

Delft University of Technology
Master's Thesis in Embedded Systems

Analysis of Remote Sensing approaches for LoRa coverage estimation

Priyanka Bhat



Analysis of Remote Sensing approaches for LoRa coverage estimation

Master's Thesis in Embedded Systems

Embedded Software Section
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands

Priyanka Bhat
p.bhat@student.tudelft.nl

26th August 2019

Author

Priyanka Bhat (p.bhat@student.tudelft.nl)

Title

Analysis of the Automated approach to coverage estimation for LoRa

MSc presentation

26th August 2019

Graduation Committee

Prof. dr. K.G. Langendoen Delft University of Technology

dr. Marco Zuniga Delft University of Technology

dr.ir. G.J.M. Janssen Delft University of Technology

Abstract

LoRa is being widely adopted by industrial communities for its long range, robustness and low power wireless communication capabilities. In fact LoRa is gaining more popularity even amongst the common people as it is an affordable solution and operates in the unlicensed radio spectrum. However, LoRa provides a widely heterogeneous coverage; it can reach hundreds of meters or up to tens of kilometers, depending on the surrounding environment. Determining the coverage of LoRa stations is key to provide a good quality of service. On one hand, the traditional method of expensive measurement campaigns can be employed to estimate LoRa's coverage; but this is impractical due to the large geographical areas involved. On the other hand, popular channel models can be adopted; but many of them are yet to be explored for LoRa or rely entirely on the user predefining the type of environment to estimate coverage. Neither of those approach are suitable for thousands of non-expert citizens and organizations around the world looking forward to understanding the coverage of their LoRa stations.

The aim of the thesis is to automatically estimate the coverage of LoRa, before the deployment of the gateway and without relying on on-site measurements or the user's perception of the environment. Moreover, the estimation must be carried out in a simple, low cost and low effort approach. Considering that the surrounding environment determines in a fundamental manner, the coverage of wireless technologies including LoRa, we use readily available remote sensing information coming from satellites to estimate the characteristics of an area. In this manner, we free up the user from providing any type of data. Based on this remote sensing approach, the thesis provides two main contributions: First we analyze a group of parametric models (ITU-R 1812 and Okumura Hata model) and determine that the Okumura Hata model is better suited for LoRa. Second we improve the performance of using the basic Okumura Hata model by proposing an automated approach that explores remotely sensed height models and land cover maps to automatically configure channel model parameters.

The performance is evaluated based on a relative comparison due to some unknown transmitter setting parameters and assess which algorithm accurately tracks the changes in the real path loss. A validation using a relative comparison approach on 18000+ samples of real LoRa data shows that the modified algorithm gives an improved performance compared to the novel approach in path loss prediction and the ITU model. The modified algorithm could improve the coverage up to a factor of 5 compared to the novel approach in free space ranges. Moreover, in an urban built-up city the modified algorithm could improve the coverage by up to 1.5 km compared to the novel approach.

Acknowledgment

I owe a deep sense of gratitude to my daily supervisor, Prof. Marco Zuniga for his constructive insights, motivation and enthusiasm throughout the course of this research work. I would like to thank Dr. Silvia Demetri for her valuable insights regarding my thesis work. I am grateful to all my friends who have helped me find a balance between my social and academic lives. A huge thank you to Thomas John, Chaitra Pai and Eunice Francis for being the wonderful friends that they are. Finally, I would like to thank my family for their unconditional love and the financial support that helped me through my journey at TU Delft.

Priyanka Bhat

Delft, The Netherlands
26th August 2019

Contents

Acknowledgment	v
1 Introduction	1
1.1 Motivation	1
1.2 Objective	4
1.3 Main contribution	4
1.4 Organization	5
2 Background and Related work: Channel models	7
2.1 LoRa: An Overview	7
2.2 Channel model classifications	10
2.3 Prior Work	11
2.3.1 Dependency on Manual measurements	11
2.3.2 Dependency on pre-defined user input	11
2.3.3 Unexplored Semi-deterministic models for LoRa	13
3 Background and Related work: Remote Sensing	15
3.1 Remote Sensing: Overview	15
3.1.1 Land cover classification	16
3.1.2 Digital Elevation Models	16
3.2 Related work using remote sensing data	19
3.2.1 Land-cover maps	20
3.2.2 Path analysis	20
3.2.3 Intersection analysis	21
4 System and Experimental Setup	25
4.1 Validation with TTN Mapper data	25
4.2 Preparing the remotely sensed data	27
5 Evaluation of ITU-R 1812 for LoRa	33
5.1 Overview of ITU-R 1812	33
5.2 Results and Conclusion: ITU-R 1812 vs. Path Analysis	36

6 Improved Pathloss Model Development	43
6.1 Approach One	43
6.1.1 Approach One: Results and Conclusion	45
6.2 Approach Two	48
6.2.1 Approach Two: Results and Conclusion	51
6.3 Approach Three	60
6.4 Relative comparison of the approaches	67
7 Conclusions and Future Work	71
7.1 Conclusions	71
7.2 Future Work	72

Chapter 1

Introduction

The Internet of Things (IoT) has created an opportunity for businesses to become more efficient and productive and with global warming on the rise [18], IoT could play a prominent role in reducing greenhouse emissions and combating climate change [19]. According to IoT analytics, the leading provider of market insights for the Internet of Things, the number of active IoT connections will rise from 7 billion in 2018 to 22 billion in 2025 globally. The key enabler for this growth is attributed to the Low Power Wide Area Networks (LPWAN) technology [26]. LPWAN represents a set of low-power wireless technologies that are designed to send small data packets a few times an hour over long distances. LPWAN allows the connectivity of billions of devices placed far apart using low-cost sensors that promote long-lasting battery life with minimal human intervention. There are several competing LPWAN technologies. Some of the leading technologies being SigFox, LoRa and NB-IoT [48].

1.1 Motivation

LoRa has been gaining popularity world-wide [16] and is a widely adopted technology for large-scale IoT deployments [51] [53]. Semtech has gathered over 600 known use cases of LoRa in different end-markets [22]. LoRa based smart agriculture has shown to provide 50% reduction in water consumption for commercial farms [30]. LoRa enabled smart water sensors has proven to reduce 20% wastewater and approximately 22% water bill savings [29]. Because LoRa is affordable and operates in the unlicensed radio spectrum, it is gaining more popularity even amongst common people. Now anyone can buy and set up a LoRa gateway, run their own applications and use the frequencies without having to shed millions of dollar for transmission rights.

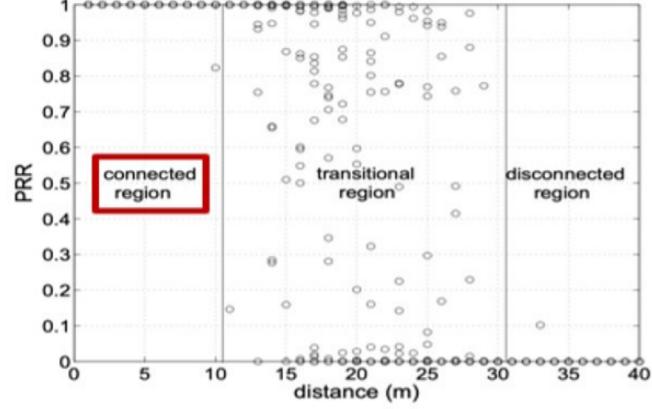
Several use cases rely on having good LoRa coverage over huge areas. For example, the zTrack LoRa GPS tracker developed by ZANE systems Kft offers fleet tracking solutions for smarter management, tracking mainten-

ance issues and real-time locations [33]. SODAQ’s solar-powered smart ear tags help keep track of cattle over huge farms for early detection of sick animals and to reduce cattle thefts [17]. LoRa claims to connect devices that are 30 miles apart in rural areas and it can penetrate through dense urban areas and deep indoor environments [32]. But considering the dynamic and unpredictable nature of the environment, estimating coverage is never easy. Several factors affect the quality of a wireless communication link. Attenuation due to the atmosphere, physical objects in the surrounding, terrain, hardware settings, antenna gain pattern, external interference, indoor/outdoor services, etc. play an important role in determining the performance of the wireless link.

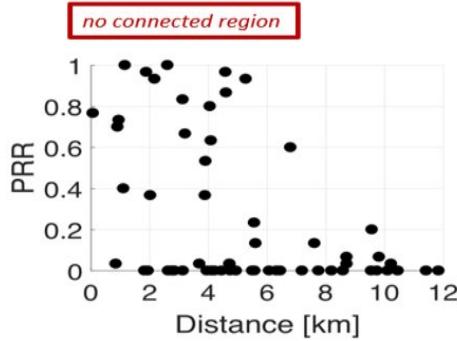
There is plenty of preliminary work done showing the high variability of LoRa’s range. The authors in [52] covered paths of more than 40 km in the Darkar Peninsula to evaluate the coverage of LoRa. The paper concluded that their gateway could reach a maximum of 10 km only. A different study [40] observed that LoRa worked effectively in urban areas of Raleigh, North Carolina up to approximately 3 km, whereas the authors in [38] claimed that LoRa reached 2.6 km in rural areas and 100 m in a built-up environment. Another study shows that LoRa communication varied between 15-30 km with a node placed on top of the car in urban areas and on a boat in open waters [50]. In Hyde Park in London, LoRa measurements could reach 450 m in built-up areas [45] while a maximum distance of 90 m was achieved in a mountain forest with dense vegetation [43]. The various studies show the wide variations of LoRa ranges from about 100 m reaching up to 30 km in different environments.

To understand the range behavior of LoRa, the authors in [41] makes a good comparison of LoRa with WiFi. LoRa can achieve a high receiver sensitivity of -140 dBm. With the maximum transmit power in Europe being +14 dBm, this would translate to a large 154 dB link budget. WiFi can achieve a receiver sensitivity of -90 dBm. This difference in the sensitivity sheds light on LoRa’s capability of decoding signals 10^5 times weaker than WiFi, significantly increasing the range. Considering the maximum transmit power allowed for WiFi in Europe (+20 dBm), the resulting link budget is 110 dB. On applying the theoretical free-space path loss model to LoRa (868Mhz) and WiFi (2.4Ghz), their link budgets map to upper-bound distances of 3 km and 1400 km respectively. But in reality, the range is limited to approximately 90 m for WiFi and ≥ 30 km for LoRa in open real environments. The extra 44 dB link budget can stretch the maximum distances of LoRa by a large degree in open scenarios. But the presence of just four brick walls along the propagation path could render this extra link budget completely pointless. Each wall could attenuate a radio signal by 10 dB, causing the LoRa signal disappear within just a 100 m, making it no different from WiFi. The fact remains that similar to any other wireless communication technology, LoRa too greatly depends on its surrounding

environment.



(a) Short range low power technology [41]



(b) Long range low power technology[41]

Figure 1.1: Link reliability (PRR) vs. distance

Figure 1.1 illustrates the link reliability as the variation of packet reception rate (PRR) with distance. As shown in Figure 1.1a short-range, low power communication technologies like Bluetooth, WiFi, cellular, etc., have three clear communication regions based on the link reliability [41]. At close distances from the transmitter, the link reliability is 1 (connected region). Considering that the connected region is just a few meters away, the probability of having a similar type of environment in every direction is high. Moving away from this region, the probability of receiving a packet begins

to vary (transitional region). Further away, the link reliability tends to 0 and reaches complete disconnection from the network. On the contrary, compared with a long-range technology like LoRa, a long transitional region is noticeable as shown in Figure 1.1b. Within only 3-4 km from the transmitter, the link reliability begins to vary. On a kilometer-scale, there exists no noticeable connected region, as the probability of a LoRa signal encompassing segments of different environment types is high.

Smart IoT applications that could make businesses more efficient and improve everyday lives would be left futile if the coverage provided by the LoRa network is unknown or insufficient. Drive testing is the most common tool used by telecommunication companies to measure and analyze the coverage, capacity, and quality of service of short-range technologies. Such expensive on-site measurement campaigns would only increase the complexity, time and requirement of manpower for the common user. There are companies that provide simulation software which use popular channel models to predict path loss for radio frequencies [13],[14],[20],[27]. These models are used for network planning and deployment and to ensure reliable service areas with minimal infrastructure. But these software tools come with a price tag and may not be affordable to common people.

1.2 Objective

The focus of this research is to completely automate the process of coverage prediction; even before investing in the deployment of a LoRa gateway and without depending on expensive on-site measurement campaigns. The goal is to be able to model the quality of a LoRa link in a simple, low cost and low effort approach. To be able to do this, local and specific features that describe the propagation environment can be obtained by leveraging existing free online sources of information. The thesis also focuses on analyzing existing mathematical channel models that can effectively utilize this information and make better predictions of the path-loss for LoRa. The key idea is to provide a tool that focuses on serving the non-technical community with a hassle-free coverage estimation.

1.3 Main contribution

The key contributions of this thesis are:

- Exploration of the ITU-R 1812 model which comes from a class of channel models (semi-deterministic models) that are expensive in terms of the complexity, computational effort and the need of key deterministic information about the propagation scenario. These models have been largely left unexplored for LoRa and this thesis will look into the

effectiveness of one such model.

- Taking inspiration from a novel automated approach [41] that uses remote sensing combined with a simpler channel model based on empirical measurements (Okumura Hata model), this thesis develops a modified algorithm that gives an improvement in predicting path loss for LoRa.

1.4 Organization

This thesis will begin with the background and related work which will be split into two chapters. Chapter 2 will delve into a brief overview of LoRa as well as channel model classifications. This chapter will also delve into some prior work for LoRa’s coverage estimation. Chapter 3 will focus on giving background information on the basics of remote sensing and the related work using remotely sensed data. Chapter 4 will describe the system and experimental setup used in the thesis. Chapter 5 will look into evaluating the ITU-R 1812 model belonging to one class of channel models that is more rich and complex in terms of the input and computational effort required for path loss estimation. Chapter 6 will then delve into the assessment of the state-of-the-art approach using a simpler channel model and provides a modified approach to improve the path loss estimation algorithm. Chapter 7 will provide conclusions and possible future work related to the thesis.

Chapter 2

Background and Related work: Channel models

This chapter aims to give background information about some topics covered in the thesis and discuss prior research work carried out to understand the coverage of LoRa. Section 2.1 will give an overview of LoRa wireless communication technology, Section 2.2 will highlight the channel model classifications. Sections 2.3 will discuss some prior work on estimating LoRa's coverage.

2.1 LoRa: An Overview

LoRa (an acronym for Long Range) is a radio modulation scheme acquired by Semtech in 2012 [21]. It is a physical layer implementation derived from the chirp spread spectrum (CSS) modulation. This modulation technique has been used for many years to provide long-range and robust communication in military and space applications, but LoRa is its first low-cost commercial implementation [21]. Unlike FSK (Frequency Shift Keying) modulation, where the carrier signal jumps between two discrete frequencies, LoRa goes further by sweeping between the two frequencies (chirps) to encode digital information. These up-chirps/ down-chirps are linearly increasing/decreasing frequencies over the entire bandwidth at a fixed rate as shown in Figure 2.1. This added complexity allows LoRa to be more robust in harsh RF environments and more resistant to interference, multipath, and Doppler effects [24].

The robust modulation scheme grants LoRa a large link budget and hence its long-range capability. A link budget accounts for all the losses and gains of the transmitted signal due to the channel medium, antenna pattern, feed line, etc. as shown in Equation 2.2. It defines the quality of the radio communication channel. The link budget is a metric commonly used for coverage evaluation. It is characterized by the receiver sensitivity and transmitter

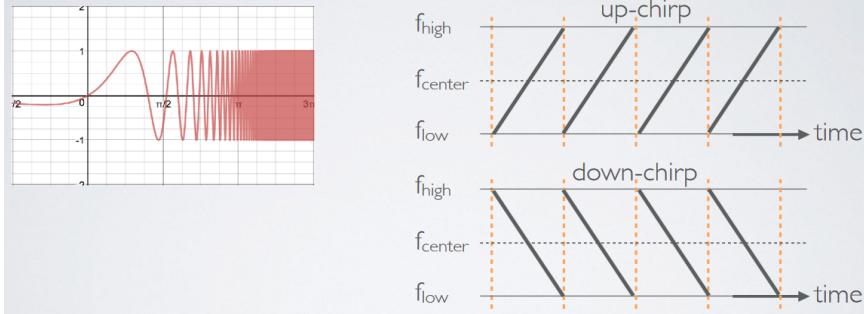


Figure 2.1: Illustration of chirps in CSS modulation [25]

power.

$$\text{LinkBudget} = \text{TransmitPower} - \text{ReceiverSensitivity} \quad (2.1)$$

If the path loss derived in Equation 2.2 is within the available link budget for the communication to occur, the signal can be decoded, else the packet is lost.

$$P_{rx} - P_{tx} = G_{tx} + G_{rx} - L_{tx} - L_p - L_{rx} \quad (2.2)$$

Where:

P_{tx}, P_{rx} = Transmit and received signal power respectively (dBm)

G_{tx}, G_{rx} = Transmitter and receiver antenna gain respectively (dBi)

L_{tx}, L_{rx} = Feedline and associated losses (dB)

L_p = Propagation loss in the channel medium (dB)

1. Receiver sensitivity

Receiver sensitivity expresses the minimum power level needed for the receiver to decode the signal. Two major parameters that influence the receiver sensitivity in LoRa are the Spreading factor (SF) and Bandwidth (BW) as shown in Equation 2.3 [23]. The combination of LoRa settings (SF, BW) can effect the range, data rate and time-on-air differently.

$$RS_{min} = -174 + 10 * log_{10}(BW) + NF + SNR_{min} \quad (2.3)$$

Where:

RS : Receiver sensitivity (dBm)

BW : Receiver Bandwidth in Hertz (Hz)

NF : Noise Figure (dB)

SNR : Signal to Noise Ratio required by the modulation scheme

NF is fixed for a particular hardware implementation, e.g. Lora transceiver chips SX1272 and SX1276, typically use a value of 6 dB [23].

- **Spreading factor :**

The Spreading factor (SF) describes the rate of change of frequencies in a chirp (or sweep rate). The SF is chosen by the user based on the trade-off between the range and data-rate requirements. In Europe, six different sweep rates or spreading factors are utilized: SF 7 - SF12 as shown in Figure 2.2. The faster the sweep rate (SF 7), the more data can be transmitted (time on air decreases), but decoding the signal becomes harder and the range decreases.

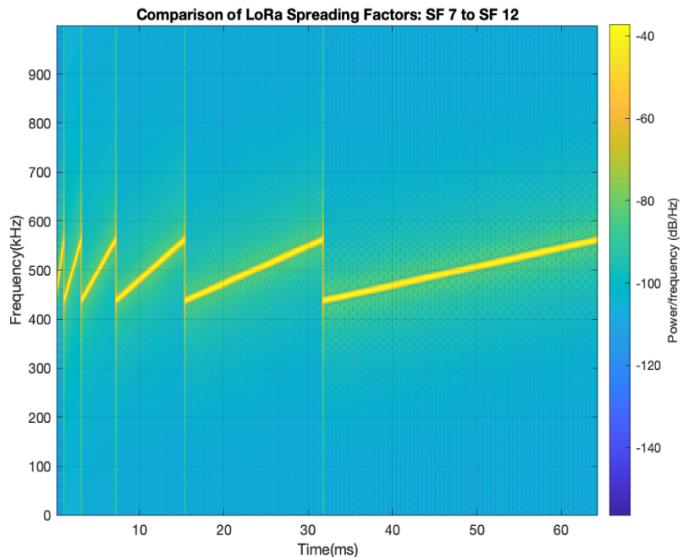


Figure 2.2: Illustration of sweep rates/ SF for a fixed bandwidth [46]

Each SF corresponds to a specific SNR limit, below which the receiver will be unable to decode the signal effectively. SNR is the ratio of the received signal power over the noise power level. For example, for a bandwidth of 125 kHz, an SF of 12 will correspond to SNR limit of -20dB, while an SF of 7 will correspond to SNR limit of -7.5dB. The higher the SF, the lower the SNR required for demodulation.

- **Bandwidth**

Bandwidth refers to the set of frequencies available for a chirp pulse. LoRa is allowed to operate on three different bandwidths: 125 kHz, 250 kHz and 500 kHz [25]. A higher bandwidth increases the data rate, hence reducing the time on-air. On the

other hand, a higher bandwidth lowers decoder sensitivity due to the integration of additional channel noise.

2. Transmission Power

Transmission power is usually expressed in dBm. For LoRa transceivers, the transmitter power can vary between -4 dBm to +20 dBm [37]. A higher transmission power results in a longer range but decreases battery life.

LoRaWAN defines the Media access control (MAC) layer built on top of LoRa which defines the communication protocol and the network architecture. The network architecture is deployed in a simple star of star topology. An end node can communicate with one or more gateways that in turn communicate with a centralized network server in a bi-directional manner.

