An investigation into the use of kriging for indoor Wi-Fi received signal strength estimation

by

PJ Joubert

A thesis submitted in partial fulfilment
of the requirements for the degree

MAGISTER INGENERIAE

in

COMPUTER AND ELECTRONIC ENGINEERING

in the

FACULTY OF ENGINEERING

at the

NORTH WEST UNIVERSITY

SUPERVISOR: Prof A.S.J. Helberg

November 2014

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT ®

Dit begin alles hier ™

Abstract

Kriging is proposed as a tool for Wi-Fi signal strength estimation for complex indoor environments. This proposal is based on two studies suggesting that kriging might be suitable for this application. Both of these studies have shortcomings in supporting this proposal, but their results encourage a more in depth investigation into this.

Even though kriging is a geostatistical method developed for geographical interpolation, it has been used successfully in a wide range of other applications as well. This further suggests that kriging might be a versatile method to overcome some of the difficulties of existing signal strength estimation methods. Two main types of signal strength estimation are deterministic methods and empirical methods. Deterministic methods are generally very complex and requires input parameters that are difficult to obtain. Empirical methods are known to have low accuracy which makes them unreliable for practical use.

Three main investigations are presented in order to evaluate the use of kriging for this application. A sampling plan is proposed as part of a generic application protocol for the practical use of kriging for Wi-Fi signal strength. It is concluded that kriging can be confidently used as an estimation technique for Wi-Fi signal strength in complex indoor environments. Kriging is recommended for practical applications, especially where insufficient information is available about a building or where time consuming site surveys are not feasible.

Keywords - Geostatistics, Kriging, Variogram, Wi-Fi, RSSI, Signal Strength Estimation, Propagation Models, Path Loss, Sampling, Coverage

Contents

List of Figures

List of Tables

1 Introduction

The high demand for mobile networking and the many applications such as coverage analysis, localization, fast hand-off, and security auditability has led to increased attention on signal strength estimation for indoor wireless communications. Signal Strength estimation is needed for generating radio signal maps for planning wireless networks [1].

A variety of approaches can be taken to estimate signal strength in an indoor environment and can be classified as either empirical model based or deterministic model based signal strength estimation. Empirical models are known to have low accuracy compared to deterministic models and are usually replaced with deterministic models. Deterministic models however can become very complex especially in an indoor environment where input parameters are difficult to obtain and not necessarily reliable [2].

A typical modern office environment can be described as a complex indoor environment containing many different wireless devices causing interference to each other. Reflections caused by different kinds of walls with different attenuation factors are unavoidable. A number of different wireless access points may be used to connect to an extensive network that can be overlapped by other networks sharing the same frequency bands.

Using a floor plan indicating the position of each access point and giving information about the walls in a building can provide the information needed to set up a simulation model to predict signal strength coverage for a building in an ideal environment. However, in a real life environment it will be impractical to take every single influence into account to set up a deterministic model for estimating signal strength throughout a building. If such a model could be constructed, it will be very processing intensive as well. In addition, a floor plan usually consists of a 2D representation of a floor in a building, thus adding a third dimension for different stories of a building only adds to the complexity.

A simple way to take every factor that has an influence on the signal strength into account is to physically measure the signal strength in the area of concern. In turn, it is also impractical to measure the signal strength at every point in a building. Since it is impractical to measure every single point, signal strength estimation methods are used.

1.1 Existing Methods

Signal strength estimation is an important aspect for planning of wireless networks and for generating radio signal maps. It can help to reduce the cost of time consuming site surveys and in addition it can be used to estimate the network coverage where samples could not be taken [2].

The following section describes methods currently used for signal strength estimation:

1.1.1 Radio Propagation Models

All radio propagation models can be grouped into three categories, based on the complexity of the model. These categories include the Simple attenuation models, Partition models, and Site-specific models [3].

Simple attenuation models usually form the basis of most other models. The partition models takes additional consideration of the attenuation effects from all indoor partitions, like walls and floors, into account. Partition models have proven great success in many cases. An example of such a model is the wall attenuation model used in RADAR [5]. Similarly, site-specific models also take partitions into account, but it relates path loss with parameters like geometrics, materials, and thickness at the specific site. Examples of site-specific models include the Hassan-Ali and Pahlavan probability model [6], and the Lot and Forkel multi-wall-and-floor model [7].

Shortcomings of Radio Propagation Models

The following is a list of shortcomings of radio propagation models:

1. Tedious time consuming measurements need to be taken to determine the attenuation coefficients of all the relevant items throughout a building.

2. The dynamic behaviour of the indoor radio propagation is not taken into account when taking the measurements.

3. Only the direct path between transmitter and receiver is considered when calculating path loss.

4. The characteristics and geometry properties of the materials need to be very detailed in order to use the site-specific models effectively.

The above implies that the radio propagation models are inconvenient to use.

1.1.2 Dynamic Signal Strength Estimation

In this section we discuss an estimation method that improves on some of the shortcomings of most radio propagation models [3]. This estimation method is called Dynamic Signal Strength Estimation and involves Floor Plan Interpretation, Ray Tracing, a Radio Propagation Model, and Parameter Estimation.

The method consists of capturing and characterising the floor plan to be able to produce 3D models necessary for ray tracing. Ray tracing is then used to determine how much each individual ray contributes to the signal strength. A propagation model is used and its parameters need to be solved.

Floor Plan Interpretation

The floor plan interpretation is used to automatically integrate the geometry acquisition process. The interpretation process extracts the structural parameters from CAD files and floor plans.

Ray Tracing

Ray tracing uses a finite number of isotropic rays emitted from a transmitting antenna to approximate radio propagation [8]. Each ray is assumed to transmit the same amount of energy when using omnidirectional antennas. The energy of each ray will attenuate individually as it goes through walls and floors. The walls and floors also cause reflections that have an influence on the energy at each point.

Radio Propagation Model

The signal strength at a receiver is the accumulated multipath strength of all individual rays of the transmitter. The attenuation of each ray is a result of the free space propagation loss, attenuation due to reflections, and attenuation due to transmission through obstacles.

Parameter Estimation

Measurements at reference positions need to be taken when estimating the radio propagation parameters inside a building. Only rays with signal strength above a certain threshold are considered in calculations to estimate the relevant parameters.

1.2 Geostatistical Estimation

Geostatistics is the study of phenomena that vary in space and consists of a number of numerical techniques that aim to characterise spatial attributes. Kriging is a geostatistical method that according to [9] provides optimal interpolation and can generate the best linear unbiased estimate at each location.

Kriging was originally developed for geographical interpolation purposes, but has proven to be a powerful tool in many other applications as well. This section provides background of different applications where kriging has been successfully implemented.

Kriging is an interpolation method based on a stochastic data model [10] that was developed to be used for geostatistical purposes, in particular to accurately predict ore reserves from the samples taken over a mining field. Since kriging is a very versatile method, it has also been implemented in a wide range of other applications.

Kriging has also been described as an accurate and fast model to assist in antenna modelling [11] and design [12] and is considered an efficient technique for design optimization of antenna structures [13]. It is also known for its use in wireless sensor networks. An example of such a wireless sensor network is where the spatial distribution of received power, at given frequencies, is estimated. These estimations are used to increase the efficiency of spectrum usage through dynamic spectrum access with Cognitive Radio networks [14].

Other unconventional uses for kriging include missing pixel recovery of digital images [15] and 3D active object recognition where a restrictive sampling budget is of concern [16].

In [17] kriging is used to lower the computational cost of computer-aided design optimization in modern microwave design and [20] proposes a technique for estimating the spatial electromagnetic field distribution by also incorporating kriging.

The above examples show the versatility of kriging and suggest that it is a reliable method containing numerous advantages for estimating values where data sampling is time consuming or only a limited amount of samples are available.

1.3 Problem Statement

In this study we evaluate the use of kriging for Wi-Fi signal strength estimation in an indoor environment. This idea originated from the fact that Wi-Fi has a geographical property that obeys the first law of geography: All places are related, but nearby places are more related than distant places [21]. The random spatial distribution of signal strength in an environment as explained above further suggests the use of a geostatistical model.

In [2] ordinary kriging was used in combination with a path loss model to estimate signal strength in wireless local area networks. This though was only done in simulation.

In [22] universal kriging was used to predict signal strength in an indoor environment. The shortcoming of this study is that the experimental setup consisted of a single hall with five access points and the results presented consisted of only six interpolated points ranging between -41.9 dBm and -49.9 dBm. This result is insufficient since Wi-Fi signal strength measured throughout a typical indoor environment, ranges between about -40 dBm and -100 dBm. The single hall also does not represent a complex indoor environment.

Even though both studies strongly recommend the use of kriging as a solution for Wi-Fi signal strength estimation, neither clearly indicated the behaviour of kriging and the accuracy that can be achieved in a complex indoor environment as described above. The methods also do not provide a practical method setting up a sampling plan.

1.4 Hypothesis

Kriging is a geostatistical tool that is well known for its reliable results in scenarios where little is known about the environment in which it is used. The hypothesis for this study is that kriging will be a suitable method for estimating signal strength in a complex indoor environment and that kriging will be more effective and convenient to use than deterministic methods while obtaining the same level of accuracy.

1.4.1 Research Objectives

Our research objectives are divided into the primary and secondary objectives presented in this section.

Primary Objective

The primary objective is to test our hypothesis that is based on results obtained from simulations and suggestions made by [2] and [22] regarding the use of kriging for signal strength estimation.

A single technique that can be applied at any site needs to be proposed in contrast to deterministic methods which require that each new site be analysed separately. The goal is to investigate whether kriging is a suitable candidate for such a technique.

The two factors that will be most important in evaluating kriging for this application are the accuracy of the results and the simplicity of the sampling plan. The outcome will be used to comment on the relevance of kriging for signal strength estimation.

Secondary Objective

The secondary objective is to define a generic application protocol for using kriging to estimate Wi-Fi signal strength indoors. The main challenge here is to specify a sampling plan that needs to be followed. This will result in a practical implementation of kriging to be used in buildings for signal strength estimation.

1.5 Scope

The scope of this study consists of empirical investigations in order to evaluate the accuracy and relevance of the use of kriging for Wi-Fi signal strength estimation in a complex, three-dimensional, indoor environment. The results will be compared to other methods for estimating signal strength where possible using information provided in literature. However, the scope of this research does not include investigating other methods in order to be compared with the results obtained from applying kriging. This research will focus specifically on the use of kriging for indoor Wi-Fi signal strength estimation.

The accuracy of the proposed method shall be determined by use of cross validation and the error will be expressed as a normalized percentage or in dB in order to get a realistic view of the accuracy of the results obtained from each investigation. In each investigation the whole environment must be measured with an appropriate resolution for cross validation to take place.

A sampling plan shall be suggested for use of this method in future applications. The sampling plan shall provide guidelines for the number of samples that will be sufficient for a given environment. Suggestions on how to determine the most effective locations for taking samples shall be made and a step by step explanation of the sampling process shall be given.

The scope of this research does not include investigating the use of kriging for estimating signal strength of other signals, even though it may be equally accurate. The scope will become too large for this study if investigations into signal strength estimation for other signals are included.

1.6 Research Methodology

Information will be obtained from literature to serve as background knowledge. This will assist in facing the challenges that need to be overcome in this study. A practical approach will be taken by applying the knowledge and learning more from the outcomes.

The practical approach will involve doing different empirical investigations, increasing in complexity with each investigation. The first investigation will be done to examine the applicability if kriging as a tool for signal strength estimation using only a single Wi-Fi access point in an indoor environment.

The rest of the investigations will be done with an approach similar to using a single access point, but in more complex environments. Investigations involving multiple access points will be done after which a generic sampling plan for complex indoor environments will be set up. The results of each investigation will be compared to the previous investigations to identify similarities and contradictions. This will provide a better understanding of the behaviour of kriging for this application.

The accuracy of the investigations will be determined through cross validation. In each investigation RSSI (Received Signal Strength Indicator) will be measured as extensively as possible to have physical measurements with which the estimated results can be compared. The results will be presented in suitable graphical presentations to show relevant information. This will allow comments to be made on the performance of the method. Once an acceptable level of confidence in the proposed method is reached, the method will be applied at other sites for validation.

Taking the results from the investigations into account, a conclusion will be reached in context of the background of the problem. Suggestions will be made on the practical use of kriging for RSSI estimation.

1.7 Publications

Peer review publications and submissions resulting from this study consist of the following:

• SATNAC 2013 - An investigation into the accuracy of the kriging method for single Wi-Fi access point received signal strength estimation

• SATNAC 2014 - An investigation into the accuracy of the kriging method for multiple Wi-Fi access point RSSI estimation

• WCNC 2015 (under review) - An investigation into the use of kriging for Wi-Fi RSSI estimation in complex indoor environments

1.8 Chapter Conclusion

The technical survey showed that kriging can be used in a wide range of applications. More specifically, simulations have been done for estimating network coverage in wireless local area networks using ordinary kriging [2]. This suggests that kriging is a valid method to take into consideration for this application, but that further investigations are necessary to be able to comment on the performance of kriging for this application. Investigations will be done by taking a practical approach in a more complex environment and in a more extensive fashion than in [22].

A practical approach will show the behaviour of kriging in real scenarios. The accuracy and practicality of the method will be validated by repeating the process at different locations increasing in complexity. The simple investigations will give an early indication of the feasibility in order to proceed with the study.

2 Literature Study

Since we propose to combine a geostatistical method with W-Fi signal strength estimation, this study requires background knowledge about a wide variety of topics. Within this study these topics are all related.

The background and history of the Wi-Fi standard will be discussed. The different ways of presenting Wi-Fi signal strength will be identified and the difference between RSS and RSSI will be explained.

It is important to understand the basics of signal propagation and the challenges present when attempting to estimate signal strength in indoor environments. Factors that influence signal strength will be discussed and the amount of attenuation caused by different materials will be shown.

Current types of estimation techniques are discussed and compared. This provides the background necessary to compare the performance and usability of kriging for this application.

Since kriging is a geostatistical method and one of the outcomes of this study involves setting up a sampling plan for the practical use of kriging for RSSI estimation, a background to statistical sampling techniques are presented. These techniques will be considered and brought in perspective with the kriging algorithm in order to define a suitable sampling plan.

2.1 Wi-Fi

The IEEE 802.11 standard, commonly known as Wi-Fi, is the first WLAN standard and so far the only one that has secured the market. The standardization of the IEEE

802.11 started in 1987 as part of the IEEE 802.4 Token Bus standard. The IEEE 802.4 is a counterpart of IEEE 802.3 Ethernet and 802.5 Token Ring. In 1990 the 802.4 WLAN group was renamed as IEEE 802.11 which formed an independent 802 standard that defines PHY and MAC layers for WLANs. The first IEEE 802.11 standard was completed in October 1997 [18] [23].

The first IEEE 802.11 standard had data rates of 1 and 2 Mbps and made use of frequency-hopping spread spectrum (FHSS), direct-sequence spread spectrum (DSSS) and diffused infrared (DFIR) physical layers for radio transmission [23].

In October 1999 the IEEE 802.11a and 802.11b amendments were approved by the IEEE 802.11 committee [18]. The IEEE 802.11a operates at the 5 GHz band and supports up to 54 Mbps using orthogonal frequency division multiplexing (OFDM). The IEEE 802.11b operates at the 2.4 GHz band and uses the high rate direct sequence spread spectrum (HR/DSSS) technique and the complementary code keying (CCK) modulation scheme to provide data rates of 5.5 and 11 Mbps. The IEEE 802.11g standard that was published in June 2003 also operates at the 2.4 GHz band and support data rates of up to 54 Mbps. It is backwards compatible with IEEE 802.11b, therefore the IEEE 802.11b and 802.11g has become mainstream standards for WLAN products.

The latest Wi-Fi standard, the IEEE 802.11n published in 2009, made significant improvement over previous standards such as the 802.11a and 802.11b/g. Improvements involved the following [24] [25]:

• The MAC layer transfer rate was increased to achieve a minimum of 100 Mbps data throughput.

• New block acknowledgements were added.

• The modified OFDM, increasing data sub-carriers from 48 to 52, improved maximum throughput from 54 to 58.5 Mbps.

• Improved forward error correction boosted the link rate from 58.5 to 65 Mbps.

• The short guard interval (GI) between OFDM intervals was decreased from 800 ns to 400 ns increasing throughput from 65 to 72.2 Mbps.

• Doubling channel bandwidth from 20 to 40 MHz slightly more than doubles the rate from 72.2 to 150 Mbps.

• Spatial multiplexing support for up to 4 spatial streams (MIMO - multiple input, multiple output antennas) increased throughput up to 4 times from 150 to 600 Mbps.

• IEEE 802.11n remains backwards compatible with existing IEEE WLAN legacy solutions.

2.2 Signal Strength in 802.11 Networks

RSSI is usually described as the average received signal strength at a given receiver during the reception of a packet, expressed in dBm [26]. However, due to a number of factors, a great deal of confusion and inconsistency occur when referring to 802.11 terms such as signal strength, signal to noise ratio, and signal quality. This section will provide a clear understanding of what is meant by RSSI in terms of Wi-Fi network coverage and how to interpret this value.

2.2.1 mW vs. dBm

Most network card vendors' utility tools for analysing 802.11 represent RF signal strength in terms of mW (milliwatts), dBm (dB-milliwatts), RSSI (Received Signal Strength Indicator) and/or a percentage. These units are all related and can all be converted from one unit to another.

The output power of a typical wireless access point is between 1 and 100 mW and the output power of a wireless client is between 1 and 30 mW. Due to the fact that signal strength fades inversely squared with distance, a receiver will practically never receive signals above 1 mW. This implies that mW is not a suitable way of representing signal strength. Since dBm is expressed as 10 times the base 10 logarithm of the power, the range of values are much more suitable to describe the power of received signal strength. For example, the difference between -85 dBm and -95 dBm is a difference of approximately 0.000 000 003 mW. This seemingly small difference might be the difference between a stable connection and a slow, unreliable connection.

2.2.2 RSSI

The IEEE 802.11 standard defines a mechanism by which RF energy needs to be measured by a wireless NIC. This specifies that a 1 byte integer called Received Signal Strength Indicator (RSSI) be used to represent the signal strength presumed by a NIC. It does not require that a vendor use all 256 values, which implies that each NIC will specify a maximum RSSI value referred to as RSSI_Max.

Cisco, for example, decided to use 101 levels for representing the RF energy with an RSSI_Max of 100. Symbol chipsets use an RSSI_Max of 31 and Atheros uses an RSSI_Max value of 60[1] [28].

Although the term 'signal strength' is generally used, the 802.11 standard does not define signal strength as measuring RF energy in mW or dBm. It rather uses the RSSI value. The reported signal strength is a value between 0 and RSSI_Max intended for use internally by the physical and data link layers. This value can be converted to represent the user of a utility tool with a signal strength measurement presented in one of the units specified above.

This conversion is done differently by different vendors and thus should not be used as an absolute reference, but should rather be considered a relative value. Since conversions are done differently, the indicated signal strength of different NICs should never be compared [29]. In [30], significant differences were found between different Wi-Fi devices. Even devices from the same vendor did not perform similarly and devices from the same model could not be proven to perform identically.

2.2.3 Signal Strength as a Percentage

When using different utility tools it is common to see signal strength represented as a percentage. Calculating the indicated percentage is done by dividing the RSSI for a particular packet by the RSSI_Max value, multiplied by 100. For example a Cisco device with an RSSI of 80 will indicate the signal strength as 80% where an Atheros device will indicate 80% when the RSSI value is 48.

