output by our placement tool based on the same *fundamental metrics* at the core of our optimization approach (Section 4.2.3), i.e., the total number $L_{acc}$ of acceptable links and the average expected received power $\overline{P}_{rx}$ across them.

Moreover, we also evaluate the *robustness* of these network from a *topological* standpoint *i)* based on well-known *connectivity* and *centrality* metrics from graph theory, and

*ii)* in terms of average number of paths to the sink. The latter is a relevant metric in, e.g., data collection applications based on a tree overlay, as it quantifies the options available to a routing protocol in building (and reconfiguring) routes from each source to the sink.

The first type of metrics deserves further explanation. Connectivity is defined as the minimum number of elements (i.e., nodes or links) that, if removed, disconnect the remaining nodes from each other; the higher the connectivity, the more robust the network. To measure this aspect, we report the *average number* $|\overline{\mathbb{L}_k}|$ *of neighbors* with acceptable quality, and compute [47]: *i)* *binary connectivity* (or *connectedness*) $k$, whose value is 1 if at least one path exists between all pairs of nodes, 0 otherwise; *ii)* *vertex connectivity* $k_v$, and *iii)* *edge connectivity* $k_e$, respectively denoting the minimum number of vertexes and edges that, if removed, disconnect the graph.

An alternative view on robustness is offered by the notion of *betweenness centrality*, defined as

$$b_x = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{s_{ij}(x)}{s_{ij}}$$

where $s_{ij}(x)$ is the number of shortest paths between $i$ and $j$ passing through $x$ and $s_{ij}$ is the total number of shortest paths between $i$ and $j$ [52]. We actually use the value normalized between 0 and 1, obtained by dividing $b_x$ by the total number $\frac{(N-1)(N-2)}{2}$ of node pairs ($x$ excluded). Betweenness centrality quantifies the extent to which each node lies on the shortest paths between other nodes, and therefore expresses the disruption induced by the removal of such node on the communications among other nodes. We report directly this metric and also exploit it in a force-based representation of the network layout which visually conveys information about the fragility of the network.

**Comparison baselines.** As mentioned in Section 4.1, existing placement approaches for WSNs neglect the impact of the environment on communication, let apart taking into account the presence of trees. Therefore, we compare the solutions found by our approach against two closely-related baselines: *i)* a *blind regular* placement where nodes are organized in an exact grid without taking into account the forest characteristics, and *ii)* a *line-of-sight* placement in which no link is obstructed by tree trunks, and therefore the standard path loss model can be applied (Section 3.3).

The blind regular placement is a common choice in the literature [24, 23, 54, 158] as it is intuitive and, in principle, of straightforward application in real deployments. Comparing against this placement strategy allows us to assess whether detailed knowledge of the tree positions is at all useful.

In contrast, the line-of-sight placement exploits precisely this information, derived from LiDAR data via our toolchain, towards a different goal. Indeed, this variant uses alternate constraints w.r.t. those in Section 4.4.2, aimed at selecting only those network

topologies whose links are unobstructed by trees; in this situation, the specialized radio attenuation model in Section 3.3 becomes superfluous, and the standard path loss model can be used instead. Therefore, comparing against this line-of-sight placement allows us to assess whether a specialized radio model is truly necessary. Interestingly, however, even in the case of a negative answer the proposed toolchain still constitutes a novel asset, enabling the selection of the topologies, if any, whose links enjoy a free line of sight.

### 4.6.2 A Closer Look at the Fitness Function

We begin our evaluation by providing quantitative evidence that our definition of the fitness function $F(\mathbb{P})$ as a linear combination of $\overline{P}_{rx}$ and $L_{acc}$ yields better performance than using either component alone. To better elicit trends, we refer to the scenario REGULAR/FREE and neglect the spatial constraints $A_{21}$ and $A_{22}$. Figure 4.7 and 4.8 show the evolution of the genetic search and the resulting network layouts when $F(\mathbb{P})$ is $\overline{P}_{rx}$, $L_{acc}$, or their linear combination in Eq. 4.10.

By optimizing only the average reception power $\overline{P}_{rx}$ we obtain, as expected, a placement with very high overall communication quality ($\overline{P}_{rx} = -71.95$ dBm) in only 121 generations; Figure 4.7a shows the corresponding evolution of the genetic search. However, the price to pay is that the total number of acceptable links is only[2] $L_{acc} = 16$. Figure 4.7b shows the evolution of $L_{acc}$ during the search; we observe that this value actually decreases as better configurations optimizing $\overline{P}_{rx}$ are found. Moreover, the resulting network layout in Figure 4.8a shows two node clusters connected by only one link; constraint $C_2$ on the minimum number of neighbors is satisfied, although by means of a very fragile topology prone to partitioning. On the other hand, by optimizing only the number of acceptable links, we more than double their number ($L_{acc} = 36$, Figure 4.7b) after only 84 generations, and remove the clustering effect from the resulting network topology (Figure 4.8b). However, this time the price to pay is a significant decrease in the overall link quality ($\overline{P}_{rx} = -76.88$ dBm, Figure 4.7a).

The definition of $F(\mathbb{P})$ in Eq. 4.10, which combines $\overline{P}_{rx}$ and $L_{acc}$ with equal weight, strikes a good balance between the two extremes above. Figure 4.7c reports the value of $F(\mathbb{P}) \in [0, 2]$ (Section 4.4.3), monotonically increasing as generation elapse, while Figure 4.7a and 4.7b show separately the evolution of its two components. The best solution is found after 136 generations, yielding $L_{acc} = 35$ and $\overline{P}_{rx} = -75.43$ dBm; one link less w.r.t. optimizing only $L_{acc}$, but with slightly higher overall power. Further, Figure 4.8c shows that the clustering effect observed when optimizing only $\overline{P}_{rx}$ is absent here, as when optimizing $L_{acc}$. However, in comparison with Figure 4.8b, we observe a marked tendency to concentrate the nodes in the center of the target area and place them

---

[2]Remember from Section 4.4.3 that links $l_{ij}$ and $l_{ji}$ are not distinguished, and counted as one.

(a) Average received signal power  (b) Number of acceptable links  (c) Fitness function

Figure 4.7: Evolution of the genetic search for $\overline{P}_{rx}$ (black), $L_{acc}$ (orange), and their linear combination (black dashed). Curves show the performance metric associated to the best individual in the corresponding generation.



(a)          (b)          (c)

Figure 4.8: Network layouts obtained by optimizing (a) $\overline{P}_{rx}$ (b) $L_{acc}$ (c) their linear combination ($\lambda = 1$). Black dots are trunk positions, red squares are node positions, black lines are acceptable links.

very close to each other, as this increases the quality of the resulting links. Nevertheless, this undesirable effect is mitigated precisely by the spatial constraints $A_{21}$ and $A_{22}$ we neglected here, but consider in the following sections.

### 4.6.3 Analyzing and Comparing Node Placements

We now analyze the performance of our placement approach when applied to a 9-node network similar to the one we used in Section 3.3, and compare it to the blind regular and line-of-sight baselines defined in Section 4.6.1. We consider all scenario combinations of REGULAR vs. IRREGULAR and FREE vs. TRUNK placement, subject to all constraints in Section 4.2. For IRREGULAR, as shown in Figure 4.10b we defined a mix of circular and squared regions examplifying spatial tiles required by the application, along with a point $F$ representing a geo-referenced position for which no spatial tolerance is allowed. Note that the specialized radio model defined in Section 3.2.3 has been validated under

the following experimental conditions: *i)* absence of snow or rain; *ii)* low temperature and humidity; *iii)* absence of leaves in the understory vegetation; and *iv)* almost flat terrain morphology. These conditions implicitly define the assumptions underpinning the evaluation discussed here.

**Core metrics.** Table 4.2 shows the final outcome in terms of fitness value $F(\mathbb{P})$, total number of acceptable links $L_{acc}$, and average received power $\overline{P}_{rx}$ on these links, while Figure 4.9 shows the evolution of these values over generations. We observe that the evolution of these metrics is comparable across the various scenarios. The fitness value (Figure 4.9a) is similar in all cases; the highest value of $F(\mathbb{P}) = 1.39$ is achieved in the case REGULAR FREE, i.e., the least constrained scenario. Similarly, the resulting topologies have the same number of links in the network ($L_{acc} = 34$) and, on average, per node ($|\overline{\mathbb{L}_k}| = 7.55$). This is somewhat surprising considering that the constraints of the four scenarios considered are quite different and, as shown in Figure 4.10a–4.10b and Figure 4.11a–4.11b, yield network layouts of different shape. On the other hand, the different complexity of the four scenarios is reflected to some extent in the average received power, which is higher in REGULAR and in FREE w.r.t. their IRREGULAR and TRUNK counterparts. Interestingly, the value of $\overline{P}_{rx}$ remains relatively stable across generations (Figure 4.9c), unlike the value of $L_{acc}$ that increases rapidly (Figure 4.9b); this is likely due due to the constraint $A_{22}$ on the minimum link length distance.

In comparison, the commonly-used and intuitive blind regular placement yields significantly worse results. We obtain this placement by forcing each node exactly in the barycenter of each tile when allowed by a FREE placement or, for a TRUNK placement, next

|  |  | $F(\mathbb{P})$ | $L_{acc}$ | $\overline{P}_{rx}$ | $|\overline{\mathbb{L}_k}|$ | $k$ | $k_v$ | $k_e$ |
|---|---|---|---|---|---|---|---|---|
| **LaPS** | | | | | | | | |
| REGULAR | FREE | 1.39 | 34 | -77.58 | 7.55 | 1 | 6 | 6 |
| | TRUNK | 1.34 | 34 | -78.32 | 7.55 | 1 | 6 | 6 |
| IRREGULAR | FREE | 1.36 | 34 | -78.05 | 7.55 | 1 | 6 | 6 |
| | TRUNK | 1.31 | 34 | -78.79 | 7.55 | 1 | 6 | 6 |
| *blind regular* | | | | | | | | |
| REGULAR | FREE | — | 17 | -78.37 | 3.77 | 1 | 2 | 2 |
| | TRUNK | — | 11 | -79.62 | 2.44 | 0 | 0 | 0 |
| *line-of-sight* | | | | | | | | |
| REGULAR | FREE | 1.29 | 32 | -77.67 | 7.11 | 1 | 6 | 6 |
| | TRUNK | 1.15 | 28 | -76.91 | 6.22 | 1 | 5 | 5 |
| IRREGULAR | FREE | 1.29 | 31 | -76.98 | 6.88 | 1 | 5 | 5 |
| | TRUNK | 1.11 | 28 | -77.57 | 6.22 | 1 | 5 | 5 |

Table 4.2: Core metrics for all scenario combinations and baselines: 9 nodes.

to the tree closest to the barycenter. However, the blind regular placement guarantees the fulfillment of $C_2$ neither in FREE nor in TRUNK; the value of $F(\mathbb{P})$ cannot therefore be computed. This is shown also in the resulting network layouts of Figure 4.10c and 4.11c; in the TRUNK case, one of the nodes does not even belong to the network, and other two are connected via a single link. Finally, the acceptable links are only one half and one third of those obtained by our approach in FREE and TRUNK, respectively, and with a lower average power in both cases.

In contrast, the performance of the line-of-sight placement is significantly better and approaches that of LaPS. The average received power $\overline{P}_{rx}$ is actually higher than LaPS (except for the REGULAR FREE case), since the line-of-sight placement enjoys communication links unobstructed by trees. On the other hand, the number of acceptable links is lower, both for the network as a whole and for individual nodes; the resulting topologies are shown in Figure 4.10d and 4.11d. Finally, it is interesting to compare the evolution of fitness value for LaPS and line-of-sight (Figure 4.12). In the FREE case, the value of $F(\mathbb{P})$ for the latter mirrors the one of the former, despite reaching only slightly worse performance at convergence. However, in the more restrictive TRUNK case, the search problem becomes more complex and the genetic optimization is unable to find solutions fulfilling the constraints for more than 300 generations.
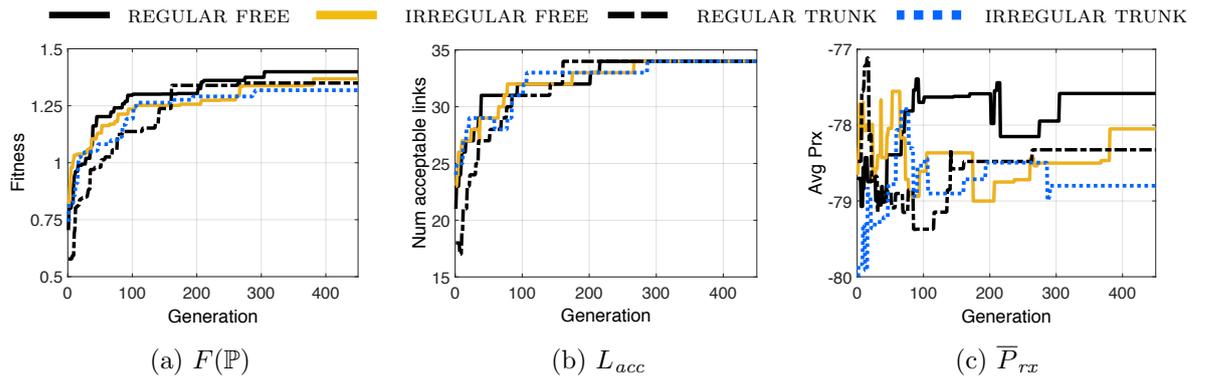


Figure 4.9: Evolution of the fitness value and its components in all scenario combinations.

(a) $F(\mathbb{P})$  (b) $L_{acc}$  (c) $\overline{P}_{rx}$

| (a) REGULAR | (b) IRREGULAR | (c) blind regular | (d) line-of-sight |
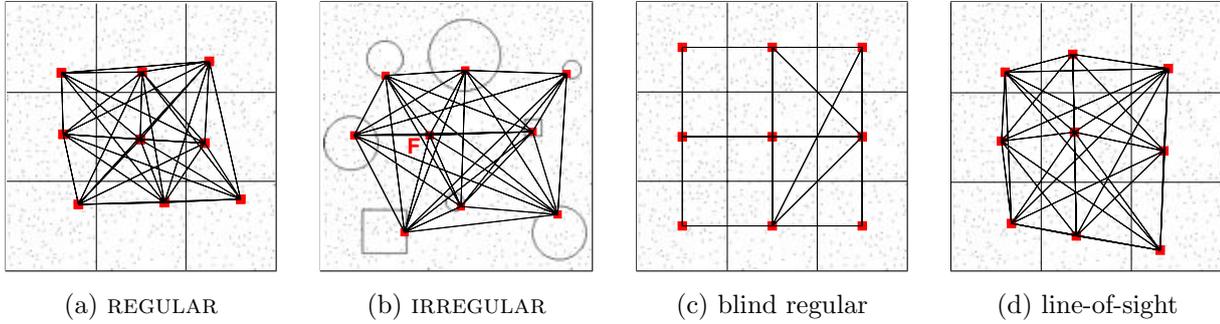
Figure 4.10: Network layouts obtained with FREE node placement.



| (a) REGULAR | (b) IRREGULAR | (c) blind regular | (d) line-of-sight |
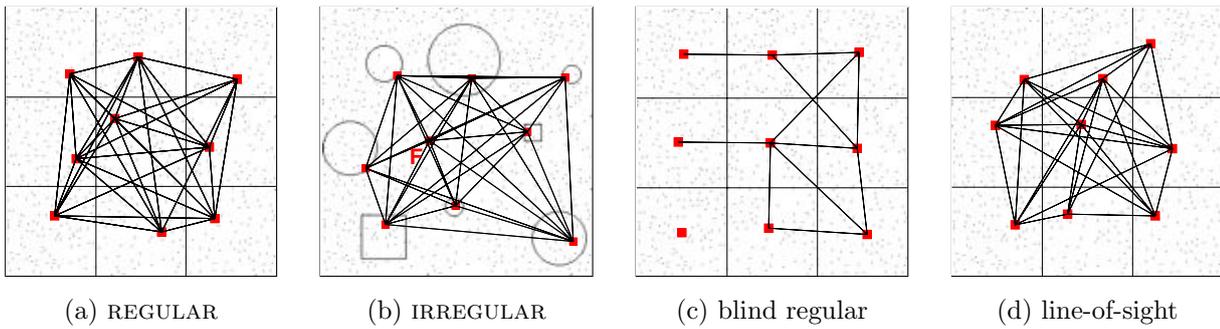
Figure 4.11: Network layouts obtained with TRUNK node placement.



| (a) $L_{acc}$ | (b) $\overline{P}_{rx}$ | (c) $F(\mathbb{P})$ |
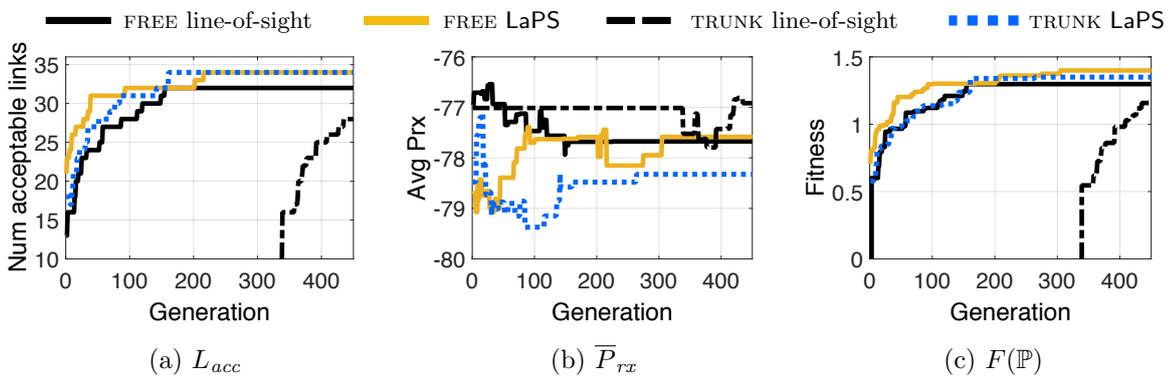
Figure 4.12: Evolution of the value of the fitness function and of its components: LaPS vs. line-of-sight in a REGULAR global placement with FREE and TRUNK node placement.

**Connectivity metrics.** From Figure 4.10 and 4.11 we can visually ascertain that the resulting network topologies are "well connected". We now quantitatively assess this aspect, which significantly affects the robustness of the network, by exploiting the graph-based metrics outlined in Section 4.6.1, whose values are shown in Table 4.2.

LaPS always yields connected networks ($k = 1$); to partition them, at least 6 nodes or links ($k_v = k_e = 6$) must fail simultaneously. In contrast, the blind regular placement yields a very vulnerable network in the FREE case, where the loss of only two nodes or links is sufficient to cause a partition; further, the network is not connected in the TRUNK case, as already noted. The line-of-sight placement yields networks that are only marginally less connected than those generated by LaPS. In the REGULAR FREE case the only difference is a slightly lower number of neighbors ($|\overline{\mathbb{L}_k}| = 7.11$), while in the other scenarios only 5 failing nodes or links suffice to partition.

Figure 4.13 offers an alternative view based on the value of the normalized betweenness centrality. We can observe that this value is similar for all nodes of the networks generated by LaPS, yielding a very robust topology; on the contrary, the blind regular placement yields highly unbalanced topologies in which a handful of nodes are critical, both in terms of reliability (e.g., causing a partition upon crash) and performance (e.g., as potential routing bottlenecks). The corresponding analysis for the line-of-sight placement yields values similar to LaPS, and is therefore omitted. On the other hand, the performance of the two approaches diverges when the network scale is increased, as discussed in Section 4.6.4.

**Paths to the sink.** We conclude our analysis by investigating the effectiveness of placement solutions from a different point of view, closer to the routing layer, and explicitly compare the number of attainable paths to a sink node. Indeed, the more paths available the more options a protocol can exploit to build and reconfigure the routing topology, therefore increasing resilience to failures and link dynamics. For each network layout



(a) LaPS FREE     (b) blind regular FREE     (c) LaPS TRUNK     (d) blind regular TRUNK
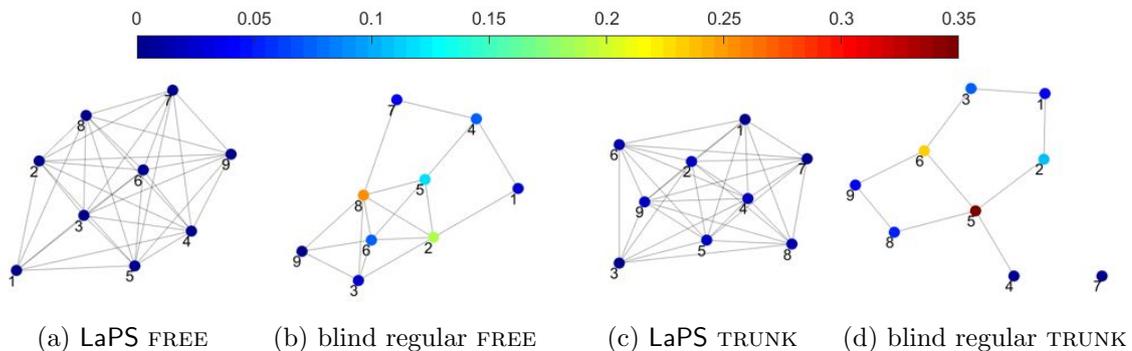
Figure 4.13: Betweenness centrality. The graphs layout emphasize readability and do not reflect the real topology.

|  |  | sink | #paths | #paths $\leq 3$ hops | #paths $\leq 2$ hops | sink | #paths | #paths $\leq 3$ hops | #paths $\leq 2$ hops |
|---|---|---|---|---|---|---|---|---|---|
| LaPS | | | | | | | | | |
| REGULAR | FREE | central | 65890 | 350 | 60 | upper left | 82200 | 300 | 48 |
| | TRUNK | central | 52516 | 324 | 58 | upper left | 59105 | 307 | 53 |
| IRREGULAR | FREE | central | 74371 | 331 | 55 | upper left | 76705 | 321 | 51 |
| | TRUNK | central | 68108 | 348 | 60 | upper left | 69206 | 332 | 58 |
| **blind regular** | | | | | | | | | |
| REGULAR | FREE | central | 390 | 59 | 19 | upper left | 511 | 27 | 8 |
| | TRUNK | central | 19 | 12 | 8 | upper left | 23 | 9 | 4 |
| **line-of-sight** | | | | | | | | | |
| REGULAR | FREE | central | 42078 | 298 | 56 | upper left | 46536 | 280 | 50 |
| | TRUNK | central | 16320 | 204 | 44 | upper left | 18032 | 190 | 38 |

Table 4.3: Number of paths to the sink for the different placement solutions and scenarios: overall number of paths, overall number of paths with cost $\leq 3$ hops and $\leq 2$ hops.

hitherto examined we compute all possible (acyclic) paths from each node to one identified as a sink. We study two sink positions yielding different path lengths: in the center and in the upper left corner.

Table 4.3 reports the overall number of paths (i.e., the sum on all the possible start-nodes in the network) for both FREE and TRUNK placement modes and both REGULAR and IRREGULAR scenarios. In addition, we report and compare the overall number of short paths (i.e., composed of 2 and 3 hops). We can observe that LaPS yields the highest number of paths for all scenarios, both in global terms as well as by considering only short (and therefore in principle more desirable) paths. Interestingly, the number of available paths to a sink node is not directly accounted for in the genetic evolution. Therefore, this is an indirect result of the definition we adopted for the fitness function. However, the number of paths for the various scenarios do not necessarily follow a precise trend, as they are strongly dependent on the specific constraints (e.g., in IRREGULAR) or the particular displacement of tree trunks. The only clear trend is the expected increase in the number of paths when the sink is in the upper left corner. Finally, the aforementioned weakness of the blind regular placement is reflected also here, where the overall number of paths is two orders of magnitude lower then those provided by LaPS, while the line-of-sight placement consistently generates less paths to the sink, as a consequence of its slightly lower connectivity.

### 4.6.4 Scaling the Network

In the previous section we showed that, even in a small 9-node network, LaPS brings remarkable advantages w.r.t. the commonly-used blind regular placement. This confirms

that the knowledge and modeling of the forest structure, acquired via LiDAR data and the LaPS toolchain, plays a key role. On the other hand, the difference with a line-of-sight placement that leverages the same information to *avoid* trees, rather than taking their attenuation into account with our specialized radio model, was less marked.