Moreover, LoRa exploits the unlicensed sub-gigahertz ISM (Industrial, Scientific, and Medical) band. Due to regulatory requirements, the specific ways in which that band is used depends on the region of operation. In Europe, LoRa works on the 868MHz frequency band with a maximum transmit power of +14 dBm.

2.2 Channel model classifications

To account for the effects on radio wave propagation considering the particular properties of the surrounding environment, radio propagation models are widely used. Modeling and simulation of radio wave propagation have been playing a critical role in coverage analysis and network planning for wireless communication systems. Channel models are classified mainly into three forms: Empirical, Semi-deterministic and Deterministic models. These classes vary based on their complexity, accuracy and computational time.

Empirical models are mathematical formulations derived from extensive field measurements in a specific environment and using statistical analysis. Some examples include Okumura Hata model, COST-231 Hata model, Ericsson 9999 model etc. These models are simplistic and easy to implement as they rely on a few simple input parameters. Since they are based on real measurements, they account for every propagation phenomenon that can affect the signal strength. This class of channel models will be dealt with in the thesis.

On the other hand, deterministic channel models are more expensive in terms of their computational demands, time and complexity but are supposed to be highly accurate. They use laws governing electromagnetic wave theory and predict the signal power at a particular location. They are highly site-specific and rely on 3D maps and building layer information for an accurate geometrical description of the environment. For instance, these

include models based on ray tracing. The main drawback of these models is that such detailed information is difficult to obtain in practice. Therefore, deterministic models will not be dealt with in the thesis.

On the contrary, semi-deterministic channel models lie in between the properties of empirical and deterministic models. They are a combination of empirical models with few key deterministic aspects describing the environment. This class of channel models will be looked into in the thesis.

2.3 Prior Work

2.3.1 Dependency on Manual measurements

A large part of the research conducted to investigate LoRa's performance in a region has involved manually deploying LoRa nodes over large areas or using drive-test measurements to measure signal strength [52],[40],[50],[43] as discussed in Section 1.1. Moreover, some researchers [55],[49],[44],[39], [50] utilize the RF measurements for tuning channel models to better represent path-loss. One such commonly used model is the log-normal shadowing model as shown in Equation 2.4 and is dependent on empirical data. This model assumes that the path-loss varies exponentially with distance and is commonly used to model loss in built-up urban areas [38]. It relies on curve fitting or regression over empirical data to obtain the radio propagation parameters of the model.

$$PL(d) = 10 * n * \log_{10}(d/d_0) + PL_0 + X_\sigma \quad (2.4)$$

Where:

n : Path-loss exponent

d : Link distance

$PL(d)$: Path-loss at distance d (dB)

PL_0 : Path-loss at reference distance d_0 (dB)

X_σ : Loss due to shadow fading with zero-mean Gaussian distribution and standard deviation σ .

The first limitation of this technique is the dependency on on-site measurements to obtain channel model parameters. Secondly, the model will provide accurate results solely for the region for which it was trained. The process would have to be repeated with new training data to get appropriate model coefficients for other regions. This is clearly against the objective of the thesis.

2.3.2 Dependency on pre-defined user input

There are certain empirical channel models like Okumura Hata model, COST-231 Hata model, Ericsson 9999 model, etc., that are mainly used for pre-

dicting path-loss for cellular transmissions. The key characteristic of these models is that they rely on an input parameter that defines the type of environment along the path. These models provide a set of equations that can be used based on the type of environment. For example, a user in the crowded street like Tokyo, will only need to select the "urban" type of environment and provide some simple input parameters for the models. The Okumura Hata (OH) model has also been used to evaluate LoRa's coverage, but this work is very limited [36], [47].

The OH empirical model is one of the most popular radio propagation models used for path-loss predictions in the frequency range of 150 MHz - 1500 MHz. The model is suitable for link distances from 1-10 km, mobile station antenna heights ranging from 1-10 m and base station antenna height ranging from 30-200 m. The OH model is known for its accurate and simplistic nature. This model is an extension of its predecessor: the Okumura model, which is based on extensive large-scale measurements collected in the built-up areas of Tokyo, Japan. The Okumura model provided a set of graphs for obtaining median attenuation (dB) relative to free-space and some correction factors to account for antenna heights and terrain type.

The Okumura model mainly used for signal prediction in urban areas served as the base for the OH model. To be able to conveniently apply the Okumura model for computer implementations, these set of curves were fit into mathematical formulations. The OH model was further developed to add the effects of diffraction, reflection, and scattering from built-up structures in urban areas. Additionally, correction factors are applied to the model for use in suburban and rural areas. The OH model depends on a few simple parameters such as frequency, link distance, transmitter and receiver antenna heights and the type of environment. The equations below highlight the four OH variants based on the type of environment.

Under the **urban** environment, the following two OH variants can exist,

$$L_U = 69.55 + 26.16 \log_{10} f - 13.82 \log_{10} h_B - C_H + [44.9 - 6.55 \log_{10} h_B] \log_{10} d \quad (2.5)$$

For small or medium-sized cities,

$$C_H = 0.8 + (1.1 \log_{10} f - 0.7) h_M - 1.56 \log_{10} f \quad (2.6)$$

and for large cities,

$$C_H = \begin{cases} 8.29 (\log_{10}(1.54h_M))^2 - 1.1 & , \text{ if } 150 \leq f \leq 200 \\ 3.2 (\log_{10}(11.75h_M))^2 - 4.97 & , \text{ if } 200 < f \leq 1500 \end{cases} \quad (2.7)$$

Under the **sub-urban** environment,

$$L_{SU} = L_U - 2 \left(\log_{10} \frac{f}{28} \right)^2 - 5.4 \quad (2.8)$$

Under the **free/open/rural** environment,

$$L_O = L_U - 4.78(\log_{10} f)^2 + 18.33(\log_{10} f) - 40.94 \quad (2.9)$$

where:

L_U = Path loss in urban areas (dB)

L_{SU} = Path loss in suburban areas (dB)

L_O = Path loss in open areas (dB)

h_B = Height of base station antenna (meter)

h_M = Height of mobile station antenna (meter)

f = Frequency of transmission (MHz)

C_H = Antenna height correction factor

d = Distance between the base and mobile stations (kilometer)

Using a model whose output is based on a user's predefined input can be disadvantageous for long-range technologies. When applied for a point-to-point path loss analysis, these models rely on the input for defining the type of environment along the path. This dependency on the user's perspective could simply lead to discrepancies in classifying the environment. In simple terms, it would require the user to physically make the effort to observe the propagation path and get a feel of how densely built-up the target area is. This is still acceptable for short-range technologies, considering the across a small area the type of surrounding environment can be considered to be the same as the type of environment can dramatically change over distances that are kilometer long. But this is impractical for long-range technologies like LoRa.

Moreover, using these models for point-to-area coverage analysis and defining one type of environment for a large region would wrongly estimate path-losses. The selection of one environment type would imply that paths at every point from the gateway will have a similar degree of urbanization over the entire area. Again, this could be valid for short ranges, but when covering huge areas, the probability of spanning different types of environment is high.

This goes against our objective which is to have a coverage prediction that is automated and can analyze the path on its own. Recently a research paper [41] published by TU Delft in collaboration with the University of Trento, provided a methodology to eliminate the need for predefined user input to classify the environment. The details of the paper will be discussed in Chapter 3 and the work of this thesis builds upon that study.

2.3.3 Unexplored Semi-deterministic models for LoRa

To the best of my knowledge, there has been no research work published that analyzes the effectiveness of semi-deterministic models like the ITU models for LoRa's coverage prediction. This is important to analyze as these models

consider more phenomena in detail compared to OH model. Chapter 5 will look into one such semi-deterministic channel model: ITU-R 1812 for LoRa. The aim is to inspect if the added complexity of the semi-deterministic model would yield more accurate results compared to the popular empirical model (Okumura Hata model).

Chapter 3

Background and Related work: Remote Sensing

Providing a completely automated estimation tool for LoRa requires knowledge from two research areas: channel models and remote sensing. The prior chapter discussed the background and work related to channel models, this chapter focuses on the remote sensing part. The remote sensing part plays a key role in providing the necessary inputs for the OH model without needing the user in the loop.

This chapter will be split into two sections; Section 3.1 will give some background information on remote sensing and Section 3.2 will dive into the details of the research work [41] where remote sensing technique is used to completely automate LoRa's link quality estimation.

3.1 Remote Sensing: Overview

Remote sensing is a technique for acquiring information about the physical characteristics of objects on the earth surface or in the atmosphere, using satellite or aircraft-based sensor technologies. It is being widely adopted in various fields for monitoring landscape transformations, for natural resource management, to increase awareness for better risk analysis and hazard management, etc. Remote sensing systems are typically classified into two categories: Passive and Active systems.

Passive systems use external sources of energy, mainly the sun's radiations. Sensors mounted on satellites collect the sun's radiations reflected off obstacles in the target area and analyze the spectral response. A spectral response refers to the reflectance/emittance of a substance with respect to the wavelength. Depending on the composition, every object can treat radiations of different wavelengths differently, leading to a unique spectral signature. On the other hand, active sensors use their own source of energy

(e.g. LiDAR, Radar) and record and analyze the intensity, phase and time delay of the backscattered pulses to obtain specific information. They can be used when a specific target needs to be examined or to investigate particular wavelengths that are not sufficiently provided by the sun's radiation [15].

The reflected energy variations are detected and recorded by light-sensitive films on the sensors and converted to images. These images are usually represented in a digital format and are subdivided into equal-sized pixels. Each pixel contains a value denoting the brightness/ intensity level of the reflected signal.

Sections 3.1.1 and 3.1.2 will highlight two important use cases where the information acquired from remote sensing is applied: Land cover classification and Digital Elevation Models.

3.1.1 Land cover classification

The land cover represents the physical material on the surface of the earth like trees, water, building, grass, etc. The method uses techniques to analyze spectral information represented as numeric values from one or more spectral bands [15].

There are numerous land cover products freely available to the public. Each product differs in their resolution, accuracy, area of coverage, method of extraction and land cover classes they can distinguish between. Table 3.1 highlights some of this information obtained from [42].

Land cover product	Pixel resolution	Accuracy	Classes
MCD12Q1	500m	71.6%	17
UMD	1km	65%	12
GlobCover2009	300m	67.5%	22
GlobeLand30	30m	80.3%	10
GLC250 m_CN	350m	75.17%	9

Table 3.1: Examples of Land cover products

3.1.2 Digital Elevation Models

A digital elevation model (DEM) is a digital representation of a terrain's elevation referenced to a particular vertical datum. Vertical datums are the zero level reference surface for vertical positions. DEM's are useful in various applications: landscape modeling, land-use studies, 3D visualizations, flood control etc.

DEM's come in two forms: Digital Surface Models (DSM) and Digital Terrain Models (DTM). DTM contains the heights of the bare-earth surface

excluding the heights of objects like buildings, vegetation, etc. On the other hand, DSM records the heights of the objects above the ground surface. Figure 3.1 below illustrates the difference between DSM and DTM.

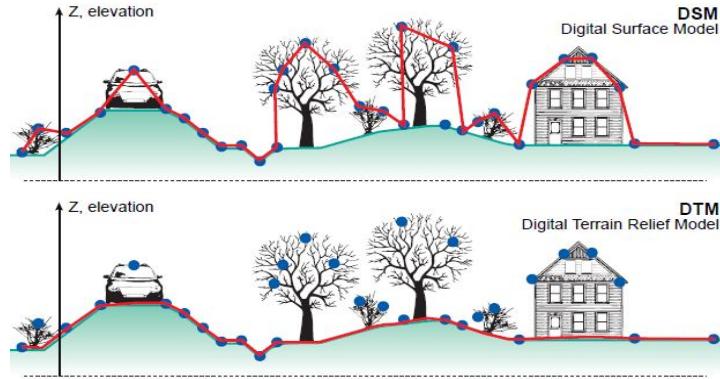


Figure 3.1: Difference between DSM and DTM [1]

DEM products provide remotely sensed surface heights, each varying with respect to their technique of extraction, accuracy, resolution, and area of coverage. Table 3.2 shows some commonly used DEM's.

DEM product	Pixel resolution	Absolute Accuracy	Coverage extent
SRTM [31]	30m x 30m	<16 meters	60.0° N -56.0°S
ASTER [2]	30m x 30m	17 meters	83.0° N -83.0°S
ALOS [3]	30m x 30m	5 meters	60.0° N -60.0°S
AHN	5m x 5m	5 centimeters	Netherlands

Table 3.2: Example of Digital Elevation Model products

Actual Heights of Netherlands (AHN) is one such elevation model that will be used throughout the thesis. Section 3.1.2.1 will describe AHN in more detail.

3.1.2.1 Actual Heights of Netherlands

The AHN project is responsible for providing high-resolution elevation data covering the entire Netherlands free of charge. The terrain height profile is obtained via airborne laser scanning or LiDAR (Light Detection And Ranging), where the target area is illuminated using laser pulses. The reflected energy is captured by a sensor and the amplitude, difference in phase and time of arrival is used to compute the surface heights. The use of LiDAR poses several advantages over other methods. LiDAR is well known for it's

higher accuracy, quick extraction time and higher sampling density. The elevation is compared to the Normal Amsterdam Level (NAP) vertical datum while the horizontal datum follows the Amersfoort/RD new coordinate system.

ahn is an initiative by government bodies: Rijkswaterstaat, the water boards, and the provinces to create up-to-date height maps for water management applications in the Netherlands. The database was made public for commercial and non-commercial purposes in 2014 [54]. The elevation database is available as a web map on the PDOK portal of the Dutch government [12]. The large data set is split into smaller tiles of size 5 x 6.25 kilometers as shown in Figure 3.2. The entire Netherlands is covered by a total of 1372 tiles [28].

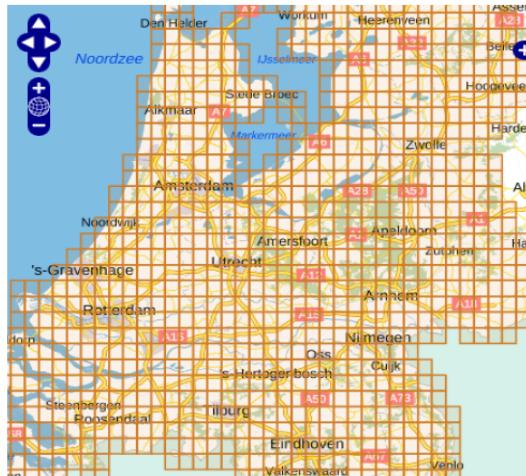


Figure 3.2: Web-map of Netherlands composed of AHN tiles [12]

Due to technological advancements and due to natural and man-made changes in the environment, there have been three public versions of the AHN: AHN1, AHN2, AHN3. While AHN1 contained 1 sample point per 16-32 square meters, AHN3 is updated to contain 8-10 points per square meter on an average. It has high vertical accuracy with ≤ 5 cm systematic error and 5 cm standard deviation [54]. These detailed high-density point clouds are difficult to work with and therefore are resampled into equidistant grids to form user-friendly raster maps. Rasters are a grid of evenly spaced cells (pixels) arranged in rows and columns representing the latitude and longitude. Each of the pixels in a DEM raster contains a numerical measurement value that represents the elevation at that location. The rasters are available in two resolutions of 0.5 x 0.5 meter and 5 x 5-meter grids.

LiDAR works on the concept of multiple returns from a single pulse transmission. As shown in Figure 3.3, depending on the number of reflective surfaces the laser pulse encounters in its path towards the ground, it gets

split into as many returns. The AHN DSM heights are defined by the first return of the pulse providing all the heights on the ground except for water. Water does not reflect laser signals and hence provides No-Data values in the raster. The AHN DTM is defined by the last return providing height information only for the exposed ground area that can be reached during laser scanning. Naturally, the ground elevation underneath a building is never going to be obtained as the pulses are blocked by the building itself. Therefore, the raster will show No-Data values where ever an object is present. So for example, for pixels that contain a building, only DSM height is available and not DTM. To obtain obstacle heights above ground level (i.e., by subtracting DTM from DSM), the No-Data values in the DTM need to be filled using interpolation techniques as will be seen in Chapter 4.

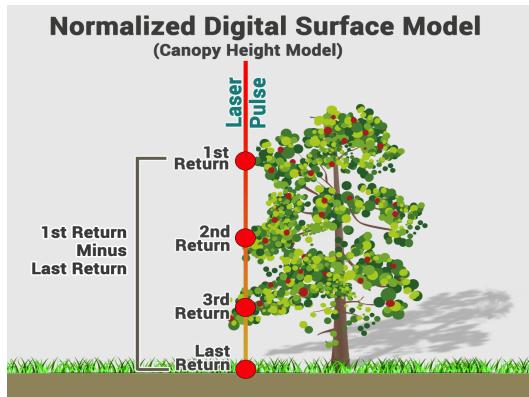


Figure 3.3: Multiple returns from a single LiDAR transmission [34]

Regarding the DSM heights for vegetation, unfortunately, the first leaf/branch that the laser hits does not necessarily have to be the top of the tree. This would mean that the first return of the laser pulse might not always represent the actual height of the tree. On the other hand, if the vegetation is sparse enough, with multiple returns the signal can likely pass through the vegetation to the ground surface and give correct DTM height values in spite of being an obstacle to the laser pulses.

3.2 Related work using remote sensing data

My thesis is inspired by the work proposed in [41]. The research paper complements the related work in Section 2.3.2 by adopting an automated approach that no longer relies on the user pre-defining the type of environment for path-loss models. The authors utilize remote sensing and propose a dedicated processing toolchain for land-cover classification. This information is used to automatically classify the type of environment and further estimate LoRa's link quality using the Okumura-Hata channel model.

3.2.1 Land-cover maps

The paper exploits multi-spectral images of the Sentinel-2 satellite operated by the European Space Agency (ESA). The images are acquired via passive remote sensing with a high spatial resolution of 10 m and having multi-spectral data with 13 bands in the visible, near-infrared, and short-wave infrared range. This data can be obtained from ESA archives free of charge. The authors analyze the spectral signature and apply supervised classification techniques based on machine learning to generate land-cover maps. Every pixel on the map corresponds to one of the seven classes of land cover as shown in Table 3.3. The paper claims to achieve high accuracy $\geq 90\%$ with a high pixel resolution of 10 m compared to the land cover products highlighted in Section 3.1.1.

Clutter category	Low-clutter category
Buildings	Field
Greenhouse structures	Soil
Trees	Road
-	Water

Table 3.3: Land cover classes distinguished by the paper

The information obtained from the sequence of land-cover between the transmitter and receiver is used to automatically define the "type of environment" and select the OH variant. This will be explained in detail in the next section. The paper chooses the small/medium city category under the urban OH variant for the analysis as it best suits the city structure of the Netherlands.

Sections 3.2.2 and 3.2.3 discuss the two prediction approaches used in the paper to decide the environment type: Path analysis and Intersection analysis.

3.2.2 Path analysis

This section will describe the logic used by the Path analysis. The tool uses the land cover map to obtain the sequence of land cover classes along the route between the transmitter and receiver. Consider a 1 km path in which the link is divided into segments of 10 meters and each segment is categorized based on the type of land cover. The land cover classes are grouped into two categories: clutter group and low-clutter group. This classification is done depending on whether they introduce significant attenuation (clutter group) or do not affect the signal propagation (low-clutter group) as seen in Table 3.3. Land cover such as soil, road, water, etc. is classified as components of the low-cluttered group, whereas trees, buildings, greenhouse fall under

the cluttered group. Finally, a simple comparison is made between the total percentage of the cluttered and low cluttered groups. If the majority of segments are "clutter", the model selects the "urban" type of environment else it uses the "sub-urban" type of environment. This classification between an urban and suburban area can then be applied to the Okumura Hata model to select the variant. The flowchart in Figure 3.4 illustrates the Path analysis algorithm.

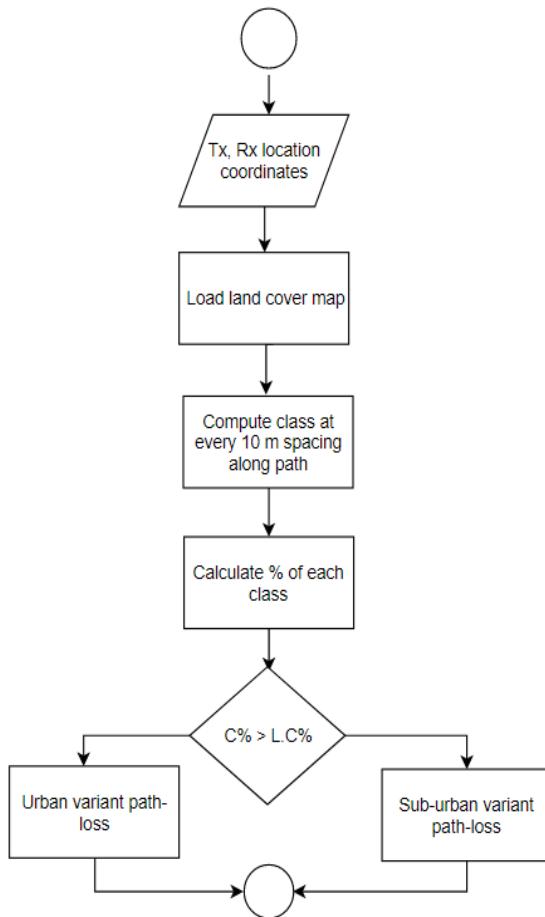


Figure 3.4: Path analysis algorithm to select the OH variant

3.2.3 Intersection analysis

While the Path analysis in Section 3.2.2 considers all the land cover between the LoRa end device and the gateway, the Intersection analysis considers only the land cover surrounding the end device. The idea behind the Intersection analysis is to capture the effect of the obstacles in the vicinity of the

end device in a better manner. Figure 3.5 illustrates the difference between Path analysis and Intersection analysis.