[1] These values are based on the vendors' 802.11b chipsets circa 2002 and are intended only as examples

One might want to assume that using the indicated percentage will allow for comparison between different vendors, but the problem with this assumption is that the RSSI_Max value is not set at the same power level for all vendors. Most NICs would report a percentage of 100% for a received power near 1 mW, but some might for example report 90% signal strength where others report 100% depending on where they chose to put their 100% position on the mW scale [28].

2.2.4 Precision in RSSI Measurements

Given that RSSI is an integer value, it must change in integer steps between 0 and RSSI_Max. Since each vendor can define its own RSSI_Max, the actual range of energy being measured must be divided into the number of integer steps provided by the RSSI range. If the RSSI value changes by 1, the actual power changed by a fraction of the measured range. A vendor with a higher RSSI_Max can indicate the received signal strength with more precision than a vendor with a lower RSSI_Max.

NICs generally do not need to measure signal strength with maximum precision since RSSI is usually used internally to determine whether another station is transmitting or to determine which antenna has the strongest signal in order to make decisions about roaming and data rate adjustments. These decisions do not need a high level of accuracy to be made effectively. If a vendors need higher precision they have the choice to use an RSSI_Max of up to 255, but vendors rarely, if ever, choose RSSI_Max to be higher than 100 [28].

2.2.5 Signal Strength Measurements

If a wireless network analysis tool is used to measure signal strength from an access point at a single location, the tool will not report a constant value. The reported value will fluctuate across a range of values. This fluctuation is a result of all the factors in the environment affecting the signal in the dynamic electromagnetic spectrum in which RF energy exists. Since the contributed power of an 802.11 device is small, noise can easily disrupt the 802.11 signal.

The term fade margin refers to the magnitude of the maximum reduction in signal strength resulting from the various environmental influences. Designing a wireless network usually involves including a fading margin of approximately 10 dB to account for possible environmental degradation. If, for example, an 802.11b device needs a signal strength of -85 dBm to operate at 11 Mbps, then the network is designed to have a signal strength of -75 dBm to compensate for environmental influences. Fading is discussed in more detail in the path loss section below.

The author of [18] suggests using WirelessMon as measurement tool. Other popular applications like inSSIDer, AirMagnet, Sniffer Wireless, or AiroPeek also provide RSSI expressed in dBm. If a specific value is required to represent the signal strength at a point, a more accurate value can be obtained by averaging the measured values over a predetermined period.

2.2.6 Data Rate vs. Signal Strength

Different data rates have different levels of complexity in encoding and modulation which results in different required receiver sensitivities. A higher data rate needs a stronger signal to operate reliably. For this reason an 802.11 NIC can change data rates according to its received signal strength.

MCS Rate Index		Modulation/ECC	Data Rate (Mbps)				Receive Sensitivity (dBm)	
			800NS GI		400NS GI			
			20 MHz	40 MHz	20 MHz	40 MHz	20 MHz	40 MHz
0	1	BPSK/1:2	6.5	13.5	7.2	15	-82	-79
1	1	QPSK/1:2	13	27	14.4	30	-79	-76
2	1	QPSK/3:4	19.5	40.5	21.7	45	-77	-74
3	1	16-QAM/1:2	26	54	28.9	60	-74	-71
4	1	16-QAM/3:4	39	81	43.3	90	-70	-67
5	1	64-QAM/2:3	52	108	57.8	120	-66	-63
6	1	64-QAM/3:4	58.5	121.5	65	135	-65	-62
7	1	64-QAM/5:6	65	135	72.2	150	-64	-61
8	2	BPSK/1:2	13	27	14.4	30	-82	-79
9	2	QPSK/1:2	26	54	28.9	60	-79	-76
10	2	QPSK/3:4	39	81	43.3	90	-77	-74
11	2	16-QAM/1:2	52	108	57.8	120	-74	-71
12	2	16-QAM/3:4	78	162	86.7	180	-70	-67
13	2	64-QAM/2:3	104	216	115.6	240	-66	-63
14	2	64-QAM/3:4	117	243	130	270	-65	-62
15	2	64-QAM/5:6	130	270	144.4	300	-64	-61
16	3	BPSK/1:2	19.5	40.5	21.7	45	-82	-79
17	3	QPSK/1:2	39	81	43.3	90	-79	-76
18	3	QPSK/3:4	58.5	121.5	65	135	-77	-74
19	3	16-QAM/1:2	78	162	86.7	180	-74	-71
20	3	16-QAM/3:4	117	243	130.7	270	-70	-67
21	3	64-QAM/2:3	156	324	173.3	360	-66	-63
22	3	64-QAM/3:4	175.5	364.5	195	405	-65	-62
23	3	64-QAM/5:6	195	405	216.7	450	-64	-61

Table 2.1: Data Rate vs. Signal Strength [27]

Table 2.1 lists the association data rates, also known as the Modulation and Coding Scheme or MCS rate indices, and associated received sensitivities for an 802.11n network [27].

2.2.7 Converting RSSI to dBm

To illustrate the different ways of converting RSSI to dBm, three different vendors' methods will be presented here. If signal strength is represented as a percentage, it must first be converted to RSSI taking each vendor's RSSI_Max value into account [28].

Atheros

Atheros uses a formula to derive dBm from RSSI. Atheros has an RSSI_Max of 60. The dBm value is calculated by subtracting 95 from the RSSI value. This gives a dBm range of -35 dBm at 100% signal strength down to -95 dBm at 0% signal strength.

$$P = RSSI - 95 \text{ for } 0 \leq RSSI \leq 60 \tag{2.1}$$

Symbol

Symbol uses an RSSI_Max of 31. Table 2.2 is used to obtain a dBm value:

Table 2.2: Symbol - RSSI Lookup Table

RSSI $\leq$ 4	RSSI $\leq$ 8	RSSI $\leq$ 14	RSSI $\leq$ 20	RSSI $\leq$ 26	RSSI $>$ 26
-100 dBm	-90 dBm	-80 dBm	-70 dBm	-60 dBm	-50 dBm

Notice that Symbol devices have a range of -50 dBm to -100 dBm, but increment only in steps of 10 dB.

Cisco

With an RSSI_Max of 100, Cisco has the least granular dBm lookup table presented as a graph in Figure 2.1.

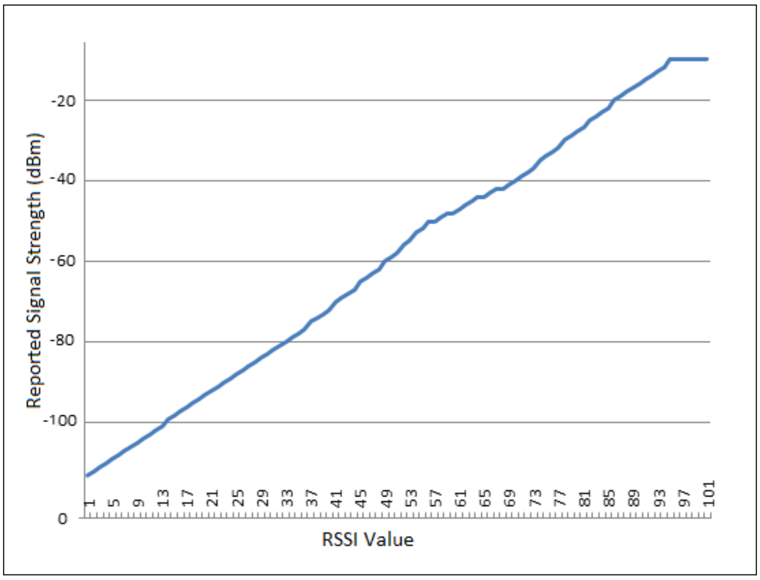

Figure 2.1: Cisco RSSI

2.3 Path Loss

Path loss is defined as the attenuation that occurs as RF signals propagate from a transmitting antenna to a receiving antenna [31]. This attenuation is caused by a number of factors including interference, reflections, refraction, scattering, and multi-path propagation [18]. These factors will be explained in this section in order to provide background information on the expected challenges that need to be overcome when estimating signal strength.

2.3.1 Path Loss Basics

In any wireless network the power received by a receiver P_{rx} can be expressed as in equation (2.2) [32] [33] [34] [35].

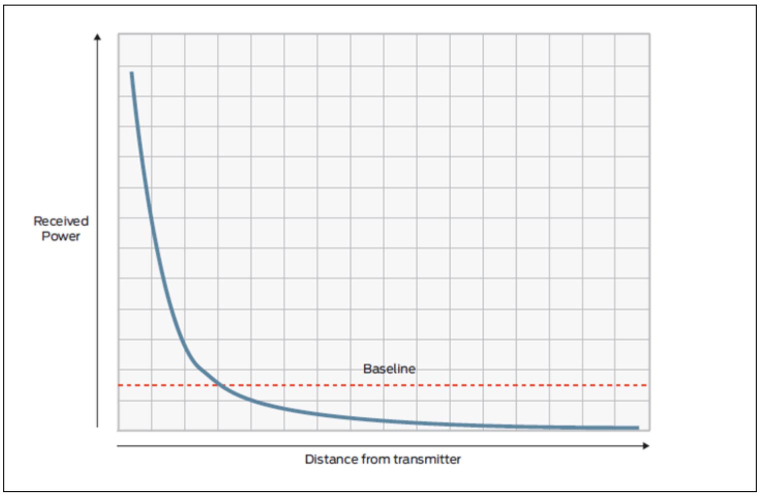

Figure 2.2: Free Space Path Loss [27]

$$P_{rx} = P_{tx} + G_{tx} + G_{rx} - PL \tag{2.2}$$

where P_{tx} is the power transmitted at the transmitter, G_{tx} is the gain of the transmitter, G_{rx} is the gain of the receiver and PL is the path loss. For any receiver there is a minimum detectable signal (MDS) that must be an acceptable level above the noise floor. Therefore the inequality in equation (2.3) must be satisfied:

$$P_{rx} = P_{tx} + G_{tx} + G_{rx} - PL \geq MDS(P_e) \tag{2.3}$$

where P_e is the probability that a bit error might occur.

The theoretical free space path loss expressed in equation (2.4) shows how the power received by a receiver is inversely squared to the distance measured from the transmitter.

$$PL(d) = 20\log_{10}(\frac{\lambda}{4\pi d}) \tag{2.4}$$

However, there is no close correlation between theoretical free space attenuation and the real signal attenuation when measured in an indoor environment [36]. This complicates the topic greatly and requires insight into the factors that influence signals in order to plan or analyse indoor wireless networks.

2.4 Interference

Interference can be seen as 'man-made noise' and is a serious issue, especially in the 2.4 GHz ISM band, which is one of the main frequency bands for Wi-Fi [18]. Performance can decline significantly when too many ISM devices operate in a small area. ISM devices include many industrial and medical equipment as well as household appliances such as microwave ovens and Bluetooth devices. The authors in [18] states that interference from Bluetooth devices in close proximity to Wi-Fi devices can degrade the throughput of IEEE 802.11b stations by 25% to 66%.

The IEEE 802.11 standard defined 14 channels, but only 3 non-overlapping channels (1, 6 and 11) are available [37] [38]. If two APs are using the same channel or adjacent channels, co-channel interference can be caused which will degrade an IEEE 802.11b WLAN by 2 Mbps [39].

In summary, interference is a problem unique to every environment, making it difficult to model or simulate without accurate knowledge about a specific environment.

2.5 Multi-Path Propagation

Multi-path propagation is caused by reflection and diffraction that commonly occur in indoor environments [40]. When signals reflect from objects they are delayed and arrive later at the receiver than signals that followed a direct path. The period between the duplicate signals arriving at the receiver is referred to as delay spread. A larger delay spread is preferable since it allows devices more time to recognise duplicate signals [41].

Other than causing duplicate signals to arrive at a receiver, reflections and diffraction also cause attenuation of signals.

2.6 Attenuation

Attenuation is the decay of a signal as it moves from a transmitter to a receiver. A few factors that contribute to attenuation include distance, obstructions, multipath effect, scattering, and absorption. Attenuation radically decreases performance of a wireless network.

Different materials that cause obstruction have different attenuation factors. The attenuation caused by a few different materials are presented for comparison.

2.6.1 Material Attenuation

Table 2.3 shows the attenuation that different materials cause to an RF signal. These values can be used in deterministic models with a floor plan indicating the different types of walls in a building.

Table 2.3: Material Attenuation [18] [35]

Material	Attenuation (dB)
Plasterboard	3 to 5
Glass wall with metal frame	6
Cinder block wall	4 to 6
Window	3
Metal door	6 to 10
Concrete wall	6 to 15
Floors of a building	12 to 27

2.6.2 Fading

If adequate information is available regarding a specific environment, the path loss at a point in a building can be calculated by applying one of the methods discussed in the estimation section below. This will be valid for steady state conditions, but in real conditions there are continuous changes in the environment. The changes in the environment cause what is called fading [28].

Fading is the inconsistent and unpredictable changes in signal strength at a point in a building that occur without changing the power transmitted. The environment between an 802.11 transmitter and receiver is very complex and constantly changing which causes the signal strength to fade unpredictably even if a client is stationary. In modern wireless networks with many mobile devices, fading can especially become excessive when for

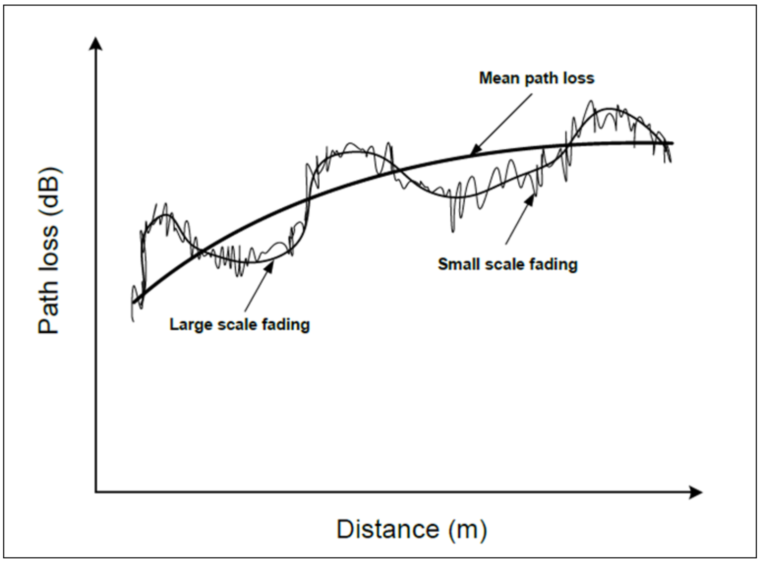

Figure 2.3: Path Loss and Fading [33]

example people walk between the device and the AP. A user of a device can turn so that he is between the device and the AP, or just move a hand over the antenna [28].

The instantaneous path loss is a combination of the mean path loss, the large scale fading (also known as shadow fading), and the small scale fading. Figure 2.3 shows an illustration of the path loss and fading in a complex environment [33].

Multipath fading and the changes in the environment play the greatest role in affecting RF signals [29]. Therefore, it is very difficult to accurately estimate signal strength by analytical modelling and simulation in a given environment [18]. Designing and analysing a wireless network, with an appropriate fading margin, can be achieved for steady state conditions and is simplified by using a suitable propagation model or by following the steps of a well defined method to estimate signal strength statistically.

2.7 RSSI Estimation

The popularity and importance of wireless networks are growing rapidly which increases the need for better methods of modelling and measuring wireless signal propagation [32]. These methods encounter numerous challenges under which multipath fading and changing environments are two of the biggest problems. In wide open areas RSSI values are related to what would be presumed, but when objects are placed in its path it has a major influence on the measured values, making them directly dependent on the environment's complexity [29].

Even though the section on 802.11 signals points out many difficulties regarding the use of RSSI for applications other than data rate adjustments or making decisions on roaming, many current applications are based on RSSI measurements. This keeps RSSI estimations relevant. If RSSI estimations can be made with acceptable accuracy compared to the accuracy of RSS measurements, it will still be relevant.

If the 802.11 standard can provide an enhanced metric that will be available to devices and users for decision making, the estimation methods discussed here might perform better. A possible metric that has been suggested by [42] is CSI (channel state information). CSI is available on some Intel NICs working with the 802.11n standard, but investigations into this are left for future studies.

2.7.1 RSSI Estimation for Indoor Localization

A popular use for RSSI estimation is to assist in streamlining indoor localization techniques. RSSI estimation speeds up the process of measuring the whole area of concern in order to draw coverage maps.

RSSI based localization techniques consist of a training phase and an estimation phase [43]. In the training phase, RSSI samples are mapped to predefined positions. These positions are usually defined by dividing the environment into cells. In the estimation phase, a target's location is estimated using the coverage map from the training phase. The estimation can be done either by probabilistic or by deterministic techniques. The deterministic techniques make use of knowledge obtained from the environment during the training phase. Probabilistic methods make use of statistical analysis to construct a probability distribution of the target's location for the area of concern [43]. This implies that there is a trade-off between precision and computational overhead in selecting the most suitable technique.

In [44] it is found that the main attraction of RSSI as a metric for localization is that the measurement and calculations involved with RSSI are very simple compared to other localization metrics. Another study specifically stresses the fact that methods using delays or angle measurements are complex and the measurements are difficult to obtain in wireless networks [45]. They also stated that RSSI can be easily extracted which made it their metric of choice. RSSI is used for different localization techniques in [36], [46] and [34] to name only a few.

2.7.2 Estimation Categories

Indoor radio propagation models are categorized mainly into four groups: deterministic models, semi-deterministic models, stochastic models, and empirical models [33]. In this section these categories are explained under the two main classes, a priori models and measurement models [32]. Examples of models fitting each class will be described and accompanied by results obtained from other studies if results are available.

Table 2.4 shows examples of propagation models from each category:

Table 2.4: Estimation Categories

Deterministic	Semi-deterministic	Stochastic	Empirical
Ray launching	Dominant Path	Rayleigh fading	One-Slope
Ray tracing	Motif	Rice fading	Dual-Slope
Finite-Distance Time-Domain	Geometry-based Channel	Nakagami-m fading	Wall and Floor Factor
ParFlow		Log-normal fading	COST231 Multi-Wall
Multi-Resolution Frequency ParFlow			Linear Attenuation

2.7.3 A Priori Models

Path loss models that make predictions based on prior knowledge obtained from an environment are called a priori models. Knowledge are gathered by using analytical expectations about propagation in a building obtained from a floor plan and parameters describing properties such as attenuation factors of different materials that are present in the environment, as discussed in the Path Loss section above.

In a survey of path loss models developed in the last 60 years [32], a priori models are subdivided into six categories:

Theoretical/Foundational Models

These models are purely analytical and derived from the theory of an ideal electromagnetic environment. Even though these models have questionable accuracy in real environments, they have been widely implemented in network simulators and are used in more complex models to serve as a minimum loss indicator.