In this section we show that the difference between these two placement approaches becomes substantial as the size of the target network increases.

**Settings.** We consider a 36-node network deployed over a target area of $300 \times 300$ m$^2$, i.e., a *fourfold* increase in both network size and target area. We assume the same spatial sampling granularity of $\sim$50 m used so far. Moreover, to minimize the bias induced by the different tree density and enable a direct comparison with the results in Section 4.6.3 we simulate the larger area by replicating and spatially recombining the 9 squared tiles composing the smaller target area previously used. The resulting target area (Figure 4.15) is four times larger than the previous one (Figure 4.10) but has, by design, the same tree density.

Hereafter, we focus only on the REGULAR scenario but consider both FREE and TRUNK variants. However, unlike Section 4.6.3, we do not report about the overall number of paths because, at this scale, an exhaustive search of all possible paths from each node to the sink cannot be performed in polynomial time (#P-complete problem).

**FREE node placement.** At the scale considered, the placement problem becomes significantly more challenging, even in the less constrained FREE case. Indeed, the *line-of-sight* approach is unable to find a solution in 1000 generations, as shown in Figure 4.14; the fitness value cannot be computed, and the genetic search essentially becomes a random search, since no solution fulfilling all the constraints is found. This is a consequence of our strict formulation of the genetic search, solutions unable to fulfill the constraints are discarded to avoid misguiding the evolution process.

In contrast, LaPS converges to a solution fulfilling all constraints, characterized by $L_{acc} = 161$ links and an average reception power of $\overline{P}_{rx} = -78.04$ dBm, as shown in Table 4.4. This demonstrates that, as the scale of the network and/or target area increases, it may become impossible to find a solution in which all links enjoy a free line of sight. LaPS removes this assumption and, thanks to its specialized radio model, can take the tree attenuation into account to identify the best topology.

The latter is shown in Figure 4.15a, evidencing a balanced structure without clusters. For comparison, Figure 4.15b reports the network layout corresponding to the line-of-sight placement providing the largest number of links, $L_{acc} = 79$. However, this topology fails to fulfill constraint $C_2$ about the minimum number of acceptable neighbors.

The connectivity metrics in Table 4.4 also show that the line-of-sight placement, although connected, is quite fragile: a single node or link failure is enough to cause a

|  |  | $F(\mathbb{P})$ | $L_{acc}$ | $\overline{P}_{rx}$ | $|\overline{\mathbb{L}_k}|$ | $k$ | $k_v$ | $k_e$ |
|---|---|---|---|---|---|---|---|---|
| **LaPS** | | | | | | | | |
| REGULAR | FREE | 0.56 | 161 | -78.04 | 8.94 | 1 | 5 | 5 |
| | TRUNK | 0.54 | 143 | -77.49 | 7.94 | 1 | 4 | 4 |
| **LaPS**, $M = 2$ | | | | | | | | |
| REGULAR | TRUNK | 0.53 | 141 | -77.48 | 7.83 | 1 | 4 | 4 |
| *line-of-sight* | | | | | | | | |
| REGULAR | FREE | — | 79 | -77.81 | 4.38 | 1 | 1 | 1 |
| | TRUNK | — | 68 | -77.51 | 3.77 | 0 | 0 | 0 |
| *line-of-sight*, $M = 2$ | | | | | | | | |
| REGULAR | TRUNK | 0.35 | 104 | -76.95 | 5.77 | 1 | 2 | 2 |

Table 4.4: Core metrics for all scenario combinations and baselines: 36 nodes.

partition. On the contrary, LaPS yields a rather robust network where $k_v = k_e = 5$ simultaneous failures are required to disconnect the network. This aspect can be better appreciated by comparing Figure 4.16a and 4.16b, which visualizes the betweenness centrality using a force-based approach, as in Figure 4.13. The line-of-sight topology is noticeably less uniform; not only three of the nodes are linked to others via a single connection, but two of the "internal" nodes are significantly more central than others, potentially becoming a traffic bottleneck or a point of failure.

**TRUNK node placement.** Next, we move to the TRUNK node placement mode, which makes the problem even more challenging at this scale because the overall number of possible node positions is significantly reduced w.r.t. the FREE case. Again, the *line-of-sight* approach is unable to find acceptable solutions in 1000 generations (Figure 4.14). In contrast, LaPS manages to find solutions fulfilling all constraints in 876 generations (see Figure 4.14), providing a topology with good quality at the $1000^{th}$ generation, as shown in Table 4.4 and in the network layout of Figure 4.15c. In LaPS, although $\overline{P}_{rx}$ is comparable to line-of-sight, the number of links in the network and per node is significantly higher (i.e., $L_{acc} = 143$ and $|\overline{\mathbb{L}_k}| = 7.94$ vs. $L_{acc} = 68$ and $|\overline{\mathbb{L}_k}| = 3.77$). As a result the LaPS network is connected with $k_v = k_e = 4$, as well as balanced and uniform (Figure 4.16c), while in the line-of-sight case two nodes are isolated and one "internal" node shows a very high betweenness centrality score, thus rendering the network prone to further partitioning (see Figure 4.16d).

(a) $L_{acc}$      (b) $\overline{P}_{rx}$      (c) $F(\mathbb{P})$
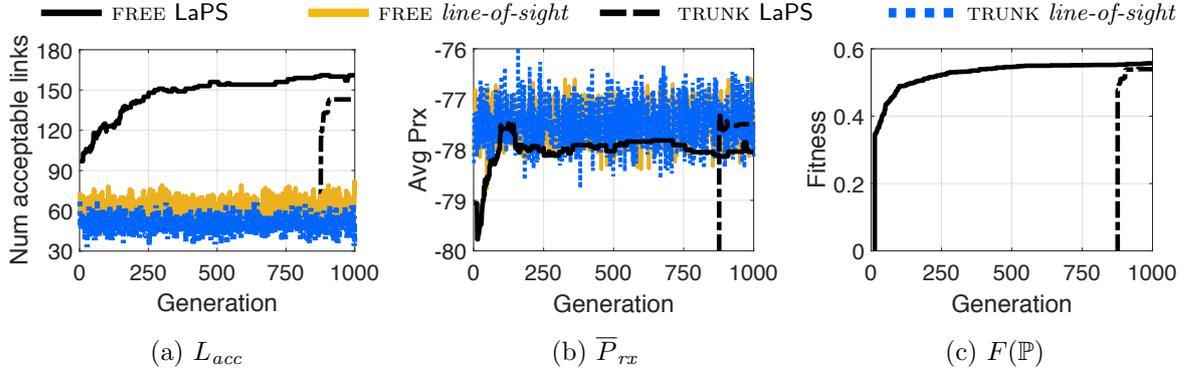
Figure 4.14: Evolution of the fitness function and of its components for 36 nodes in a $300{\times}300$ m$^2$ forest area, scenario REGULAR.



(a) LaPS FREE      (b) line-of-sight FREE      (c) LaPS TRUNK      (d) line-of-sight TRUNK
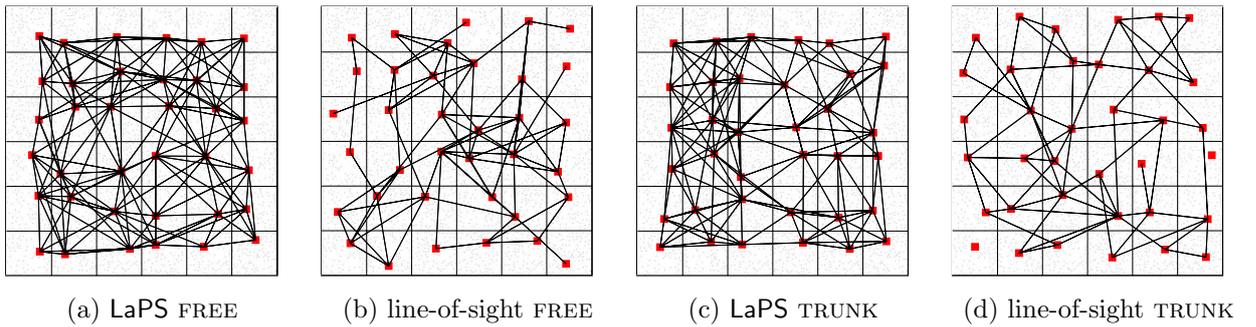
Figure 4.15: Network layouts for 36 nodes in a $300{\times}300$ m$^2$ forest area, scenario REGULAR.



(a) LaPS FREE      (b) line-of-sight FREE      (c) LaPS TRUNK      (d) line-of-sight TRUNK
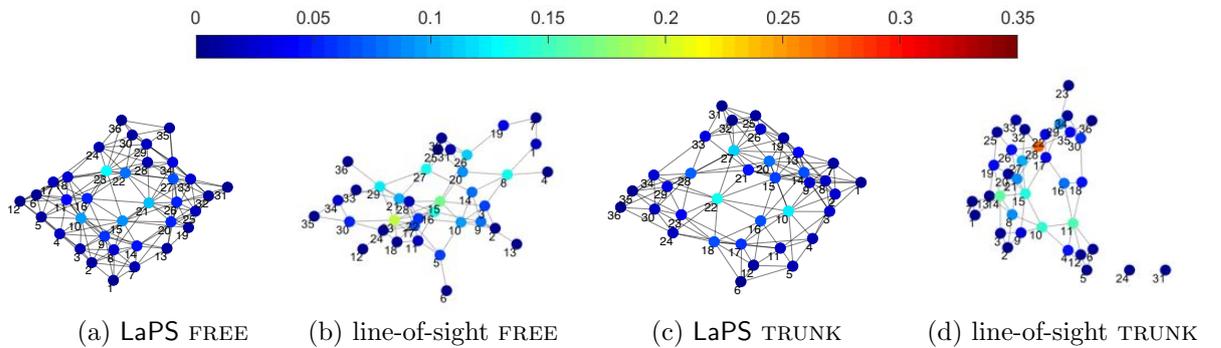
Figure 4.16: Betweenness centrality for 36 nodes in a $300{\times}300$ m$^2$ forest area, scenario REGULAR.

63

**Weakening the network constraints.** These results suggest that, to obtain a good line-of-sight placement in the 36-node scenario, we have to accept some compromise on the constraints. In principle, any of the spatial and network constraints we defined in Section 4.5 could be relaxed: the ability to quickly explore different tradeoffs is precisely one of the strengths of our approach.

For instance, we show here the effect of weakening constraint $C_2$ by setting the minimum number of acceptable neighbors to $M = 2$. We consider the more restrictive TRUNK node placement and investigate the effect of weaking $C_2$ on both line-of-sight and LaPS. The line-of-sight approach converges to a placement solution that fulfills this relaxed constraint and that provides $L_{acc} = 104$ links overall. However, the corresponding network, although connected, can be visually ascertained (Figure 4.18b) as more sparse and fragile than the LaPS solutions, both with $M = 3$ (Figure 4.16c) and $M = 2$ (Figure 4.18a). Indeed, the connectivity metrics show that it is sufficient to remove two nodes or links ($k_v = k_e = 2$) to partition the network; further, some nodes appear to be critical to performance and reliability (Figure 4.18b). Nevertheless, it is interesting to note that, despite the relaxed $C_2$ allows nodes with only $M = 2$ acceptable neighbors, in practice this situation occurs only for 2 nodes (i.e., node 24 and 30). All of the other nodes have at least 3 neighbors and the average number of neighbors $|\overline{\mathbb{L}_k}| = 5.77$ is relatively high.

With the same relaxed constraint, LaPS yields a balanced network with a rather uniform betweenness centrality (Figure 4.18a). The network layout is shown in Figure 4.17a and provides $L_{acc} = 141$ acceptable links (Table 4.4), only slightly less than those obtained in both FREE and TRUNK settings with $M = 3$, as well as comparable average received power and, most important, connectivity. Interestingly, in this case all of the nodes have at least 3 neighbors. This suggests that: *i)* the constraints actually specify a *lower bound* on the desired solutions that can be overreached along the evolutionary search (as in this



| (a) LaPS | (b) line-of-sight |
|:---:|:---:|
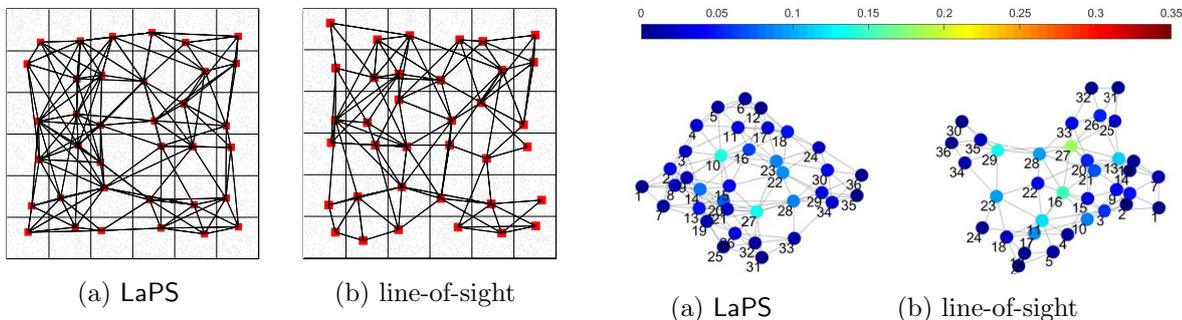
Figure 4.17: Network layouts for 36 nodes in a $300 \times 300$ m$^2$ forest area, scenario REGULAR TRUNK, $M = 2$.



| (a) LaPS | (b) line-of-sight |
|:---:|:---:|

Figure 4.18: Betweenness centrality for 36 nodes in a $300 \times 300$ m$^2$ forest area, scenario REGULAR TRUNK, $M = 2$.

case); and *ii)* the lower is such bound the higher is the number of acceptable topologies that is explored, therefore relaxing a constraint may also lead to the identification of a (near-)optimal solution, depending on the constraint and on the specific forest at hand (e.g., in this case LaPS with $M = 3$ and LaPS with $M = 2$ converge to comparable solutions in the same number of generations).

### 4.6.5 Summary and Discussion

The analysis we presented in this section confirms that LaPS is a powerful tool to understand and examine, automatically and prior to the in-field deployment, what are the attainable tradeoffs between the application and network requirements and the effectiveness of the placement that can be actually achieved in the target forest. Further, it also shows that LaPS is able to find very good placements that would be impractical, if not impossible, to find with the trial-and-error approach typically applied in these cases, due to the necessarily limited exploration; the placements yielded by the latter in-field campaigns are therefore more prone to inefficiency and over-provisioning.

However, are these in-field topologies *significantly* worse than the best one output by LaPS? This aspect can be quantified by analyzing the value of the fitness function $F(\mathbb{P})$. Indeed, if we assume that the in-field network topologies are contained in the search space explored by our evolutionary optimization, we can characterize them with a value of $F(\mathbb{P})$. By looking at the actual range of values for $F(\mathbb{P})$ generated during the evolutionary search, we can have a measure of the "distance" from the optimal case.

To this end, we focus on the 9-node network we analyzed in Section 4.6.3, as this represents a rather simple scenario where one would expect that the benefits provided by LaPS are somehow reduced—although our comparison against a common blind regular placement already showed this is not the case. Table 4.5 reports, for all acceptable solutions (i.e., fulfilling all constraints) and across all combinations we considered, the maximum (best) value of $F(\mathbb{P})$ corresponding to the final output placement, along with the average and standard deviation; Figure 4.19 also shows the empirical cumulative distribution function (CDF) of $F(\mathbb{P})$ values.

These statistics show that, for all considered scenarios, the final solutions to which LaPS converges are significantly better than the average of all those explored. In other words, the distribution of $F(\mathbb{P})$ values is rather broad, and therefore the probability to

| | REGULAR | | IRREGULAR | |
|---|---|---|---|---|
| | FREE | TRUNK | FREE | TRUNK |
| average | 0.94 | 0.94 | 0.90 | 0.90 |
| stddev | 0.26 | 0.23 | 0.22 | 0.26 |
| best | 1.39 | 1.34 | 1.36 | 1.31 |
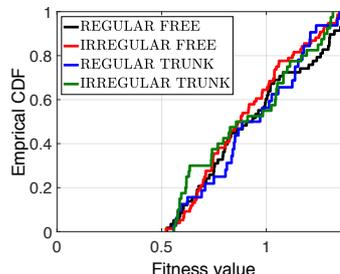
Table 4.5: Statistics and best value of the fitness for all the feasible solutions obtained in all scenarios and placement modes



Figure 4.19: Empirical CDF of the fitness values obtained with 450 generations in all scenarios and placement modes

select, with an in-field deployment, a severely under-performing placement is quite high; for instance, Figure 4.19 shows that the probability to obtain $F(\mathbb{P}) < 1$ is higher than 50% for all scenarios. Table 4.6 further analyzes the quality of placements for given values of $F(\mathbb{P})$, using the combination REGULAR FREE as an example; it is easy to see that the difference between the best value and the others is significant. Further, the value of $F(\mathbb{P})$ is computed only for network topologies that fulfill all constraints; ensuring that this is the case for an in-field deployment is a rather laborious task in itself, even for the small-scale 9-node network considered.

Of course, there is a computational overhead associated to LaPS. On a rather low-scale laptop (Apple MacBook 2008, MacOSX 10.7.5) computing a single generation for the 9-node setup we considered takes about 3 minutes, which become 15 minutes for the 36-node one. These figures are the consequence of the many variables at stake, which in turn witness the complexity of the problem. However, we observe that the current implementation can be significantly optimized, as our goal was simply to build a prototype to demonstrate the feasibility and effectiveness of our techniques. In particular, a distributed implementation, amenable for exploiting parallelism on multiple servers in a data center or in the cloud, can be devised by leveraging the vast literature on distributed

Table 4.6: Quality of placement for specific values of $F(\mathbb{P})$ (REGULAR FREE).

| $F(\mathbb{P})$ | $\overline{P}_{rx}$ (dBm) | $L_{acc}$ | $|\overline{\mathbb{L}_k}|$ |
|---|---|---|---|
| 0.7 (min) | $-78.72$ | 21 | 4.2 |
| 0.94 (median) | $-78.63$ | 26 | 5.6 |
| 1.39 (max) | $-77.58$ | 34 | 7.55 |

genetic algorithms [95]. In any case, we observe that the *offline, unsupervised* computing time required by LaPS, even on the scale we reported, is a negligible cost when compared with the human effort currently required by in-field deployments.

Based on these considerations, we argue that the availability of LiDAR data, properly processed in LaPS via the synergistic application of several techniques, constitutes a formidable asset in effectively and efficiently exploring the space of alternative placement *before* tackling an effort-demanding in-field deployment. Interestingly, finding the best placement is not the only possible use of LaPS, and not necessarily the most interesting or useful. Indeed, LaPS can be exploited also to quickly explore the implications of slight changes in the spatial or network constraints, as we exemplified at the end of Section 4.6.4, therefore navigating the space of inherently conflicting tradeoffs. Needless to say, a similar analysis would be prohibitive to perform in-field, and impossible to carry out offline in the absence of precise information about the tree positions and a specialized radio model exploiting it.

## 4.7 Conclusion

The effective deployment of a WSN in a real forest is known to be difficult and costly to achieve. Methods of practical applicability, capable of providing realistic placement guidelines while ensuring a connected and efficient network, are essentially lacking. Deployments are often performed in-field via trial and error, a process likely to yield networks with a quality inferior to what potentially achievable, and involving a very high effort.

In contrast, in this chapter we presented LaPS, a node placement approach that exploits the increasing availability of LiDAR data in the context of forestry applications to provide an informed node placement layout *automatically* and *prior to the deployment*, accounting for both the actual forest structure and its effect on the network.

LaPS enables off-line automatic exploration and evaluation of placement options in a target forest via three main components: *i)* a representation of the forest derived from LiDAR data, yielding the position and diameter of each tree *ii)* a specialized radio attenuation model exploiting this knowledge to predict the power received on each link *iii)* an optimization strategy based on a genetic evolution process that, along with the other two components, drives the exploration of candidate node placements towards an efficient (sub-)optimal solution. We have shown how, in a real forest scenario, LaPS outperforms alternative placement strategies based on a regular placement or a purely topological free line-of-sight by yielding networks that are significantly more connected and robust. Our results also show that the LiDAR-based forest representation and the specialized radio model are key in achieving this superior performance, especially as the problem size scales, as they drive the search for the best placement with fundamental variables of the target

environment. Further, our tool is flexible and open, as it allows great expressiveness in specifying application and network constraints, as well as the desired quality of the solution, encoded in the fitness function. More generally, the ability to quickly explore the impact of changes in the constraints provided by LaPS is an invaluable asset in improving the quality of WSN operation while minimizing the effort of its in-field deployment.

The contribution presented here also offers opportunities for future work along at least two dimensions. On one hand, although LaPS takes into account the main disruption to communication in a forest (i.e., the position and dimension of its trees), other factors affecting connectivity are currently not considered (e.g., temperature and humidity, or terrain morphology) that can be derived by LiDAR as well. One possible avenue of future research is the integration into LaPS of these variables, some of which are already partially addressed in the literature [99, 28, 27, 154]. This (rather complex) endeavour could likely exploit also the aforementioned flexibility in the (re)definition of constraints and fitness function, whose concrete use in other types of applications is actually an interesting research topic per se.

The other dimension concerns the conceptual contribution put forth in this dissertation, namely, the observation that the vagaries of low-power wireless communication can be tamed, or at least mitigated, automatically and prior to deployment with the help of remote sensing technology. We showed concretely that this can be achieved for the short-range low-power radios that have been commonplace in the last decade. Nevertheless, new radios are appearing that offer different tradeoffs between range, bandwidth, and energy consumption. An example is LoRa [93], whose sensitivity to the environment in general, and vegetation in particular, has already been observed [73, 30, 120], as we discuss in Chapter 5. The surge of 5G [4] may further exacerbate this problem. These contexts are clearly different from the specific ones we tackled in this chapter. However, we argue that the contributions we put forth can be adapted and in general serve as inspiration for techniques that similarly exploit remote sensing to reduce the human effort involved in deploying in-field several low-power wireless devices and, at the same time, improve the expected quality of the resulting placement.

# Chapter 5

# Analyzing the Impact of Land Cover Type on LoRa

LoRa, which stands for <u>Lo</u>ng <u>Ra</u>nge Wireless, is a trademark of Semtech communication technology based on chirp spread spectrum modulation (CSS) [109]. CSS spreads the signal on a wide frequency band by varying, in time, the frequency of the *chirp* carrier signal. This modulation technique improves the robustness against interference and enables the reception of very low power signals (i.e., up to 20 dB lower than the noise floor), thus rendering the communication more resilient to the power loss induced by the signal propagation. As a consequence, it provides comparable low power characteristics as Frequency Shifting Keying (FSK), but significantly increases the communication range (i.e., in the order of kilometers).

LoRa long range and low power characteristics enable to build Low-Power Wide-Area Networks (LPWANs) based on simple single-hop topologies. It is exploited to achieve single-hop end-device to gateway communication in LoRaWAN, an open software protocol standardized by the LoRa Alliance [7, 8, 93], which covers the Data Link and Network layer on top of LoRa. LoRaWAN typically follows a star-of-stars topology, i.e., end-devices communicate by single hop with gateways, which in turn are connected to network servers via IP connection. It is widely adopted in the context of IoT networks, e.g., The Things Network (TTN) [1] is an open free to use IoT network based on LoRaWAN.

**LoRa communication range.** LoRa communication range is typically in the kilometers, with LoRA chirp spread spectrum enabling the selection of several settings that define different tradeoffs between communication range, data rate and power consumption. The LoRa configurable parameters [109] are:

- *Spreading Factor* (SF), representing the number of chips used to encode each symbol. SF may vary between 6 and 12, the larger the more robust is the communication w.r.t. noise. By increasing the SF the receiver sensitivity threshold improves, provid-

---

The contents of this chapter are planned to be extended and submitted to a major conference.

ing longer communication range but also higher energy consumption, in that longer packets are sent (i.e., each symbol is encoded into more chips).

- *Bandwidth* (BW) can be 125, 250 or 500 kHz. The larger the bandwidth the higher the data rate. In turn, increasing BW reduces sensitivity and range, as well as increases the time in the air (i.e., energy consumption).

- *Coding Rate* (CR) for error detection and correction. CR can take the values 4/5, 4/6, 4/7 or 4/8. The larger is CR the more robust is the communication, but increasing CR also requires to send longer packets, thus consuming more energy.

