

Aalto University  
School of Science  
Degree Programme in Computer Science and Engineering

Srikanth Gadicherla

# Data Analysis and Memory Based Methods for RSS Bluetooth Low Energy Indoor Positioning

Master's Thesis  
Espoo, !FIXME **July 26, 2017** FIXME!

**DRAFT! — September 26, 2018 — DRAFT!**

Supervisor: Professor Aki Vehtari  
Advisors: D.Sc. (Tech.) Roland Hostettler, Aalto University  
M.Sc. (Tech.) Heikki Pulkkinen, Helvar Oy Ab

<b>Author:</b>	Srikanth Gadicherla		
<b>Title:</b>	Data Analysis and Memory Based Methods for RSS Bluetooth Low Energy Indoor Positioning		
<b>Date:</b>	<b>!FIXME July 26, 2017 FIXME!</b>	<b>Pages:</b>	xi + 91
<b>Major:</b>	Machine Learning and Data Mining	<b>Code:</b>	SCI3044
<b>Supervisor:</b>	Professor Aki Vehtari		
<b>Advisors:</b>	D.Sc. (Tech.) Roland Hostettler, Aalto University M.Sc. (Tech.) Heikki Pulkkinen, Helvar Oy Ab		
<p>The thesis is a proof of concept solutions to Bluetooth low energy indoor positioning (BLE-IP) including comprehensive data analysis of the signal strength values. The data analysis of received signal strength indication (RSSI) values was done to understand different factors influencing the RSSI values so as to improve the data model. The positioning task is accomplished using a methodology called <i>fingerprinting</i>. The fingerprinting based positioning involves two phases namely <i>calibration phase</i> and <i>localization phase</i>. The calibration phase deals generation of radio-maps with the recorded received signal strength indication fingerprints at pre-destined location in the state space. The localization phase involves memory or non memory based methods for positioning. In this thesis we focus on memory based methods: <i>particle filters</i> and <i>unscented Kalman filter</i>. The radio-map generated in the calibration phase is used as the measurement model in the filtering context. The optimal calibration phase parameters were determined and the filtering methods were evaluated in terms root mean square error.</p>			
<b>Keywords:</b>	Indoor Positioning, Bayesian Filtering, Gaussian Processes, Fingerprinting		
<b>Language:</b>	English		

## Acknowledgements

मातृ देवो भव।	Honour thy Mother as God.
पितृ देवो भव।	Honour thy Father as God.
आचार्य देवो भव।	Honour thy Teacher as God.
अतिथि देवो भव॥	Honour thy Guest as God.

My master's journey has been no less dramatic than a roller coaster ride. There are many people who always stood by my side but most foremost and above all I would like to acknowledge the endless support of my family, my mother, father, brother and sister-in-law, who've always stood like a wall. I'm in debt to my brother and friend *Badri Kakulavarapu* for his constant words of motivation throughout.

I express my deepest gratitude to my supervisors, Professor *Aki Vehtari* and Professor *Simo Särkka* for being kind and having faith in me in various project I've worked with them. I can't thank much my advisor and guide *Roland Hostettler* for many wonderful discussions both offline and online.

I thank Helvar for giving me this opportunity to work on my thesis, and especially my bosses *Henri Júslen*, *Max Björkgren* and my daily advisor *Heikki Pulkkinen* for giving me enough freedom in my research and explore different avenues in my thesis. I would also like to thank my colleagues *Toni Ajo* and *Guo Haipeng* for helping regarding floorplans and measurements application.

I can't thank more my lab-mates *Juho Piironen*, *Eero Siivola*, *Marko Järvenpää*, *Markus Paasiniemi*, *Olli-Pekka Koistinen*, *Aakash Kumar Dhaka* and *Michael Riis Andersen* in PML group for being accommodating and being readily available for brainstorming and discussion in times when I was struck and out of ideas.

Last but not the least, I thank my friends and life gurus *Nonappa Billinele*, *Jussi Gillberg*, *Kalyan G.V.S.*, *Bhanu Prakash Reddy*, *Karthik Upadhyay*, *Purendar Venkateshan*, *Srikanth Balasubramanian*, *Afaque Hussain*, *Eliza Barkane*, *Pradeep Kumar Eranti*, *Sriharsha Kuchimanchi*, *Manoj Kumar Tyagi*, *Mayank #nevermind Khandelwal*, *Sunny Vijay*, *Kunal Ghosh*, *Shishir Bhattacharai*, *Maria Valdimirovna Pochtar*, *Abheeshta Putta*, *Ashish Sultania* for everything.

Otaniemi, !FIXME **July 26, 2017** FIXME!

Srikanth Gadicherla



# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>Preface</b>	<b>ix</b>
<b>Symbols and Abbreviations</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	2
1.2 Indoor Positioning . . . . .	2
1.3 Bluetooth Low Energy . . . . .	3
1.4 Challenges Plaguing the RSS based methods . . . . .	4
1.5 Goal & Contribution of the thesis . . . . .	5
1.6 Outline of the thesis . . . . .	5
<b>2 Location Fingerprinting</b>	<b>7</b>
2.1 Location Fingerprinting . . . . .	7
2.1.1 Radiomap . . . . .	8
2.1.1.1 Reference Table . . . . .	9
2.1.2 Raw data: RSSI . . . . .	10
2.2 RSS Localization algorithms . . . . .	11
2.2.1 Non-memory based methods . . . . .	11
2.2.1.1 k-Nearest Neighbour . . . . .	12
2.2.2 Memory based methods . . . . .	13
2.3 Related Work . . . . .	13
2.4 Performance metrics . . . . .	14
<b>3 Gaussian Process Model for Bayesian Filtering</b>	<b>15</b>
3.1 Gaussian Processes . . . . .	15
3.1.1 Application in Indoor Postioning . . . . .	16
3.2 Motivation for using the GPs . . . . .	16
3.3 Advantages of GPs . . . . .	17

3.4	GP modeling for indoor positioning . . . . .	17
<b>4</b>	<b>Memory Based Filtering Methods</b>	<b>21</b>
4.1	Bayesian Filtering . . . . .	21
4.1.1	State Space Modeling . . . . .	22
4.1.2	Advantages of Bayesian Filtering . . . . .	24
4.2	Filtering . . . . .	24
4.2.1	Particle Filters . . . . .	24
4.2.1.1	Importance Sampling . . . . .	24
4.2.2	Algorithm . . . . .	27
4.2.3	Unscented Kalman Filters . . . . .	30
4.2.3.1	Unscented Transform . . . . .	30
4.2.3.2	Algorithm . . . . .	31
<b>5</b>	<b>Measurement Setup &amp; Data Collection</b>	<b>35</b>
5.1	Experimental Testbed . . . . .	35
5.2	Measurement Configuration . . . . .	35
5.3	Bluetooth Low Energy . . . . .	36
5.3.1	Physical Layer . . . . .	36
5.3.2	Advertisement channels . . . . .	37
5.3.3	Data channels . . . . .	38
5.3.4	Different BLE protocols . . . . .	38
5.4	Assumptions on the RSSI values . . . . .	39
5.5	Radio Analyzer . . . . .	39
5.6	Measurement Application . . . . .	39
5.6.1	Smartphones . . . . .	40
5.7	Data Collection . . . . .	41
5.7.1	For Data Analysis . . . . .	41
5.7.1.1	Different User Orientations . . . . .	41
5.7.1.2	Different Phone Orientations . . . . .	43
5.7.1.3	Outdoor Measurements . . . . .	43
5.7.2	For Positional Algorithms . . . . .	43
5.7.2.1	Setup for Calibration Phase and Test Phase .	43
5.7.2.2	Calibration Data . . . . .	44
5.7.2.3	Test data . . . . .	44
<b>6</b>	<b>Data Analysis of RSSI</b>	<b>47</b>
6.1	RSSI as a measure of distance . . . . .	49
6.2	Statistical Hypothesis Testing . . . . .	50
6.2.1	Two Sample Z-test . . . . .	50
6.2.2	Two Sample Kolmogorov-Smirnov test . . . . .	51

6.2.3	Levene's test . . . . .	52
6.2.4	Wilcoxon rank-sum test . . . . .	53
6.2.5	Experiments . . . . .	54
6.2.5.1	Bias due to user's presence . . . . .	54
6.2.5.2	Comparisons of Smartphone . . . . .	56
6.2.5.3	Orientation of Smart phone . . . . .	58
6.2.5.4	Material of the Luminaires . . . . .	59
6.2.5.5	Comparison with Radio Analyzer . . . . .	61
6.3	Received Signal Strength Indication: Revisited . . . . .	61
6.3.1	Measurement Outdoors with Radio analyzer . . . . .	63
<b>7</b>	<b>Experiments and Results</b>	<b>65</b>
7.1	Effect of parameters in calibration phase on positioning . . . . .	65
7.1.1	Separate parameter evaluation . . . . .	65
7.1.1.1	Calibration points . . . . .	65
7.1.1.2	Calibration time . . . . .	66
7.1.2	Combined parameter evaluation . . . . .	67
7.2	Filtering location estimation algorithms . . . . .	67
7.2.1	Summary of the location estimation algorithms . . . . .	70
<b>8</b>	<b>Discussion &amp; Future Work</b>	<b>73</b>
8.1	Discussion . . . . .	73
8.1.1	Data Analysis . . . . .	73
8.1.2	Radio map . . . . .	74
8.1.3	Indoor Positioning . . . . .	74
8.2	Future Work . . . . .	74
<b>References</b>		<b>77</b>
<b>A</b>	<b>Appendix</b>	<b>85</b>
A.1	Tables for Data Analysis . . . . .	85



# Preface

(*Bayesian*) Probability is a way-point between ignorance and knowledge.  
-Pierre-Simon Laplace

Uncertainty is endemic to the problems related to decision analysis. It can range from making a personal choice to making a policy decision for multi-billion dollar company. In that sense, the concept of uncertainty is extremely innate to the field of Machine Learning which involves making predictions.

Formalizing the concept of uncertainty and to generate valid measures for making decision in terms of probability distributions, we use a methodology called *Bayesian Inference*, in which our beliefs are updated based on model evidence.

One example of how powerful this methodology is locating the wreckage of *Air France Flight AF447* (Stone et al., 2014) in 2009. The aforementioned methodology took just a week for the search operation. The analysis not only included the flight data but also the failures points of the previous methods to correctly estimate the location of the wreckage. Subsequently, it was also used in the search operations for the *Malaysian Airliner MH370* (Davey et al., 2016) in 2014.

From above examples it is evident that *Bayesian Inference* is a powerful tool to combine knowledge from multiple sources and gives opportunity to the modeller to criticize and update ones own model. The above analysis particularly involved an iterative algorithm called *Particle Filters*.



# Symbols and Abbreviations

## Symbols

<b>B</b>	magnetic flux density
$c$	speed of light in vacuum $\approx 3 \times 10^8$ [m/s]
$\omega_D$	Debye frequency
$\omega_{\text{latt}}$	average phonon frequency of lattice
$\uparrow$	electron spin direction up
$\downarrow$	electron spin direction down

## Operators

$\nabla \times \mathbf{A}$	curl of vector in $\mathbf{A}$
$\frac{d}{dt}$	derivative with respect to variable $t$
$\frac{\partial}{\partial t}$	partial derivative with respect to variable $t$
$\sum_i$	sum over index $i$
$\mathbf{A} \cdot \mathbf{B}$	dot product of vectors $\mathbf{A}$ and $\mathbf{B}$

## Abbreviations

GP	Gaussian Process
APLAC	an object-oriented analog circuit simulator and design tool (originally Analysis Program for Linear Active Circuits)
BCS	Bardeen-Cooper-Schrieffer
DC	direct current
TEM	transverse electromagnetic



# Chapter 1

## Introduction

*Self-Awareness is the one of the significant parts of human evolution and location awareness is one of its component. As the field of artificial intelligence is moving towards autonomy, the location awareness of intelligent machines would become increasingly more vital.*

The quest to accurately navigate through unknown terrains has plagued the mankind since forever. The sixteenth century approach to solve this problem moved from being deterministic to one being deduced. The deterministic methods involved rigorous application of astronomy and mathematics with a spate of tools like celestial globe, astrolabe, quadrant, cross-staff etc. The method by deduction, popularly known today as *dead reckoning*, used the prior information in the form of initial position to determine the future locations. This method mainly involved usage of magnetic compass, which was pivotal and it completely changed how mankind navigated. Barring the fact that the dead reckoning accumulated errors, under certain constraints, it was quite accurate.

*New age, new learnings.*

Come twenty-first century, there are successful accomplished efforts to connect mankind in the outdoor space. The outdoor navigation systems includes the ubiquitous Global Positioning System (GPS) and regional systems like Russian Global Navigation Satellite System (GLONASS), DORIS, Galileo, China's BeiDou (now BedDou-2 or COMPASS), India's NAVIC (or IRNSS) and QZSS. With the explosion of usage of smartphones and cheap data connectivity has made it convenient and easy to navigate outdoors.

The peregrination for such a kind of connectivity has moved indoors. In the pursuit for cheap and reliable indoor positioning system, the research

community has tried solving the problem using Wi-Fi, Zigbee, GSM mobile network signal, RADAR, etc., but now it is lingering upon the Bluetooth Low Energy technology as it has the capability to become the de facto IoT connectivity device (Ahmad, 2015).

## 1.1 Motivation

It is well founded that we spend 80% of our time indoors and given the advancement in the radio-based communication today there is ample opportunity for infrastructure based positioning. The indoor setting is complex, dynamic and modeling such a data which is coupled with the stochasticity of signal propagation makes it a hard problem to solve.

It is also well known fact that GPS operations (like Assisted GPS (Djuknic and Richton, 2001)) are limited indoors and there is a void for solution which can seamlessly work both indoors and outdoors. There have been efforts to increase the GPS accuracy indoors using GPS-repeaters (Jardak and Samama, 2009) but this solution has greater initial costs and longer start-up time. In short, there is no wide-spread indoor positioning system (IPS) which integrates different solutions together.

*A scalable, seamless solution which performs equally well in both indoor and outdoor environment is the need of the hour.*

As the technology is becoming more context and location aware in the world of Internet of Things, for sensor (device) connectivity the location is imperative for intelligent solutions. The IPS opens up a great avenues for this with applications like assets tracking, product flow optimization, product recommendation etc., which would have market value of 10 billion US dollars by the year 2020 (He and Chan, 2016).

Unlike technologies like Ultra Wide Band solutions, which need heavy investments on infrastructural installations, *Wireless Local Area Networks* and *Bluetooth* are existential and cheap. With smartphones having Bluetooth capabilities, it calls for building a positioning solution on smartphone and make effective use of the sensor suite present in them (Langlois et al., 2017).

## 1.2 Indoor Positioning

Over two decades now, the research in the field *Indoor Positioning (IP; or Indoor Localization)* has seen both infrastructure based and infrastructure free solutions (Hazas et al., 2004). IPS is either device aided (Liu, 2008a)

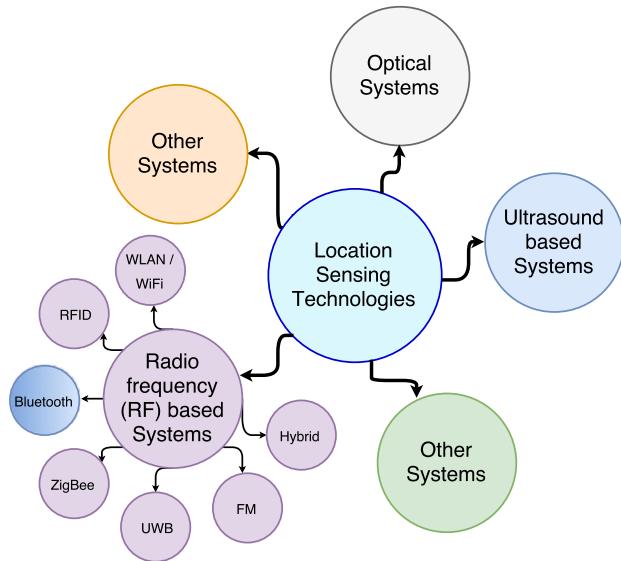


Figure 1.1: Technologies used in the Indoor Localization systems

or device-free (Patwari and Wilson, 2010) location sensing system which can accurately estimate the physical location of object (or person) indoors. These solutions are used for both research and commercial applications or to aid other location based services.

Indoor positioning systems in vague terms can be referred to as *indoor GPS*. Indoor environment can be office, hospital, shopping mall, school, airport, metro station etc., and positioning technology using BLE makes much more sense for a larger arena as the accuracy of such a system fall within *10 meters* accuracy (refer Figure 1.2). Research and development in indoor positioning would drastically proliferate the location aware and location-based applications (Hazas et al., 2004) making them a cornerstone in the world of IoT.

### 1.3 Bluetooth Low Energy

*Bluetooth Low Energy* (BLE) is a rather new low-powered wireless technology maneuvering in the 2.4 GHz ISM band space for short range communication. Unlike its predecessor, the BLE is configured for low-power solutions (Gomez et al., 2012). Hence, the BLE makes a trade off between power and performance. BLE has dragged the attention towards itself due to its low installation time and for being economical. Most of the other solutions require expensive infrastructure installations.

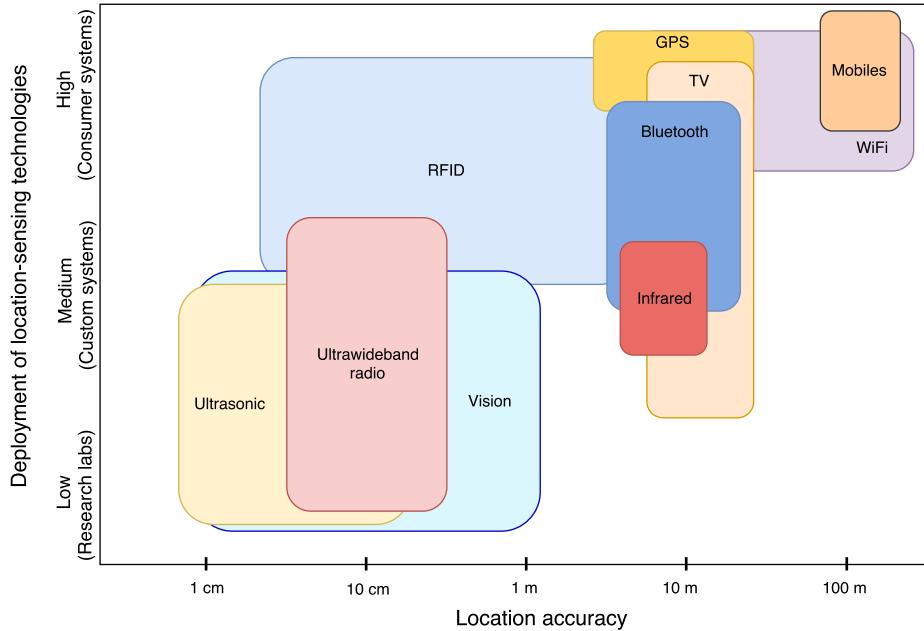


Figure 1.2: Technologies used in the Indoor Localization systems

In general, the BLE based positioning involves recording the signal strength data. This signal strength metric is called *received signal strength indication*. The RSSI data could be either directly triangulated to find the location or radio-maps could be learnt to estimate the location based on that.

Estimote's *nearables* is a new term for BLE plus the sensors. We discuss in detail about BLE's in Chapter 5 and on power-performance trade off in Chapter 6.

## 1.4 Challenges Plaguing the RSS based methods

Going about solving RSS based IP problem can be challenging as it is tagged by high non line-of-sight occurrences, effects of multiple obstacles, density and movement of human beings. This is exacerbated by signal fluctuations, channel interference and reflection leading to multipath, signal attenuation (Kaemarungsi and Krishnamurthy (2012), He and Chan (2016)). The problem of indoor positioning is inherently challenging (Roos et al., 2002) and it is compounded due to the stochastic nature of the indoor radio signal waves characterized by temporal and spatial non-stationarity (Hashemi, 1993). The dual non-stationarity of the radio signals is due to large-scale fading & multipath, reflection, refraction (Hashemi, 1993) and small-scale fading due to

dynamic nature of environment (Kaemarungsi and Krishnamurthy (2012), Luo et al. (2011a)). The problem is exacerbated by co-channel interference given the fact that BLE radio signals follows 2.4 GHz ISM band (Hashemi, 1993). As the radio waves are readily absorbed by water, a single human can attenuate the signal by -3.5 dBm (Bahl et al., 2000). It is also seen that BLE RSSI values are troubled wit rapid fluctuations and this could be attributed to the low bandwidth and low transmission power of BLE protocol in contrast to WiFi. This also makes BLE signals vulnerable to fast fading Faragher and Harle (2014).

## 1.5 Goal & Contribution of the thesis

Given the of infrastructure of the Bluetooth modules <sup>1</sup>, this thesis aims at finding optimal calibration parameters and best memory based method for positioning.

The main contributions of this thesis are:

- Comprehensive data analysis for the signal strength pertaining to Bluetooth low energy beacons.
- An overview of Gaussian processes in the field of indoor positioning.
- Using smartphone based measurements for getting the optimal calibration parameters and evaluating the memory based methods.

## 1.6 Outline of the thesis

This monologue is structured as follows. Chapter 2 covers the concept of fingerprinting and methodology for generating the radio-maps for the filtering based positioning methods. It also covers a brief look back into non-memory based algorithms and related technologies concerning IP. The theoretical distinction between memory and non-memory based methods is also given. In Chapter 3, a detailed background of Gaussian processes for IP is given. In Chapter 4, we introduce and explain the memory based methods i.e., the non-linear Bayesian filtering algorithms. The measurement setup, including the measurement application and BLE modules, for the experiments is described in the Chapter 5 and Chapter 6 delves into statistical analysis of the RSSI data from the beacons and explores how different factors affect the

---

<sup>1</sup>or access point as it can simultaneously advertise and read the BLE signals, hence the word *module*. We will interchangeably use module or beacon in the thesis.

signal indoors. In Chapter 7, we evaluate the memory based methods using the different optimal calibration parameters discussed therein and results are stated. We discuss the results in Chapter 8 and also state the future course of action.

# Chapter 2

## Location Fingerprinting

In this chapter, we delineate the background for the thesis with the relevant previous work. In Section 2.1, we introduce the concept of location fingerprinting technique and then move on to preliminary concepts building the radio-maps from the reference table. In Section 2.2, we introduce the RSS based methods and its categorization in terms of deterministic and probabilistic algorithms. Here, we also describe a non-memory based method. In Section 2.3 does a review of previous IPS's. In Section 2.4, we look at the performance metrics for evaluating the positioning algorithms.

### 2.1 Location Fingerprinting

Fingerprinting is a method for signal pattern recognition (Aguilar-Garcia et al., 2015). It exploits relationship between the signal based characteristics to the location for positioning (Kaemarungsi and Krishnamurthy, 2012). It was first introduced for wireless local area network based positioning (Bahl and Padmanabhan, 2000). A typical fingerprinting based localization technique juxtaposes the RSSI to the already ones present in the form of radiomap or a reference table (Gu et al., 2016). No matter how advanced the algorithm is, deep down a RSS based method would performs a similar logic. Yiu et al. (2017) calls localization using the reference map as traditional fingerprinting.

The dense network of beacons would allow the fingerprints to be unique corresponding to a unique location Huang et al. (2011). Fingerprinting is much researched because of its ease in implementation and economical in terms of usage of existing infrastructure (Yiu et al., 2017).

Being a laborious task, one of the few proposed reduced fingerprinting techniques like compressive sensing based *sparsity rank singular value decomposition* in conjunction with k-Nearest Neighbour to compensate for the

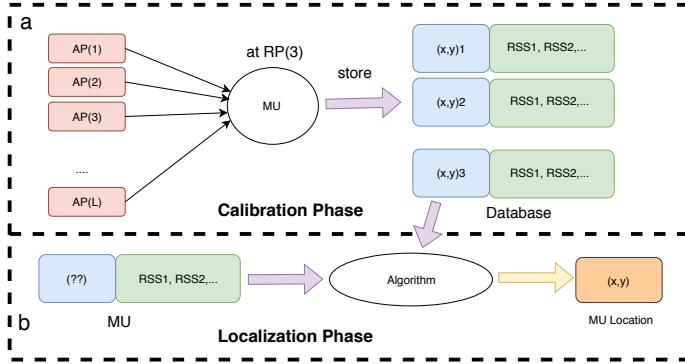


Figure 2.1: An overview of fingerprinting technique.

missing values and mitigate the redundancies caused due to multipath and interference (Zhang et al., 2013). Yiu et al. (2017) enumerates the effects of outdated radiomaps shows that the positioning accuracies are overestimated. They show it by using realistic and unrealistic scenarios.

Fingerprinting based positioning method is categorized into two phases: *offline calibration phase* and *online localization phase*. We describe them next.

1. *Calibration Phase:* It is an offline training phase where we measure the RSSI values of the radio-frequency signal from all the detectable BLE modules at a particular *calibration point* for a certain time called *calibration time*. These measurements are called *fingerprints* (or signatures (Yiu et al., 2017)) and are location specific. As illustrated in the Figure 2.1(a), this phase involves creating a database of fingerprints called *reference table*. This database is later used for generating a *radiomap*.
2. *Localization Phase:* It is a online testing phase where the real-time signal strengths are recorded and location is estimated using an algorithm. This phase makes use of the radiomap generated in the calibration phase. The same is illustrated in the Figure 2.1(b).

### 2.1.1 Radiomap

As already mentioned, the fingerprinting methodology involves learning the spatial properties of the signal in the calibration phase. Radiomaps are defined as the continuous interpolation of the signal strength values over the state space for a particular access point. There are other notable characteristics like *Signal to Noise ratio (SNR)* but the signal strength has stronger

correlation to the distance. In fingerprinting technique, we are generally interested in the variation of RSSI values over the localization space, hence, the units of RSSI is irrelevant<sup>1</sup>. We discuss the relation of the signal strength and distance in detail in Chapter 6.

The design of radiomap is quite critical to the performance of the positioning system. It involves selecting the variables like calibration points and calibration time. These variables vary for different indoor settings. For example, a ware house might need more accurate state estimation i.e., less than few centimeters but locating a room inside a office would allow us for error upto few meters.

The radiomap can be achieved in different ways either based on propagation model (Klepal et al., 2007) or based on fingerprint model (Gu et al., 2016) or based on Gaussian process regression model (Schwaighofer et al., 2004). Previous notion is that the reference table (discussed in 2.1.1.1) is a radiomap which looks quite inappropriate.

Bahl and Padmanabhan (2000) in their pioneering paper showed the usage of radio-based localization. Many distributed systems for creating and updating the radiomap have been proposed, for example Kriz et al. (2016). Alternative to the conventional radiomap, only the difference of the RSSI values over the raw data could also be used. This was devised to mitigate the receiver gain difference in the devices (Wang et al., 2011).