Examples of theoretical/foundational models include:

- Free Space Between Isotropic Antennas [47]
- Flexible Path Loss Exponent [32]
- Ground Reflection [48] [49]

Basic Models

Basic models are considered to be the most popular model type. Path loss is computed along a single path and corrections are made based on measurements. They use distance, carrier frequency, and antenna heights as input. The following models are categorized as Basic Models:

- Egli [50]
- Green Obaidat [51]
- Edwards-Durkin [52]
- Blomquist-Ladell [53]
- Allsebrook-Parson [54]
- deSouza-Lins [55]
- TM90 [56]

- Hata-Okumura [57]
  - COST-Hata/Extended Hata
  - Hata-Davidson
  - ECC-33
  - ITU-R/CCIR
  - Rural Hata
- Flat-Edge [58]
- Walfisch-Bertoni [59]
- Walfisch-Ikegami [60]
- Herring [61]
- Erceg-Greenstein [62]
- IMT-2000: Pedestrian Environment [63]

Terrain Models

Terrain models are similar to basic models, but take diffraction losses, due to obstacles, along the line-of-sight into account. They are more complex than basic models and are usually used for long distances at high power in the VHF band. Terrain models include:

- ITU Terrain [64] [65]
- Longley-Rice Irregular Terrain Model [66]

Supplementary Models

Supplementary models are used in combination with other models. They aim to correct for weaknesses in existing models. In [32] the phenomenon these models are attempting to correct for are subdivided as follows:

- Frequency Coverage [67]
- Obstructions [68] [69] [70]
  - Atmospheric Gases
  - Statistical Terrain Diffraction Estimate
  - Building-Transmission
  - Durgin-Rapaport
  - Vegetation
- Directivity [71] [72]
  - Gain Reduction Factor
  - EDAM

Stochastic Fading Models

Stochastic fading models account for additional fading by adding a random variable to a path loss model. The additional fading is caused by scattering and multipath effects that are uncorrelated in measurements at distances less than a wavelength. Attenuation due to fading can be a function of time or frequency. These models are specifically useful for designing the physical layer and data-link layer of wireless networks.

Stochastic fading models can be subdivided into large scale and small scale models [56]:

- Large Scale
  - Lognormal Shadowing Model [48]
- Small Scale - Used with the following distributions:
  - Rayleigh [73]

- Ricean [74]
- Nakagami [75]
- Barclay-Okumura a simple stochastic fading model proposed by Barkley in [76] based on data collected by Okumura.

2.7.4 Many-Ray Models

Many ray models are used to refer to ray-launching or ray tracing models, but are named so by [32] to emphasize how they differ from the previous types of models. Instead of only calculating the line-of-sight path loss, they sum the loss along many distinct paths.

These models require very detailed information about the environment. Vector models of buildings in 2D and 3D accompanied by interfering structures are commonly used to be able to trace the interaction of many individual paths. The effect of obstacles, reflections, refraction, and diffraction are calculated using the Uniform Theory of Diffraction (UTD), or an equivalent numerical approximation. This enables these methods to calculate the median path loss as well as the delay, spread and frequency shift of signals arriving at a receiver.

Even though these models are the most advanced of all the above models, their major concern is the amount of pre-processing and extensive number of calculations necessary to make estimations. A substantial amount of optimization is needed to make these models feasible for complex environments.

2.8 Model Comparison

In order to get a sense of the functioning of current estimation methods and their level of accuracy, this section briefly presents a comparison of existing methods. In [4] a comparison was made between five different indoor coverage models. The following path loss propagation models were compared:

2.8.1 One-Slope Model

In the One-Slope model the path loss is given in dB by equation (2.5):

$$L_{dB} = L_{0,dB} + 10n\log_{10} d \tag{2.5}$$

where $L_{(0,dB)}$ is the path loss measured 1m from the transmitter and n is the path loss exponent, which is determined experimentally using an interpolation technique [77].

2.8.2 Dual-Slope Model

The Dual-Slope model is similar to the One-Slope model, but elaborates on it by dividing the distance into a line of sight (LOS) and an obstructed LOS section. The path loss is calculated in dB by equation (2.6) [78]:

$$L_{dB} = L_{0,dB} + \begin{cases} 10n_1\log_{10} d, 1 < d \leq d_{bp} \\ 10n_1\log_{10} d_{bp} + 10n_2\log_{10}(\frac{d}{d_{bp}}), d > d_{bp} \end{cases} \tag{2.6}$$

where n_1 and n_2 are determined experimentally and the break point distance d_bp is calculated by equation (2.7):

$$d_{bp} = \frac{4h_b h_m}{\lambda} \tag{2.7}$$

where h_b and h_m are the shortest distance from the ground or wall from the AP and client respectively.

2.8.3 Partitioned Model

The Partitioned model makes use of previous field measurement campaigns to determine values for the path loss exponents and break point distances. Equation (2.8) is used to calculate the path loss in dB:

$$L_{dB} = L_{0,dB} + \begin{cases} 20\log_{10} d, & 1m < d \leq 10m \\ 20 + 30\log_{10}\frac{d}{10}, & 10m < d \leq 20m \\ 29 + 60\log_{10}\frac{d}{20}, & 20m < d \leq 40m \\ 47 + 120\log_{10}\frac{d}{40}, & 40m < d \end{cases} \tag{2.8}$$

2.8.4 COST231-Multi-Wall Model

In equation (2.9) the path loss is calculated, in dB, using the COST231-Multi-Wall model.

$$L_{dB} = L_{(0,dB)} + 20\log_{10} d + k_f^{[\frac{k_f+2}{k_f+1}-b]} L_f + \sum_{i=1}^{k_w} k_{wi} L_{wi} \tag{2.9}$$

where k_f is the number of penetrated floors and b is used to empirically fit the non-linear effects of the floors. L_f indicates the loss between adjacent floors. The number of wall types is denoted by k_w. L_0 is the free space path loss and is calculated as in equation (2.10):

$$L_0 = (\frac{4\pi d_0}{\lambda})^2 \tag{2.10}$$

Table 2.5 lists the values as an example of a building with two types of walls [4].

Table 2.5: Wall Attenuation [4]

Wall Type	Description	Value (dB)
k_{w1}	Light wall: plasterboard or light concrete wall ($< 10cm$)	3.4
k_{w2}	Heavy wall: thick ($> 10cm$), concrete or brick	6.9

2.8.5 Average Walls Model

The Average Walls model is based on the COST-231, but the loss caused by walls is combined into one parameter. The path loss after i walls is determined as in equation (2.11):

$$L_i = L - L_{0,dB} - 20\log_{10} d - \sum_{j=1}^{i-1} L_j \tag{2.11}$$

where L denotes the measured loss at 1m behind each wall.

2.8.6 Comparison Results

In [4] RSSI values were measured using free test software on a notebook with an 802.11a/b/g card adapter. The accuracy of the five different propagation models was compared and they concluded that the Average Walls model provided the best results.

The results were obtained using cross validation and the standard deviation was within the shadowing standard deviation that is typically between 5 and 12 dB [79]. The performance of all five models as obtained from [4] and [78] is listed in Table 2.6.

Table 2.6: Model Comparison

Model	Mean Error (dB)	Standard Deviation (dB)
Dual Slope [78]	12.38	9.68
Partitioned [78]	6.46	2.10
Log-Normal Shadowing [78]	7.74	5.34
P. 1238-1 [78]	6.79	3.54
Adjusted Motley-Keenan [78]	7.70	5.86
COST-231 Multi-wall [78]	7.87	6.16
COST-231 [4]	AP1 2.05 AP2 12.70	AP1 1.74 AP2 5.93
One-Slope [4]	AP1 2.29 AP2 3.91	AP1 1.45 AP2 2.99
Dual-Slope [4]	AP1 6.41 AP2 9.25	AP1 4.73 AP2 7.89
Average Walls [4]	AP1 2.67 AP2 7.52	AP1 2.56 AP2 6.84

2.9 Measurement Models

Taking every factor that affects the signal strength in an indoor environment into account might be simplified by physically taking measurements, thus capturing information of real life conditions in the measurements.

The final category is more about defining a method for taking measurements that will be representative of an environment than having a model to represent an environment. It assumes that no sufficient data is available for constructing an a priori model [32], which is usually the case in a complex indoor environment, and that taking a number of samples is unavoidable. The samples will be used to predict or interpolate the values throughout the rest of the area.

Three examples of measurement models will be briefly discussed after which geostatistical methods will be discussed in more detail.

2.9.1 Explicit Mapping

Explicit mapping involves taking as many measurements as possible throughout the whole area. A GPS can be used to determine position in large areas, but GPS have very low accuracy indoors. A surveyor's wheel, or 'clickwheel', can be used to determine position in a building, since no indoor localization techniques will be available.

A typical network planning procedure involves placing a transmitter at a temporary position. A surveyor takes measurements and maps them to the corresponding positions. The transmitter is then moved to a different temporary position and the measurement results are compared. This process is repeated until satisfactory results are obtained. A less tedious approach is to divide the area into cells and only take a sample in each cell.

2.9.2 Partition Models

To use partition models, the key obstructions in a building, such as walls and floors, must be identified. A static path loss value is then fitted to each key obstruction by taking measurements. The static path loss value of each object is then generalized to be used in different environments [70].

2.9.3 Iterative Heuristic Refinement

In attempt to find coverage holes in large wireless networks, Robinson et al. [80] combined an a priori model with a fitted partition model. This enabled them to make corrections from measurements.

Robinson's approach requires taking 'pilot' measurements and then applying the model to estimate the signal strength of each AP at a large number of equally spaced points. A signal strength threshold must be specified to consider a point as 'covered' or not.

2.9.4 Active Learning and Geostatistics

Active learning can be seen as a generalization of an iterative refinement process. Neural networks, mixed Gaussian, and locally weighted regression are three types of learning systems [81]. Active learning differs from passive learning in the sense that an algorithm is used to choose training data as opposed to using a set of random observations.

In geostatistics the term optimised sampling relates to active learning. The main idea is that each sample added to a trained model must improve the accuracy of the model. In terms of geostatistics the variance between samples must be minimized.

Geostatistical Methods

Geostatistics is described as a spatial statistical tool that can be employed in any practical problem where predictions of a random variable are needed in a 1D, 2D or 3D application. The basic theoretical framework of geostatistics was developed in the early 1950s mainly for mining purposes to estimate ore reserves. Ever since, this theoretical framework has been applied to other types of applications. Today geostatistics is widely recognized as a tool for accurate spatial estimation [31].

The authors of [31] claim that there is a close correlation between path loss data sampling, acquisition, and estimation and standard problems encountered in ore-body or reservoir forecasting. The goal is to quantify the variability of measurements with respect to the distance between the different measurements. For example, two path loss measurements in close proximity are more likely to have similar values than two path loss measurements further apart.

The first attempt to demonstrate that geostatistical techniques can be effectively and pragmatically applied in the domain of signal strength estimation is presented in [82]. They consider these methods powerful and statistically rigorous and encourage researchers to further investigate this concept when approaching the problem of empirical radio environment mapping.

Kriging is a popular geostatistical method that has been proposed for use in the domain of signal strength estimation [2]. In the next chapter kriging will be discussed more thoroughly.

2.9.5 Conclusion on RSSI Estimation

In literature there are contradictions regarding operational efficiency in terms of effort (ease of use) and computational cost between measurement models and a priori models. In [83] empirical approaches are said to have low accuracy, but that deterministic models become very complex. In [32] the process of measuring is referred to as "the burden of taking samples".

In [32], taking the last 60 years' methods for estimating signal strength into account, it is concluded that whether a network operator does a small random sampling and basic fit, or carefully tunes an a priori model to their environment, they can still expect predictions with low accuracy.

That being said, only preliminary work has been done in terms of applying geostatistical modelling to radio environment predictions [2] and there are still a great deal of open questions on this subject [32].

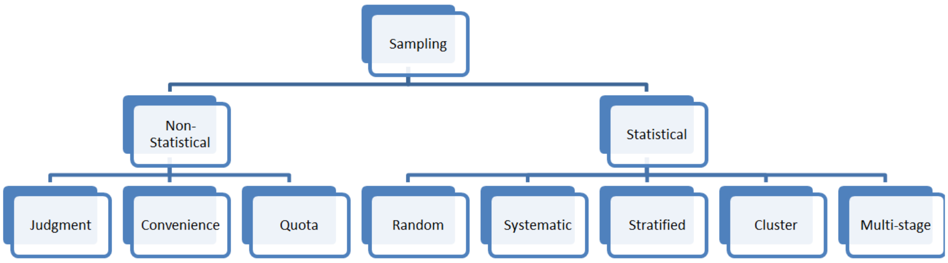

Figure 2.4: Sampling Procedures

2.10 Sampling

Statistics is used to describe a population as an estimation of the corresponding population using only a fraction of the population. In any statistical model, the sampling scheme is crucial to the success of the outcome. Sampling can save a lot of time and reduce cost significantly, but in order to maintain an acceptable level of accuracy the best sampling strategy for the given problem must be applied. Figure 2.4 illustrates the different sampling procedures that can be implemented.

2.10.1 Statistical Sampling

Any sampling procedure that is involved in the laws of probability to calculate sampling risk is considered statistical sampling. This section defines the different statistical sampling procedures and statistical terms regularly used [84] [85] [86].

• Random sampling - Each item has an equal chance of being selected.

• Systematic sampling - One or two items are selected randomly, but the remaining items are selected by adding the average sampling interval to the previous item.

• Stratified sampling - After separating the population into groups, systematic sampling is applied to each cell.

• Cluster sampling - Samples are taken in groups of items in the same area.

• Multi-stage sampling - Cluster sampling is applied, but instead of using all the items in a cluster, items are randomly selected within each cluster.

• Population size - The size of the entire collection from which conclusion needs to be drawn.

• Sample size - The amount of samples taken from the population.

• Sampling fraction - The ratio between sample size and population size. A higher sampling fraction will result in higher accuracy, but will increase the time and cost needed to do an investigation.

2.10.2 Geostatistical Sampling

The concept of spatial dependence is based on the notion that observations made in close proximity to each other are more likely to be similar than observations separated by large distances. In order to apply a spatial prediction model, a spatially dependent data set needs to be constructed. This allows for prediction of the observed metric at other locations, where samples were not taken, on the basis of their position relative to the actual observations [87].

Estimations are required in a wide variety of scenarios and at different landscapes. This prevents a single standard sampling plan, for geostatistics in general, to be established. Though there are some corresponding concepts to guide the development of a sampling plan. The effectiveness of a sampling plan depends on the spatial variability of the metric being measured. Since the spatial variability is initially unknown, conventional, single phase, sampling plans are usually inefficient. Several phases must be incorporated in designing a sampling plan where later phases include information obtained from previous phases about the spatial variability of the metric that is sampled [88].

The input data to any model needs to be of acceptable quality for the model to be able to provide acceptable results. The authors of [89] specifically stress the fact that even the most sophisticated geostatistical tools cannot save data sets of poor quality. In their practical guide to geostatistical mapping they also provide a list of questions that need to be answered in order to evaluate the input data [87]:

• Is it large enough - Statistical analysis requires large enough data sets for the given population. Reliability of a variogram model decreases significantly as n approaches small numbers.

• Is it representative - The collected samples need to represent the area of interest. The geographical coverage and the diversity of the environmental features must be considered. Samples must be both representative of the geographical area of interest as well as the range of the values measured in that area.

• Is it independent - An objective sampling technique needs to be used when selecting samples. The locations must be chosen in an unbiased way so that no special preference is given to certain locations. Recommended sampling designs for selecting independent point locations include simple random sampling, regular sampling, and stratified random sampling.

• Is it produced using a consistent methodology - A consistent methodology for field sampling and laboratory analysis must be established and described in detail. The methods described must be practical and reproducible.

• With what precision was it measured - Field measurements need to be more precise than the natural variation of the variables that are measured.

Geostatistical mapping using either small data sets, inconsistent point samples, or subjectively selected samples can lead to unreliable estimates of the model in parts or in the whole area of interest. Repetition of a mapping project should be considered if the prediction error of the output map exceeds the total variance in more than 50% of the study area.

2.10.3 Evaluating the Quality of a Sampling Plan

For each sample point the clustering of points can be evaluated by comparing the sampling plan with a random design. The level to which the samples represent the full data set can be evaluated in both geographical and feature space by means of a histogram comparison and the consistency of the sampling intensity needs to be evaluated.

Sampling Strategies

Two main groups that are commonly utilized for environmental mapping are [87]:

Regular Sampling Regular sampling has the advantage of systematically covering the whole area of interest which minimizes the overall prediction variance. It has the disadvantage of misrepresenting distances smaller than the grid size.

Randomized Sampling Randomized sampling has the advantage of representing all distances between points which is beneficial for variogram estimation. The drawback of this technique is that it has a lower spreading of points in geographical space than regular sampling. This causes the overall precision of the final maps to be lower.

Both these sampling strategies belong to the group of design-based sampling. In [89] a combination of the two strategies is recommended since none of them are universally applicable. They suggest obtaining half the points using regular sampling and obtaining the other half of the points with randomised sampling.

Smarter allocation of points can also greatly increase accuracy and reduce survey costs by minimizing the sampling points. An approach might be to spread samples around extremes of the feature space and maximize their spreading in the area of interest. The number of sampling points is mainly dictated by the precision requirements as more accurate and detailed maps require higher sampling densities.

3 Kriging

Tobler's first law of geography states that all places are related, but nearby places are more related than distant places [21]. This usually refers to natural phenomena such as the occurrence of minerals, elevation and rainfall. However, this analogy can also be applied to signal strength and in this case it will be used to estimate signal strength in an indoor environment. The estimation must be done by taking a limited number of measurements and using the kriging interpolation method to calculate the signal strength at other positions in the building.

3.1 The Origin of Kriging

In 1951, a South African Mining Engineer, D.G. Krige, published a seminal paper in the Journal of the Chemical, Metallurgical and Mining Society of South Africa, where he pursued a statistical explanation of the conditional biases in ore block valuations. This formed the basis of the interpolation method known today as kriging [1].

Kriging is a geostatistical interpolation technique used as a tool in computer software solutions. It is a linear weighted-averaging method, but differs from inverse weighted distance methods by depending on a model of spatial correlation. The spatial correlation model is known as the variogram model and is used to estimate the variability of an attribute as a function of its distance from neighbouring points. The value of an unknown data point is calculated by using the weights and the values from neighbouring points. Kriging eliminates bias by accounting for the spatial correlation of neighbouring points in addition to their distances from the interpolated point [31].

3.2 Types of Kriging

Literature distinguishes between three main types of kriging: simple kriging, ordinary kriging, and universal kriging [90]. This section will discuss the principles and differences between the three types of kriging.

Kriging is a mathematical method based on statistical principles to model quantities in a geographical region. Kriging differs from the deterministic models described above by including statistical probability in the method. Since probability is associated with the predictions made by kriging the values are not predicted perfectly even with large numbers of samples. However the method does accommodate the assessment of the error of each prediction.

Kriging is based on the concept of autocorrelation and the basic principle of geography that nearby places are more similar than distant places. The rate at which correlation decreases can be expressed as a function of the distance between the points of interest. This is contrasting with classical statistics which assumes that there is no correlation between observations. With geostatistics the information is located at specific locations which enable one to calculate the distance between them. This allows an autocorrelation model to be constructed as a function of distance.

As the correlation between data differs as a function of distance it forms a trend. The following formula expresses the trend that will be taken into account when deciding on a kriging method:

$$Z(s) = \mu(s) + \epsilon(s), \tag{3.1}$$

In equation (3.1) $Z(s)$ is the value of the new point to be interpolated in the point s. The deterministic trend is indicated by $\mu(s)$ and a random autocorrelated error is added as $\epsilon(s)$. All the different types of kriging are variations of this formula. The error is expected to be zero on average and the autocorrelation between two errors depends on the distance between them rather than the actual position s.

The trend can either be constant, where $\mu(s) = m$ for all locations s, or it can be composed of a linear function based on spatial coordinates:

$$\mu(s) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy \tag{3.2}$$

If $\mu(s)$ is constant with an unknown value, the model forms the basis of ordinary kriging. A linear function with unknown regression coefficients form the basis of universal kriging. When the trend is completely known, it forms the model for simple kriging [90].

3.2.1 Simple Kriging

Simple kriging assumes the model:

$$Z(s) = \mu(s) + \epsilon(s), \tag{3.3}$$

where μ is a known constant [91].