- *Transmission Power* (TP) ranges from -4 to 20 dBm. The higher is TP the higher is the signal strength at the cost of increasing the power consumption.

The capability of configuring several parameters differentiates LoRa transceivers from other popular mote radios, like those considered in other chapters, which enable one to vary the communication range only by tuning the transmission power TP. The impact of LoRa settings on the achievable communication range has been empirically observed in several studies (e.g., [11, 73, 38, 29]). Moreover, it has been observed that the communication range is strongly dependent on the specific environment. In particular, very different, and sometimes contradictory, communication ranges are reported in outdoor scenarios [30, 156, 120, 73, 81].

This variability in the communication range affects the coverage that a gateway is able to provide, depending on its placement and on the characteristics of the surrounding environment. In the context of an IoT scenario, e.g., as implemented in TTN, it is important to deploy each gateway so as to ensure the coverage of a target area of interest (e.g., to provide connectivity to a university campus, or to a specific part of it). Therefore, a better understanding and modeling of the effect of the environment characteristics on LoRa communication can support the definition of environment-representative LoRa connectivity models, as well as gateway coverage models. In this respect, a *quantitative representation of the static characteristics of the environment and of their impact on communication* can support the development of effective network planning strategies. Information about such characteristics can be retrieved by processing Remote Sensing data, thus aiding the analysis and the modeling of their effect on communication. This is the focus of this chapter.

We *exploit multispectral satellite images to automatically derive information about the static characteristics of the communication environment, thus enabling the joint analysis of such characteristics and LoRa communication performance.*

More precisely, we automatically classify Sentinel-2 multispectral images, as described in Section 5.3, to derive *land cover* maps with spatial resolution of 10 meters. The land cover

classes (e.g., Building, Trees or Fields) describe the type of landscape which is present on the ground and that, in this context, represent the communication environment. In other words, the land cover classes describe, with granularity $10 \times 10$ m$^2$, the landscape which is traversed by the LoRa signal along its propagation.

We analyze two LoRa connectivity datasets collected through the TTN gateways in The Netherlands. The first dataset is representative of a *free space like communication environment*, with a LoRa end-device flying on top of an high altitude weather balloon that communicates with the TTN gateways on the ground[1]. In this scenario, which is analyzed in Section 5.2, the communication path is mostly traversing the free space, thus enabling to achieve a communication range in the order of hundreds kilometers (i.e., a maximum range of 245 km is reached). The second dataset is collected with the end-device transmitting at few meters above the ground and across both urban areas and farming fields[2], as described in Section 5.4. It represents a *diversified ground communication environment*, which shows very different communication ranges compared to the previous free space like case (i.e., a maximum range of ∼10 km is achieved). In addition, the communication performance appear very susceptible to the specific (i.e., per-link) characteristics of the surrounding environment. The land cover information retrieved from the multispectral satellite images is analyzed together with the connectivity traces. The analysis is performed in terms of both Packet Reception Rate (PRR) and Received Signal Strength Indicator (RSSI), and provides several insights about the effect of the land cover on LoRa communication. In particular, the predominant land cover class that characterizes the communication path results effective in relating the signal power decay rate with the distance. Moreover, the high resolution land cover maps provide detailed information about the relative position of the land cover classes w.r.t. both end-device and gateway. This information, combined with the respective height of end-device and gateway, can be related with significant differences in the communication performance, e.g., we experimentally observe that a group of trees in the vicinity of the end-device completely obstruct the line-of-sight with the gateway, causing no packet reception, while just 300 meters apart, the communication path traverses only fields and we measure a PRR of 80%. This preliminary analysis provides promising results that encourage to further investigate (e.g., on bigger connectivity datasets) the usefulness of land cover maps to define LoRa connectivity models that account for the land cover type characteristics.

---

[1]Mr. Thomas Telkamp performed the experiments with the balloon and gathered the data.

[2]Members of the Embedded Software Group at TU Delft performed the experiments and gathered these data, with special thanks to Lichen Yao, Minfeng Li, Lu Liu and Xin Liu.

## 5.1 Related Work

The availability of off-the-shelf LoRa long-range and low-power transceivers fostered the adoption of the technology in a number of commercial and trial deployments. The performance analysis of LoRa-based networks, both in terms of coverage and scalability, is therefore gaining interest in the context of IoT LPWAN [39, 120, 11, 149, 59]. The coverage that LoRa gateways can provide affect the performance that protocols can achieve in real deployments and greatly impact the design of IoT systems and applications. Therefore, realistic coverage estimation is crucial to provide coverage guarantees, satisfactory service and resource optimization (e.g., number of deployed gateways) [39], as well as to enable reliable scalability analysis, which are inherently based on assumptions on the coverage and on the communication performance (e.g., range and reliability) [149, 59]. To provide a coverage planning for the city of Padova (Italy), Centenaro et al. [39] estimate the number of gateways required to enable city-wide LoRaWAN coverage. They experimentally obtain a coverage estimate of 2 km in a high buildings area and therefore conservatively assume a nominal range of 1.2 km, thus planning to cover an area of about 1000 km$^2$ with 30 gateways.

However, the actual LoRa communication range is under debate [107, 30, 156, 120, 73, 81], in that real-world observations show both a significant gap w.r.t. theoretical expectations [75, 76] and significant variability depending on the specific environment at hand. Bor et al. [30] observed a range of 2.6 km in rural areas and of 100 m in a built-up environment, while in the central business district of Glasgow (Scotland) the communication range varies from 1 to 20 km [156]. In Hyde Park (London), Kartakis et al. [81] achieve 2.4 km with semi-line of sight conditions, while reaching 450 m in a built-up area. Petajajarvi et al. [120] report a range of 15 to 30 km in a urban/maritime environment and a maximum range of 90 m is measured in a mountain forest with dense vegetation [73].

In addition, LoRa configurable parameters (i.e., PHY settings) determine different trade-offs between range, consumption and data rate. The impact of LoRa settings is indeed explicitly considered in several studies [11, 73, 38, 29]. The authors of [29] evaluate the effectiveness of channel separation using SF and test the reliability of Carrier Activity Detection (CAD) with different SF/BW combinations, which may enable power-efficient duty-cycling exploiting non-destructive concurrent transmission. In [38], instead, a quantitative assessment of the impact of PHY settings on Packet Reception Rate (PRR) is presented. The authors perform extensive experimental campaigns indoor, outdoor, and underground observing: *i)* a drastic decrease of LoRa links reliability at high temperatures; and *ii)* the convenience of using energy hungry PHY setting to increase the link quality instead of relying on retransmission schemes. In particular, they conclude that at

the edge of the communication range it is more effective to use fastest PHY settings with high transmission power. However, the edge of the communication range is not quantified and its value still remains an environment-dependent variable. In this respect, the selection of the environments for the experimental campaigns is usually narrow and/or generic (i.e., indoor, outdoor or urban, sub-urban) and does not reflect the actual location-specific limitations posed by the real-world deployments targeted by IoT applications. Indeed, similar type of environments (e.g., urban) often give rise to different and/or contradictory experimental results in terms of communication range (e.g., [30, 156, 120, 81]) and the need of deployment guidelines steadily fosters further observations.

There is a lack of systematic understanding about how the technology behaves under real conditions and about the impact of real conditions on the technology. Guidelines in this respect could significantly support IoT developers to face the challenges posed by placing LoRa technologies in real spaces, including devices and technologies coexistence and scalability issues.

In [59], the authors analyze the scalability limits of LoRa due to co-spreading factor interference (i.e., interference due to concurrent transmission of devices using the same SF), concluding that it is worse than what was originally promised. They evaluate the coverage probability by exploiting stochastic geometry and observe that when collision occur between packets with the same SF, the stronger signal can be successfully received if it is at least 6 dB stronger than any other. Therefore, the capability to accurately (i.e., with 6 dBm accuracy) predict the expected received power is fundamental to enable reliable collision modeling and realistic scalability analysis. The expected received signal power is estimated by considering the path loss attenuation, which in the paper is computed assuming the path loss exponent equal to 2.7 in sub-urban environments and to 4 in urban environments. However, these approximations may not be representative enough for real scenarios, in that the attenuation as a function of distance usually varies also within environments belonging to the same category (e.g., urban and sub-urban) and it is not isotropic in practice, due to the intrinsic non-homogeneity of the propagation environment.

A complementary perspective is presented in [149], which focuses on analyzing the impact of inter-network interference due to independent LoRa networks operating over the same deployment area. The authors investigate the effectiveness of solutions aimed at reducing such impact, namely exploiting directional antennae and/or increasing the number of gateways. They conclude that the use of multiple gateways improves LoRa performance in the presence of inter-network interference, besides being also an efficient way to scale LoRa networks [30]. In the simulations, the best improvement is obtained by placing the additional gateways so to ensure that all devices are in reach of at least

one of them, under the assumption of circular (i.e., isotropic) coverage. However, in real scenarios, the desired interference mitigation can be guaranteed only by properly accounting for the non-uniform spatial coverage that gateways can provide.

The development of models and tools capable to provide realistic coverage estimates (i.e., a realistic representation of the problem) is of paramount importance to enable performance analysis and planning methodologies that, taking into account the actual deployment environment, represent more closely the behavior of real networks, thus leading to conclusions and insights applicable in practice.

## 5.2 Free-space Like Communication

In this section, LoRa end-device to gateway communication is analyzed considering a free-space like communication environment. The end-device is placed on a flying balloon and communicates with the TTN gateways on the ground. More precisely, a high altitude weather balloon is launched on $15^{th}$ March 2017 at around 1 PM in The Netherlands. The balloon follows the wind direction until it lands 3 hours later, when the helium is consumed, covering a land distance (i.e., the distance between the launch position and the landing position) of 164.4 km. The balloon carries a GPS receiver and a LoRa device, which periodically broadcasts one packet. The packets are received by the TTN gateways on the ground. The 3D trajectory of the balloon is shown in Figure 5.1a, together with the gateways (red dots) that received the packets sent from the balloon. Figure 5.1b shows the same trajectory in 2D reporting also the balloon altitude in color scale. The altitude of the balloon, which reaches a maximum of 22.61 km, is also reported as a function of time in Figure 5.3a.

The communication dataset is composed of all of the packets that are received by TTN gateways. TTN servers provide, for each received packet, RSSI and SNR measured by the gateway, reception timestamp, spreading factor, bandwidth, coding rate, frequency channel and the gateway coordinates (when available). In addition, the payload of the balloon packet is provided, which includes the balloon GPS coordinates, battery voltage, as well as temperature, pressure and altitude measured by a BMP280 barometer. Overall the dataset is composed of 8578 packet receptions. All of the received packets have Spreading Factor (SF) 7, Bandwidth (BW) 125 KHz and Coding Rate (CR) 4/5. The frequency channel, instead, varies. For every transmission the end-device selects the channel in a pseudo-random fashion respecting local regulation (i.e., the frequency plan EU 863-870) and the maximum transmit duty cycle imposed by LoRaWAN on each sub-band (i.e., 1%). Eight channels with central frequencies 867.1, 867.3, 867.5, 867.7, 867.9, 868.1, 868.3, 868.5 MHz are used quite uniformly. The occurrence of channel usage is depicted in Figure 5.2.

(a) 3D balloon trajectory

(b) 2D balloon trajectory
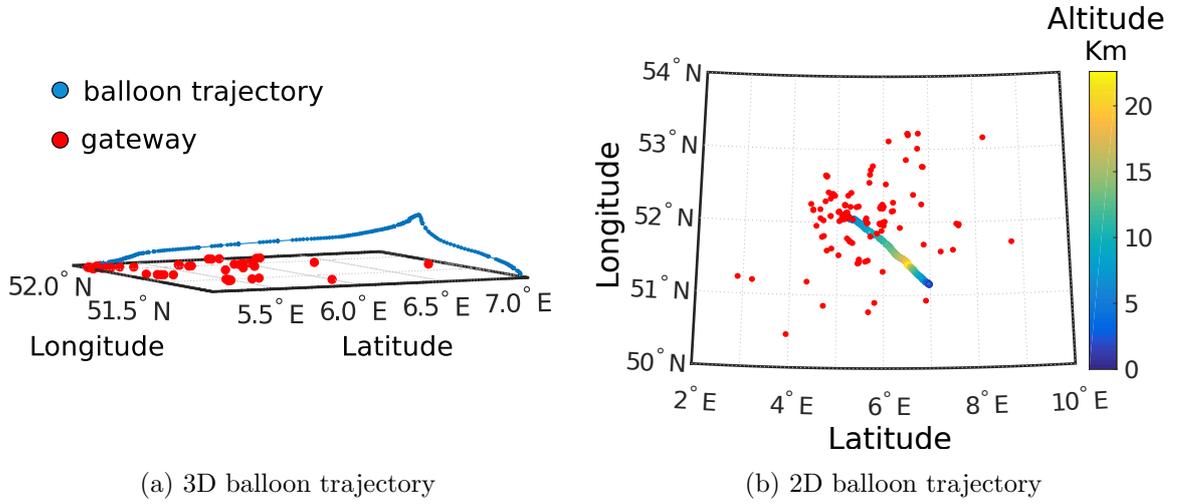
Figure 5.1: Balloon trajectory and receiving TTN gateways (red dots) in Latitude/Longitude geographic coordinate system a) 3D representation b) 2D representation with balloon altitude color label.
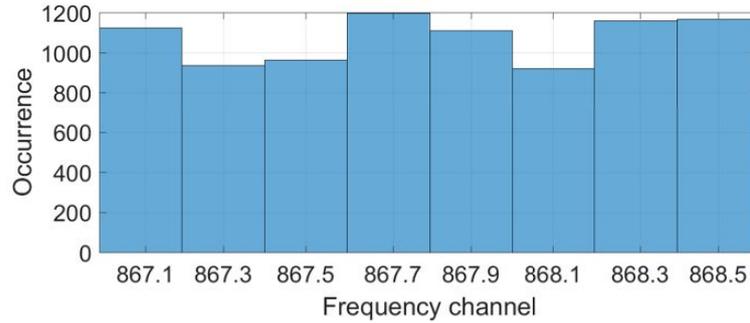


Figure 5.2: Occurrence of frequency channel.

### 5.2.1 Analysis of LoRa Communication Performance

The communication between the LoRa transmitter on the balloon and the TTN gateways on the ground is analyzed along the balloon flight. The balloon connected to 141 gateways overall. The number of visible gateways $N_{vg}$, i.e., the number of gateways receiving the balloon packets, is reported as a function of time in Figure 5.3b, which shows that $N_{vg}$ varies along the flight. The balloon altitude is reported in Figure 5.3a as a function of time. By comparing Figure 5.3a and Figure 5.3b, we can observe that in the first part of the flight, $N_{vg}$ tends to increase as the balloon altitude increases. In particular $N_{vg}$ rapidly increases from 2 to 30, between 13:00 and 13:04, i.e., when the balloon altitude rises from 0.157 km to 0.359 km. Then, the increasing trend of $N_{vg}$ continues between 13:05 and 13:55, despite being less pronounced, with the balloon altitude varying from

75

(a) Balloon altitude



(b) Number of receiving gateways $N_{vg}$ in time
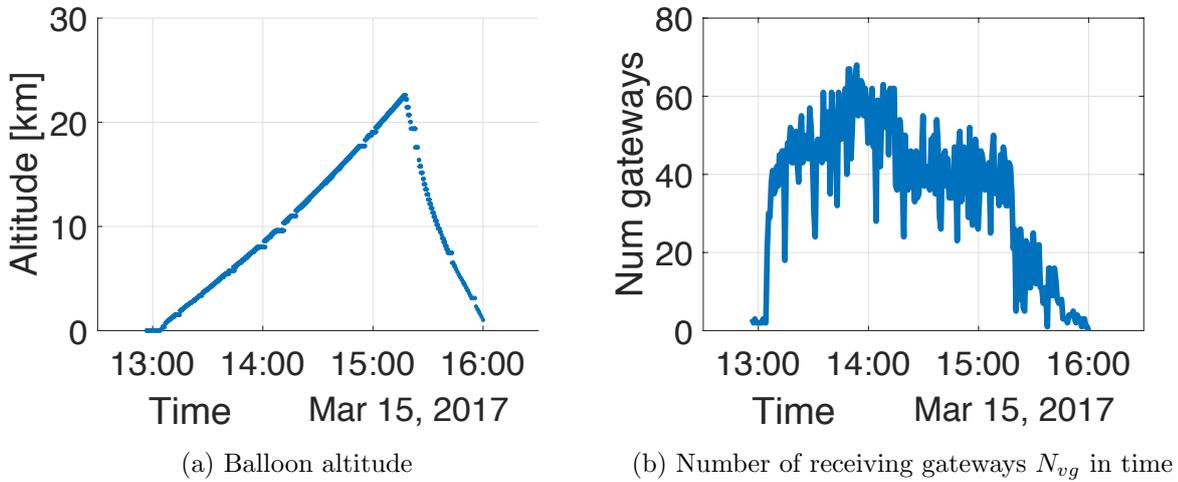
Figure 5.3: Balloon altitude in time and number of gateways receiving the balloon data in time.



(a) Distance from each gateway



(b) Maximum and average distance $\pm$ standard deviation from visible gateways in time
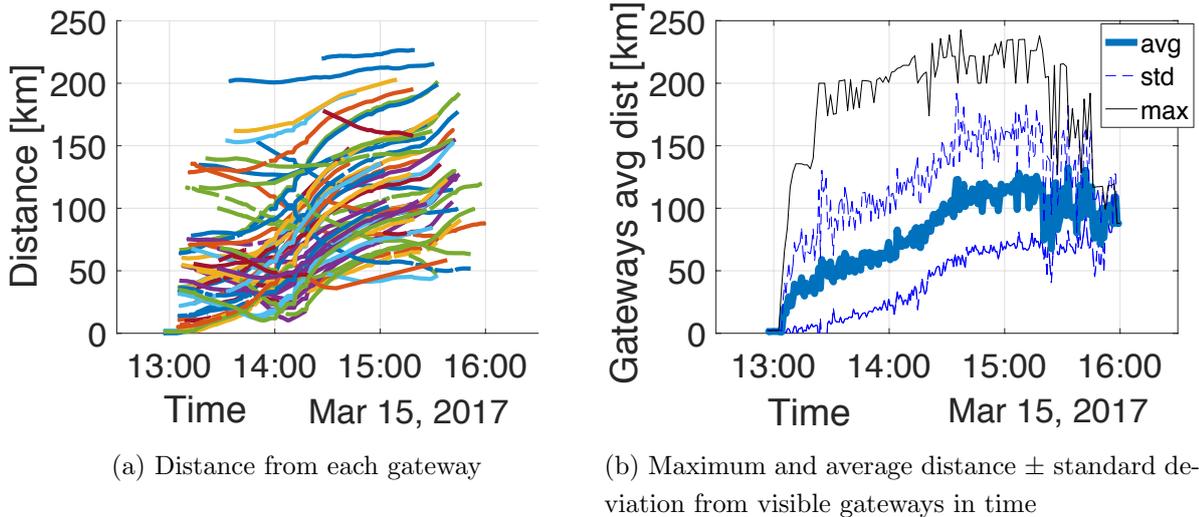
Figure 5.4: Distance of the balloon from the visible gateways in time

0.36 km up to 7.42 km and $N_{vg}$ achieving value 68. After, from $\sim$14:00 to 15:30, $N_{vg}$ tends to decrease, while the balloon altitude is continuously increasing up to $\sim$20 km. Finally, in the last part of the trajectory, during the landing phase, $N_{vg}$ shows a quick drop when the balloon rapidly moves from $\sim$20 km altitude to the ground.

A first observation is that when the balloon rises from the ground up to $\sim$7.5 km, as well as when it lands on the ground, its *altitude* appears to significantly *affect the capability of the device to establish connections with the gateways*. As the balloon rises, the transmitting LoRa device gets line-of-sight communication with more and farther gateways, getting rid of the obstructions that typically characterize the communication

at the ground level (e.g., buildings).

Next, we investigate the effect of the balloon-gateway distance on the established connections. The average balloon-gateway distance, computed on the visible gateways, is shown in time in Figure 5.4b. The balloon-gateway distance is computed as a 3D distance between the balloon and the gateway GPS coordinates and altitudes, considering the World Geodetic System 1984 (WGS84) as reference ellipsoid. Note that the distance is computed only for those gateways that report their GPS coordinates (i.e., 126 out of 141).

Interestingly, the average distance between end-device and visible gateways reaches 130 km along the flight, with a *maximum communication range of 245 km* being achieved (see Figure 5.4b). Note that the spreading factor is 7 and the coding rate is 4/5, i.e., the *less* favorable settings for the communication range. This result significantly differs from every other "ground" observation reported in the literature, thus underlining once again, to which extent the environment can affect the communication range. In particular, a LoRa device flying on a balloon represents a close to ideal *scenario in which the communication path is mostly traversing the free space* and such a condition enables the achievement of a communication range in the order of hundreds kilometers.

Figure 5.4b shows that the average distance between the balloon and the visible gateways tends to increase over time, apart from the very last part of the flight (i.e., the balloon landing). Indeed, in Figure 5.1 we can observe that most of the visible gateways are closer to the launch position of the balloon than to the landing position. However, it is worth noting that the standard deviation of the distances between the balloon and the visible gateways, which is also reported in Figure 5.4b, is considerable. This means that gateways at very different distances are receiving the same balloon packets. We then look at the distance from *each* individual gateway as a function of time, which is shown with a different color in Figure 5.4a. We can observe a diversified behavior depending on the relative position and motion of the balloon w.r.t. to each gateway. Mostly the distances increase in time, i.e., the balloon gets farther from the gateway while proceeding along its trajectory. Sometimes the distances first decrease and then increase, i.e., the balloon first gets close to some gateways and then it moves apart. In few cases the distance is consistently decreasing, i.e., the balloon is approaching the gateway. Therefore, the balloon connects with gateways placed at significantly different distances in the same time interval and this holds for almost the whole flight. In addition, the relative motion between the balloon and the gateways is diversified. These factors should be taken into consideration while investigating the communication performance in this scenario in order to better understand the underlying phenomena affecting LoRa connectivity. Indeed, in Section 5.2.4 we analyze more in detail the received signal strength for a subset

of representative gateways, which are selected by explicitly considering (also) different balloon-gateway distances, as well as different balloon-gateway relative motions.

In summary: *i)* the LoRa communication range in the considered free space like scenario exceeds 200 km; *ii)* the number of visible gateways $N_{vg}$ appears to increase with the device altitude; and *iii)* along the balloon flight, very diversified conditions occur in terms of relative position and motion between the transmitting device and the gateways.

### 5.2.2   Packet Reception

The packet reception for each visible gateway is shown as a function of time and as a function of the balloon-gateway distance in Figures 5.5a and 5.5b, respectively. Each gateway is represented by a different color. The 126 considered gateways (i.e., those which reported their location and for which the distance from the balloon can be computed) are represented in the y-axis of the charts with an integer identifier ranging from 1 to 126. We observe a very diversified packet reception behavior for the various gateways. Some remain connected for almost the whole flight (e.g., the gateway identified by number 78), some others show an intermittent connection (e.g., 31) and other gateways are connected for only one or some time intervals (e.g., 7 and 43). Similarly, by comparing the packet reception with the distance, we find gateways very far ($\sim$200 km) which remain connected for almost the whole flight (e.g., gateway 1) and gateways very close ($\sim$20 km) which remain connected for only half an hour, (e.g., gateway 11).