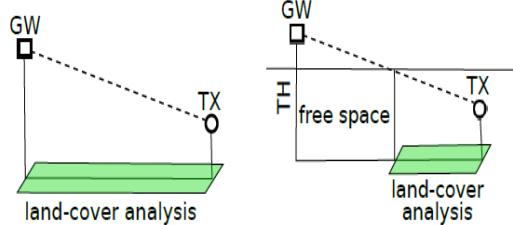


Figure 3.5: Difference between Path and Intersection Analysis [41]

The threshold T_H is assumed to be 10 meters and represents the average height of the buildings in the test area. Figure 3.5 shows the threshold intersecting the line-of-sight between the gateway and the LoRa node. In this case, only the land cover between the node and the intersection is considered to select the OH variant (suburban or urban). The path between the intersection and the gateway is considered free space with no obstructions. Moreover, if no intersection is detected, the rural/free/open space path loss variant is selected directly. The flowchart in Figure 3.6 illustrates the Intersection Analysis algorithm.

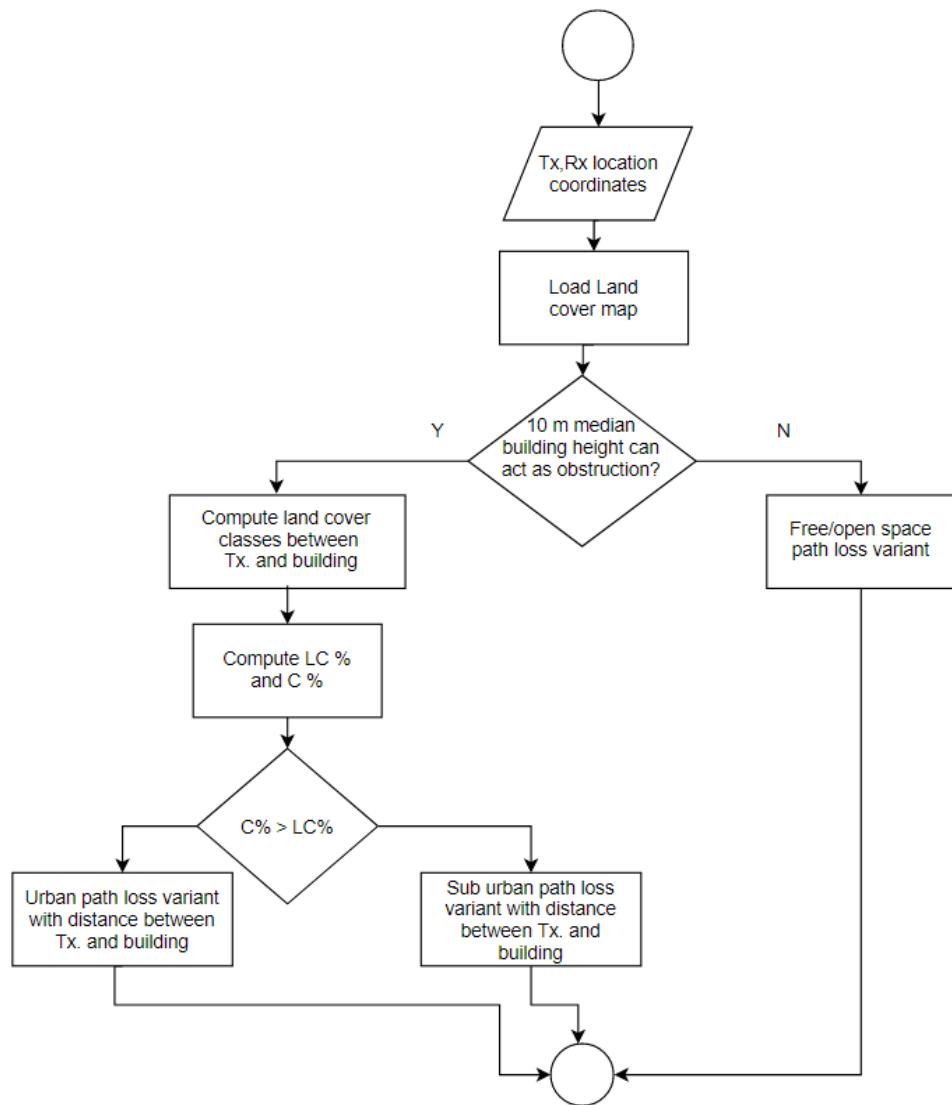


Figure 3.6: Intersection Analysis algorithm to select the OH variant

Chapter 4

System and Experimental Setup

The software environment used for the thesis is MATLAB. The Mapping toolbox package must be downloaded to utilize the functions for analyzing and editing geographic data. Section 4.1 will discuss the TTN data used for validation. Section 4.2 will discuss the setup method to prepare the remotely sensed data.

4.1 Validation with TTN Mapper data

The data used in the thesis is provided by The Things Network (TTN). This data serves as the ground truth for validating the results.

TTN initiated an open-sourced and crowdfunded Internet Of Things network based on LoRaWAN [4]. The company has around 7603 gateways operating across 140 countries. TTN Mapper is an application integration provided by the TTN [5]. The application can be used to visualize the coverage of the TTN network and analyze the performance of the gateways. It is a worldwide community effort which allows people to upload the uplink packet information from their LoRa node to the TTN Mapper through the network's back-end directly.

The data is provided as .csv files. The file includes information of each LoRa node transmission received by the gateways such as the Received Signal Strength Indication (RSSI), the Signal-to-Noise Ratio (SNR), the operating frequency and the location coordinates, height and IDs of the LoRa nodes. Only gateways located outdoors are used for validation in the thesis. Information about the placement of the gateways i.e. if the gateway is indoors or outdoors can be acquired using the following API provided by TTN: [`https://account.thethingsnetwork.org/gateways/<gateway-id>`](https://account.thethingsnetwork.org/gateways/<gateway-id>)

The following API can be used to get the location coordinates and altitude of the gateway:

<http://noc.thethingsnetwork.org:8085/api/v2/gateways/<gateway-id>>

The API provides the location reported by the gateway from its Global Positioning System (GPS) or as configured statically on the gateway. GPS uses the World Geodetic System (WGS84) as its reference coordinate system. Hence, the elevation of the gateway is likely relative to the WGS84 spheroid and not with respect to the ground. Moreover, GPS altitudes are known to have low accuracy compared to the latitude and longitude values. To obtain a better estimate of the height of the gateway from the ground, digital elevation models are used to compute the building height and the altitude value obtained from the API is only used as a sanity check.

To ensure that transmitter nodes are not present indoors, the samples from the TTN data files with node height less than and equal to 2 meters is used for evaluation as in the paper [41]. This is done assuming that as the node height increases, the probability of a node being indoors increases.

The packet information does not include the transmit power of the node and antenna gains of the transmitter and receiver. The paper [41] assumes a transmit power of +14 dBm which is the maximum allowable limit in Europe and a typical half-wave dipole antenna (gain of 2 dBi) for the evaluation and the thesis uses the same settings. This assumption will tend to overestimate the path loss and the thesis will later delve into the issue of doing so in Chapter 6.

Before using the data for testing purpose, the samples are filtered using python. The first step of filtering involves separating the large data set into individual gateways. Then the data for each gateway is further separated into individual nodes. Testing is done for each node and gateway combination. The second step of filtering removes samples that are outside the boundaries of the available land cover map. The third step is to exclude gateways that do not meet the antenna heights and link distance as stated by the OH model as stated in Section 2.3.2.

The sample space included 8 gateways of varying heights between 30m to 110m, 105 transmitter nodes with a total of 18000+ recorded packets. The gateways in the sample space are located in the following regions: Delft, Utrecht, and Arnhem. The data was collected between the end of 2016 to the beginning of 2019.

Since LoRa can decode signals with power up to 20dB below the noise floor, the Expected Signal Power (ESP) value is used as the metric to denote the received power of the signal Figure 4.1 shows a graph illustrating the variation of ESP for different values of SNR assuming a fixed value for RSSI. ESP has a major effect when the SNR tends to get more negative, while for positive SNR the ESP becomes similar to RSSI.

ESP can be calculated from the RSSI and SNR as shown below:

$$ESP(dBm) = RSSI(dBm) + SNR(dB) - 10 * \log_{10}(1 + 10^{SNR(dB)/10}) \quad (4.1)$$

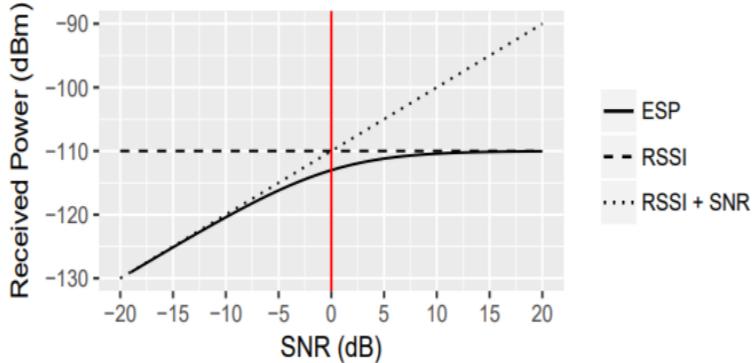


Figure 4.1: Graph comparing ESP with RSSI for different SNR [35]

Root-Mean-Square-Error (RMSE) is the metric used to evaluate the difference between the measured pathloss and the predicted pathloss. This metric is preferred as it is more sensitive to large errors which are highly undesirable. The equation is as follows :

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^T (x_{1,t} - x_{2,t})^2}{T}}. \quad (4.2)$$

Where:

x_1 : Actual measured value

x_2 : Predicted value

T: total number of samples

4.2 Preparing the remotely sensed data

Two types of remote sensing data are used in the thesis: Landcover maps and Digital Elevation Models (DEM).

The land cover maps are provided by the authors of the paper [41]. Figure 4.2 shows the Google map screenshot marking the boundary of the land cover map available for testing purposes.

The AHN digital elevation models need to be processed before using it for coverage predictions. For validation using the TTN data, a python script is prepared to get the limits of the test area for each gateway. Each tile (DSM and DTM) within the test area is manually downloaded from [12].

The horizontal datum is based on the Amersfoort/RD new coordinate system. The first step in preparing the DEMs is to convert the horizontal datum to the more universally accepted WGS84 coordinate system. WGS84 is a geographic coordinate system that uses latitude and longitude to measure the location on the earth surface. There were two options to convert the

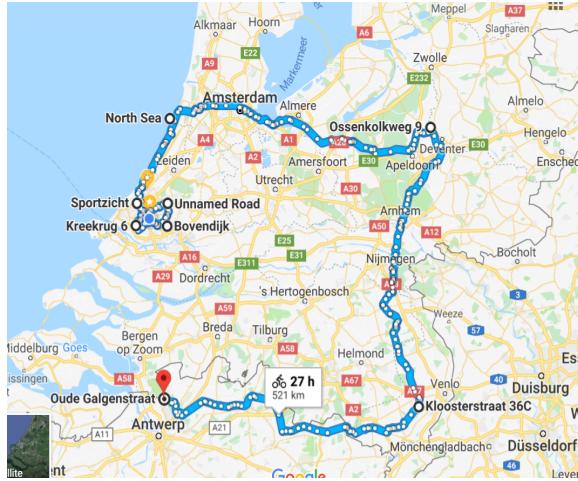


Figure 4.2: Extent of the available land cover map

coordinate system: using the default functions provided by MATLAB or by using QGIS software. QGIS is a free and open-source geographic information system (GIS) application that is used to analyze and manage geospatial data [6]. Several points were transformed from the RD new coordinate system to the universal coordinate system using both the methods separately. It was observed that the transformed points did not match and there was more than a 100-meter difference between them. To decide which tool is more precise in its conversion, Google-maps is used as a reference. For example, the elevation of one such point in the Dutch coordinate system in a DSM raster is noted and the point is converted into the WGS84 coordinate system using both tools. The tools generate two different location coordinates as shown in Table 4.1. The points lie about a 100 meters apart as can

Tool	
Test point (RD new coordinate system)	X: 456230.014000000, Y: 85004.997500000
Height of test point from DSM	24.19 meters
MATLAB	52.090888763158638, 4.366147468400354
QGIS	52.089892933394005, 4.365825794955500

Table 4.1: Example test point transformed into WGS84 coordinate system

be seen in Figure 4.3a. It is clear that the point generated by MATLAB does not map to an object of height 24 meters and instead points to the ground as shown in Figure 4.3c, while the point generated by QGIS points to a building in Figure 4.3b. Several such cases are analyzed and eventually QGIS software is chosen as it is more reliable in transforming the coordinate system.



(a) Difference in locations generated by QGIS and MATLAB



(b) Close-up of coordinate by QGIS



(c) Close-up of coordinate by MATLAB

Figure 4.3: Example illustrating the reliability of the tools coordinate conversion

An additional step must be carried out for the DTM tiles using QGIS. As discussed in Section 3.1.2.1, the DTM raster contains No-Data values in pixels that are occupied by an obstacle. To calculate obstacle heights (i.e., obstacle height = DSM-DTM), the missing DTM height data for the No-Data pixels is filled using Inverse Distance weighting (IDW) interpolation. This technique uses height values from the surrounding pixels to make predictions for the No-Data pixel. As the name suggests, IDW assumes that points closer to the No-Data pixel are more similar and have more influence compared to points that are farther away. Approximating the ground heights using interpolation is acceptable considering the flatlands of the Netherlands.

On changing the coordinate system using QGIS, the raster restructures and re-projects itself according to the WGS84 coordinate system as shown in Figure 4.4. The elevation values shift under the new coordinate system and more No-Data values are noticed in the rasters. In Figure 4.4, the green area represents the elevation data, whereas the white area represents No-Data values. The elevation information for the points with No-Data value in a tile is now available in the adjacent tile instead. The tiles now partly overlap with its adjacent tiles.

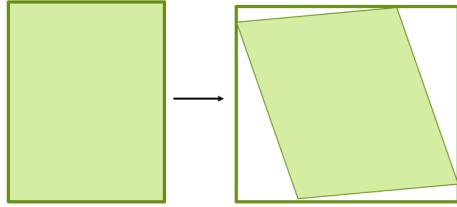
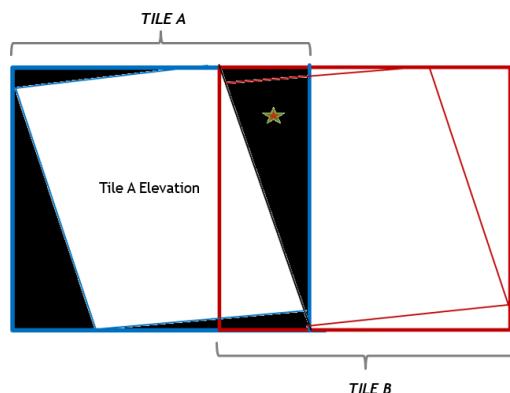
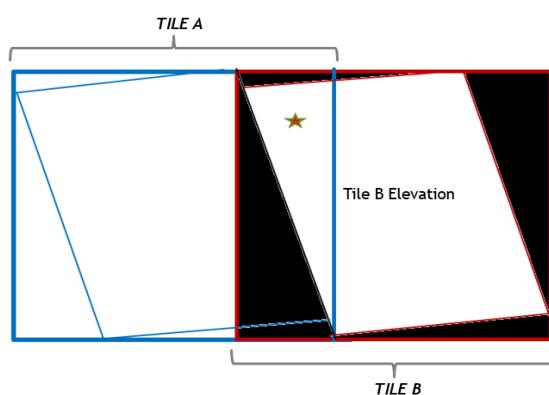


Figure 4.4: Restructuring of the raster after coordinate transformation

Figure 4.5 gives an example of two adjacent tiles (tile A outlined in blue and tile B outlined in red) and the black area for each tile are the No-Data pixels while the white area represents pixels with the actual elevation. For example, assume that the simulator requires the height information of a point (marked as a star in Figure 4.5) that lies in the overlapped region of the two tiles. The same point in tile A will give a No-Data value as shown in 4.5a while tile B will give the correct elevation value as seen in 4.5b. This additional processing to ensure that the right tile is chosen will need to be done for every point along the path and can increase the processing time and complexity. This issue is avoided by merging all the tiles and processing them as one single unit.



(a) Point in the black region gives No-Data value while processing Tile A separately



(b) Point in the white region gives height value while processing Tile B separately

Figure 4.5: Transformed rasters overlap and fill each others elevation values

Chapter 5

Evaluation of ITU-R 1812 for LoRa

This chapter will analyze one of the ITU-R recommendations of a semi-deterministic model: ITU-R 1812 for LoRa. Section 5.1 will give an overview of the model followed by a comparison of the model with real data in Section 5.2.

The International Telecommunication Union (ITU) is an agency of the United Nations. The organization is responsible for radio spectrum allocation, improving telecommunication infrastructure and managing satellite orbit resources. ITU-R is one of the divisions of ITU that is responsible for developing technical standards for radio communication [7]. ITU models are chosen for two reasons for this thesis: Firstly, these models are found to be commonly used by commercial coverage simulators [13],[14],[20],[27]. Secondly, ITU provides free MATLAB implementations for most of its recommendations [8]. Among the various recommendations produced by ITU-R for path-loss estimation, ITU-R 1812 was selected for evaluation. This model is a good fit with respect to the frequency range for LoRa but also fits the link distances, antenna height requirements of the OH model which is important to do a comparison of the two models for LoRa. It also considers attenuation due to the atmosphere and surrounding obstacles which is important to analyze for a long-range technology like LoRa.

5.1 Overview of ITU-R 1812

The ITU-R 1812 recommendation is developed for propagation prediction for terrestrial services in the VHF and UHF bands (30 MHz - 3 GHz). It predicts a median transmission loss which should not be exceeded for p% of an average year [9]. This parameter is set by the user and can be in the range $1\% \leq p \leq 50\%$. The model allows for path lengths between 0.25 km up to 3000 km and terminal antenna heights of up to 3 km above the

ground. It considers the path to be symmetric with respect to the terminals (i.e, the positions of the transmitter and receiver can be interchangeable).

The model accounts for various losses as follows:

- Line-of-sight (with short term multi-path effects)
- Diffraction (including smooth-Earth, irregular terrain and sub-path cases)
- Tropospheric scatter
- Anomalous propagation due to ducting and layer reflection/refraction
- Height-gain variation in terminal clutter

Table 5.1 displays the basic input parameters that must be prepared by the user to estimate these effects.

Input Parameter	Description
Frequency	0.03 Ghz - 3 Ghz
Transmitter, Receiver coordinates	In degrees
Antenna heights	1 - 3000 meters
p%	% of average year the calculated signal level should not exceed
Clutter and Ground heights	In meters above sea level
Radio climatic zones	Sea, coastal, inland
Terminal distances from sea coast	In meters
Radio meteorological parameters	$N_0, \Delta N$
Width of the street	Default 27 meters
Antenna polarization	Vertical, Horizontal

Table 5.1: Basic input required by ITU-R 1812 model

1. The link distance, terminal heights, and frequency are common inputs used to calculate the Line-of-Sight loss with some short term multi-path effects. These inputs variables are similar to that of the OH model. The input defining the “type of environment” in the OH model is taken into account by the ITU-R 1812 in finer detail by considering losses due to obstacles and the atmosphere separately.
2. The model uses the terrain elevation above mean sea level for each point along the great circle path. The great circle path is the shortest path traced between any two points on the surface of the earth. The clutter and bare-ground heights are mainly used to account for the following:
 - Diffraction losses due to the bulge of the earth and the obstacles along the path:

The recommendation uses the Bullington diffraction model. Details of this model will be addressed in Section 5.2.

- Terminal clutter losses:

This loss is considered when a clutter surrounding the terminal is located at a height above the terminal. The model distinguishes between two cases: for smaller terminal clutter the dominant mechanism is identified as scattering and reflection, whereas for bigger clutter heights diffraction is considered to be the dominant phenomenon.

The remotely sensed terrain heights can be obtained from the AHN digital elevation models as discussed in Section 3.1.2.1.