### 2.1.1.1 Reference Table

The first step in the creation of the radiomap is to generate a *reference table* (refer figure (2.2)). A reference table is a collection of calibration points and RSSI from all the access points. Mathematically, it is

$$\mathcal{R} = \bigcup_{i=1}^N \bigcup_{j=1}^M \bigcup_{k=0}^{P_{ij}} \{(x_i, y_i), r_{jk}\}, \quad \text{if } k = 0, \quad r = r_{min} \quad (2.1)$$

where  $N$  is number of calibration points,  $M$  is number of access points,  $P$  is location and access point depend RSSI values and  $r_{min}$  is empirical minimum of all RSSI values in case an access point is not heard. It should be noted that due to indoor stochasticity the number of signal strength values vary for different access points and may vary in time even from the same access point. For that purpose, we introduce two reference tables. *Reference table 1* ( $R_1$ ), where there is only single entry of each calibration point. It is a common practice to use the mean estimate of the all the RSSI values collected for a particular access point at a certain calibration point, but there

---

<sup>1</sup>RSSI are mentioned in terms of signal power i.e., mW or dBm.

	1	2	3	4	5	6	7	8	9	10	1
1	1.5000	0	-72.3333	-80	-90.6667	-92.8571	-94.5000	-91.5000	-90.8333	-88.1250	-86
2	3	2.3500	-93	-93	-93	-93	-93	-94.2500	-93	-93.2000	
3	-3	1.1700	-93	-93	-93	-93	-93	-92	-93	-98	
4	6	2.1500	-60.2857	-88.9000	-90.8333	-96	-93.6250	-89.0714	-94.1250	-87.4545	-89
5	8	0	-67.4000	-80.8462	-90.0769	-85.0833	-91	-92.3333	-88.6667	-94	
6	10.5000	2.3500	-76.2222	-68.0667	-74.8000	-79	-84	-75.0909	-84.3333	-79.5833	-63
7	13	0.3000	-69	-65.5000	-76.8750	-85.4286	-86	-80.2308	-78.5385	-72.8667	-71
8	15	2.2500	-67.2222	-67.1222	-60.6222	-78.2000	-85.7500	-75.5000	-82.0000	-75.0000	-62

Figure 2.2: Reference table generated as part of the fingerprinting technique. The first two columns represent the location in the form of *x coordinate* and *y coordinate* from a specific reference point. Depending on the number of access point the rest of columns are filled. !FIXME replace this with two columned table. first columns showed mean estimated reference table. second non mean estimated reference table. FIXME!

is the down side of losing variance in the data. Hence, this is also called the *mean reference table* ( $\mathcal{R}_{mean}$ ).

$$\mathcal{R}_{mean} = \bigcup_{i=1}^N \bigcup_{j=1}^M \{(x_i, y_i), r_j^{mean}\}, \quad \text{if } r_j^{mean} \in \emptyset, \quad r_j^{mean} = r_{min} \quad (2.2)$$

Next, *reference table 2* ( $\mathcal{R}_2$ ), where the equation of reference table remains as is as defined in Equation 2.1, therefore, there are multiple entries of RSSI values  $r_{jk}$  for a single calibration point and the number for that could be variable. The advantage of  $R_2$  over  $R_1$  is that the variance of the RSSI for a particular beacon is intact and that adds to the accuracy of the constructed GP radiomap.

Yiu et al. (2016) suggests to use the minimum power device sensitivity level in case of unheard access point, but in practice the mobile unit can record a RSSI value below minimum sensitivity level. In this thesis if a particular module is unheard or is below device sensitivity level, we update them to be the device sensitivity level which is -110 dBm. So, the  $r_{min}$  from the Equation 2.1 is equal to the device sensitivity level.

### 2.1.2 Raw data: RSSI

Received Signal Strength Indication (RSSI) is a measurement of power present in the radio signal at the point of reception. RSSI is one of the mostly used

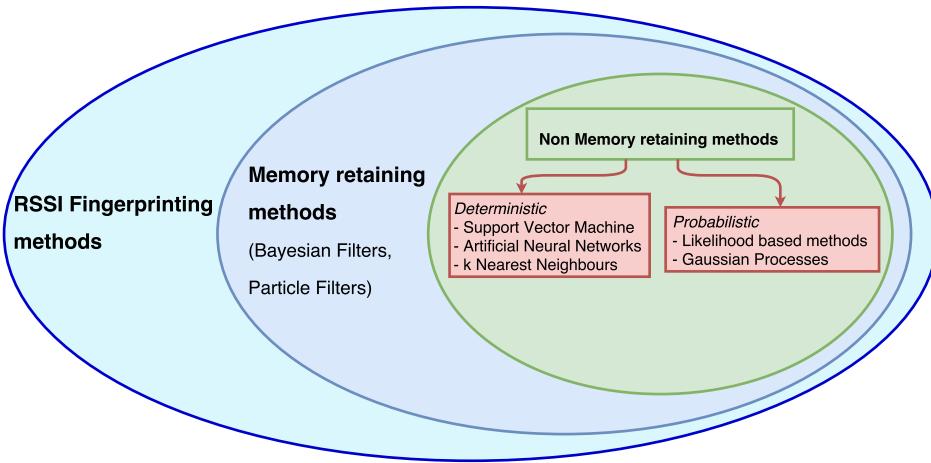


Figure 2.3: Illustration of received signal strength indication value decreases with increase in the distance of the luminaire

parameters from the radio signals for indoor positioning. It is widely used as a part of 802.11 standards. It is a *pseudo location sensor* and is the de facto measurement in the applications related to indoor positioning. As shown by Bahl and Padmanabhan (2000) that the signal strength shows proportional decrease with distance.

## 2.2 RSS Localization algorithms

The RSS based methods can be categorized as *Non-memory based* and *Memory based*. The non-memory based methods are a special case of memory based methods (refer Figure 2.3). The various non-memory based methods could be used as the measurement models which can sufficiently capture the data generating process.

### 2.2.1 Non-memory based methods

The non-memory based methods are range based methods (Aravecchia and Messelodi, 2014), where the RSSI is converted to a distance measure between the mobile unit and the access points. This measure is used for localizing the mobile unit. As seen from the Figure 2.3, the non-memory methods could either deterministic or probabilistic, or even non-parameteric.

Various approaches have been employed for eg. like using polynomial curve to learn non-linearities in the RSSI (Feldmann et al., 2003), or by using space grids method with the mean of RSSI values (Elnahrawy et al. (2004),

Luo et al. (2011b)), or by deploying the path loss model (Zanca et al. (2008), Pivato et al. (2011)). Bahl and Padmanabhan (2000) used *Cohen-Sutherland line clipping algorithm*(Foley, 1990) to obtain the accurate building layout information to estimate the radio propagation model.

The memory based methods are the range-free solutions (Aravecchia and Messelodi, 2014). These stochastic methods use a likelihood model to compare the current RSSI measurements to predicted RSSI from the previously built knowledge base in the training phase. The knowledge base can be either the reference table or radio maps, and the precision of measurements depends on the type of measurement model used. The predictions are computed over the parameter state space, which could be a grid of points, or dispersed location particles. We describe one of the popular non-memory based deterministic method namely k-Nearest Neighbour in the next section. We would describe about Gaussian processes in Chapter 3.

### 2.2.1.1 k-Nearest Neighbour

In the *k-Nearest Neighbour (k-NN)* (Honkavirta, 2008) method, the measurements from multiple beacons are collected and they are compared to the mean reference table generated in the calibration phase. The comparison could be as simple as *euclidean distance* (or  $l^2$  norm) in between the measurement samples is

$$l^2 = \left( \sum_{i=1}^M |r_1 - r_2| \right)^{1/2}, \quad (2.3)$$

where  $M$  is the dimension of the vector equal to the number of access points,  $r_1$  and  $r_2$  are RSSI fingerprints. The top "k" nearest calibration points in terms of calculated euclidean distance are selected and the mean of their physical location is taken as the estimated locations. Let a vector of measurement  $r_t$  be recorded at time  $t$  be compared to measurements from mean reference table  $\mathcal{R}^{mean}$ , then

$$L_N^2 = [l_1^2, \dots, l_N^2] \quad (2.4)$$

is the vector of ascending distances of  $r_t$  to all fingerprints in  $\mathcal{R}^{mean}$  and corresponding calibration points be

$$(\mathcal{X}, \mathcal{Y}) = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}. \quad (2.5)$$

Then the estimation location taking  $\{(x_1, y_1), \dots, (x_k, y_k)\}$  is

$$(\hat{x}, \hat{y}) = \frac{1}{k} \sum_{i=1}^k (x_i, y_i) \quad (2.6)$$

A variation of this method called *weighted k-nearest neighbour (wk-NN)* uses a weighted mean as shown

$$(\hat{x}, \hat{y}) = \frac{1}{k} \sum_{i=1}^k w_i \cdot (x_i, y_i) \quad (2.7)$$

The weights denote the reliability of the measurements as shorter distance gives accurate measurements. Any criteria for calculating the weights can be used, for example, Rizos et al. (2007) uses the inverse of the distance as weights. Bahl and Padmanabhan (2000) gives the justification for using k-NN and tells that if the orientation of the error vector is different for each beacon, then averaging would lead us closer to the true location, given that we have dense network of APs.

### 2.2.2 Memory based methods

The memory based methods are probabilistic methods used in conjunction with non-memory based methods. The probabilistic nature of the positioning solution would allow seamless integration of inertial sensor data for increased performance. Refer to Chapter 4 for detailed discussion on these methods.

## 2.3 Related Work

The first IPS the *infrared-based Active Badge system* was developed by Roy Want, Andy Hopper, and others in 1989 (Hazaras et al., 2004). This used wall-mounted sensors which would read the infrared ID advertisements by occupants wearable tags. An ultrasonic based application *Cricket indoor system* was developed at MIT (Priyantha, 2005) which uses a method similar to triangulation for localization. This method like the RFID requires the occupant to wear a tag. Another ultrasonic system AT&T's Active Bats systems (Addlesee et al., 2001) also used tags in the form of ultrasonic badges and needed ultrasound receiver installed. Ubisense, a ultra-wide band based real time localization system uses pulsed signals which can accurately measure the time of arrival and has accuracy of 15 centimeters (Steggles and Gschwind, 2005). The ultra-wide band technology doesn't suffer from non line of sight issues. Ekahau uses the WLAN signal strength measurements for positioning (Aittola et al., 2003).

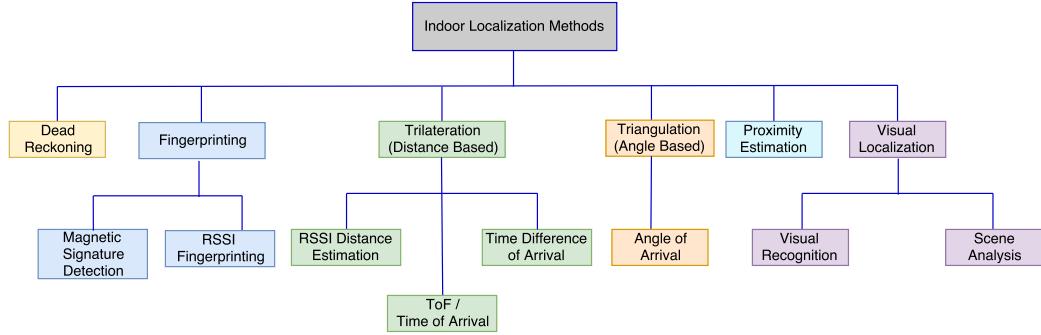


Figure 2.4: Taxonomy of Indoor Positioning

Accurate positioning systems required extensive auxiliary infrastructural services with sensors and time expensive calibration. Subsequently, the research community focused on radio based methods including WiFi, Bluetooth, radio frequency identification tags and making the short range technologies solutions fine-grained for accurate (less than 30 cms) location estimation. But these technologies couldn't compete with wide coverage technologies like GPS, cellular networks, TV broadcasts.

taxonomy of indoor positioning methods (Langlois et al., 2017).

## 2.4 Performance metrics

The performance of the positioning methods are evaluated in terms of the absolute error of the prediction. This thesis aims at one-shot prediction of the position and the optimal criterion is averaged root mean square error (RMSE). The other criteria used are the mean, 90th percentile, maximum error and variance of the error estimates.

## Chapter 3

# Gaussian Process Model for Bayesian Filtering

*Neural networks are popular learning methodology to fit any non-linear function. Gaussian Processes can be looked on as Bayesian neural networks, where the neural network model is the prior distribution and learning in the form of weights, the posterior distribution.*

- Radford M. Neal, 1996

Gaussian Processes (GPs) are the non-memory based non-parametric<sup>1</sup> method. In this chapter we formally introduce GPs in Section 3.1. Next we give the motivation for using GPs in Section 3.2. Then we enumerate the advantage of using GPs in Section 3.3. Finally, we explain in detail the different GP modeling techniques for IP in Section 3.4.

### 3.1 Gaussian Processes

GPs are random processes for describing distributions over functions (Rasmussen and Williams, 2005). The term 'process' originates from signal processing while 'random fields' could also be used in spatial statistics (Solin, 2016) context. After the modeling is done, the spatial properties of the learnt function are inferred depending on the kernel function used (Solin, 2016).

**Definition 3.1.1.** A Gaussian process is a collection of random variables, of

---

<sup>1</sup>going by the definition of basic Gaussian process, mean of Gaussian is non-parametric, but the conditional distribution is Gaussian, i.e, parametric. Here, as the computation increases with the number of data points, GPs are strictly non-parametric.

which any finite collection of variable follow a joint Gaussian distribution (Rasmussen and Williams, 2005).

For a latent stochastic process  $f(\mathbf{x})$ , we define mean function  $m(\mathbf{x})$  and covariance function  $k(\mathbf{x}, \mathbf{x}')$  as

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}_p[f(\mathbf{x})] \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}_p[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \end{aligned} \quad (3.1)$$

and then we can draw inferences over function  $f(x)$  by putting GP prior as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (3.2)$$

on a function.

GPs are generalization of multivariate Gaussian distribution (Rasmussen and Williams, 2005). Contrary to sampling from distributions which yields finite dimensional vectors, sampling here would yield infinite dimensional vectors of which any finite section would follow a multivariate Gaussian distribution.

### 3.1.1 Application in Indoor Postioning

GPs were first used by Schwaighofer et al. (2004) and later on by Li et al. (2005) as error correction map for the non-line-of-sight problem for positioning in the cellular networks. Localization was performed using GPs and k-NN algorithm on a cellular network. They showed that GP's gave an accuracy of 7.5 m, bit worse than kNN's 7 m. Matérn covariance function was used for GP modeling of individual stations and maximum likelihood was used for estimation of location. Yiu et al. (2016) talks about hybrid method i.e., log path loss model as the mean function in the GP's which is in similar lines with Schüssel and Pregizer (2015).

Atia et al. (2013) proposed a solution like the GPS using the RSSI of the WiFi network. The method had no overhead in the form of fingerprinting or had any additional infrastructural needs. They used a hybrid propagation model using GPs and showed that it could model RSS residuals better than path loss models. Their proposed system gave 2-3 m accuracy.

## 3.2 Motivation for using the GPs

Due to complexity of the indoor environment coupled with interference, multipath propagation of radio-frequency signals, obstacles, leading to distorted

spatial distributions of the signal strength values. The simple parametric distributions are inadequate in modeling the complex RSS distributions (Seco et al., 2010). Hence, we need flexible models for tackling this problem and Gaussian processes are perfect solution for that.

### 3.3 Advantages of GPs

There are various advantages of GP's which fit modeling signal strength based localization problems. Few of the important advantages is enumerated below (Ferris et al., 2007).

1. *Continuous Locations*: Traditionally, GPs were known as *kriging*, which is an regression task, hence they can predict the signal strength values (with the uncertainty estimates). GPs have excellent capabilities of interpolating over other test locations. They are flexible as they don't need any designated training points for accomplishing this task.
2. *Arbitrary likelihood models*: A wide variety of complex data models can be approximated given the non-parametric nature of GPs: multiple kernels could be used in conjunction with each other Rasmussen and Williams (2005). Hence, GPs are can model highly non-linear signals such as RSSI (Aravecchia and Messelodi, 2014).
3. *Correct uncertainty handling*: As the GPs come with a Bayesian flavor, along with the mean estimates they also spit out the uncertainty estimates for each value in the state space. This is mainly dependent on amount of data and the estimated noise around the test points (Faragher and Harle, 2014).
4. *Consistent parameter estimation*: The model selection problem in GPs helps solving the obtaining the optimal (hyper-) parameters. This is done via the maximizing the marginal likelihood Rasmussen and Williams (2005). These point to spatial correlation between measurements and learn the measurement noise (Ferris et al., 2007).

### 3.4 GP modeling for indoor positioning

In a broader sense, the GPs could be modeled in following ways:

1. *Indirect modeling* (Aravecchia and Messelodi, 2014): A widely used approach to apply GP for positioning is through the following the equation:

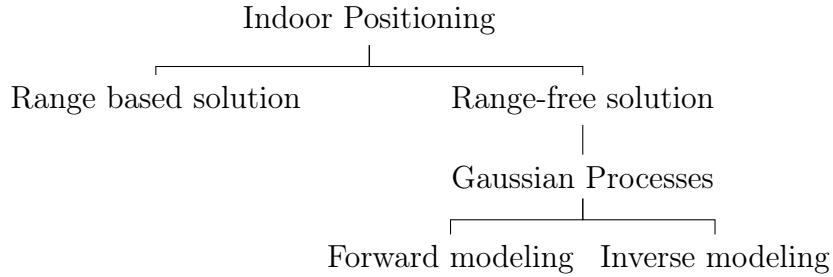


Figure 3.1

$$s_j = f(x) + \epsilon \quad (3.3)$$

where  $s_j$  is the value of RSSI at the location  $x$  for  $j$ -th access point. Hence, the GPs could be modeled inversely from metric space to signal space.

$$\begin{aligned} GP : \mathbf{R}^d &\rightarrow \mathbf{R} \\ x &\mapsto s \end{aligned} \quad (3.4)$$

This might look unremarkable but works for most of the problems and could be directly applied from the filtering point of view. It enables us to model the signal strengths as *latent variables* and learn its characteristics over the position state space. The characteristics are recorded in the form *radio-maps*. Radio maps are discussed in Chapter 2. GP here could be exploited in the form of measurement model using the learnt radio maps, as in (Ferris et al., 2007).

With the ease comes along few limitations, like the quality and amount of the fingerprint data for constructing the radio map, which is a laborious task. This approach has been called *Forward* modeling by Schwaighofer et al. (2004) which is quite counter intuitive.

2. *Direct* modeling (Aravecchia and Messelodi, 2014): Logically, it would be suitable if we could get the location estimate directly from the RSSI values i.e., from signal space to metric space

$$x = f(\mathbf{s}) + \epsilon \quad (3.5)$$

where  $\mathbf{s} = s_{1:j}$  is an array of RSSI measurements from  $j$  access points riddled with the noise  $\epsilon$  at the location  $x$ . Hence, the GPs could be modeled directly from signal space to metric space.

$$\begin{aligned} GP : \mathbf{R}^q &\rightarrow \mathbf{R}^d \\ \mathbf{s} &\mapsto x \end{aligned} \tag{3.6}$$

This could be achieved through *Maximum Likelihood estimation (MLE)*, which is entrenched by type and convexity of likelihood function, and its initialization. It is not uncommon fact that MLE innately suffers from over-fitting (Bishop, 2006). This approach has been called *Inverse* modeling by Schwaighofer et al. (2004), which is again quite counter intuitive. One observation from equation 3.5, evidently a vice, is that it assumes that the input RSSI values are noise free i.e., we tend to ignore the stochasticity of the signal propagation.

3. *Hybrid* modeling: !FIXME add reference to the previous equations  
FIXME! The *hybrid* modeling tries to overcome the limitations of *indirect* and *direct* modeling. It is an augmented form of direct model and is constructed two fold. This model overcomes the problem of initialization by intelligently using the indirect model to overcome its problem of initialization Aravecchia and Messelodi (2014). This forms the first GP fold. The second fold uses these vague location estimates and runs it through the indirect model using the MLE to get the updated location estimates. The hybrid model shows a crude mimicry of Bayesian filtering approach Särkkä (2013). The first fold mimics the prediction step which is formed using the dynamic model and the second fold mimics the update step which is formed using the measurement model.

$$\begin{aligned} \tilde{x} &= f_{GP_1}(\mathbf{s}) \\ x &= g_{GP_2}(\tilde{x}) \end{aligned} \tag{3.7}$$

where  $f$  is function which follows  $GP_1$  from the equation 3.6 while  $g$  follows  $GP_2$  from the equation 3.4,  $\tilde{x}$  is the predicted estimate of location from signal strengths  $\mathbf{s}$  from luminaires whereas  $x$  is the updated location estimate.

!FIXME change it; for reference only. FIXME!

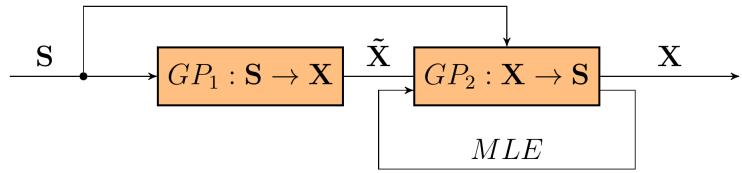


Figure 3.2: Hybrid model. courtesy Aravecchia and Messelodi (2014).

To incorporate the stochasticity in the RSSI values, we can also model the inputs as *Noisy Input GP (NIGP)* and also include a Gaussian prior on the state (Aravecchia and Messelodi, 2014).

$$\begin{aligned}
 NIGP : & \mathbf{R}^q \rightarrow \mathbf{R}^d \\
 \mathbf{s} & \mapsto x
 \end{aligned} \tag{3.8}$$

## Chapter 4

# Memory Based Filtering Methods

*When the facts change, I change my mind. What do you do, sir?*  
- John Maynard Keynes

In this chapter we introduce the Bayesian filtering and describe the state space modeling equation and derive a simple prediction and update equations. Then we will explain the memory based filtering methods Particle and Unscented Kalman filters.

## 4.1 Bayesian Filtering

Most of the real-world problems circumscribe evaluating an unknown quantity given some external measurements. In most of the cases, sufficient prior information about the data generating and dynamic processes are available. This a classic setup for using Bayesian inference and forms another view for optimal filtering Särkkä (2013). It involves incorporating the initial notion of the unknown quantity giving rise to updated belief about the quantity. Precisely, the prior distribution is updated in the light of new evidence to give rise to posterior distribution which allows for inference on the quantity of interest (for e.g. location of person). For comprehensive overview on Bayesian inference, refer (Gelman et al., 2014).

In the field of *sensor informatics* and *time varying systems*, the data arrives sequentially and the marginal posterior needs to be updated simultaneously. This Bayesian recursive state estimation can be termed as *Bayesian Filtering* (Särkkä, 2013). Bayesian Filtering resembles the statistical inversion problem (Särkkä, 2013) with states as latent variables and measurements as observed variables, figure 4.1. This finite state-space Markov chain with data partially observed can be termed as hidden Markov model filter (Doucet

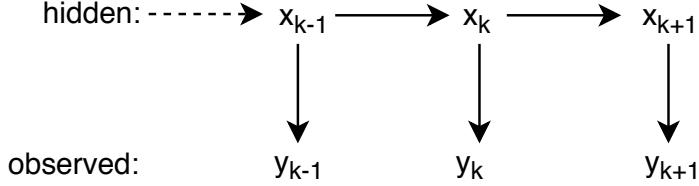


Figure 4.1: The Hidden Markov model illustrating the states  $x$  being indirectly estimated from the measurements  $y$ .

et al., 2001). Applications can be found in the field of navigation, telecommunication, economics, etc.

### 4.1.1 State Space Modeling

We consider the following state space model

$$\begin{aligned} \mathbf{x}_t &= f(\mathbf{x}_{t-1}, \boldsymbol{\theta}) + \mathbf{q}_{t-1}, \\ \mathbf{y}_t &= h(\mathbf{x}_t, \boldsymbol{\theta}) + \mathbf{r}_t. \end{aligned} \quad (4.1)$$

Restricting ourselves to Markovian, non-linear and non-Gaussian scenario, (Doucet et al., 2001), the unobserved states (hidden or latent variables) are  $\{\mathbf{x}_t; t \in \mathbb{N}\}$ ,  $\mathbf{x} \in \mathbb{R}^n$  are modeled as a *Markov process*. Using the initial distribution  $p(\mathbf{x}_0)$ , transition equation  $p(\mathbf{x}_t|\mathbf{x}_{t-1})$  and marginal distribution  $p(\mathbf{y}_t|\mathbf{x}_t)$  given the noise riddled observed measurements  $\mathbf{y} \in \mathbb{R}^m$ , we write our model as

$$\begin{aligned} p(\mathbf{x}_0) \\ p(\mathbf{x}_t|\mathbf{x}_{t-1}) \text{ for } t \geq 1 \\ p(\mathbf{y}_t|\mathbf{x}_t) \text{ for } t \geq 1 \end{aligned} \quad (4.2)$$

We aim at iteratively computing the *posterior distribution*  $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ , principally the *filtering distribution*  $p(\mathbf{x}_t|\mathbf{y}_{1:t})$  and its expectation

$$I(g_t) = \mathbb{E}_{p(\mathbf{x}_t|\mathbf{y}_{1:t})}[g(\mathbf{x}_t)] \triangleq \int g(\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t}) d\mathbf{x}_t \quad (4.3)$$

where  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is an arbitrary function,  $\mathbf{x}_{0:t} \triangleq \{\mathbf{x}_0, \dots, \mathbf{x}_t\}$  and  $\mathbf{y}_{1:t} \triangleq \{\mathbf{y}_1, \dots, \mathbf{y}_t\}$ . The filtering solution involves first computing the joint posterior distribution of the states and this can be accomplished using the Bayes' rule:

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t})}{p(\mathbf{y}_{1:t})} \quad (4.4)$$

where

- $p(\mathbf{x}_{0:t})$ , is the dynamic model which forms the prior distribution,
- $p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})$ , is the measurement model<sup>1</sup> which forms the likelihood function,
- $p(\mathbf{y}_{1:t})$ , is the evidence and is a normalizing constant.

$$p(\mathbf{y}_{1:t}) = \int p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t}) \quad (4.5)$$

To obtain the recursive equation lets consider a new observation  $y_{t+1}$ , hence, the updated joint posterior is

$$p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1}) = p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{x}_t)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})}. \quad (4.6)$$

The filtering distribution  $p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1})$  can be recursively solved by

- *Prediction:* Using the *Chapman-Kolmogorov equation*, we get

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) = \int p(\mathbf{x}_{t+1}|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t}) d\mathbf{x}_t \quad (4.7)$$

- *Updating:* Using the current measurement  $\mathbf{y}_{t+1}$ ,

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1}) = \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t})}{\int p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) d\mathbf{x}_{t+1}} \quad (4.8)$$

Though the filtering equations look fairly straightforward, the evidence  $p(\mathbf{y}_{1:t})$  is unavailable and  $I(g_t)$  involves integrating highly non-linear function in high dimensions (Doucet et al., 2001).