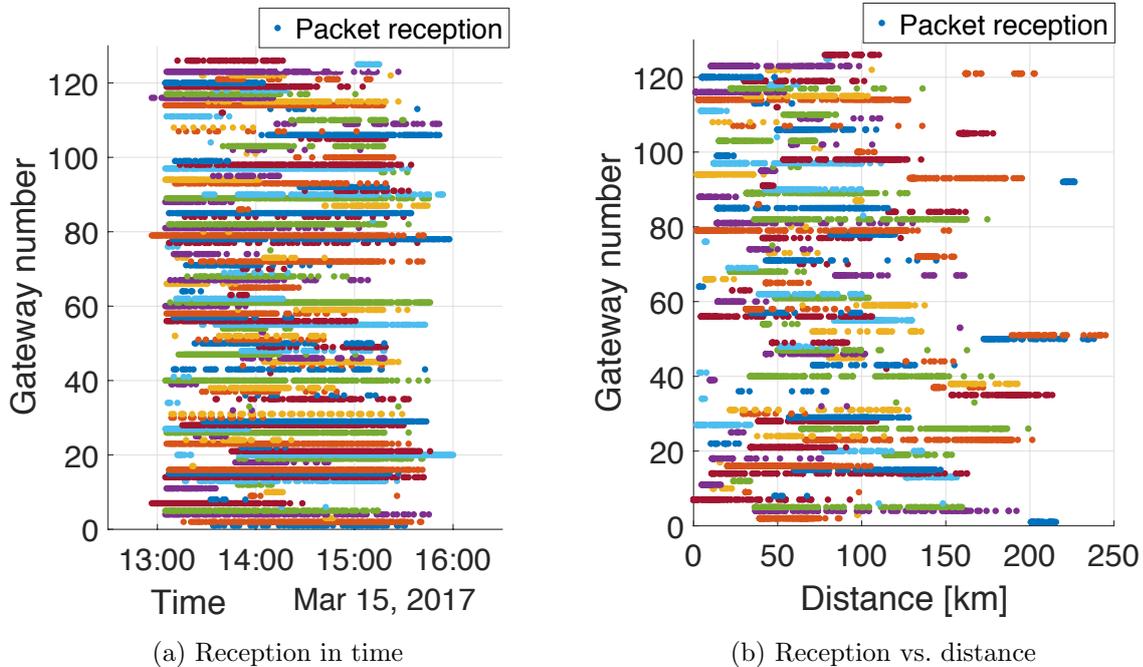(a) Reception in time

(b) Reception vs. distance

Figure 5.5: Reception in time and vs. distance for all the visible gateways.

By analyzing the receptions we can only ascertain that different gateways show different behaviors both in time and w.r.t. to distance. However, by analyzing the reception power we can make some interesting observations and considerations related to the possible factors determining such discrepancies. We therefore, move the investigation of our dataset at the very physical level, namely analyzing the dataset from the point of view of the reception power.

### 5.2.3 Average Received Signal Strength

Figure 5.6a shows the RSSI value as a function of distance for all of the receptions in the dataset, while Figure 5.6b represents the average RSSI and standard deviation vs. the distance. In addition, the expected received power computed according to the Free Space Model [126] is depicted (black dashed line). More precisely, the expected received power $P^{rx}$ is computed by applying the Friis formula of Equation (5.1),

$$P^{rx} = P^{tx} + G^{tx} + G^{rx} - FSPL \qquad (5.1)$$

where FSPL is the free space path loss, $P^{tx}$ is the transmission power, $G^{tx}$ and $G^{rx}$ are the transmitting and receiving antenna gains. The transmission power is not encoded in the packet, since it is not part of the protocol. We therefore assume $P^{tx} =14$ dBm, i.e., the maximum transmission power and default setting for RN2483 LoRa-based wireless module. We also assume no antenna gains $G^{tx} = G^{rx} = 0$, thus reproducing a worst-case scenario. FSPL is given by Equation (5.2), where $d$ is the distance in meters and $f$ is the frequency in MHz, 868 MHz in our case.

$$FSPL = 20log(d) + 20log(f) - 27.55 \qquad (5.2)$$

In addition, Figure 5.7 reports the average absolute error between the Free Space Model and the measurements $\pm$ the error standard deviation. By comparing Figures 5.6a, 5.6b and 5.7 we see that in this free space-like scenario the Free Space curve well captures the trend of the measured RSSI for distances greater than 10 km (i.e., for such distances the error is in average lower than 10 dBm). For distances shorter than 10 km the average error tends to be larger, i.e., the Free Space model does not accurately represent the actual attenuation experienced by the signal. Moreover, for distances shorter than 10 km the standard deviation of the measured RSSI tends to be larger, thus denoting high variability in the RSSI measurements. We further investigate this discrepancy in the next section, by analyzing the behavior of RSSI vs. distance on a per-gateway basis.
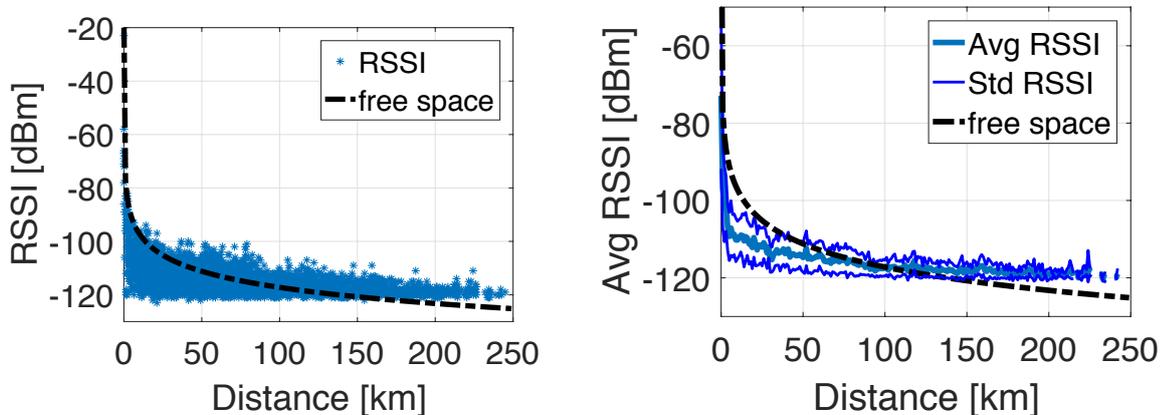
(a) RSSI vs. distance and Free Space Model



(b) Average RSSI ± standard deviation vs. distance and Free Space Model

Figure 5.6: Measured RSSI vs. distance
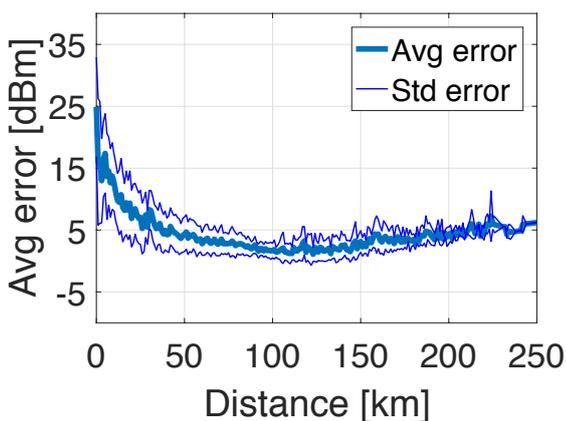


Figure 5.7: Average absolute error between Free Space prediction and measurements ± standard deviation
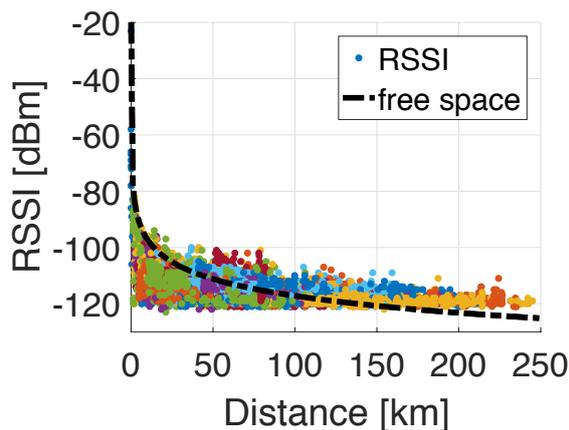


Figure 5.8: RSSI vs. distance for different gateways

### 5.2.4 Per-gateway Received Signal Strength

Observing the RSSI vs. distance for the different gateways, which is depicted with different colors in Figure 5.8 (i.e., each color is associated to one gateway), we can ascertain the diversity characterizing the signal attenuation with the distance for the different gateways. This diversity is reflected also in the different per-gateway behaviors we observed previously analyzing the reception patterns both in time and w.r.t. distance. Moreover, the RSSI variability appears particularly emphasized for the shorter distances. A variety of factors and physical parameters contribute in determining such diversified behavior.

Among them we have differences in the hardware of the gateways (e.g., different antennas with different gains and hardware losses), differences in the relative antenna orientation between the balloon device and the gateway, and therefore between their correspondent non-isotropic radiation patterns, in addition, the relative movement of the balloon w.r.t. the gateways is neither regular nor uniform, the weather (e.g., rain) may be changed in time, and we likely have different scenarios on the ground where the gateways are placed.

Next, we further investigate the aforementioned diversity analyzing more in detail the communication for some representative gateways. We focus on a subset $G$ composed of 7 gateways. $G = \{7, 50, 61, 72, 79, 106, 116\}$, where the numbers refer to the integer gateway identifiers we previously adopted. Table 5.1 summarizes some relevant characteristics related to the gateways in $G$, namely, the gateway identifier we use in this analysis $G_{id}$ and the correspondent TTN unique identifier $TTN_{id}$, the overall number of received packets $N_{rx}$, the duration of the time interval during which the gateway received the balloon packets $T_c$ (i.e., the time elapsed between the first and the last reception) and the range of distances for which the packets are received $D_c$ (i.e., the range between the shorter

| $G_{id}$ | $TTN_{id}$ | $N_{rx}$ | $T_c$ | $D_c$ |
|----------|------------|----------|-------|-------|
| 7 | eui-0000024b080602ed | 98 | 1h 30' | 0.01 - 93 |
| 50 | eui-1dee0855a73c5652 | 75 | 1h 55' | 172 - 238 |
| 61 | eui-1dee15a874d24644 | 179 | 2h 38' | 40 - 104 |
| 72 | eui-84eb18fffee38ec6 | 120 | 2h 20' | 133 - 154 |
| 79 | eui-aa555a00080605b7 | 173 | 2h 45' | 2 - 151 |
| 106 | eui-b827ebfffee7e242 | 123 | 1h 48' | 49 - 108 |
| 116 | eui-fffeb827eb5d8d35 | 79 | 1h 13' | 1 - 53 |

Table 5.1: Selected gateways and correspondent connection characteristics: gateway integer identifier ($G_{id}$), TTN identifier ($TTN_{id}$), number of receptions ($N_{rx}$), connection duration ($T_c$) and connection distances from-to [Km] ($D_c$).
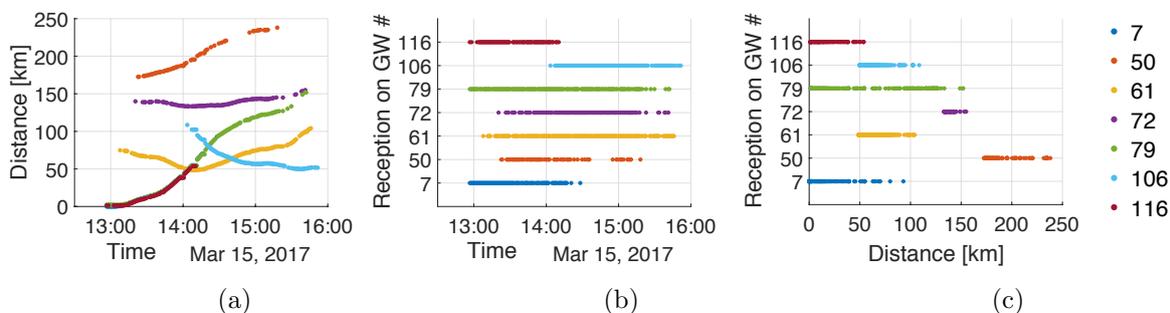


Figure 5.9: Distance in time (a), reception in time (b) and reception vs. distance (c) for the 7 gateways in subset $G$.

distance of reception and the greater distance of reception). Note that the packet reception rate is not necessarily constant in $T_c$ and $D_c$.

These gateways have been selected since they show different and representative characteristics, in terms of time span between the first and the last reception, range of distances and relative movement of the balloon w.r.t. them. The different combinations of these characteristics can be visually ascertained by comparing Figure 5.9a, Figure 5.9b and Figure 5.9c, which show, for each gateway, the distance in time (i.e., the balloon-gateway relative motion), the reception in time and the reception as a function of distance, respectively. Some receive packets at very long distances (e.g., $G_{id} = \{72, 50\}$), others do not reach such communication range (e.g., $G_{id} = \{7, 116\}$). The balloon is moving away from gateways $G_{id} = \{7, 79, 116\}$, which are the closest to the launch location and receive packets from the beginning of the flight. In particular, $G_{id} = 79$ is the gateway that maintained the contact for the longest period of time (i.e., 2h 45'), as well as for the widest range of distances (i.e., from 2 to 151 km). The relative movement of the balloon w.r.t. $G_{id} = 61$ undergoes an inversion along the fligth: the balloon first approaches the gateway and after moves away, with the distance ranging from 50 to 100 km and the contact being maintained for a long time (i.e., 2h 38'). Finally, the ballon is approaching $G_{id} = 106$ (i.e., the distance decreases in time), with the communication starting at ~100 km.

Figure 5.10 shows the RSSI as a function of distance for the various gateways in $G$, which are distinguished by different colors. In addition, Figure 5.11 shows the average error and standard deviation between the RSSI measurements and the Free Space Model for each gateway in $G$, separately. First, we observe that for the gateways with which the communication occur only for distances $\geq 50$ km , i.e., $G_{id} = \{50, 72, 61, 106\}$, RSSI matches well the Free Space Model, with average error consistently lower than 10 dBm. This underlines that in this case the communication scenario is actually "free space like". In the case of $G_{id} = 61$, the balloon first approaches the gateway and then moves apart (as shown in Figure 5.12a). Figure 5.12b, which reports the RSSI in time for $G_{id} = 61$, shows that the relationship between RSSI and distance is reflected in time as an effect of the balloon-gateway relative motion. More precisely, as the balloon approaches the gateway and the distance decreases in time, the RSSI value tends to increase in time. The trend is then inverted when the balloon moves apart from the gateway and RSSI tends to decrease.

Next, we move to the gateways that are close to the launch position of the balloon, namely $G_{id} = \{116, 7, 79\}$, to observe the RSSI vs. distance behavior for distances shorter than 50 km. This is particularly interesting because it is on short distances (i.e., $\leq 10$ km) that we previously observed a significant average error between the measured RSSI and the
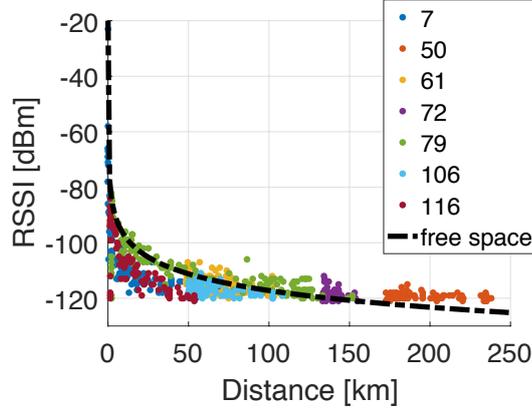
Figure 5.10: RSSI vs. distance for the subset of gateways $G$ and Free Space Model.
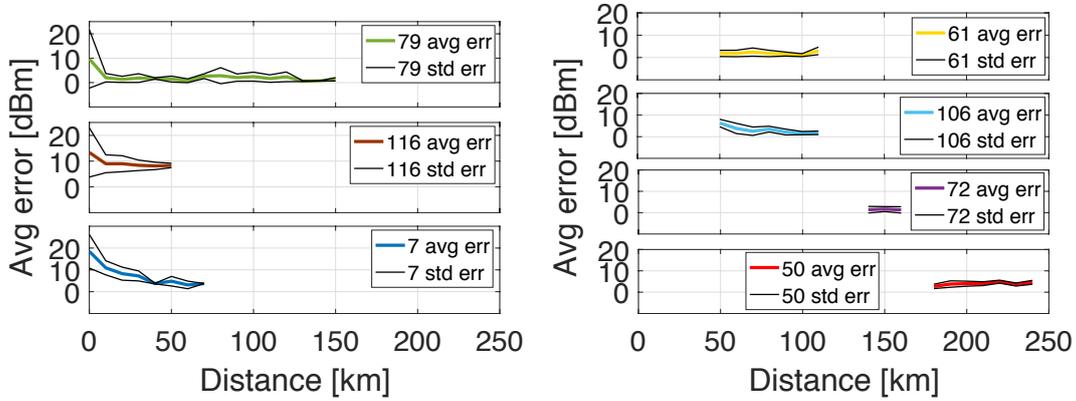


Figure 5.11: Average absolute error $\pm$ standard deviation between RSSI measurements and Free Space prediction for each gateway in $G$.

Free Space Model (Section 5.2.4). The same behavior is encountered also in this subset of data, as shown in Figure 5.11. We separately analyze the communication with these three gateways. The distance in time, the RSSI in time and the RSSI vs. distance are shown for gateways $G_{id} = \{116, 7, 79\}$ in Figures 5.13, 5.14 and 5.15, respectively. The RSSI measured by $G_{id} = 79$ fits very well the Free Space Model (see Figure 5.15c), apart from some samples in the very first kilometers, whose value spans in the range [-120 -80] dBm. In contrast the trend observed for $G_{id} = 7$ and $G_{id} = 116$ present stronger decay rates, with the communication being lost at shorter distances (i.e., 90 and 50 km, respectively) and faster (see Figures 5.14b and 5.13b).

These three different trends in the decay of RSSI with the distance, which can be better compared in Figure 5.16, can be considered as representative of the wide variability observed in the overall RSSI samples for distances $\leq 50$ km. The three gateways have likely different hardware (e.g., antennas), however it is interesting to note that the environment surrounding them, as well as the height at which they are placed, differ (see Figure 5.17).
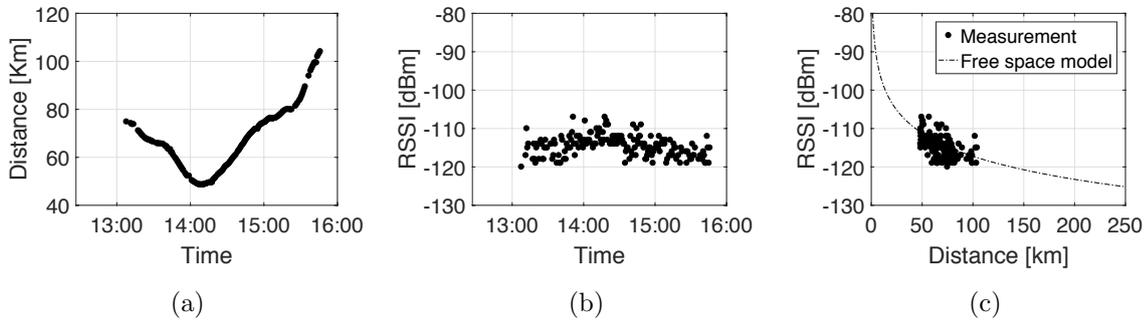
Figure 5.12: $G_{id} = 61$ (a) distance vs. time (b) RSSI vs. time (c) RSSI vs. distance
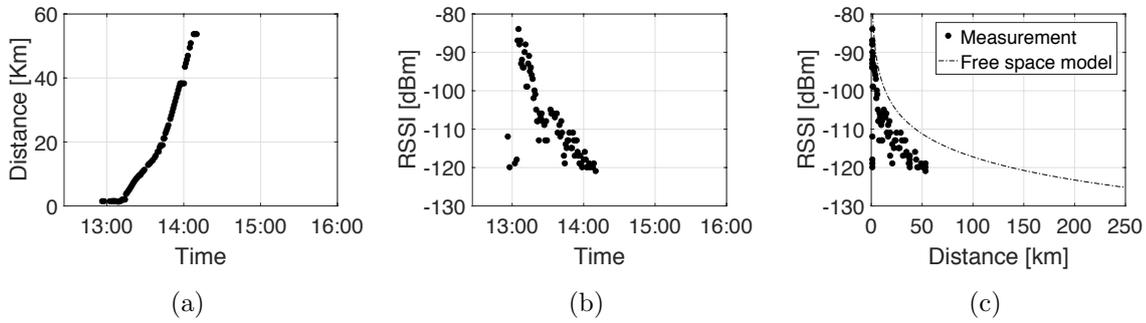


Figure 5.13: $G_{id} = 116$ (a) distance vs. time (b) RSSI vs. time (c) RSSI vs. distance
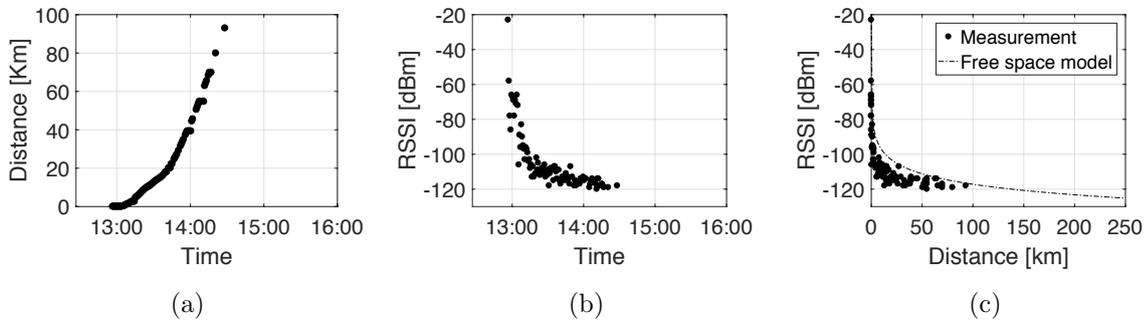


Figure 5.14: $G_{id} = 7$ (a) distance vs. time (b) RSSI vs. time (c) RSSI vs. distance
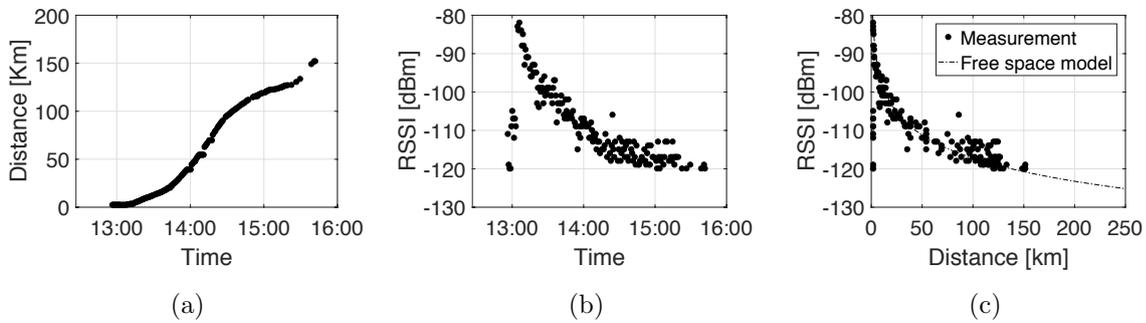


Figure 5.15: $G_{id} = 79$ (a) distance vs. time (b) RSSI vs. time (c) RSSI vs. distance
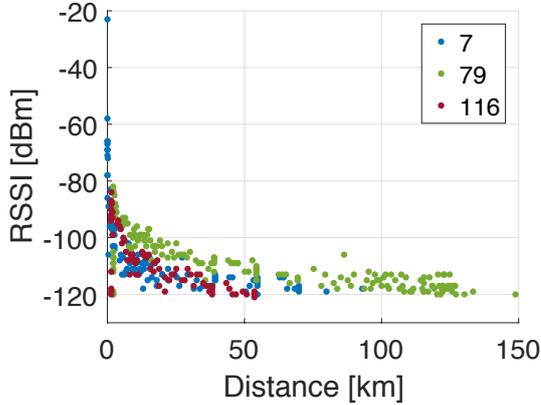
Figure 5.16: RSSI vs. distance for a gateways $G_{id} = \{116, 7, 79\}$
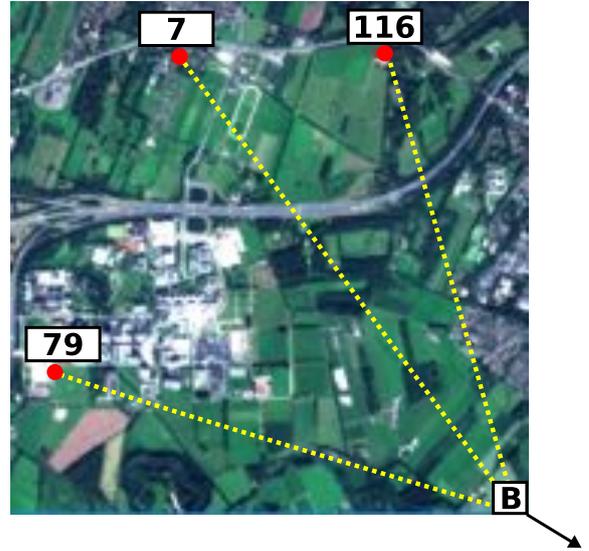
Figure 5.17: Gateways $G_{id} = \{116, 7, 79\}$, direction of the balloon trajectory (B) and communication paths (yellow dashed lines)

More precisely, $G_{id} = 79$ is placed at 38 m height and in its vicinity, considering the direction of the balloon trajectory, there are agricultural fields. In contrast, $G_{id} = 7$ is placed at only 5 m height in a residential area (i.e., it is surrounded by residential buildings). and $G_{id} = 116$ is at 11 m height in front of trees w.r.t. the balloon movement. The combination of the environment *static* characteristics surrounding the gateway w.r.t. the balloon position are likely concurrent factors in determining their different power decay rate with the distance and therefore their different communication range. In Section 5.4 we show how this kind of effect is exacerbated when the transmitting device is placed close to the ground, where much stronger power decay rates and much shorter communication ranges are observed.