3. Scattering due to the troposphere layer of the atmosphere and atmospheric ducting account for the attenuation by the atmosphere.

Ducting is a radio propagation phenomenon where the lower layer of the atmosphere acts as a horizontal duct guiding the signal to follow the earth curvature. The refractive index decreases with higher altitudes which can bend the signal away from the anticipated direction. Due to the effects of reflection and refraction from the boundary of a region with a lower refractive index, the signal tends to remain within this duct.

The following highlights the input parameters that account for atmospheric attenuation.

- Radio climatic zones: The model distinguishes between three radio climatic zones: Sea, Coastal land and Inland. It requires data on what lengths of the propagation path lies in each zone. If a point on the path lies at least 50 km away from the nearest sea, the point can be assigned to the inland zone.
- Terminal distance from the coast: If a portion of the path lies over the sea, the model requires the distance of each terminal from the coast in the direction of the other terminal. If a terminal is on the sea, the distance is considered to be 0.
- Basic Radio-meteorological parameters: To describe the variability of the atmosphere's refractivity, the model uses two parameters:
 - N_0 (N-units) - represents sea-level surface refractivity
 - ΔN (N-units/km) - represents the lapse-rate of the refractive index on an average, through the lowest 1 km of the atmosphere.

This data can be extracted from digital maps provided by ITU. If a location needed for the prediction process does not match

any grid point in the map, bilinear interpolation using the four closest values must be performed by the user. These global maps are derived from a ten-year analysis from 1983 to 1992.

4. All of the above losses are calculated such that they do not exceed the value for p% of the year.

5.2 Results and Conclusion: ITU-R 1812 vs. Path Analysis

On running a few initial sample points with the ITU model and comparing the predicted path loss to the measured path loss as shown in Figure 5.1, it is observed that the path loss predicted is unusually high in a few cases. Especially the samples marked in red show path loss values reaching up to 170 dB for distances within just 3 km.

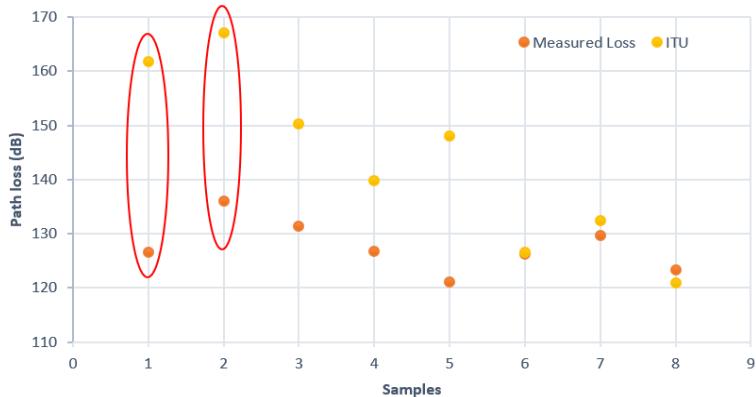


Figure 5.1: ITU pathloss with diffraction problem

To be certain there is no error caused in the way the input is prepared for the model, the path loss components of the ITU model is evaluated. On close analysis, it is observed that the diffraction component causes the high path losses even though the path does not have any dominant obstacle blocking the line-of-sight. This occurs because the receiver acts as an obstacle to itself. To understand this better, a brief understanding of the Bullington diffraction model is required. Firstly, the diffraction model finds the first obstacle obstructing the line-of-sight from the transmitter to the receiver end. Similarly, the first obstacle from the receiver to the transmitter end is detected. Then the model constructs a single equivalent knife-edge obstacle at the intersection of the transmitter and receiver lines as shown in Figure 5.2 and calculates the diffraction loss associated to that height.

Figure 5.3 shows a google map screenshot of the building on which the

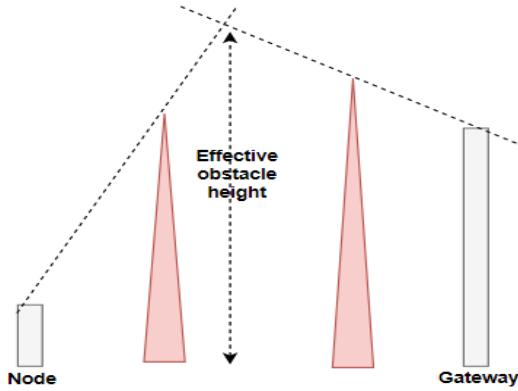


Figure 5.2: Illustration of the Bullington diffraction model

gateway is placed for the same samples as shown in the Figure 5.1. Figure 5.4 gives an idea of the shape and the height information of the building obtained from the digital elevation models. The elevation values are not constant and vary by a few meters in different pixels. This is likely due to the objects on the surface of the roof.

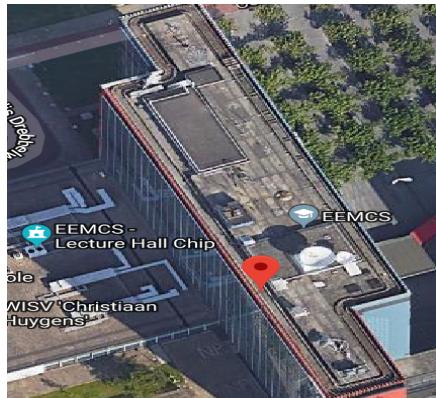


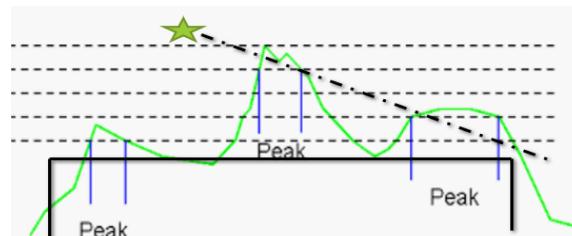
Figure 5.3: Google map image of the gateway's location

We use the DEM's to get an approximate height of the gateway above the ground, using the GPS heights in the TTN data files for a sanity check. The pixel selected in the DEM in Figure 5.4 represents the location of the gateway. Assuming the gateway height is 69 meters, the surrounding pixels with greater height behaves like an obstruction.

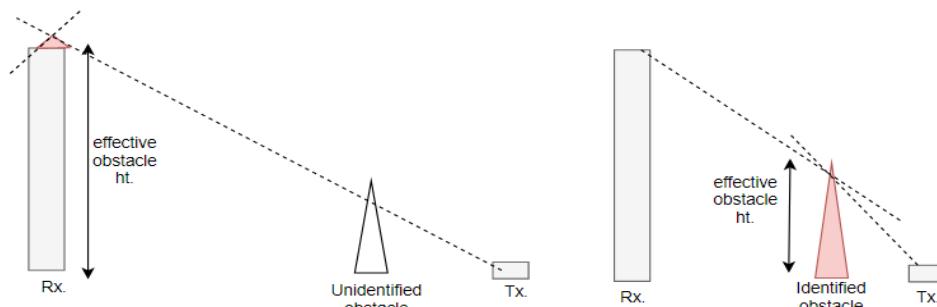
Figure 5.5a tries to illustrate the varying height values of the roof obtained via DEM and the star represents the gateway. As shown in the Figure 5.5b, with the points on the building acting as an obstacle to itself, the effective obstacle height for the Bullington model increases and hence the diffraction loss increases.

0	0.0220000	8.0450001	30.8770008	0.4060000	0.1900000	0.2780000	0.2780000	0.1360000	0.1270000	0.2420000
0	0	73.5370026	71.3669968	16.1630001	0.3010000	0.1580000	0.1580000	0.1330000	0.2150000	0.1490000
0.0060000	-0.4270000	44.1716014	68.0910034	66.4840012	6.9000001	0.2810000	0.2810000	0.2330000	0.0910000	0.3310000
1.2350000	2.9530001	67.2870026	67.2639999	64.3970032	31.3910007	0.1850000	0.1850000	0.1850000	0.1810000	0.1040000
1.2350000	2.9530001	67.2870026	67.2639999	62.4150009	56.2700005	1.8060000	1.8060000	0.1480000	0.0990000	0.0040000
9.2620001	8.7440004	61.5079994	68.9110031	62.4150009	56.2700005	1.8060000	1.8060000	0.1480000	0.0990000	0.0040000
9.8889999	9.8839998	41.3480011	65.2689972	66.8190002	63.4220009	63.4220009	20.7669983	0.1510000	0.0200000	0.0180000
9.4989996	9.3350000	10.3719997	64.5400009	70.6350021	59.9860001	59.9860001	47.8330002	0.7800000	1.9029999	0.2840000
9.5369997	9.3290005	-0.0250000	51.7729988	67.7799988	66.3960037	66.3960037	60.2760040	11.7480001	2.8680000	1.2520000
9.3120003	9.7900000	3.7830000	27.0070000	65.5989990	68.9540024	68.9540024	60.3190000	36.8730011	0.4560000	0
9.2980003	9.5500002	7.5679998	8.2570000	54.0099983	71.3089981	71.3089981	60.0079994	60.4980011	3.2890000	0
9.2770004	9.2919998	6.3179998	0	34.1930008	62.1520004	62.1520004	65.2789993	59.0919991	26.1919994	0
0.2270000	0.8820000	2.6589999	0.5460000	0	47.2929993	47.2929993	69.9020004	51.6839981	14.3459997	7.3320003
2.5250001	0.4760000	1.2029999	1.0360000	0	42.3969994	42.3969994	63.6959991	56.5879974	3.3479998	0.1580000
0.6570000	0.9630000	0.3690000	1.0520000	6.5970001	2.2670002	2.2670002	47.4619980	60.4050982	22.7849998	0.0060000
0.0590000	0	0.0050000	0.8190000	0.0670000	0.0670000	0.0180000	28.6459999	48.3349991	20.5569992	0

Figure 5.4: Illustration of the building on which the gateway is mounted on using the digital elevation model



(a) Points on the receiver obstructing the line-of-sight



(b) Before and after filtering the points on the receiver shows a change in the effective obstacle height

Figure 5.5: Illustration of the diffraction problem

The assumption is made that the gateway will be placed in a position that avoids getting blocked by the building it is mounted on. To eliminate the possibility of such occurrences, the program makes sure that points within a certain distance from the gateway and having a height close to the gateway's height is not considered as an obstruction during the diffraction analysis. Figure 5.6 shows the reduction in path loss after discarding points on the building itself. Samples from 1 to 5 show predicted path loss closer to the real loss. But this also has a negative effect on the path loss of other samples (6-8), bringing it farther away from the real loss.

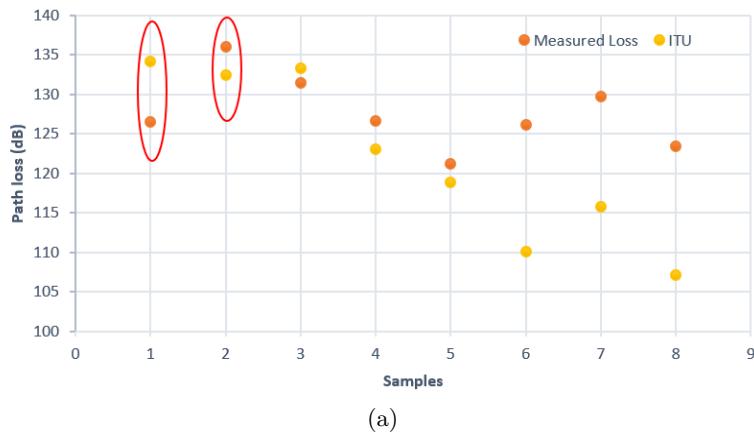


Figure 5.6: ITU pathloss eliminating diffraction problem

Figure 5.7 gives an overall comparison of the ITU model with the Path analysis [41] for five gateways (GW1-GW5). The comparison is made for the ITU model with input parameter $p\%$ set as 50%, 25%, 10%, and 1%. The higher the $p\%$ (which represents the percentage of the average year the estimated path loss should not be exceeded), the more pessimistic the model is and the path loss is comparatively higher. The graphs plot the average, maximum and minimum RMSE among all the user nodes under each gateway.

In general, the average RMSE of the Path analysis is lower or comparable to the ITU model with a drop in the maximum RMSE except for GW2. The ITU model does not show any significant improvement compared to the OH model in spite of all the extensive input information that the ITU model requires. The accuracy of path loss predictions does not seem to increase on using a more complex semi-deterministic model. Furthermore, the model heavily depends on the accuracy of its inputs for e.g; heights of the nodes and gateways for the diffraction model, making it difficult to fully validate this model with large LoRa datasets without a controlled measurement setup.

Since the Path analysis performs better than the ITU model in our analysis, the next chapter will delve into a deeper analysis of the automated

approach used by the paper [41] for LoRa's link quality estimation using the simple OH model.

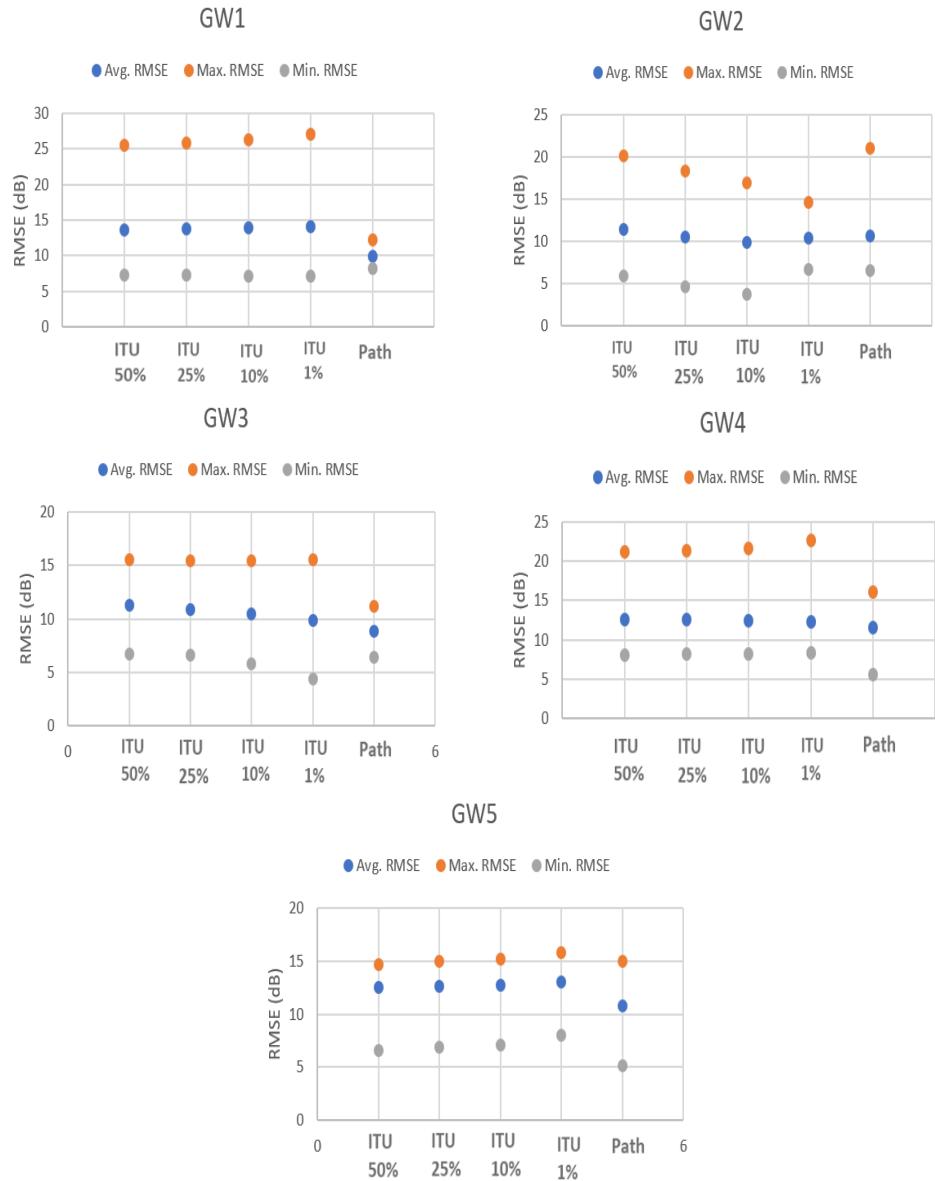


Figure 5.7: Comparison of ITU model with varying p% against Path Analysis

Chapter 6

Improved Pathloss Model Development

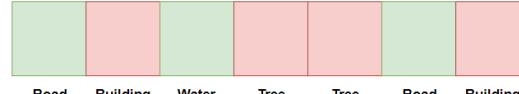
This chapter takes inspiration from the paper [41] and covers the development of the improvised algorithm for path loss prediction using the OH model. We have developed the final algorithm in a series of stages and at every stage, the approach is evaluated and improved upon. Sections 6.1, 6.2 and 6.3 will discuss the different approaches and the results and conclusion of each approach. Section 6.4 will describe the relative comparison as an alternative to evaluate the performance of the models.

6.1 Approach One

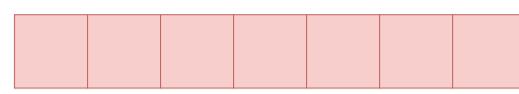
This section will discuss a different approach to apply the land cover map information to the OH model compared to the Path algorithm discussed in Section 3.2.2.

The paper [41] follows a binary decision-making process to distinguish between the urban and suburban-type of environment. This causes a point in between the path-loss curves defined by the urban and suburban OH variants to move to the closest curve, causing a quantization error as shown in Figure 6.2a. Instead, a weighted average approach can be implemented, using the percentages of the cluttered and low cluttered groups as weights. To explain this idea better, consider two paths as shown in Figure 6.1 that have the same link distance, frequency, and antenna settings. According to the Path analysis, both paths qualify as urban scenarios because the majority of the cells have clutter. But the two scenarios can be differentiated based on the degree of urbanization along the path. The path on the bottom shows more cluttered components compared to the path on the top. Hence, the percentages can be used as weights/ indicators to factor in this degree of urbanization. Depending on the percentage values, the path loss can now lie anywhere between the sub-urban and urban scenario as shown in Figure

6.2b and is no longer limited to the two discrete values on the path-loss curves of the model. This method no longer depends on a binary decision and deviates from the traditional method of using the OH model.



(a) Lower cluttered path



(b) Higher cluttered path

Figure 6.1: Example of two paths with different degree of urbanization

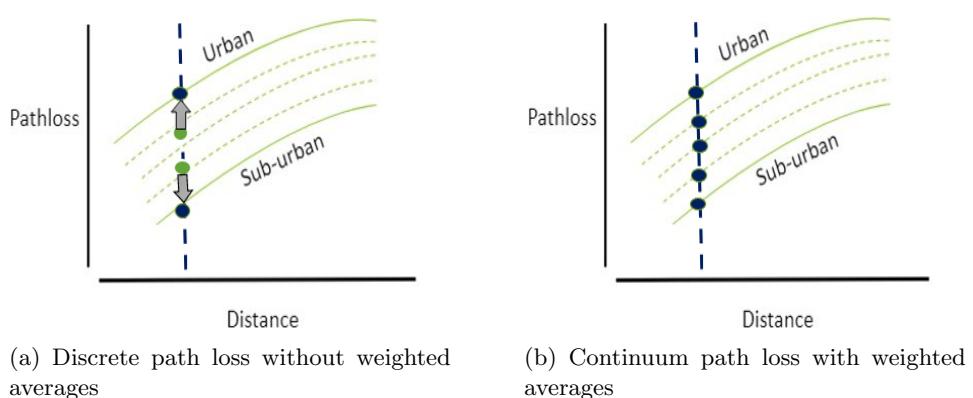


Figure 6.2: Pathloss versus distance graphs for OH model

Equation 6.1 illustrates the path loss obtained using the weighted average between the urban and sub-urban OH variants.

$$PathLoss(dB) = C\% \times L_{urb} + L.C\% \times L_{sub-urb} \quad (6.1)$$

where:

L_{urb} : Pathloss using the urban OH variant (dB)

$L_{sub-urb}$: Pathloss using the sub-urban OH variant (dB)

C% : Cluttered components percentage

L.C% : Low-cluttered components percentage

The flowchart in Figure 6.3 explains the modified algorithm using the weighted average approach.

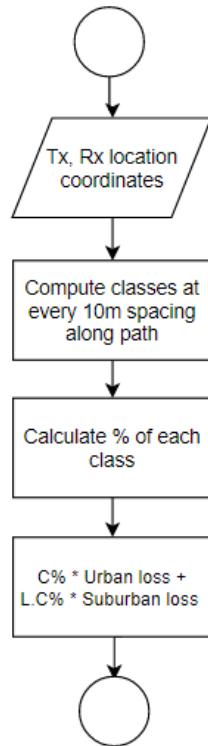


Figure 6.3: “Approach One” Algorithm

6.1.1 Approach One: Results and Conclusion

6.1.1.1 Approach One: Analysis

This section compares the results from the Path Analysis and Approach One. The graphs highlight few samples of user nodes and gateway combinations and the associated RMSE in the path loss.