3.2.2 Ordinary Kriging

Ordinary kriging assumes the model:

$$Z(s) = \mu(s) + \epsilon(s), \tag{3.4}$$

where μ is an unknown constant. The issue here is whether it is reasonable to assume what the constant of the mean is [92].

3.2.3 Universal Kriging

Universal kriging assumes the model:

$$Z(s) = \mu(s) + \epsilon(s), \tag{3.5}$$

where $\mu(s)$ is a deterministic function. Universal kriging is also known as kriging with a trend which is usually represented by a polynomial. The error, $\epsilon(s)$, is the difference between the measured data and the polynomial with a mean of 0 [93].

3.3 Kriging Algorithm

Kriging can be divided into three main types: simple kriging, ordinary kriging, and universal kriging. Cross validation of initial empirical investigations showed that universal kriging gave the best and most consistent results for the problem scenario presented here. This section will briefly explain how universal kriging is applied.

The first step in universal kriging is to record a scatter point set to be interpolated and to construct an experimental variogram from the data [94]. The experimental variogram is used to construct a model variogram which will be used to determine the weights used in kriging. A variogram is a representation of the variance in data as a function of the distance between samples.

The general formula used in universal kriging to interpolate a value F in point (x, y) is shown in equation (3.6) [94].

$$F(x,y) = \sum_{i=1}^{n} w_i f_i \tag{3.6}$$

In equation (3.6), n is the number of scatter points used, w_i is the weight of each scatter point with each $w_i < 1$ and $\sum_{i=1}^{n} w_i = 1$ and f_i is the corresponding value of the scatter point.

The weights are determined using the model variogram and solving the matrix in 3.7:

$$
\begin{bmatrix}
w_1S(d_{11}) & w_2S(d_{12}) & \dots 1 \; x_1 \; y_1 \\
w_1S(d_{21}) & w_2S(d_{22}) & \dots 1 \; x_2 \; y_2 \\
w_1S(d_{31}) & w_2S(d_{32}) & \dots 1 \; x_3 \; y_3 \\
\vdots & \vdots & \vdots \\
1 & 1 & \dots 0 \; 0 \; \; 0 \\
x_1 & x_2 & \dots 0 \; 0 \; \; 0 \\
y_1 & y_2 & \dots 0 \; 0 \; \; 0
\end{bmatrix}
\begin{bmatrix}
w_1 \\ w_2 \\ w_3 \\ \vdots \\ \lambda \\ \alpha_1 \\ \alpha_2
\end{bmatrix}
=
\begin{bmatrix}
S(d_{1p}) \\ S(d_{2p}) \\ S(d_{3p}) \\ \vdots \\ 1.0 \\ x_p \\ y_p
\end{bmatrix}
\tag{3.7}
$$

In the matrix above, $S(d_{nm})$ represents the model variogram value at the distance that samples n and m are from each other and $S(d_{np})$ represents the model variogram value at the distance between samples and the new interpolated point. Variables w_1 .. w_n represent the weighting of each sample when calculating the value of the new interpolated point. The constraint of the weights summing to 1.0 introduces one more equation than there are variables. In order to solve this matrix, the variable λ is added to obtain a unique solution and is called a Lagrange multiplier, used to minimize the possible estimation error. The variables α_1 and α_2 are the local trend coefficients of the first order trend. Including the x and y coordinates in the matrix is unique to universal kriging and is used when a trend is present in the data.

Solving (3.7) for w_1 .. w_n enables one to calculate from (3.6) the specific interpolation point f_p as in (3.8).

$$
f_p = w_1 f_1 + w_2 f_2 + w_3 f_3 + \cdots + w_n f_n \tag{3.8}
$$

3.4 Characterising a Variogram

A semivariogram, also referred to as a variogram, is a representation of how much spatial locations relate as a function of the distance between them. The difference between a variogram and a semivariogram is that the semivariance equation is the variance divided by two. Since this has no effect on the use of the kriging algorithm in this case, this study will not distinguish between the two.

Figure 3.1 shows the characteristics of a variogram and Figure 3.2 shows the suitable variogram models as explained in the following subsections:

3.4.1 Characteristics of a Variogram

The following characteristics are considered when interpreting a variogram [19]:

Sill

A sill is the value of a variogram at which the variance levels off. This is the value on the y-axis at which the autocorrelation is essentially zero.

Range

The range indicates the lag distance at which the variogram reaches the sill value. This implies that this is the distance at which the autocorrelation becomes zero.

Nugget

In theory the value of the variogram at the origin should be zero. If lag values close to zero have large variances, then this variogram value is referred to as the nugget. The nugget represents the variance at distances smaller that the sampling distance, including the typical error of measurements.

3.4.2 Modelling a Variogram

In order to use an experimental variogram for practical purposes, it needs to be replaced by a model variogram. The following shows the most frequently used variogram models, where h represents the lag distance, a represents the range, and c represents the sill:

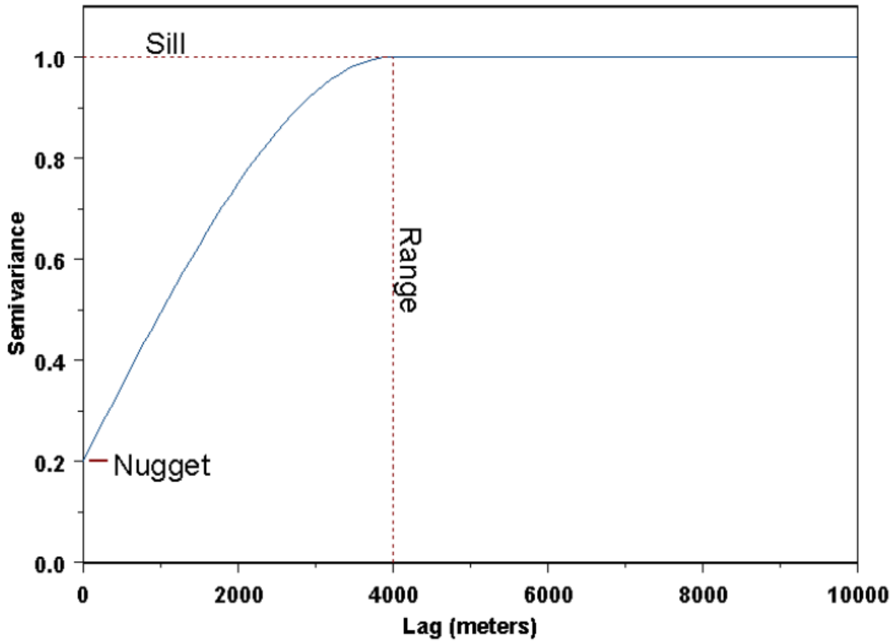

Figure 3.1: Characteristics of a Variogram [19]

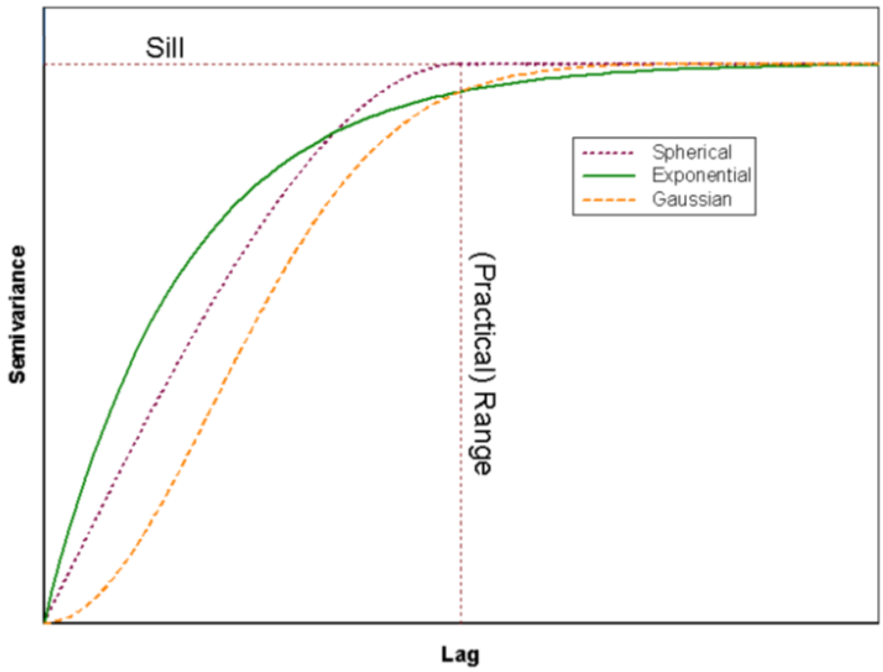

Figure 3.2: Suitable Variogram Models [19]

Nugget

$$g(h) = \begin{cases} 0, & if h = 0 \\ c, & otherwise \end{cases} \tag{3.9}$$

Spherical

$$g(h) = \begin{cases} c(1.5(\frac{h}{a} - 0.5\frac{h}{a}^3)), & if h \leq a \\ c, & otherwise \end{cases} \tag{3.10}$$

Exponential

$$g(h) = c(1 - \exp(\frac{-3h}{a})) \tag{3.11}$$

Gaussian

$$g(h) = c(1 - \exp(\frac{-3h^3}{a^2})) \tag{3.12}$$

Power

$$g(h) = ch^{\omega}, 0 < \omega < 2 \tag{3.13}$$

3.5 Chapter Conclusion

In this chapter we explained and discussed the critical building blocks of kriging that will form the basis of the investigations specified for this study. The concepts and algorithms explained here will be implemented in MATLAB before experiments can be done.

Spatial interpolation is done by assigning weights to samples according to their distances from the position of the interpolated point. These weights are a measure of how much the value of each sample is related to the value of the interpolated point. This is in line with Tobler's first law of geography [21].

The variance between samples as a function of their distance from each other are indicated by constructing an experimental variogram. The experimental variogram is replaced by fitting a model variogram. From the model variogram, a set of linear equations are derived and solved to obtain the weight of each sample. These weights then determine how much influence the value of each sample will have on the value of the interpolated point.

Wi-Fi signal strength can be viewed as a geographical entity that, in an indoor environment, occurs randomly yet spatially correlated. Since kriging is an established interpolation method designed specifically for estimating spatially correlated geographical entities, it is a suitable candidate for estimating signal strength.

4 Preliminary Investigations

In order to become familiar with using kriging, some preliminary investigations were done. This chapter shows the experiments and results that were done in a simple indoor environment using a single access point. Various unconventional model variograms are compared and evaluated.

4.1 Preliminary Experimental Setup

An omnidirectional Wi-Fi hotspot was set up in a building and placed at a point (0,0). Signal strength was measured at 60 random positions throughout the building. The number of measurements was influenced by the size of the building and the sample density. Since no significant change in Wi-Fi signal strength could be measured between positions about 2.5 m apart in the same room, samples were taken approximately 2 m apart. This allowed the values of the measurements to represent the environment without having to take an infinite number of samples throughout the area.

The signal strength at each position was determined by taking four measurements, about 20 cm apart. The average value of the four measurements was calculated to represent the value of each sample [7]. This was done in an attempt to minimise the effect of local interference patterns that can influence the samples.

The goal was to eliminate a number of samples from the data set and to estimate their values by applying the kriging interpolation method to the remaining samples. The interpolated samples were then compared to the original measured samples to determine the accuracy of kriging for RSSI estimation.

4.1.1 Evaluating Consistency

In order to determine the consistency of the method for this application it first had to be tested in a number of different scenarios. The most practical approach for determining the consistency was to do cross validation [8].

The consistency was evaluated by choosing 40 random samples of the 60 originally measured samples to represent a test case. The 40 samples were then divided into two subsets of 20 samples each. One subset of 20 samples was used as known samples to interpolate the remaining 20 unknown samples. Choosing 40 samples from the total number of 60 samples allows for a number of unique subsets in excess of 4×10^{15}, but 100 repetitions were sufficient to practically illustrate the consistency when using random samples.

4.1.2 Choosing a Model Variogram

A number of models are available when constructing a model variogram and it is important to choose the model that is most representative of the given data [9]. To better understand the role of a variogram in practice, a number of unconventional model variograms were tested on the data gathered for this investigation. This was done instead of using conventional model variograms in a black box approach.

We proposed six different candidate model variogram curve fittings to be fitted to the experimental variogram. Their results were compared in order to find the curve fitting that delivered the most consistent and accurate results.

A linear and a logarithmic relation were tested as candidate model variograms. These model variograms do not correlate well in this scenario since distant positions are also strongly related as a result of low signal strength far away from the Wi-Fi access point in all directions.

The candidate model variogram curve fittings that were considered for experimentation were a parabola [95], a 4th degree polynomial, a piecewise linear curve fitting [96], and a linear combination of the three. The accuracy obtained using these curve fittings will be discussed in the results section.

4.1.3 Evaluating Accuracy

To address the goal, one must consider that in a real scenario, a human will select samples to represent the full data set. The final test was to determine the minimum number of samples necessary to obtain an accurate estimation when a human selects the samples.

After evaluating the consistency of the method for different model variograms, a sensible level of accuracy of which the method is capable of, could be identified. This level was used as a benchmark for this building to determine the minimum number of samples necessary to maintain the same level of accuracy.

Figure 4.1 shows a 3D view of the distribution of the Wi-Fi signal strength throughout the area where the data was sampled.

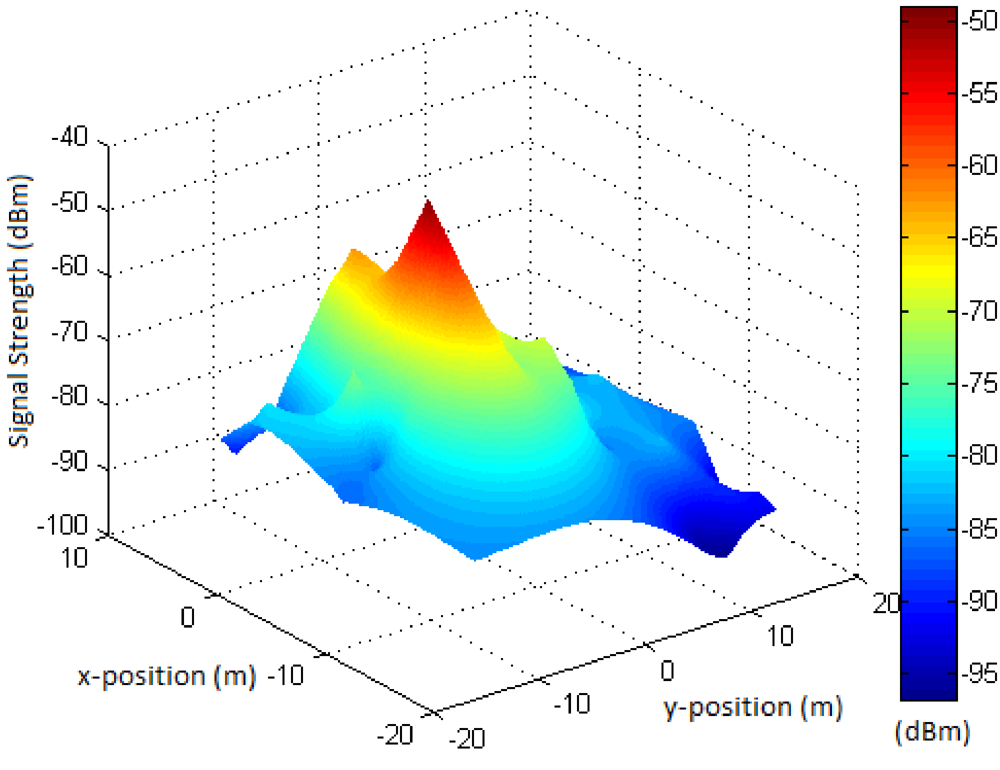

Figure 4.1: Measured Signal Strength

4.2 Results

The results of the random samples will be shown first, demonstrating the consistency of the method, followed by the results obtained from the hand picked samples.

4.2.1 Consistency

Figure 4.2 shows the average error of each instance where a random subset of 20 samples were used to interpolate another random subset of 20 unknown samples. Since the samples were selected randomly, the larger errors occurred when the subset of known samples were locally grouped. In these cases they were not representative of the full data set.

If we assume that in practice there is some intelligence behind the sample selection process, these cases can be considered as outliers. An average accuracy of 87% was calculated for this experiment with these outliers included.

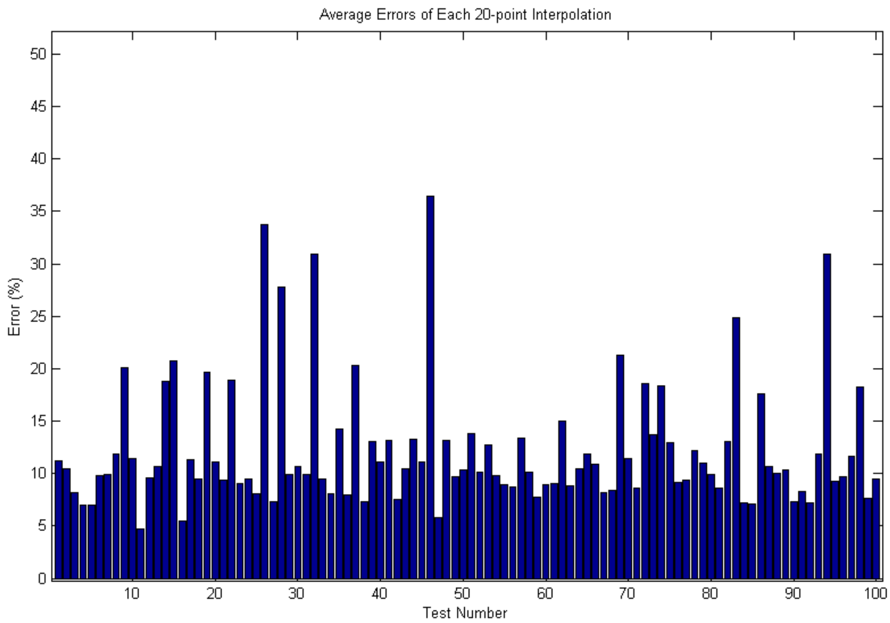

Figure 4.2: Average Errors of each 20-point interpolation

4.2.2 Choosing a Model Variogram

Figure 4.2 resulted from using the combination curve since it provided the most consistent results over a series of tests. An example of a combined curve fitting used as a model variogram is shown in Figure 4.3.