### 5.2.5 Summary

From this analysis, we can conclude that: *i)* increasing the altitude of the LoRa transmitter (in the order of kilometers) we obtain a free space like communication environment, thus increasing the capability to establish connections with more gateways; *ii)* in a free space like environment LoRa communication range can achieve the order of hundreds kilometers, i.e., much greater than what commonly reported on the ground; *iii)* the Free Space Model captures reasonably well the RSSI vs. distance behavior for distances $\geq 10$ km (in the considered scenario); *iv)* for distances $\leq 10$ km (in the considered scenario) the combination of gateway altitude and surrounding environment play a role in affecting the attenuation rate of the received power with the distance and therefore the communication range.

## 5.3 Mapping the Land Cover with Satellite Multispectral Images

This section describes how information about the land-cover characteristics is automatically derived by processing Sentinel-2 (S2) multispectral (MS) images. Satellite RS technologies have been widely applied to large-scale monitoring of the Earth surface and latest-generation equipment (e.g., S2 satellite constellation) can acquire and provide optical images at global scale with high spatial and temporal resolutions. These images are exploited for a variety of services as land-cover and land-use mapping, detection of changes, water management, forest monitoring, risk management (e.g., floods, landslides) [71, 94, 19, 91]. In particular S2 MS images, which are available through the Copernicus Open Access Hub [2], recently proved to be effective in the context of greenhouse detection [110], build-up areas identification [119] and water bodies mapping [51].

Our goal is to automatically derive a classification map representing the land-cover types (i.e., the land-cover classes) that characterize the area where the LoRa communication traces we analyze have been collected. Accurate land-cover maps are derived from S2 MS images, exploiting the spectral response of the different land-cover classes and applying supervised classification techniques [134, 71, 103, 131]. We apply a pixel-based classification approach based on kernel methods and in particular Support Vector Machines (SVM) [106, 70, 35, 61, 103, 32], due to their good generalization capabilities, high classification accuracy and relatively simple design through few control parameters.

This section is organized as follows. First, we describe the Sentinels and in particular the Sentinel-2 constellation, which provides the MS images we use (Section 5.3.1). We then describe our MS dataset and define the land-cover classes of interest (Section 5.3.2). Next, we define the features we use for classification, specifying the image pre-processing we perform (Section 5.3.3), and detail the SVM classification approach we apply (Section 5.3.4). Finally, the obtained classification results are reported and discussed in Section 5.3.5.

### 5.3.1 The Sentinel Constellation

The Sentinels are a family of satellites developed and launched by the European Space Agency (ESA) in the context of the Copernicus programme, which is aimed at providing global monitoring information for environment and security applications. More precisely, the Sentinels are a series of dual satellite constellations that supply satellite data at European and international levels, with Copernicus providing an open access platform through which such data are available for services and applications [3, 2].

In particular, the Sentinel-2 constellation [100, 50] is dedicated to the pair of twin polar-orbiting satellites Sentinel-2A and Sentinel-2B, launched in June 2015 and March 2017, respectively. This satellite pair, orbiting sun-synchronously at 786 km altitude,

| S2 Band | CW [nm] | B [nm] | R [m] |
|---|---|---|---|
| B1 - Coastal aerosol | 443 | 20 | 60 |
| B2 - Blue | 490 | 65 | 10 |
| B3 - Green | 560 | 35 | 10 |
| B4 - Red | 665 | 30 | 10 |
| B5 - Vegetation Red Edge | 705 | 15 | 20 |
| B6 - Vegetation Red Edge | 740 | 15 | 20 |
| B7 - Vegetation Red Edge | 783 | 20 | 20 |
| B8 - Near InfraRed (NIR) | 842 | 115 | 10 |
| B8A - Vegetation Red Edge | 865 | 20 | 20 |
| B9 - Water vapour | 945 | 20 | 60 |
| B10 - Short Wave InfraRed (SWIR) - Cirrus | 1375 | 30 | 60 |
| B11 - Short Wave InfraRed (SWIR) | 1610 | 30 | 20 |
| B12 - Short Wave InfraRed (SWIR) | 2190 | 180 | 20 |

Table 5.2: Bands of the Sentinel-2 Mutispectral Imager (MSI) with the correspondent central wavelength CW [nm], bandwidth B [nm] and spatial resolution R [m].

operate simultaneously to realize a global multispectral Earth-observation system with high revisit frequency. They systematically acquire MS images over land and coastal areas (from -56° to 84° latitude), with revisit frequency of 5 days at the equator, which becomes 2-3 days at mid-latitudes. S2 MS images have 13 spectral bands in the visible, near-infrared and short wave infrared range, with spatial resolution ranging between 10 meters and 60 meters, depending on the spectral band. The available bands are reported in Table 5.2 together with the correspondent central wavelength [nm], bandwidth [nm] and spatial resolution [m].

### 5.3.2 Multispectral Dataset and Land-cover Classes

The classification process is performed on three S2 MS images, or tiles, covering the target area of interest in The Netherlands (i.e., where the LoRa communication traces we analyze are collected[3]). Each tile represents a ground area of approximately $100 \times 100$ km$^2$ and the area covered by the three tiles, identified as 31UET, 31UFT and 31UFU, is shown in Figure 5.18. These MS images have been acquired on $26^{th}$ May 2017 with no cloud coverage. Given the exploratory purpose of our analysis, in such a first stage we do not consider multitemporal series of images but single-date images, assuming that the land cover of our target area does not significantly change in the time period we focus on (i.e., July 2017). The true color composition (i.e., Red, Green and Blue bands) of the S2 MS image representing a portion of tile 31UET, which covers a part of the west coast of The Netherlands, is shown in Figure 5.19a.

---

[3]Note that the LoRa connectivity dataset analyzed in this dissertation covers only one portion of one tile (i.e., 31UET). We plan to analyze a much larger LoRa dataset covering all of the three tiles we refer to in this section.

Figure 5.18: Area covered by the considered Sentinel-2 MS images (tiles 31UET, 31UFT and 31UFU).

**Land-cover classes.** Seven land-cover classes of interest are defined. The definition of these classes is based on two main criteria: i) usefulness in characterizing the LoRa communication environment, since they may differently affect the LoRa propagation; and ii) possibility to discriminate them in the spectral domain of MS images, due to their different spectral signatures. The seven classes are: Water, Field (i.e., farming field or grassland), Soil (i.e., bare soil), Building, Greenhouse (i.e., plastic covered greenhouses for indoor farming), Road and Trees. In particular, Building, Greenhouse and Trees are supposed to significantly affect the radio signal attenuation at 868 MHz, when the communication path traverse them relatively close to the ground. In contrast, Field, Soil and Road are expected not to cause additional attenuation other than the loss due to communication distance. We specifically consider the classes Water and Greenhouse, since they are significantly present in the study area. Note that the effectiveness of S2 MS images for detecting some of these classes (i.e., Water, Building and Greenhouse) has been specifically observed in recent studies [51, 119, 110].

### 5.3.3   Features and Pre-processing

The goal of the classification task is to associate each pixel of the MS images to the class that optimizes a predefined classification criterion, to obtain a land-cover map $M$. A MS image is composed of $m \times n$ pixels. Each pixel is represented by F features and the classification task is accomplished in the feature space. In our case, a MS image (i.e., a tile) is composed of $10980 \times 10980$ pixels (i.e., $m = n = 10980$) and we represent each pixel with F = 13 spectral features. Eleven spectral features are defined by the S2 MS

bands (i.e., all apart from Band 10, as it is used to perform cirrus detection and does not contain surface information, and Band 8A, which is mainly used as reference channel to perform atmospheric corrections). Moreover, two indexes are computed and used as additional features: the Normalized Difference Vegetation Index (NDVI) [133], and the Normalized Difference Water Index (NDWI) [102]. These indexes are based on spectral ratios that emphasize the spectral properties of vegetation and water, thus rendering more easy their detection. NDVI is computed as per Equation 5.3, exploiting the Near InfraRed (NIR) and Red bands (i.e., B8 and B4), while Equation 5.4 defines NDWI as a function of the Green and NIR bands (i.e., B3 and B8).

$$NDVI = \frac{NIR - R}{NIR + R} = \frac{B8 - B4}{B8 + B4} \qquad (5.3) \qquad NDWI = \frac{G - NIR}{G + NIR} = \frac{B3 - B8}{B3 + B8} \qquad (5.4)$$

The spectral features are retrieved from Sentinel-2 level 1C products: MS images with radiometric corrected radiances, geometrically corrected (i.e., ortho-rectified) and geo-referenced (i.e., spatially registered on a global reference system). In addition, the following additional pre-processing steps are performed: 1) the bands with 20 and 60 m resolution (see Table 5.2) are resampled to 10 m in the spatial domain to ensure that each band has the same pixel size; and 2) all the features are quantile normalized (98%) in the interval [0 1], thus removing outliers.

### 5.3.4 Classification Approach

The considered classification problem involves multiple non-linearly-separable classes. Therefore, the classification task is accomplished by applying nonlinear kernel methods, in particular SVM [36], and adopting the one-against-all (OAA) strategy, which represents a state-of-the-art multiclass approach used for SVMs [31]. This classification method is supervised, i.e., it is based on a training phase that exploits a set of labeled samples. A set of image pixels is manually associated to the correspondent class label via photo interpretation. In particular, a training set of labeled samples is used for the learning phase of SVM and a test set of labeled samples is used to evaluate the classification accuracy.

The multiclass OAA strategy involves as many parallel SVMs as the number of considered classes, i.e., seven in this case. Each SVM solves a binary classification problem defined by one class against all the others. The final (multiclass) classification map is derived according to a "winner-takes-all" rule on the intermediate results. Each SVM finds the optimal discriminant function $f(\mathbf{x})$ that best separates the considered two classes in the feature space, while properly penalizing the misclassified samples [147, 33]. The vector $\mathbf{x} \in \mathbb{R}^F$ represents a pixel in the F-dimensional feature space. The non-linear SVM approach is based on the mapping of $\mathbf{x}$ (i.e., $\Phi(\mathbf{x})$) into a higher dimensional space where the classes, which are not linearly separable in the original feature space, can be linearly

separated by an hyperplane. The discriminant function $f(\mathbf{x})$ in the original feature space is defined in Equation 5.5 as a function of a weight vector $\mathbf{w}$ and a bias $b \in \mathbb{R}$.

$$f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b \tag{5.5}$$

The optimal hyperplane minimizes the cost function $\Psi(\mathbf{w}, \xi)$ defined in Equation 5.6, subject to the constraints of Equation 5.7, where $\xi_i$ are the slack variables introduced to account for non-separable data, $N$ is the number of training samples and $y_i \in \{\omega_A, \omega_B\}$ are the binary class labels targeted by each binary SVM. In our case the set of seven classes of interest is $\Omega = \{\omega_1, \ldots, \omega_7\} = \{$ Water, Field, Soil, Building, Greenhouse, Road, Trees $\}$. Each of the seven parallel SVMs solves the binary problem between classes $\omega_A = \omega_i$ and $\omega_B = \Omega - \omega_i$ with $\omega_i \in \Omega$. $C$ is a regularization parameter that tunes the shape of the discriminant function determining the amount of penalty to associate to misclassified samples (i.e., increasing the value of C increases the penalty associated to misclassified samples).

$$\Psi(\mathbf{w}, \xi) = \frac{1}{2}\|\mathbf{w}\| + C \sum_{i=1}^{N} \xi_i \tag{5.6}$$

$$\begin{cases} y_i(\mathbf{w} \cdot \Psi(\mathbf{x}) + b) \geq 1 - \xi_i & i = 1, 2, \cdots N \\ \xi_i \geq 0 & i = 1, 2, \cdots N \end{cases} \tag{5.7}$$

The Lagrangian formulation enables the dual representation of the discriminant function as per Equation 5.8 in terms of Lagrange multipliers $\alpha_i$ and kernel function $K(\cdot, \cdot)$, which should satisfy the Mercer's theorem.

$$f(\mathbf{x}) = \sum_{i \in S} \alpha_i\, y_i\, K(\mathbf{x}_i, \mathbf{x}) + b \tag{5.8}$$

The Lagrange multipliers $\alpha_i$ weight each training sample depending on its importance in determining the discriminant function and they can be estimated by quadratic programming methods [147]. In particular, the training samples associated to nonzero Lagrange multipliers (i.e., $\mathbf{x}_i$ with $i \in S$ and $S \subset N$) are the support vectors that identify the most important samples in determining the discriminant function. In this case a radial basis function (RBF) kernel [36] is adopted, which is defined as per Equation 5.9 and tuned by the parameter $\gamma$, which determines its width.

$$K = exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2) \tag{5.9}$$

This classification approach requires the selection of few key parameters, namely the kernel parameters (i.e., $\gamma$ in our case) and the regularization parameter $C$. $\gamma$ determines the smoothing of the discriminant function, while $C$ determines the amount of penalty to

associate to misclassified samples. The control parameters $(C, \gamma)$ are estimated through a model selection process aimed at i) accurately discriminating the classes; and ii) minimizing the expected generalization error. The model selection is accomplished by applying a grid search strategy, in which the grid is defined by a discretized range of $C$ and $\gamma$ values. Each pair of $(C, \gamma)$ values in the grid is used to train an SVM and evaluate its classification performance by $n$-fold cross-validation [20]. Namely, the training samples are randomly divided into $n$ folds. $n-1$ folds are used to train the SVM and the classification accuracy is evaluated on the $n^{th}$ fold, according to the OAA procedure. The process is repeated $n$ times exchanging the training/evaluation folds and the average classification accuracy of the $n$ results is computed. The $(\gamma, C)$ pair providing the best cross-validated estimate of the classification accuracy is selected and used for the final SVM training. Finally, the trained model is applied to predict the class labels of the whole MS image and obtain the land-cover classification map $M$. This result is evaluated on an independent set of test samples in terms of Overall Accuracy (OA), Producer's Accuracy (PA) and User's Accuracy (UA). OA is the percentage of test pixels correctly classified. PA and UA are per-class metrics that relate to the error of omission and to the error of commission, respectively. PA is the percentage of correctly classified pixels for the given class. UA is the percentage of correctly classified pixels computed w.r.t. the overall number of pixels that have been associated to the given class.

### 5.3.5 Classification Results

Each tile is classified independently after collecting tile-specific reference data. In particular, a training set composed of 200 samples for each class and a test set with 100 samples for each class are prepared for each tile. The training sets are used to perform the model selection and the SVM training, while the test sets are exploited to assess the classification accuracy.

A grid-search model selection based on 5-fold cross-validation (i.e., $n = 5$) is performed on each tile, testing $C$ between 100 and 1000 with a step size increment of 20 and $\gamma$ between 0.1 and 2 with a step size increment of 0.1 [46, 150]. Table 5.3 summarizes the best performing tuning parameters $(C, \gamma)$ obtained for each tile, together with the

| Tile | $C$ | $\gamma$ | Accuracy (%) |
|------|-----|----------|--------------|
| 31UET | 620 | 1.2 | 96.21 |
| 1UFT | 980 | 0.6 | 94.68 |
| 31UFU | 660 | 1.4 | 90.78 |

Table 5.3: Selected regularization parameter $C$ and width of the RBF kernel $\gamma$ for the three tiles and average 5-fold cross-validation accuracy.

correspondent average 5-fold cross-validation accuracy. These $(C, \gamma)$ values are used to train the SVM classifier for the correspondent tile. The resulting classification performances computed on the test sets are reported in Table 5.4 in terms of Overall Accuracy (OA), Producer's Accuracy (PA) and User's Accuracy (UA). Table 5.4 shows good overall accuracy for the three tiles (i.e., OA $\geq$ 90% ), despite OA for tile 31UFT (i.e., 92%) is slightly lower than OA for the other two tiles (i.e., 95% and 98.3%). We can observe that for tile 31UFT the most penalized classes are Building and Road, with PA = 70% and PA = 80%, respectively. Comparing these PA values and the corresponding UA values we can deduce that these two classes tend to be confused (i.e., not well discriminated), likely due to relatively similar spectral signatures. A similar behavior, but less pronounced, is found for tile 31UFU. The analysis we want to perform is aimed at relating the land-cover characteristics with LoRa communication performance. In this respect, confusing Building and Road pixels might be critical in principle, in that these two classes are expected to affect differently LoRa communication. However, the percentage error is not major (see Table 5.4) and it tends to occur on sparse pixels (see Figure 5.20b) representing a ground area of 10 $\times$ 10 m$^2$ each, while LoRa communication range is in the order of kilometers. Therefore, such an error is not expected to critically affect the analysis of the land-cover types that characterize LoRa communication paths.

| | Tile 31UET | | | Tile 31UFT | | | Tile 31UFU | | |
|---|---|---|---|---|---|---|---|---|---|
| | OA | PA | UA | OA | PA | UA | OA | PA | UA |
| Water | | 100 | 100 | | 100 | 100 | | 100 | 99 |
| Filed | | 99.5 | 100 | | 95 | 96 | | 99 | 94.2 |
| Soil | | 99.5 | 98.5 | | 98 | 88.3 | | 98 | 85.2 |
| Building | 98.3 | 92.5 | 96.3 | 92 | 70 | 77.7 | 95 | 86 | 92.4 |
| Greenhouse | | 100 | 100 | | 100 | 100 | | 100 | 99 |
| Road | | 97 | 94.2 | | 80 | 82.5 | | 86 | 98 |
| Trees | | 100 | 99.5 | | 100 | 97 | | 97 | 99 |

Table 5.4: Classification results: Overall Accuracy % (OA) for each tile, Producer's Accuracy % (PA) and User's Accuracy % (UA) for each class in each tile.

The classification result can be visually assessed by comparing Figure 5.19a and Figure 5.19b. Figure 5.19a shows the true color composition of the RGB bands representing a portion of tile 31UET, namely a $35 \times 35$ km$^2$ area covering the west coast of the Netherlands, around Delft, where the communication experiments we describe in Section 5.4 are performed. Figure 5.19b depicts the correspondent classification map. In addition, a detail of tile 31UFT, representing a $4 \times 5$ km$^2$ area, is shown in Figure 5.20a, with the correspondent classification map being reported in Figure 5.20b. We can visually assess that the classification result is satisfactory, despite some sparse misclassified pixels in the classes Building and Road can be noticed (e.g., along the highway in Figure 5.20b). These classification results can be further improved by applying more sophisticated approaches than the pure pixel-based approach we use. For instance, the contextual information [136] can be taken into account and/or morphological profiles [21] can be exploited.

Figure 5.19: True color composition of the S2 MS Red, Green and Blue bands (a) and classification map (b) of a $35 \times 35$ km$^2$ area in the west coast of The Netherlands centered at 5779530N, 591970E UTM Zone 31 North (WGS-84).
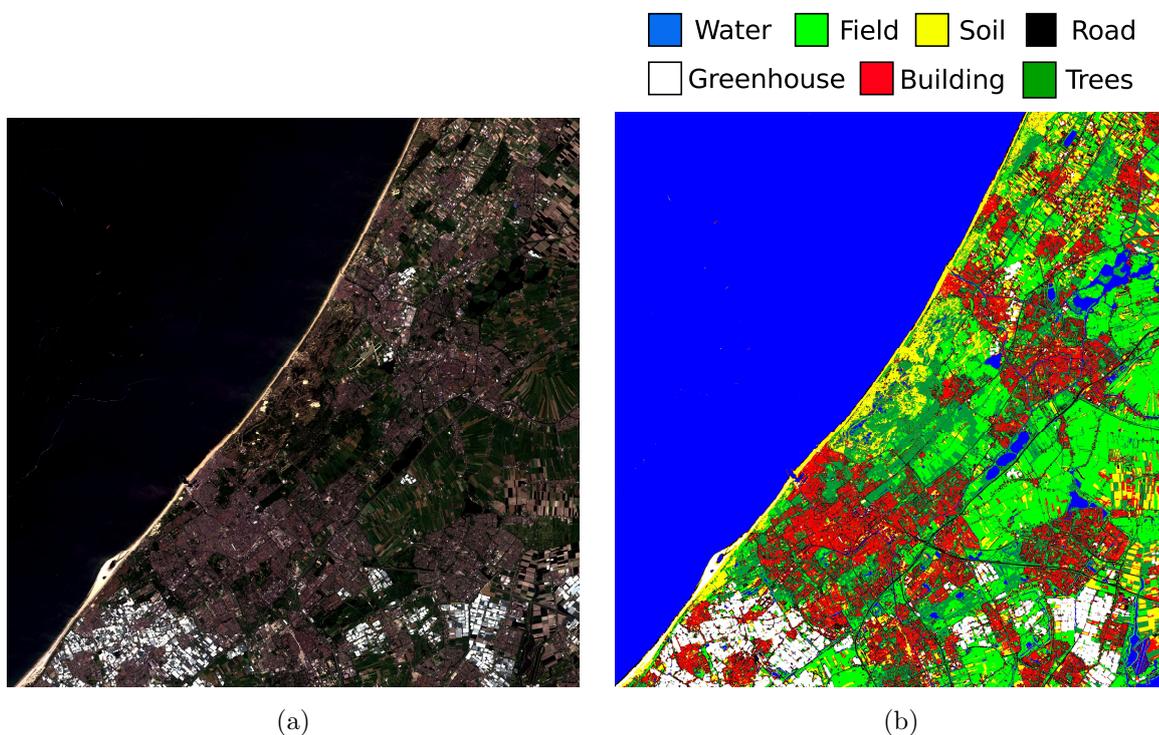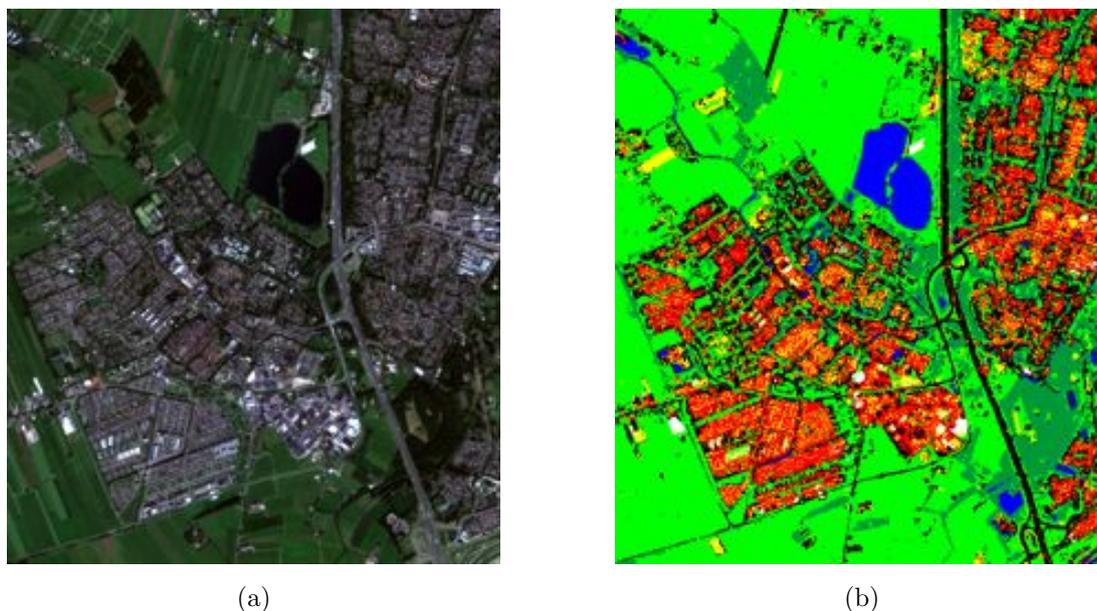


Figure 5.20: True color composition of the S2 MS Red, Green and Blue bands (a) and classification map (b) of a $4 \times 5$ km$^2$ area in the center of The Netherlands centered at 5765970.00N 640730E UTM Zone 31 North (WGS-84).