If assumed Gaussian state-space model, the filtering solution gives rise to the optimal *Kalman filter* which has a neat closed form analytical expression (Särkkä (2013), Doucet et al. (2001)). This assumption at times doesn't hold as the empirical data suffers from non-Gaussianity, high dimensionality and non-linearity (Doucet et al., 2001) which requires filtering techniques with statistical workarounds and posterior approximations. These methods include extended and unscented Kalman filters, Gaussian filter, Gauss-Hermite Kalman filter, Cubature Kalman filter and Particle filters (Särkkä, 2013). Particle filters are the sequential Monte Carlo methods of which Bayesian filtering is a special case.

---

<sup>1</sup>also called data model or observation model. Here, We would use the terms interchangeably.

### 4.1.2 Advantages of Bayesian Filtering

One major advantage is that the solution degrades gracefully.

## 4.2 Filtering

In this section we will describe the probabilistic filtering methods. Based on the previous literature, we select only promising filtering

### 4.2.1 Particle Filters

*Sequential Monte Carlo* (SMC) methods are simulation-based methods for generating draws from target distribution through sequentially generating weighted particles from intermediate sampling distributions. The weights of the particles are corrective measures for bias reduction with respect to corresponding *auxiliary distribution* (Liu (2008b), Doucet et al. (2001)).

#### 4.2.1.1 Importance Sampling

*Importance Sampling* (Gelman et al., 2014) is an efficient Monte Carlo integration method when the sampling from the target distribution is implausible. It was named "importance" so as to underscore the important regions which becomes critical in the high dimensional posterior space (Liu, 2008b). In this method, weighted samples are drawn from an approximating *importance distribution* and expectation is obtained by weighted mean calculation.

Let  $\pi(\mathbf{x}_t|\mathbf{y}_{1:t})$  be the importance distribution and  $p(\mathbf{x}_t|\mathbf{y}_{1:t})$  be the target distribution. We are naturally interested in the

$$E_{p(\mathbf{x}_t|\mathbf{y}_{1:t})}[g(\mathbf{x}_t)] = \int g(\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t}) d\mathbf{x}_t = \int \left[ g(\mathbf{x}_t) \frac{p(\mathbf{x}_t|\mathbf{y}_{1:t})}{\pi(\mathbf{x}_t|\mathbf{y}_{1:t})} \right] \pi(\mathbf{x}_t|\mathbf{y}_{1:t}) d\mathbf{x}_t. \quad (4.9)$$

The choice of importance distribution is critical and has to be non-zero in the important posterior regions. The Monte Carlo approximation of  $N$  samples drawn from  $\pi(\mathbf{x}_t|\mathbf{y}_{1:t})$  is:

$$\begin{aligned} E[g(\mathbf{x}_t)|\mathbf{y}_{1:t}] &\approx \frac{1}{N} \sum_{i=1}^N \frac{p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t})}{\pi(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t})} g(\mathbf{x}_t^{(i)}) \\ &= \sum_{i=1}^N w_t^{(i)} g(\mathbf{x}_t^{(i)}) \end{aligned} \quad (4.10)$$

where the weights  $w_t$  are defined as:

$$w_t^{(i)} = \frac{1}{N} \frac{p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t})}{\pi(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t})} \quad (4.11)$$

Now, the approximate filtering distribution can be written as:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx \sum_{i=1}^N w_t^{(i)} \delta(\mathbf{x}_t - \mathbf{x}_t^{(i)}) \quad (4.12)$$

where the  $\delta(\cdot)$  is Dirac delta function.

As we don't have the filtering distribution readily available, we compute it using the Bayes's rule (ref equation 2.3) !FIXME **use ref** !FIXME!. Hence,

$$p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)})}{\int p(\mathbf{y}_{1:t}|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)})} \quad (4.13)$$

Using equation 4.13 in equation 4.9, we arrive at

$$\begin{aligned} E[g(\mathbf{x}_t)|\mathbf{y}_{1:t}] &= \sum_{i=1}^N \left[ \frac{\frac{p(\mathbf{y}_{1:t}|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)})}{\pi(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t})}}{\sum_{j=1}^N \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_t^{(j)})p(\mathbf{x}_t^{(j)})}{\pi(\mathbf{x}_t^{(j)}|\mathbf{y}_{1:t})}}} \right] g(\mathbf{x}_t^{(i)}) \\ &= \sum_{i=1}^N \tilde{w}^{(i)} g(\mathbf{x}_t^{(i)}). \end{aligned} \quad (4.14)$$

Notice that the importance weights are self-normalized. The stability of importance weights and existence of higher moments ensure the convergence of the estimates as the central limit theorem holds. In case of ill fit of the proposal distribution to the posterior distribution, the weights would eventually have infinite variance due to the presence of heavy right tail (Vehtari et al., 2015), we need sophisticated methods to solve this issue.

**Algorithm 1:** Importance Sampling

1 Sample from the importance distribution:

$$\mathbf{x}_t^{(i)} \sim \pi(\mathbf{x}_t | \mathbf{y}_{1:t}), \quad i = 1, \dots, N$$

2 Calculate the unnormalized weights  $w_t$ :

$$w_t^{(i)} = \frac{p(\mathbf{y}_{1:t} | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)})}{\pi(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t})}, \quad i = 1, \dots, N$$

3 Normalize the weights:

$$\tilde{w}^{(i)} = \frac{w^{(i)}}{\sum_{j=1}^N w^{(j)}}, \quad i = 1, \dots, N$$

4 The approximate filtering distribution and its expectation of  $g(\mathbf{x}_t)$  is given by

$$p(\mathbf{x}_t | \mathbf{y}_{1:t}) \approx \sum_{i=1}^N \tilde{w}_t^{(i)} \delta(\mathbf{x}_t - \mathbf{x}_t^{(i)})$$

where the  $\delta(\cdot)$  is Dirac delta function and

$$E[g(\mathbf{x}_t) | \mathbf{y}_{1:t}] \approx \sum_{i=1}^N \tilde{w}_t^{(i)} g(\mathbf{x}_t^{(i)})$$

.

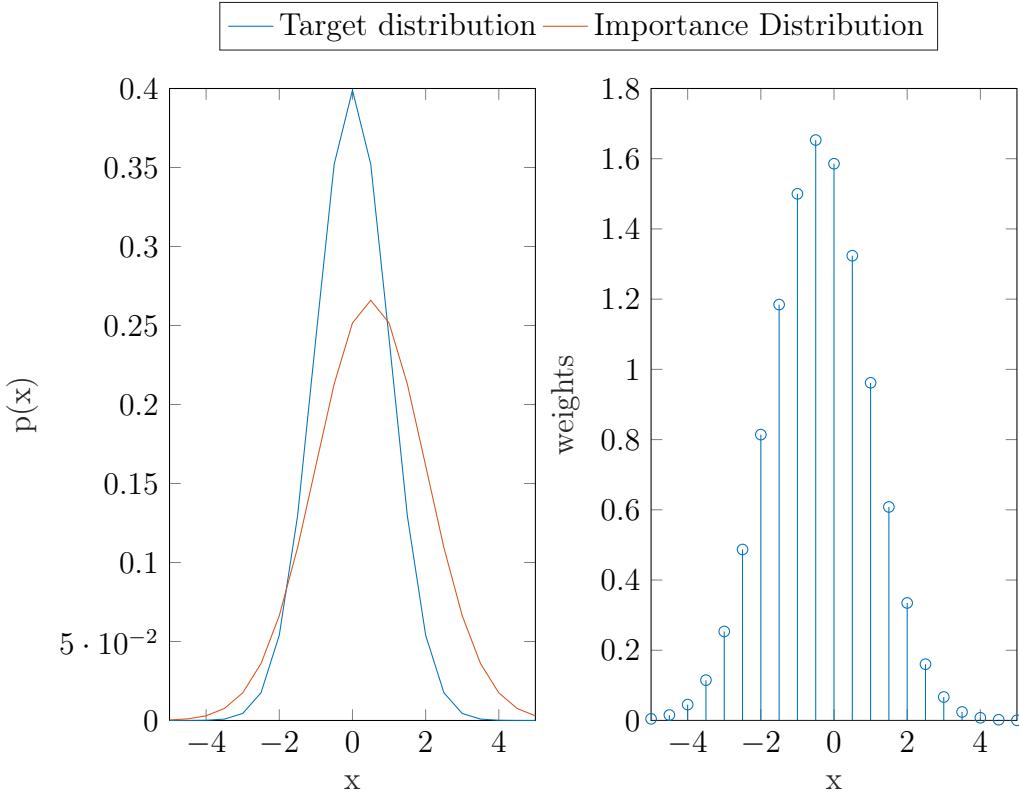


Figure 4.2: (left) The importance distribution and target distribution. (right) Approximate posterior distribution in the form of weights.

### 4.2.2 Algorithm

Importance sampling algorithm when recursively iterated with resampling gives the *sequential importance resampling (SIR)* or *particle filters*. *Resampling* is a rejuvenating step for dealing with the degeneracy problem (Vehtari et al., 2015). Resampling can be defined as a procedure in which the particles are re-selected from the particle distribution with probability equal to their weights from the importance sampling (Doucet et al., 2001). Hence, more prominent particles would move to the next filtering sequence. Resampling is performed using the criterion *effective sample size* ( $S_{eff}$ ; Liu (2008a)). The effective sample size could be found using

$$S_{eff} = \frac{1}{\sum_{i=1}^N (w_t^{(i)})^2}. \quad (4.15)$$

Effective sample size is a way of determining the exact samples from the target distribution. It interprets the number of particles which effectively

contributes to the estimation of the state and shows the efficiency of the estimation (Martino et al., 2017). Mathematically, it is a ratio estimate (Kong, 1992) but it can also be explained as a discrepancy measure i.e., the euclidean distance between the probability mass function of the normalized weights to its discrete uniform probability mass function (Martino et al., 2016). A detailed discussion on alternative  $S_{eff}$  criteria are discussed in Martino et al. (2017).

We have summarized the resampling algorithm in algorithm 2.

**Algorithm 2:** Resampling

- 1 Sample from the current state particles  $\{\mathbf{x}_t^{(i)}, i = 1, \dots, N\}$  with probability equal to  $\{w_t^{(i)}, i = 1, \dots, N\}$ .
- 2 Substitute the old particle set with the newly drawn particles.
- 3 Reweight the new particles as  $w_t^{(i)} = 1/N$

The particle filter algorithm forms weighted set of particles from the importance distribution at every time step  $t$ , i.e.,  $\{(w_t^{(i)}, x_t^{(i)}) : i = 1, \dots, N\}$ , which approximates the filtering distribution  $p(x_t | y_{1:t})$ . To derive the algorithm, we consider the full posterior distribution consisting of all previous states and measurements. The recursion goes as

$$p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}) \propto p(\mathbf{y}_t | \mathbf{x}_{0:t}, \mathbf{y}_{1:t-1}) p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t-1}) \quad (4.16a)$$

$$= p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1}) p(\mathbf{x}_{0:t-1} | \mathbf{y}_{1:t-1}) \quad (4.16b)$$

$$= p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{0:t-1} | \mathbf{y}_{1:t-1}). \quad (4.16c)$$

The equations 4.16 use the Markov properties that probability distribution of current measurement only depends on current state and probability distribution of current state depends only on previous state.

Using equations 4.13 and 4.14, we can similarly write here:

$$w_t^{(i)} \propto \frac{p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)} | \mathbf{x}_{t-1}^{(i)}) p(\mathbf{x}_{0:t-1}^{(i)} | \mathbf{y}_{1:t-1})}{\pi(\mathbf{x}_{0:t}^{(i)} | \mathbf{y}_{1:t})} \quad (4.17)$$

The importance distribution can be conveniently split as:

**Algorithm 3:** Particle Filter (Sequential Importance Resampling)

- 1 Draw samples from the prior distribution and set all the weights  $w_0^{(i)} = 1/N$ . Set a threshold  $\mathcal{N}$  for resampling.
$$\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0), \quad i = 1, \dots, N.$$
- 2 **for** each time step  $t = 1, \dots, T$ : **do**
- 3     Sample particles from the importance distribution using previous state particles and all the measurements.
$$\mathbf{x}_t^{(i)} \sim \pi(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t}), \quad i = 1, \dots, N.$$
- 4     Update and normalize the weights using
$$w_t^{(i)} \propto \frac{p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{\pi(\mathbf{x}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} w_{t-1}^{(i)}.$$
- 5     Calculate the effective sample size  $S_{eff}$
$$S_{eff} = \frac{1}{\sum_{i=1}^N (w_t^{(i)})^2}.$$
- 6     Get the state estimate. If the effective sample  $S_{eff}$  is less than threshold  $\mathcal{N}$  perform resampling and set all the weights to  $1/N$ .
- 7 **end**

$$\pi(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}) \propto \pi(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{1:t}) \pi(\mathbf{x}_{0:t-1} | \mathbf{y}_{1:t-1}). \quad (4.18)$$

Using equation 4.18 in the equation 4.17, we get

$$w_t^{(i)} \propto \frac{p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{\pi(\mathbf{x}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} \frac{p(\mathbf{x}_{0:t-1}^{(i)} | \mathbf{y}_{1:t-1})}{\pi(\mathbf{x}_{0:t-1}^{(i)} | \mathbf{y}_{1:t-1})} \quad (4.19a)$$

$$= \frac{p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{\pi(\mathbf{x}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} w_{t-1}^{(i)} \quad (4.19b)$$

This recursive expression leads us to particle filter algorithm summarized in algorithm 3.

### 4.2.3 Unscented Kalman Filters

#### 4.2.3.1 Unscented Transform

The *unscented transform* utilizes deterministically selected *sigma points* for approximating a transformed target random variable (Julier and Uhlmann, 1997). Consider the random variables  $\mathbf{x}$  and  $\mathbf{y}$  defined as

$$\begin{aligned}\mathbf{x} &\sim N(\mathbf{m}, \mathbf{P}) \\ \mathbf{y} &= g(\mathbf{x}).\end{aligned}\tag{4.20}$$

First, we form a set of sigma points which sufficiently captures the random variable  $\mathbf{x}$ . The sigma points are then propagated through the non-linearity via the function  $g(\mathbf{x})$ . These transform sigma points are used to determine the first two moments of the random variable  $\mathbf{y}$ . The basic idea of unscented transform is that the sigma points retain sufficient moment information even through non-linear transformation. For convenience, the transformed variable is approximated as a Gaussian distribution.

We follow the following procedure for forming the approximation:

- Get the  $2n + 1$  sigma points:

$$\begin{aligned}\mathcal{X}^{(0)} &= \mathbf{m}, \\ \mathcal{X}^{(i)} &= \mathbf{m} + \sqrt{n + \lambda}[\sqrt{\mathbf{P}}]_i, \\ \mathcal{X}^{(i+n)} &= \mathbf{m} - \sqrt{n + \lambda}[\sqrt{\mathbf{P}}]_i, \quad i = 1, \dots, n\end{aligned}\tag{4.21}$$

where  $n$  is the dimensions of the state,  $[\cdot]_i$  is the  $i$ th column of the matrix,  $\lambda$  is the scaling parameter which is defined as

$$\lambda \triangleq \alpha^2(n + \kappa) - n.\tag{4.22}$$

$\alpha$  and  $\kappa$  are the user set parameters which dictate the spread of the sigma points around the mean.

- Transform the sigma points using the non-linear function  $g(\cdot)$ :

$$\mathcal{Y}^{(i)} = g(\mathcal{X}^{(i)}), \quad i = 0, \dots, 2n.\tag{4.23}$$

- Compute the mean and the covariance using the transformed sigma points.

$$\begin{aligned} E[g(\mathbf{x})] &\simeq \boldsymbol{\mu}_{ut} = \sum_{i=0}^{2n} W_i^{(m)} \mathcal{Y}^{(i)}, \\ Cov[g(\mathbf{x})] &\simeq \mathbf{S}_{ut} = \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{Y}^{(i)} - \boldsymbol{\mu}_{ut})(\mathcal{Y}^{(i)} - \boldsymbol{\mu}_{ut})^T, \end{aligned} \quad (4.24)$$

where the weights  $W_i^{(m)}$  and  $W_i^{(c)}$  can be computed as:

$$\begin{aligned} W_0^{(m)} &= \frac{\lambda}{n + \lambda}, \\ W_0^{(c)} &= \frac{\lambda}{n + \lambda} + (1 - \alpha^2 + \beta), \\ W_i^{(m)} &= \frac{1}{2(n + \lambda)}, \quad i = 1, \dots, 2n, \\ W_i^{(c)} &= \frac{1}{2(n + \lambda)}, \quad i = 1, \dots, 2n, \end{aligned} \quad (4.25)$$

where  $\beta$  is a parameter which can be used to incorporate additional moment information (skewness, kurtosis, etc).

#### 4.2.3.2 Algorithm

The *unscented Kalman filter* or *sigma point filter* is a non-optimal Bayesian filter which utilizes unscented transform for approximating the filtering distribution. It is a better alternative to *extended Kalman filter* as it is Jacobian and Hessian free.

$$p(\mathbf{x}_t | \mathbf{y}_{1:t}) \simeq N(\mathbf{x}_t | \mathbf{m}_t, \mathbf{P}_t), \quad (4.26)$$

where  $\mathbf{m}_t$  and  $\mathbf{P}_t$  are the estimated mean and covariance. Refer Chapter 5 of Särkkä (2013) for a comprehensive introduction.

**Algorithm 4:** Unscented Kalman Filter

1 **for** each time step  $t = 1, \dots, T$ : **do**

2     *Prediction:*

- Get the sigma points:

$$\begin{aligned}\mathcal{X}_{t-1}^{(0)} &= \mathbf{m}_{t-1}, \\ \mathcal{X}_{t-1}^{(i)} &= \mathbf{m}_{t-1} + \sqrt{n + \lambda} [\sqrt{\mathbf{P}_{t-1}}]_i, \\ \mathcal{X}_{t-1}^{(i+n)} &= \mathbf{m}_{t-1} - \sqrt{n + \lambda} [\sqrt{\mathbf{P}_{t-1}}]_i, \quad i = 1, \dots, n\end{aligned}$$

where  $n$  is state dimension and  $\lambda$  is defined in (4.22).

- Transform the sigma points using the dynamic model:

$$\hat{\mathcal{X}}_t^{(i)} = f(\mathcal{X}_{t-1}^{(i)}), \quad i = 0, \dots, 2n.$$

- Get the predicted mean  $\mathbf{m}_t^-$  and the predicted covariance  $\mathbf{P}_t^-$ :

$$\begin{aligned}\mathbf{m}_t^- &= \sum_{i=0}^{2n} W_i^{(m)} \hat{\mathcal{X}}_t^{(i)}, \\ \mathbf{P}_t^- &= \sum_{i=0}^{2n} W_i^{(c)} (\hat{\mathcal{X}}_t^{(i)} - \mathbf{m}_t^-) (\hat{\mathcal{X}}_t^{(i)} - \mathbf{m}_t^-)^T + \mathbf{Q}_{t-1},\end{aligned}$$

where  $W_i^{(m)}$  and  $W_i^{(c)}$  are the weights defined in the equation (4.25)

3     *Update:*

- Get the sigma points:

$$\begin{aligned}\mathcal{X}_t^{(0)} &= \mathbf{m}_t^-, \\ \mathcal{X}_t^{(i)} &= \mathbf{m}_t^- + \sqrt{n + \lambda} [\sqrt{\mathbf{P}_t^-}]_i, \\ \mathcal{X}_t^{(i+n)} &= \mathbf{m}_t^- - \sqrt{n + \lambda} [\sqrt{\mathbf{P}_t^-}]_i, \quad i = 1, \dots, n\end{aligned}$$

**3**

- Transform the sigma points using the measurement model:

$$\hat{\mathcal{Y}}_t^{(i)} = h(\mathcal{X}_t^{(i)}), \quad i = 0, \dots, 2n.$$

- Get the predicted mean  $\boldsymbol{\mu}_t$ , the predicted covariance  $\mathbf{S}_t$ , and the cross-covariance of the state and the measurement  $\mathbf{C}_t$

$$\begin{aligned}\boldsymbol{\mu}_t &= \sum_{i=0}^{2n} W_i^{(m)} \hat{\mathcal{Y}}_t^{(i)}, \\ \mathbf{S}_t &= \sum_{i=0}^{2n} W_i^{(c)} (\hat{\mathcal{Y}}_t^{(i)} - \boldsymbol{\mu}_t) (\hat{\mathcal{Y}}_t^{(i)} - \boldsymbol{\mu}_t)^T + \mathbf{R}_t, \\ \mathbf{C}_t &= \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{X}_t^{(i)} - \boldsymbol{\mu}_t^-) (\hat{\mathcal{Y}}_t^{(i)} - \boldsymbol{\mu}_t)^T.\end{aligned}$$

**4**

- Evaluate the filter gain  $\mathbf{K}_t$ , the state mean  $\mathbf{m}_t$  and the covariance  $\mathbf{P}_t$  given the current measurement  $\mathbf{y}_k$ :

$$\begin{aligned}\mathbf{K}_t &= \mathbf{C}_t \mathbf{S}_t^{-1}, \\ \mathbf{m}_t &= \mathbf{m}_t^- + \mathbf{K}_t [\mathbf{y}_t - \boldsymbol{\mu}_t], \\ \mathbf{P}_t &= \mathbf{P}_t^- - \mathbf{K}_t \mathbf{S}_t \mathbf{K}_t^T.\end{aligned}$$

**5 end**



## Chapter 5

# Measurement Setup & Data Collection

In this chapter, we discuss about the experimental area, the BLE beacon used in our project. Then we will describe the measurement setup on how the experiments are conducted. Then we will describe in detail the experimental test bed. The android application used for recording the RSSI values.

### 5.1 Experimental Testbed

Our experiment area is an office space located on fourth floor at Helvar office. The area is a vacant space with luminaires fitted with the beacons. The stem part of L-shaped floored space has the dimensions is 32 meters by 5.5 meters and the leg part is 10 meters by 7 meters in dimension, refer the Figures 5.1 and 5.6a. For simplicity and accurate results, the movement of personnel during the experiments were from minimal to zero. The low right corner in the Figure 5.6a is considered the origin and the locations mentioned are with respect to this.

### 5.2 Measurement Configuration

The beacons in the test bed are present inside the luminaires. The luminaires present at 1.37 meters from the floor, so the height of the beacons could be approximated to the same height as the luminaire as the exact location of the beacon can not be estimated inside the luminaire. As shown in the Figure 5.1, there are 28 beacons and the measurements include from all the beacons.

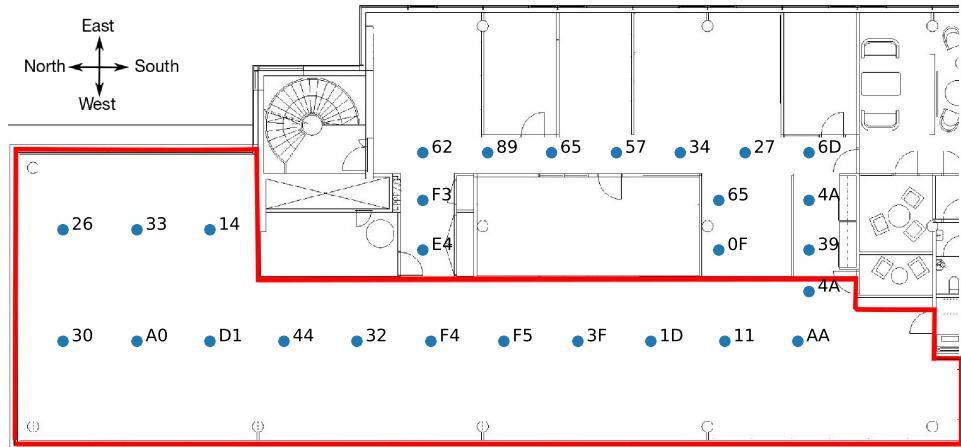


Figure 5.1: Floor plan of Helvar's R&D section in Keilaniemi office with the location of the beacons.

### 5.3 Bluetooth Low Energy

Bluetooth Low Energy (BLE) is a revolutionary short-range, wireless, radio technology operating in the free license 2.4 GHz ISM band. It was developed and announced by Bluetooth Special Interest Group (SIG) on 30th June, 2010 (Gomez et al., 2012). These are coin-cell battery operated devices ranging from 40 mAh to 620 mAh and can last between few months to up to 5 years.

The BLE beacons are the *peripheral* devices capable of connecting to *master* or *central* device for carrying out a specific task. The indoor positioning application is possible as a single peripheral can advertise to multiple mobile devices.

In terms of positioning techniques, any RSS based method reads the beacon packets by the *Wireless Access Points (WAPs)* at regular intervals.

#### 5.3.1 Physical Layer

Bluetooth Low Energy has 40 physical radio channels with each radio channel spaced out of 2 MHz in between them (refer Figure 5.2). The BLE enjoys the data rate of 1 Mbit/sec and like the *classic Bluetooth*, uses the *Gaussian Frequency Shift Keying* modulation. But both the technologies have different spacing of radio channels, hence are incompatible and therefore, can't communicate.

The physical channels are categorized based on type of data they transmit and they described next.

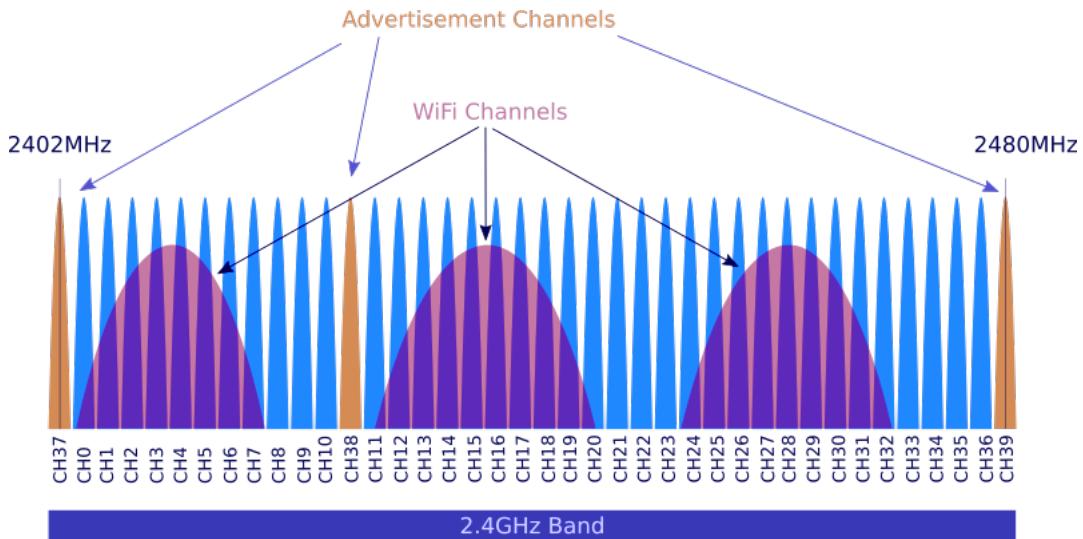


Figure 5.2: Channels in Bluetooth low energy.

### 5.3.2 Advertisement channels

The channels 37, 38 and 39 are the advertisement channels. These radio channels are strategically placed to avoid interference from the WiFi (see figure 5.3). The advertisement channels are critical as they are responsible for making a connection to a mobile devices (like smartphones, smartwatches, etc.) and three channels are used to increase the probability of central device reading the advertisement packets. This mode serves for uni-directional communication. The BLE technology also gives the option of masking any channel.