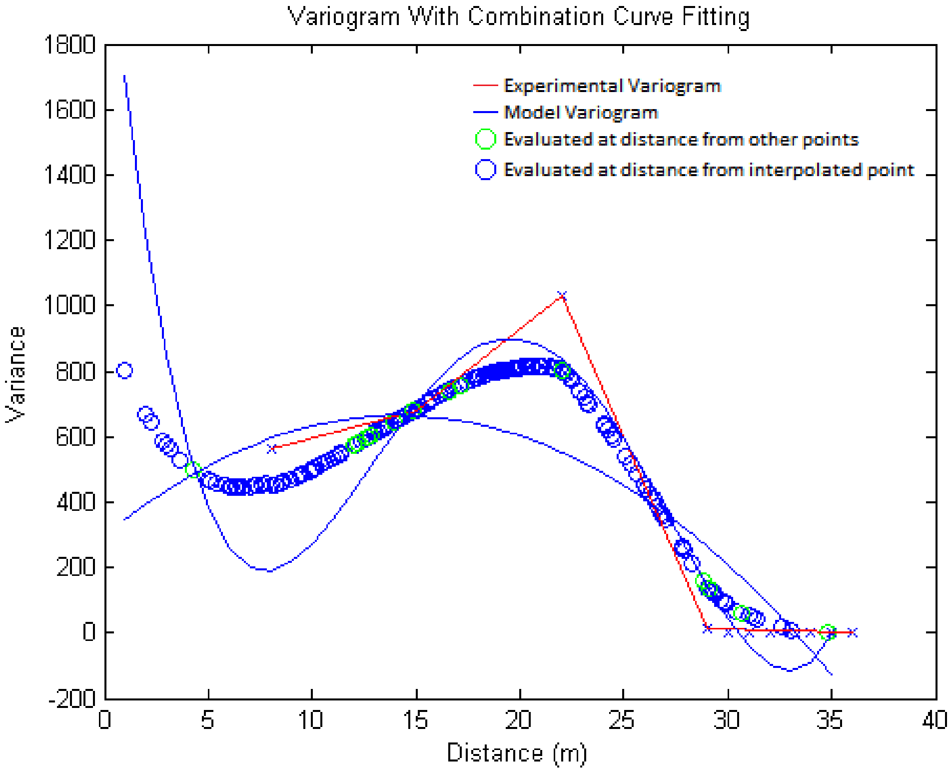

Figure 4.3: Combination curve used as model variogram

Even though the parabola did not result in the most consistent estimations, in some instances it resulted in the most accurate estimations. In general, the combined curve stretches the polynomial shapes towards the piecewise defined linear curve. This tends to lower the maximum error when compared to the other individual curves over a number of runs, but in cases where spatially distributed samples were selected, the parabola outperformed the other model variograms. An example of a parabola used as a model variogram is shown in Figure 4.4.

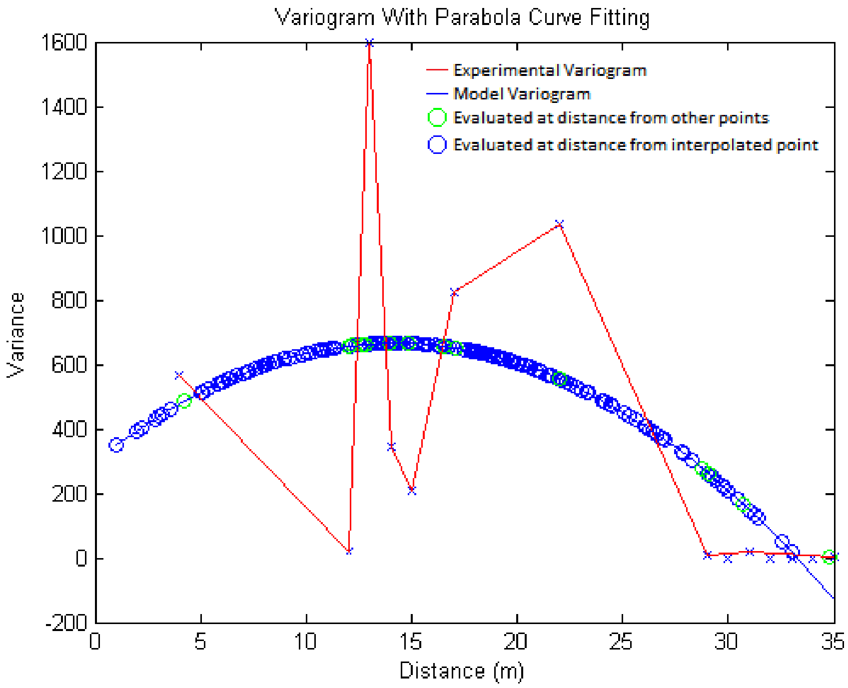

Figure 4.4: Parabola used as model variogram

Figure 4.5 shows a comparison between the measured values of the samples and the estimated values of the samples arranged by the size of the measured values. These estimations were obtained by choosing spatially distributed samples and using a parabola as model variogram.

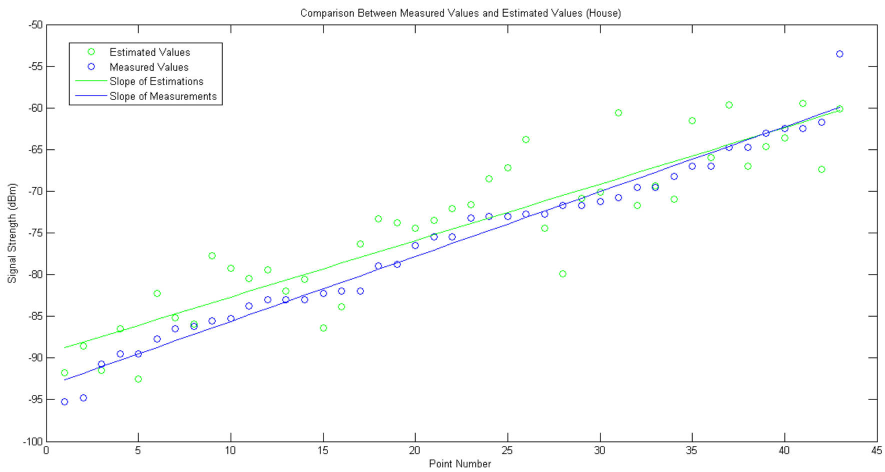

Figure 4.5: Comparison between interpolated points and measured points

4.2.3 Accuracy

Evaluation of the consistency of kriging for RSSI estimation resulted in an average accuracy of 87%. Furthermore, the comparison between the different model variograms showed that spatially distributed samples with a parabola as model variogram can provide a higher level of accuracy. Therefore, by using a parabola as model variogram and selecting spatially distributed samples, an accuracy of 90% could be achieved using 6 samples to estimate the remaining 54 samples.

The error percentage ranged from 0.7% to 29% with an average error of 10%. Therefore, as few as six spatially distributed samples are sufficient to describe the interpolated area with an accuracy of 90%. The samples used to obtain these results are shown in blue in Figure 4.6.

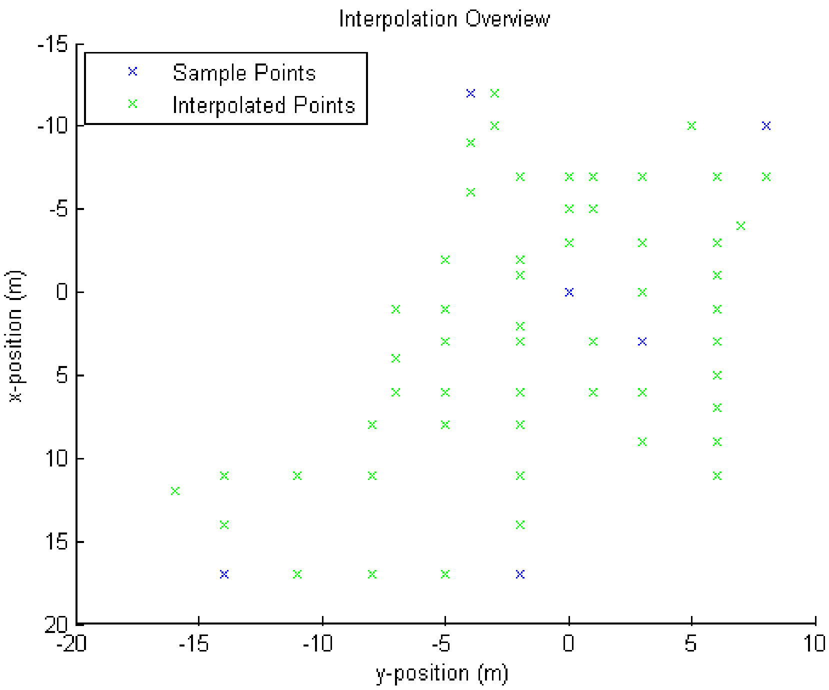

Figure 4.6: Comparison between interpolated points and measured points

4.3 Chapter Conclusion

The results indicated that the accuracy provided by kriging was satisfactory for single access point Wi-Fi signal strength estimation. Kriging is recommended for further investigation in more complex environments.

Using a suitable model variogram is important and has a significant influence on the accuracy of the method. In this investigation using a parabola as model variogram provided the most accurate results. However, this will only be a suitable model variogram for environments with a single access point. In an environment with a single access point the most distant samples all have low signal strength. This makes those samples just as related as nearby samples which made the parabola a suitable model variogram for this investigation.

If, for example, elevation needs to be interpolated in a geographical area, there will not be only one mountain. Likewise in complex indoor environment where signal strength estimation needs to be done there will be multiple Wi-Fi access points. This suggests that the conventional model variograms from [9] should be used. In the next investigation the most suitable model variogram from [9] will be used in a multiple access point environment. The benefit of using different variogram models in this investigation was to obtain a better understanding of how they work in practice.

For this investigation, a simple sampling method was followed by taking the nearest and furthest points with a few scattered points in between. The results of the cross validation using random points of the measured data showed that the method can provide consistent results, but a more formal sampling plan needs to be put in place in future investigations.

Randomly selecting samples resulted in an average accuracy of 87%, but with the possibility of significantly lower accuracies in worst case scenarios. An accuracy of 90% can be achieved with as few as six prudently selected samples that is spatially distributed.

The six samples were empirically selected, but a more formal analysis is needed to determine the minimum number of samples necessary for a given level of accuracy. It is also important to find a way to determine the convergence of the level of accuracy while taking samples, to prevent unnecessarily measuring the environment throughout.

5 Complex Indoor Environments

For this investigation, two different Wi-Fi environments were surveyed. The steps taken in the most complex environment will be discussed in detail after which differences in the remaining environment will be pointed out.

5.1 Experimental Setup

Wi-Fi signal strength, in the form of RSSI expressed in dBm, was measured at 110 random positions throughout the publicly accessible areas of the main building at the NWU engineering campus. The building is three storeys high, approximately 20 m in width, 80 m in length and has open volumes shared for three floors in two areas. Figures 5.1, 5.2 and 5.3 shows the floor plan of the building.

Each floor has two Wi-Fi access points and the top floor has a third access point. All of these access points enable a Wi-Fi device to connect to the same network. Seamless handover is enabled for devices moving from one access point to another.

RSSI was sampled by taking four measurements, about 20 cm apart, per sample. The average value of the four samples was calculated to represent the value of a sample point [97]. This was done in an attempt to minimise the effect of local interference patterns that can influence the samples.

Even though the RSSI of the different access points were available throughout the building, the highest RSSI value of the SSID of concern was sampled. Consequently, the sample represents the signal strength that a roaming Wi-Fi device would be receiving at that position while connected to the network using that specific SSID.

Accuracy and consistency in the results obtained by using kriging are two important

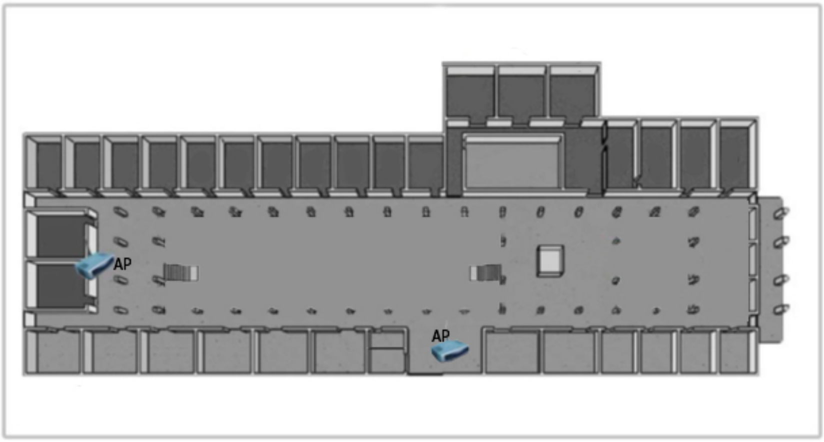

Figure 5.1: Engineering building - Ground Floor

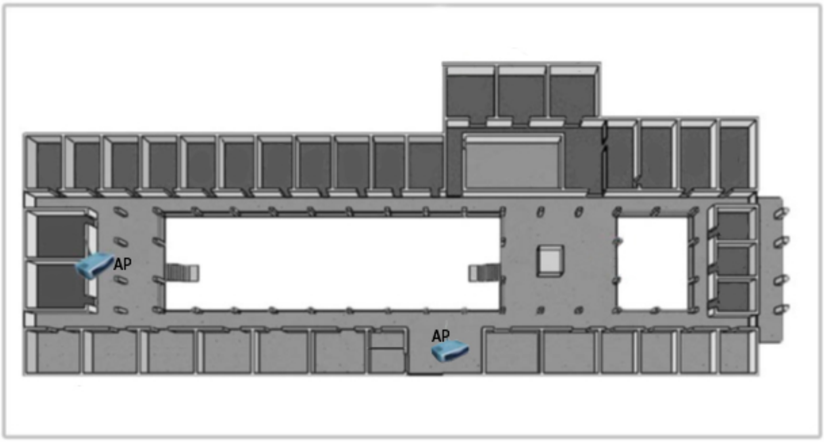

Figure 5.2: Engineering building - First Floor

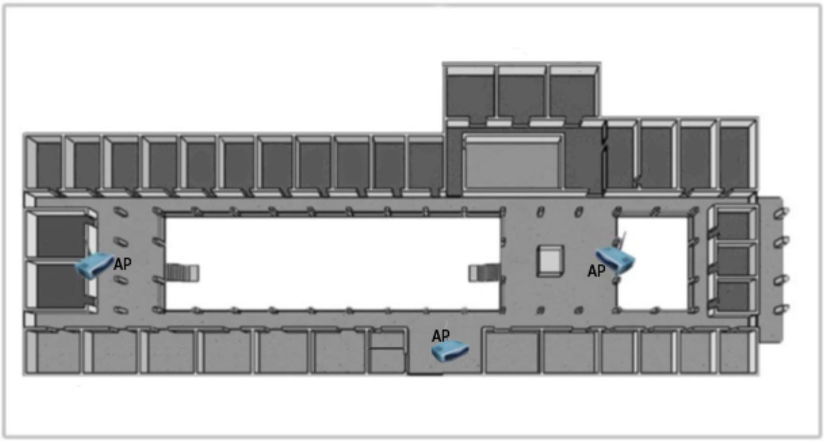

Figure 5.3: Engineering building - Second Floor

aspects of this investigation. The accuracy and the consistency of the results were evaluated through cross validation [98].

A subset of the 110 measurements was considered to be known samples that a surveyor would have measured. The remaining measurements were considered to be unknown and had to be estimated using universal kriging.

For this investigation the model variograms as suggested by [9] were used. The results from the preliminary investigation were obtained in an environment with only a single access point and were found not to be valid in environments with multiple access points. The exponential model variogram provided the most accurate and consistent results and was used to obtain the results presented in this chapter.

The set of known samples was chosen randomly and used as input values for the universal kriging algorithm to estimate the values of the remaining samples that was considered unknown. The consistency of the results could be assessed by comparing the estimated values with the original measurements that were taken at corresponding positions and repeating the process, each time using a different set of known samples within the 110 measurements.

From the preliminary investigations it was found that the number of known samples, used as inputs, has a significant influence on the accuracy of the interpolated results. For this investigation the process of using a random subset, with a fixed size, to interpolate remaining points was repeated while increasing the number of known samples with each repetition. This was done in attempt to find the correlation between the number of samples used and the accuracy of the results.

The correlation between the number of samples and the accuracy was examined in order to determine a favourable number of samples required to represent the given environment. Using fewer samples than the favourable number will result in poor accuracy. In contrast, using more samples than the favourable number, will not produce a significant improvement on the accuracy of the model.

The next step of this investigation was to examine the distribution error made in the interpolated points when using the favourable number of samples suggested by the results of the previous process. This assisted in determining the level of confidence with which kriging can be used to estimate Wi-Fi signal strength.

5.2 Coverage Maps

The final step of this investigation was to construct a Wi-Fi coverage map of the environment, using the results obtained from the previous steps. In order to verify that a reliable coverage map is constructed it has to be compared to a coverage map that is known to be accurate.

Since a heat map or contour plot needs continuous data which is unrealistic to measure, interpolation is still needed to construct a coverage map. From the previous steps the favourable number of samples needed to represent an area, and a confidence level at which this number of samples can be used to interpolate signal strength at other positions, was determined. Consequently, the full set of 110 measurements can be considered as an oversampled set that will represent the area with a confidence level that is larger than the confidence level achieved with the favourable number of samples.

The 110 measurements were used as inputs for the universal kriging algorithm to calculate nearly continuous values for the signal strength throughout the building. These values were used to construct a coverage map for each floor in the building. Coverage maps were constructed separately for each floor in order to overcome the difficulty of visually presenting spatial data of a 3D area. The coverage maps were regarded as the benchmark to which coverage maps, constructed using the favourable number of samples, could be compared.

5.3 Comparison With Previous Environment

We returned to the same building where the preliminary investigation was done, but in this investigation the same procedures were followed as in the main building of the engineering campus. A four bedroom residential house with only a single Wi-Fi access point were surveyed.

A total of 60 measurements were taken throughout the house and in the yard and the favourable number of samples were determined for this building. This was done in order to compare the results obtained from the complex office environment to the results obtained from a less complex environment while following the same approach.

5.4 Results

The results of investigations explained above are presented and discussed in this section. To illustrate the measurements that were taken throughout the main building of the engineering campus, only the measurements of the ground floor will be shown here. Since the access points on the other floors are placed in the same (x, y) positions, the distribution looks very similar to that of the ground floor. Figure 5.4 shows a 3D view of the distribution of the Wi-Fi signal strength on the ground floor of the building.

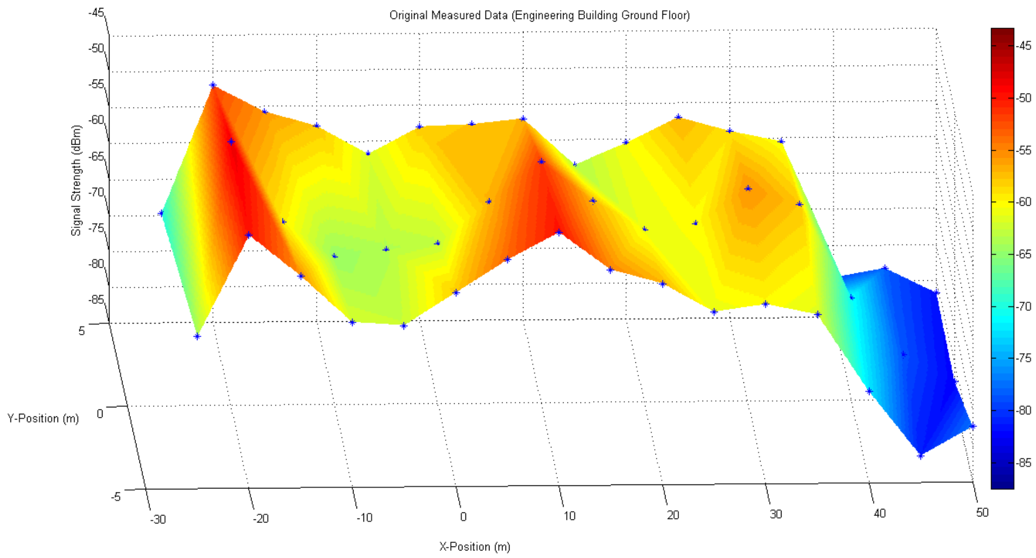

Figure 5.4: Measured Wi-Fi signal strength at the ground floor

Figure 5.5 shows how the average error decreases for each instance where a larger random subset of the 110 measurements are used to interpolate the remaining points. Since the points used as samples are randomly chosen, the average error is influenced. This can be due to random samples that are locally grouped. These selections of samples are not representative of the full data set. If we assume that in practice there will be some intelligence behind the sample selection process, cases where the average error is very high as a result of locally grouped samples can be considered as outliers.