## 5.4 Ground-level Communication

In this section, the outdoor performance of LoRa communication is evaluated considering a "ground-level" scenario, i.e., with LoRa devices transmitting at few meters from the ground. This is a typical real-world application scenario, which usually presents a diversified communication environment, in that the landscape characteristics may vary from link to link, as well as along each link. For instance, the communication path can traverse build-up areas or farming fields, or both. We focus on analyzing the relationship between the landscape characteristics and the communication performance. To this end, we represent and model the different landscape characteristics in terms of *land cover*, i.e., the land cover type, or class, describes the type of landscape which is traversed by the LoRa signal along its propagation. This analysis is accomplished taking advantage of the *land cover maps* obtained by automatically classifying Sentinel-2 multispectral images, as described in Section 5.3. The land cover maps represent the seven classes characterizing the environment, namely {Water, Field, Building, Greenhouse, Road, Trees}, with 10 m spatial resolution.

In order to analyze and compare the LoRa communication performance w.r.t. the different land cover types, we perform a set of experiments aimed at collecting connectivity traces over different type of areas (Section 5.4.1). We then analyze such traces in combination with the land cover maps to understand and model their correlation (Sections 5.4.2 and 5.4.3). In particular, the focus is on analyzing the correlation between PRR/RSSI and *i)* the distance between the transmitting device and the gateway; and *ii)* the land cover type.

### 5.4.1 Experimental Settings

The LoRa communication performance is tested across urban and rural areas of Delft (The Netherlands). More precisely, the aim of the experiment is to sample and examine the uplink communication (i.e., end-device to gateway) for different distances *and* in the presence of different land cover types. The end-device is composed of a LoRa transceiver based on a Dragino LoRa shield v1.3 embedding a RF96 radio chip and mounting an external antenna with gain 2 dBi. The device is registered on TTN and configured to connect to the in-range TTN gateways. TTN servers provide information about the packets that are received by the TTN gateways. This information is retrieved from TTN servers through MQTT protocol. The LoRa settings we use are: Spreading factor (SF) 7, Bandwidth (BW) 125 KHz and Coding Rate (CR) 4/5. The transmission power is set to 14 dBm.
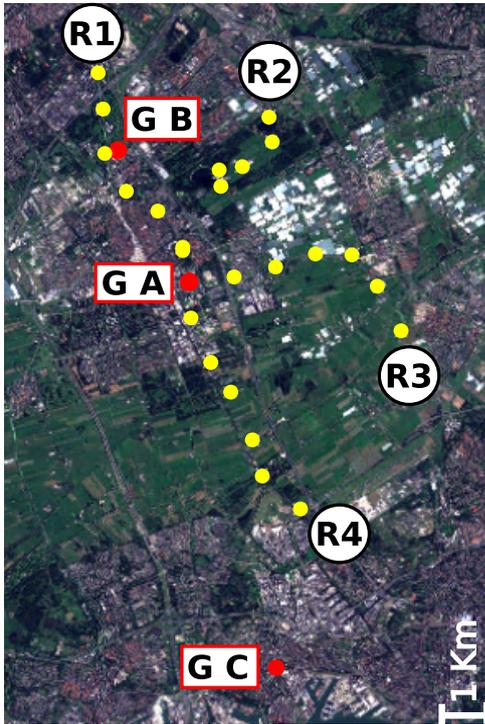
Figure 5.21: Measurement sites (yellow dots) along the four routes R1, R2, R3, R4 and position of gateways G A, G B and G C

We focus on communication up to 6 km with sampling granularity of ∼1 km on the link length. The measurement campaign is planned as follows: we take as reference position the location of the gateway placed at the $22^{nd}$ floor of the EEMCS department building of TU Delft. Hereafter, we refer to this gateway as GA. GA is placed at 62 m above the ground, therefore it is expected to be a good receiver, since, due to its height, has greater chances to favor line-of-sight communication. Starting from the reference position four routes in different directions are considered. The four routes (labeled as R1, R2, R3, and R4) and GA are shown in Figure 5.21. The routes are approximately radial w.r.t. to the reference. The aim was to cover with radial routes different directions, however, due to practical obstacles (e.g., fences preventing the access to farming fields) the routes had to be adapted accounting for the accessible areas. The end-device stops every ∼1 km, along each route, to transmit 30 packets. One packet is sent every 40 seconds. We test 23 measurement sites in total, 6 for each route (apart from R2 where measurements are taken for only 5 sites). The GPS coordinates of each measurement site is recorded. The tests are performed on the 7th of June 2017 for R1 and R3 and on the 12th of June 2017 for R2 and R4. Figure 5.21 shows the four routes together with the measurement sites (yellow dots) and the reference position of GA (red dot).

Across the measurement campaign, also other TTN gateways received the transmitted packets, depending on the transmission site at hand. In particular, two other gateways received our packets and they are reported in Figure 5.21 with labels GB and GC. The dataset collected with this campaign is composed of the GPS position of each measurement site, together with the packets received by each of the three gateways involved and the related information provided by TTN. This information consists in application and device identifiers, packet payload, time of reception, frequency, modulation scheme, data rate, coding rate, gateway identifier, RSSI, SNR and GPS coordinates of the gateway. Overall, the dataset is composed of 482 receptions.

In addition, another set of measurements is performed inside the EEMCS building, very close to gateway GA. More precisely, the LoRa transmitter is placed 1 meter far from the gateway, sending one packet every 45 seconds for 2 hours and 30 minutes, i.e., from 11:30 to 14 on the $15^{th}$ June 2017. All the packets are received from GA resulting in 200 communication samples provided by TTN. These measurements are used to experimentally derive a reference path loss to which refer the path loss analysis for all the other measurements. This reference path loss is the term $PL(d_0)$ in the log-normal path loss model, as defined per Equation (3.1). In this model the effect of the propagation environment is represented by the path loss exponent $n$ and by the gaussian random variable $X_\sigma$. The value of $PL(d_0)$ is sometimes set to the free space path loss at distance $d_0$, or, alternatively it can be determined by measurements performed at $d_0$ [126, 60], which is the approach we follow. $PL(d_0)$ is measured in the far-field region of the gateway antenna and at a close-in reference distance, $d_0 = 1$m, i.e., at a distance smaller than any practical distance used in the communication system. As we describe in detail in the next Section, the value of $PL(d_0)$ that we experimentally derive is used as anchor point in the fitting of the other measurements with the aim of estimating the path loss exponent $n$ according to the different land cover characteristics.

### 5.4.2 Analyzing the Effect of Land Cover on LoRa Links

The goal is to analyze the quality of communication with the three gateways GA, GB and GC for the different measurement sites. The quality of communication is evaluated both in terms of Packet Reception Rate (PRR) and Received Signal Strength Indicator (RSSI).
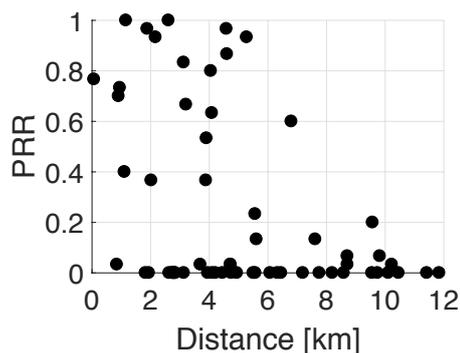


Figure 5.22: PRR vs. distance for all the measurements sites (R1, R2, R3, and R4) and all the gateways (GA, GB and GC)

First, we consider PRR as a function of distance, which is reported for all the measurement sites in Figure 5.22. We can observe that PRR is: *i)* highly variable for distances ranging from 0 to 7 km; *ii)* consistently $\leq 0.2$ between 7 and 11 km; and *iii)* consistently 0 for distances greater than 11 km. In addition, Figure 5.22 shows that no packet reception (PRR = 0) occurs at almost every distance $\geq 2$ km, and that when packet reception occurs, PRR shows a decreasing trend with the distance.

Next, we look at each gateway and each route, separately. Figure 5.23a reports the PRR of GA as a function of distance showing a diversified reception behavior along the different routes. R1, which passes exclusively through buildings, shows a drastic drop in
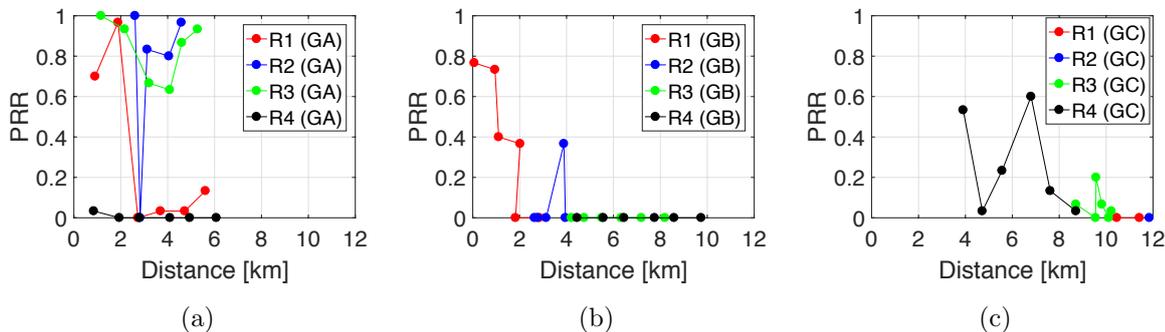
Figure 5.23: PRR vs. distance for GA (a) GB (b) and GC (c), all routes (R1, R2, R3 and R4).

PRR after 3 km, while R2 and R3, which mostly traverse farming fields, exhibit better reception, apart from one specific site in R2 (at $\sim$3 km) where no packet is received by GA. We will further investigate this specific case in Section 5.4.4. Interestingly, GA receives no packets from the farming fields in R4. It only receives very few packets from one single R4 measurement spot at $\sim$1 km. In this respect, we must underline that GA is placed indoor and on the very north-west side of the building, i.e., on the side of R1 in Figure 5.21. Therefore, the body of the building is exactly and entirely in between the gateway and the measurement spots along R4. The shielding effect that the building is causing on the communication from R4 can likely explain the almost complete absence of reception for this route. [4]

GB is placed at 6 m height in the built-up area we traverse along route R1. As it can be noticed in Figure 5.21, we pass very close to it (i.e., 67 meters apart) while performing the experiments. As reported in Figure 5.23b, GB shows decent reception for very short distances in the urban route R1 (i.e., PRR $\geq$ 0.7 for distance $<$ 1 km), with a rapid drop to PRR=0.4 at 1 km and a complete loss of connection farther than 2 km. In addition, GB receives around the 40% of the packets sent from one site in R2, located in the nearby field area.

Finally, GC receives several of the packets sent from the farming fields along R3 and R4, reaching the communication range of $\sim$10 km, as shown in Figure 5.23c. However, the PRR is variable, reaching 0.6 at maximum and remaining consistently $<$0.2 for distances greater than 7 km. Unfortunately, the altitude above the ground of this gateway is not available from TTN. We may only argue it is relatively high, given the long range reception capability it shows.

Overall, we observe a diversified reception behavior for the different gateways as well as for the different characteristics of the environment where the communication occur (e.g.,

---

[4]This same shielding effect w.r.t. the south-east direction has been observed also in other TTN data apart from this measurement campaign.
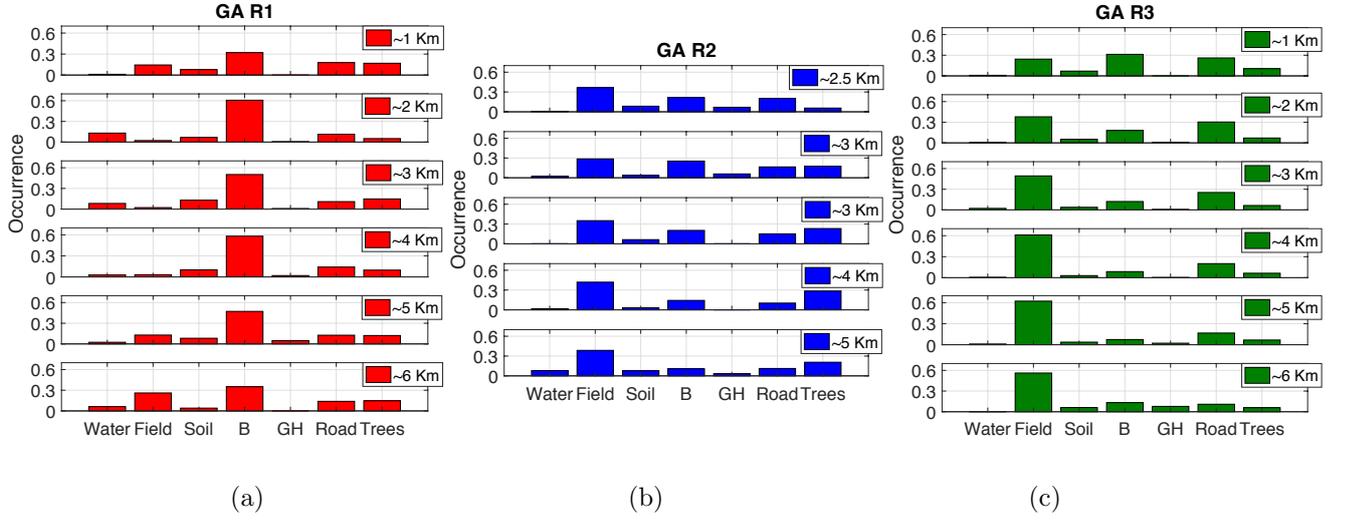
Figure 5.24: Percentage occurrence of the seven land cover classes {Water, Field, Building (B), Greenhouse (GH), Road, Trees} along the link paths connecting the end-device with GA along R1 (a) R3 (b) and R2 (c) for each measurement site. The correspondent distance is reported in the legend.

built-up areas vs. open field areas). Therefore, we analyze more in detail the relationship between the PRR and the characteristics of the environment exploiting the land cover map we derived from Sentinel-2 images. More precisely, we retrieve from the map the sequence of land cover classes that are present along the path connecting the transmitting device position and the receiving gateway position, for each link. The spatial granularity of the sequence is defined by the geometric resolution of the Sentinel-2 multispectral images (i.e., 10 m), from which the land cover map is derived. Then, we compute the occurrence (i.e., the percentage presence) of each land cover class in the sequence that is associated to each link. Figure 5.24a shows the occurrence of the seven considered land cover classes along the link paths that connect the sites in R1 with GA. Each subplot refers to one site and the subplots are ordered by increasing site-GA distance. The approximated distances are reported in the legends for comparison. Class Building (abbreviated as B in the plot) is the prevailing class in the link paths for all of the six measurement sites in R1, with the percentage of presence ranging from 32% to 61%, depending on the site. In contrast, Figure 5.24c shows that for route R3 the prevailing class is Field (from 38% to 62%), apart from the closest measurement site at ~1 km from GA. For this site Building prevails (31%), and the classes Field and Road are also significantly present, covering the 24% and 26% of the link path, respectively. Figure 5.24b shows that for the links established along R2 the Field class is the most present (i.e., consistently ≥30%). In this case, however, the predominance is less pronounced with a considerable presence of Building at 3 km

(25%) and Trees at 4 km (28%). We previously observed (see Figure 5.23a) that the GA PRR vs. distance behavior diverge significantly on R1 and R3, for distances >3 km, with PRR > 0.6 and PRR < 0.2, respectively. These trends reflect a strong predominance of different land cover classes along the links (i.e., Building on R1 and Field on R3) as shown in Figures 5.24a and 5.24c. A similar consideration can be done by comparing the PRR behavior for distances >3 km of R1 (Building dominated) and R2 (Field dominated).
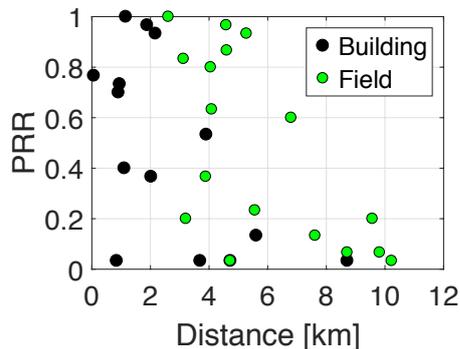


Figure 5.25: PRR vs. distance for Building and Filed dominated links in black and green, respectively.

By computing the land cover sequence and the predominant land cover class for all of the links involved in the experimental campaign, we observe that all of the the links showing PRR > 0 are characterized by prevailing class Building or Field. We classify and label these links according to the correspondent prevailing class and then compare the PRR vs. distance behavior w.r.t. these classes, as shown in Figure 5.25. The PRR for Building dominated links is represented in black and the PRR for Field dominated links in green. Since the gateways are usually placed in built-up areas, we do not have short range measurements for the Field class to directly compare with Building. However, we see that the Field packets are received at longer distances and, mostly, with better PRR than the Building packets at comparable distance. Since examining the PRR we can only ascertain a difference in the packet reception capability in the presence of different land cover types, we move the investigation at the very physical level to analyze the causes determining such reception behavior, namely we analyze the impact of the land cover on the signal propagation and attenuation.

**RSSI analysis.** Figure 5.26 reports the RSSI vs. distance for all the available RSSI measurements (i.e., when PRR > 0). The expected reception power according to the Free Space Model is also reported for comparison (black dashed line). We note that the Free Space Model significantly differ from the measurements, overestimating the expected reception power with an average error in the order of 20 dBm. This quantitatively shows the difference in the attenuation experienced by the signal between this "ground" scenario and the previously considered free space like scenario. In addition, Figure 5.26 reports the expected reception power according to the path loss model proposed by Bor et al. [30]. The authors estimate the log-normal path loss model parameters, namely $PL(d_0)$, $n$ and $\sigma$, from a set of experimental measurements performed in a built-up envi-

Figure 5.26: RSSI vs. distance for all the measurements (blue), Free Space model (black dashed line) and Bor's model (plain black line).
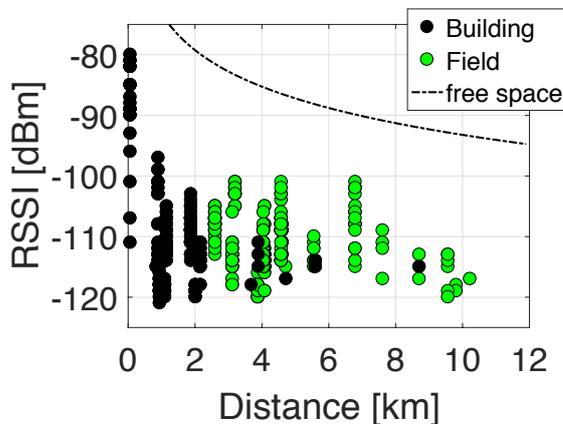
Figure 5.27: RSSI vs. distance for Building dominated links (black dots) and Field dominated links (green dots) and Free Space model (black dashed line).

ronment with a Semtech SX1272 LoRa transceiver and Spreading Factor 12. In particular, they estimate $PL(d_0) = 17.41$ dB, with $d_0 = 40$m, $n = 2.08$ and $\sigma = 3.57$. The Bor's model significantly underestimates our RSSI measurements, likely due to the different (e.g., more densely built-up) communication environment where the measurements have been collected. However, as we show and discuss later on, this discrepancy is much less pronounced w.r.t. one subset of our measurements, namely the Building dominated communication traces between the transmission sites along R1 and the low-altitude gateway GB.

Next, we compare the RSSI vs. distance of Building dominated links and Field dominated links, which are reported in Figure 5.27 in black and green, respectively. The communication range is larger for Fileds than for Building, as we noticed also in the PRR analysis. In addition, we can identify a difference in the RSSI decay rate with the distance between the two classes. It is particularly interesting to further investigate this aspect by comparing the different classes for the different gateways. Figure 5.28 shows the RSSI vs. distance distinguishing both the class Building/Field and the receiving gateway. We observe the strongest decay rate for GB Building (i.e., gateway at 6 m height), followed by GA Building (i.e., gateway at 62 m height). GA Field decays more smoothly and GC Field exhibits the less pronounced decay rate, reaching the longest communication range.

### 5.4.3 Analyzing the Power Decay Rate

The different trends in the power decay rate are quantified by separately fitting the clusters of measurements corresponding to each gateway/class combination. The fitting is based
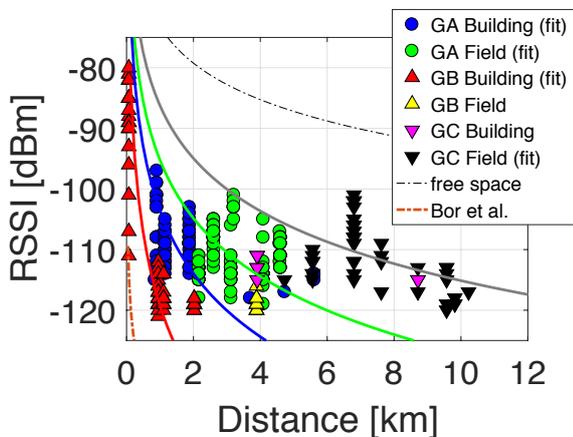
Figure 5.28: RSSI vs. distance for the different gateways and link dominating land cover class Building/Field, Free Space model (black dashed line) and fitting curves for GA Building, GA Field, GB Building and GC Field.
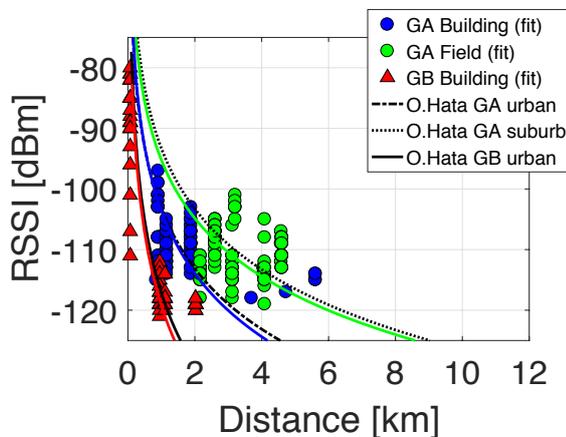
Figure 5.29: RSSI vs. distance and fitting curves for GA Building, GA Field, GB Building and Okumura-Hata model urban/suburban according to the dominating land cover class.

on the log-normal path loss model. More precisely, the fitting function $f(n, d)$ is defined by Equation (5.10), which relates the RSSI measurements with the model parameters $n$ and $d$ through the Friis Equation (5.1).

$$f(n, d) = P_{tx} + G_{tx} + G_{rx} - PL(d_0) - 10 \cdot n \cdot log_{10}\left(\frac{d}{d_0}\right) \qquad (5.10)$$

$d$ is the distance in meters, $n$ is the path loss exponent we want to determine through fitting (i.e., the fitting parameter), $P_{tx} = 14$ dBm is the transmission power we used in the experiments, $G_{tx} = 2$ dBi is the gain of the transmitting antenna and $G_{rx} = 2$ dBi is the gain of the receiving antenna. GA mounts a typical half-wave dipole antenna with gain 2 dBi; we assume that also GB and GC mount a similar antenna. We use as anchor point the path loss at the reference distance $d_0 = 1$m ($PL(d_0)$) that we experimentally measured for GA. The experiment provided a stable connection with average path loss $PL(d_0) = 23.9$ dB and standard deviation 1.1 dB.

We solve the non linear curve fitting problem by least squares. The value of $n$ that best fits the nonlinear function $f(n, d)$ to the RSSI measurements is searched starting from $n_0 = 2$ (i.e., the free space path loss exponent). We do not fit the data for GB Field and GC Building, due to the very few measurements available, i.e., only 11 samples at one single distance for the former and 17 samples for the latter. Table 5.5 reports the fitting results we obtain for the different gateway-land cover class combinations, in terms of path loss exponent $n$, shadowing $\sigma$, i.e., the standard deviation of the gaussian random variable $X_\sigma$ in Equation (3.1), and average error in dB between the fitted

model and the measurements. We see that $n$ is much larger for Building than for Field ([3.46, 3.98] vs. [2.88, 3.18]), indicating a stronger attenuation rate in the former class.

| subset | $n$ | $\sigma$ | avg err [dB] |
|---|---|---|---|
| GA Building | 3.46 | 3.73 | 5.34 |
| GB Building | 3.98 | 8.05 | 8.94 |
| GA Field | 3.18 | 3.13 | 4.38 |
| GC Field | 2.88 | 2.64 | 4.03 |

Table 5.5: Least square estimate of the path loss exponent $n$, standard deviation $\sigma$ of the gaussian random variable $X_\sigma$ and average error between the fitted model and the measurements.