While the nodes in Figure 6.4 show a reduction of the RMSE (dB) in Approach One compared to the Paper Path, the graph in Figure 6.5 shows an opposite effect for other nodes. In Figure 6.5 the Path analysis gives better results than the weighted average approach. To debug this issue, Section 6.2 will delve into investigating the output of the land cover map.

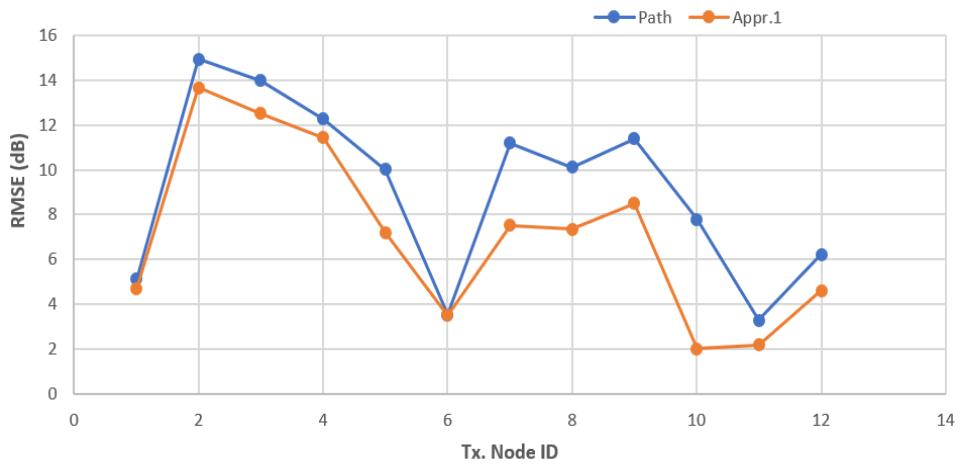


Figure 6.4: Nodes showing improvement with Approach One

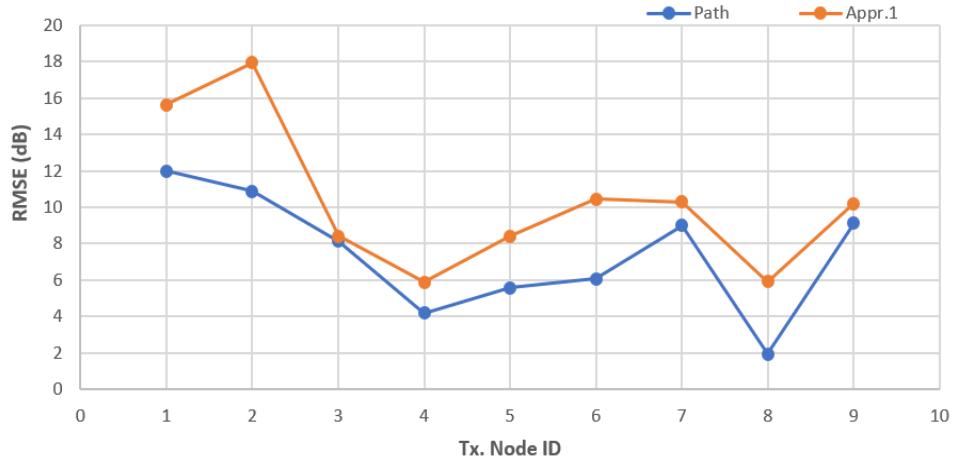


Figure 6.5: Nodes showing higher errors with Approach One

6.1.1.2 Approach One: Overall comparison

The graphs in Figure 6.6 give an overall comparison of the average, minimum and maximum RMSE of each gateway across all transmit nodes for the Approach One, Path and Intersection analysis.

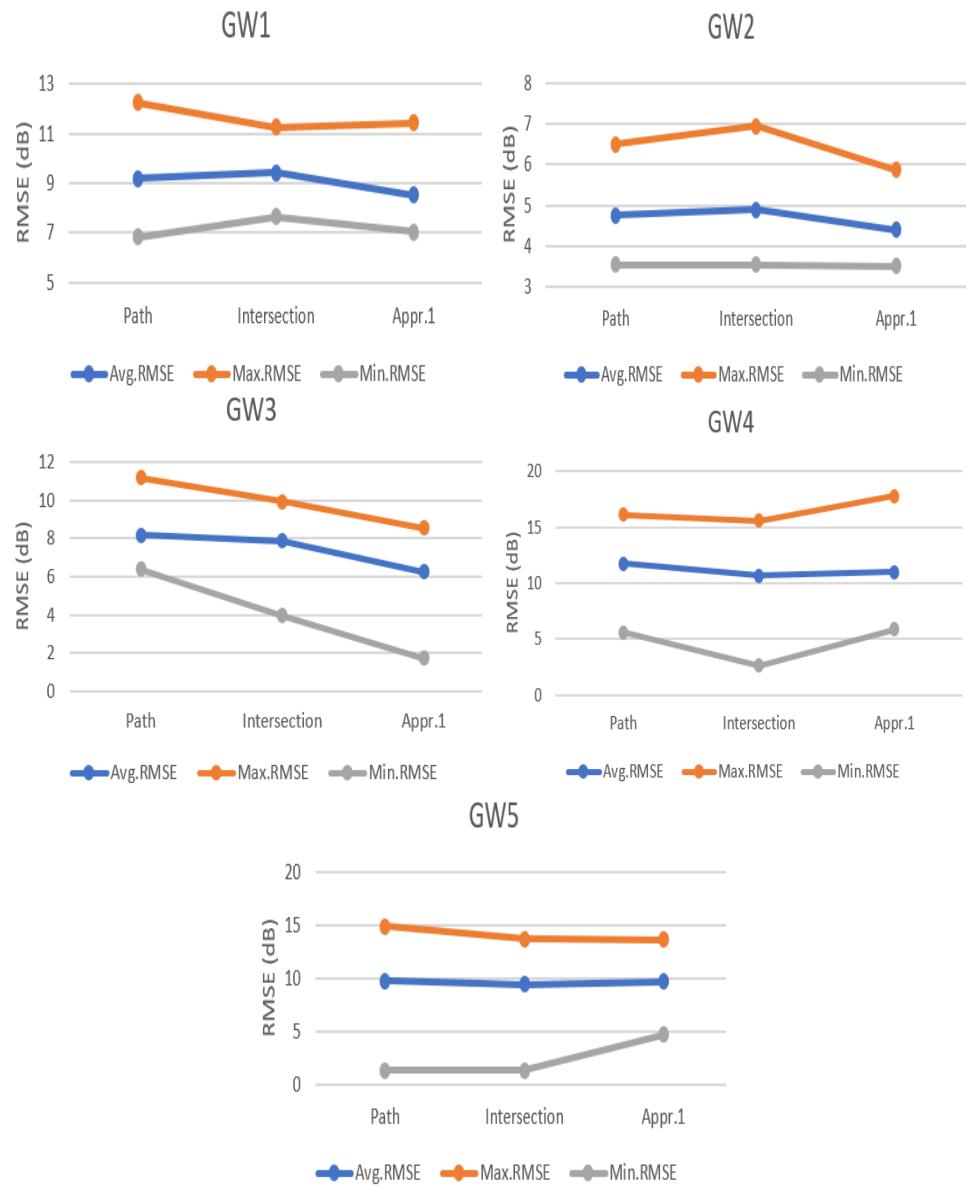


Figure 6.6: Comparison of Approach One with Path and Intersection Analysis

6.2 Approach Two

On closely analyzing points on the land cover map and cross-referencing the predicted land cover with Google Maps, many cases are observed where pixels are not classified correctly. As an example, three such cases are listed in Table 6.1. The corresponding Google map location of the pixels is shown in Figure 6.7. At coordinate location A, Google map points to the Aula building at the TU Delft campus. The land cover map wrongly classifies the pixel at this location as "Fields", whereas the obstacle height obtained from the digital elevation models gives a value of 14.20 meters. Similarly, the water body at location B is classified as "Trees" by the land cover map, whereas the height models provide a No-Data value. This is consistent with the fact that LiDar pulses do not get reflected off water bodies and therefore gets stored as a No-Data height value. These three examples are particularly highlighted as such cases of misclassification will have an adverse effect on the decision making the process for selecting the OH variant. Furthermore, the paper also mentions that the tool misclassifies "building" as "road" and vice-versa for certain pixels. The authors state that the cause is likely due to the similar spectral signature of the related materials used for the road and building. Confusing a cluttered component (e.g: Building) for a low-cluttered component (e.g: Bare soil), will reflect on the percentages and hence, affect the weighted average approach. Thus, Approach Two uses freely available remotely sensed surface heights, eliminating the need for land-cover maps to classify the type of environment.

	Coordinates	Google maps	Land Cover Map	DEM
A	52.002202, 4.373419	Building	Fields	14.20 meters
B	52.002962, 4.373823	Water	Tree	No-data
C	52.001367, 4.374419	Building	Bare soil	11.61 meters

Table 6.1: Examples illustrating wrongly classified pixels of the land cover map

We make use of LiDAR generated digital elevation models (AHN) as discussed in Section 3.1.2.1 and propose an alternate method to classify components in the cluttered and low-cluttered category.

The sequence of obstacle heights at every 5m spacing is obtained along the path between the transmitter and receiver. Obstacle heights can be obtained by subtracting the altitudes of the interpolated DTM from the DSM. If the obstacle height is above a threshold, the point is associated with the clutter category, whereas if the height lies below a threshold, it contributes to the low-cluttered category. Sources [10], [11] state that the height of greenhouses must be approximately 4 meters or taller. Keeping this

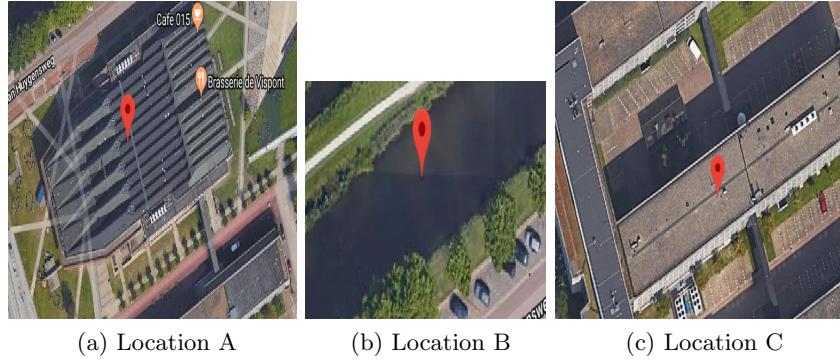


Figure 6.7: Screenshots of mis-classified locations on Google map

in mind and on further analyzing the heights of buildings from the DEMs, the threshold is set to 4 meters. Moreover, the results of the approach when the threshold is set lower than 4 meters did not do as well as when the threshold is set to 4 meters.

The Path analysis utilizes only two variants of the OH model i.e. the sub-urban and urban variants as discussed in Section 3.2.2. Since the height information is available using DEMs, the free/open/rural variant of the OH model is incorporated in Approach Two. This variant is applied when two conditions are met: Firstly, the path has no city-like built-up structures and secondly there is no obstruction to the line-of-sight between the transmitter and receiver. For the first condition, the algorithm checks if all the heights along the path remain below the threshold (low cluttered = 100%). This will indicate that the possibility of having a free space scenario is high. For the second condition, along the entire path at each point, the elevation angles from the transmitter to the obstacle and the transmitter to the receiver are used to analyze possible obstructions to the line-of-sight as shown in Figure 6.8. The scenario on the left indicates no obstruction, whereas the scenario on the right indicates an obstruction to the line-of-sight.

If the above two conditions are met, the free/open path loss variant is applied, otherwise, the program resorts to the weighted average between the sub-urban and urban variant losses.

The flow chart in Figure 6.9 illustrates the modified algorithm that replaces land cover maps with DEMs.

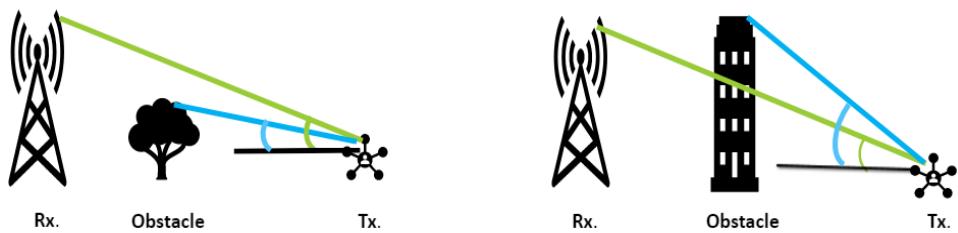


Figure 6.8: Using the elevation angles to check for obstruction to the line-of-sight

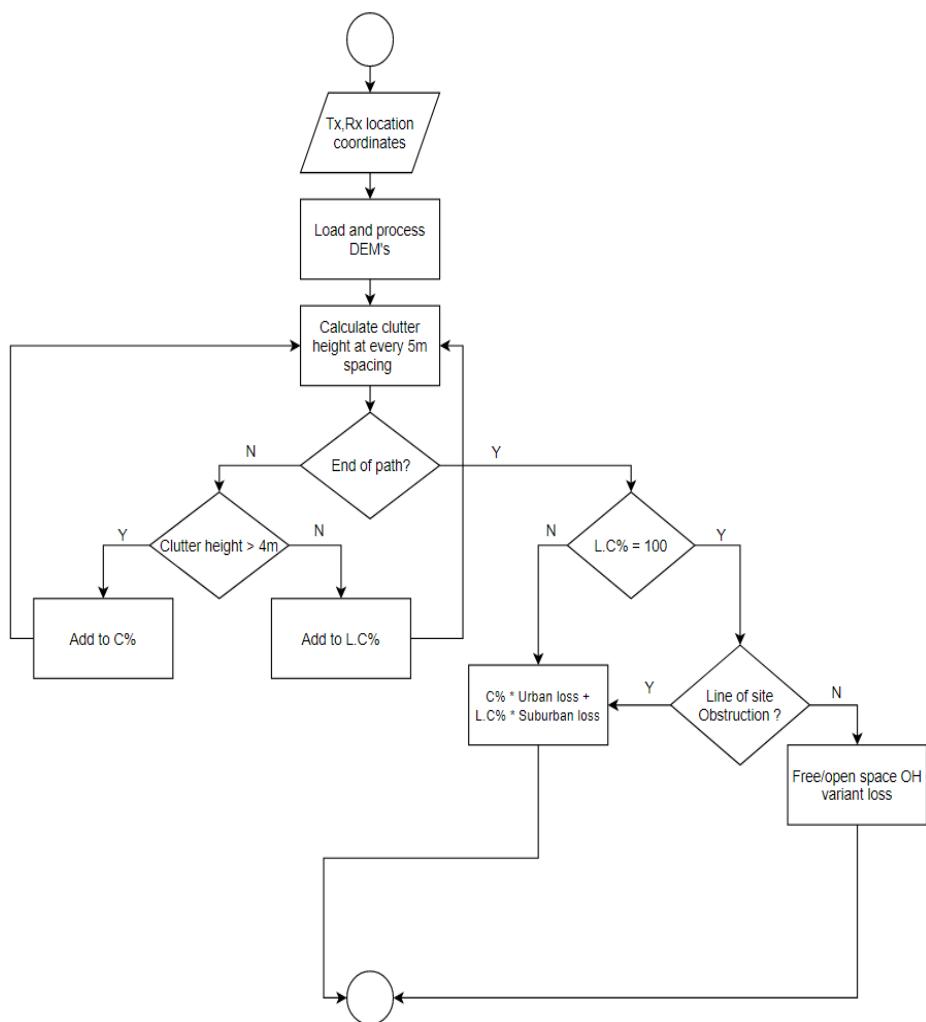
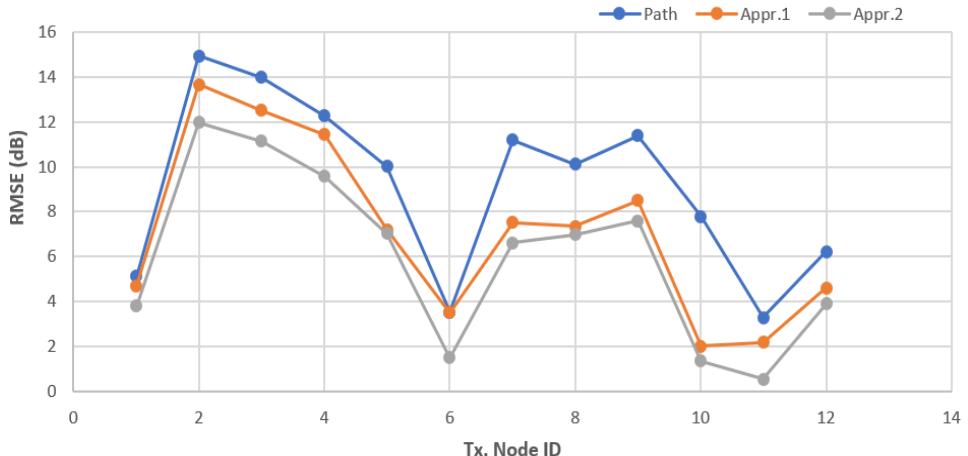


Figure 6.9: Algorithm for “Approach Two”

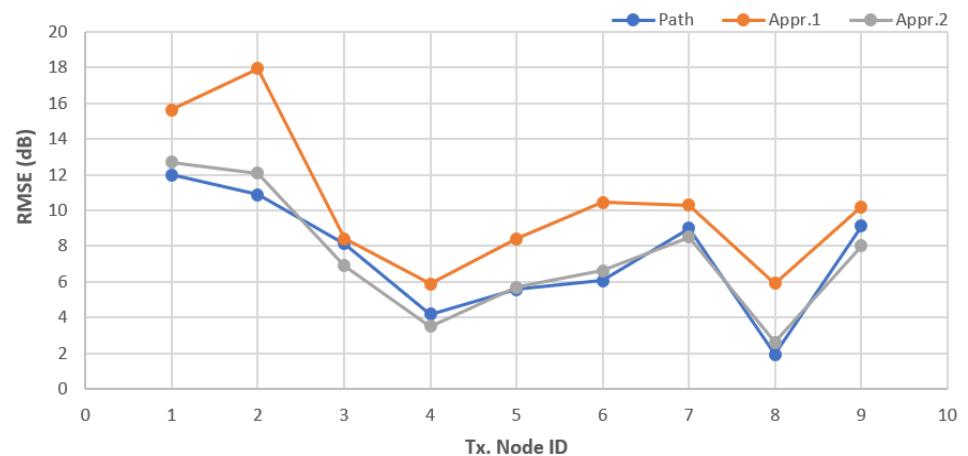
6.2.1 Approach Two: Results and Conclusion

6.2.1.1 Approach Two: Analysis

This section compares the path loss RMSE of Approach Two with the Path analysis and Approach One. The same set of nodes used to analyze the performance of Approach One in Section 6.1.1 is plotted here. The nodes in Figure 6.10a show a further reduction in the RMSE compared to Approach One. Also, Figure 6.10b shows the effect of Approach Two on the set of nodes that had performed poorly in Approach One. All the nodes show a reduction in the error compared to Approach One. Moreover, the majority of the nodes also do better than the Path analysis.



(a) Nodes showing further improvement with Approach Two



(b) Nodes showing improvement with Approach Two compared to Approach One

Figure 6.10: Comparison of Approach Two against Path analysis and Approach One

The main advantage of this approach is that the model no longer relies on the availability of land cover maps for different regions of the Netherlands. The method of replacing land cover maps with free and publicly available high-resolution height models eliminates the effort to undergo land cover classification for different regions. But in spite of the reduction in the RMSE shown by Approach Two, the improvement is limited. This is expected because compared to the Path analysis which chooses between the suburban and urban variants, the Approach Two now performs a weighted average between the same two variants.

6.2.1.2 Approach Two: Overall comparison

This section will give an overall comparison of the average, minimum and maximum RMSE of each gateway across all transmit nodes. The graphs in Figure 6.11 show an overall reduction in errors from Approach One to Approach Two.