In Figure 5.5 the blue line indicates the average error. The green line represents the standard deviation and the red lines show the upper and lower bounds of the error. The average accuracy of the interpolated points increases as the number of samples increases up to a point at around 23 samples where it approaches an asymptote. This implies that a sample size of more than 23 samples in this environment will not have a significant improvement on the accuracy of the method. At this point the average error is 4.68 dB with a standard deviation of 4.08.

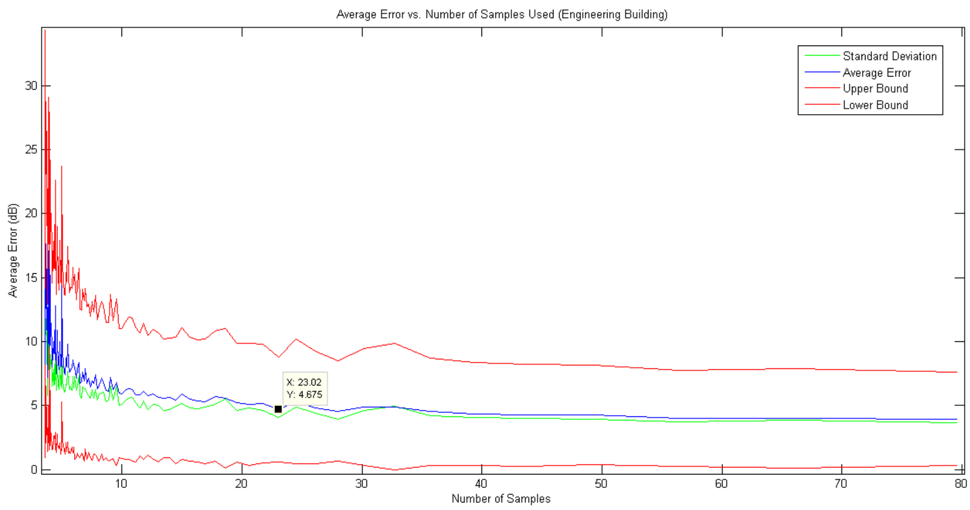

Figure 5.5: Average Error vs. Number of Samples - Engineering Building

Figure 5.6 shows a histogram of the distribution of the errors when using the recommended number of samples for the given environment. This is a strong indication that most of the interpolated points are interpolated with acceptable accuracy when compared to the accuracy of the models listed in Table 2.6 in the model comparison section. Even though some interpolated points can have low accuracy, there are very few points with high errors when compared to the number of points with higher accuracy.

The distribution shown in Figure 5.6 further suggest that these interpolated points will be suitable for constructing coverage maps. The purpose of a coverage map is to show the distribution of signal strength through a given area. The few high error points will have very little impact on coverage maps when compared to the majority of low error points.

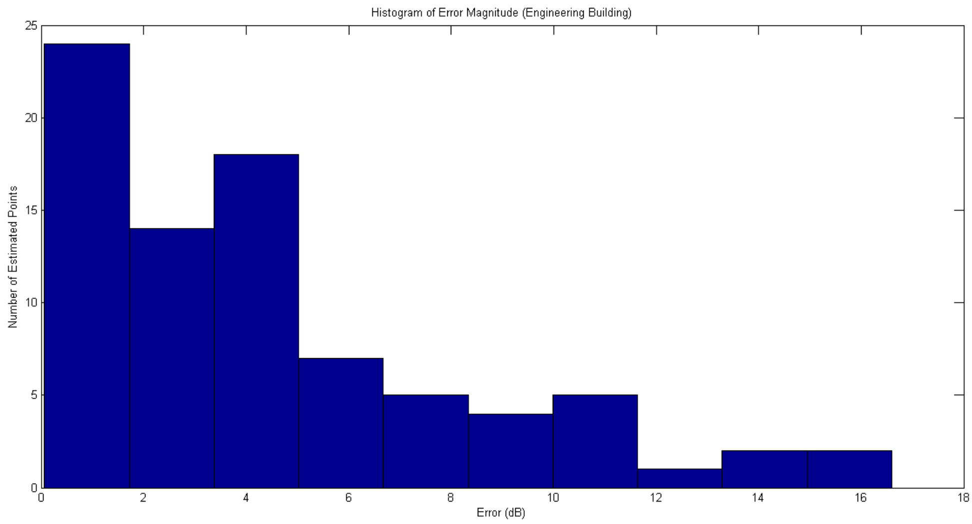

Figure 5.6: Histogram of Error Magnitude

In Figure 5.7 a coverage map, constructed with data from the recommended number of samples from Figure 5.5, is compared to a coverage map constructed with data from the full measurement survey of the building. These are coverage maps of the ground floor and can be compared to the signal strength distribution in Figure 5.4. For this building the coverage maps of the first and second floors look very similar to the ground floor since the access points on each floor are in the same (x, y) positions. The red zones indicate the high signal strength close to the two access points on the ground floor. A third zone that indicates high signal strength is a result of the third access point on the second floor that is shared for the three floors by an open volume. The blue zones are outside the building, moving away from all access points.

The two buildings that were surveyed provided remarkably similar results. The most accurate coverage map resulted from the single access point survey that was done in a house. A comparison between the coverage maps constructed from the fully measured survey and the limited samples is shown in Figure 5.8. The number of samples used was determined by a graph similar to the graph in Figure 5.5.

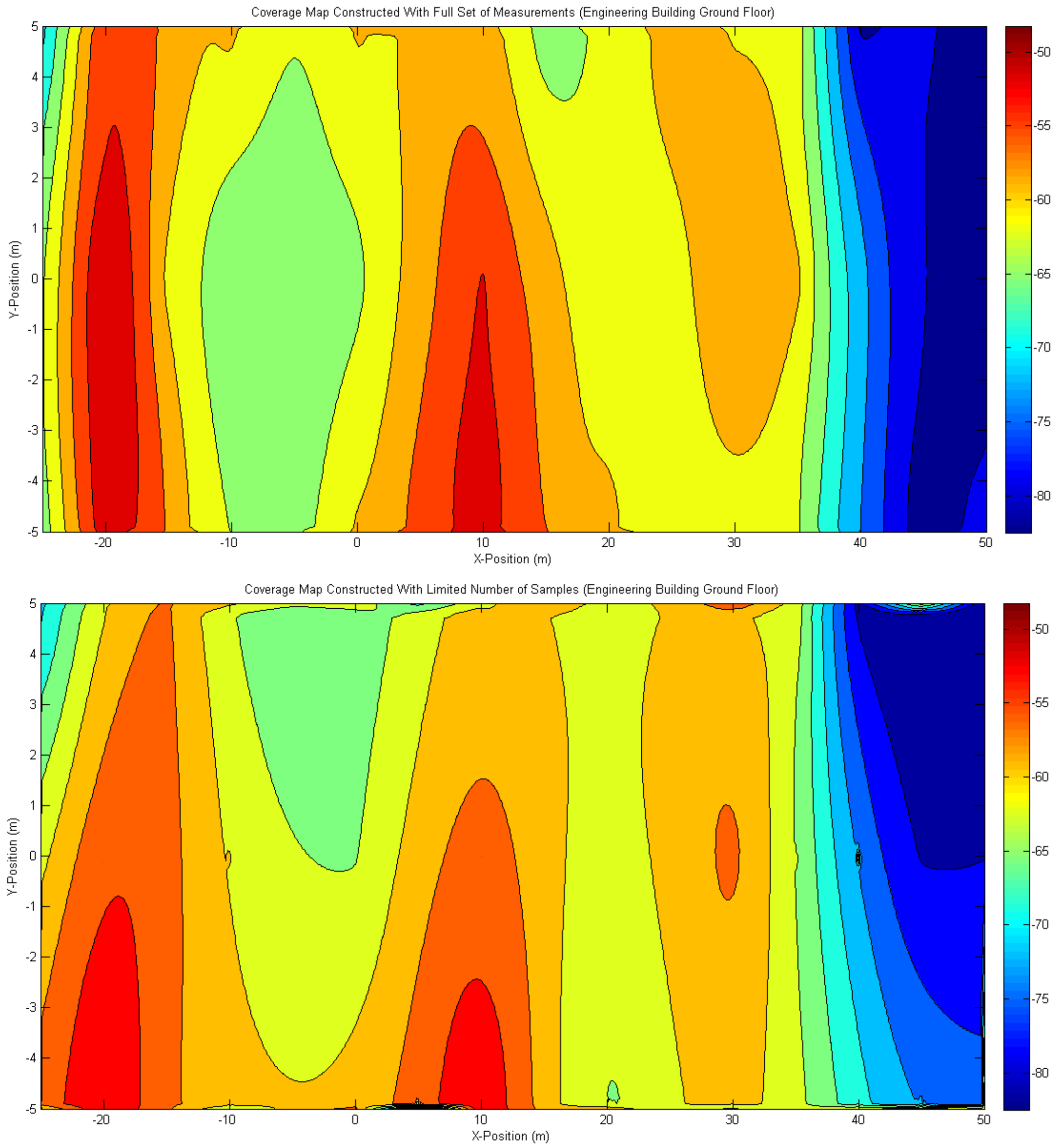

Figure 5.7: Coverage Map Comparison - Engineering Building

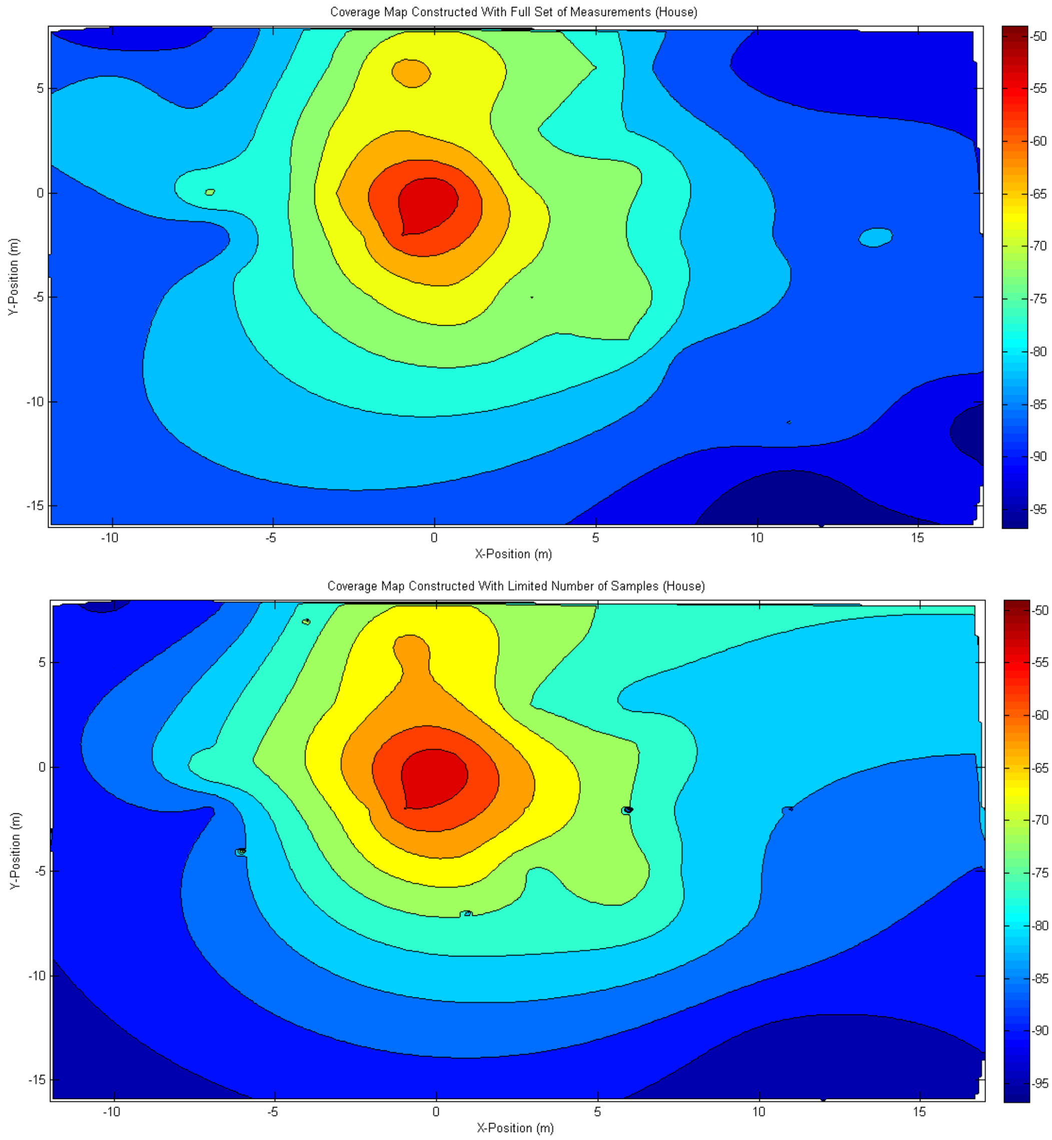

Figure 5.8: Coverage Map Comparison - House

5.5 Chapter Conclusion

This investigation provided an indication of the value that kriging can add to the field of signal strength estimation. During this investigation the data from the fully surveyed building was available at all times and could be used as a benchmark for the interpolated results. When estimating signal strength in real scenarios, this will not be the case.

A way to determine the number of samples needed to obtain acceptable interpolation results without any prior information has to be specified. This specification is required before kriging can be considered a valid method for practical applications.

The results shown in this chapter were obtained by using a simple sampling strategy. All samples were chosen randomly. It is, however, important that samples be representative of both the area of interest as well as the range of the values present in that area.

In an area where the minimum number of samples needed to represent the environment is unknown, we propose an iterative approach for taking samples. Each sample in a set of samples can be estimated using the rest of the samples in order to do cross validation. The sample size can then be increased until an acceptable level of accuracy is achieved.

Our next investigation will focus on testing our proposed sampling method and applying kriging at another building where no prior information is available about the building. This will also provide a better perspective on the practical use of kriging for signal strength estimation in real scenarios.

6 Sampling and Validation

In real life applications, a surveyor needs to know the favourable number of samples necessary to represent an environment without the luxury of having a full sample set that can be used to verify the results. In this chapter we introduce an iterative sampling plan that is based on the idea of the cross validation done in the previous chapter.

The proposed iterative sampling plan is used at a different building in order to validate the results obtained in the previous chapter. The proposed sampling plan will serve as a generic application protocol for using kriging for RSSI estimation.

The results of this chapter will indicate whether the accuracy of kriging for Wi-Fi signal strength estimation was achieved by chance in the previous building or whether kriging in different buildings will provide accuracy of the same level.

6.1 Sampling

From the sampling section in chapter two it was concluded that two main factors need to be considered when sampling. The samples must be representative of the range of values present in the environment and they must be representative of the area of concern.

Furthermore, from the theory of the variogram it can be derived that sampling in a grid should be avoided since it will cause lag distances at fixed intervals. The lag distances between these intervals will not be included in the experimental variogram and can have a negative influence on the accuracy of the method.

Taking the two main factors into account and avoiding sampling in a grid suggest that a spatially distributed random sampling scheme should be followed. This leaves the question of the number of samples needed to represent the environment with a high enough confidence level to be able to provide estimations with acceptable accuracy.

6.1.1 Iterative Sampling Plan

To address the above mentioned problem of how to determine the number of samples necessary we introduce an iterative sampling plan. The idea is to take a small number of spatially distributed samples through the building. The exact positions of the samples are not important as long as they are not locally grouped and they don't create a pattern that might have distances of fixed intervals.

In the process of evaluating the accuracy of a given number of samples, each sample will be used in a set of known values to estimate a sample that is considered to have an unknown value. This will allow for cross validation, but instead of measuring the whole building, as in the previous chapter, the samples that will be used for constructing the final coverage map are used as reference values.

The level of accuracy is estimated by using leave-one-out cross validation (LOOCV) [99]. This is a special case of leave-p-out cross validation (LpOCV) where $p = 1$.

LOOCV has the advantage over LpOCV by significantly lowering the computational cost needed for cross validation. Using LpOCV requires to learn and validate C_n^p times, but with $p = 1$ the LOOCV requires to learn and validate only $C_n^1 = n$ times.

Equation (6.1) shows how the error is calculated for each sample in a given set of n samples:

$$\epsilon_i = |m_i - p_i|, \tag{6.1}$$

where m_i is the measured value and p_i is the predicted value. This equation is repeated n times. The expected error for n samples is calculated by equation (6.2):

$$\epsilon = \frac{1}{n} \sum_i \epsilon_i \tag{6.2}$$

Equations (6.1) and (6.2) is then repeated for $n + 1$ samples and the expected error is compared to the expected error when using n samples. This process is repeated until an acceptable level of accuracy is achieved. At this point the error ϵ of the last iteration is considered to be the confidence level of the model.

6.1.2 Number of Iterations

When examining a graph similar to Figure 5.5 it is easy to see when enough samples were available after which the accuracy did not significantly increase. However, in order to know when to stop the iterative sampling process, while still taking samples, we need to define what an acceptable level of accuracy is.

At first one would like to state that sampling must continue until the error reaches below some level α. This will assume that all buildings have the same complexity and that kriging will perform the same for all environments. The problem with this assumption is that the error might not get below the level α. Also, if α is set too high, better estimations could have been made.

The second consideration is to state that the accuracy has reached an asymptotic level when the moving average of the slope of the error approaches zero. In practice the moving average will not reach exactly 0, which leaves us with the same problem as in the previous scenario. A certain value β close to 0, needs to be chosen to indicate when the slope has stabilised.

Choosing a value for β has the advantage over choosing a value for α in the sense that it indicates a stable condition for the results of the method for the given scenario, regardless of the level of accuracy. A suitable value for β still has to be selected and a method to determine this value can be derived by empirically investigating the behaviour of the accuracy as the number of samples increases.

A mathematical definition would state that stability is reached at the limit as the slope approaches zero. Since this method will be used in practical applications, a practical approach is needed to choose a value for β instead. By repeating the process of drawing graphs similar to Figure 5.5 using random samples, a few assumptions can be made to guide one in choosing a suitable value for β.

It was noticed that when the absolute value of the slope is larger than average, the asymptotic level is reached faster, with regards to the number of samples, than it would when the slope initially has an absolute value that is smaller than average. This suggests that β can be given a higher value for larger slopes since the error has a strong trend towards the minimum error where it will stabilise. A smaller initial slope needs a smaller value for β to make sure the variance in the error is small enough to be considered stable.

In cases where the initial slope had a lower absolute value, one could still see a clear spike in the values of the slope before it approaches zero. The phenomenon of the spike in the value of the slope for the first few samples has lead to setting a 'rule of thumb' for choosing β in practical applications.

In our investigations we found that β must be chosen so that it will define the boundaries between which the slope must eventually stabilise. The value of β is chosen to be one quarter of the absolute value of the largest spike. The slope must then stabilise to values between β and $-\beta$.

The last requirement is that at least the last quarter of the values, with a minimum of 10, of the calculated slope must be between these boundaries for the accuracy to be considered close enough to the asymptotic level. At this stage the samples represent the Wi-Fi signal strength in the building and can be used to construct a coverage map.

6.2 Experimental Setup

The final investigation was done as a blind test. No prior information was available about the building regarding the number of access points as well as their positions. The iterative sampling process as described above was followed to obtain the results as shown in the results section of this chapter.

The first set of measurements consisted of only three spatially distributed positions that were chosen randomly and RSSI were measured at those positions. Two of the samples were used to estimate the third sample for al three the combinations. The average error of the three estimations were used to represent the level of accuracy when using two samples. This process was repeated and the number of samples was increased with each evaluation.