During the advertisement interval, the three advertisement channels transmit the packets sequentially in under 1 ms then followed by sleep period. The sleep period consists of fixed interval and a pseudo-random delay. Based on the target application, the fixed interval can be varied from 20 ms to 10.24 seconds and the pseudo-random delay from 0 ms to 10 ms. In case of overlap of advertisement interval and central devices' *scanning interval*, the pseudo-random delay helps avoiding central device missing the packets. In this phase, as most of the time, the device is in sleep mode and transmission power in between -20 dBm to +10 dBm, guarantees low power consumption. For more technical details refer Lindh (2015).

BLE beacons which are used to advertise and communicate the data. The advertisement and data transmission window is 1 sec. The beacons advertise at the start of the second which is approximately 1 ms.

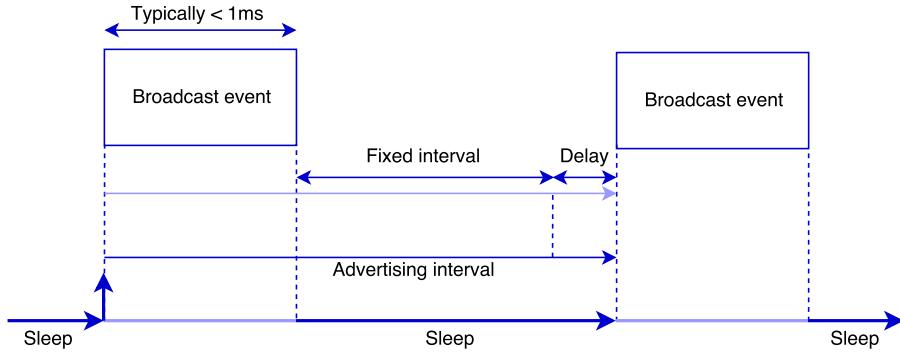


Figure 5.3: Advertisement in BLE

### 5.3.3 Data channels

The rest of the radio channels i.e., channel 0 to channel 36, are dedicated data channels. These channels are used once a device is discovered and a connection with a master device is established. This is a bi-directional mode as shown in Figure 5.4 where we can see two-way communication between master and slave. For positioning application, the connection between device's is rarely seen and hence, this mode is hardly used.

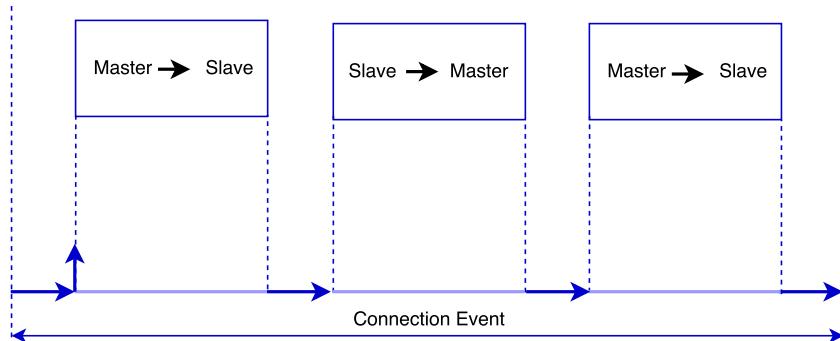


Figure 5.4: Data communication in BLE

### 5.3.4 Different BLE protocols

Largely, the BLE beacons are configured into two types of protocols.

- **iBeacon:** The iBeacon (Newman, 2014) is communication protocol for BLE technology developed by Apple Inc. in 2013. This protocol supports both iOS (over version 7) and Android (over Jelly Beans 4.3) devices with a minimum requirement of Bluetooth 4.0.

- **Eddystone:** The Eddystone (Eddystone, 2016) is a free BLE communication software from Google Inc. announced in 2015. It is also compatible with both iOS and Android devices that have Bluetooth 4.0 and above.

## 5.4 Assumptions on the RSSI values

RSSI is designed for wireless communication and designing receiver antenna but not for positioning applications. The RSSI values are quantized i.e., they are estimated in the steps of 1 dBm, so in theory the quantized value of RSSI represent an area rather than a point Kaemarungsi and Krishnamurthy (2012). We assume that in-luminaire BLE beacons are omni-directional with the luminaires always being located over the head of the user. The variation of RSSI due to change in the location of the beacon inside the luminaires was assumed negligible.

We study the caveats of RSSI for positioning thoroughly in Chapter 6, with the luminaires always being located over the head of the user. The variation of RSSI due to change in the location of the beacon inside the luminaires was assumed negligible. The interference of in-luminaire working and other radio signal is consider negligible. Also, for the data analysis, we consider the data distribution to be continuous. The bias due to direction of the phones when taking user-free measurements was also assumed negligible.

## 5.5 Radio Analyzer

The Frontline Sodera LE !FIXME **cite it** !FIXME! radio-analyzer was used for reading the RSSI values. The radio-analyzer accurately reads the signal strength values and gives additional data like channel information, radio channel information !FIXME **correct it** !FIXME!, universally unique identifier (UUID) and data type. A comprehensive discussion on the various ways to measure RSSI values is presented in Kaemarungsi and Krishnamurthy (2012).

## 5.6 Measurement Application

The RSSI measurement application was built using the AltBeacon Android library (Networks, 2017). This library incorporates extra features like foreground and background interval. The in-luminaire beacons we used didn't follow either the iBeacon (Newman, 2014) or Eddystone (Eddystone, 2016)

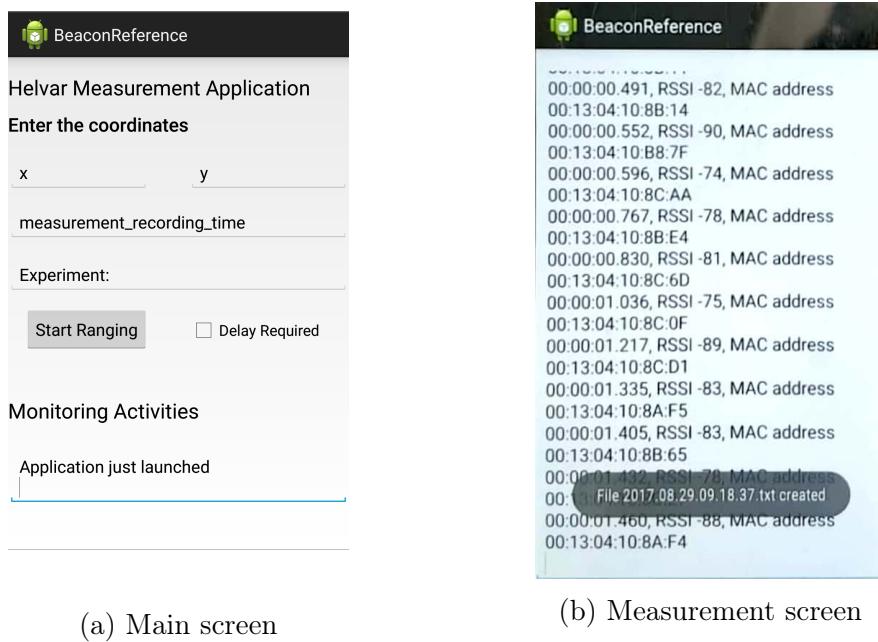


Figure 5.5: Caption place holder

protocol and this library made it convenient to read the radio signals from non-standard advertising beacons.

The application allows for adding the location i.e., x and y coordinates, measurement recording time, and the experiment description. It also has the option to delay recording the measurements in the case of without user measurements. The application buzzes after the completion of measurement time. The application displays relative time from start, absolute time-stamp, the beacon's media access control (MAC) address and RSSI value. The same data is logged in a text file with the absolute time-stamp when the file was opened as the file name.

### 5.6.1 Smartphones

All the measurements were taken with a smartphone unlike the other studies where the experiment were conducted using a laptop (Kaemarungsi and Krishnamurthy (2004); Bahl and Padmanabhan (2000)). The various smartphones used during the project were Samsung S7, Samsung S4, Samsung S4 mini. Due to instability of the measurement application in reading the RSSI the Sony and LG devices were abstained from using. The reason is beyond the scope of this thesis.

## 5.7 Data Collection

The data collection is categorized in two phases. The first phase is for data analysis (in Chapter 6) and second for the position algorithms (in Chapter 7). The measurements for the various analyses mentioned next were collected for 50 seconds.

### 5.7.1 For Data Analysis

For better understanding the data generating process we collect the data from different smartphone and human orientation to better construct the measurement model in the Bayesian filtering context. The different user orientations will be discussed in the following section. We selected two location for conducting the following experiments location 1 and 2, which we hereinafter refer to as L1 and L2 respectively.

#### 5.7.1.1 Different User Orientations

The main idea here is to cover different aspects of the the user directions while recording the measurements from the WAP's. In this thesis, we recorded the user measurements in the directions north, east, west and south (for directions refer Figure 5.1). For the user measurements, the user the held the smartphone device at a 45°angle, with head parallel to the screen of the phone (Hansraj, 2014). We refer this as *standard usage configuration*. By observation, this most common way of usage when the user is walking. With the same configuration, we recorded the shadow and rotate measurements.

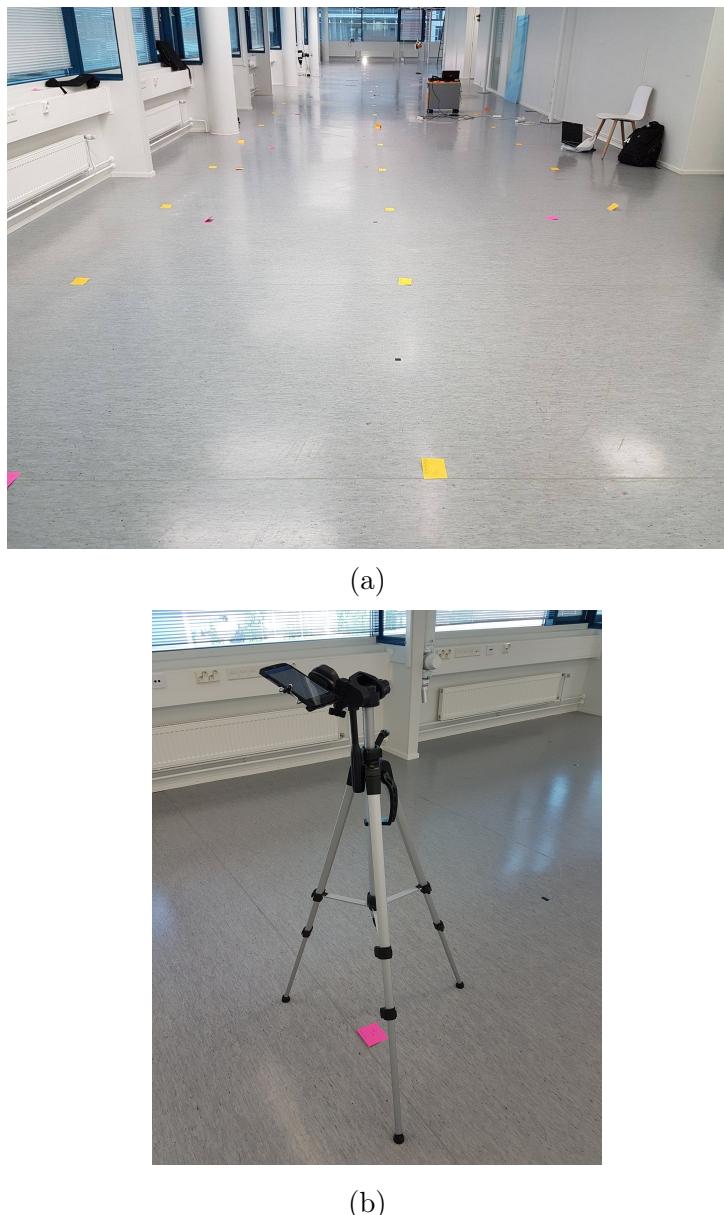


Figure 5.6: (a) measurement arena (b) Tripod with mobile holder for user free measurements at one of the measurement landmarks.

In the shadow mode, the user stands in between the signal generating beacon and the smartphone shadowing the measurements cutting the line of sight to the beacon. In the rotation mode (Honkavirta, 2008), the user rotates while recording the measurements.

### 5.7.1.2 Different Phone Orientations

In addition to the different user orientations, we also experimented with different phone orientation, as enumerated in Hansraj (2014). The different configuration of phone angles used were  $0^\circ$ ,  $45^\circ$  and  $90^\circ$ . The signals strengths recorded in this configurations were user free. For the user free measurements, the smartphone device was clamped and mounted on the tripod as show in Figure 5.6b. In user mode, we took measurements with the smartphone in the side pocket of the trousers for juxtaposing against the standard usage configuration.

### 5.7.1.3 Outdoor Measurements

To avoid the possible attenuation of signal due to reflection from walls indoors, we recorded measurements in a outdoor setting. At a fairly empty car parking area behind Jämeräntaival 1, Espoo, we arranged the setup as shown the Figure 5.7. We mounted a luminaire with beacon on a tripod at a height of 0.6 meter and connected it to a power source. At the same height and 1 meter away, we placed the radio-analyzer (refer Section 5.5) on thermocol boxes and tripod clamped smartphone. The radio analyzer was powered up by connecting it to a power source and was later connected to a laptop for running radio-analyzer's measurement application. The radio-analyzer's and smartphone's respective applications were set to record measurements simultaneously. In case of mismatch, the absolute time-stamps were used to get the accurate time measurements for further data analysis.

FIXME update this section... FIXME!

## 5.7.2 For Positional Algorithms

The data collection for the position algorithms is done in two phases. The first phase is for collection of the fingerprinting calibration data and second phase is the collection of the test data.

### 5.7.2.1 Setup for Calibration Phase and Test Phase

The calibration phase involves the fingerprinting process. In this process, we identify predestined locations in the experiment test bed and collect measurements. As shown in the Figure 5.6a, we landmarked the locations with respect to the origin as described in Section 5.1. We selected a total of 63 calibration points and recorded measurements for 50 seconds.

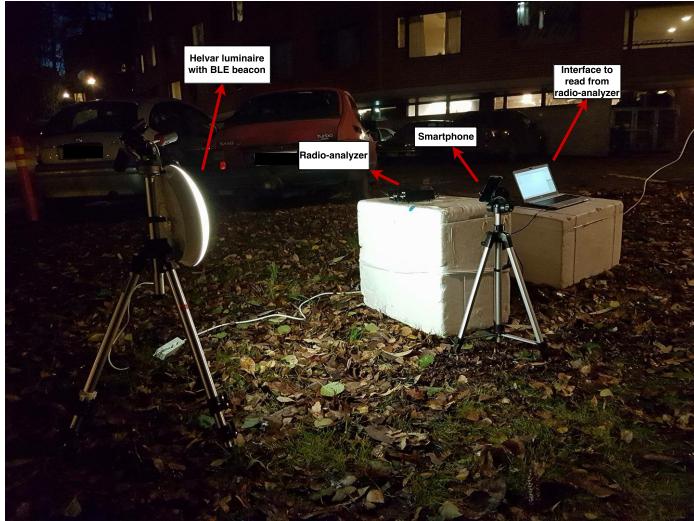


Figure 5.7: Measuring the RSSI values outdoors using the smartphone and radio analyzer.

The test phase we landmarks the exact locations on the floor and using the application the track Figure 5.8 was followed. The measurement application was used for time flagging whenever the landmarks were reached.

### 5.7.2.2 Calibration Data

The calibration data is used for generating the reference maps which in turn is used in creating the radiomaps. The radiomaps are used as the measurement model in the filtering process as described in the Chapter 4. For collecting the data the user measures the signal strength at a particular calibration point at a random direction. This was done keeping in mind the final product, where the user would not be forced to choose a direction but rather take measurements as is.

### 5.7.2.3 Test data

Obtaining the data was done using the mobile device walking at a constant speed in the test setup. The measurements were collected using the smartphone device Samsung S7. The data was then ported to MATLAB where it was converted to appropriate data structure for to be used evaluating the filtering methods. The test device Samsung S7 was the same device as was used for fingerprinting and generating the radiomaps. The test data collected can seen in the Figure 5.8.

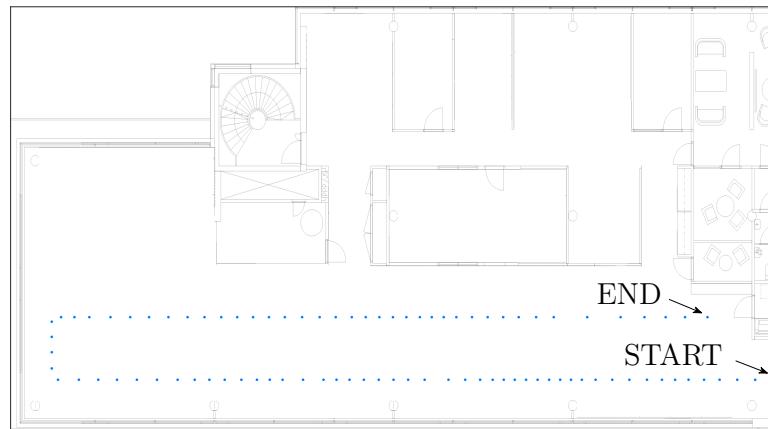


Figure 5.8: Test data from Helvar R&D



# Chapter 6

## Data Analysis of RSSI

*You attract the right things when you have a sense of who you are.*

- Amy Poehler

Before we start exploring the different methods for solving the indoor positioning problem, the initial challenge lies in getting the right measurement (data) model as the RSSI varies due to various factors like signal attenuation due to obstacles, human beings, signal interference, type of smartphone hardware, orientation of the phone and algorithm related factors and other factors like malfunctioning BLE modules. The challenges have been discussed in detail in Chapter 1.

The data analysis of RSSI is vital to the understanding and formulating the location dependent features in fingerprinting methodology and hence designing an accurate indoor positioning algorithm (Kaemarungsi and Krishnamurthy, 2004). A thorough knowledge about the BLE signal's data generating process can help in understanding the signal variations in different locations of the space and effect of obstacles and walls. This understanding can be learnt via studying the statistical properties of the RSSI values. This knowledge about data generating process and RSSI's can help in efficient modeling of the measurement model which could lead to better indoor positioning system.

The *initial data analysis*<sup>1</sup> is the primary task in statistical analysis and modeling which yields critical statistical properties about data generating process. It veers us towards finding the right solution i.e., finding the right data model for our problem and learn the peculiarities in the data. It helps to efficiently design and analyze an indoor positioning system. A comprehensive data analysis of RSSI data from the beacons is hardly done (for WiFi, check

---

<sup>1</sup>not to be confused with exploratory data analysis. Check Chatfield (2006) for more details.

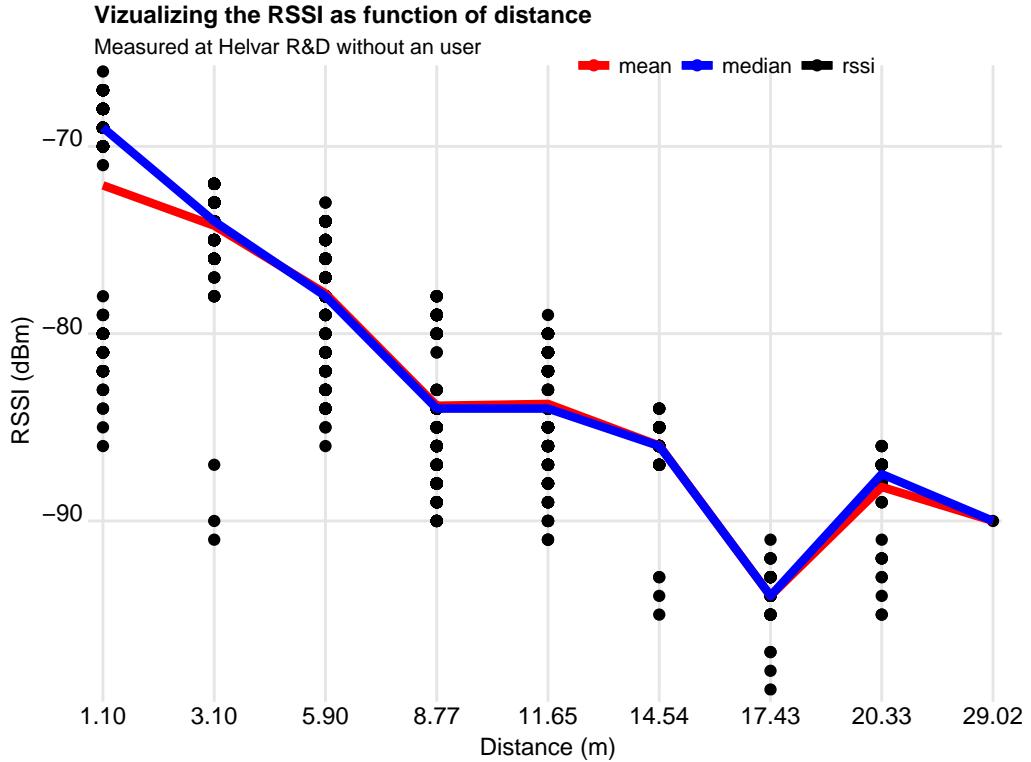


Figure 6.1: Illustration of received signal strength indication value decreases with increase in the distance of the luminaire.

Kaemarungsi and Krishnamurthy (2012)) in previous literature with most of the researchers focusing on the algorithms. Based on our review, there was no in depth analysis of statistical properties of RSSI values of the BLE beacons, so in this thesis, we take closer look at the RSSI and investigate different factors which affect its variation. Due to paucity of time, RSS analysis based on time is not studied, and it was taken care that all the measurements were taken in close successions in time. We take into consideration the factors such as user's presence, smart-phones and orientation of smart-phones and material of the luminaire. First, in 6.1, we look at the raw RSSI data for it to be suitable as distance criterion. Next, in Section 6.2, we explain the different hypothesis testing used. In Section 6.2.5, we explain the results of different experiments conducted to understand the factor biasing the RSSI values. Last, in Section 6.3, we go deeper in understanding the RSSI with the perspective of its architecture.

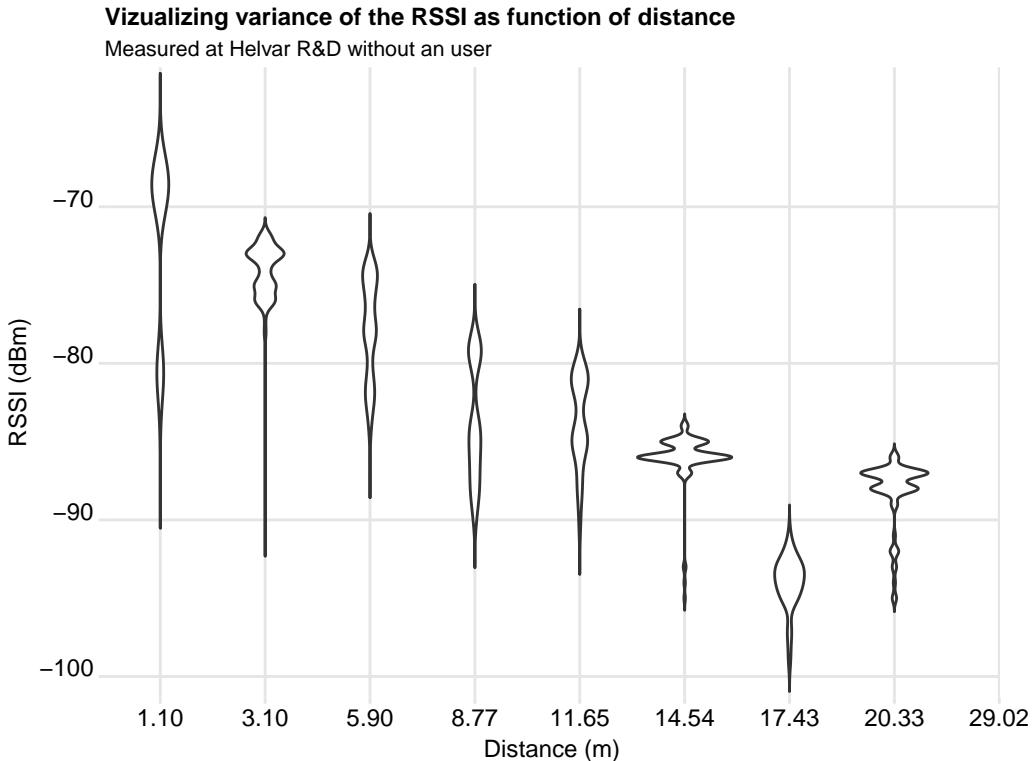


Figure 6.2: The figure shows the variation of signal strength as distance increases.

## 6.1 RSSI as a measure of distance

The vital proposition to the thesis is that RSSI is a valid measure convenient for solving the indoor positioning problem. The use of RSS indication was proposed in the pioneering work by Bahl and Padmanabhan (2000) where the wireless local area network (WLAN) was used. Similar to Bahl and Padmanabhan (2000), in Figure 6.1 we show that the RSSI is a realistic criteria with the measure inversely correlating with the distance. The measurements were recorded with the Samsung S7 device.

The experiment included the tripod clamped mobile unit which recorded the signal strength for a particular period of time. Evidently, conforming to the common knowledge that the signal strength at the beacon closest to the mobile unit has the highest signal strength and reduces as the beacon is farther away.

Additionally we can interestingly observe from Figure 6.2, that how the variance of the RSSI values decreases as the distance increases. We discuss

more about this in the Section 6.3

## 6.2 Statistical Hypothesis Testing

For the empirical (sample) data collected from different experiments (refer 6.2.5) we need to draw conclusions in order to understand the data generating process. This allows us to hypothesize about the data population and utilize them to model our problem. Hence, hypothesis testing is also called *confirmatory data analysis*.

The hypothesis testing procedure goes as follows:

1. A *null hypothesis*  $H_0$  and an *alternative hypothesis*  $H_1$  are formulated.
2. Get the sample statistics in view of the hypothesis considered.
3. Select another sample from the population for one sample test.
4. Determine the test statistic and infer from the results.

Generally, the null hypothesis states the status quo and the alternative hypothesis states otherwise. For example, null hypothesis might state that two sample data have equal means and the alternative hypothesis put forth's otherwise. The *region of acceptance* defined by the *critical values* are used as evidence to accept or reject the null hypothesis and the *p-values* denote the unusuality of the computed test statistic, and signify our confidence on our decision. If the computed test statistic falls in the region of acceptance, which is a range of sample statistics, then the null hypothesis is not rejected. The critical values are the threshold from region of acceptance to region of rejection, also sometimes referred to as *critical region*. In this thesis, we investigate the statistical significance for the parameters sample mean, variance and median from the different samples recorded using the smartphone. The different tests used are described below.