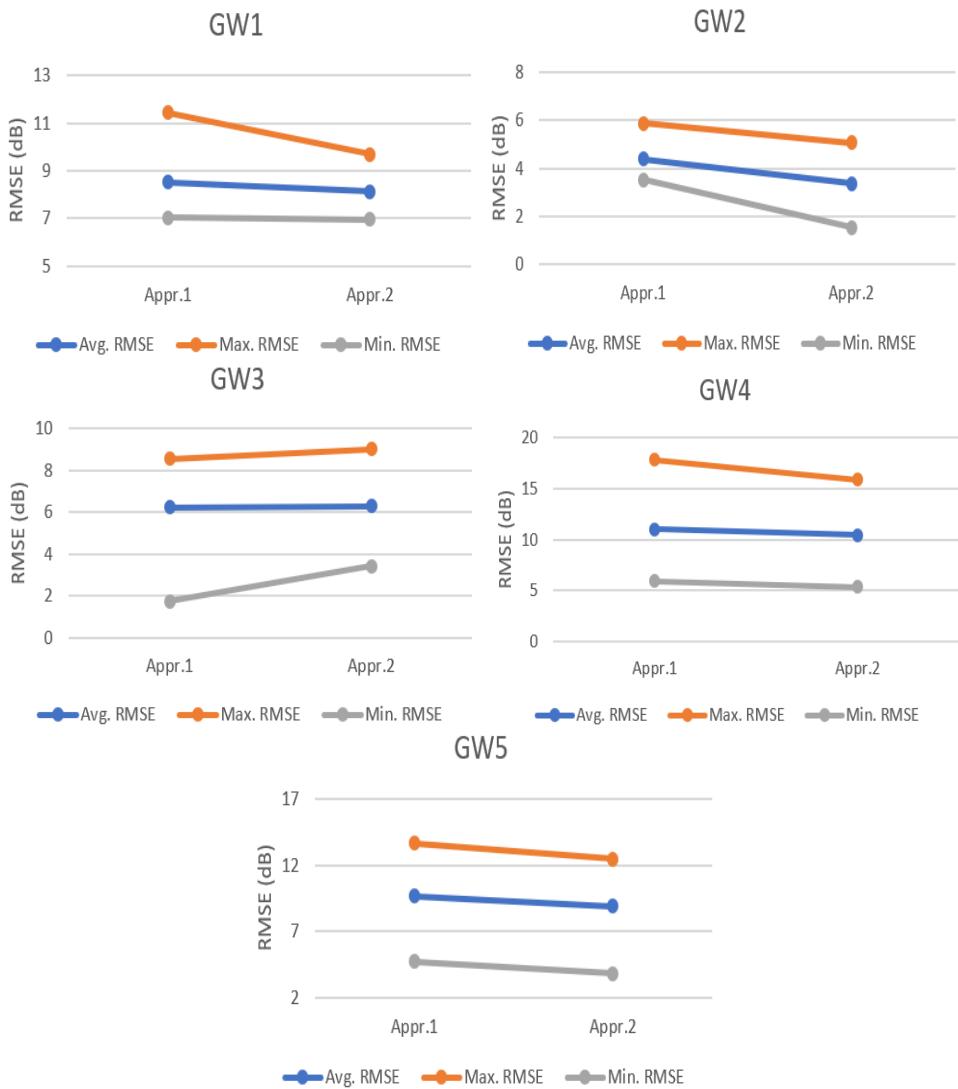


Figure 6.11: Comparison of Approach Two with Approach One

The graphs in Figure 6.12 compares the Approach Two with the Path and Intersection Analysis.

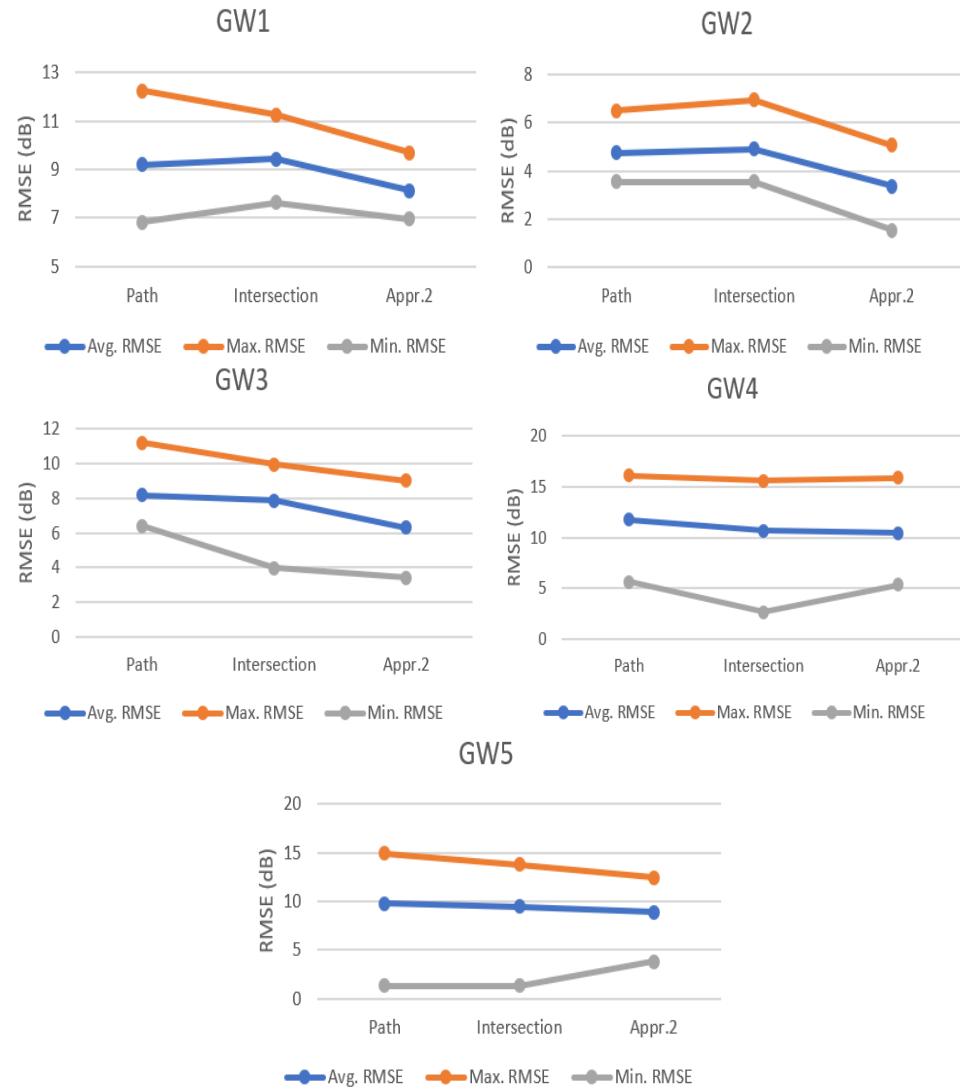


Figure 6.12: Comparison of Approach Two with the Path and Intersection Analysis

To validate the final approach further, the sample space is increased to include three more gateways (GW8 in Arnhem and GW6, GW7 in Utrecht). Figure 6.13 shows the result of the 3 gateways.

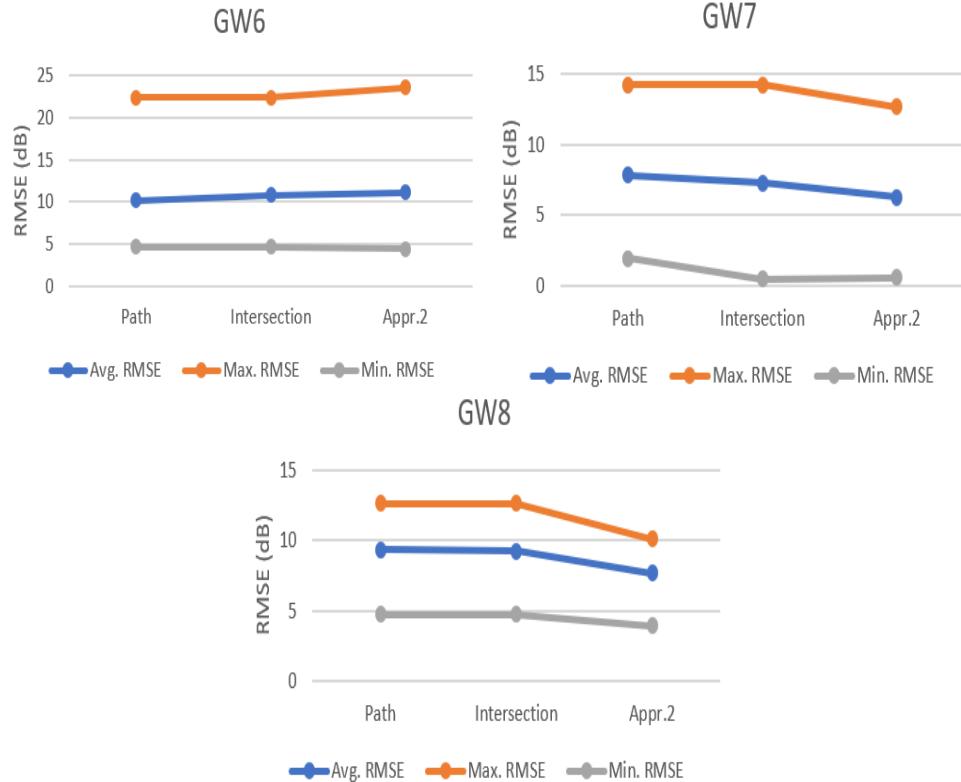


Figure 6.13: Comparison of Approach Two with the Path and Intersection Analysis for 3 more gateways

As shown in the Figures 6.12 and 6.13, overall the Approach Two shows improved performance compared to the Path and Intersection analysis except for GW6 giving higher errors in terms of the average and maximum RMSE. Also, GW4 and GW5 shows an increase in the minimum RMSE with Approach Two. These gateways (GW4, GW5, GW6) will be further looked into in Section 6.3.

Table 6.2 and Table 6.3 quantifies the errors of Approach Two with Path analysis and Approach Two with Intersection analysis respectively.

	Path analysis			Approach 2		
	(dB)			(dB)		
	Avg. RMSE	Max. RMSE	Min. RMSE	Avg. RMSE	Max. RMSE	Min. RMSE
GW1	9.19	12.25	6.84	8.12	9.68	6.95
GW2	4.75	6.51	3.54	3.37	5.07	1.52
GW3	8.17	11.18	6.40	6.29	9.02	3.43
GW4	11.76	16.08	5.62	10.45	15.88	5.34
GW5	9.79	14.93	1.33	8.89	12.47	3.80
GW6	10.19	22.37	4.70	11.11	23.56	4.41
GW7	7.81	14.24	1.93	6.25	12.69	0.56
GW8	9.38	12.64	4.71	7.70	10.12	3.91

Table 6.2: Comparison of Approach Two with Path Analysis

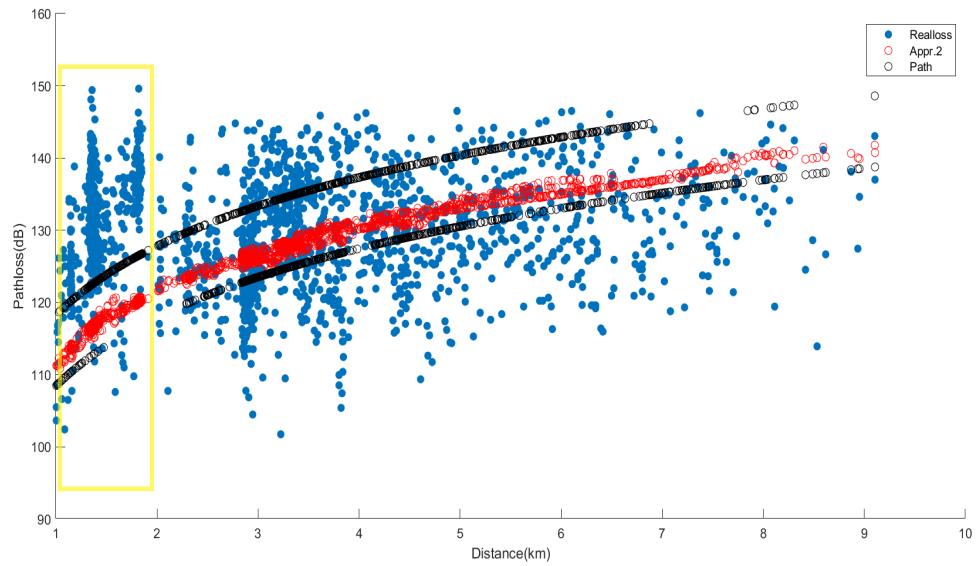
	Intersection analysis			Approach 2		
	(dB)			(dB)		
	Avg. RMSE	Max. RMSE	Min. RMSE	Avg. RMSE	Max. RMSE	Min. RMSE
GW1	9.43	11.25	7.64	8.12	9.68	6.95
GW2	4.90	6.95	3.54	3.37	5.07	1.52
GW3	7.87	9.95	3.97	6.29	9.02	3.43
GW4	10.66	15.60	2.65	10.45	15.88	5.34
GW5	9.43	13.75	1.38	8.89	12.47	3.80
GW6	10.79	22.37	4.70	11.11	23.56	4.41
GW7	7.26	14.24	0.46	6.25	12.69	0.56
GW8	9.27	12.64	4.71	7.70	10.12	3.91

Table 6.3: Comparison of Approach Two with Intersection Analysis

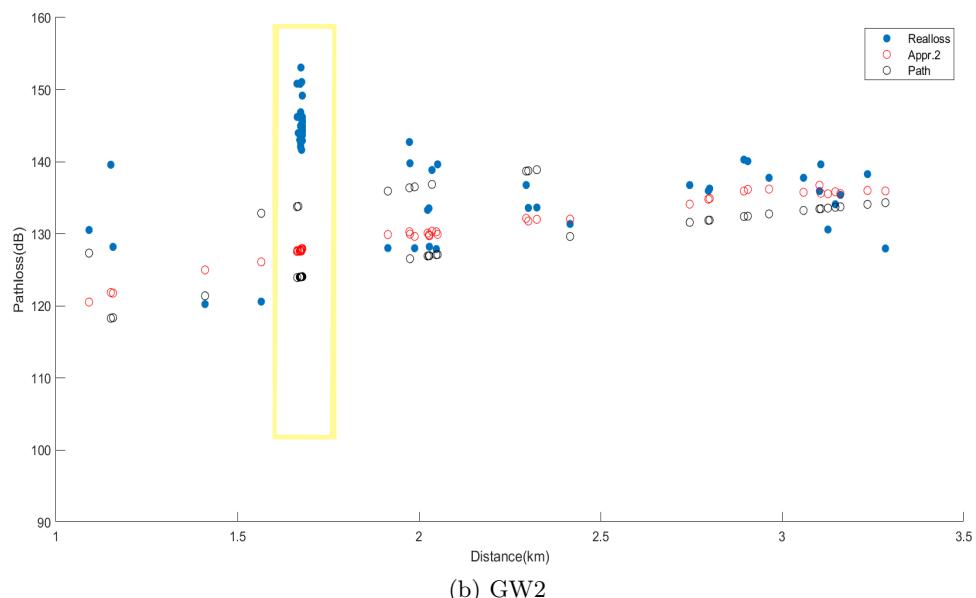
6.2.1.3 Effect of samples at very short distances (1-2 km)

The OH model is valid for link distances between 1-10 km only. The results for the overall comparison is shown for samples from 2-10 km, excluding samples between 1-2 km. This is done because for samples at short distances, the prediction of the OH model does not map well to the measured path loss. To illustrate this, the path loss versus distance graphs for GW1 and GW2 are shown in Figure 6.14. The graphs show the measured samples (Blue), the predicted loss of the Path analysis (Black) and the Approach Two (Red). For samples at very short distances from the gateway, the measured path loss is almost equivalent to or higher than the path loss measured at longer distances. Moreover, the measured path loss goes much higher than the path loss curve of the urban variant of the OH model and does not map well to the measured losses for samples within 1-2 km. This unusual spike in the path loss for short distances could be the result of nodes being indoors.

The Figure 6.15 illustrates the effect on the maximum RMSE for Approach Two, Path and Intersection Analysis with samples from 1-10 km compared to with samples from 2-10 km. Only the effect on the maximum RMSE is shown as the number of samples within the 1-2 km only form a small part of the overall samples for most gateways and hence, its effects can be witnessed in the change in maximum RMSE compared to that of the average RMSE. In some cases, the modified algorithm does not work well for samples between 1-2 km. Hence, the thesis performs all the evaluation for samples from 2-10 km range.



(a) GW1



(b) GW2

Figure 6.14: Plot showing a spike in the path loss for distances between 1km-2km

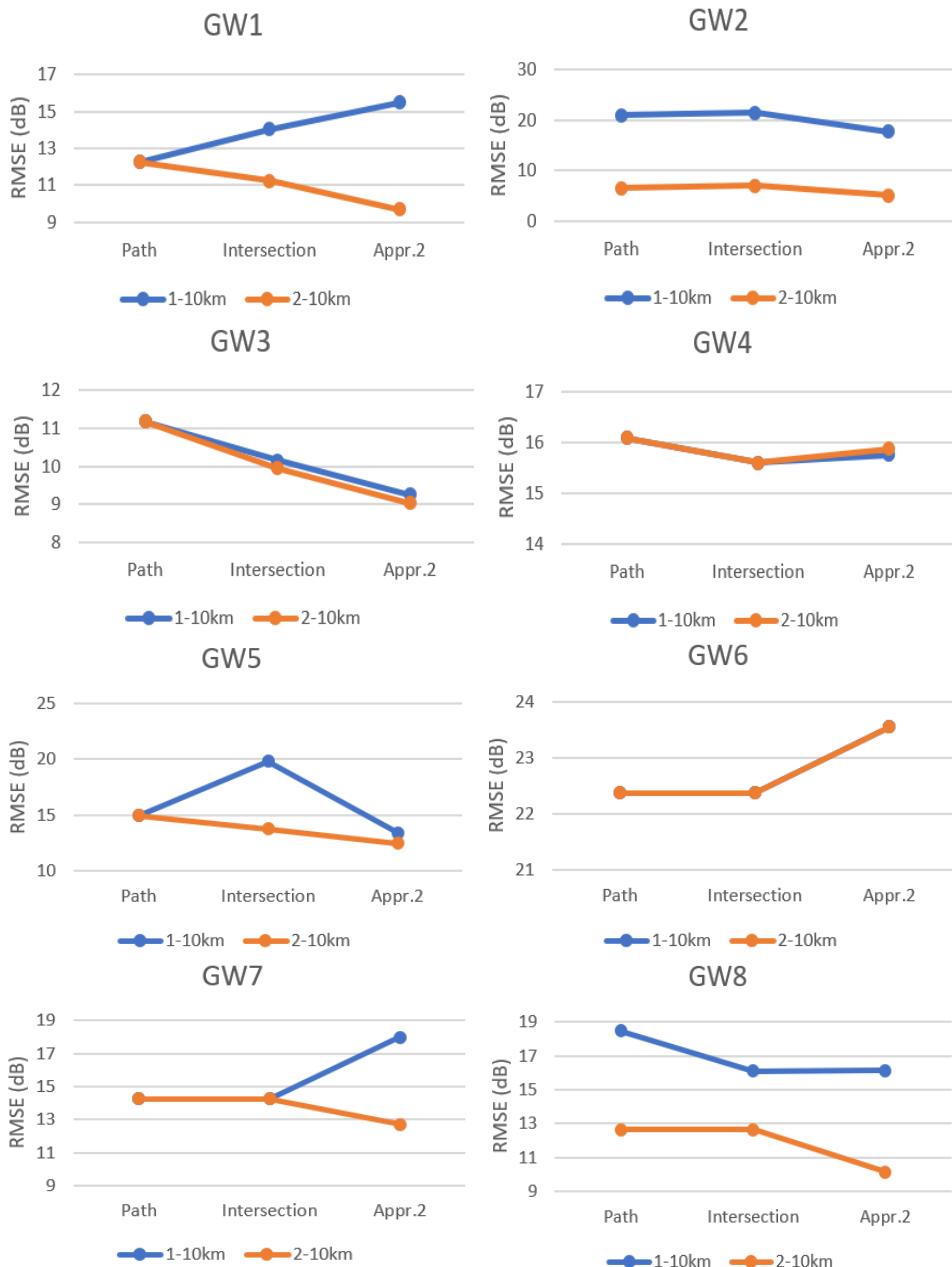


Figure 6.15: Comparison for Path, Intersection and Approach Two with and without samples from 1-2 km

6.3 Approach Three

To further investigate the gateways (GW4, GW5, GW6), Figure 6.16, Figure 6.18 and Figure 6.17 displays the real path loss against the predicted losses obtained from the Path analysis and the Approach Two for the gateways. The path loss curve of the pure rural/free-space OH variant is also included in the graphs.