While increasing the number of samples an evaluation of the stability of the results was done and a decision was made to either continue with the sampling process or to stop. At the point where the algorithm as explained above showed that enough samples were taken, the number of samples was marked as the favourable number of samples where one would stop in practice, but to illustrate the position of this point on the graph, sampling was continued.

6.3 Results

While performing the iterative sampling process, accuracy increased as the number of samples increased up to a point after which it flattened, resulting in a graph similar to Figure 5.5. Figure 6.1 shows the average error, standard deviation and range of the estimations as the number of samples increase.

Figure 6.2 shows the moving average of the slope obtained from the average error. The boundaries obtained by calculating β is shown by the green horizontal lines. The number of samples at which the algorithm signalled to stop the process is indicated by the vertical red line on both Figures 6.1 and 6.2.

Figure 6.3 and Figure 6.4 was drawn by selecting different samples from the same building to show that similar results can be obtained by selecting a different set of spatially distributed random samples. The number of samples needed to represent the building differed by only a few samples when using different sets. When visually examining the graphs it is also clear that the error stabilises between 20 and 30 samples as indicated in both the examples in this chapter.

Figure 6.5 shows a comparison of a coverage map constructed using a full set of measurements for this building and a coverage map constructed using 25 samples. The 25 samples were obtained from the iterative sampling process as explained above.

For the two examples presented here, the stop position was signalled at 23 samples with an average error of 3.96 dB and at 27 samples with an average error of 3.81 dB respectively. It is therefore appropriate to say that the coverage map in Figure 6.5 using 25 samples, can be constructed with a confidence level that will have an average error of about 4 dB throughout the building.

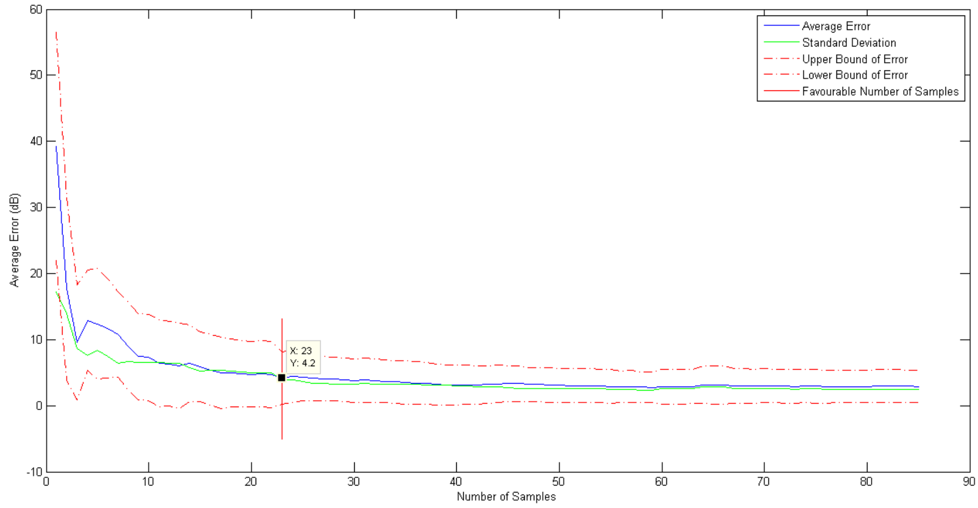

Figure 6.1: Average Error vs. Number of Samples - Cafeteria

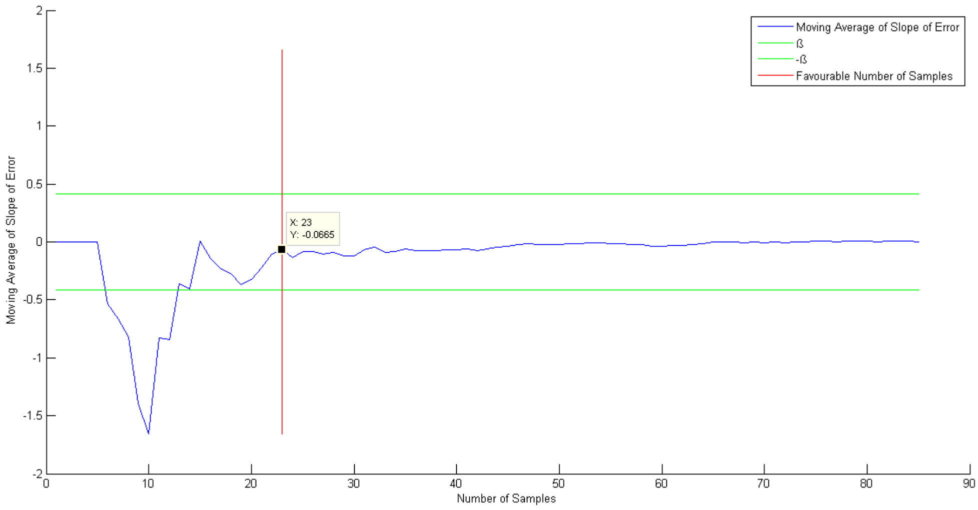

Figure 6.2: Moving average of the slope of the error - Cafeteria

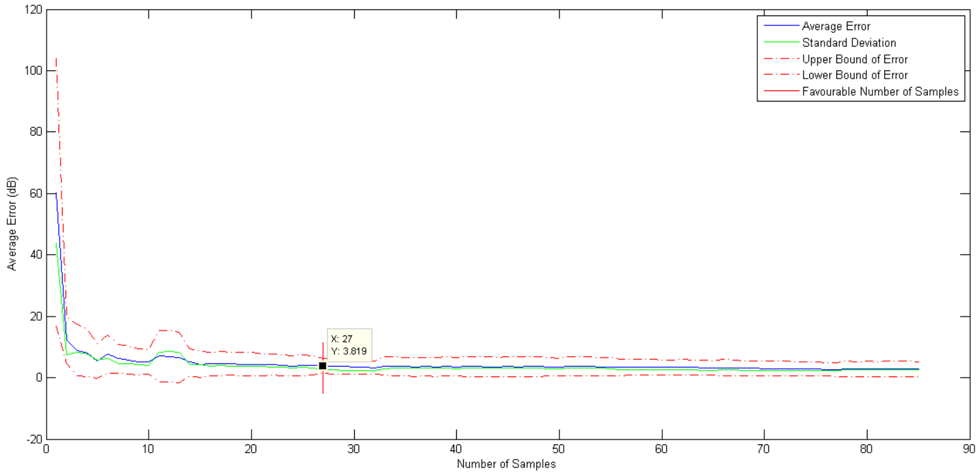

Figure 6.3: Average Error vs. Number of Samples - Cafeteria

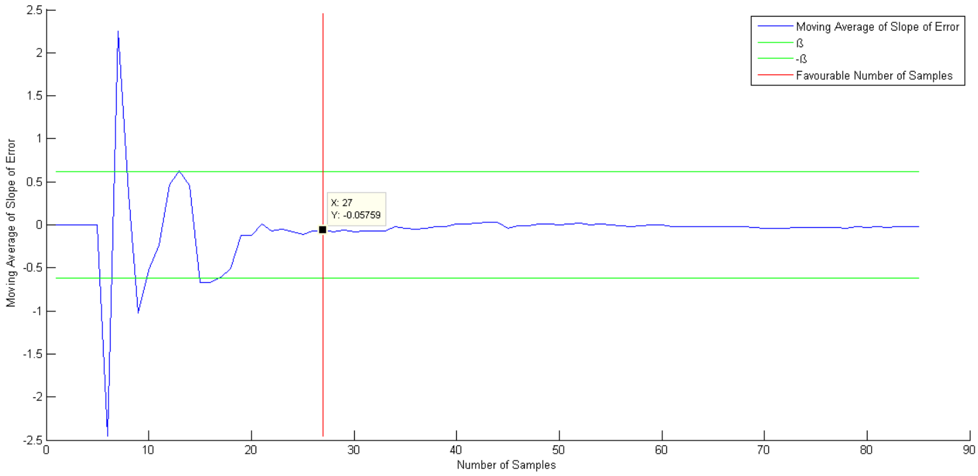

Figure 6.4: Moving average of the slope of the error - Cafeteria

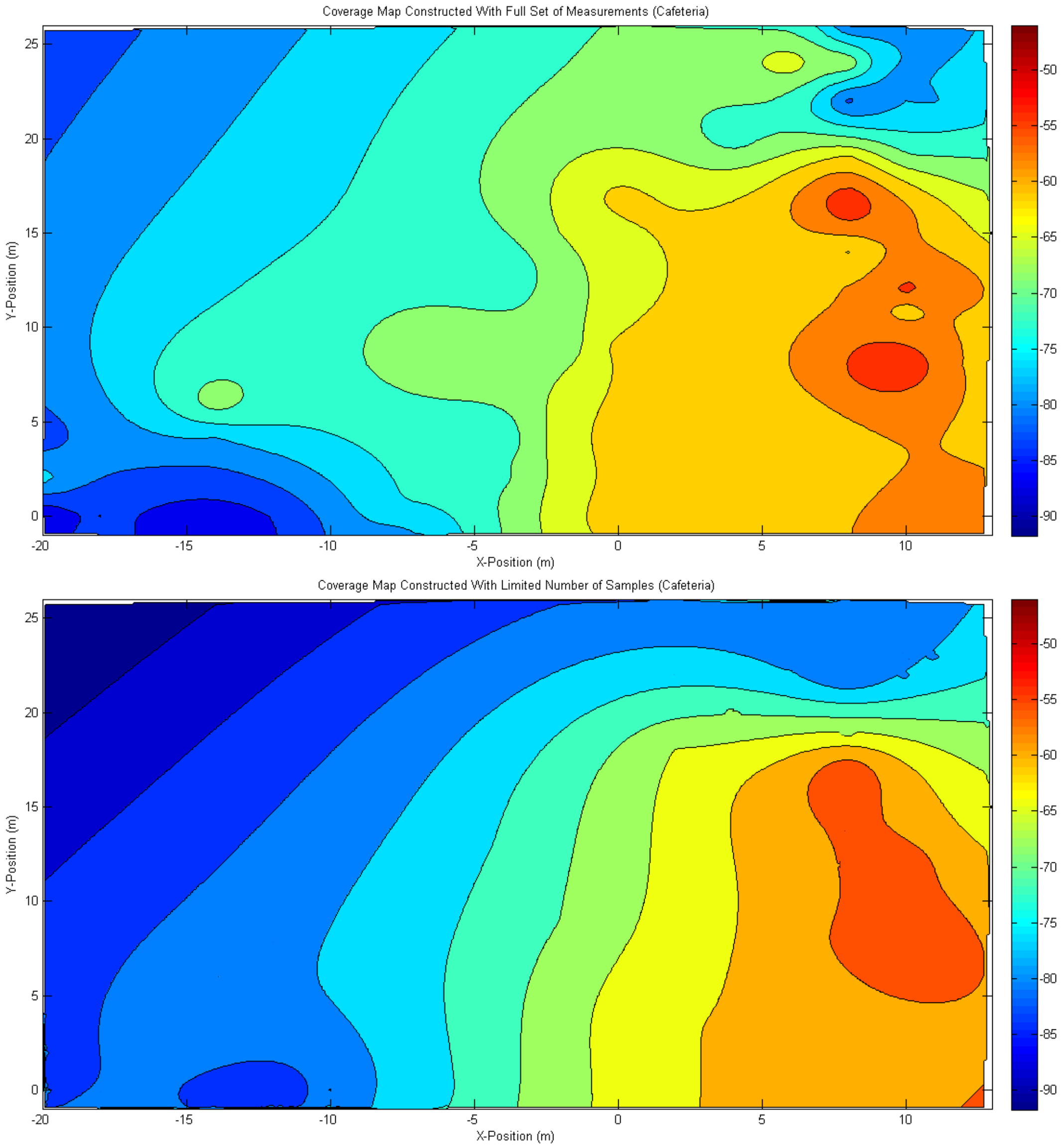

Figure 6.5: Coverage Map Comparison - Cafeteria

6.4 Chapter Conclusion

In order to thoroughly validate the results, Wi-Fi signal strength needs to be measured in an indefinite number of buildings where interpolation must be done and evaluated through cross validation. However, by staying within the scope and time constraints of this study the results of the third building is presented for validation. The strong similarities in both the behaviour as well as the level of accuracy of this method point to the relevance of kriging for Wi-Fi signal strength estimation for all the buildings that were surveyed.

From the results in this chapter it can be concluded that the performance of kriging is consistent for different buildings. It was shown that the samples used to construct a coverage map, can also be used for cross validation which provides a level of confidence with which the set of samples represent the building.

The blind test provided estimations that are on par with existing methods without increasing the effort it would take for a surveyor to take samples. In real life, this process can be optimised further by starting with more samples and increasing the number of samples with more than one sample with each iteration. This will decrease the total number of iterations necessary to find the favourable number of samples.

In this chapter a 'rule of thumb' was explained to determine the number of samples needed for constructing a reliable coverage map. At first glance this does not seem to be a very scientific approach to determine when to stop taking samples. No justification was given for choosing β as a quarter of the maximum slope in the error curve. However, choosing β is not nearly as critical as it would have been to choose a value for α which would define a minimum value for the error to reach before stopping the sampling process. Unlike α, the value of β has no correlation to the level of accuracy with which the method is estimating signal strength. It only indicates when the accuracy has stabilised for a number of iterations.

7 Conclusion

Kriging has been proposed for Wi-Fi signal strength estimation by [2] and [22]. In these two studies, tests were done by either using simulated data or by doing an experiment in a single hall that does not correspond to our definition of a complex indoor environment.

In [32] the use of geostatistical models for signal strength estimation is also encouraged, but it is stated that only preliminary work has been done in this field. In our study we did a more extensive investigation into using the geostatistical model kriging for Wi-Fi signal strength estimation in complex indoor environments.

Kriging for RSSI estimation was tested and evaluated based on prior knowledge obtained from literature and compared to existing propagation models. The use of kriging for Wi-Fi signal strength estimation can be described as convenient, accurate and simple to apply. This description is the result of multiple investigations.

7.1 The Investigations

This study was divided into three main investigations. Each investigation served as a foundation on which the next investigation was based.

All code used for the experiments in this study was written by the author. The kriging algorithm as explained in chapter 3 was also implemented in MATLAB without using a MATLAB toolbox. This was done to have maximum customizability in the process as a whole and to be able to make fine adjustments. Once the implemented algorithm provided satisfactory results, no further changes were made between the different investigations.

7.1.1 Investigation One

The first investigation involved getting familiar with kriging and to test the algorithm as implemented in MATLAB. RSSI was measured in a simple environment with only one floor and a single access point.

During this investigation, different unconventional variogram models were evaluated. For this specific case the parabola curve fitting, used as a model variogram, provided the best results. However, it was concluded that this is only valid for an environment with a single access point. This is due to the fact that the most distant positions were also strongly related. This is due to the weak signal strength measured far away from the access point. The decision was made to return to the conventional variogram models for the rest of the investigations.

7.1.2 Investigation Two

The next investigation was done in more complex environments to further our understanding of the behaviour of kriging for RSSI estimation. This investigation also provided a measure of the accuracy of kriging for this application to be compared to existing methods.

At this stage, it was already clear that the accuracy of kriging can compete with current methods for estimating signal strength. The small number of samples needed to obtain these accurate results suggested that kriging can be used confidently with little effort. The question that still needed to be answered was how to determine the number of samples that will provide satisfactory results, without measuring the whole building first.

The results obtained from this investigation assisted in preparing the sampling plan that followed. Most of the knowledge that was needed to establish the generic application protocol as described in the final investigation was obtained from these experiments.

7.1.3 Investigation Three

The final investigation consisted of finding a practical way to take samples and to provide some form of validation for the results from the previous investigations. The goal was to define a sampling plan that can be incorporated into a generic application protocol for technicians and engineers to use when planning a wireless network.

This investigation was done in a different building where no prior information was available about the number of access points or their positions. The level of accuracy was consistent when compared to the results of the previous investigations.

7.2 Kriging Compared to Existing Models

The drawbacks of current signal strength estimation models are that they are either very complex and require a lot of information about an environment, or that they have low accuracy. The geostatistical method kriging is suggested to address these problems for Wi-Fi signal strength estimation. The accuracy of kriging was found to be amongst the best of the existing estimation models. The effort of taking only a few measurements to represent an area is much less than the standard estimation models require.

To support this statement, the reader is referred back to Table 2.6 where the results of other comparative studies are presented. Kriging achieved an average error of about 4.5 dB with a standard deviation of 4 when using the recommended number of samples for the buildings that were surveyed.

By simply comparing the values of the results in Table 2.6 with the results obtained in this study, it is clear that kriging is amongst the top methods for this application. The complexity of the buildings with regards to size, number of walls, and number of Wi-Fi access points in their investigations was much simpler than the buildings surveyed in this study.

7.3 Sampling Plan

From our investigations we found that even though the accuracy of the estimated points increases as the number of samples increases, it reaches a point where the accuracy stabilises and no significant improvement can be seen by increasing the number of samples.

In our sampling plan the biggest challenge was knowing when enough samples have been taken to represent a building. This also raised the question of the level of confidence with which these samples can be used to construct a coverage map. To answer these questions an iterative sampling plan was suggested and implemented using leave-one-out cross validation (LOOCV).

During this study it became clear that instead of choosing a minimum level for the error to reach before sampling is stopped, the slope of the error should rather be examined. When the moving average of the slope reaches zero, the accuracy is stable and the sampling can stop. In practice this is not that simple since the moving average of the slope does not become exactly zero.

A formal mathematical definition would explain that stability is reached at the limit where the variation in accuracy approaches zero. In this case we needed a practical way to determine when the accuracy is stable enough. This was achieved by choosing boundaries within which the variation of the error must stabilise before sampling can be stopped. This involved drawing a graph of the error as the number of samples increases and taking the derivative of the graph for each new sample. As the number of samples increases, the moving average of the slope approaches zero. If the last quarter of the values of the slope, with a minimum of 10 values, lie between the chosen boundaries, the accuracy is considered to be stable.

This approach allowed the process of deciding on the number of samples needed to represent a building, to be performed independent from the actual level of accuracy that the model is providing. It shows that the accuracy is stable and the level of accuracy obtained using this number of samples is the level at which the model will settle. This is the confidence level of the model for the building that is surveyed.

7.4 Future Work

With estimation models there will always be optimisations. Accuracy can be improved and smarter sampling techniques can be defined.

Our sampling plan stated that spatially distributed random samples must be taken throughout a building. In future work, research might suggest more specific details to look for when choosing the exact positions for taking samples. This might complicate the method and implies that more information must be available about each building before it can be surveyed. This contradicts the genereric model that was the aim for this study, but the trade-off can be considered to decide whether this will be feasible.

Future work can also consider using kriging for channel state information (CSI) estimation. CSI is a more reliable and accurate metric for the quality of a connection available in 802.11n networks.

7.5 Closure

From the results of the surveys it is concluded that kriging can be confidently used as an estimation technique for RSSI in complex indoor environments. Kriging is recommended for practical applications, especially where insufficient information is available about a building or where time consuming site surveys are not feasible.

Bibliography

[1] P. R. Minnitt and D. W. Assibey-Bonsu. (2010) Professor D.G. Krige - his contributions to research and engineering. [Online]. Available: http://www.gasa.org.za/wp-content/uploads/Tribute-to-Prof-Danie-Krige-2010-Edited.pdf

[2] A. Konak, "Estimating path loss in wireless local area networks using ordinary kriging," in *Simulation Conference (WSC), Proceedings of the 2010 Winter*, Dec 2010, pp. 2888–2896.