### 6.2.1 Two Sample Z-test

The *two sample z-test* is a statistical test to determine the relationship between two sample means given that variance of the two distributions is unknown (Daniel, 1999). The different possibilities of formulating the null hypothesis  $H_0$  and alternative hypothesis  $H_1$  are

1.  $H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2,$

2.  $H_0 : \mu_1 \geq \mu_2$        $H_1 : \mu_1 < \mu_2,$
3.  $H_0 : \mu_1 \leq \mu_2$        $H_1 : \mu_1 > \mu_2.$

As the data is greater than 30, we use the Z test in opposed to t test and *central limit theorem* allows us to use the Equation 6.1 and adds robustness. We assume that the samples are drawn from normal distribution with equal but unknown variances and null hypothesis being that the means are equal. Consider two samples with sample means  $\bar{x}_1$  and  $\bar{x}_2$ , the test statistic  $z$  is given by

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (6.1)$$

where the  $\mu_1, \mu_2$  with subscript zero are hypothesized parameters,  $\sigma_1^2, \sigma_2^2$  are known variances, if unknown sample variance are calculated and  $n_1, n_2$  are number of data points in each corresponding samples. With the *level of significance*  $\alpha$  as 0.05, and assumed variance of 10 and equal known means (i.e.,  $\mu_1 = \mu_2$ ) for the samples, the region of acceptance lies in between  $\pm 1.980$ .

### 6.2.2 Two Sample Kolmogorov-Smirnov test

The two-sample *Kolmogorov-Smirnov* (KS) test is non-parametric hypothesis test for the equality of two unknown but independent continuous probability distributions (Daniel, 1999). Consider two probability distributions  $\mathcal{P}_1(x)$  and  $\mathcal{P}_2(x)$ , the null and alternate hypothesis are defined as

- $$\begin{aligned} H_0 : \mathcal{F}_1(x) &= \mathcal{F}_2(x) & \forall x \in \{-\infty, +\infty\} \\ H_1 : \mathcal{F}_1(x) &\neq \mathcal{F}_2(x) & \exists x \end{aligned}$$

where  $\mathcal{F}_1(x), \mathcal{F}_2(x)$  are the corresponding *cumulative distribution function*. The KS test reports the maximum discrepancy over the range of the random variable  $x$  between the distributions. It uses the test statistic

$$D = \sup_x |\mathcal{F}_1(x) - \mathcal{F}_2(x)| \quad (6.2)$$

where *sup* is the supremum function which gives maximum of all values.  
!FIXME add to abbreviations FIXME!

If the computed test statistic  $D$  exceeds the pre-computed values (check Appendix table M of Daniel (1999)) for a particular  $\alpha$  *level of significance* we reject the null hypothesis.

The assumptions in using the test are that the distributions sampled be random and distribution be continuous. Even though the RSSI values are quantized (discretized), we assume them to be continuous.

### 6.2.3 Levene's test

Levene's test is a hypothesis test to evaluate the equality of variance of any number distributions with an assumption that the random variables don't follow normality. For any random variable of sample size  $N$  with  $k$  subgroups with  $i$ -th group having  $N_i$  sample size. In this test, we define null and alternate hypothesis as

$$H_0 : \sigma_1^2 = \sigma_2^2 \dots = \sigma_n^2$$

$$H_1 : \sigma_i^2 \neq \sigma_j^2, \quad \exists(i, j).$$

The test statistic is defined as

$$W = \frac{(N - k)}{(k - 1)} \frac{\sum_{i=1}^k N_i (\bar{Z}_{i\cdot} - \bar{Z}_{..})^2}{\sum_{i=1}^k \sum_{j=1}^{N_i} (\bar{Z}_{ij} - \bar{Z}_{i\cdot})^2} \quad (6.3)$$

where  $Z_{ij}$  can have either of the following three definition:

1.  $Z_{ij} = |Y_{ij} - \bar{Y}_{i\cdot}|$ , where  $\bar{Y}_{i\cdot}$  is the mean of the  $i$ -th subgroup.
2.  $Z_{ij} = |Y_{ij} - \tilde{Y}_{i\cdot}|$ , where  $\tilde{Y}_{i\cdot}$  is the median of the  $i$ -th subgroup.
3.  $Z_{ij} = |Y_{ij} - \bar{Y}'_{i\cdot}|$ , where  $\bar{Y}'_{i\cdot}$  is the 10% mean of the  $i$ -th subgroup.

Here,  $Z_{i\cdot}$  are the group means of the  $Z_{ij}$  and  $\bar{Z}_{..}$  is the overall mean of  $Z_{ij}$ .

The aforementioned options give the robustness and power to the test. Robustness means not wrongly detect the equality of variance and power means to find equality when one exist in the non-normal random variable. Given a level of significance  $\alpha$ , the Levene's test defines the *critical region* to reject the null hypothesis if

$$W > F_{\alpha, k-1, N-k}$$

where  $F_{\alpha, k-1, N-k}$  is the *upper critical value* of the F-distribution with  $k - 1$  and  $N - k$  degrees of freedom at  $\alpha$  significance level.

One way to infer about the variances of two distribution is to use its ratio estimate i.e.,  $\sigma_1^2/\sigma_2^2$  (Daniel, 1999). The *F distribution*, given its definition, conveniently allows us to conclude if the two distributions have same variance or not. Given two independent samples, assumed to be drawn from different normal distributions, then random variable  $X$ ,

$$X = \frac{(s_1^2/\sigma_1^2)}{(s_2^2/\sigma_2^2)} \implies X \sim F(d_1, d_2) \quad (6.4)$$

follows the F distribution. Here,  $d_1$  and  $d_2$  are degrees of freedom of the samples and are respectively used to evaluate the sample variances  $s_1^2$  and  $s_2^2$ . The degrees of freedom  $d_1$  and  $d_2$  are traditionally referred to as the *numerator degrees of freedom* and the *denominator degrees of freedom*. These are computed using the respective number of sample data points as  $n_1 - 1$  and  $n_2 - 1$ . It is easier to infer if  $s_1^2$  represents the larger of the two sample variances.

To find the confidence interval for the interval estimate  $\sigma_1^2/\sigma_2^2$  given a significance level  $\alpha$ , we use the following expression

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} < \frac{F_{(\alpha/2)}}{F_{(1-\alpha/2)}} < \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \quad (6.5)$$

which can be re-written as

$$\frac{s_1^2/s_2^2}{F_{(1-\alpha/2)}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2/s_2^2}{F_{(\alpha/2)}} \quad (6.6)$$

where calculating the values of  $F_{(\alpha/2)}$  and  $F_{(1-\alpha/2)}$  requires F distribution tables (refer Daniel (1999), Appendix G) and generally significance level of 5 percent is used. If the confidence interval includes 1, then it is likely that the two distributions have equal variance.

#### 6.2.4 Wilcoxon rank-sum test

The *Wilcoxon signed-rank test* for median is a non-parametric statistical test when either the  $t$  or  $z$  statistic is not applicable due to non-normality of the random variable leading to non-applicability of the central limit theorem (Daniel, 1999). The two main assumption of the test are that the random variable  $X$  is continuous and the probability density function is symmetric.

The null hypothesis  $H_0$  states that the two medians scores  $M_1, M_2$  from the sampled data are same while the alternative hypothesis  $H_1$  states otherwise. The level of significance  $\alpha$  considered is 0.05.

$$\begin{aligned} H_0 : M_1 &= M_2 \\ H_1 : M_1 &\neq M_2 \end{aligned} \quad (6.7)$$

The assumptions in the test given the ordinal, unequal sample data size are

Table 6.1: A  $2 \times 2$  Contingency table

Second criterion	First criterion			Total
	data 1	data 2		
#data above common median	a	b	a + b	
#data below common median	c	d	c + d	
Total	a + c	b + d	n	

1. that they are independently and randomly selected from their respective populations,
2. their functional form of population is similar but vary only location-wise and
3. the parameter being inferred is continuous.

An estimate of chi square statistic  $X^2$  is computed as the test statistic using the equation

$$X^2 = \frac{n(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)} \quad (6.8)$$

where a, b, c, d and n are defined in the Table 6.1. It is important to note that in the construction of table a *common median* needs to be computed for segregating the data. It is accomplished by getting the median of the combined data.

For the null hypothesis to be true  $X^2$  needs to be approximately  $\chi^2$  distributed with degree of freedom 1. It implies that computed  $X^2$  should be less than 3.841 given that  $\alpha$  is defined as 0.005.

### 6.2.5 Experiments

The main aim of these experiments was to find the bias of different factors, hence, it was taken care to change only single factor keeping the other factors unchanged. Measurements were taken with one phone at a time to avoid any kind of unknown interference. We accomplish the statistical hypothesis testing methods discussed in section 6.2.

#### 6.2.5.1 Bias due to user's presence

In this experiment, we investigated the effect of user's presence on the recorded RSSI values. The smartphone Samsung S7 phone was used for the experiments. It was assumed that when collecting the data without the user,

the direction of the phone had minimal bias on the RSSI value. Hence, a random direction was chosen and this direction was consistent over all the experiments with different smartphones. In addition, we also checked for the *shadowing effect* and *rotation effect* (Honkavirta, 2008). The shadowing effect deals with the bias when the user completely blocks the signals from a certain beacon. The rotation effect was to check for decreased bias in smartphones from user presence and shadowing. Here, the measurements were collected while the user rotated at the calibration point. From the location of measurement, we selected seven beacons with increasing distance to observe the effects.

Overall for locations L1 and L2, the null hypothesis was not rejected for most of the experiments including shadowing and rotating experiments. We can say the bias due to user presence can have maximum mean error of 11.646 dBm. We can see the bias in the Figures 6.3 and 6.4.

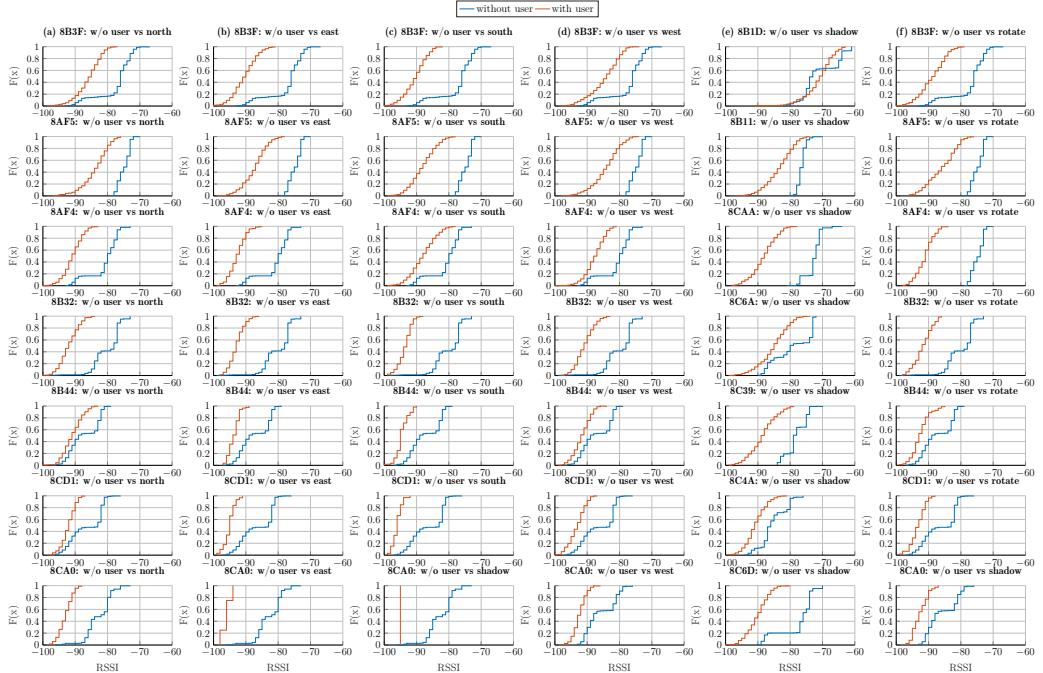


Figure 6.3: Experiment 1: Helvar user presence

On a granular level the observations are:

- **Directions:** One observation for L2, though unusual, is that it consistently records higher RSSI counts when compared to its counterpart user-free experiment. It is hypothesized that as L2 is a corner location, the signal reflection might be source of such aberration.

- **Shadowing:** As mentioned above, the shadowing experiment is performed to quantify the decrease in signal strength due to presence of human body in between mobile device and BLE beacon. It can be seen from the Tables A.2 and A.3
- **Rotating:** As mentioned above, the rotation experiment was conducted to see whether this could reduce the bias due to human presence and perform better when compared shadowed RSSI measurement. As per the results in the Tables A.2(6) and A.3(6), the numbers suggest that rotation doesn't significantly improve the RSSI measurement in terms of signal strength or signal measurement count. This can be observed from the Figures 6.3(f) and 6.4(f) !FIXME put it in numbers the error. FIXME!

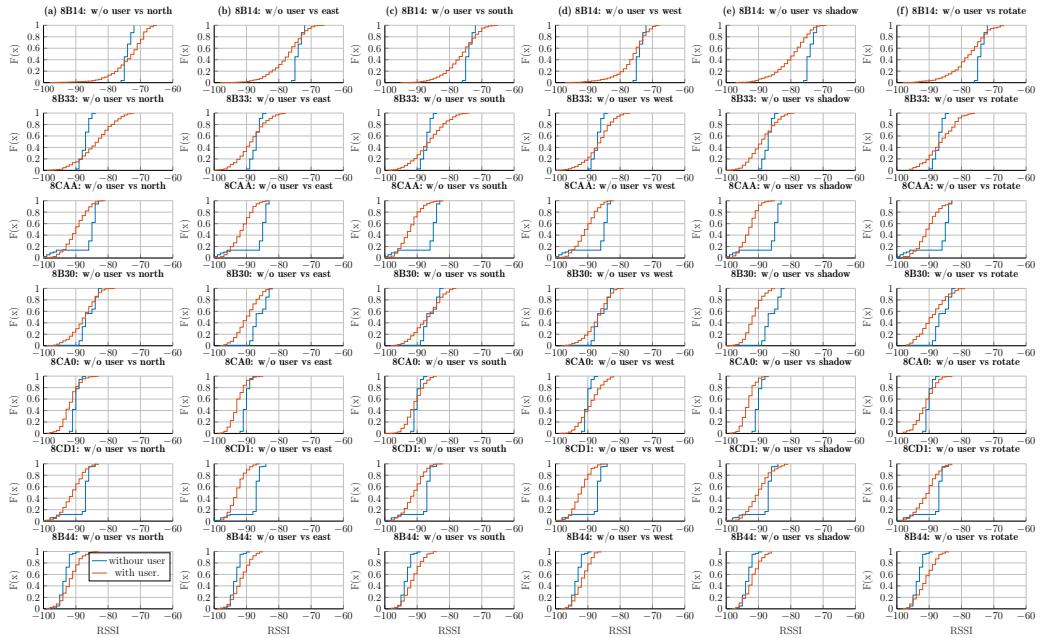


Figure 6.4: Helvar User presence experiment 2.

### 6.2.5.2 Comparisons of Smartphone

In this experiment, we want to quantify the difference in the hardware of the smartphones and its effect on the RSSI values. This is an user free experiment where the recorded measurements over the smartphones using the standard usage configuration. A random direction in relation to the room was chosen and the it was kept consistent over the different smartphone measurements.

Overall for the locations L1 and L2, the null hypothesis was not rejected for the experiments S4 vs S4 mini and S4 vs Nexus 5 but from the Tables A.4 and A.5 we can infer that this can cause a maximum bias of 4.63 dBm. The similarity of the distributions can be seen from Figures 6.5 and 6.6

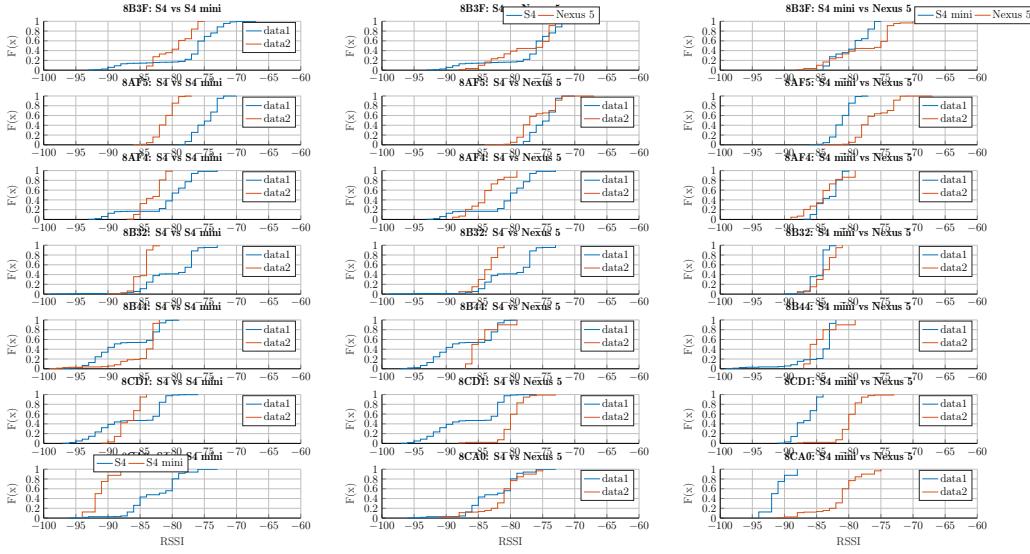


Figure 6.5: Smartphone helvar location 1

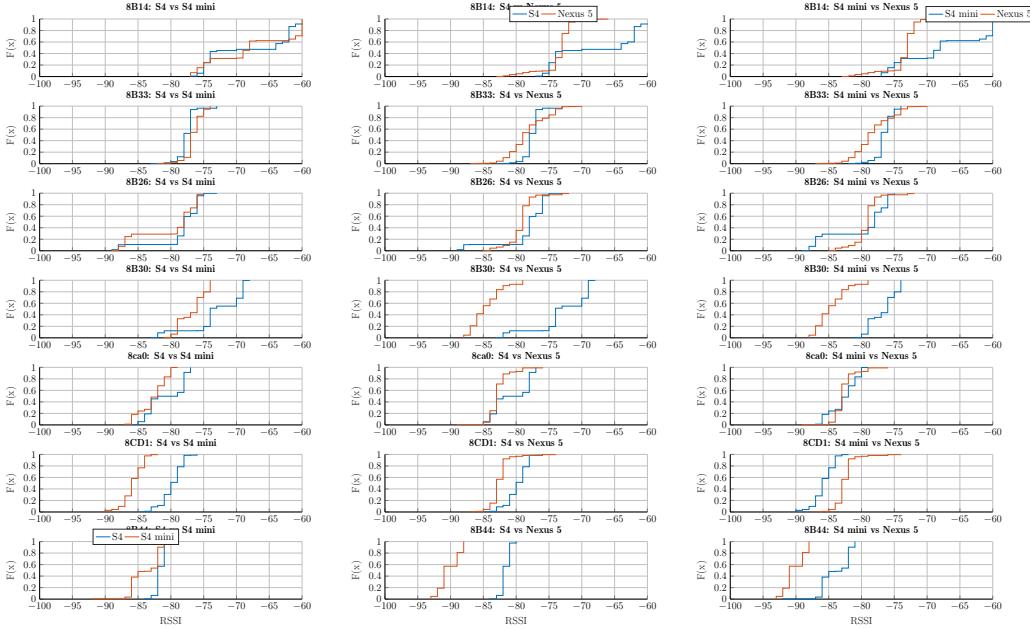


Figure 6.6: Smartphone helvar location 2.

### 6.2.5.3 Orientation of Smart phone

We recorded the RSSI values for the smartphones over different orientation of the phones. We chose different orientations like for example,  $0^\circ$ ,  $45^\circ$  and  $90^\circ$  using the S4 mini smartphone. From the Figures 6.7 and 6.8 we don't see significant difference in the distributions of RSSI values. The Tables A.7 and A.7 suggest that user free orientation bias not more than 3.2 dBm and in terms of *hand vs pocket* the bias is not more than 2.2 dBm.

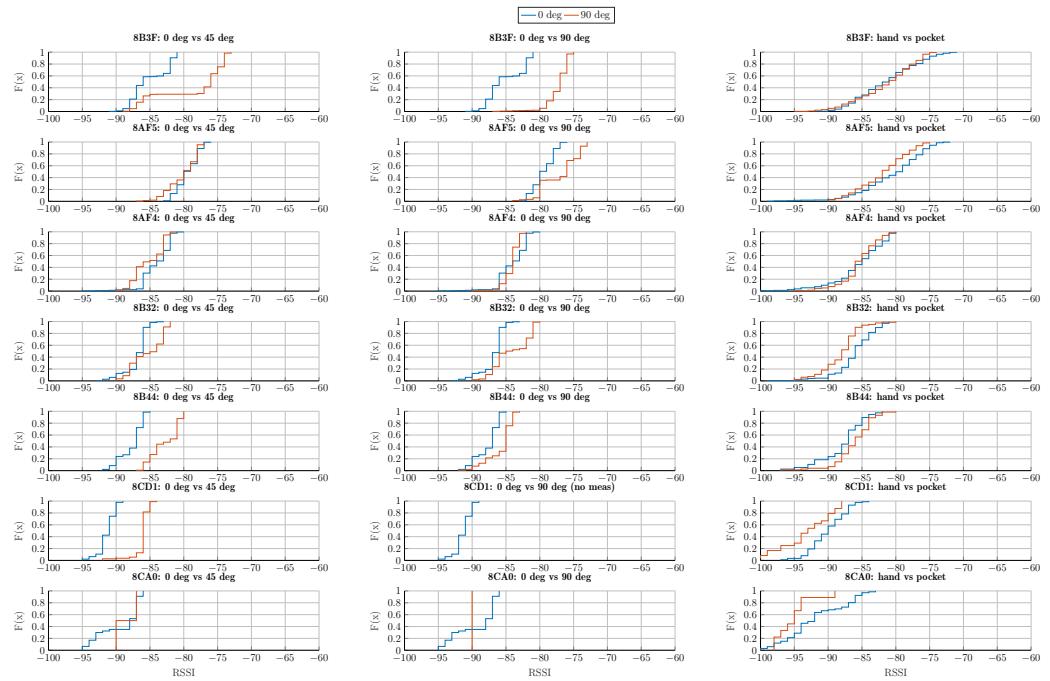


Figure 6.7: Orientation Helvar location 1

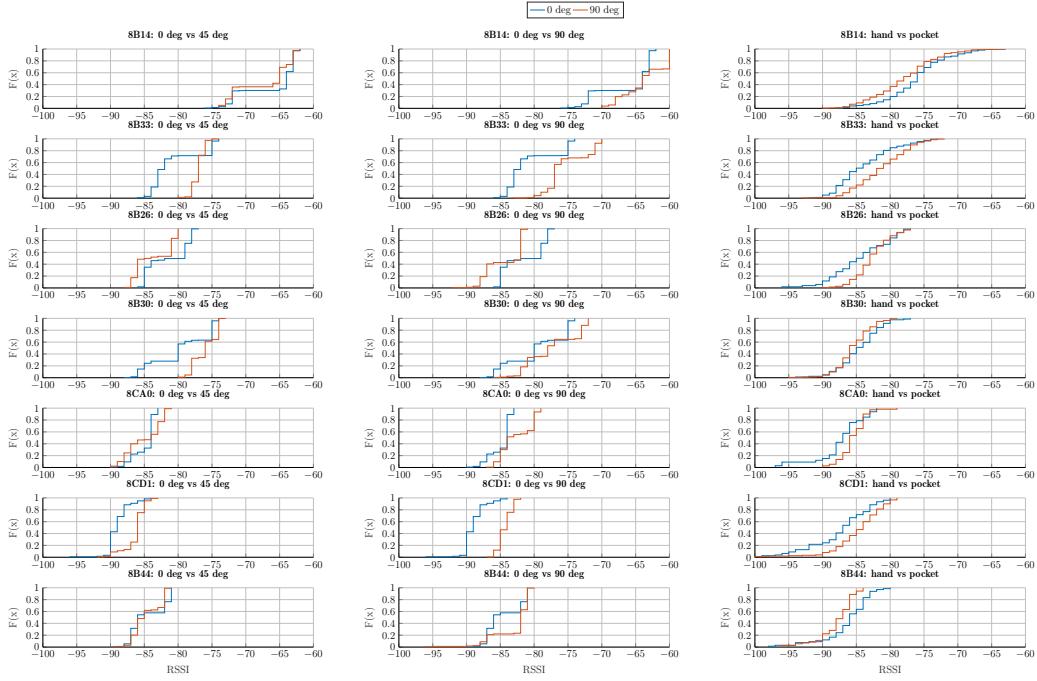


Figure 6.8: Helvar orientation location 2

#### 6.2.5.4 Material of the Luminaires

In this experiment, we investigate if the material of the luminaire adds any bias to the signal strength values from the BLE access point. In order to eliminate any other factor affecting the measurements, a single smartphone namely Samsung S4 mini was chosen and measurements were recorded without any user. The phone was present at 90 degrees angle at both the locations right below the access point at a height of 2.898 meters for *metal-plastic* luminaire and 2.707 meters for plastic luminaire. It was assumed that the orientation of the Bluetooth access point or mobile unit have minimal effect on the signal strength values.

Compared the RSSI values from different luminaires types at different distances.

We conducted two sets of experiments, with one luminaire made of plastic and metal, and other one made only of plastic with the access point inside the ceiling. The results from the Table A.8 shows that till 5.5 meters the metal cum plastic luminaire add not more than 10 dBm bias when compared against the plastic luminaires. The bias is quite evident in the Figure 6.9 in experiment 1: A0 vs 89 and D1 vs 65 plots. The KS test and median test confirm for both the experiments that the underlying distribution is different but the t-test and f-test infer conflicting results.

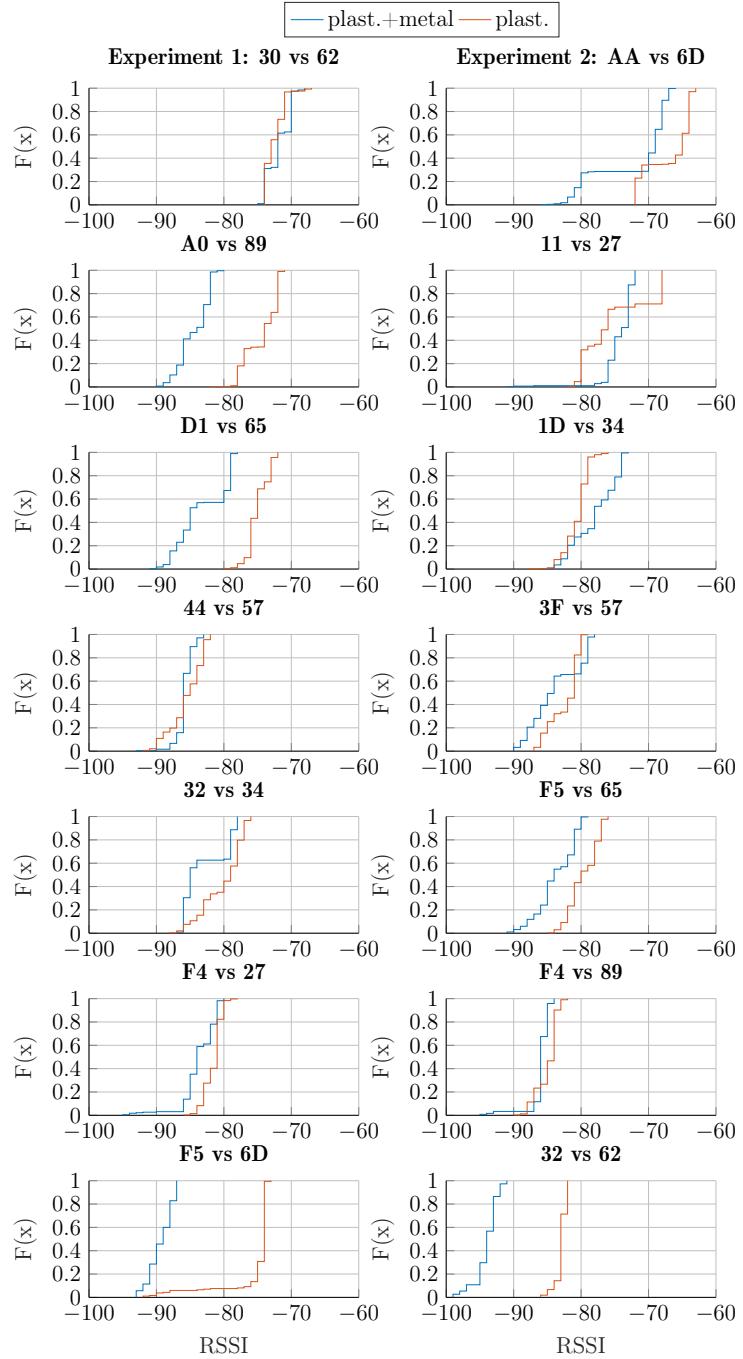


Figure 6.9: Comparing the RSSI values from luminaires with different material at different distances (check the table 6.3). !FIXME **add distance on x axis** FIXME!