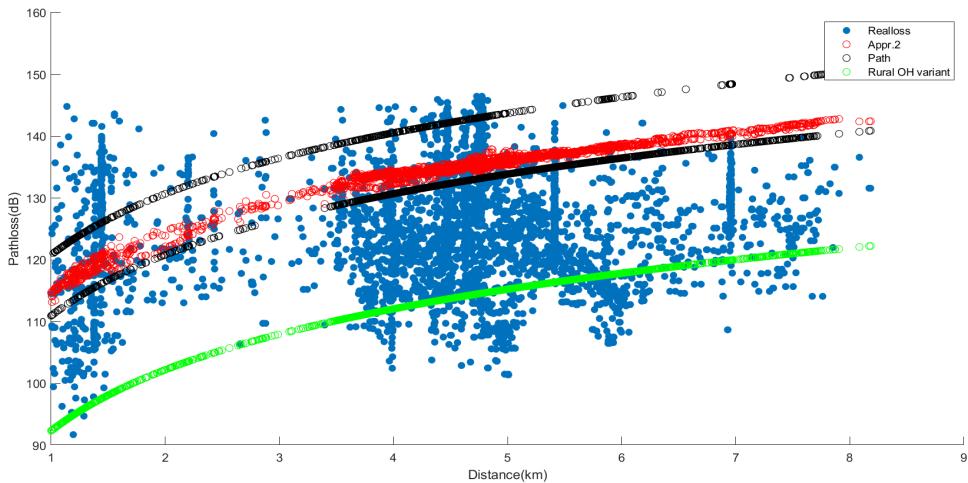


Figure 6.16: Pathloss(dB) versus Distance(km) plots for GW4

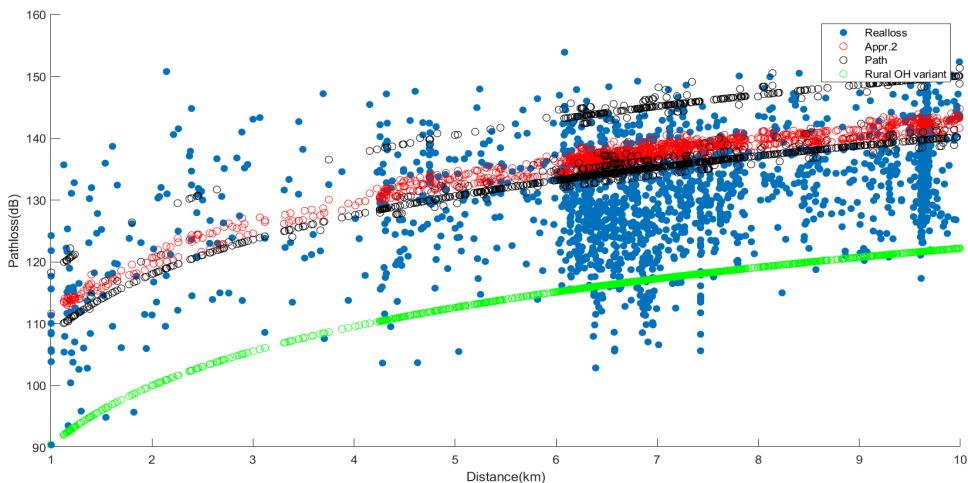


Figure 6.17: Pathloss(dB) versus Distance(km) plots for GW6

The black lines in the graphs represent the Path analysis losses that shift among the two variants (urban and sub-urban), while the red lines represent the Approach Two's losses that vary within the limits set by the two variants. We notice that a big portion of the real path loss samples lies further

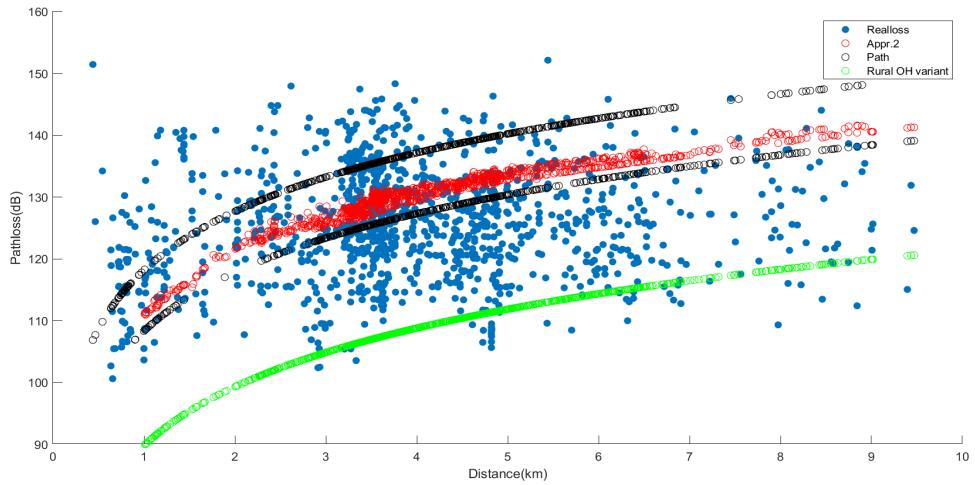


Figure 6.18: Pathloss(dB) versus Distance(km) plots for GW5

below the lower limit (suburban variant). The weighted average approach between the suburban and urban variant path loss did show an improvement compared to the Path and Intersection analysis for the gateways as seen in Figure 6.15. But this improvement is limited as the region within which the loss can vary is limited between the path loss given by the two variants. In other words, the plots show that the Approach Two is a more pessimistic approach for the three gateways. To provide more room for the path loss to vary, the range for the weighted averages is increased to include paths that can lie between a suburban and rural/open/free space scenario (represented by the green curve in the graphs).

The free space OH variant in Approach Two was considered only when there was a certainty that the propagation path has only low cluttered components with no line-of-site obstruction. To increase the range of the weighted averages, setting the lower extreme limit for the weighted averages as the free space OH variant path loss will underestimate the loss to a great extent. If a path has a higher percentage of low cluttered components ($L.C\% > C\%$), a weighted average between the free space and urban variants extremes will give a very low path loss value. This tends to represent a path close to open/free space in spite of some clutter presence in the path. Hence, the lower limit for the averages is set to a value between that of the suburban and free-space variant limits.

Considering that for a particular propagation setting, the difference in path loss predicted by the urban and suburban variant is around 10 dB and between the suburban and free-space variant is around 20dB, the lower limit can be set to a value between the path loss represented by the suburban and free-space variant. Approach Three calculates the weighted average between the urban variant and the path loss between a suburban and free

space-variant as shown in Equation 6.2.

$$PathLoss(dB) = C\% \times L_{urb} + L.C\% \times (L_{sub-urb} - L_{free})/2 \quad (6.2)$$

where:

L_{urb} : Pathloss using the urban OH variant (dB)

$L_{sub-urb}$: Pathloss using the sub-urban OH variant (dB)

L_{free} : Pathloss using the free/open space OH variant (dB)

C% : Cluttered components percentage

L.C% : Low-cluttered components percentage

The Table 6.4 compares the performance of the Path analysis, Approach Two and Approach Three in terms of the average RMSE and the maximum RMSE observed across all nodes for each gateway. The cells marked in green show the best performance (lowest errors).

	Path analysis		Approach 2		Approach 3	
	(dB)		(dB)		(dB)	
	Avg. RMSE	Max. RMSE	Avg. RMSE	Max. RMSE	Avg. RMSE	Max. RMSE
GW1	9.19	12.25	8.12	9.68	10.44	15.90
GW2	4.75	6.51	3.37	5.07	8.40	10.67
GW3	8.17	11.18	6.29	9.02	9.03	14.31
GW4	11.76	16.08	10.45	15.88	10.08	15.01
GW5	9.79	14.93	8.89	12.47	9.97	16.70
GW6	10.19	22.37	11.11	23.56	8.34	17.42
GW7	7.81	14.24	6.25	12.69	8.51	19.86
GW8	9.38	12.64	7.70	10.12	13.86	17.18

Table 6.4: Comparison of Approach Three with Path Analysis and Approach Two

In general, the Approach Three does not show any improvement in the average and maximum errors except for GW4 and GW6. Moreover, GW5 does not do better with Approach Three and the improvement shown by GW4 using Approach Three is not significant compared to that of Approach Two. Observing the graphs of the path loss versus distance in Figure 6.16, Figure 6.18 and Figure 6.17 for GW4, GW5, and GW6, the expectation was to have a more significant improvement in the performance on increasing the range for the weighted average approach. This was expected because majority of the samples have losses lower than the limits set for the weighted averages in Approach Two.

Figure 6.19 and Figure 6.20 plots the RMSE of the path loss for individual nodes for these gateways (GW4, GW5, GW6). Due to the high number of transmitting nodes for GW6, the graph is split into two: Figure 6.20a includes the nodes that show a reduction in the RMSE with Approach Three and Figure 6.20b includes nodes that show worse performance with Approach Three.

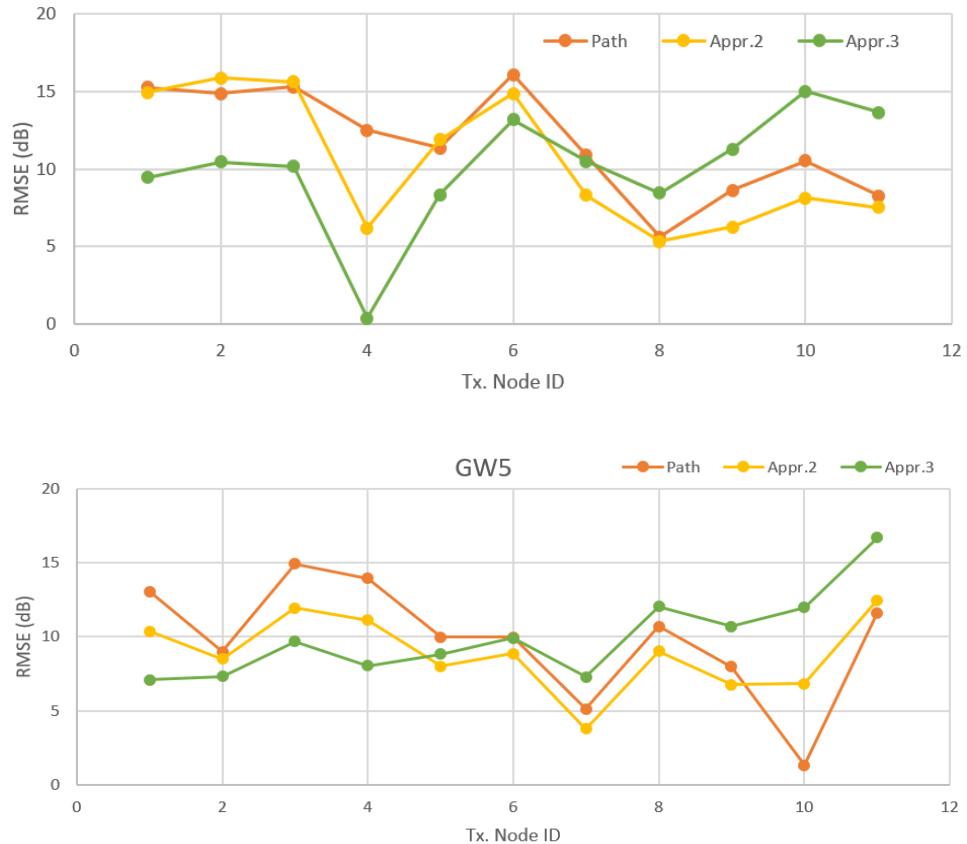
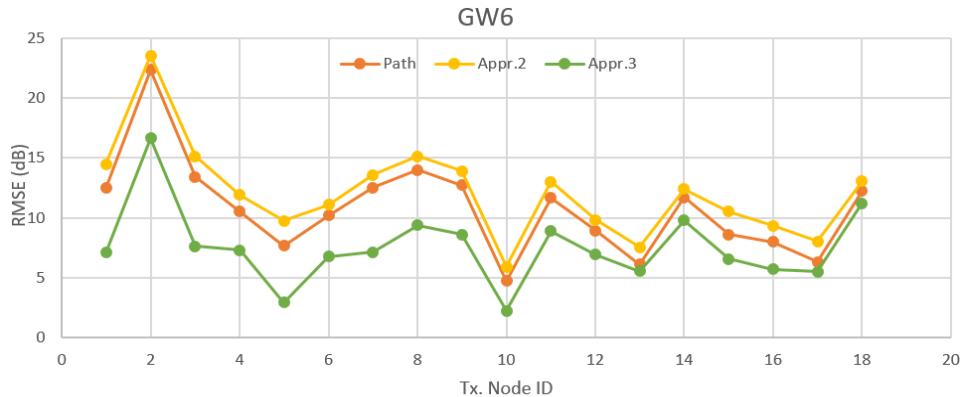


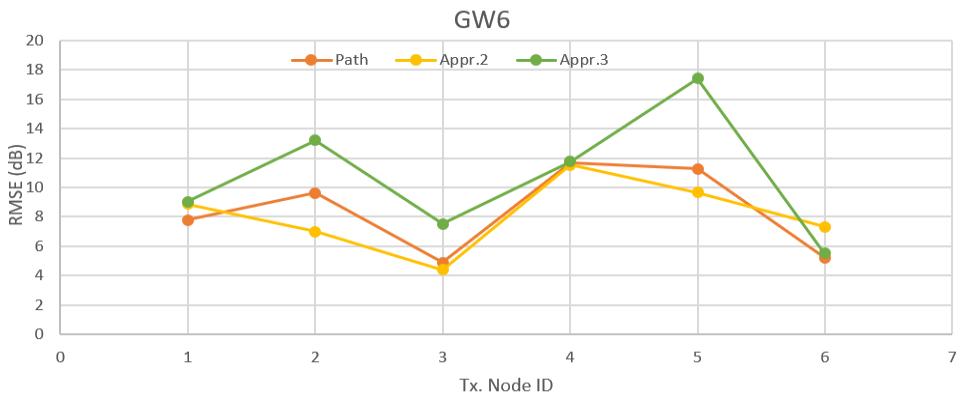
Figure 6.19: Plot of RMSE(dB) of each user node for GW4 and GW5

On closer analysis of the performance of the individual nodes for the gateways, we observe that the Approach Three does show a significant improvement for around half of the nodes for GW4 and GW5 and 75% of the nodes in GW6. While the reduction in errors of Approach Three is significantly higher compared to that of the Approach Two for the first half of the nodes in GW4 and GW5, the rest of the nodes shoot up the errors leading to an increase in the overall maximum and average RMSE. The significant improvement in some nodes gets hidden while computing the overall performance of the approach for the gateways.

With Approach Three there is a possibility of having much lower path losses than what the Approach Two predicts since the range for the weighted



(a) Nodes showing improved performance for GW6



(b) Nodes showing increased errors for GW6

Figure 6.20: Plot of RMSE of each user node for GW6

averages is increased by setting the lower limit further below the sub-urban variant. While these lower predicted path losses tend to improve the performance significantly for some nodes, the errors shoot up for samples that have measured path losses that are much higher.

There could be two reasons for this varied performance of Approach Three among different nodes.

One of the reasons could be the possibility of having some nodes inside buildings which results in a low received signal power (or high measured path loss). The thesis tried to eliminate this issue by considering only nodes with height less than and equal to 2 meters under the assumption that as the nodes go higher the possibility of them being indoors increases. The testing was carried out in a similar manner in the paper [41]. But this does not ensure that all nodes within 2 meters high will be used outdoors. So samples from nodes that lie on a very low cluttered path but was transmitted from indoors, will result in a higher path loss than what the model predicts.

Such samples will show a better performance with Approach Two compared to Approach Three as the losses vary between the higher limits set by the urban and sub-urban variants.

To verify this claim, shapefiles for each node is created, imported and viewed with QGIS to get an idea of the route taken by the nodes. Shapefile is a common storage format for locations and other attributes of geographic features for use in GIS (geographic information systems). Some nodes as shown in Figure 6.21 show a clear indication of nodes being outdoors and forming an obvious route. On the other hand, some nodes in Figure 6.22 show samples that form a cluster as though they could be transmitted from within a building. Moreover, some nodes show samples that are scattered and located far from each other and do not form an obvious route. These samples could have been transmitted during different times of the year. Also, many nodes show a mixture of these patterns, hence making it difficult to assign a node as being indoors or outdoors. The Figure 6.22 shows the Google-map location of the clusters spotted in the shapefiles.

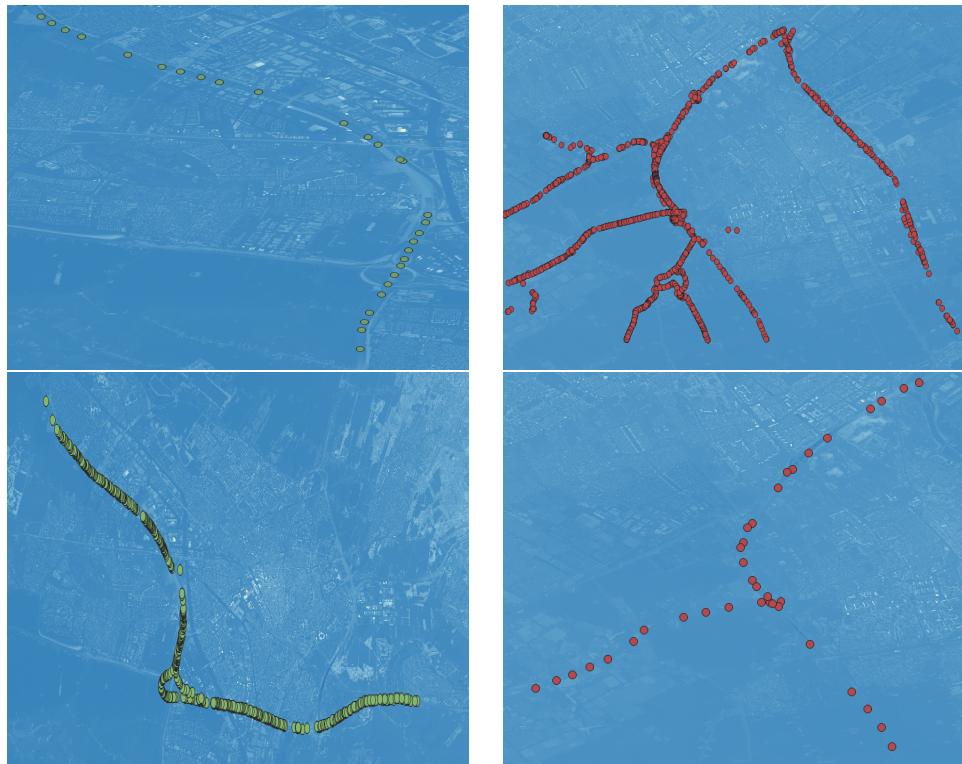


Figure 6.21: Shape files of nodes with a higher probability of being outdoors as they follow a clear path (e.g. roads)

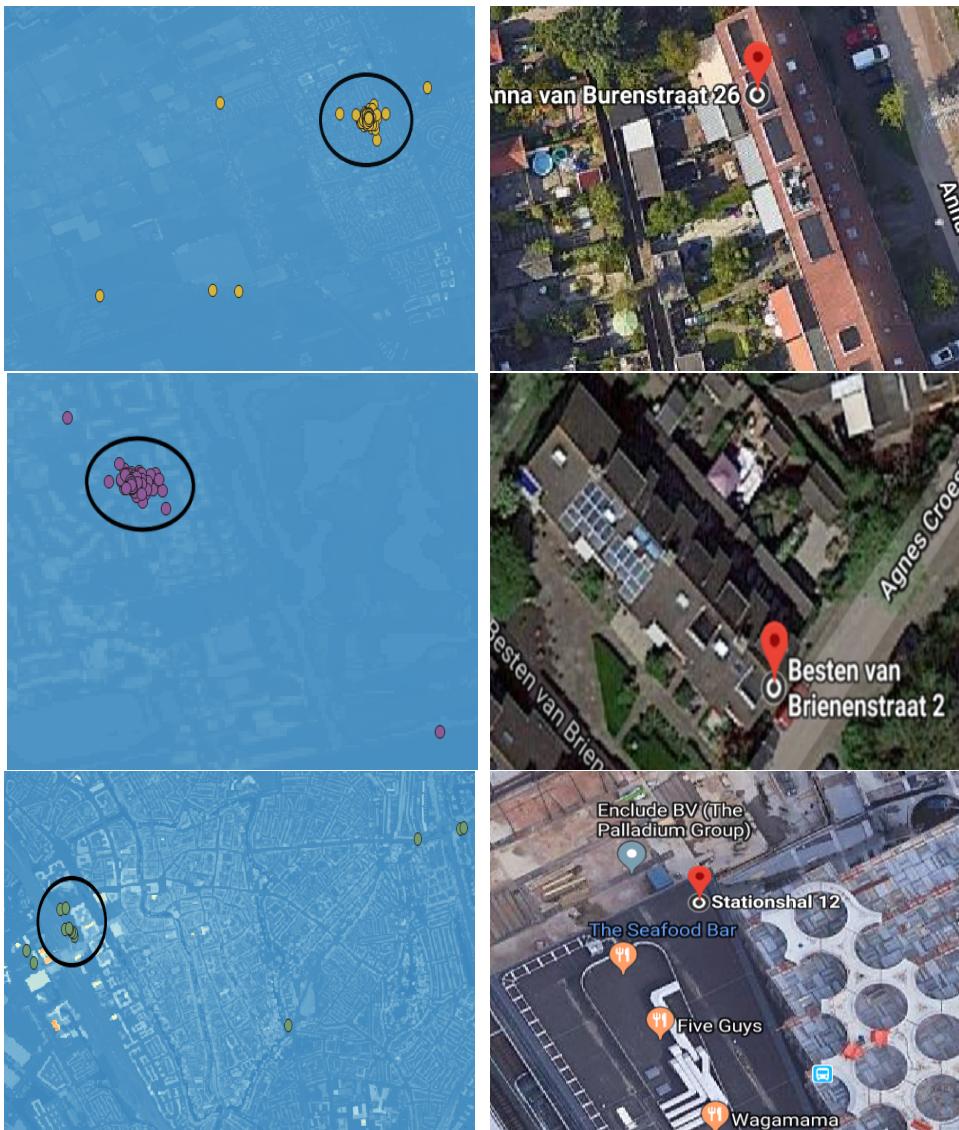


Figure 6.22: Shape files of nodes with a higher probability of being indoors

Moreover, the graphs in Figure 6.16 and Figure 6.18 can contribute to the reasoning of the possibility of having nodes indoors. Ideally, the path loss should increase with distance, showing a curve similar to that of the Path analysis with the OH model. But the graphs show a spike in the path loss for samples at short distances which then tends to decrease with an increase in the distance. This trend in the path loss for shorter paths could be the effect of having nodes indoors.