[3] S. Biaz, J. Yiming, B. Qi, and S. Wu, "Dynamic signal strength estimates for indoor wireless communications," in *Wireless Communications, Networking and Mobile Computing, 2005. Proceedings. 2005 International Conference on*, vol. 1, Sept 2005, pp. 602–605.

[4] C. Andrade and R. Hoefel, "Ieee 802.11 wlans: A comparison on indoor coverage models," in *Electrical and Computer Engineering (CCECE), 2010 23rd Canadian Conference on*, May 2010, pp. 1–6.

[5] P. Bahl and V. Padmanabhan, "Radar: an in-building rf-based user location and tracking system," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, 2000, pp. 775–784 vol.2.

[6] M. Hassan-Ali and K. Pahlavan, "A new statistical model for site-specific indoor radio propagation prediction based on geometric optics and geometric probability," *Wireless Communications, IEEE Transactions on*, vol. 1, no. 1, pp. 112–124, Jan 2002.

[7] M. Lott and I. Forkel, "A multi-wall-and-floor model for indoor radio propagation," in *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd*,

vol. 1, 2001, pp. 464–468 vol.1.

[8] H. Kim and H. Ling, "Electromagnetic scattering from an inhomogeneous object by ray tracing," *Antennas and Propagation, IEEE Transactions on*, vol. 40, no. 5, pp. 517–525, May 1992.

[9] V. Gandhi. Semivariogram modeling. [Online]. Available: http://www-users.cs.umn.edu/~gandhi/courses/CS8701/g4_semivariogram_final_draft.pdf

[10] W. Kerwin and J. Prince, "The kriging update model and recursive space-time function estimation," *Signal Processing, IEEE Transactions on*, vol. 47, no. 11, pp. 2942–2952, Nov 1999.

[11] S. Koziel, S. Ogurtsov, I. Couckuyt, and T. Dhaene, "Variable-fidelity electromagnetic simulations and co-kriging for accurate modeling of antennas," *Antennas and Propagation, IEEE Transactions on*, vol. 61, no. 3, pp. 1301–1308, March 2013.

[12] S Koziel, "Cost-efficient electromagnetic-simulation-driven antenna design using co-kriging," *Microwaves, Antennas Propagation, IET*, vol. 6, no. 14, pp. 1521–1528, November 2012.

[13] S Ogurtsov, "Efficient simulation-driven design optimization of antennas using co-kriging," in *Antennas and Propagation Society International Symposium (APSURSI), 2012 IEEE*, July 2012, pp. 1–2.

[14] G. Boccolini, G. Hernandez-Penaloza, and B. Beferull-Lozano, "Wireless sensor network for spectrum cartography based on kriging interpolation," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, Sept 2012, pp. 1565–1570.

[15] C. Guvendik, A. Genc, O. Tamer, and M. Nil, "Improving the performance of kriging based interpolation application with parallel processors," in *Signal Processing and Communications Applications Conference (SIU), 2012 20th*, April 2012, pp. 1–4.

[16] J. Defretin, J. Marzat, and H. Piet-Lahanier, "Learning viewpoint planning in active recognition on a small sampling budget: A kriging approach," in *Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on*, Dec 2010, pp.

169–174.

[17] S. Koziel, "Surrogate-based optimization of microwave structures using space mapping and kriging," in *Microwave Conference, 2009. EuMC 2009. European*, Sept 2009, pp. 1062–1065.

[18] E. C.-C. Lo, "An investigation of the impact of signal strength on wi-fi link throughput through propagation measurement," Ph.D. dissertation, Auckland University of Technology, 2007.

[19] G. Bohling, "Introduction to geostatistics and variogram analysis," *Kansas geological survey, 20p*, 2005.

[20] I. Sayin, F. Arikan, and O. Arikan, "Synthetic tec mapping with kriging and random field priors," in *Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th*, June 2007, pp. 1–4.

[21] H. J. Miller, "Toblers first law and spatial analysis," *Annals of the Association of American Geographers*, vol. 94, no. number, pp. 284–289, June 2004.

[22] Y. Mezali and P. Jacquet, "On indoor wifi signal statistical properties," in *Wireless and Mobile Networking Conference (WMNC), 2011 4th Joint IFIP*, Oct 2011, pp. 1–8.

[23] K. Pahlavan, *Principles of wireless networks: A unified approach.* John Wiley & Sons, Inc., 2011.

[24] R. Leutert, "Inside 802.11n - technical details about the new wlan standard." [Online]. Available: http://www.wireshark.ch/download/Cisco_PSE_Day_2009.pdf

[25] Broadcom, "802.11n: Next-generation wireless lan technology." [Online]. Available: http://www.broadcom.com/collateral/wp/802_11n-WP100-R.pdf

[26] A. Ramachandran and S. Jagannathan, "Spatial diversity in signal strength based wlan location determination systems," in *Local Computer Networks, 2007. LCN 2007. 32nd IEEE Conference on*, Oct 2007, pp. 10–17.

[27] J. Networks, "Coverage or capacity - making the best use of 802.11n."

[Online]. Available: http://www.adtechglobal.com/Data/Sites/1/marketing/juniperwhitepaperwlancoverageorcapacity.pdf

[28] J. Bardwell, “You believe you understand what you think i said,” *The truth about*, vol. 802, 2004.

[29] R.-H. Wu, Y.-H. Lee, H.-W. Tseng, Y.-G. Jan, and M.-H. Chuang, “Study of characteristics of rssi signal,” in *Industrial Technology, 2008. ICIT 2008. IEEE International Conference on*, April 2008, pp. 1–3.

[30] G. Lui, T. Gallagher, B. Li, A. Dempster, and C. Rizos, “Differences in rssi readings made by different wi-fi chipsets: A limitation of wlan localization,” in *Localization and GNSS (ICL-GNSS), 2011 International Conference on*, June 2011, pp. 53–57.

[31] J. Arpee, S. Gutowski, and M. Touati, “Apparatus and method for geostatistical analysis of wireless signal propagation,” Mar. 23 2004, uS Patent 6,711,404. [Online]. Available: http://www.google.com/patents/US6711404

[32] C. Phillips, D. Sicker, and D. Grunwald, “A survey of wireless path loss prediction and coverage mapping methods,” *Communications Surveys Tutorials, IEEE*, vol. 15, no. 1, pp. 255–270, First 2013.

[33] M. Luo, “Indoor radio propagation modeling for system performance prediction,” Ph.D. dissertation, INSA de Lyon, 2013.

[34] X. Song, F. Yang, L. Ding, and L. Qian, “Weight adjust algorithm in indoor fingerprint localization,” in *Signal Processing and Communication Systems (ICSPCS), 2012 6th International Conference on*, Dec 2012, pp. 1–5.

[35] Sputnik, “Rf propagation basics.” [Online]. Available: https://www.sputnik.com/resources/support/deployment/rf_propagation_basics.pdf

[36] P. Catherwood, T. Zech, and J. McLaughlin, “Cost-effective rssi wi-fi positioning solution for ambulatory patient monitoring devices,” in *Antennas and Propagation Conference (LAPC), 2010 Loughborough*, Nov 2010, pp. 557–560.

[37] I. . W. Group *et al.*, “Ieee standard for information technology–telecommunications and information exchange between systems–local and metropolitan area networks–

specific requirements–part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 6: Wireless access in vehicular environments," *IEEE Std*, vol. 802, p. 11p, 2010.

[38] K. Ogunjemilua, J. N. Davies, V. Grout, and R. Picking, "An investigation into signal strength of 802.11 n wlan," in *Proc. Fifth Collaborative Research Symposium on Security, E-Learning, Internet and Networking (SEIN 2009)*, 2009.

[39] J.-A. Park, S.-K. Park, P.-D. Cho, and K.-R. Cho, "Analysis of spectrum channel assignment for ieee 802.11 b wireless lan," in *Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on*, vol. 3. IEEE, 2002, pp. 1073–1077.

[40] N. Prasad and A. Prasad, *802.11 WLANs and IP Networking.* Artech House, 2005.

[41] N. P. Reid and R. Seide, *802.11 (Wi-Fi): networking handbook.* McGraw-Hill Osborne Media, 2003.

[42] Z. Z. Z Yang and Y. Liu, "From rssi to csi: Indoor localization via channel response." [Online]. Available: http://tns.thss.tsinghua.edu.cn/iotbook/resources/RSSI2CSI.pdf

[43] V. Seshadri, G. Zaruba, and M. Huber, "A bayesian sampling approach to indoor localization of wireless devices using received signal strength indication," in *Pervasive Computing and Communications, 2005. PerCom 2005. Third IEEE International Conference on*, March 2005, pp. 75–84.

[44] A. T. Parameswaran, M. I. Husain, S. Upadhyaya *et al.*, "Is rssi a reliable parameter in sensor localization algorithms: An experimental study," in *Field Failure Data Analysis Workshop (F2DA09)*, 2009.

[45] S. Mazuelas, A. Bahillo, R. Lorenzo, P. Fernandez, F. Lago, E. Garcia, J. Blas, and E. Abril, "Robust indoor positioning provided by real-time rssi values in unmodified wlan networks," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 3, no. 5, pp. 821–831, Oct 2009.

[46] D. Vilaseca and J. Giribet, "Indoor navigation using wifi signals," in *Embedded Systems (SASE/CASE), 2013 Fourth Argentine Symposium and Conference on*,

Aug 2013, pp. 1–6.

[47] H. T. Friis, “A note on a simple transmission formula,” *proc. IRE*, vol. 34, no. 5, pp. 254–256, 1946.

[48] T. S. Rappaort, “Wireless communications: principles and practice,” 2002.

[49] H. P. Pfeifer, “On the validation of radio propagation models,” 2010.

[50] D. Sharma, P. K. Sharma, V. Gupta, and R. Singh, “A survey on path loss models used in wireless communication system design,” *International J. of Recent Trends in Engineering and Technology*, vol. 3, no. 2, 2010.

[51] D. B. Green and A. Obaidat, “An accurate line of sight propagation performance model for ad-hoc 802.11 wireless lan (wlan) devices,” in *Communications, 2002. ICC 2002. IEEE International Conference on*, vol. 5. IEEE, 2002, pp. 3424–3428.

[52] C. E. Dadson, J. Durkin, and R. Martin, “Computer prediction of field strength in the planning of radio systems,” *Vehicular Technology, IEEE Transactions on*, vol. 24, no. 1, pp. 1–8, 1975.

[53] A. Blomquist and L. Ladell, “Prediction and calculation of transmission loss in different types of terrain,” in *In AGARD Electromagnetic Wave Propagation Involving Irregular Surfaces and Inhomogeneous Media 17 p (SEE N75-22045 13-70)*, vol. 1, 1975.

[54] K. Allsebrook and J. D. Parsons, “Mobile radio propagation in british cities at frequencies in the vhf and uhf bands,” *Vehicular Technology, IEEE Transactions on*, vol. 26, no. 4, pp. 313–323, 1977.

[55] R. S. de Souza, “A new propagation model for 2.4 ghz wireless lan,” *Communications*, 2008.

[56] T. S. Rappaport *et al.*, *Wireless communications: principles and practice.* Prentice Hall PTR New Jersey, 1996, vol. 2.

[57] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, “Field strength and its variability in vhf and uhf land-mobile radio service,” *Rev. Elec. Commun. Lab*, vol. 16, no. 9, pp. 825–73, 1968.

[58] S. Saunders and F. Bonar, "Explicit multiple building diffraction attenuation function for mobile radio wave propagation," *Electronics Letters*, vol. 27, no. 14, pp. 1276–1277, 1991.

[59] J. Walfisch and H. L. Bertoni, "A theoretical model of uhf propagation in urban environments," *Antennas and Propagation, IEEE Transactions on*, vol. 36, no. 12, pp. 1788–1796, 1988.

[60] D. Har, A. M. Watson, and A. G. Chadney, "Comment on diffraction loss of rooftop-to-street in cost 231-walfisch-ikegami model," *Vehicular Technology, IEEE Transactions on*, vol. 48, no. 5, pp. 1451–1452, 1999.

[61] K. T. Herring, J. W. Holloway, D. H. Staelin, and D. W. Bliss, "Path-loss characteristics of urban wireless channels," *Antennas and Propagation, IEEE Transactions on*, vol. 58, no. 1, pp. 171–177, 2010.

[62] V. Erceg, L. J. Greenstein, S. Y. Tjandra, S. R. Parkoff, A. Gupta, B. Kulic, A. A. Julius, and R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *Selected Areas in Communications, IEEE Journal on*, vol. 17, no. 7, pp. 1205–1211, 1999.

[63] V. Garg, *Wireless Communications & Networking.* Morgan Kaufmann, 2010.

[64] R. ITU-R, "P. 452-10: Prediction procedure for the evaluation of microwave interference between stations on the surface of the earth at frequencies above about 0.7 ghz," 2001.

[65] S. Seybold John, "Introduction to rf propagation," 2005.

[66] G. A. Hufford, A. G. Longley, W. A. Kissick *et al.*, *A guide to the use of the ITS irregular terrain model in the area prediction mode.* US Department of Commerce, National Telecommunications and Information Administration, 1982.

[67] M. Riback, J. Medbo, J.-E. Berg, F. Harrysson, and H. Asplund, "Carrier frequency effects on path loss," in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, vol. 6. IEEE, 2006, pp. 2717–2721.

[68] A. Dissanayake, J. Allnutt, and F. Haidara, "A prediction model that combines

rain attenuation and other propagation impairments along earth-satellite paths," *Antennas and Propagation, IEEE Transactions on*, vol. 45, no. 10, pp. 1546–1558, 1997.

[69] Y. L. De Jong, M. H. Koelen, and M. H. Herben, "A building-transmission model for improved propagation prediction in urban microcells," *Vehicular Technology, IEEE Transactions on*, vol. 53, no. 2, pp. 490–502, 2004.

[70] G. Durgin, T. S. Rappaport, and H. Xu, "Measurements and models for radio path loss and penetration loss in and around homes and trees at 5.85 ghz," *Communications, IEEE Transactions on*, vol. 46, no. 11, pp. 1484–1496, 1998.

[71] L. J. Greenstein and V. Erceg, "Gain reductions due to scatter on wireless paths with directional antennas," *Communications Letters, IEEE*, vol. 3, no. 6, pp. 169–171, 1999.

[72] E. Anderson, C. Phillips, D. Sicker, and D. Grunwald, "Modeling environmental effects on directionality in wireless networks," *Mathematical and Computer Modelling*, vol. 53, no. 11, pp. 2078–2092, 2011.

[73] B. Sklar, "Rayleigh fading channels in mobile digital communication systems. i. characterization," *Communications Magazine, IEEE*, vol. 35, no. 7, pp. 90–100, 1997.

[74] N. Youssef, T. Munakata, and M. Takeda, "Fade statistics in nakagami fading environments," in *Spread Spectrum Techniques and Applications Proceedings, 1996., IEEE 4th International Symposium on*, vol. 3. IEEE, 1996, pp. 1244–1247.

[75] M. Nakagami, "The m-distribution-a general formula of intensity distribution of rapid fading," *Statistical Method of Radio Propagation*, 1960.

[76] L. W. Barclay, *Propagation of radiowaves.* Iet, 2003, vol. 502.

[77] S. Zvanovec, M. Valek, and P. Pechac, "Results of indoor propagation measurement campaign for wlan systems operating in 2.4 ghz ism band," in *Antennas and Propagation, 2003.(ICAP 2003). Twelfth International Conference on (Conf. Publ. No. 491)*, vol. 1. IET, 2003, pp. 63–66.

[78] R. Mardeni and Y. Solahuddin, “Path loss model development for indoor signal loss prediction at 2.4 ghz 802.11n network,” in *Microwave and Millimeter Wave Technology (ICMMT), 2012 International Conference on*, vol. 2, May 2012, pp. 1–4.

[79] R. Hoefel, “Ieee wlans: 802.11, 802.11e mac and 802.11a, 802.11b, 802.11g phy cross layer link budget model for cell coverage estimation,” in *Electrical and Computer Engineering, 2008. CCECE 2008. Canadian Conference on*, May 2008, pp. 001 877–001 882.

[80] J. Robinson, R. Swaminathan, and E. W. Knightly, “Assessment of urban-scale wireless networks with a small number of measurements,” in *Proceedings of the 14th ACM international conference on Mobile computing and networking.* ACM, 2008, pp. 187–198.

[81] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, “Active learning with statistical models,” *arXiv preprint cs/9603104*, 1996.

[82] C. Phillips, M. Ton, D. Sicker, and D. Grunwald, “Practical radio environment mapping with geostatistics,” in *Dynamic Spectrum Access Networks (DYSPAN), 2012 IEEE International Symposium on*, Oct 2012, pp. 422–433.

[83] S. Latif, M. Ghazanfar, A. Memon, B. Chowdhry, and J. Ahmed, “Comparison of d-model and wall-attenuation model for signal strength estimations in indoor environment,” in *Computational Intelligence, Modelling and Simulation (CIMSiM), 2012 Fourth International Conference on*, 2012, pp. 336–340.

[84] L. Jaisingh, “Statistics for the utterly confused,” 2000.

[85] “Explorable,” 2014. [Online]. Available: https://explorable.com/statistical-sampling-techniques

[86] S. DiCalogero, “Statistical sampling,” 2009. [Online]. Available: http://www.tcc.edu/vml/documents/StatisticalSampling.pptx

[87] A. B.M. Whelan and B.Minasny, “Spatial prediction software for precision agriculture.” [Online]. Available: http://sydney.edu.au/agriculture/pal/documents/vesperpaper.pdf

[88] B. Marchant and R. Lark, "Adaptive sampling and reconnaissance surveys for geostatistical mapping of the soil," *European journal of soil science*, vol. 57, no. 6, pp. 831–845, 2006.

[89] T. Hengl, "A practical guide to geostatistical mapping of environmental variables," *JRC Scientific and Technichal Reports. Office for Official Publication of the European Communities, Luxembourg*, 2007.

[90] ArcGIS, "What are the different kriging models?" [Online]. Available: http://resources.arcgis.com/fr/help/main/10.2/index.html#//00310000003q000000

[91] ArcGIS, "Understanding simple kriging." [Online]. Available: http://resources.arcgis.com/es/help/main/10.1/index.html#//003100000040000000

[92] ArcGIS, "Understanding ordinary kriging." [Online]. Available: http://help.arcgis.com/de/arcgisdesktop/10.0/help/index.html#//00310000003s000000

[93] ArcGIS, "Understanding universal kriging." [Online]. Available: http://resources.arcgis.com/es/help/main/10.1/index.html#//003100000048000000

[94] Ordinary kriging. [Online]. Available: http://www.ems-i.com/gmshelp/Interpolation/Interpolation_Schemes/Kriging/Ordinary_Kriging.htm

[95] F. P. Agterberg. Georges atheron - founder of spatial statistics. [Online]. Available: http://www.geostatscam.com/Adobe/G_Matheron.pdf

[96] A. Cantoni, "Optimal curve fitting with piecewise linear functions," *Computers, IEEE Transactions on*, vol. C-20, no. 1, pp. 59–67, 1971.

[97] D. Tang, G. Zhang, and J. Qin, "On the combination of spatial diversity and multiuser diversity under spatial correlation," in *Mobile Technology, Applications and Systems, 2005 2nd International Conference on*, 2005, pp. 4 pp.–4.

[98] K. Yang, H. Wang, G. Dai, S. Hu, Y. Zhang, and J. Xu, "Determining the repeat number of cross-validation," in *Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on*, vol. 3, 2011, pp. 1706–1710.

[99] A. W. Moore and M. S. Lee, "Efficient algorithms for minimizing cross validation error." in *ICML*, 1994, pp. 190–198.