Moreover, for comparable classes, $n$ increases as the gateway height decreases, e.g., $n = 3.46$ for GA Building (62 m height) vs. $n = 3.98$ for GB Building (6 m height), thus indicating a stronger attenuation for the lower gateway. $\sigma$ is in the order of 3 dB, and the error is in the order of 5 dB, apart from GB Building, where $\sigma$ and the error reach more than 8 dB. This is due to the high variability in the RSSI measured in the site at 64 m from GB (see Figure 5.28). This variability is likely caused by dynamic components as the presence of mobile obstructors or reflectors (e.g., cars) in the very proximity of the transmitter/receiver.

The fitting results can be visually evaluated in Figure 5.28, which reports the curves corresponding to the expected received power according to the log-normal path loss model tuned with the different values of $n$. Each curve is represented with the color of the corresponding measurements, i.e., red for GB Building, blue for GA Building, green for GA Field and black for GC Field. We see that the trends are well captured. In addition, we can note that the fitting curve for GB building (red) is closer to the Bor's model (orange) than the other fitting curves, thus indicating that the signal attenuation characteristics observed by Bor et al. are closer to those we observed in our built-up environment when a low altitude gateway is considered. This comparison underlines that $i)$ a number of factors affect the communication, including the gateway altitude combined with the specific environment characteristics; and $ii)$ it is hard to define one general model able to effectively capture the complexity and variety of real-world scenarios, without explicitly taking into account their different characteristics and the related effects.

Finally, we compare the fitting results with the Okumura-Hata model [126, 112, 64]. It is an empirical path loss model defined as per Equation (5.11), which takes explicitly into account the height of the transmitter ($h_m$) and of the receiver ($h_b$). $h_m$ and $h_b$ are in meters, the distance $d$ is in kilometers and the term $a(h_m)$ is a correction factor for $h_m$, which is defined as per Equation 5.12 for small-medium cities.

$$L_U[dB] = 69.55 + 26.16\ log_{10}(f) - 13.82\ log_{10}(h_b) - a(h_m) + (44.9 - 6.55\ log_{10}(h_b))\ log_{10}(d) \tag{5.11}$$

$$a(h_m)[dB] = (1.1\ log_{10}(f) - 0.7)h_m - (1.56\ log_{10}(f) - 0.8) \tag{5.12}$$

In addition, the corrections in Equation 5.13 and Equation 5.14 are applied for suburban

and rural environments, respectively, with $K \in [35.94 \; 40.94]$.

$$L_{SU}[dB] = L_U - 2 \left( log_{10} \left( \frac{f}{28} \right) \right)^2 - 5.4 \tag{5.13}$$

$$L_R[dB] = L_U - 4.78 \; (log_{10}(f))^2 + 10.33 \; log_{10}(f) - K \tag{5.14}$$

We apply the urban (small-medium cities) model for the Building dominated links and the suburban correction for the Field dominated links. In the latter case, we select the suburban correction (instead of the rural correction), since the analysis of the land cover occurrence along the link paths (see Figure 5.24) shows that, despite the predominance of Field, Building is anyhow present. The transmitter height is $h_m = 1.5$m and the receiver height $h_b$ is the height of the considered gateway, i.e., $h_b = 62$m for GA and $h_b = 6$m for GB. The height of GC is not available, we therefore do not compute the Okumura-Hata model for this gateway. In Figure 5.29 we compare our fitted models with the Okumura-Hata predictions, driven by the land cover dominating class. We see that the fitted curves are very similar to the Okumura-Hata curves, i.e., the difference is in the order of only 1 dB. This result underlines that the signal attenuation rates described by Okumura and Hata are very similar to those observed for the LoRa signal, as long as the communication environment (e.g., the predominant land cover class along the link paths) is taken into account.

This analysis provides several insights about the effect of the static characteristics of the environment on the communication and shows how some of these characteristics can be *quantitatively* represented and accounted for by leveraging multispectral satellite images. In addition, this analysis suggests that *i)* the predominant land cover class along the link is a reasonable representation that enables to capture the general trend of the attenuation rate in the considered dataset; and *ii)* information about the predominant land cover can be exploited to drive the selection of a proper Okumura-Hata model with the aim of predicting the expected LoRa signal attenuation rate in real outdoor scenarios.

However, despite the predominant land cover shows to be useful to capture the general trend of the signal attenuation, it is not precise enough to represent all the links we observed, or more precisely, the links we could not observe since PRR = 0. We further investigate the land cover characteristics for these links in the next section.

### 5.4.4 Fine-grained Land Cover Analysis

Figure 5.23a shows that one measurement spot along R2 (i.e., at 2.8 km from GA) exhibits no packet reception from GA (PRR = 0), despite presenting the same land cover dominating class (i.e., Field) as the other spots in R2. To further investigate this behavior we analyze the land cover more in detail and observe that the percentage occurrence of the land cover classes in the vicinity of the transmitter in this measurement site presents peculiar characteristics, which do not reflect the overall occurrence for the whole link.
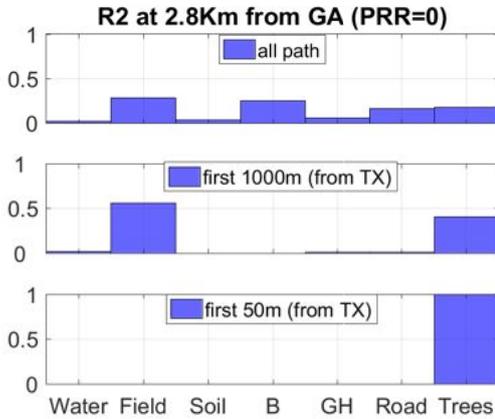
Figure 5.30: Occurrence of the land cover class in the whole link path, in the first kilometer from the transmitter and in the first 50m from the transmitter, for the measurement site at 2.8 km from GA in R2, where PRR = 0.
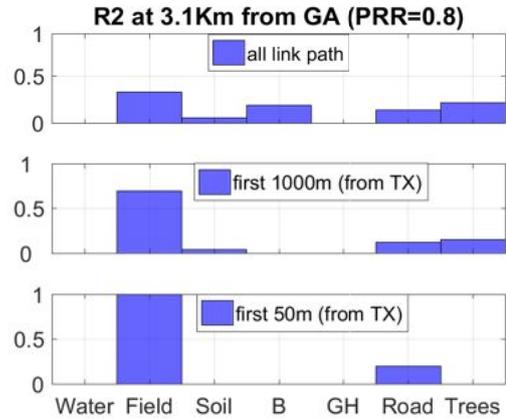
Figure 5.31: Occurrence of the land cover class in the whole link path, in the first kilometer from the transmitter and in the first 50m from the transmitter, for the measurement site at 3.1 km from GA in R2, where PRR = 0.8.

Figure 5.30 shows the land cover occurrence in the whole link path (first subplot), in the first kilometer from the transmitter (second subplot) and in the very first 50 m from the transmitters (third subplot). We see that Field is predominant in the overall path, but in the first kilometer Trees is 41% present and in the very first 50m we find 100% Trees, i.e., trees are exactly in front of the transmitting device (1.5 m high) w.r.t. GA, thus obstructing the line-of-sight communication with the gateway. Indeed, no packet is received from this location. On the contrary, Figure 5.31 shows the same analysis for the close measurement site in R2, at 3.1 km from GA. In this case, 80% of packets are received (PRR = 0.8) and Field remains predominant in the whole path, as well as in the first kilometer from the transmitter and in the first 50 m from the transmitter. Similarly, the presence of the entire EEMCS department building in front of GA is completely obstructing the communication with the sites along R4.

We can conclude that the relative position of the land cover type w.r.t. both transmitter (i.e., end-device) and receiver (i.e., gateway) matters, and it acts in combination with the respective height of transmitter and receiver. Clearly, the height of buildings and trees is relevant and could be accounted for through the availability of LiDAR data or other sources of information as cadastral maps. However, also in the absence of such data, high resolution land cover maps provide a detailed representation of the horizontal structure of the scene, in that they provide information about the spatial distribution of objects in the environment and about their relative positions w.r.t. the communication devices. This information can be exploited to further investigate the interaction between the structure

of the communication environment and the communication performance, with the aim of improving the modeling and prediction of LoRa connectivity.

### 5.4.5 Conclusion

Land cover classification maps derived from multispectral satellite images can support both the analysis and the modeling of LoRa outdoor communication performance, by *quantitatively* introducing the representation of the static characteristics of the environment. More precisely, these land cover maps allow to *automatically* assess what kind of objects are present and where they are placed in the communication environment, thus enabling the development of systematic analysis accounting for their presence and effect on communication. We showed that the predominant land cover along the link path represents a first indicator for capturing the signal attenuation trend. However, in order to better model LoRa connectivity, the characteristics of the environment should be accounted for with more detail. To this end, the high level of spatial detail characterizing the land cover maps we are considering represents a powerful instrument. It remains an open issue how to take full advantage of such level of detail in this context, opening interesting venues for further investigation. We plan to analyze bigger datasets of LoRa connectivity traces, presenting more variability in the land cover characteristics. One option is to investigate the applicability of machine learning techniques to learn and model the relationship between the transmission parameters (e.g., LoRa settings as Spreading Factor), the end device and gateway altitudes, the fine-grained land cover characteristics along the link paths, as well as other environmental factors as temperature and weather conditions. This approach could lead to the development of connectivity models for outdoor LoRa communication to be applied for the prediction of the expected communication performance.

# Chapter 6

# Conclusion and Outlook

In this dissertation, we addressed the challenges posed by real-world outdoor environments on the deployment of low-power wireless networks. In particular, we focussed on the effect of the static characteristics of the environment on low-power wireless links. The main contribution we put forward is to demonstrate the potential carried by remote sensing systems and the related data analysis techniques for the definition of realistic low-power connectivity models in outdoor scenarios and for the development of effective deployment strategies. In this context, we showed that remote sensing systems and data represent a powerful mean for *i)* identifying the static properties of the environment that affect low power communication performance; *ii)* automatically and quantitatively represent such properties; and *iii)* develop connectivity models and deployment strategies that take them into account.

We exploited LiDAR data to analyze the impact of the forest structure on short-range low-power wireless at 2.4 GHz. This analysis enabled the definition of a specialized LiDAR-based radio attenuation model for trunk-level communication in forests. In addition, we showed how prior knowledge about the forest structure derived from LiDAR can be exploited to develop informed node placement planning strategies accounting for both the actual forest structure and its effect on the network, automatically and prior to the deployment. The evolutionary optimization implemented in LaPS combined with the LiDAR-derived inputs (i.e., a representation of the forest attributes and the specialized radio model) provided significant gains w.r.t. the simplistic approaches commonly used. Moreover, LaPS proved to be a valuable instrument for assessing the connectivity of a target forest and evaluate the attainable trade-offs between user requirements and robustness of the network, prior to the actual in-field deployment.

Then, we have quantified the gap in communication range that the presence of objects in the environment causes when LoRa low-power technology is exploited in the context of IoT networks. Specifically, we compared the communication performance achieved in a free-space like scenario (i.e., when a LoRa transmitter on an high altitude balloon communicates with gateways on the ground) against those achieved in a ground level

communication scenario characterized by a diversified environment. Moreover, we have analyzed the effect of the type of environment (i.e., land-cover type) on LoRa communication performance by leveraging multispectral satellite images. We showed how high-detailed land-cover maps (i.e., with spatial resolution 10 m) can enable effective analysis that relate the land cover characteristics with the signal attenuation, which in turn affects LoRa communication range. These contributions can be applied to plan the deployment of WSNs in forests and to enable the development of LoRa connectivity models that account for the actual communication environment.

Despite the promising and valuable results, further investigation can significantly improve the connectivity models in both considered contexts. Clearly, such further development requires to gather additional connectivity measurements performing more in-field experiments, which was a very effort-demanding part of this work. The signal attenuation model for low-power short-range communication in forests can be refined by considering other environmental variables as temperature and weather conditions, as well as other important static features as the terrain morphology and the presence of foliage. In this respect, LiDAR can support the quantitative representation of both these static features. In addition, optical images can be exploited to derive indicators as the Normalized Difference Vegetation Index (NDVI) and the Leaf Area Index (LAI) that can provide further information about the status of vegetation. These additional sources of information can be integrated in the formulation of the node placement problem and included in the LaPS methodology. LaPS is indeed open and flexible, in that new constraints can be easily added and the fitness function, which drives the search for (near-)optimal placement solutions, can be enriched with additional terms. Moreover, the genetic algorithms that carry out the evolutionary search process are widely applied in the context of multi-objective optimization problems, thus rendering LaPS suitable for encompassing further aspects of the problem.

In the context of LoRa connectivity modeling, the preliminary analysis we presented is mostly based on the predominant land cover class along the link path. One relevant topic for future research is the refinement of this preliminary analysis to take full advantage of the high level of spatial detail that characterizes the land cover maps we derive from Sentinel-2 multispectral images. However, this fine detail increases the dimensionality of the problem and the complexity of the modeling. Methods that are capable to properly handle such level of detail and effectively take advantage of it must be delineated. In this respect, a promising line of investigation is the evaluation of the effectiveness of machine learning techniques for learning the complex relation between the numerous parameters in play, e.g., the fine-grained land cover information, the different LoRa communication settings, and other environmental variables as the weather conditions. The modeling

of this complex relation can lead to the definition of accurate connectivity models that can support the effective deployment of LoRa gateways in the context of real-world IoT applications, accounting for the actual environment surrounding the gateways and for the specific coverage requirements posed by the application.

We argue that the analysis and the methodologies we presented advance the state of the art on low power connectivity modeling and network planning in real world environments. We envision that our contributions can pave the way for further improvements in the understanding and quantitative modeling of the factors that make the link behavior difficult to predict, thus fostering the development of realistic simulations and rendering the deployment of low power networks more predictable.

# Bibliography

[1] https://www.thethingsnetwork.org/.

[2] http://scihub.copernicus.eu/.

[3] https://sentinel.esa.int.

[4] Mamta Agiwal, Abhishek Roy, and Navrati Saxena. Next Generation 5G Wireless Networks: A Comprehensive Survey. *IEEE Communications Surveys Tutorials*, 18(3):1617–1655, 2016.

[5] Vahab Akbarzadeh, Albert Hung-Ren Ko, Christian Gagné, and Marc Parizeau. Topography-aware sensor deployment optimization with CMA-ES. In *International Conference on Parallel Problem Solving from Nature*, pages 141–150. Springer, 2010.

[6] Fadi Al-Turjman, Hossam S. Hassanein, and Mohamed A. Ibnkahla. Connectivity optimization for wireless sensor networks applied to forest monitoring. In *IEEE International Conference on Communications, 2009 (ICC'09)*, pages 1–6.

[7] LoRa Alliance. LoRaWAN specification. 2015.

[8] LoRa Alliance. A technical overview of LoRa and LoRaWAN. 2015.

[9] Giuseppe Anastasi, Alessio Falchi, Andrea Passarella, Marco Conti, and Enrico Gregori. Performance measurements of motes sensor networks. In *Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems*, pages 174–181. ACM, 2004.

[10] Imen Arfaoui, Ramzi Bellazreg, and Noureddine Boudriga. A novel realistic irregular radio model to enhance coverage evaluation in wireless sensor networks. In *22nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 160–164. IEEE, 2014.

[11] Aloÿs Augustin, Jiazi Yi, Thomas Clausen, and William Mark Townsley. A study of LoRa: Long range & low power networks for the internet of things. *Sensors*, 16(9):1466, 2016.

[12] Abdalkarim Awad, Thorsten Frunzke, and Falko Dressler. Adaptive distance estimation and localization in WSN using RSSI measures. In *10th Euromicro Conference on Digital System Design Architectures, Methods and Tools (DSD 2007)*, pages 471–478. IEEE, 2007.

[13] Peter Axelsson. DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing*, 33(B4/1):111–118, 2000.

[14] Joaquim A. R. Azevedo and Filipe E. S. Santos. An empirical propagation model for forest environments at tree trunk level. *IEEE Transactions on Antennas and Propagation*, 59(6), 2011.

[15] Nouha Baccour, Anis Koubâa, Luca Mottola, Marco Antonio Zúñiga, Habib Youssef, Carlo Alberto Boano, and Mário Alves. Radio link quality estimation in wireless sensor networks: A survey. *ACM Transactions on Sensor Networks (TOSN)*, 8(4):34, 2012.

[16] Nouha Baccour, Daniele Puccinelli, Thiemo Voigt, Anis Koubaa, Claro Noda, Hossein Fotouhi, Mario Alves, Habib Youssef, Marco Antonio Zuniga, Carlo Alberto Boano, et al. External radio interference. In *Radio Link Quality Estimation in Low-Power Wireless Networks*, pages 21–63. Springer, 2013.

[17] Emmanuel P. Baltsavias. A comparison between photogrammetry and laser scanning. *ISPRS Journal of photogrammetry and Remote Sensing*, 54(2-3):83–94, 1999.

[18] Kenneth Bannister, Gianni Giorgetti, and Sandeep K. S. Gupta. Wireless sensor networking for hot applications: Effects of temperature on signal strength, data collection and localization. In *Proceedings of the 5th Workshop on Embedded Networked Sensors (HotEmNets 08)*. Citeseer, 2008.

[19] W.G.M. Bastiaanssen, E.J.M. Noordman, H. Pelgrum, G. Davids, B.P. Thoreson, and R.G. Allen. SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. *Journal of irrigation and drainage engineering*, 131(1):85–93, 2005.

[20] A.I. Belousov, S.A. Verzakov, and J. Von Frese. A flexible classification approach with optimal generalisation performance: support vector machines. *Chemometrics and intelligent laboratory systems*, 64(1):15–25, 2002.

[21] Jon Atli Benediktsson, Martino Pesaresi, and Kolbeinn Amason. Classification and feature extraction for remote sensing images from urban areas based on morpho-

logical transformations. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1940–1949, 2003.

[22] Jan Beutel, Stephan Gruber, Andreas Hasler, Roman Lim, Andreas Meier, Christian Plessl, Igor Talzi, Lothar Thiele, Christian Tschudin, Matthias Woehrle, et al. PermaDAQ: A scientific instrument for precision sensing and data recovery in environmental extremes. In *Proceedings of the 2009 International Conference on Information Processing in Sensor Networks*, pages 265–276. IEEE Computer Society, 2009.

[23] Amol P. Bhondekar, Renu Vig, Madan Lal Singla, C. Ghanshyam, and Pawan Kapur. Genetic algorithm based node placement methodology for wireless sensor networks. In *Proceedings of the international multiconference of engineers and computer scientists*, volume 1, pages 18–20. Citeseer, 2009.

[24] Edoardo S. Biagioni and Galen Sasaki. Wireless sensor placement for reliable and efficient data collection. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, pages 10–pp. IEEE, 2003.

[25] Carlo Alberto Boano, Kay Römer, Fredrik Österlind, and Thiemo Voigt. Realistic simulation of radio interference in cooja. In *European Conference on Wireless Sensor Networks (EWSN 2011), February 2011, Bonn, Germany*, 2011.

[26] Carlo Alberto Boano, Nicolas Tsiftes, Thiemo Voigt, James Brown, and Utz Roedig. The impact of temperature on outdoor industrial sensornet applications. *IEEE Transactions on Industrial Informatics*, 6(3):451–459, 2010.

[27] Carlo Alberto Boano, Hjalmar Wennerstrom, Marco Zuniga, James Brown, Chamath Keppitiyagama, Felix Oppermann, Utz Roedig, Lars-Ake Norden, Thiemo Voigt, and Kay Romer. Hot packets: A systematic evaluation of the effect of temperature on low power wireless transceivers. In *Proceedings of the 5th Extreme Conference on Communication (ExtremeCom)*, 2013.

[28] Carlo Alberto Boano, Marco Zúñiga, James Brown, Utz Roedig, Chamath Keppitiyagama, and Kay Römer. Templab: A testbed infrastructure to study the impact of temperature on wireless sensor networks. In *Proceedings of the 13th international symposium on Information processing in sensor networks (IPSN)*, pages 95–106. IEEE Press, 2014.

[29] Martin Bor, John Edward Vidler, and Utz Roedig. LoRa for the Internet of Things. pages 361–366, 2016.

[30] Martin C. Bor, Utz Roedig, Thiemo Voigt, and Juan M. Alonso. 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*, pages 59–67. ACM, 2016.

[31] Léon Bottou, Corinna Cortes, John S. Denker, Harris Drucker, Isabelle Guyon, Lawrence D. Jackel, Yann LeCun, Urs A. Muller, Edward Sackinger, Patrice Simard, et al. Comparison of classifier methods: a case study in handwritten digit recognition. In *Pattern Recognition, 1994. Vol. 2-Conference B: Computer Vision & Image Processing., Proceedings of the 12th IAPR International. Conference on*, volume 2, pages 77–82. IEEE, 1994.

[32] Lorenzo Bruzzone, Mingmin Chi, and Mattia Marconcini. A novel transductive SVM for semisupervised classification of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(11):3363–3373, 2006.

[33] Christopher J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[34] James B. Campbell and Randolph H. Wynne. *Introduction to remote sensing*. Guilford Press, 2011.

[35] Gustavo Camps-Valls and Lorenzo Bruzzone. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1351–1362, 2005.

[36] Gustavo Camps-Valls and Lorenzo Bruzzone. *Kernel methods for remote sensing data analysis*. John Wiley & Sons, 2009.

[37] Benji Capsuto and Jeff Frolik. A system to monitor signal fade due to weather phenomena for outdoor sensor systems. In *Fifth International Conference on Information Processing in Sensor Networks (IPSN 2006)*, 2006.

[38] Marco Cattani, Carlo Alberto Boano, and Kay Römer. An experimental evaluation of the reliability of LoRa long-range low-power wireless communication. *Journal of Sensor and Actuator Networks*, 6(2):7, 2017.

[39] Marco Centenaro, Lorenzo Vangelista, Andrea Zanella, and Michele Zorzi. Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios. *IEEE Wireless Communications*, 23(5):60–67, 2016.

[40] Matteo Ceriotti, Matteo Chini, Amy L. Murphy, Gian Pietro Picco, Francesca Cagnacci, and Bryony Tolhurst. Motes in the jungle: lessons learned from a short-term WSN deployment in the Ecuador cloud forest. In *Real-World Wireless Sensor Networks*, pages 25–36. Springer, 2010.

[41] Matteo Ceriotti, Michele Corrà, Leandro D'Orazio, Roberto Doriguzzi, Daniele Facchin, Gian Paolo Jesi, Renato Lo Cigno, Luca Mottola, Amy L. Murphy, Massimo Pescalli, et al. Is there light at the ends of the tunnel? Wireless sensor networks for adaptive lighting in road tunnels. In *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on*, pages 187–198. IEEE, 2011.

[42] Alberto Cerpa, Naim Busek, and Deborah Estrin. SCALE: A tool for simple connectivity assessment in lossy environments. *Center for Embedded Network Sensing*, 2003.

[43] Alberto Cerpa, Jennifer L. Wong, Louane Kuang, Miodrag Potkonjak, and Deborah Estrin. Statistical model of lossy links in wireless sensor networks. In *Proceedings of the 4th international symposium on Information processing in sensor networks*, page 11. IEEE Press, 2005.

[44] Yin Chen and Andreas Terzis. On the implications of the log-normal path loss model: an efficient method to deploy and move sensor motes. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, pages 26–39. ACM, 2011.

[45] Michele Dalponte, Lorenzo Bruzzone, Loris Vescovo, and Damiano Gianelle. The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment*, 113(11):2345–2355, 2009.

[46] Begüm Demir, Francesca Bovolo, and Lorenzo Bruzzone. Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1):300–312, 2013.

[47] Reinhard Diestel. Graph theory, volume 173 of. *Graduate texts in mathematics*, page 5, 2010.

[48] Laigen Dong and Jie Shan. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, 84:85–99, 2013.

[49] David M. Doolin and Nicholas Sitar. Wireless sensors for wildfire monitoring. In *Proc. SPIE*, volume 5765, pages 477–484, 2005.