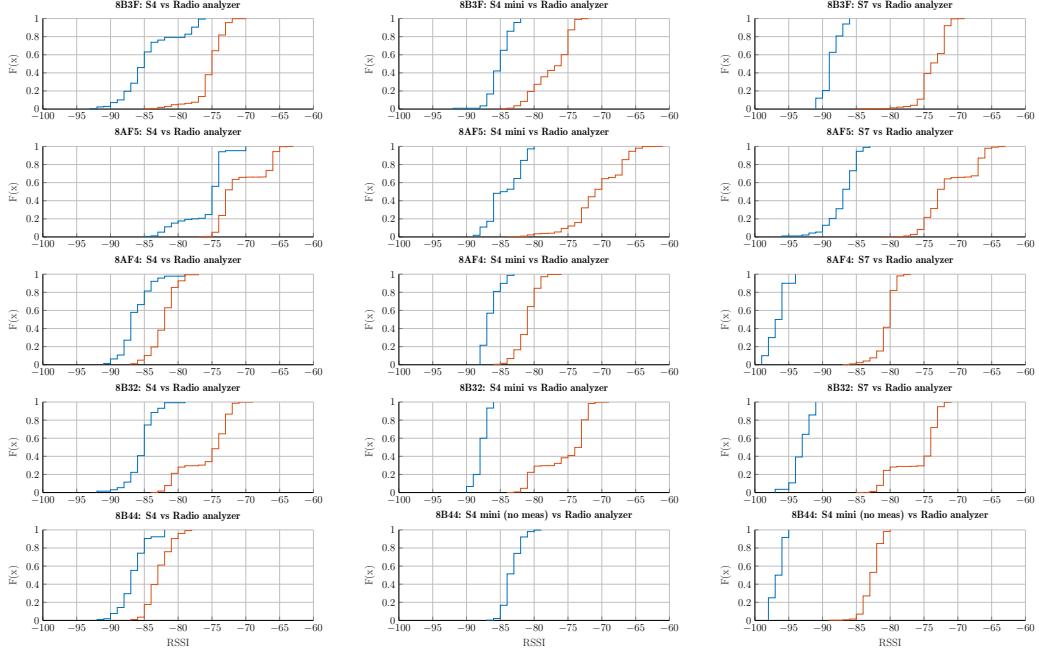


Figure 6.10: Radio analyzer bias from smartphones location 1

#### 6.2.5.5 Comparison with Radio Analyzer

In this experiment we investigate the bias induced by the smartphones in comparison to the radio-analyzer. From the figure 6.10, the bias is evident. The experiment involved putting the radio analyzer and the phone in the standard usage configuration. The results from the Table A.1 claim that the bias could be as worse as 15.6 dBm.

## 6.3 Received Signal Strength Indication: Revisited

As mentioned in chapter 5, we record the signal strength along with their MAC address with absolute and relative time of recording using the android measurement application. For simplicity, we investigate only three closest luminaires with user-free setting. We look at few nearest beacons from Kontakt and Helvar, then plot them against time. We used the smartphones Nexus 6, Samsung S4 and Samsung S4 mini. A detailed description about BLE beacons and its architecture is given in chapter 5.

The figure 6.11 shows a peculiar but periodic characteristic of the RSSI signal for Helvar and Kontakt beacons. It is too systematic to be attributed to the signal fluctuation. This calls for a detailed investigation of the RSSI

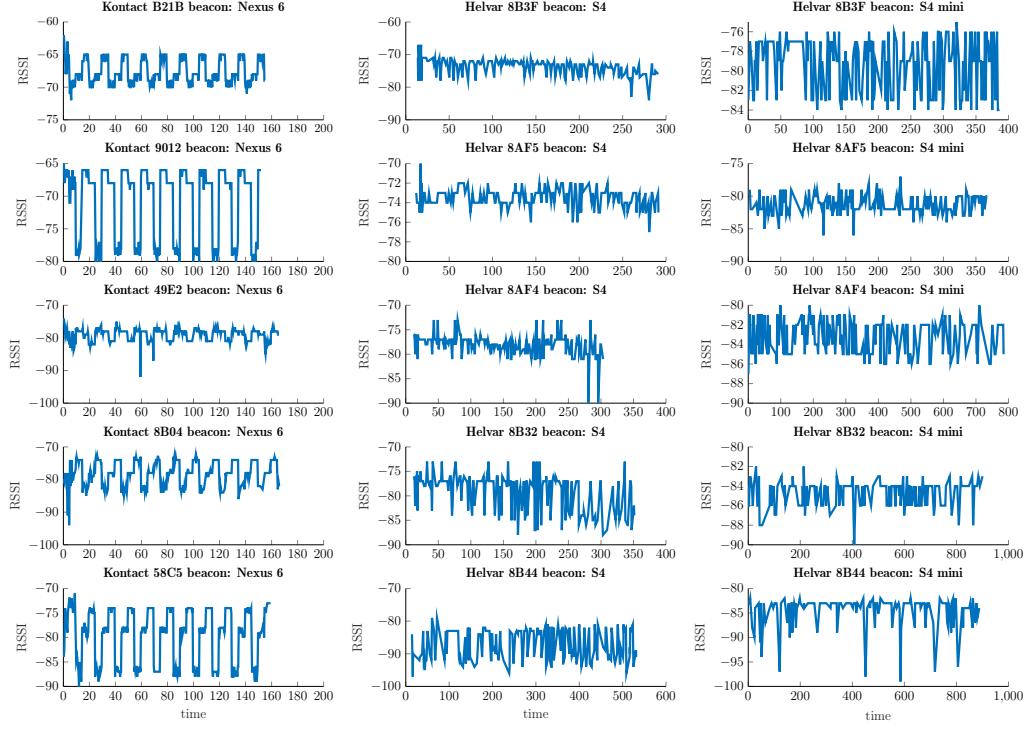


Figure 6.11: The figure illustrates the RSSI versus the time from Kontakt and Helvar beacons. It shows a systematic periodic pattern.

values. In order to avoid all the different environmental factors which could add bias to RSSI values we take the measurements of the beacons in a outdoor setting with the android phone and radio-analyzer.

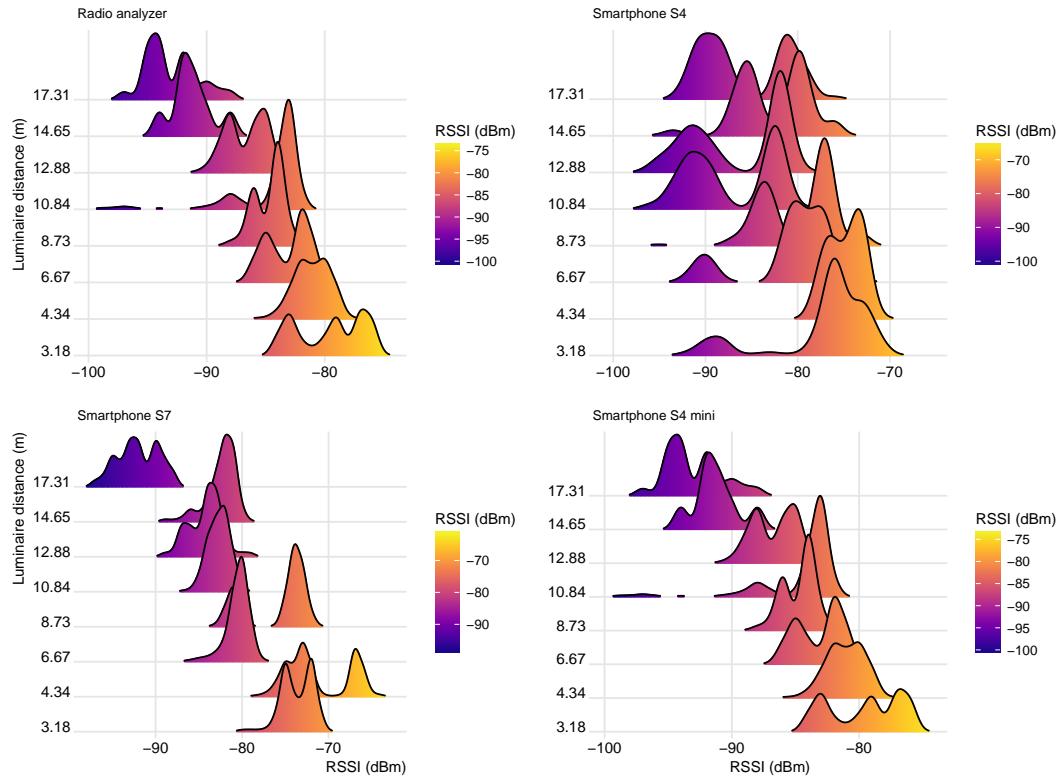


Figure 6.12: The figure shows the density of signal strength as distance increases.

### 6.3.1 Measurement Outdoors with Radio analyzer

To completely eliminate the multipath, interference or any other factors inducing bias in the RSSI, we decided to get user-free measurements in the outdoor environment. We synced the measurements with the radio-analyzer by manually checking the absolute timestamps in both smartphone and the radio analyzer.

From the figure 6.13, it is evident that the periodic trend in the RSSI values is due to the difference in the performance of the channels of the BLE beacon. Due to the mismatch in the performance of the channels the variation in the RSSI values were over 20 dBm. Testing with the different beacons showed that the channels which performed the worse always wasn't same. Now we understand that the multimodality of the RSSI values (refer Figure 6.12) is due to the unequal performance of the advertisement channels. This is the low-power and performance tradeoff.

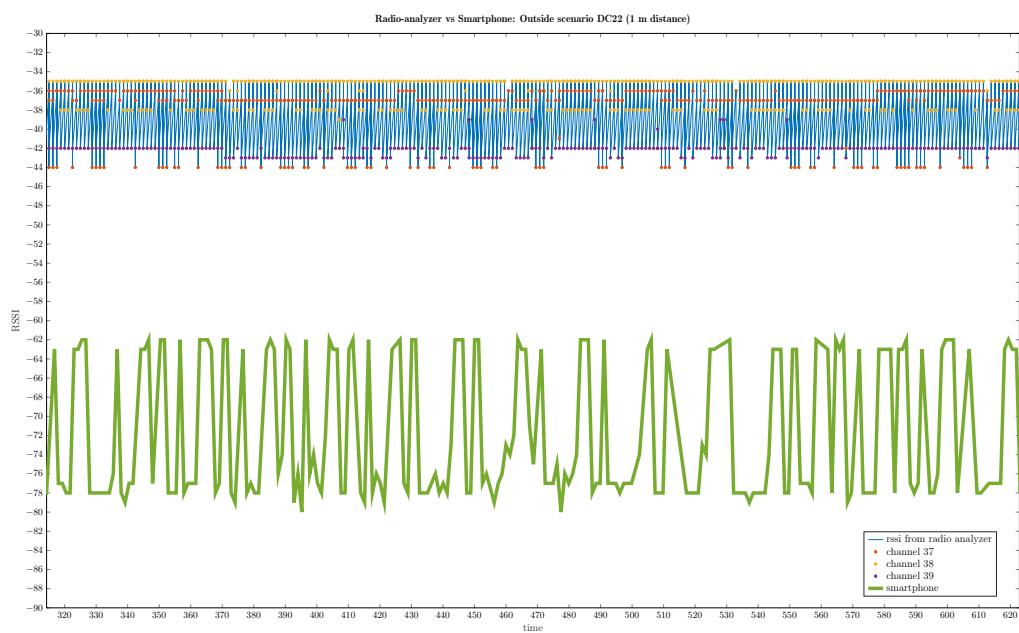


Figure 6.13: Helvar beacon RSSI measurement outside using radio analyzer and smartphone.

# Chapter 7

## Experiments and Results

The algorithms explained in Chapters 4 are simulated using the test data collected from the Helvar R&D (refer Chapter 5). In the section 7.1, we aim to investigate the optimum number of parameter value i.e., calibration points and calibration time for the current experiment test-bed and use them for the positioning evaluation as done in section 7.2. The simulations were implemented using Matlab.

### 7.1 Effect of parameters in calibration phase on positioning

In this section the performance of the non-linear filtering methods are examined. There are two categories of parameters one related to the calibration phase and other related to filtering phase viz-a-viz the filtering methods. Only the calibration phase parameters i.e., calibration points and calibration time are varied. Unlike Au (2010) where the number of AP's was also experimented, we can't experiment the same as the number of luminaires and their embedded BLE beacons are fixed. The RMSE error plotted in the figures is average over 100 Monte Carlo iterations. We refer to optimal parameters inferred from this experiments as *optimal parameters 1*.

#### 7.1.1 Separate parameter evaluation

##### 7.1.1.1 Calibration points

First, we experiment with the increasing number of calibration points keeping the calibration time constant to the maximum. As discussed in the Chapter 5, we have 63 calibration points. We start with 6 points and increase the points,

depending on the location of the space in our testbed. We see an elbow at 15 calibration points from the figure 7.1 for all the algorithms. We also see an unexplained spike at 40 calibration points for UKF-1 using the *radio-map 1* which is not visible in the *radio-map 2*, showing the shortcomings of the *radio-map 1*.

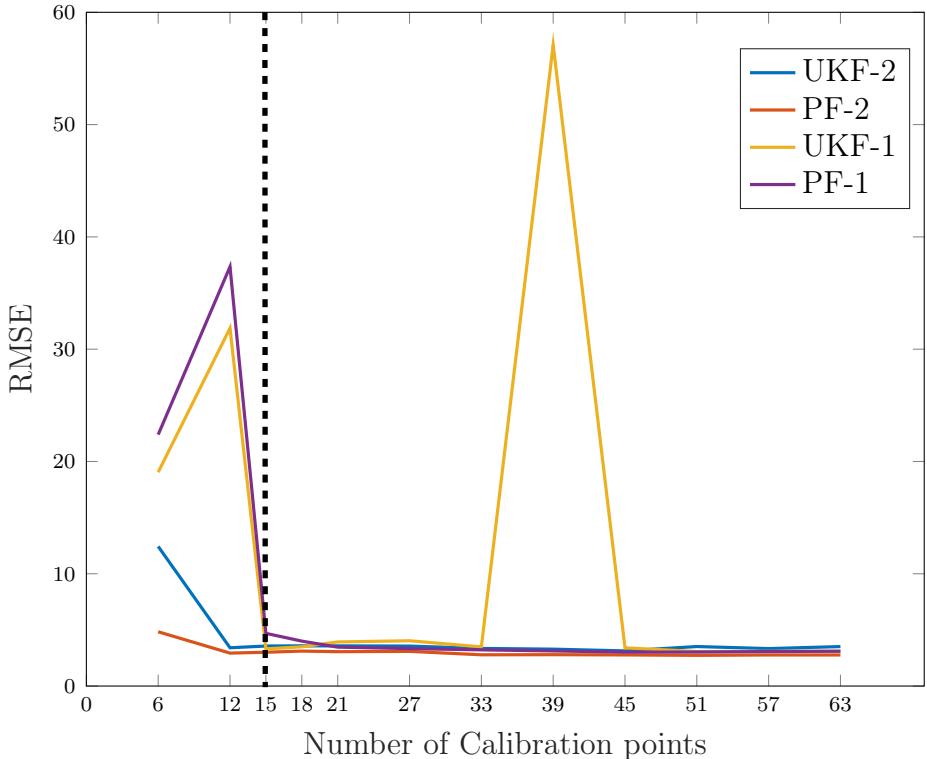


Figure 7.1: Experiment with increasing calibration points for GP-PF and GP-UKF

### 7.1.1.2 Calibration time

Second, we experiment with the increasing number of calibration time keeping calibration points to maximum constant i.e., 63 points. In the experiments we recorded the signal strength data for 50 seconds and the experiments were divided with a gap of 2 seconds each. From the figure 7.2, we infer a clear elbow at 4 secs. As observed in the previous section of calibration points, we see the spike in UKF when used with *radio-map 1*.

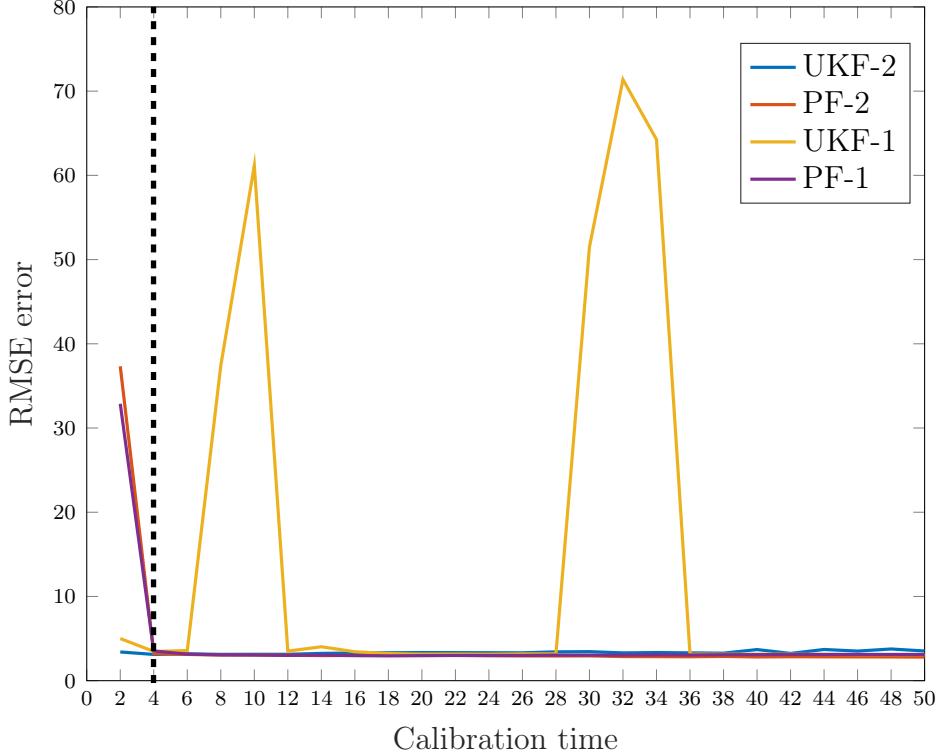


Figure 7.2: Experiment with increasing calibration time for GP-PF and GP-UKF.

### 7.1.2 Combined parameter evaluation

Lastly, we vary both the parameters i.e., calibration points and calibration time. Following the similar strategy as previous two experiments. We can see that the *radio-map 2*, gives stable results than its counter part *radio-map 1*. The results have been displayed in the Figures 7.3 and 7.4. We observe that Figure 7.3a gives best results and we infer that optimal parameters are 12 calibration points and 24 calibration time. We refer to these inferred parameters as *optimal parameters 2*. The spike in the Figures 7.1 and 7.2 can also be seen in the Figure 7.4b.

## 7.2 Filtering location estimation algorithms

The memory based methods described in the Chapter 4 were implemented and simulated in MATLAB. The test data and testbed as discussed in Chap-

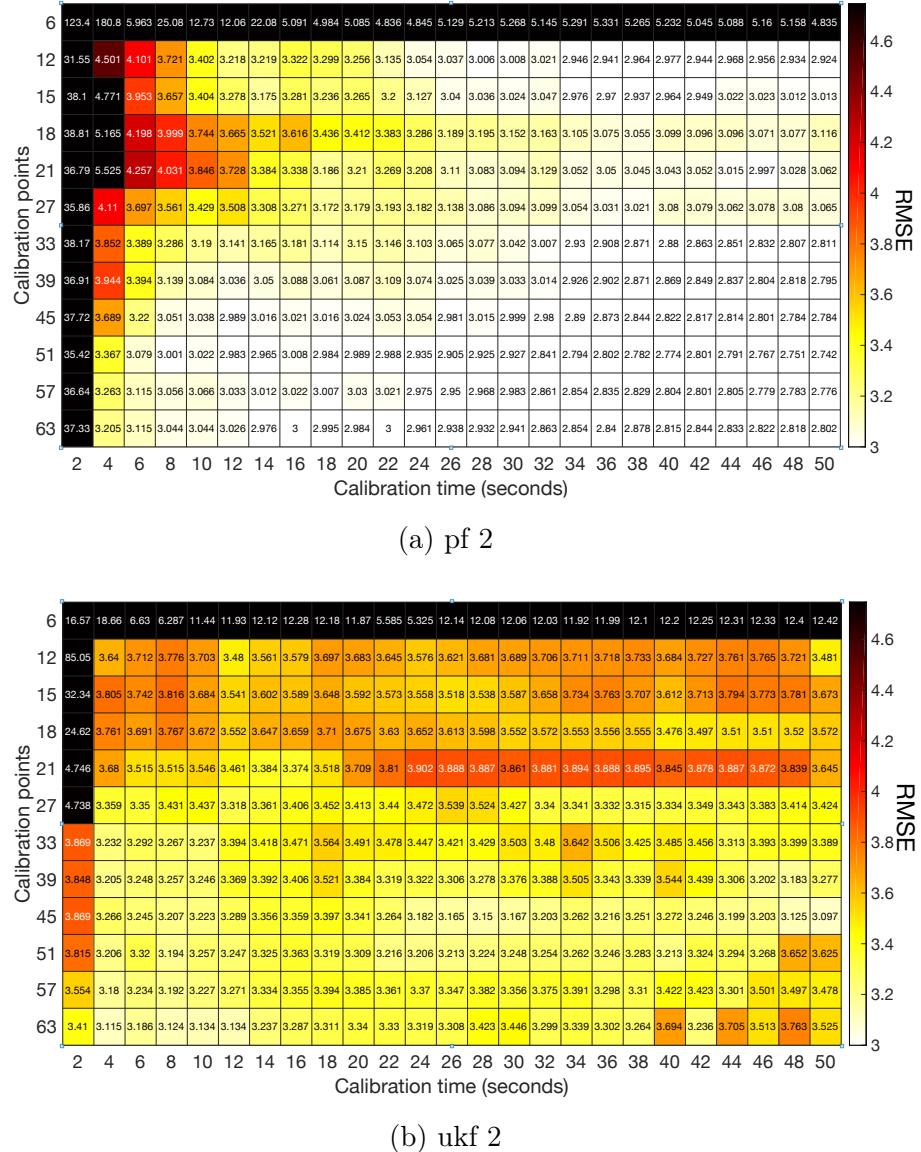


Figure 7.3: radio-map 1

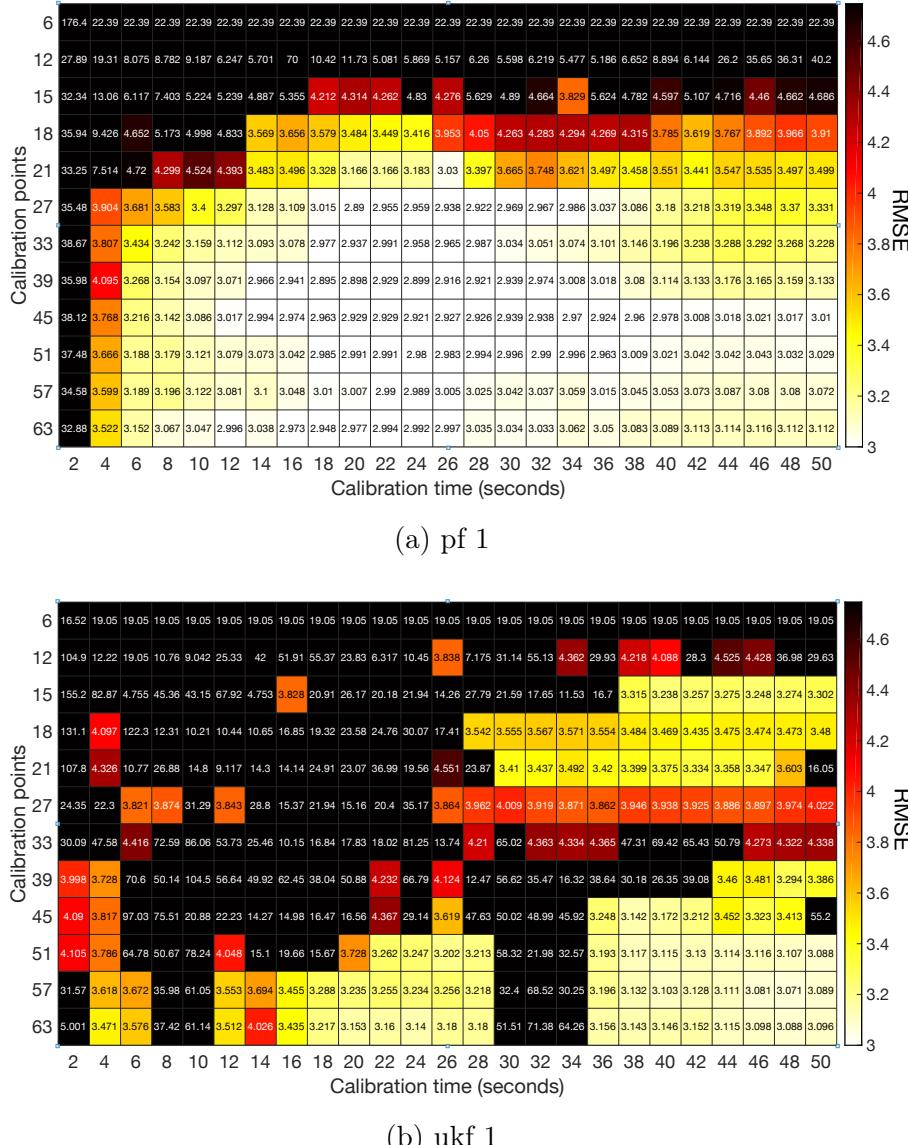


Figure 7.4: radio-map 2

Table 7.1: Summary of performance of the memory based indoor positioning methods with radio-map 1 & 2

Method	RMSE (m)	Mean (m)	90th Percentile (m)	Max (m)	Variance(m <sup>2</sup> )
<i>Optimal Calibration Parameters - 1</i>					
UKF-1	82.8703	62.4112	139.4404	172.1567	2975.7
PF - 1	8.9948	6.6894	16.4693	25.8883	36.1994
UKF - 2	3.8054	3.3731	5.7768	9.4516	3.1066
PF - 2	4.7603	3.9037	7.6029	13.3894	7.4300
<i>Optimal Calibration Parameters - 2</i>					
UKF - 1	10.4546	8.3055	18.4604	22.5404	40.3636
PF - 1	6.0635	4.8835	8.8260	18.4443	12.9323
UKF - 2	3.5764	3.1876	5.5935	8.3233	2.6330
PF - 2	2.9997	2.6962	4.9132	7.1569	1.7305
<i>Full Calibration Parameters</i>					
UKF - 1	3.0958	2.8017	4.6421	6.6037	1.7365
PF - 1	3.1429	2.7994	4.7821	8.9032	2.0434
UKF - 2	3.5248	3.1425	5.2881	9.8346	2.5516
PF - 2	2.8213	2.4834	3.8989	7.7688	1.7947

ter 5 were used for evaluating these methods. The estimation of the location was done at every measurement step using all the 28 BLE AP's. These simulations produced the one-shot position estimate and these methods didn't average the values of the RSSI. The simulations were run in conjunction with the radio-map 1 and radio-map 2 described in the Chapter 2.