The second reason behind the varied performance of Approach Three for a set of nodes could be the assumption made about the transmit power. Since we cannot obtain any information about the transmit power for the nodes,

similar to the paper [41] we assume a transmit power set to the maximum limit of 14 dbm in Europe. This will give an upper bound to the path loss and hence would lead to an overestimation of the measured path loss for a set of nodes. Therefore, as an example, as shown in Figure 6.23, assume a packet is transmitted with a transmit power of +4 dbm, but the comparison is done with an assumption for the transmit power as +14 dbm. The direct comparison of the predicted path loss of Approach Two with the assumed measured path loss will give better results than Approach Three, in spite of Approach Three being closer to the actual measured path loss. Hence directly comparing the true value of the losses in this case will not be the right approach to verify whether the Path analysis or the weighted average approaches perform better.

Since we cannot do a direct comparison of the predicted loss with the real loss, it is difficult to estimate the suitable range for the weighted averages to vary between i.e, the Approach Two or Approach Three. The analysis using Approach Three is mainly a means to state that while the Path analysis and Approach Two limits its evaluation between the urban and suburban variants, the possibility of utilizing the other variants of the OH model must not be completely excluded due to the assumptions made regarding the transmitter settings and the indoor nodes.

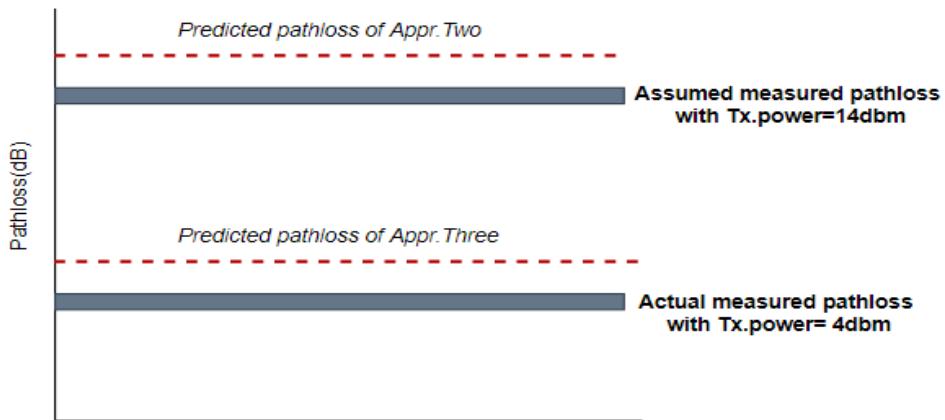


Figure 6.23: Illustration of the issue with direct comparison of pathloss while assuming the Tx.power of +14dbm

6.4 Relative comparison of the approaches

To avoid the issue of the possibility of having indoor nodes and having unknown transmit power and antenna gains, a relative comparison can instead be adopted. Here the difference in the predicted path loss of a particular

approach is compared to the difference in the received signal powers. This comparison is done for every combination of samples of one node and gateway pair. Equation 6.3 and Equation 6.4 represent the link budget equation for two samples from the same node sent from two locations with variables being P_{rx1} , P_{rx2} , PL_1 and PL_2 . The dependency on common terms such as the transmit power and antenna gains can be eliminated by subtracting the two equations to give Equation 6.5.

The relative comparison will be able to track the changes in the predicted pathloss of the models and compare them to the trend in the real received power. To some extent, the relative comparison between samples of the same node and same gateway will cancel out the dependency of antenna gains and transmit powers. This is assuming the best case scenario that the same user node will have the same antenna and will transmit packets at the same power for all the samples. Moreover, it will also cancel out the dependency of additional loss from buildings due to samples sent from an indoor node to some extent. This is assuming that all the samples from one node were transmitted from indoors. This kind of comparison is one way to analyze the performance of the approaches but it will not be able to detect if the absolute values of the predictions are comparable to the real measurements.

$$P_{tx} = P_{rx1} + G_{tx} + G_{rx} - PL_1 - L_{indoor} \quad (6.3)$$

$$P_{tx} = P_{rx2} + G_{tx} + G_{rx} - PL_2 - L_{indoor} \quad (6.4)$$

$$\Delta P_{rx} = \Delta PL \quad (6.5)$$

where:

P_{tx} : Transmit power (dBm)

P_{rx1}, P_{rx2} : Received power at location 1 and 2 (dBm)

PL_1, PL_2 : Path loss at location 1 and 2 (dB)

G_{tx}, G_{rx} : Transmit and receiver antenna gains

L_{indoor} : Additional loss due to indoor nodes (dB)

This can be used as a metric for comparing the trend captured by the binary decision making approach using land cover map of the Path and Intersection analysis of the paper [41] versus the trend captured by the weighted average approaches using digital elevation models.

The Table 6.5 compares the percentage change in the average, maximum and minimum RMSE of the approaches from the Path analysis. The cells marked in green indicate the better performing approach for the gateway. The more negative is the percentage, the higher is the reduction in errors.

	% Improvement of Approach Two			% Improvement of Approach Three		
	(From Path analysis)			(From Path analysis)		
	Max. RMSE	Avg. RMSE	Min. RMSE	Max. RMSE	Avg. RMSE	Min. RMSE
GW1	-17.8	-14.4	-9.5	-17.4	-15	-10.9
GW2	-38.9	-25.8	-10.5	-41.9	-28	-15
GW3	-5.8	-18.4	-0.1	1.8	-15.1	0.2
GW4	-11.8	-5.4	-10.9	-12.5	-5.7	-8.5
GW5	-18.7	-12.9	-5.5	-19.8	-14.1	-6.5
GW6	2.2	-9.5	-60	4.2	-8.6	-45.9
GW7	-11.3	-18.9	-36.4	-12.7	-20.7	-93.1
GW8	-1.2	-3.1	-0.8	-3.9	-6.2	-3.3

Table 6.5: Comparison of the % improvement of Approach Two Approach Three from Path Analysis

Approach Two shows an improvement from the Path analysis of up to 25.8% for the average RMSE, 38.9% for the maximum RMSE and 60% for the minimum RMSE. While Approach Three shows an improvement from the Path analysis of up to 28% for the average RMSE, 41.9% for the maximum RMSE and 93.1% for the minimum RMSE.

The Table 6.6 compares the percentage change in the average, maximum and minimum RMSE of the approaches from the Intersection analysis. The cells marked in green indicate the better performing approach for the gateway. The more negative is the percentage, the higher is the reduction in errors.

Compared to the Intersection analysis, Approach Two gives an improvement of up to 25% for the average RMSE, 48.9% for the maximum RMSE and 60% for the minimum RMSE. Similarly, the Approach Three gives an improvement of up to 27.3%, 49.7% for the maximum RMSE and 45.9% for the minimum RMSE.

We notice the majority of the cells in Approach Three is marked green, hence showing a better improvement compared to Approach Two from both the Path and Intersection analysis.

	% Improvement of Approach Two			% Improvement of Approach Three		
	(From Intersection analysis)			(From Intersection analysis)		
	Max. RMSE	Avg. RMSE	Min. RMSE	Max. RMSE	Avg. RMSE	Min. RMSE
GW1	-30.7	-18.6	-4.3	-30.4	-19.2	-5.7
GW2	-37.6	-25.1	-10.5	-40.7	-27.3	-15
GW3	-1.6	-18.8	0.1	6.3	-15.5	0.2
GW4	-4.1	-3.6	5	-4.9	-3.9	7.8
GW5	-48.9	-19.2	48.4	-49.7	-20.3	46.9
GW6	2.2	-9.3	-60	4.2	-8.4	-45.9
GW7	-8.2	-8.2	1887.3	-9.7	-10.1	117.6
GW8	3.2	-3.4	-0.8	0.4	-6.5	-3.3

Table 6.6: Comparison of the % improvement of Approach Two Approach Three from Intersection Analysis

Chapter 7

Conclusions and Future Work

7.1 Conclusions

This thesis provides an analysis of the automated approaches used to estimate the coverage of LoRa using remotely sensing methods. First, the thesis delves into analyzing the suitability of the semi-deterministic channel model ITU-R 1812, largely neglected by LoRa literature and compares its performance with the Path analysis using the empirical OH model. The motive was to understand if a more complex and rich model like the ITU-R 1812 can fulfill the objective of estimating accurately LoRa's coverage or if we can use simpler empirical channel models by effectively utilizing their capabilities to provide more accurate predictions. We observe that the ITU model does not show any significant improvement compared to the OH model in spite of the extensive input information it requires. The accuracy of path loss predictions does not seem to increase by using a more complex semi-deterministic model. We also notice that the ITU model heavily depends on the accuracy of its inputs e.g. heights of the nodes and gateways for the diffraction model.

The thesis then delves into the assessment of the state-of-the-art that uses a combination of the land cover map obtained from remote sensing satellite images and the Okumura Hata model for automating LoRa's coverage estimation. The thesis introduces the concept of weighted averages between the OH variants (urban and sub-urban) to factor in the degree of urbanization along a path (Approach One). The approach no longer depends on a binary decision to select the type of variant like the Path analysis and also deviates from the traditional method of using the OH model. The thesis further tries to improve performance by replacing the use of land cover maps with more accurate digital elevation models (Approach Two). The approach uses heights to understand the clutter composition along a path. Furthermore,

the logic to select the third OH variant (rural/free-space) using height data is introduced in the algorithm to account for paths with only low cluttered components and no line-of-sight obstruction.

The approach is further modified to investigate the performance on increasing the limits for the weighted averages to include paths that could have a very low degree of urbanization and represent scenarios between that of the sub-urban and free-space environment (Approach Three). The analysis using this approach is done as a means to state that while the Path analysis and Approach Two limits its evaluation between the urban and sub-urban variants, the possibility of utilizing the other variants of the OH model must not be completely excluded due to the assumptions made regarding the transmitter settings and the indoor nodes.

The thesis further resorts to a relative comparison to eliminate the dependencies on transmitter settings and the placement of the nodes. In conclusion, the thesis finds that the Approach Three improves the performance from the state-of-the-art approaches and the ITU model and can better track the changes in the received signal power. Moreover, the evaluation in the thesis is carried out for link distances from 2-10 km.

7.2 Future Work

LoRa is proprietary and relatively new wireless communication technology. Hence, not much work has been implemented in automatically predicting it's coverage. During the study and analysis of this thesis, many potential topics have come across to improve the accuracy and performance of the thesis work. The following represents some possible future work:

- Exploration of other semi-deterministic models for LoRa.
- Combining the land cover and digital elevation model data to gain more insight into the obstacles in the propagation path i.e, for instance including the excess loss due to foliage.
- The data in this thesis is limited to the Netherlands. Further evaluation and verification of the modified approach with other areas can be implemented using data from controlled measurement campaigns.
- Using machine learning techniques to automatically understand the type of environment classification to predict the attenuation.

Bibliography

- [1] <http://charim.net/datamanagement/32>. Accessed:2019-08-12.
- [2] https://ssl.jspacesystems.or.jp/ersdac/GDEM/ver2Validation/Summary_GDEM2_validation_report_final.pdf. Accessed:2019-07-23.
- [3] https://www.eorc.jaxa.jp/ALOS/en/aw3d/index_e.htm. Accessed:2019-07-23.
- [4] <https://www.thethingsnetwork.org/>. Accessed:2019-07-23.
- [5] <https://www.thethingsnetwork.org/docs/applications/ttnmapper/>. Accessed:2019-07-23.
- [6] <https://qgis.org/en/site/>. Accessed:2019-07-23.
- [7] <https://en.wikipedia.org/wiki/ITU-R>. Accessed:2019-07-23.
- [8] <https://www.itu.int/en/ITU-R/study-groups/rsg3/Pages/iono-tropo-spheric.aspx>. Accessed:2019-07-23.
- [9] https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.1812-4-201507-I!!PDF-E.pdf. Accessed:2019-07-23.
- [10] <https://www.dpi.nsw.gov.au/agriculture/horticulture/greenhouse-structures-and-technology/height>. Accessed:2019-07-23.
- [11] <https://www.greenhousemag.com/article/choose-the-right-greenhouse-gutter-height/>. Accessed:2019-07-23.
- [12] Ahn. <https://downloads.pdok.nl/ahn3-downloadpage/>. Accessed:2019-07-23.
- [13] Atdi | advanced radiocommunications. <http://www.atdi.com/>. Accessed:2019-07-23.
- [14] Atoll radio planning software overview (Rf planning and optimisation) | forsk. <https://www.forsk.com/atoll-overview>. Accessed:2019-07-23.
- [15] Fundamentals of remote sensing. https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/earthsciences/pdf/resource/tutor/fundam/pdf/fundamentals_e.pdf. Accessed:2019-07-24.
- [16] Ground breaking world record! LoRaWAN packet received at 702 km (436 miles) distance. <https://www.thethingsnetwork.org/article/ground-breaking-world-record-lorawan-packet-received-at-702-km-436-miles-distance>. Accessed:2019-07-23.
- [17] Home. <https://sodaq.com/>. Accessed:2019-07-23.
- [18] Interactives climate change: vital signs of the planet. <https://climate.nasa.gov/interactives/climate-time-machine>. Accessed: 2019-07-23.
- [19] The IoT and climate change | Interact. <https://www.interact-lighting.com/global/iot-insights/the-iot-and-climate-change>. Accessed: 2019-07-23.
- [20] Lmr planner by rla communications engineering llc. <http://lmrplanner.com/>. Accessed:2019-07-23.

- [21] Lora. <https://en.wikipedia.org/w/index.php?title=LoRa&oldid=906893249>. Accessed:2019-07-23.
- [22] Lora applications | semtech lora technology | semtech. <https://www.semtech.com/lora/lora-applications>. Accessed:2019-07-23.
- [23] Lora desgin guide. https://www.semtech.com/uploads/documents/LoraDesignGuide_STD.pdf. Accessed:2019-07-23.
- [24] Lora modulation basics. <https://www.semtech.com/uploads/documents/an1200.22.pdf>. Accessed:2019-07-23.
- [25] LoRa LoRa documentation. <https://lora.readthedocs.io/en/latest/>. Accessed:2019-07-23.
- [26] LPWAN emerging as fastest growing IoT communication technology 1.1 billion IoT connections expected by 2023, LoRa and NB-IoT the current market leaders - IoT Analytics. <https://iot-analytics.com/lpwlan-market-report-2018-2023-new-report/>. Accessed:2019-07-23.
- [27] Network planning software. <https://www.lstelcom.com/en/solutions-for/critical-communications/network-planning-software/>. Accessed:2019-07-23.
- [28] A seamless ahn2 download service. https://3d.bk.tudelft.nl/pdfs/3dsm/matahn_3dbgt-dag_2014.pdf.
- [29] Semtech lora-based iot solution saves 22% in water management. <https://www.semtech.com/company/press/semtech-lora-based-iot-solution-saves-22-in-water-management>. Accessed:2019-07-23.
- [30] Smart agriculture | applications | semtech lora technology | semtech. <https://www.semtech.com/lora/lora-applications/smart-agriculture>. Accessed:2019-07-23.
- [31] Srtm mission statistics. <https://www2.jpl.nasa.gov/srtm/statistics.html>. Accessed:2019-07-23.
- [32] What is LoRa? <https://www.semtech.com/lora/what-is-lora>. Accessed:2019-07-23.
- [33] Ztrack - lora gps tracker. <http://ztrackmap.com/>. Accessed:2019-07-23.
- [34] A complete guide to lidar: light detection and ranging, Aug. 2015.
- [35] Fernando Kuipers Andri Rahmadhani. Understanding collisions in a lorawan. MSc thesis.
- [36] N. S. Bezerra, C. hlund, S. Saguna, and V. A. de Sousa. Propagation model evaluation for lorawan: planning tool versus real case scenario. In *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pages 1–6, Apr. 2019.
- [37] M. Bor and U. Roedig. Lora transmission parameter selection. In *2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 27–34, June 2017.
- [38] 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*, MSWiM ’16, pages 59–67, New York, NY, USA, 2016. ACM. event-place: Malta, Malta.
- [39] R. El Chall, S. Lahoud, and M. El Helou. Lorawan network: radio propagation models and performance evaluation in various environments in lebanon. *IEEE Internet of Things Journal*, 6(2):2366–2378, Apr. 2019.

- [40] V. A. Dambal, S. Mohadikar, A. Kumbhar, and I. Guvenc. Improving lora signal coverage in urban and sub-urban environments with uavs. In *2019 International Workshop on Antenna Technology (iWAT)*, pages 210–213, Mar. 2019.
- [41] S. Demetri, M. Zuniga, G. P. Picco, F. Kuipers, L. Bruzzone, and T. Telkamp. Automated estimation of link quality for lora: a remote sensing approach. In *2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 145–156, Apr. 2019.
- [42] George Grekousis, Giorgos Mountrakis, and Marinos Kavouras. An overview of 21 global and 43 regional land-cover mapping products. *International Journal of Remote Sensing*, 36(21):5309–5335, Nov. 2015.
- [43] 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 *Proceedings of the 2017 International Conference on Embedded Wireless Systems and Networks*, EWSN 17, pages 317–322, USA, 2017. Junction Publishing. event-place: Uppsala, Sweden.
- [44] P. Jrke, S. Becker, F. Liedmann, and C. Wietfeld. Urban channel models for smart city IoT-networks based on empirical measurements of LoRa-links at 433 and 868 MHz. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–6, Oct. 2017.
- [45] Sokratis Kartakis, Babu D. Choudhary, Alexander D. Gluhak, Lambros Lambbrinos, 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*, WiNTECH ’16, pages 2–8, New York, NY, USA, 2016. ACM. event-place: New York City, New York.
- [46] Dong-Hoon Kim, Eun-Kyu Lee, and Jibum Kim. Experiencing lora network establishment on a smart energy campus testbed. *Sustainability*, 11(7):1917, Jan. 2019.
- [47] Markus Knoll, Philipp Breitegger, and Alexander Bergmann. Low-Power Wide-Area technologies as building block for smart sensors in air quality measurements. *e & i Elektrotechnik und Informationstechnik*, 135(6):416–422, Oct. 2018.
- [48] Kais Mekki, Eddy Bajic, Frederic Chaxel, and Fernand Meyer. A comparative study of LPWAN technologies for large-scale IoT deployment. *ICT Express*, 5(1):1–7, Mar. 2019.
- [49] Gianni Pasolini, Chiara Buratti, Luca Feltrin, Flavio Zabini, Cristina De Castro, Roberto Verdone, and Oreste Andrisano. Smart city pilot projects using lora and ieee802. 15. 4 technologies. *Sensors (Basel, Switzerland)*, 18(4), Apr. 2018.
- [50] J. Petajajarvi, K. Mikhaylov, A. Roivainen, T. Hanninen, and M. Pettissalo. On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology. In *2015 14th International Conference on ITS Telecommunications (ITST)*, pages 55–59, Dec. 2015.
- [51] K. Phung, H. Tran, Q. Nguyen, T. T. Huong, and T. Nguyen. Analysis and assessment of LoRaWAN. In *2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications Computing (SigTelCom)*, pages 241–246, Jan. 2018.

- [52] M. R. Seye, B. Gueye, and M. Diallo. An evaluation of lora coverage in dakar peninsula. In *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 478–482, Oct. 2017.
- [53] Rashmi Sharan Sinha, Yiqiao Wei, and Seung-Hoon Hwang. A survey on lpwa technology: lora and nb-iot. *ICT Express*, 3(1):14–21, Mar. 2017.
- [54] Hugo van Meijeren. Assessing the differences between dutch elevation datasets ahn2 and ahn3. MSc thesis.
- [55] A. M. Yousuf, E. M. Rochester, B. Ousat, and M. Ghaderi. Throughput, coverage and scalability of lora lpwan for internet of things. In *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, pages 1–10, June 2018.