[50] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment*, 120:25–36, 2012.

[51] Yun Du, Yihang Zhang, Feng Ling, Qunming Wang, Wenbo Li, and Xiaodong Li. Water bodies mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sensing*, 8(4):354, 2016.

[52] Wendy Ellens and Robert E Kooij. Graph measures and network robustness. *arXiv preprint arXiv:1311.5064*, 2013.

[53] Christopher D. Elvidge and Zhikang Chen. Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote Sensing of Environment*, 54(1):38–48, 1995.

[54] Konstantinos P. Ferentinos and Theodore A. Tsiligiridis. Adaptive design optimization of wireless sensor networks using genetic algorithms. *Computer Networks*, 51(4):1031–1051, 2007.

[55] Richard Figura, Matteo Ceriotti, Chia-yen Shih, Margarita Mulero-Pázmány, Songwei Fu, Roberta Daidone, Sascha Jungen, Juanjo José Negro, and Pedro José Marrón. Iris: Efficient visualization, data analysis and experiment management for wireless sensor networks. 2014.

[56] Luca Filipponi, Andrea Vitaletti, Giada Landi, Vincenzo Memeo, Giorgio Laura, and Paolo Pucci. Smart city: An event driven architecture for monitoring public spaces with heterogeneous sensors. In *Sensor Technologies and Applications (SENSORCOMM), 2010 Fourth International Conference on*, pages 281–286. IEEE, 2010.

[57] Deepak Ganesan, Bhaskar Krishnamachari, Alec Woo, David Culler, Deborah Estrin, and Stephen Wicker. Complex behavior at scale: An experimental study of low-power wireless sensor networks. Technical report, Citeseer, 2002.

[58] Jose Antonio Gay-Fernandez et al. Propagation analysis and deployment of a wireless sensor network in a forest. *Progress In Electromagnetics Research*, 106:121–145, 2010.

[59] Orestis Georgiou and Usman Raza. Low power wide area network analysis: Can LoRa scale? *IEEE Wireless Communications Letters*, 6(2):162–165, 2017.

[60] Andrea Goldsmith. *Wireless communications*. Cambridge university press, 2005.

[61] J.A. Gualtieri, Samir R. Chettri, R.F. Cromp, and L.F. Johnson. Support vector machine classifiers as applied to AVIRIS data. In *Proc. Eighth JPL Airborne Geoscience Workshop*, 1999.

[62] Jane K. Hart and Kirk Martinez. Environmental sensor networks: A revolution in the Earth system science? *Earth-Science Reviews*, 78(3):177–191, 2006.

[63] Carl Hartung, Richard Han, Carl Seielstad, and Saxon Holbrook. FireWxNet: A multi-tiered portable wireless system for monitoring weather conditions in wildland fire environments. In *Proceedings of the 4th international conference on Mobile systems, applications and services*, pages 28–41. ACM, 2006.

[64] Masaharu Hata. Empirical formula for propagation loss in land mobile radio services. *IEEE transactions on Vehicular Technology*, 29(3):317–325, 1980.

[65] Tian He, Chengdu Huang, Brian M Blum, John A Stankovic, and Tarek Abdelzaher. Range-free localization schemes for large scale sensor networks. In *Proceedings of the 9th annual international conference on Mobile computing and networking*, pages 81–95. ACM, 2003.

[66] Tian He, Sudha Krishnamurthy, John A Stankovic, Tarek Abdelzaher, Liqian Luo, Radu Stoleru, Ting Yan, Lin Gu, Jonathan Hui, and Bruce Krogh. Energy-efficient surveillance system using wireless sensor networks. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pages 270–283. ACM, 2004.

[67] Mohamed Hefeeda and Majid Bagheri. Wireless sensor networks for early detection of forest fires. In *Mobile adhoc and sensor systems, 2007. MASS 2007. IEEE international conference on*, pages 1–6. IEEE, 2007.

[68] Marco Heurich. Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park. *Forest Ecology and Management*, 255(7):2416–2433, 2008.

[69] Matthew M. Holland, Ryan G. Aures, and Wendi B. Heinzelman. Experimental investigation of radio performance in wireless sensor networks. In *Wireless Mesh Networks, 2006. WiMesh 2006. 2nd IEEE Workshop on*, pages 140–150. IEEE, 2006.

[70] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 13(2):415–425, 2002.

[71] C. Huang, L.S. Davis, and J.R.G. Townshend. An assessment of support vector machines for land cover classification. *International Journal of remote sensing*, 23(4):725–749, 2002.

[72] Juha Hyyppa et al. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and remote Sensing*, 39(5):969–975, 2001.

[73] Oana Iova, Amy L. Murphy, Gian Pietro Picco, Lorenzo Ghiro, Davide Molteni, Federico Ossi, and Francesca Cagnacci. LoRa from the City to the Mountains: Exploration of Hardware and Environmental Factors. In *Proc. of the Int. Workshop on New Wireless Communication Paradigms for the Internet of Things Workshop (MadCom 2017), co-located with the 14th Int. Conf. on Embedded Wireless Systems and Networks (EWSN)*, February 2017.

[74] Timofei Istomin et al. Trident: In-field connectivity assessment for wireless sensor networks. *Proc. of ExtremeCom*, 2014.

[75] Recommedation ITU-R P ITU-R. 525-3, calculation of free-space attenuation. *International Telecommunication Union*, 2016.

[76] Recommedation ITU-R P ITU-R. 837-7, characteristics of precipitation for propagation modelling. *International Telecommunication Union*, 2017.

[77] Seo Jae-Hyun, Kim Yong-Hyuk, Ryou Hwang-Bin, Cha Si-Ho, and Jo Minho. Optimal sensor deployment for wireless surveillance sensor networks by a hybrid steady-state genetic algorithm. *IEICE transactions on communications*, 91(11):3534–3543, 2008.

[78] Sujuan Jiang, Carlos Portillo-Quintero, Arturo Sanchez-Azofeifa, and Mike H MacGregor. Predicting RF path loss in forests using satellite measurements of vegetation indices. In *Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on*, pages 592–596. IEEE, 2014.

[79] Damien B. Jourdan and Olivier L. de Weck. Layout optimization for a wireless sensor network using a multi-objective genetic algorithm. In *Vehicular technology conference, 2004. VTC 2004-Spring. 2004 IEEE 59th*, volume 5, pages 2466–2470. IEEE, 2004.

[80] Srinivasan Kannan and Philip Levis. RSSI is under appreciated. In *Proceedings of the Third Workshop on Embedded Networked Sensors, Cambridge, MA, USA*, volume 3031, page 239242, 2006.

[81] Sokratis Kartakis, Babu D Choudhary, Alexander D Gluhak, Lambros Lambrinos, and Julie A McCann. Demystifying low-power wide-area communications for city IoT applications. In *Proceedings of the Tenth ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation, and Characterization*, pages 2–8. ACM, 2016.

[82] Akira Kato et al. Capturing tree crown formation through implicit surface reconstruction using airborne LiDAR data. *Remote Sensing of Environment*, 113(6):1148–1162, 2009.

[83] Andreas Konstantinidis and Kun Yang. Multi-objective k-connected deployment and power assignment in WSNs using a problem-specific constrained evolutionary algorithm based on decomposition. *Computer Communications*, 34(1):83–98, 2011.

[84] Andreas Krause, Carlos Guestrin, Anupam Gupta, and Jon Kleinberg. Near-optimal sensor placements: Maximizing information while minimizing communication cost. In *Proceedings of the 5th International Conference on Information Processing in Sensor Networks*, pages 2–10. ACM, 2006.

[85] Andreas Krause, Carlos Guestrin, Anupam Gupta, and Jon Kleinberg. Robust sensor placements at informative and communication-efficient locations. *ACM Transactions on Sensor Networks (TOSN)*, 7(4):31, 2011.

[86] R.M. Krzanowski and J. Raper. Hybrid genetic algorithm for transmitter location in wireless networks. *Computers, Environment and Urban Systems*, 23(5):359–382, 1999.

[87] Koen Langendoen, Aline Baggio, and Otto Visser. Murphy loves potatoes: Experiences from a pilot sensor network deployment in precision agriculture. In *Parallel and Distributed Processing Symposium, 2006. IPDPS 2006. 20th International*, pages 8–pp. IEEE, 2006.

[88] HyungJune Lee, Alberto Cerpa, and Philip Levis. Improving wireless simulation through noise modeling. In *Information Processing in Sensor Networks, 2007. IPSN 2007. 6th International Symposium on*, pages 21–30. IEEE, 2007.

[89] Jing Li, George Taylor, David Kidner, and Mark Ware. Prediction and visualization of GPS multipath signals in urban areas using LiDAR digital surface models

and building footprints. *International Journal of Geographical Information Science*, 22(11-12):1197–1218, 2008.

[90] Kai Lieska, Erkki Laitinen, and Jaakko Lahteenmaki. Radio coverage optimization with genetic algorithms. In *Personal, Indoor and Mobile Radio Communications, 1998. The Ninth IEEE International Symposium on*, volume 1, pages 318–322. IEEE, 1998.

[91] Thomas Lillesand, Ralph W Kiefer, and Jonathan Chipman. *Remote sensing and image interpretation*. John Wiley & Sons, 2014.

[92] Yunhao Liu, Yuan He, Mo Li, Jiliang Wang, Kebin Liu, and Xiangyang Li. Does wireless sensor network scale? A measurement study on GreenOrbs. *IEEE Transactions on Parallel and Distributed Systems*, 24(10):1983–1993, 2013.

[93] LoRa Alliance™. Wide area networks for IoT. `https://www.lora-alliance.org`, 2018.

[94] Dengsheng Lu, P Mausel, E Brondizio, and Emilio Moran. Change detection techniques. *International journal of remote sensing*, 25(12):2365–2401, 2004.

[95] Gabriel Luque and Enrique Alba. *Parallel genetic algorithms: theory and real world applications*, volume 367. Springer, 2011.

[96] Alan Mainwaring, David Culler, Joseph Polastre, Robert Szewczyk, and John Anderson. Wireless sensor networks for habitat monitoring. In *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, pages 88–97. Acm, 2002.

[97] Matti Maltamo, Erik Næsset, and Jari Vauhkonen. Forestry applications of airborne laser scanning. *Concepts and case studies. Manag For Ecosys*, 27:2014, 2014.

[98] Carsten Maple, Liang Guo, and Jie Zhang. Parallel genetic algorithms for third generation mobile network planning. In *Parallel Computing in Electrical Engineering, 2004. PARELEC 2004. International Conference on*, pages 229–236. IEEE, 2004.

[99] Ramona Marfievici, Amy L. Murphy, Gian Pietro Picco, Federico Ossi, and Francesca Cagnacci. How environmental factors impact outdoor wireless sensor networks: a case study. In *Mobile Ad-Hoc and Sensor Systems (MASS), 2013 IEEE 10th International Conference on*, pages 565–573. IEEE, 2013.

[100] P Martimor, Olivier Arino, Michael Berger, Roberto Biasutti, Bernardo Carnicero, Umberto Del Bello, Valérie Fernandez, Ferran Gascon, Pierluigi Silvestrin, François

Spoto, et al. Sentinel-2 optical high resolution mission for GMES operational services. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pages 2677–2680. IEEE, 2007.

[101] COST 235. Radio propagation effects on next-generation fixed-service terrestrial telecommunication systems. 1996. Luxemburg. Final Report.

[102] Stuart K McFeeters. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*, 17(7):1425–1432, 1996.

[103] Farid Melgani and Lorenzo Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8):1778–1790, 2004.

[104] Herve Meunier, E-G Talbi, and Philippe Reininger. A multiobjective genetic algorithm for radio network optimization. In *Evolutionary Computation, 2000. Proceedings of the 2000 Congress on*, volume 1, pages 317–324. IEEE, 2000.

[105] Luca Mottola, Gian Pietro Picco, Matteo Ceriotti, Ştefan Guna, and Amy L Murphy. Not all wireless sensor networks are created equal: A comparative study on tunnels. *ACM Transactions on Sensor Networks (TOSN)*, 7(2):15, 2010.

[106] Giorgos Mountrakis, Jungho Im, and Caesar Ogole. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247–259, 2011.

[107] Bryon Moyer. Low power, wide area: A survey of longer-range iot wireless protocols. *Electronic Engineering Journal*, pages 09–07, 2015.

[108] Erik Næsset. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80(1):88–99, 2002.

[109] AN1200.22 LoRa™ Modulation Basics Application Note. *Semtech Corporation*, Revision 2, May 2015.

[110] Antonio Novelli, Manuel A Aguilar, Abderrahim Nemmaoui, Fernando J Aguilar, and Eufemia Tarantino. Performance evaluation of object based greenhouse detection from Sentinel-2 MSI and Landsat 8 OLI data: A case study from Almería (Spain). *International Journal of Applied Earth Observation and Geoinformation*, 52:403–411, 2016.

[111] Kimmo Nurminen, Mika Karjalainen, Xiaowei Yu, Juha Hyyppä, and Eija Honkavaara. Performance of dense digital surface models based on image matching in the estimation of plot-level forest variables. *ISPRS Journal of Photogrammetry and Remote Sensing*, 83:104–115, 2013.

[112] Yoshihisa Okumura. Field strength and its variability in VHF and UHF land-mobile service. *Rev. Elec. Comm. Lab.*, 16(9):825–873, 1968.

[113] Carlos A Oroza, Ziran Zhang, Thomas Watteyne, and Steven D Glaser. A machine-learning-based connectivity model for complex terrain large-scale low-power wireless deployments. *IEEE Transactions on Cognitive Communications and Networking*, 3(4):576–584, 2017.

[114] Carlos A Oroza, Zeshi Zheng, Steven D Glaser, Devis Tuia, and Roger C Bales. Optimizing embedded sensor network design for catchment-scale snow-depth estimation using LiDAR and machine learning. *Water Resources Research*, 52(10):8174–8189, 2016.

[115] Claudia Paris and Lorenzo Bruzzone. A three-dimensional model-based approach to the estimation of the tree top height by fusing low-density LiDAR data and very high resolution optical images. *IEEE Transactions on Geoscience and remote Sensing*, 53(1):467–480, 2015.

[116] Claudia Paris, Davide Valduga, and Lorenzo Bruzzone. A hierarchical approach to three-dimensional segmentation of LiDAR data at single-tree level in a multilayered forest. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7):4190–4203, 2016.

[117] Neal Patwari and Piyush Agrawal. Effects of correlated shadowing: Connectivity, localization, and RF tomography. In *Proceedings of the 7th international conference on Information processing in sensor networks*, pages 82–93. IEEE Computer Society, 2008.

[118] Asa Persson, Johan Holmgren, and Ulf Soderman. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, 68(9):925–932, 2002.

[119] Martino Pesaresi, Christina Corbane, Andreea Julea, Aneta J Florczyk, Vasileios Syrris, and Pierre Soille. Assessment of the added-value of Sentinel-2 for detecting built-up areas. *Remote Sensing*, 8(4):299, 2016.

[120] 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 *ITS Telecommunications (ITST), 2015 14th International Conference on*, pages 55–59. IEEE, 2015.

[121] Gian Pietro Picco, Davide Molteni, Amy L Murphy, Federico Ossi, Francesca Cagnacci, Michele Corrà, and Sandro Nicoloso. Geo-referenced proximity detection of wildlife with WildScope: design and characterization. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, pages 238–249. ACM, 2015.

[122] Pyari Mohan Pradhan and Ganapati Panda. Connectivity constrained wireless sensor deployment using multiobjective evolutionary algorithms and fuzzy decision making. *Ad Hoc Networks*, 10(6):1134–1145, 2012.

[123] Suzanne Prange, Trevor Jordan, Colin Hunter, and Stanley D Gehrt. New radio-collars for the detection of proximity among individuals. *Wildlife Society Bulletin*, 34(5):1333–1344, 2006.

[124] Dave Ramsey, Nick Spencer, Peter Caley, Murray Efford, Keith Hansen, Mary Lam, and Des Cooper. The effects of reducing population density on contact rates between brushtail possums: implications for transmission of bovine tuberculosis. *Journal of Applied Ecology*, 39(5):806–818, 2002.

[125] CJ Rankine, G Arturo Sanchez-Azofeifa, and Mike H MacGregor. Seasonal wireless sensor network link performance in boreal forest phenology monitoring. In *Sensing, Communication, and Networking (SECON), 2014 Eleventh Annual IEEE International Conference on*, pages 302–310. IEEE, 2014.

[126] Theodore S. Rappaport. *Wireless communications: Principles and Practice*, volume 2. Prentice Hall PTR New Jersey, 1996.

[127] Colin Reeves. Genetic algorithms. In *Handbook of metaheuristics*, pages 55–82. Springer, 2003.

[128] Niels Reijers, Gertjan Halkes, and Koen Langendoen. Link layer measurements in sensor networks. In *Mobile Ad-hoc and Sensor Systems, 2004 IEEE International Conference on*, pages 224–234. IEEE, 2004.

[129] Charles Reis, Ratul Mahajan, Maya Rodrig, David Wetherall, and John Zahorjan. Measurement-based models of delivery and interference in static wireless networks.

In *ACM SIGCOMM Computer Communication Review*, volume 36, pages 51–62. ACM, 2006.

[130] Neil C Rogers et al. A generic model of 1-60 GHz radio propagation through vegetation-final report. *Radio Agency, UK*, 2002.

[131] Fabio Roli and Giorgio Fumera. Support vector machines for remote sensing image classification. In *Image and Signal Processing for Remote Sensing VI*, volume 4170, pages 160–167. International Society for Optics and Photonics, 2001.

[132] A Roullier-Callaghan, M Al-Nuaimi, and D Kidner. A topological feature extraction system from LiDAR data with the application of radiowave propagation modelling. In *Antennas and Propagation, 2003.(ICAP 2003). Twelfth International Conference on (Conf. Publ. No. 491)*, volume 2, pages 522–525. IET, 2003.

[133] J_W Rouse Jr, 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, Goddard Space Flight Center, NASA SP-351, Science and Technical Information Office, NASA, Washington, DC*, pages 309–317, 1974.

[134] Robert A Schowengerdt. *Remote sensing: models and methods for image processing.* Elsevier, 2006.

[135] Leo Selavo, Anthony Wood, Qing Cao, Tamim Sookoor, Hengchang Liu, Aravind Srinivasan, Yafeng Wu, Woochul Kang, John Stankovic, Don Young, et al. Luster: wireless sensor network for environmental research. In *Proceedings of the 5th international conference on Embedded networked sensor systems*, pages 103–116. ACM, 2007.

[136] Shashi Shekhar, Paul R Schrater, Ranga Raju Vatsavai, Weili Wu, and Sanjay Chawla. Spatial contextual classification and prediction models for mining geospatial data. *IEEE Transactions on Multimedia*, 4(2):174–188, 2002.

[137] Amit Sheth, Cory Henson, and Satya S Sahoo. Semantic sensor web. *IEEE Internet computing*, 12(4), 2008.

[138] Byungrak Son, Yong-sork Her, and Jung-Gyu Kim. A design and implementation of forest-fires surveillance system based on wireless sensor networks for south korea mountains. *International Journal of Computer Science and Network Security (IJCSNS)*, 6(9):124–130, 2006.

[139] Rebecca A Spriggs, David A Coomes, Trevor A Jones, John P Caspersen, and Mark C Vanderwel. An alternative approach to using LiDAR remote sensing data to predict stem diameter distributions across a temperate forest landscape. *Remote Sensing*, 9(9):944, 2017.

[140] Kannan Srinivasan, Prabal Dutta, Arsalan Tavakoli, and Philip Levis. An empirical study of low-power wireless. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):16, 2010.

[141] Kannan Srinivasan, Maria A Kazandjieva, Mayank Jain, Edward Kim, and Philip Levis. SWAT: enabling wireless network measurements. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 395–396. ACM, 2008.

[142] Robert Szewczyk, Joseph Polastre, Alan Mainwaring, and David Culler. Lessons from a sensor network expedition. In *European Workshop on Wireless Sensor Networks*, pages 307–322. Springer, 2004.

[143] Theodor Tamir. On radio-wave propagation in forest environments. *IEEE Transactions on Antennas and Propagation*, 15(6):806–817, 1967.

[144] John Thelen, Dann Goense, and Koen Langendoen. Radio wave propagation in potato fields. In *1st workshop on wireless network measurement, Riva del Garda, Italy, April 2005*, pages np–np, 2005.

[145] Gilman Tolle, Joseph Polastre, Robert Szewczyk, David Culler, Neil Turner, Kevin Tu, Stephen Burgess, Todd Dawson, Phil Buonadonna, David Gay, et al. A macroscope in the redwoods. In *Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems*, pages 51–63. ACM, 2005.

[146] David M Tralli, Ronald G Blom, Victor Zlotnicki, Andrea Donnellan, and Diane L Evans. Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(4):185–198, 2005.

[147] V Vapnik. Statistical learning theory. *NY: Wiley*, 1998.

[148] Pascal Vicaire, Tian He, Qing Cao, Ting Yan, Gang Zhou, Lin Gu, Liqian Luo, Radu Stoleru, John A Stankovic, and Tarek F Abdelzaher. Achieving long-term surveillance in vigilnet. *ACM Transactions on Sensor Networks (TOSN)*, 5(1):9, 2009.

[149] Thiemo Voigt, Martin Bor, Utz Roedig, and Juan Alonso. Mitigating inter-network interference in LoRa networks. 2017.

[150] Michele Volpi, Devis Tuia, Francesca Bovolo, Mikhail Kanevski, and Lorenzo Bruz-zone. Supervised change detection in VHR images using contextual information and support vector machines. *International Journal of Applied Earth Observation and Geoinformation*, 20:77–85, 2013.

[151] Xiaorui Wang, Guoliang Xing, Yuanfang Zhang, Chenyang Lu, Robert Pless, and Christopher Gill. Integrated coverage and connectivity configuration in wireless sensor networks. In *Proceedings of the 1st international conference on Embedded networked sensor systems (SenSys)*, pages 28–39. ACM, 2003.

[152] Mark A Weissberger. An initial critical summary of models for predicting the at-tenuation of radio waves by trees. Technical report, DTIC Document, 1982.

[153] Matt Welsh. Sensor networks for the sciences. *Communications of the ACM*, 53(11):36–39, 2010.

[154] Hjalmar Wennerström, Frederik Hermans, Olof Rensfelt, Christian Rohner, and Lars-Åke Nordén. A long-term study of correlations between meteorological condi-tions and 802.15. 4 link performance. In *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2013 10th Annual IEEE Communications Society Confer-ence on*, pages 221–229. IEEE, 2013.

[155] Geoffrey Werner-Allen, Konrad Lorincz, Mario Ruiz, Omar Marcillo, Jeff Johnson, Jonathan Lees, and Matt Welsh. Deploying a wireless sensor network on an active volcano. *IEEE internet computing*, 10(2):18–25, 2006.

[156] Andrew J Wixted, Peter Kinnaird, Hadi Larijani, Alan Tait, Ali Ahmadinia, and Niall Strachan. Evaluation of LoRa and LoRaWAN for wireless sensor networks. In *SENSORS, 2016 IEEE*, pages 1–3. IEEE, 2016.

[157] Alec Woo, Terence Tong, and David Culler. Taming the underlying challenges of reliable multihop routing in sensor networks. In *Proceedings of the 1st international conference on Embedded networked sensor systems*, pages 14–27. ACM, 2003.

[158] Mohamed Younis and Kemal Akkaya. Strategies and techniques for node placement in wireless sensor networks: A survey. *Ad Hoc Networks*, 6(4), 2008.

[159] Marco Zúñiga Zamalloa and Bhaskar Krishnamachari. An analysis of unreliability and asymmetry in low-power wireless links. *ACM Transactions on Sensor Networks (TOSN)*, 3(2):7, 2007.

[160] Jerry Zhao and Ramesh Govindan. Understanding packet delivery performance in dense wireless sensor networks. In *Proceedings of the 1st international conference on Embedded networked sensor systems*, pages 1–13. ACM, 2003.

[161] Gang Zhou, Tian He, Sudha Krishnamurthy, and John A Stankovic. Impact of radio irregularity on wireless sensor networks. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pages 125–138. ACM, 2004.

[162] Gang Zhou, Tian He, Sudha Krishnamurthy, and John A Stankovic. Models and solutions for radio irregularity in wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 2(2):221–262, 2006.