### 7.2.1 Summary of the location estimation algorithms

The optimal parameters gathered from the experiments in the Section 7.1 have been simulated with memory based particle and unscented Kalman filters. The results of using the parameters is juxtaposed against the simulation with the *full calibration parameters* setting of 63 calibration points and 50 calibration time in the Table 7.1. The table portrays the absolute error metrics RMSE, mean, 90th percentile, maximum error and variance.

In comparison to radio-map 1, the results of radio-map 2 for the different parameter zones shows significant improvement in terms of performance for both particle filter and unscented Kalman filter. The preserved variance in the reference table 2 actually translates into improved performance.

In terms of the performance of particle filter and unscented Kalman filter for radio-map 2, the particle filter beats unscented Kalman in every simulation except in the case of optimal parameters 1. Given the poor performance of radio-map 1, the results are considered sub-optimal.

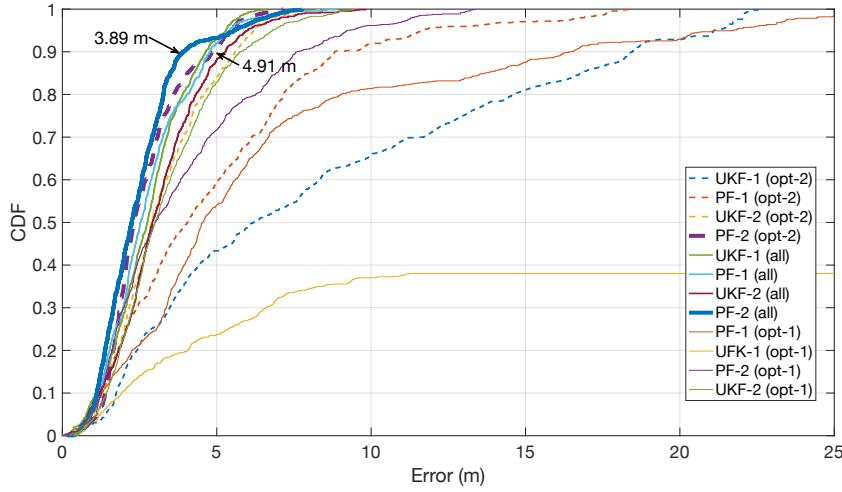


Figure 7.5: Indoor position cumulative error distribution evaluation for particle and unscented Kalman filter using radio-map 1 & 2

We have additionally plotted Figure 7.5, showing the cumulative error for the error from all the evaluated filtering methods with different calibration parameters. The best performance can be seen from PF-2 when all the calibration points and full calibration time is used and it showed 90% of the errors to be under 3.89 meters. This makes intuitive sense as given more information and given the flexibility of the filter, the particle filter would perform better. In terms of least effort in the calibration phase, which could be attributed to *optimal parameters 2*, we again see the PF-2 perform here with 90% of the errors to be under 4.91 meters.



# Chapter 8

## Discussion & Future Work

### 8.1 Discussion

#### 8.1.1 Data Analysis

This study takes a step backwards in understanding the data and in that process we understood that the variation in RSSI is inherent in its architecture and technology. This thesis made an **invention** in understanding the unique characteristic of the BLE devices which shows a distinction in working of the three different channels. The BLE architecture has 40 channels and has three advertisement channels which are strategically placed to avoid interference with the WiFi channels. The advertisement channels are spread over the frequency channels and the performance of the channels vary. The difference in the performance channel affects the signal strength value to var as much as 20 dBm. The variation in the performance of the channel could attributed to that it is low energy device.

With this characteristic of the RSSI signal, we can say that averaging the signal strength values could lead to misleading results. As learnt from the Chapter 6, the distance interpreting non-memory methods if not implemented with caution the prediction could be biased. As seen from the experiments in Chapter 6 the data analysis shows the difference is quite huge when the device near with variance spreading 20 dBm.

The bias due unequal performance of channels could be two fold. One in the construction of the radio map 1. And next in the running of the filter, this effect is evident when we try to combine the signal information from various beacons in the calculation of weights in the particle. This bias confuses the filter into selecting an estimate outside the state space.

### 8.1.2 Radio map

This thesis looked at generation of radiomaps using the GPs. We used  $\mathcal{R}_1$  and  $\mathcal{R}_2$  reference tables, where  $\mathcal{R}_2$  preserved the variance of the RSSI values due to unequal channel performance, unlike avergaing them in  $\mathcal{R}_1$ . The  $\mathcal{R}_2$ , though based on a hunch, proved blessing in disguise for mitigating the unequal channel performance.

The advantages of GP based radiomaps are:

- more flexible spatial distribution as GPs learnt the presence of the wall / obstruction based on the RSSI values and was evident in the radiomaps.
- considerably good accuracy even with less calibration points.

But the usage of non-memory measurement model GP makes the quantization of the RSSI values insignificant, making the area / cell based location estimation methodology irrelevant.

### 8.1.3 Indoor Positioning

In this thesis, we test the memory based methods in the form of Bayesian filtering methods: particle filters and unscented Kalman filters. We experimented to find the optimal calibration parameters, for least manual effort and found that particle filters are the best in the office space setting. These methods add additional capabilities of integrating multiple data information for e.g. inertial sensors and can pave path for a unified solution indoor and outdoor positioning solution.

## 8.2 Future Work

Using the prior information in terms of bias from the channel performance, smart-phones, orientation, material of the luminaire and we were able to incorporate this information in data model of the particle filter.

This study also looked at how the masking of the channel (or multiple channels) in BLE beacons would help in improvement of accuracy of the localization problem. We were in an advantageous position to implement this as our aim was to implement an IP solution which involved BLE beacons of same make. The case would be different if we had to create a single standalone solution for beacons of different make. Additionally, we have the opportunity to incorporate the standards like iBeacon or Eddystone into the in-luminaire beacons.

Incorporate the AA follow strengths and IMU sensors for bettering the accuracy. The BLE beacons in the house could be standardized for adding additional information like type of the luminaire material. How Wi-Fi's wide range and higher TX power affects the RSSI values from the beacons. Rigorously scrutinize the different parameter's that affect the RSSI values and how they fare against the fluctuations of the channels.

Use the truncated singular value decomposition (TSVD) to map the signal strength measure to spatial locations. TSVD encapsulates the dynamics of the indoor environment and our indoor setting fulfills the criteria of dense APs (Lim et al., 2005). Lim et al. (2005) also discusses about zero configuration positioning.

Inertial Navigation systems (Solin et al., 2017)

Aguilar-Garcia et al. (2015) talks about self-optimization and self-healing methods for Self-Organizing networks (SON) applied to cellular networks. These concepts are highly relevant to IPS and potentially could make finger-printing techniques redundant.

Gaussian process with a mean function. !FIXME **cite it** FIXME!

How to the advertisement and reading interval affect the likelihood model.

Malfunction BLE devices.

### **Masking the channels**

Our next challenge was that we either use the prior information in the form of frequency of the channel switching and the extent to which the RSSI value vary or find a way to ignore the RSSI values which bias the prediction of the locations. The BLE technology allows the masking of the channels, hence (or so), we masked one channel i.e., channel 39, and measured the signal strength values from the other two channels.

The measurement model could be tested with Deep Gaussian Processes as the RSSI is highly non-linear.

different dynamic models.



# References

- Addlesee, M., Curwen, R., Hodges, S., Newman, J., Steggles, P., Ward, A., and Hopper, A. (2001). Implementing a sentient computing system. *Computer*, 34(8):50–56.
- Aguilar-Garcia, A., Fortes, S., Molina-García, M., Calle-Sánchez, J., Alonso, J. I., Garrido, A., Fernández-Durán, A., and Barco, R. (2015). Location-aware self-organizing methods in femtocell networks. *Computer Networks*, 93:125–140.
- Ahmad, U. (2015). iBeacon Localization. Master’s thesis, Eindhoven University of Technology, Netherlands.
- Aittola, M., Ryhänen, T., and Ojala, T. (2003). Smartlibrary—location-aware mobile library service. In *International Conference on Mobile Human-Computer Interaction*, pages 411–416. Springer.
- Aravecchia, M. and Messelodi, S. (2014). Gaussian process for RSS-based localisation. In *2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 654–659.
- Atia, M. M., Noureldin, A., and Korenberg, M. J. (2013). Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks. *IEEE Transactions on Mobile Computing*, 12(9):1774–1787.
- Au, A. W. S. (2010). *Rss-based wlan indoor positioning and tracking system using compressive sensing and its implementation on mobile devices*. PhD thesis.
- Bahl, P. and Padmanabhan, V. N. (2000). 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*, volume 2, pages 775–784. IEEE.

- Bahl, P., Padmanabhan, V. N., and Balachandran, A. (2000). A software system for locating mobile users: Design, evaluation, and lessons. *Microsoft Research, MSR-TR-2000-12*.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Chatfield, C. (2006). Initial data analysis. *Encyclopedia of Statistical Sciences*.
- Daniel, W. W. (1999). Nonparametric and distribution-free statistics. *Biostatistics: A foundation for analysis in the health sciences*, pages 596–7.
- Davey, S., Gordon, N., Holland, I., Rutten, M., and Williams, J. (2016). *Bayesian Methods in the Search for MH370*. Springer.
- Djuknic, G. M. and Richton, R. E. (2001). Geolocation and assisted gps. *Computer*, 34(2):123–125.
- Doucet, A., De Freitas, N., and Gordon, N. (2001). Sequential monte carlo methods in practice. series statistics for engineering and information science.
- Eddystone, I. (2016). The new beacon format from google. URL <http://developer.radiusnetworks.com/2015/07/14/introducing-eddystone.html>.
- Elnahrawy, E., Li, X., and Martin, R. P. (2004). The limits of localization using signal strength: A comparative study. In *Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on*, pages 406–414. IEEE.
- Faragher, R. and Harle, R. (2014). An analysis of the accuracy of bluetooth low energy for indoor positioning applications. In *Proceedings of the 27th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ '14)*, pages 201–210.
- Feldmann, S., Kyamakya, K., Zapater, A., and Lue, Z. (2003). An indoor bluetooth-based positioning system: Concept, implementation and experimental evaluation. In *International Conference on Wireless Networks*, pages 109–113.
- Ferris, B., Hahnel, D., and Fox, D. (2007). *Gaussian processes for signal strength-based location estimation*, volume 2, pages 303–310. MIT Press Journals.

- Foley, J. D. (1990). A. vandam, sk feiner, and jf hughes. *Computer Graphics: Principles and Practice*.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2014). *Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science)*. Chapman and Hall/CRC, London, third edition.
- Gomez, C., Oller, J., and Paradells, J. (2012). Overview and evaluation of bluetooth low energy: An emerging low-power wireless technology. *Sensors*, 12(9):11734–11753.
- Gu, Z., Chen, Z., Zhang, Y., Zhu, Y., Lu, M., and Chen, A. (2016). Reducing fingerprint collection for indoor localization. *Computer Communications*, 83:56–63.
- Hansraj, K. (2014). Assessment of stresses in the cervical spine caused by posture and position of the head. 25:277–9.
- Hashemi, H. (1993). The indoor radio propagation channel. *Proceedings of the IEEE*, 81(7):943–968.
- Hazas, M., Scott, J., and Krumm, J. (2004). Location-aware computing comes of age. *Computer*, 37(2):95–97.
- He, S. and Chan, S. H. G. (2016). Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons. *IEEE Communications Surveys Tutorials*, 18(1):466–490.
- Honkavirta, V. (2008). Location Fingerprinting Methods In Wireless Local Area Networks. Master’s thesis, Tampere University of Technology, Finland.
- Huang, J., Millman, D., Quigley, M., Stavens, D., Thrun, S., and Aggarwal, A. (2011). Efficient, generalized indoor wifi graphslam. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1038–1043. IEEE.
- Jardak, N. and Samama, N. (2009). Indoor positioning based on gps-repeaters: performance enhancement using an open code loop architecture. *IEEE Transactions on Aerospace and electronic systems*, 45(1):347–359.

- Julier, S. J. and Uhlmann, J. K. (1997). New extension of the kalman filter to nonlinear systems. In *Signal processing, sensor fusion, and target recognition VI*, volume 3068, pages 182–194. International Society for Optics and Photonics.
- Kaemarungsi, K. and Krishnamurthy, P. (2004). Properties of indoor received signal strength for wlan location fingerprinting. In *Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004. The First Annual International Conference on*, pages 14–23. IEEE.
- Kaemarungsi, K. and Krishnamurthy, P. (2012). Analysis of wlan’s received signal strength indication for indoor location fingerprinting. *Pervasive and mobile computing*, 8(2):292–316.
- Klepal, M., Pesch, D., et al. (2007). Influence of predicted and measured fingerprint on the accuracy of rssi-based indoor location systems. In *Positioning, Navigation and Communication, 2007. WPNC’07. 4th Workshop on*, pages 145–151. IEEE.
- Kong, A. (1992). A note on importance sampling using standardized weights. Technical report no. 348, Department of Statistics, University of Chicago.
- Kriz, P., Maly, F., and Kozel, T. (2016). Improving indoor localization using bluetooth low energy beacons. *Mobile Information Systems*, 2016.
- Langlois, C., Tiku, S., and Pasricha, S. (2017). Indoor localization with smartphones: Harnessing the sensor suite in your pocket. *IEEE Consumer Electronics Magazine*, 6(4):70–80.
- Li, B., Rizos, C., and Lee, H. K. (2005). Utilizing kriging to generate a nlos error correction map for network based mobile positioning. *Positioning*, 1(09):0.
- Lim, H., Kung, L.-C., Hou, J. C., and Luo, H. (2005). Zero-configuration, robust indoor localization: Theory and experimentation. Technical report.
- Lindh, J. (2015). Bluetooth low energy beacons. *Texas Instruments*, page 15.
- Liu, J. S. (2008a). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.
- Liu, J. S. (2008b). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.

- Luo, X., O'Brien, W. J., and Julien, C. L. (2011a). Comparative evaluation of received signal-strength index (rss) based indoor localization techniques for construction jobsites. *Advanced Engineering Informatics*, 25(2):355–363.
- Luo, X., O'Brien, W. J., and Julien, C. L. (2011b). Comparative evaluation of received signal-strength index (rss) based indoor localization techniques for construction jobsites. *Advanced Engineering Informatics*, 25(2):355–363.
- Martino, L., Elvira, V., and Louzada, F. (2016). Alternative effective sample size measures for importance sampling. In *Statistical Signal Processing Workshop (SSP), 2016 IEEE*, pages 1–5. IEEE.
- Martino, L., Elvira, V., and Louzada, F. (2017). Effective sample size for importance sampling based on discrepancy measures. *Signal Processing*, 131:386–401.
- Networks, R. (2017). Altbeacon: Beacon android library.
- Newman, N. (2014). Apple ibeacon technology briefing. *Journal of Direct, Data and Digital Marketing Practice*, 15(3):222–225.
- Patwari, N. and Wilson, J. (2010). Rf sensor networks for device-free localization: Measurements, models, and algorithms. *Proceedings of the IEEE*, 98(11):1961–1973.
- Pivato, P., Palopoli, L., and Petri, D. (2011). Accuracy of rss-based centroid localization algorithms in an indoor environment. *IEEE Transactions on Instrumentation and Measurement*, 60(10):3451–3460.
- Priyantha, N. B. (2005). *The cricket indoor location system*. PhD thesis, Massachusetts Institute of Technology.
- Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.
- Rizos, C., Dempster, A. G., Li, B., and Salter, J. (2007). Indoor positioning techniques based on wireless lan.
- Roos, T., Myllymäki, P., Tirri, H., Misikangas, P., and Sievänen, J. (2002). A probabilistic approach to wlan user location estimation. *International Journal of Wireless Information Networks*, 9(3):155–164.

- Särkkä, S. (2013). *Bayesian filtering and smoothing*, volume 3. Cambridge University Press.
- Schüssel, M. and Pregizer, F. (2015). Coverage gaps in fingerprinting based indoor positioning: The use of hybrid gaussian processes. In *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*, pages 1–9. IEEE.
- Schwaighofer, A., Grigoras, M., Tresp, V., and Hoffmann, C. (2004). Gpps: A gaussian process positioning system for cellular networks. In *Advances in Neural Information Processing Systems*, pages 579–586.
- Seco, F., Plagemann, C., Jiménez, A. R., and Burgard, W. (2010). Improving rfid-based indoor positioning accuracy using gaussian processes. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, pages 1–8. IEEE.
- Solin, A. (2016). *Stochastic Differential Equation Methods for Spatio-Temporal Gaussian Process Regression*. Phd dissertation, Aalto University.
- Solin, A., Cortes, S., Rahtu, E., and Kannala, J. (2017). Inertial odometry on handheld smartphones. *arXiv preprint arXiv:1703.00154*.
- Steggles, P. and Gschwind, S. (2005). The ubisense smart space platform.
- Stone, L. D., Keller, C. M., Kratzke, T. M., Strumpfer, J. P., et al. (2014). Search for the wreckage of air france flight af 447. *Statistical Science*, 29(1):69–80.
- Vehtari, A., Gelman, A., and Gabry, J. (2015). Pareto smoothed importance sampling. *arXiv preprint arXiv:1507.02646*.
- Wang, J., Gao, Q., Wang, H., Chen, H., and Jin, M. (2011). Differential radio map-based robust indoor localization. *EURASIP Journal on Wireless Communications and Networking*, 2011(1):17.
- Yiu, S., Dashti, M., Claussen, H., and Perez-Cruz, F. (2016). Locating user equipments and access points using rssI fingerprints: A gaussian process approach. In *Communications (ICC), 2016 IEEE International Conference on*, pages 1–6. IEEE.
- Yiu, S., Dashti, M., Claussen, H., and Perez-Cruz, F. (2017). Wireless rssI fingerprinting localization. *Signal Processing*, 131:235–244.

- Zanca, G., Zorzi, F., Zanella, A., and Zorzi, M. (2008). Experimental comparison of rss-based localization algorithms for indoor wireless sensor networks. In *Proceedings of the workshop on Real-world wireless sensor networks*, pages 1–5. ACM.
- Zhang, Y., Zhu, Y., Lu, M., and Chen, A. (2013). Using compressive sensing to reduce fingerprint collection for indoor localization. In *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 4540–4545.



# Appendix A

## Appendix

### A.1 Tables for Data Analysis

Table A.1: Summary of smartphone bias to the RSSI measurements using the radio analyser.

RSSI bias in Smartphones: Radio analyzer vs S4							
Access Point	Radio analyzer RSSI (count)	S4 RSSI (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-75.2315 (1326)	-84.4048 (168)	9.1732	0	3.53E-97	4.61E-31	3.60E-94
8AF5	-70.7684 (1956)	-75.7176 (170)	4.9492	1.21E-09	3.84E-69	7.32E-13	1.20E-69
8AF4	-82.1404 (178)	-86.3286 (140)	4.1881	5.73E-09	2.92E-37	2.49E-02	1.92E-39
8B32	-75.8115 (1751)	-85.4 (130)	9.5885	0	1.46E-90	7.05E-27	1.48E-80
8B44	-82.8472 (386)	-86.5905 (105)	3.7433	5.01E-08	1.33E-39	<b>0.3496</b>	2.19E-40
8CD1	-82.8471 (157)	-92.0886 (79)	9.2415	0	1.36E-46	<b>0.3919</b>	1.02E-36
8CA0	-82.8815 (540)	-88.0294 (34)	5.1479	0	1.82E-24	7.26E-08	1.49E-22
<i>Mean</i>	<i>-78.9325</i>	<i>-85.5085</i>	<i>6.576</i>				
RSSI bias in Smartphones: Radio analyzer vs S4 mini							
Access Point	Radio analyzer RSSI (count)	S4 mini RSSI (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-77.3256 (1689)	-85.1228 (114)	7.7972	0	6.14E-81	1.7044e-24	1.79E-71
8AF5	-70.6342 (2452)	-84.2727 (110)	13.6386	0	8.00E-88	2.57E-04	2.58E-70
8AF4	-81.042 (405)	-86.481 (79)	5.439	0	1.94E-50	0.2888	5.60E-45
8B32	-75.5629 (2288)	-87.8 (15)	12.2371	0	2.20E-14	2.41E-09	8.90E-12
8B44	-83.3467 (424)	NA	NA	NA	NA	NA	NA
8CD1	-84.619 (21)	-88.2143 (14)	3.5952	0.0434	7.61E-05	2.73E-05	1.45E-04
8CA0	-82.975 (519)	-89 (1)	-6.025	0	0.0997	NA	8.13E-02
<i>Mean</i>	<i>-79.3579</i>	<i>-86.8151</i>	<i>7.4572</i>				
RSSI bias in Smartphones: Radio analyzer vs S7							
Access Point	Radio analyzer RSSI (count)	S7 RSSI (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-73.6596 (1416)	-88.6988 (83)	15.0392	0	1.41E-70	2.73E-06	1.48E-55
8AF5	-71.3179 (2120)	-86.9892 (93)	15.6713	0	5.54E-80	3.25E-20	2.82E-61
8AF4	-80.5172 (435)	-96.7 (10)	16.1828	<b>0</b>	8.84E-10	0.7497	1.60E-08
8B32	-75.8498 (1977)	-93.0714 (28)	17.2217	0	1.04E-25	2.82E-10	1.12E-20
8B44	-82.7293 (548)	-96.6667 (12)	13.9374	<b>0</b>	1.66E-11	0.38	1.14E-09
8CD1	-84.1875 (64)	NA	NA	NA	NA	NA	NA
8CA0	-82.3946 (299)	NA	NA	NA	NA	NA	NA
<i>Mean</i>	<i>-76.8148</i>	<i>-92.4252 (till 8b44)</i>	<i>15.6104 (till 8b44)</i>				

Table A.2: Summary of user presence bias to the RSSI measurements at L1.

(1) Without User vs User facing north direction							
Access Point	w/o user RSSI (count)	User RSSI (count)	Error	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-77.1349 (645)	-85.8808 (453)	8.7459	0	2.22E-156	6.45E-04	1.50E-103
8AF5	-74.8347 (629)	-83.7002 (487)	8.8655	0	9.49E-203	1.54E-53	3.58E-175
8AF4	-80.8054 (596)	-90.7882 (321)	9.9828	0	2.82E-188	1.21E-21	3.53E-142
8B32	-79.9181 (415)	-91.9278 (291)	12.0098	0	3.05E-134	1.47E-14	1.65E-108
8B44	-87.1852 (351)	-90.8154 (298)	3.6303	1.90E-04	9.57E-29	6.82E-41	2.88E-20
8CD1	-86.3827 (405)	-92.1137 (387)	5.731	7.38E-11	3.61E-50	4.10E-133	8.15E-48
8CA0	-82.6421 (461)	-93.0089 (224)	10.3668	0	7.91E-124	1.58E-28	2.81E-94
Mean	-81.2719	-89.7479	8.476				
(2) Without User vs User facing east direction							
Access Point	w/o user RSSI (count)	User east (count)	Error	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-77.1349 (645)	-90.4377 (345)	13.3028	0	1.33E-139	0.0024	2.02E-118
8AF5	-74.8347 (629)	-87.2577 (388)	12.4231	0	1.55E-210	1.66E-41	6.92E-160
8AF4	-80.8054 (596)	-92.6728 (162)	11.8675	0	6.86E-115	3.47E-08	1.26E-87
8B32	-79.9181 (415)	-92.8242 (165)	12.9062	0	1.93E-99	1.05E-22	8.85E-77
8B44	-87.1852 (351)	-93.7024 (84)	6.5172	4.73E-11	7.88E-30	8.09E-49	2.96E-31
8CD1	-86.3827 (405)	-95.0484 (62)	8.6657	4.44E-16	1.22E-30	1.19E-50	1.18E-30
8CA0	-82.6421 (461)	-96 (4)	13.3579	4.31E-07	2.22E-04	0.0124	5.28E-04
Mean	-81.2719	-92.5633	11.2914				
(3) Without User vs User facing south direction							
Access Point	w/o user RSSI (count)	User south (count)	Error	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-77.1349 (645)	-90.4205 (440)	13.2856	0	7.83E-166	1.62E-05	6.97E-137
8AF5	-74.8347 (629)	-88.0415 (434)	13.2068	0	3.13E-224	3.41E-57	1.02E-170
8AF4	-80.8054 (596)	-92.8765 (170)	12.0711	0	2.76E-119	<b>0.2467</b>	6.51E-91
8B32	-79.9181 (415)	-93.6351 (222)	13.7171	0	1.50E-124	3.95E-34	1.12E-93
8B44	-87.1852 (351)	-94.4821 (56)	7.297	1.96E-10	1.79E-22	3.09E-33	1.83E-25
8CD1	-86.3827 (405)	-95.9697 (33)	9.587	0	3.65E-22	2.71E-34	3.23E-20
8CA0	-82.6421 (461)	-95 (2)	12.3579	0	1.15E-02	0.0115	1.40E-02
Mean	-81.2719	-92.9179	11.646				
(4) Without User vs User facing west direction							
Access Point	w/o user RSSI (count)	User west (count)	Error	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-77.1349 (645)	-84.6492 (553)	7.5143	4.44E-16	2.34E-151	<b>0.7018</b>	8.46E-109
8AF5	-74.8347 (629)	-84.1693 (561)	9.3347	0	1.27E-210	3.79E-62	9.40E-185
8AF4	-80.8054 (596)	-87.7208 (480)	6.9155	0	1.54E-240	5.74E-29	6.97E-181
8B32	-79.9181 (415)	-90.7911 (426)	10.873	0	2.58E-153	3.44E-19	2.64E-129
8B44	-87.1852 (351)	-91.6395 (319)	4.4543	1.58E-06	1.55E-31	5.93E-59	1.88E-29
8CD1	-86.3827 (405)	-93.1667 (252)	6.784	2.04E-12	2.81E-41	1.60E-77	9.79E-51
8CA0	-82.6421 (461)	-92.4114 (299)	9.7693	0	2.27E-70	2.50E-108	1.49E-91
Mean	-81.2719	-89.2211	7.9492				
(5) Without User vs User facing shadowing direction							
Access Point	w/o user RSSI (count)	User shadow (count)	Error	Z-test	KS test	Levene test	Wilcoxon test
8B1D	-70.8425 (546)	-71.1298 (524)	0.2873	0.7705	3.08E-19	1.56E-24	<b>0.2479</b>
8B11	-76.4113 (586)	-83.3536 (509)	6.9424	0	1.10E-168	2.78E-101	3.60E-149
8CAA	-72.7231 (585)	-88.1457 (405)	15.4226	0	3.04E-209	8.22E-31	7.55E-160
8C6A	-79.1505 (651)	-85.2004 (489)	6.0499	1.48E-09	1.63E-49	8.17E-27	1.15E-49
8C39	-78.0507 (670)	-88.5194 (387)	10.4686	0	8.40E-146	1.45E-09	3.68E-148
8C4A	-85.2737 (464)	-90.7217 (357)	5.448	3.50E-11	5.65E-68	4.74E-05	9.97E-64
8C6D	-77.581 (568)	-90.1753 (405)	12.5943	0	3.26E-133	1.04E-13	1.11E-121
Mean	-77.1475	-85.3208	8.1733				
(6) Without User vs User facing rotating direction							
Access Point	w/o user RSSI (count)	User rotate (count)	Error	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-77.1349 (645)	-88.8976 (203)	11.6927	0	3.32E-94	<b>0.3101</b>	3.81E-74
8AF5	-74.8347 (629)	-85.8068 (208)	10.9721	0	6.61E-129	1.15E-78	6.67E-104
8AF4	-80.8054 (596)	-91.1449 (138)	10.3396	0	3.79E-101	2.14E-09	2.38E-77
8B32	-79.9181 (415)	-91.7045 (132)	11.7865	0	8.86E-81	8.53E-10	3.72E-65
8B44	-87.1852 (351)	-92.5375 (80)	5.3523	2.25E-05	1.56E-13	1.68E-26	1.33E-19
8CD1	-86.3827 (405)	-93.3084 (107)	6.9257	1.11E-10	2.32E-26	6.68E-55	4.34E-31
8CA0	-82.6421 (461)	-92.6267 (75)	9.9846	1.96E-10	2.27E-23	5.17E-31	5.67E-31
Mean	-81.2719	-90.8509	9.579				

Table A.3: Summary of user presence bias to the RSSI measurements for L2.

Without User vs User facing north direction							
Access Point	w/o user RSSI (count)	User RSSI (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-74.034 (235)	-73.6262 (650)	0.4079	<b>0.5554</b>	3.47E-28	5.41E-44	3.47E-06
8B33	-87.1358 (243)	-83.7008 (625)	3.435	3.63E-06	2.26E-50	3.93E-56	6.05E-25
8B26	-86.5161 (124)	-89.9404 (554)	3.4243	<b>0.0158</b>	1.06E-42	<b>0.1128</b>	1.86E-26
8B30	-86.0913 (252)	-87.6348 (564)	1.5435	<b>0.023</b>	1.37E-13	6.85E-14	7.78E-08
8CA0	-90.1701 (241)	-91.7809 (324)	1.6107	0.0011	3.30E-32	5.49E-29	5.22E-22
8CD1	-87.8323 (155)	-90.9436 (461)	3.1113	0.0023	1.64E-48	0.0081	9.53E-29
8B44	-93.4215 (121)	-91.6318 (402)	1.7896	0.0035	2.79E-15	4.34E-12	2.54E-12
<i>Mean</i>	<i>-86.4573</i>	<i>-87.0369</i>	<i>2.1889</i>				
Without User vs User facing east direction							
Access Point	w/o user RSSI (count)	User east (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-74.034 (235)	-77.6172 (674)	3.5832	8.76E-08	9.65E-54	4.04E-48	1.75E-26
8B33	-87.1358 (243)	-88.5488 (605)	1.413	<b>0.0227</b>	1.34E-24	7.27E-45	5.91E-08
8B26	-86.5161 (124)	-91.4196 (491)	4.9034	5.42E-04	4.85E-58	<b>0.0238</b>	8.91E-34
8B30	-86.0913 (252)	-89.7476 (519)	3.6563	2.15E-08	2.24E-44	1.67E-10	1.22E-44
8CA0	-90.1701 (241)	-92.7592 (245)	2.5891	2.46E-07	6.26E-50	2.10E-22	8.85E-42
8CD1	-87.8323 (155)	-92.8976 (332)	5.0653	3.31E-07	1.09E-65	<b>0.275</b>	9.28E-43
8B44	-93.4215 (121)	-91.6407 (334)	1.7808	<b>0.0049</b>	4.72E-14	6.59E-13	1.76E-11
<i>Mean</i>	<i>-86.4573</i>	<i>-89.2329</i>	<i>3.2844</i>				
Without User vs User facing south direction							
Access Point	w/o user RSSI (count)	User south (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-74.034 (235)	-76.5601 (666)	2.526	1.33E-04	4.45E-38	8.40E-50	1.15E-12
8B33	-87.1358 (243)	-86.2746 (590)	0.8612	<b>0.2238</b>	1.66E-20	2.19E-52	0.0065
8B26	-86.5161 (124)	-92.1648 (449)	5.6487	7.56E-05	1.62E-57	<b>0.024</b>	2.56E-33
8B30	-86.0913 (252)	-86.9623 (610)	0.871	<b>0.222</b>	3.95E-15	9.34E-25	<b>0.013</b>
8CA0	-90.1701 (241)	-90.7882 (439)	0.618	<b>0.1768</b>	2.71E-20	8.77E-39	1.37E-04
8CD1	-87.8323 (155)	-90.9836 (488)	3.1513	0.0017	2.26E-50	4.09E-02	3.41E-31
8B44	-93.4215 (121)	-91.3266 (346)	2.0949	9.98E-04	7.12E-17	2.13E-13	1.98E-14
<i>Mean</i>	<i>-86.4573</i>	<i>-87.8657</i>	<i>2.2530</i>				
Without User vs User facing west direction							
Access Point	w/o user RSSI (count)	User west (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-74.034 (235)	-76.3689 (656)	2.3349	1.78E-04	1.29E-34	1.38E-32	6.42E-14
8B33	-87.1358 (243)	-86.6918 (623)	0.444	<b>0.4529</b>	1.61E-17	8.98E-40	8.38E-04
8B26	-86.5161 (124)	-90.5759 (514)	4.0597	0.0043	3.07E-49	<b>0.0977</b>	1.65E-30
8B30	-86.0913 (252)	-87.1509 (603)	1.0596	<b>0.1007</b>	1.44E-09	2.27E-10	5.60E-04
8CA0	-90.1701 (241)	-89.2592 (517)	0.9109	7.75E-02	1.12E-19	1.24E-51	6.43E-05
8CD1	-87.8323 (155)	-92.1986 (423)	4.3663	1.08E-05	1.33E-65	<b>0.6872</b>	3.40E-41
8B44	-93.4215 (121)	-91.6941 (340)	1.7274	<b>0.0033</b>	5.79E-14	1.45E-10	<b>7.49E-13</b>
<i>Mean</i>	<i>-86.4573</i>	<i>-87.7056</i>	<i>2.1289</i>				
Without User vs User facing shadowing direction							
Access Point	w/o user RSSI (count)	User shadow (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-74.034 (235)	-79.8323 (650)	5.7983	0	3.23E-85	1.46E-50	2.78E-65
8B33	-87.1358 (243)	-88.931 (536)	1.7952	0.0037	2.32E-28	4.55E-47	2.76E-11
8B26	-86.5161 (124)	-92.9302 (344)	6.4141	5.66E-06	8.62E-60	7.67E-06	2.84E-34
8B30	-86.0913 (252)	-91.508 (374)	5.4168	0	2.12E-82	<b>0.2536</b>	7.96E-78
8CA0	-90.1701 (241)	-93.25 (172)	3.0799	4.50E-09	1.14E-57	3.08E-18	2.53E-47
8CD1	-87.8323 (155)	-89.9053 (433)	2.0731	<b>0.0488</b>	9.36E-36	8.45E-04	3.55E-17
8B44	-93.4215 (121)	-91.9349 (215)	1.4866	<b>0.0325</b>	1.04E-08	2.69E-08	1.62E-08
<i>Mean</i>	<i>-86.4573</i>	<i>-89.756</i>	<i>3.7234</i>				
Without User vs User facing rotating direction							
Access Point	w/o user RSSI (count)	User rotate (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-74.034 (235)	-77.9962 (261)	3.9621	4.04E-04	1.06E-38	7.02E-43	9.13E-21
8B33	-87.1358 (243)	-86.3686 (236)	0.7672	<b>0.4328</b>	1.30E-13	3.16E-46	1.84E-02
8B26	-86.5161 (124)	-90.3661 (183)	3.85	<b>0.0142</b>	2.40E-35	<b>0.2417</b>	1.78E-19
8B30	-86.0913 (252)	-89.0408 (196)	2.9495	0.0041	3.24E-23	1.72E-16	5.38E-17
8CA0	-90.1701 (241)	-91.2 (130)	1.0299	<b>0.2296</b>	4.87E-13	7.37E-33	9.03E-05
8CD1	-87.8323 (155)	-90.6932 (176)	2.8609	<b>0.0189</b>	1.34E-31	1.47E-02	5.10E-18
8B44	-93.4215 (121)	-91.1071 (140)	2.3143	<b>0.0111</b>	1.96E-15	1.03E-13	3.75E-11
<i>Mean</i>	<i>-86.4573</i>	<i>-88.1103</i>	<i>2.5334</i>				

Table A.4: Summary of smartphone bias to the RSSI measurements at L1.

Smartphone: S4 vs S4 mini							
Access Point	S4 RSSI (count)	S4 mini RSSI (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-77.1349 (645)	-79.5686 (503)	2.4337	0.0018	2.02E-65	1.01E-13	2.42E-55
8AF5	-74.8347 (629)	-81.0689 (450)	6.2342	0	5.10E-226	2.91E-16	2.43E-174
8AF4	-80.8054 (596)	-83.1289 (225)	2.3235	8.78E-04	7.72E-54	3.00E-15	9.94E-40
8B32	-79.9181 (415)	-84.7287 (188)	4.8107	9.19E-12	2.86E-50	5.90E-52	7.12E-47
8B44	-87.1852 (351)	-84.4841 (157)	2.7011	0.0159	4.72E-18	3.82E-34	1.11E-04
8CD1	-86.3827 (405)	-86.6853 (197)	0.3026	0.7337	5.49E-33	6.46E-112	0.0653
8CA0	-82.6421 (461)	-91.25 (8)	8.6079	2.42E-05	2.09E-07	0.0016	3.41E-06
<i>Mean</i>	<i>-81.2719</i>	<i>-84.4163</i>	<i>3.1444</i>				
Smartphone: S4 vs Nexus 5							
Access Point	S4 RSSI (count)	Nexus 5 (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-77.1349 (645)	-77.7444 (90)	0.6096	0.7189	1.98E-05	0.3105	0.3761
8AF5	-74.8347 (629)	-76.1156 (467)	1.281	0.004	6.05E-28	3.16E-28	9.87E-19
8AF4	-80.8054 (596)	-83.7955 (44)	2.9901	0.0371	1.43E-13	0.0119	2.07E-09
8B32	-79.9181 (415)	-83.75 (20)	3.8319	0.0052	1.72E-06	6.96E-07	3.32E-05
8B44	-87.1852 (351)	-84.5 (10)	2.6852	0.2907	0.0046	3.28E-07	<b>0.194</b>
8CD1	-86.3827 (405)	-79.7979 (94)	6.5848	4.97E-11	1.18E-40	5.08E-68	4.24E-42
8CA0	-82.6421 (461)	-81.1889 (90)	1.4532	0.2477	2.11E-07	6.13E-06	0.0163
<i>Mean</i>	<i>-81.2719</i>	<i>-80.9846</i>	<i>0.2873</i>				
Smartphone Orientation: S4 mini vs Nexus 5							
Access Point	S4 mini (count)	Nexus 5 (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-76.176 (375)	-77.7444 (90)	1.8241	<b>0.2553</b>	1.02E-19	3.41E-25	3.21E-06
8AF5	-83.8701 (154)	-76.1156 (467)	4.9533	0	9.79E-141	1.75E-51	1.89E-136
8AF4	-84.5294 (102)	-83.7955 (44)	0.6666	<b>0.6237</b>	<b>0.0128</b>	0.0018	<b>0.0535</b>
8B32	-84.6071 (84)	-83.75 (20)	0.9787	<b>0.4326</b>	0.0036	<b>0.1504</b>	0.0025
8B44	-87.303 (33)	-84.5 (10)	0.0159	<b>0.995</b>	<b>0.0818</b>	<b>0.7005</b>	<b>0.1428</b>
8CD1	-87.359 (78)	-79.7979 (94)	6.8874	0.00	3.66E-55	<b>0.0156</b>	8.32E-42
8CA0	-86.2576 (66)	-81.1889 (90)	10.0611	8.43E-06	4.40E-06	<b>0.2322</b>	3.99E-06
<i>Mean</i>	<i>-84.3003</i>	<i>-80.9846</i>	<i>3.6267</i>				

Table A.5: Summary of smartphone bias to the RSSI measurements at L2.

Smartphone: S4 vs S4 mini							
Access Point	S4 RSSI (count)	S4 mini RSSI (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-67.7138 (622)	-67.5916 (404)	0.1222	<b>0.9252</b>	1.27E-10	2.04E-08	<b>0.1851</b>
8B33	-77.5363 (606)	-76.5158 (380)	1.0205	6.16E-05	4.26E-36	<b>0.146</b>	7.22E-53
8B26	-78.4482 (589)	-80.1535 (443)	1.7053	<b>0.0403</b>	1.01E-07	5.90E-27	1.88E-06
8B30	-72.8802 (668)	-76.6959 (467)	3.8157	3.47E-11	5.20E-88	1.12E-31	1.07E-80
8CA0	-80.6734 (646)	-82.7034 (381)	2.03	2.73E-05	4.27E-53	4.60E-61	5.42E-21
8CD1	-79.7937 (630)	-85.7704 (318)	5.9768	0	2.41E-173	<b>0.1582</b>	3.81E-140
8B44	-81.6128 (545)	-83.8438 (256)	2.2309	1.43E-07	4.54E-36	5.89E-199	4.04E-51
<i>Mean</i>	<i>-76.9512</i>	<i>-79.0392</i>	<i>2.4144</i>				
Smartphone: S4 vs Nexus 5							
Access Point	S4 RSSI (count)	Nexus 5 (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-67.7138 (622)	-73.5583 (609)	5.8445	3.14E-11	5.13E-79	0	1.93E-14
8B33	-77.5363 (606)	-78.2458 (419)	0.7095	<b>0.1255</b>	8.49E-39	2.19E-58	1.45E-14
8B26	-78.4482 (589)	-79.303 (330)	0.8548	<b>0.1335</b>	1.04E-51	9.89E-11	1.93E-42
8B30	-72.8802 (668)	-84.5116 (43)	11.6314	0	1.31E-28	4.59E-04	4.71E-28
8CA0	-80.6734 (646)	-82.7299 (211)	2.0565	2.01E-05	1.51E-26	3.78E-111	1.56E-15
8CD1	-79.7937 (630)	-82.5678 (273)	2.7741	2.22E-16	1.16E-110	1.82E-06	7.25E-85
8B44	-81.6128 (545)	-90.1905 (21)	8.5776	2.89E-15	3.90E-19	1.64E-26	1.07E-17
<i>Mean</i>	<i>-76.9512</i>	<i>-81.5867</i>	<i>4.6355</i>				
Smartphone Orientation: S4 mini vs Nexus 5							
Access Point	S4 mini (count)	Nexus 5 (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-67.5916 (404)	-73.5583 (609)	5.9667	9.96E-09	2.39E-98	6.01E-117	1.95E-34
8B33	-76.5158 (380)	-78.2458 (419)	1.73	2.92E-04	3.45E-56	7.81E-40	1.76E-28
8B26	-80.1535 (443)	-79.303 (330)	0.8505	<b>0.2637</b>	5.09E-24	6.01E-76	8.88E-08
8B30	-76.6959 (467)	-84.5116 (43)	7.8157	1.87E-11	6.61E-31	<b>0.4783</b>	5.29E-27
8CA0	-82.7034 (381)	-82.7299 (211)	0.0264	<b>0.9552</b>	1.08E-06	5.89E-12	0.0164
8CD1	-85.7704 (318)	-82.5678 (273)	3.2027	0	2.88E-88	0.0024	41.51E-85
8B44	-83.8438 (256)	-90.1905 (21)	6.3467	4.39E-08	2.86E-18	1.08E-04	4.04E-15
<i>Mean</i>	<i>-79.0392</i>	<i>-81.5867</i>	<i>2.5475</i>				

Table A.6: Summary of smartphone orientation bias to the RSSI measurements at L1.

(1) Smartphone Orientation: 0°vs 45°							
Access Point	0°RSSI (count)	45°RSSI (count)	Error	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-85.0298 (369)	-78.6019 (422)	6.4279	4.13E-12	7.74E-88	1.76E-48	1.88E-55
8AF5	-79.4463 (419)	-80.1002 (499)	0.6539	<b>0.1006</b>	1.96E-06	1.30E-10	2.13E-04
8AF4	-84.0418 (287)	-85.2217 (221)	1.1799	<b>0.0538</b>	6.66E-16	1.93E-04	2.97E-12
8B32	-86.9134 (254)	-85.3123 (285)	1.6011	0.0053	2.03E-29,	2.04E-29	3.76E-10
8B44	-87.7042 (144)	-82.7621 (269)	4.9421	1.11E-15	4.44E-59	9.14E-09	5.61E-58
8CD1	-91.3404 (47)	-86.1193 (109)	5.2211	4.89E-15	3.39E-28	0.0416	7.59E-24
8CA0	-89.3506 (77)	-88.5 (2)	0.8506	0.8615	0.9132	0.1212	0.8718
<i>Mean</i>	<i>-86.2609</i>	<i>-83.8025</i>	<i>2.4584</i>				
(2) Smartphone Orientation: 0°vs 90°							
Access Point	0°RSSI (count)	90°(count)	Error	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-85.0298 (369)	-77.2751 (458)	7.7547	0.00	2.62E-173	3.32E-70	7.48E-133
8AF5	-79.4463 (419)	-76.9649 (484)	2.4814	1.57E-07	3.33E-66	9.16E-40	8.73E-43
8AF4	-84.0418 (287)	-84.189 (328)	0.1472	7.38E-01	1.54E-11	5.73E-25	0.027
8B32	-86.9134 (254)	-84.1456 (309)	2.7678	4.19E-06	2.98E-29	1.52E-43	3.05E-25
8B44	-87.7042 (144)	-85.7639 (144)	1.9403	0.0048	2.70E-28	0.4608	1.06E-19
8CD1	-91.3404 (47)	NA (0)	NA	NA	NA	NA	NA
8CA0	-89.3506 (77)	-90 (2)	0.6494	0.5661	0.2337	0.0016	0.4691
<i>Mean</i>	<i>-86.2609</i>	<i>-83.0564</i>	<i>3.2045</i>				
(3) Smartphone Orientation: Hand vs Pocket							
Access Point	Hand (count)	Pocket (count)	Error	Z-test	KS test	Levene test	Wilcoxon test
8B3F	-81.4135 (445)	-81.5547 (375)	0.1412	<b>0.8871</b>	<b>0.1043</b>	<b>0.3591</b>	<b>0.8251</b>
8AF5	-80.5614 (399)	-81.9297 (256)	1.3683	<b>0.193</b>	4.06E-07	4.31E-04	2.06E-06
8AF4	-85.5031 (161)	-85.4275 (131)	0.0756	<b>0.9505</b>	<b>0.386</b>	6.64E-04	<b>0.5429</b>
8B32	-86.0152 (197)	-88.1098 (82)	2.0945	<b>0.0735</b>	9.15E-08	<b>0.9584</b>	1.65E-09
8B44	-87.8684 (38)	-86.1622 (74)	1.7063	<b>0.3853</b>	<b>0.0476</b>	<b>0.3643</b>	0.0066
8CD1	-90.0795 (88)	-93.2083 (24)	3.1288	<b>0.25</b>	0.0062	0.0082	5.38E-04
8CA0	-91.9394 (66)	-95.1111 (9)	3.1717	<b>0.3534</b>	<b>0.0712</b>	<b>0.0417</b>	0.0258
<i>Mean</i>	<i>-86.1972</i>	<i>-87.3576</i>	<i>1.1604</i>				

Table A.7: Summary of smartphone orientation bias to the RSSI measurements at L2.

Smartphone Orientation: 0° vs 45°							
Access Point	0°RSSI (count)	45°RSSI (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-66.1402 (428)	-67.209 (421)	1.0688	<b>0.2375</b>	2.70E-25	<b>0.2359</b>	1.58E-09
8B33	-80.6326 (430)	-77.0043 (460)	3.6282	5.08E-10	7.32E-100	6.54E-137	1.41E-30
8B26	-81.531 (226)	-83.5867 (225)	2.0557	<b>0.0213</b>	5.85E-26	4.36E-06	3.29E-20
8B30	-79.5641 (390)	-75.9783 (415)	3.5858	1.04E-06	1.36E-57	1.99E-53	1.07E-36
8CA0	-84.8239 (159)	-85.0227 (220)	0.1988	<b>0.7729</b>	1.29E-09	6.97E-27	<b>0.2264</b>
8CD1	-88.9256 (121)	-86.2967 (246)	2.6289	2.10E-06	1.10E-41	<b>0.5421</b>	8.14E-31
8B44	-84.4265 (272)	-84.6272 (228)	0.2007	<b>0.7653</b>	4.40E-06	2.17E-15	<b>0.5018</b>
<i>Mean</i>	<i>-80.8634</i>	<i>-79.9607</i>	<i>0.9027</i>				
Smartphone Orientation: 0° vs 90°							
Access Point	0°RSSI (count)	90°(count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-66.1402 (428)	-63.6649 (370)	2.4753	0.0027	8.86E-21	1.06E-17	1.84E-10
8B33	-80.6326 (430)	-75.259 (417)	5.3736	6.77E-13	1.20E-92	1.01E-05	3.54E-55
8B26	-81.531 (226)	-84.3929 (168)	2.8619	0.0025	2.69E-22	1.94E-07	9.97E-15
8B30	-79.5641 (390)	-77.3325 (382)	2.2316	0.0158	1.67E-20	<b>0.1976</b>	6.88E-11
8CA0	-84.8239 (159)	-82.5994 (362)	2.2245	9.42E-05	7.01E-20	7.40E-30	3.00E-13
8CD1	-88.9256 (121)	-84.32 (90)	4.6056	0	2.34E-66	<b>0.0012</b>	2.22E-57
8B44	-84.4265 (272)	-82.9281 (167)	1.4983	0.0811	4.35E-12	0.0062	2.55E-06
<i>Mean</i>	<i>-80.8634</i>	<i>-78.6424</i>	<i>3.0386</i>				
Smartphone Orientation: Hand vs Pocket							
Access Point	Hand (count)	Pocket (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
8B14	-76.176 (375)	-77.9639 (360)	1.7879	<b>0.0991</b>	1.36E-06	<b>0.0317</b>	2.81E-07
8B33	-83.8701 (154)	-81.3208 (293)	2.5493	<b>0.0457</b>	1.02E-07	<b>0.271</b>	1.80E-10
8B26	-84.5294 (102)	-82.728 (250)	1.8014	<b>0.2106</b>	3.71E-06	1.10E-11	1.40E-04
8B30	-84.6071 (84)	-85.4365 (197)	0.8294	<b>0.4949</b>	0.0228	<b>0.0273</b>	<b>0.0181</b>
8CA0	-87.303 (33)	-85.4865 (111)	1.8165	<b>0.384</b>	0.0072	0.0014	0.0034
8CD1	-87.359 (78)	-84.7133 (150)	2.6456	<b>0.158</b>	1.37E-04	<b>0.0537</b>	7.89E-06
8B44	-86.2576 (66)	-87.831 (71)	1.5734	0.3467	9.49E-04	<b>0.1321</b>	8.23E-05
<i>Mean</i>	<i>-84.3003</i>	<i>-83.64</i>	<i>2.221</i>				

Table A.8: Summary of luminaire material bias to the RSSI measurements.

Experiment 1							
Distance	Metal + plastic RSSI (count)	Plastic RSSI (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
1.84074	-71.8418 (392)	-72.5564 (532)	0.7146	<b>0.0359</b>	6.90E-24	9.50E-07	1.09E-10
3.68148	-84.4118 (272)	-74.3834 (493)	10.0284	0	2.18E-154	<b>0.1446</b>	3.52E-118
5.52222	-83.25 (392)	-74.9864 (368)	8.2636	0	2.29E-161	1.53E-112	8.51E-126
7.36296	-85.8042 (240)	-85.529 (293)	0.2752	<b>0.5922</b>	2.36E-12	5.38E-29	0.0013
9.2037	-82.9 (230)	-80.1208 (356)	2.7792	0.0011	5.20E-28	3.13E-06	7.61E-24
11.12811	-83.6786 (224)	-81.5825 (424)	2.096	1.64E-04	8.12E-34	3.93E-14	1.09E-34
13.05252	-89.3429 (35)	-75.3946 (185)	13.9483	0	3.18E-24	<b>0.3137</b>	2.45E-22
<i>Mean</i>	<i>-83.0327</i>	<i>-77.7933</i>	<i>5.2394</i>				
Experiment 2							
Distance	Metal + plastic RSSI (count)	Plastic RSSI (count)	MAD	Z-test	KS test	Levene test	Wilcoxon test
1.84074	-72.0806 (434)	-66.9766 (471)	5.104	2.85E-07	5.93E-83	5.22E-36	7.01E-40
3.68148	-74.233 (279)	-75.1373 (437)	0.9043	<b>0.2879</b>	5.25E-31	6.46E-56	1.46E-11
5.52222	-77.8531 (422)	-80.5909 (506)	2.7378	5.68E-07	3.62E-77	3.61E-42	1.74E-41
7.36296	-83.8486 (370)	-82.3724 (427)	1.4763	0.0337	1.20E-18	5.28E-30	1.70E-05
9.2037	-83.7517 (286)	-79.6709 (471)	4.0808	3.19E-11	6.70E-43	3.46E-08	3.51E-61
11.12811	-85.975 (120)	-84.9881 (420)	0.9869	<b>0.0598</b>	1.47E-20	1.08E-06	1.89E-13
13.05252	-94 (37)	-82.9429 (105)	11.0571	0	1.73E-25	6.13E-05	3.70E-21
<i>Mean</i>	<i>-81.6774</i>	<i>-78.9542</i>	<i>2.7232</i>				